RECAP15
Re-thinking the Efficacy of International Climate Change Agreements Post COP15

Financial support from the German Federal Ministry of Education and Research within its programme on 'Economics of Climate Change' is gratefully acknowledged (Grant No 01LA1139A)

www.europa-uni.de/recap15

Discussion Paper Series recap15

No 24 - November 2016:

Spatio-temporal statistical assessment of anthropogenic CO2 emissions from satellite data

Patrick Vetter, Wolfgang Schmid, Reimund Schwarze

Contact:

Patrick Vetter
Department of Statistics, European University Viadrina, Frankfurt (Oder), Germany

vetter@europa-uni.de

Spatio-temporal statistical assessment of anthropogenic CO_2 emissions from satellite data

Patrick Vetter*
Department of Statistics
European University Viadrina, Frankfurt (Oder), Germany

Wolfgang Schmid
Department of Statistics
European University Viadrina, Frankfurt (Oder), Germany

Reimund Schwarze
Department of International Environmental Economics
European University Viadrina, Frankfurt (Oder), Germany

Abstract

The analysis of sources and sinks of CO_2 is a dominant topic in diverse research fields and in political debates these days. The threat of climate change fosters the research efforts in the natural sciences in order to quantify the carbon sequestration potential of the terrestrial ecosystem and CO_2 mitigation negotiations strengthens the need for a transparent, consistent and verifiable Monitoring, Verification and Reporting infrastructure. This paper provides a spatio-temporal statistical modeling framework, which allows for a quantification of the Net Ecosystem Production and of anthropogenic sources, based on satellite data for surface CO_2 concentrations and source and sink connected covariates. Using spatial and temporal latent random effects, that act as space-time varying coefficients, the complex dependence structure can be modeled adequately. Finally, spatio-temporal smoothed estimates for the sources and sinks can be used to provide dynamic maps on $0.5° \times 0.5°$ grid for the Eurasien area in intervals of 16 days between September 2009 and August 2012. Finally, the self-reported CO_2 emissions within the UNFCCC can be compared with the model results.

Keywords: Anthropogenic CO_2 emissions, Net Ecosystem Production, Linear mixed effects, Spatio-temporal model

*Corresponding author. *E-mail address:* vetter@euv-frankfurt-o.de

1 Introduction

The increasing concentration of the greenhouse gas CO_2 in the earth atmosphere has been identified as one of the main drivers of climate change. Likewise, the study of sources and sinks of CO_2 has attracted much interest in many research disciplines and in political debates on climate change and CO_2 emission mitigation. Current efforts to limit and reduce fossil CO_2 emissions, whether they are voluntary or part of international agreements, are self-reported statistical data by emitters. They are used to define baselines and assess the effectiveness of climate and energy policies over time. Self-reported inventories of CO_2 emissions are primarily based on energy-use statistics collected for different sectors. They offer limited transparency and their accuracy and completeness cannot be assessed independently. In order to monitor the effectiveness of an international climate agreement, self-reported emissions data will need to be independently assessed for their accuracy and reliability. Besides the comparably small anthropogenic CO_2 source, the terrestrial ecosystem is responsible for the majority of CO_2 fluxes between the surface and the atmosphere. The carbon sequestration potential of the terrestrial ecosystem is therefore another important topic in the recent research, that can be monitored through the Net Ecosystem Exchange (NEE), which captures the net carbon flux between an ecosystem and the atmosphere [Watson et al., 2000]. Several studies deal with the decomposition of NEE into its components (e.g. Reichstein et al. [2005] and an overview is given in Stoy et al. [2006]), others rather focus on particular regional ecosystems (e.g. Suyker and Verma [2001] and Saigusa et al. [2002]). Another branch of literature is focused on the integration of ecosystem models with remotely-sensed data, as the Normalized Difference Vegetation Index (NDVI) (e.g. Veroustraete et al. [1996]).

The aim of the present article is multifold. Based on satellite data on surface CO_2 concentrations a spatio-temporal statistical model is used to perform spatio-temporal smoothing and predictions. More importantly, sources and sinks of CO_2 are quantified by exploiting the spatio-temporal dependence structure between the CO_2 data and covariates, representing common sources and sinks of CO_2. Using spatial and temporal latent random variables, as space-time varying coefficients for the covariates, anthropogenic and ecosystem emissions can be derived. The model is able to capture the seasonal behavior and the spatial distribution of CO_2 and its sources and sinks, such as the Net Ecosystem Production and anthropogenic emissions. Finally, the inferred anthropogenic emissions allow for a comparison with the self-reported data of the emitters.

Spatio-temporal statistical models have become very popular in environmental studies and a lot of model classes have been developed in this direction, whereas an overview can be found in Cressie and Wikle [2011]. Through the technological progress in the last decades, environmental data can be collected from space. However, in that way the amount of data becomes huge and statistical applications are very high-dimensional. Despite the technological development in computing power, computational

bottlenecks are often encountered. Therefore recent efforts were focused on the development of efficient statistical models, that are able to handle large data sets. Prominent approaches here are dimension reduced models (e.g. González-Manteiga et al. [2012], Cressie et al. [2010] and Katzfuss and Cressie [2011]), covariance tapering [Furrer et al., 2006] and the full-scale approximation [Zhang et al., 2013], as a combination of both previous approaches. For a comparison study of these approaches with an efficiency evaluation, the reader is referred to Vetter et al. [2014]. A recent development are dynamic spatio-temporal models including latent variables. In a state-space framework rapid computations can be achieved through the application of the Kalman filter. Examples can be found in Mardia et al. [1998], Wikle and Cressie [1999], Farrell and Ioannou [2001], Cressie and Wikle [2002], Wikle and Hooten [2006] and Voutilainen et al. [2007]. A combination of all the aforementioned large data approaches is the spatio-temporal linear mixed effects model with space-time varying coefficients outlined in Fassò and Finazzi [2010, 2011, 2013], whereas a Matlab implementation can be found on `http://code.google.com/p/d-stem/`.

The present article is structured as follows. In section 2, the theoretical background, related to the global CO_2 cycle and the statistical model, is outlined. Furthermore the utilized data is presented and an exploratory data analysis is conducted. Section 3 defines the model used for the analysis of the surface CO_2 concentrations and a discussion of the estimated parameters follows. Using the estimated model, the inferred CO_2 sources and sinks are analyzed and discussed. A thorough investigation of the estimated anthropogenic fluxes is done in section 4, which also comprises a comparison of the model results with the CO_2 emissions obtained from the self-reports agreed on in international agreements ([UNFCCC, 2015]) and from scientific estimates based on energy-use statistics ([Olivier et al., 2015]).

2 Spatio-temporal statistical analysis of the global CO_2 cycle

The main contribution of the present article is to analyze the CO_2 cycle and its sources and sinks and to provide an alternative approach to the quantification of global anthropogenic CO_2 emissions based on a state-of-the-art spatio-temporal statistical model applied to satellite data. In that way, the spatial and temporal distribution of ecosystem and anthropogenic sources and sinks of CO_2 are inferred from data on surface CO_2 concentrations. This section is structured as follows. First a brief description of the global CO_2 cycle is given in section 2.1. Subsequently, the spatio-temporal data sets, which consist of remotely sensed satellite observations, are discussed in section 2.2 and an exploratory data analysis follows in section 2.3. Finally the statistical methodology is outlined in section 2.4.

2.1 The global carbon cycle

Local surface CO_2 concentrations represent the outcome of a complex system of carbon exchange between sources and sinks at the ground level and the atmosphere. In order to derive the spatial and temporal distribution of CO_2 sources and sinks from data on local concentrations it is necessary to identify the drivers in the global carbon cycle. As Vetter et al. [2015] point out, the largest carbon fluxes result from the terrestrial ecosystem. According to Prentice et al. [2001], the driving processes here are photosynthesis, autotrophic and heterotrophic respiration and the removal of living biomass (e.g. through fires or harvest). The amount of carbon converted during photosynthesis is called Gross Primary Production (GPP). Since around half of the fixed carbon is used and released back into the atmosphere by the plant, the Net Primary Production (NPP) is the difference between GPP and autotrophic respiration. Furthermore, dead organic matter provides another source of CO_2, since it is decomposed by organisms in the soil (heterotrophic respiration). Reducing the NPP by the heterotrophic respiration results in the Net Ecosystem Production or Exchange (NEP/NEE), which represents the total carbon budget for a certain local vegetation. Finally, after accounting for losses of carbon stock through fires and the removal of vegetation through mankind yields the Net Biome Production. In a long term equilibrium system, this budget should be balanced. However there are impact factors for the carbon cycle, which lead to unbalances, as [Prentice et al., 2001][p. 186] state. Another major driver in the global carbon cycle are ocean carbon processes. Oceans are large carbon pools and through dissolution processes they are able to take up a large fraction of anthropogenic CO_2. In the absence of the oceanic system, atmospheric CO_2 concentrations would be around 200 ppm higher [Prentice et al., 2001][p. 198]. However, since this article examines CO_2 source/sink behavior at the surface on land area, the interested reader is referred to Prentice et al. [2001] for a thorough explanation of the carbon exchange mechanisms in the oceans. The remaining source of CO_2 is the human influence, which is given a special focus in this article. The majority of anthropogenic emissions result from burning fossil fuels. With a share of 67.4% in 2013, fossil fuels are the dominant source of energy [OECD/IEA, 2015][p. 24] and mainly contribute to the increase in atmospheric CO_2 concentrations. Another anthropogenic CO_2 source results from human's impact on terrestrial ecosystems through changes in land-use, especially deforestation. These types of anthropogenic emissions make up a share of 10-30% of total emissions [Prentice et al., 2001][p. 204].

2.2 Satellite Data

The present study is based on remotely sensed data from satellites. Observations of surface CO_2 concentrations are used and linked to data sets related to common sources and sinks of CO_2. The region of interest of this study is bounded between $30°N$ and $60°N$ latitude and $10°W$ and $150°E$

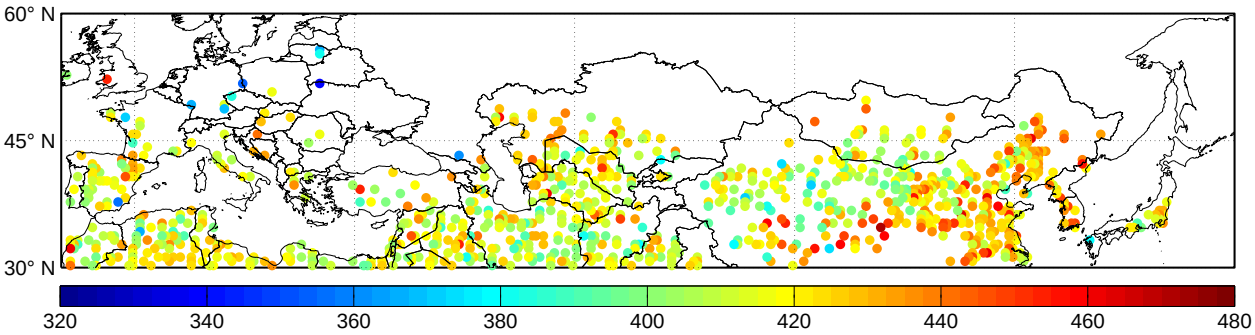

Figure 1: ACOS-GOSAT surface CO_2 concentrations for the $3^{rd} - 19^{th}$ of December 2009. Unit: ppm

longitude and comprises most of the land mass of Europe and Asia. The time span was chosen from the 30^{th} of September 2009 to the 13^{th} of August 2012, whereas time intervals of 16 days are applied. All data sets are averaged into a $0.5° \times 0.5°$ grid, whereas only grid cells on land will be considered for the analysis. The resulting observation grid contains 3629 cells and is observed 67 times. Important to note is that the data on surface CO_2 concentrations are very sparse. There are only around 400 observations per time interval, resulting in about 11% non-missing grid cells. All variables have been standardized in advance with respect to their overall mean and standard deviation, in order to improve numerical stability of the estimation procedure. Furthermore the CO_2 data have been logarithmized before the standardization. A short description of the data sets is given in the following paragraphs.

Surface CO_2 concentrations - Japan's Greenhouse Gases Observing Satellite, GOSAT, together with the retrieval algorithms developed by the "Atmospheric CO2 Observations from Space" (ACOS) [O'Dell et al., 2012] team are providing global remotely sensed measurements of the vertical column of CO_2 concentrations, expressed in Parts Per Million (PPM). Besides total column and column average data, the CO_2 concentration is derived for 20 different pressure levels. For a thorough description of the retrieval algorithm, the reader is referred to Osterman et al. [2013]. The layer closest to the surface is used in this study, due to it's proximity to the CO_2 sources and sinks. The satellite is orbiting the earth synchronous to the sun, guaranteeing an almost constant local-sun time for each observation at around 13pm. Consequently, the measurements are not affected by intra-day variations. The satellite returns almost to the same footprint after 3 days [Yokota et al., 2009]. A sample of the data for the 3^{rd}-19^{th} of December 2009, transformed into the aforementioned observation grid and averaged over time intervals of 16 days, is depicted in Figure 1. As can be seen, the data coverage is very sparse. Measurability is limited through cloud coverage and also snow cover leads to missing data, as can be seen in Figure 1, where a large fraction in the northern part of the map is colored white, indicating no available data.

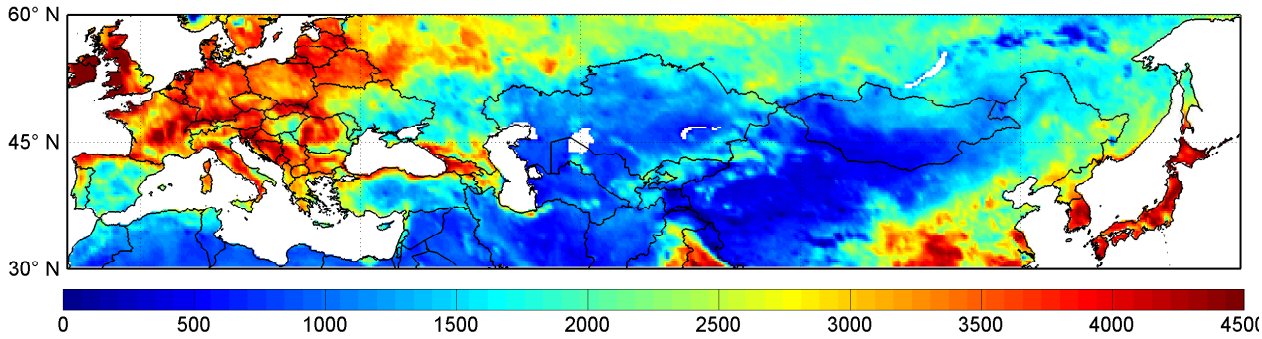

Figure 2: Normalized Difference Vegetation Index for the 30^{th} of September - 15^{th} of October 2009

Normalized Difference Vegetation Index The most influential driver in the global carbon cycle in terms of intra-annual CO_2 fluxes is the terrestrial ecosystem. In order to adequately capture it's source and sink behavior, the relationship between surface CO_2 concentrations and the local vegetation has to be modeled. For that purpose, vegetation indices can be used. The Normalized Difference Vegetation Index (NDVI) is the most frequently applied index and captures vegetation activity at the land surface or the density of living vegetation in an area through the measurement of wavelengths of red and near-infrared light that is absorbed and reflected by plants [Huete et al., 1999]. NDVI data can be obtained from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS)[Huete et al., 1999] in various global grids with averages over 16 days intervals. In Figure 2 the NDVI is depicted for the interval between the 30^{th} of September and the 15^{th} of October 2009. In this study the Net Ecosystem Exchange, as outlined in Section 2.1, is assessed through modeling the space-time varying relationship between the NDVI and surface CO_2 concentrations, as it was shown in Vetter et al. [2015].

Nitrogen Dioxide Anthropogenic CO_2 emissions mainly result from the combustion of fossil fuels and therefore occur during energy production, road traffic and industrial production. In order to capture anthropogenic CO_2 sources, the relationship between surface CO_2 concentrations and other air pollutants, which are produced alongside CO_2 during fossil fuel combustion, is modeled. The first pollutant is Nitrogen Dioxide (NO_2), which is a by-product of all processes of fossil fuel combustion, mainly road traffic and electricity production. This study uses vertical tropospheric column densities of NO_2, which track NO_2 pollution near the surface, available from the Ozone Monitoring Instrument on-board the EOS-Aura satellite[1] and documented in Chance [2002]. The NO_2 data has been averaged over time and space, according to the 16-day interval and the $0.5° \times 0.5°$ grid, and standardized for numerical stability reasons. Despite the high data coverage there is still a small amount of missing data,

[1]URL: `http://projects.knmi.nl/omi/research/product/product_generator.php?&info=intro&product=NO2`

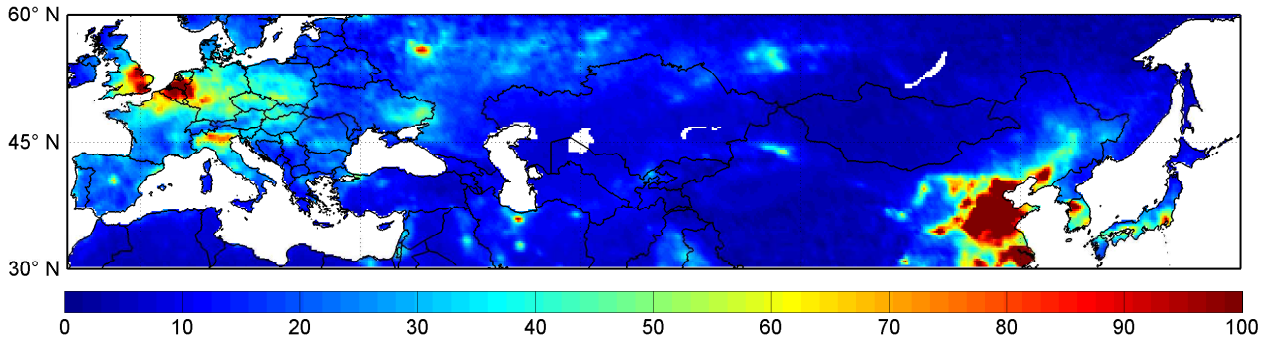

Figure 3: Tropospheric NO_2 column density for the 30^{th} of September - 15^{th} of October 2009. Unit: $10^{15} molec./cm^2$

especially in winter, where snow cover limits the measurability. For that reason a nearest neighbor interpolation was applied to assure that the data set can be used as a covariate. A sample map for the period 30^{th} of September and 15^{th} of October 2009 is provided in Figure 3. High concentrations of NO_2 can be mainly found in urban and industrialized ares, such as Europe and China, where fossil fuel combustion is high.

Carbon Monoxide The combustion of carbon-based fuels leads to the emission of CO_2. Since the combustion is not always complete, Carbon Monoxide (CO) is also a frequent by-product of these processes. Therefore its emissions are characterized by the same spatio-temporal pattern as for the anthropogenic CO_2 emissions. The CO data used in this study can be obtained from the Measurement of Pollution in the Troposphere (MOPITT) instrument flying on NASA's Earth Observing System Terra spacecraft[2] and represents CO surface mixing ratios. A sample of the data for the period between 30^{th} of September and 15^{th} of October 2009 is illustrated in Figure 4, whereas missing data due to snow coverage have been interpolated using the nearest neighbor method. A comparable hot spot pattern as for the NO_2 sample can be identified with increased concentrations in industrialized urban areas. However CO hot spots are less pronounced in Europe.

2.3 Exploratory Data Analysis

A closer look at the main characteristics of the data is required in order to develop a suitable statistical model in space and time. In Figure 5 the spatial average of the CO_2 surface concentration data is depicted over time. Noticeable here is the strong seasonal behavior with lower concentrations in the summer months and elevated values from autumn to spring. The seasonal cycle of CO_2 concentrations

[2]URL: `https://www2.acom.ucar.edu/mopitt`

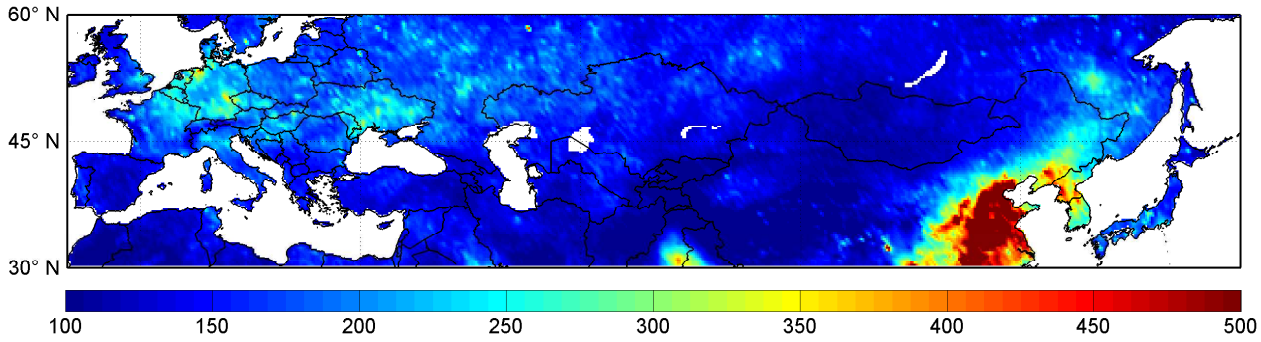

Figure 4: Carbon Monoxide surface mixing ratio for the 30^{th} of September - 15^{th} of October 2009. Unit: ppbv

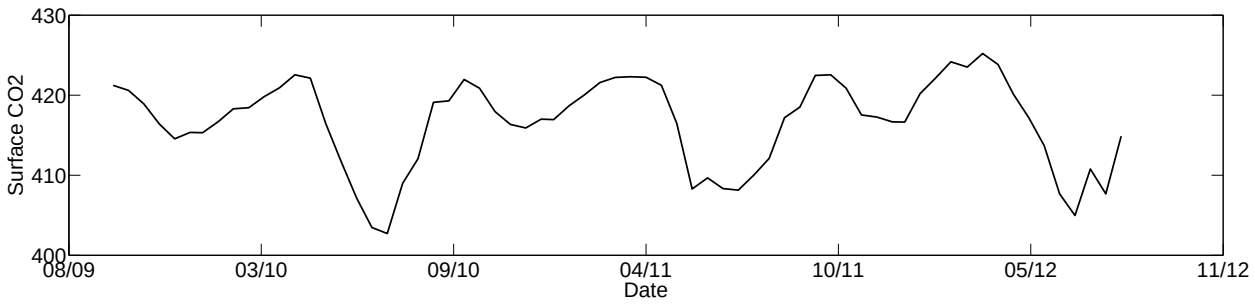

Figure 5: Spatial average of surface CO_2 concentrations over time, Unit: ppm

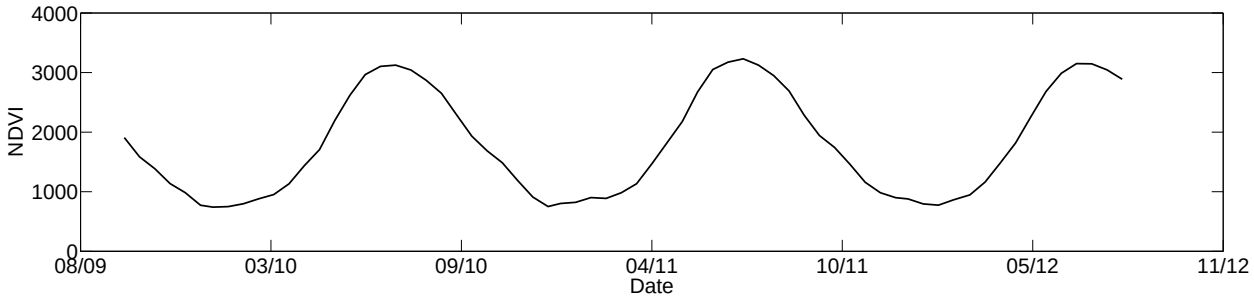

Figure 6: Spatial average of the NDVI over time

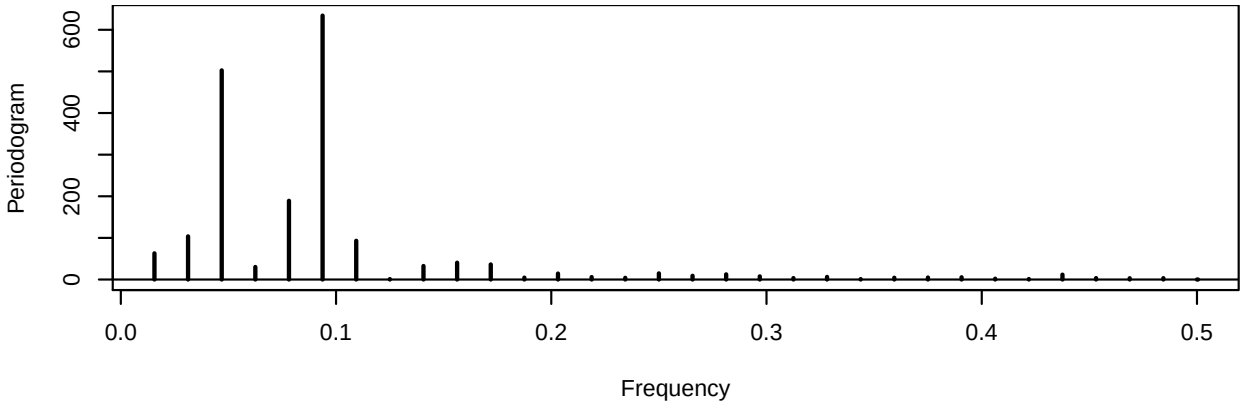

Figure 7: Sample Periodogram for the spatial average surface CO_2

is closely linked to the growth cycle of the terrestrial vegetation. This becomes apparent in comparison with Figure 6, where the spatial average of the amount of living vegetation, represented by the NDVI, is illustrated over time. Here an opposite seasonal cycle is visible with a peak in the NDVI in July to August and lower levels from November till March. However the shape of the CO_2 cycle is different from the NDVI cycle, which can be seen from the two-peak behavior in autumn and winter. This is a result from two overlapping cycles, as can be seen in the sample periodogram for the CO_2 data in Figure 7. The two peaks in the periodogram are at frequencies, which correspond to an annual and a semi-annual cycle. However, these two mixed cycles both result from the terrestrial ecosystem. As described in section 2.1, vegetation acts both as a CO_2 source and sink through photosynthesis and ecosystem respiration. Jiang et al. [2012] find out, that the cycles of GPP and ecosystem respiration are shifted by 6 months, resulting in a two-peak behavior of the NEP, which translates into the same behavior for the CO_2 surface concentration. However, since data on NEP and ecosystem respiration is not available on a global scale for small time intervals, the space-time varying relationship between vegetation, as the source of GPP and respiration, and CO_2 is modeled directly. In Figure 8 the empirical

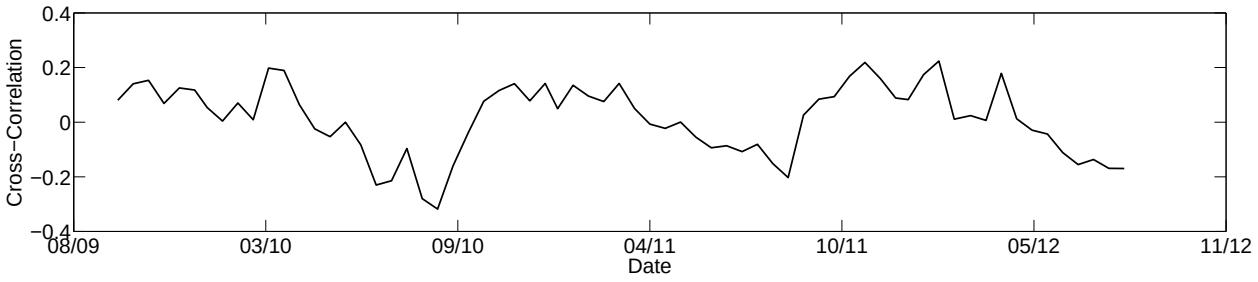

Figure 8: Empirical linear correlation coefficient of the spatial averages of CO_2 and NDVI over time

linear correlation coefficient of the spatial averages of CO_2 and NDVI is depicted over time. Positive values indicate a net source behavior and negative correlations correspond to a net CO_2 sink of the terrestrial ecosystem. In effect, depending on whether photosynthesis or respiration predominates the terrestrial ecosystem seasonally turns into a CO_2 sink or source. The main driver behind these seasonal cycles here is the rate of incoming solar radiation and this quantity not only changes in time, but also varies with space. Depending on the latitude region, the amplitude and phase of the seasonal growth cycle of plants change, whereas higher amplitudes can be observed in northern regions. Consequently, capturing the space-time varying dependence structure between the CO_2 surface concentration and the NDVI, in order to cover the Net Ecosystem Production, becomes a critical requirement of the statistical model.

2.4 Spatio-temporal linear mixed effects model with varying coefficients

As outlined above, the dependence structure between the surface CO_2 concentrations and the explanatory variables is changing in space and time. In order to capture the distribution of the CO_2 sources and sinks resulting from the terrestrial ecosystem and from human activities a suitable statistical model is needed, which incorporates a spatio-temporal dependence and cross-dependence structure. For this purpose the spatio-temporal linear mixed effects model with varying coefficients proposed in Fassò and Finazzi [2013] is applied. The general univariate version of the model is summarized below.

Let $\mathbf{y}(\mathcal{S},t)$ denote the vector of observed responses at the set of spatial locations $\mathcal{S}=\{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$ and at time point $t=1,\ldots,T$, where $\mathbf{s}_i\in\mathcal{D}, i=1,\ldots,n$ and $\mathcal{D}\subset\mathbb{R}^2$, then

$$\begin{aligned}\mathbf{y}(\mathcal{S},t) &= \mathbf{X}_\beta(\mathcal{S},t)\boldsymbol{\beta}+\sum_{i=1}^{c}\alpha_i\cdot\mathbf{x}_{w,i}(\mathcal{S},t)\odot\mathbf{w}_i(\mathcal{S},t)+\sum_{j=1}^{p}\mathbf{x}_{z,j}(\mathcal{S},t)\cdot z_j(t) \\ &\quad +z(t)\mathbf{1}+\delta\mathbf{w}(\mathcal{S},t)+\boldsymbol{\epsilon}(\mathcal{S},t),\end{aligned} \tag{1}$$

where the $n\times k$ matrix $\mathbf{X}_\beta$ and the $n\times 1$ vectors $\mathbf{x}_{w,i}, i=1,\ldots,c$ and $\mathbf{x}_{z,j}, j=1,\ldots,p$ consist of known covariates measured at the set of locations $\mathcal{S}$ and time t. The scalars $\boldsymbol{\alpha}=(\alpha_1,\ldots\alpha_c)', \boldsymbol{\beta}=(\beta_1,\ldots,\beta_k)'$

and δ are unknown scale coefficients and '$\odot$' is the Hadamard product of two matrices. The $n \times 1$ vectors $\mathbf{w}_i, i = 1, \ldots, c$ and $\mathbf{w}$ represent spatial random fields. Assuming multivariate normality, the independent, but spatially colored, Gaussian processes are

$$\begin{aligned} \mathbf{w}_i(\mathcal{S}, t) &\sim N_n\left(\mathbf{0}, \mathbf{\Sigma}_i(\mathbf{H}^{obs}, \theta_i)\right); i = 1, \ldots, c \\ \mathbf{w}(\mathcal{S}, t) &\sim N_n\left(\mathbf{0}, \mathbf{\Sigma}(\mathbf{H}^{obs}, \theta_y)\right), \end{aligned} \tag{2}$$

where $\mathbf{\Sigma}_i, i = 1, \ldots, c$ and $\mathbf{\Sigma}$ are spatial covariance matrices and $\mathbf{H}^{obs}$ is the matrix of pairwise geographic distances between each observation location $\mathbf{s}_i, i = 1, \ldots, n$. The spatial dependence is determined through a valid spatial covariance function $C(d(\mathbf{s}_h, \mathbf{s}_l), \theta_i)$ or $C(d(\mathbf{s}_h, \mathbf{s}_l), \theta_y)$ with $d(\cdot, \cdot)$ as the geographic distance function and parameterized by θ_i or θ_y, respectively. In this study the exponential covariance function is applied

$$C(d(\mathbf{s}_i, \mathbf{s}_j), \theta) = \exp(-d(\mathbf{s}_i, \mathbf{s}_j)/\theta). \tag{3}$$

In order to account for regional differences in the cross-dependence between CO_2 and the explanatory variables, the spatial random effects $\mathbf{w}_i, i = 1, \ldots, c$ represent space varying coefficients. The latent random variable $\mathbf{w}$ models the residual spatial auto-correlation of the data. Likewise the temporal random effects $z_j, j = 1, \ldots, p$ and z are Gaussian processes with lagged cross-correlation and auto-correlation, respectively. The vector of temporal random effects $\mathbf{z}(t) = (z(t), z_1(t), \ldots, z_p(t))'$ is assumed to follow a vector autoregressive process of order one

$$\mathbf{z}(t) = \mathbf{G}\mathbf{z}(t-1) + \boldsymbol{\eta}(t), \tag{4}$$

where $\mathbf{G}$ is a stable $(p+1)\times(p+1)$ transition matrix and $\boldsymbol{\eta}(t) \sim N_{p+1}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\eta}})$ is a Gaussian vector of temporal innovations with covariance matrix $\mathbf{\Sigma}_{\boldsymbol{\eta}}$. The initial state is distributed as $\mathbf{z}(0) \sim N_{p+1}(\boldsymbol{\nu}_0, \mathbf{\Sigma}_0)$. Finally, the vector $\boldsymbol{\epsilon}(\mathcal{S}, t) \sim N_n(\mathbf{0}, \sigma_\epsilon^2 \cdot \mathbf{I})$ models the measurement error of the data process at the observed locations $\mathcal{S}$ and time t and represents a Gaussian white noise process in space and time. The $n \times 1$ vector $\mathbf{1}$ and the $n \times n$ matrix $\mathbf{I}$ denote the unit vector and the identity matrix, respectively. The estimation of the model parameters shown in equation 5 is performed though a maximum likelihood procedure, which is described in detail in Fassò and Finazzi [2013].

$$\Psi = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta, \sigma_\epsilon^2; \boldsymbol{\nu}_0, \mathbf{\Sigma}_0; \mathbf{G}, \mathbf{\Sigma}_{\boldsymbol{\eta}}; \theta_y, \theta_1, \ldots, \theta_c\} = \{\Psi_y; \Psi_{\mathbf{z}_0}; \Psi_z; \Psi_w\}. \tag{5}$$

In a complete-data log-likelihood representation, estimates are obtained through an EM-algorithm, which results in an iterative algorithm with mostly closed form updating formulas for the model parameters. Within this setting, the explanatory variables are connected with a fixed effect and two random effects, which account for the space-time varying dependence with the independent variable.

3 Spatio-temporal statistical assessment of CO_2 sources and sinks

In order to adequately describe the CO_2 cycle and its sources and sinks, the methodology outlined in section 2.4 is now used to model the CO_2 surface concentration data and its space-time dependent relationship with ecosystem and anthropogenic source and sink data. The main focus of the modeling approach is to capture the data characteristics shown in section 2.3. In particular the space-time dependent relationship of CO_2 and the NDVI is of importance, since the ecosystem is the main driver behind variations in surface CO_2 concentrations and is responsible for the two-peak seasonal behavior, shown in section 2.3. At first, the model for the surface CO_2 concentration is defined and the estimation results are presented and discussed in section 3.1. Afterwards, the results of the spatio-temporal smoothing are outlined and the influence of ecosystem and anthropogenic processes on surface CO_2 is illustrated in section 3.2.

3.1 Model estimation

The linear mixed effects model for the surface CO_2 concentration with space and time varying coefficients for the covariates is shown in equation 6. Apart from the explanatory variables shown in section 2.2 additional covariates were introduced. These are the product of CO and $NDVI$ and of NO_2 and $NDVI$. Emissions of CO and NO_2 do not solely result from anthropogenic sources, but also occur in relation with ecosystem processes. Both pollutants are produced during wild fires and CO is also produced to a small proportion during the heterotrophic respiration. Since we are mainly interested in the anthropogenic proportion of CO and NO_2 the two interaction effects with the NDVI were introduced in order to capture the CO and NO_2 sources resulting from the ecosystem.

$$\begin{aligned}
\mathbf{CO_2}(\mathcal{S},t) \;=\; & \mathbf{NDVI}(\mathcal{S},t)\beta_{NDVI} + \mathbf{CO}(\mathcal{S},t)\beta_{CO} + \mathbf{NO_2}(\mathcal{S},t)\beta_{NO_2} \\
& + \mathbf{NO_2}(\mathcal{S},t) \odot \mathbf{NDVI}(\mathcal{S},t) \cdot \beta_{NN} + \mathbf{CO}(\mathcal{S},t) \odot \mathbf{NDVI}(\mathcal{S},t) \cdot \beta_{CN} \\
& + \alpha_{NDVI} \cdot \mathbf{NDVI}(\mathcal{S},t) \odot \mathbf{w}_{NDVI}(\mathcal{S},t) + \alpha_{CO} \cdot \mathbf{CO}(\mathcal{S},t) \odot \mathbf{w}_{CO}(\mathcal{S},t) \\
& + \alpha_{NO_2} \cdot \mathbf{NO_2}(\mathcal{S},t) \odot \mathbf{w}_{NO_2}(\mathcal{S},t) + \mathbf{NDVI}(\mathcal{S},t) \cdot z_{NDVI}(t) \\
& + z(t)\mathbf{1} + \delta\mathbf{w}(\mathcal{S},t) + \boldsymbol{\epsilon}(\mathcal{S},t)
\end{aligned} \tag{6}$$

In the first two lines the fixed effects $\beta_{NDVI}, \beta_{CO}, \beta_{NO_2}, \beta_{NN}$ and β_{CN} for the covariates and the interaction terms are introduced. The corresponding indices consist of abbreviations for the related covariate, whereas $NDVI$ denotes the Normalized Difference Vegetation Index, CO stands for Carbon Monoxide, NO_2 for Nitrogen Dioxide, NN marks the product of NO_2 and $NDVI$ as a covariate and CN abbreviates the product of CO and $NDVI$. The relationship between the ecosystem and CO_2 is expected to depend on the region, due to different vegetation types and climate conditions,

	β_{NDVI}	β_{CO}	β_{NO_2}	β_{NN}	β_{CN}	α_{NDVI}	θ_{NDVI}	α_{CO}	θ_{CO}
Estimate	-0.250	0.250	0.069	0.086	0.157	0.525	1000.84	0.152	204.26
Stand. deviation	0.026	0.069	0.008	0.013	0.027	0.010	53.80	0.008	46.60
	α_{NO_2}	θ_{NO_2}	δ	θ	σ_ϵ^2				
Estimate	0.089	176.83	0.323	876.33	0.629				
Stand. deviation	0.006	42.44	0.047	22.19	0.070				

Table 1: Maximum Likelihood Estimates

which results in a spatial cross-correlation. This behavior is accounted for by the spatial random effect $\mathbf{w}_{NDVI}$. Likewise the relationship between CO_2 and the trackers for anthropogenic processes CO and NO_2 is expected to be spatially dependent, because of regional differences in technology. The resulting spatial cross-correlation is covered through $\mathbf{w}_{CO}$ and $\mathbf{w}_{NO_2}$. Additionally a temporal random effect for the ecosystem influence z_{NDVI} is used, in order to capture the seasonally changing source and sink behavior. Residual spatial and temporal auto-correlation in the CO_2 data is accounted for by $\mathbf{w}$ and z, respectively. Using the EM-Algorithm, described in Fassò et al. [2009], Maximum likelihood estimates of the parameter set and smoothed values for the latent random variables can be iteratively obtained. The parameter estimates and their standard deviation are summarized in table 1 and equations 7 and 8. The fixed effects for $NDVI$,CO and NO_2 cannot be interpreted separately, since the total effects of the covariates on CO_2 results from the combination of both fixed and random effects. The space-time varying effects are discussed in detail in section 3.2. However, interpretable are the fixed effects for the interaction terms β_{NN} and β_{CN}. Carbon Monoxide and Nitrogen Dioxide emissions originating from the ecosystem through wildfires and heterotrophic respiration have a positive and significant impact on local CO_2 concentrations. Concerning the scale coefficients α_{NDVI},α_{CO},α_{NO_2} and δ it can be stated, that a large weight is given to the spatial random effect of the $NDVI$, which confirms its high influence on the variability of surface CO_2 concentrations. As can be seen from the weights, the influence of CO and NO_2 is much smaller, as expected. The intermediate value for δ shows, that there is still a sizable amount of residual spatial auto-correlation, which indicates that either there are influence factors not considered in the model or that the fit of the spatial random fields could be improved by a more complex spatial correlation structure (e.g. through anisotropic correlation). Furthermore from the spatial range parameters θ_{NDVI},θ_{CO},θ_{NO_2} and θ it can be seen, that the spatial influence of anthropogenic sources is more local, than for the ecosystem. However, as the high value of the nugget variance of the measurement error process shows, much uncertainty is left in the residuals, which is mainly a result of the large measurement uncertainty of the instrument of the Greenhouse Gases Observing Satellite. The estimated transition matrix and the variance-covariance matrix of the

temporal innovations related to the temporal state vector $\mathbf{z}(t) = \{z_{NDVI}(t), z(t)\}$ shown in equations 7 and 8 reveals a very persistent temporal cross-dependence and auto-correlation behavior, due to the seasonality in the data.

$$\hat{\mathbf{G}} = \begin{pmatrix} 0.801_{(0.03)} & 0.063_{(0.002)} \\ -0.221_{(0.03)} & 0.902_{(0.04)} \end{pmatrix} \tag{7}$$

$$\hat{\mathbf{\Sigma}}_{\boldsymbol{\eta}} = \begin{pmatrix} 1.2 \cdot 10^{-4}_{(3.2 \cdot 10^{-7})} & -1.6 \cdot 10^{-4}_{(3.2 \cdot 10^{-8})} \\ -1.6 \cdot 10^{-4}_{(1.9 \cdot 10^{-8}))} & 8.3 \cdot 10^{-4}_{(3.2 \cdot 10^{-7})} \end{pmatrix} \tag{8}$$

Using $\widehat{\Psi}$, as the Maximum-Likelihood estimate of the parameter set

$$\Psi = \{\beta_{NDVI}, \beta_{CO}, \beta_{NO_2}, \beta_{NN}, \beta_{CN}, \alpha_{NDVI}, \theta_{NDVI}, \alpha_{CO}, \theta_{CO}, \alpha_{NO_2}, \theta_{NO_2}, \delta, \theta, \sigma^2_\epsilon, \mathbf{G}, \mathbf{\Sigma}_\eta\},$$

then the surface CO_2 concentration can be predicted at a set of spatial locations $\mathcal{S}_0$ and for time point $t = 1, \ldots, T$ as

$$\begin{aligned} \widehat{\mathbf{CO_2}}(\mathcal{S}_0, t) = \; & \mathbf{NDVI}(\mathcal{S}_0, t)\hat{\beta}_{NDVI} + \mathbf{CO}(\mathcal{S}_0, t)\hat{\beta}_{CO} + \mathbf{NO_2}(\mathcal{S}_0, t)\hat{\beta}_{NO_2} \\ & + \mathbf{NO_2}(\mathcal{S}_0, t) \odot \mathbf{NDVI}(\mathcal{S}_0, t) \cdot \hat{\beta}_{NN} + \mathbf{CO}(\mathcal{S}_0, t) \odot \mathbf{NDVI}(\mathcal{S}_0, t) \cdot \hat{\beta}_{CN} \\ & + \hat{\alpha}_{NDVI} \cdot \mathbf{NDVI}(\mathcal{S}_0, t) \odot \mathbf{w}^T_{NDVI}(\mathcal{S}_0, t) + \hat{\alpha}_{CO} \cdot \mathbf{CO}(\mathcal{S}_0, t) \odot \mathbf{w}^T_{CO}(\mathcal{S}_0, t) \\ & + \hat{\alpha}_{NO_2} \cdot \mathbf{NO_2}(\mathcal{S}_0, t) \odot \mathbf{w}^T_{NO_2}(\mathcal{S}_0, t) + \mathbf{NDVI}(\mathcal{S}_0, t) \cdot z^T_{NDVI}(t) \\ & + z^T(t)\mathbf{1} + \hat{\delta}\mathbf{w}^T(\mathcal{S}_0, t), \end{aligned} \tag{9}$$

where $z^T_{NDVI}(t) = E_{\widehat{\Psi}}\left[z_{NDVI}(t)|\mathbf{CO_2}^{obs}\right]$ and $z^T(t) = E_{\widehat{\Psi}}\left[z(t)|\mathbf{CO_2}^{obs}\right]$ is the Kalman smoother output with $\mathbf{CO_2}^{obs}$ as the $n \times T$ observation matrix and

$$\begin{aligned} \mathbf{w}^T_{NDVI}(\mathcal{S}_0, t) &= E_{\widehat{\Psi}}\left[\mathbf{w}_{NDVI}(\mathcal{S}_0, t)|\mathbf{CO_2}^{obs}\right] \\ \mathbf{w}^T_{CO}(\mathcal{S}_0, t) &= E_{\widehat{\Psi}}\left[\mathbf{w}_{CO}(\mathcal{S}_0, t)|\mathbf{CO_2}^{obs}\right] \\ \mathbf{w}^T_{NO_2}(\mathcal{S}_0, t) &= E_{\widehat{\Psi}}\left[\mathbf{w}_{NO_2}(\mathcal{S}_0, t)|\mathbf{CO_2}^{obs}\right] \\ \mathbf{w}^T(\mathcal{S}_0, t) &= E_{\widehat{\Psi}}\left[\mathbf{w}(\mathcal{S}_0, t)|\mathbf{CO_2}^{obs}\right] \end{aligned} \tag{10}$$

are the estimated conditional expectations of the latent spatial random variables. The according prediction variance-covariance matrix $Var_{\widehat{\Psi}}\left[\widehat{\mathbf{CO_2}}(\mathcal{S}_0, t)|\mathbf{CO_2}^{obs}\right]$ can be obtained by evaluating the conditional variances and covariances of the latent random variables in equation 9. The smoothed predictions and prediction errors can be used to produce dynamic maps. The set of prediction locations used in this study consists of a $0.5° \times 0.5°$ grid, bounded by the same latitude and longitude limits, as the observed data, and an applied land/ocean mask. In figure 9 the temporal average prediction of the surface CO_2 concentration and its prediction standard error is depicted over the european area.

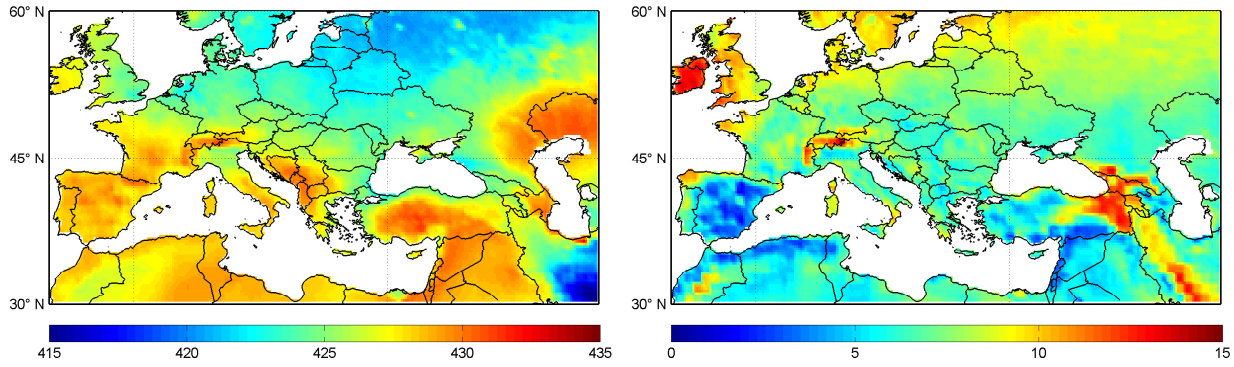

Figure 9: Predicted temporal average surface CO_2 concentration (Left) and its prediction standard error (Right) in ppm

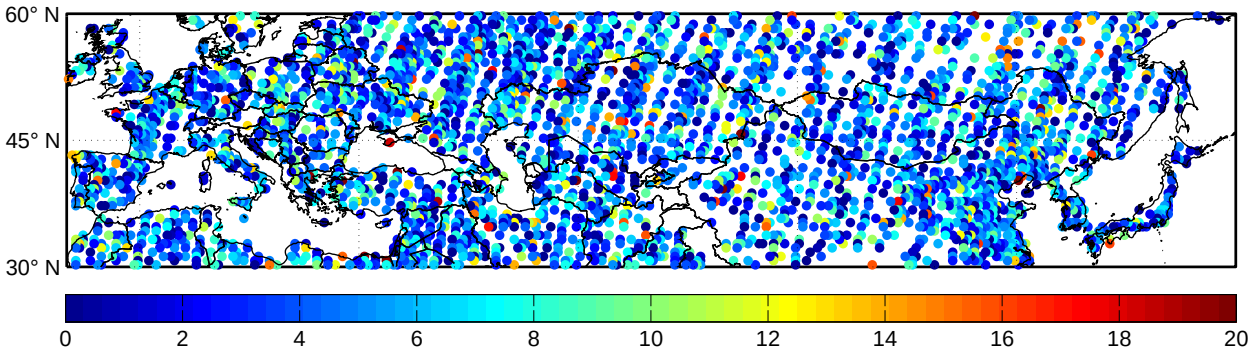

Figure 10: Temporal average RMSE, Unit: ppm

As can be seen in the left panel of figure 9, on average lower CO_2 concentrations can be observed in the northern region above $45°N$, in areas with much vegetation and in coastal areas. In contrast, higher concentrations are dominant in southern areas with less vegetation and in mountain areas. The pattern of the average prediction standard error, shown in the right panel of figure 9, is characterized by elevated prediction uncertainty in the northern prediction area, where the CO_2 surface data is very sparse, especially in winter times. Furthermore a higher prediction error can be observed in sparse vegetation regions, like in mountains, and as well in areas with much vegetation. Besides the estimated prediction uncertainty, a 10-fold cross-validation study was performed in order to evaluate the model fit and the out-of-sample prediction performance. In figure 10 the resulting RMSE is depicted over space, whereas no clear spatial pattern of elevated prediction errors can be identified. As can be seen from table 2 the total RMSE from the cross-validation study amounts to $6.201ppm$. Compared with the expected total prediction error (RMSPE) of $8.828ppm$ the model performs even better than expected. The corresponding adjusted R^2 is with 0.894 very high, especially when compared with

RMSPE	RMSE	R^2_{adj}
8.828	6.201	0.894

Table 2: Predictive power

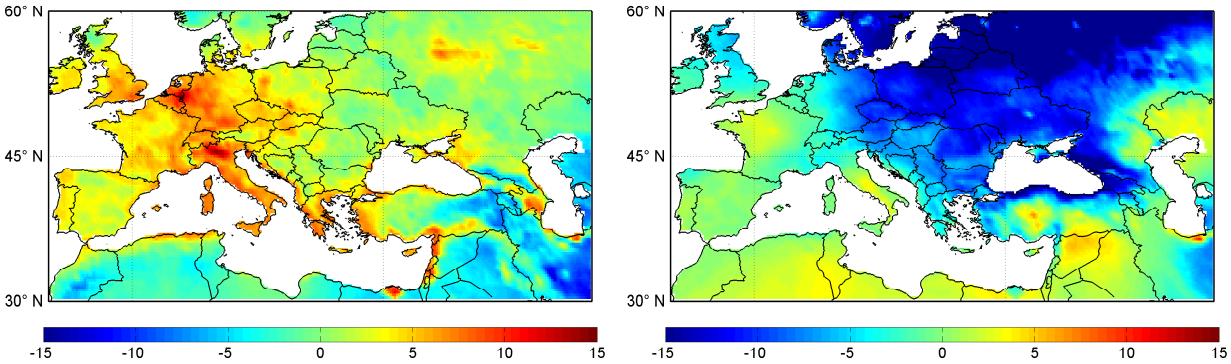

Figure 11: Ecosystem influence on surface CO_2 concentrations in January 2010 (Left) and July 2010 (Right)

other applications in the field of environmental processes. The model was able to capture around 90% of the variability in the surface CO_2 concentration data.

3.2 Spatio-temporal smoothing and CO_2 sources and sinks

The special focus of this article is to quantify CO_2 sources and sinks and their spatio-temporal distribution. For this purpose the estimated fixed and random effects for $NDVI$,CO,NO_2 and the product terms have to be analyzed separately. In order to obtain an estimate for the influence of the ecosystem on surface CO_2 concentrations ($\widehat{\mathbf{ECO}}$) at the set of prediction locations $\mathcal{S}_0$ and at time point t, the terms related to the $NDVI$ in equation 9 are collected:

$$\begin{aligned}\widehat{\mathbf{ECO}}(\mathcal{S}_0,t) \;=\; & \mathbf{NDVI}(\mathcal{S}_0,t)\hat{\beta}_{NDVI} + \mathbf{NO_2}(\mathcal{S}_0,t) \odot \mathbf{NDVI}(\mathcal{S}_0,t)\cdot\hat{\beta}_{NN} \\ & + \mathbf{CO}(\mathcal{S}_0,t) \odot \mathbf{NDVI}(\mathcal{S}_0,t)\cdot\hat{\beta}_{CN} + \hat{\alpha}_{NDVI}\cdot\mathbf{NDVI}(\mathcal{S}_0,t)\odot\mathbf{w}^T_{NDVI}(\mathcal{S}_0,t) \\ & + \mathbf{NDVI}(\mathcal{S}_0,t)\cdot z^T_{NDVI}(t).\end{aligned} \tag{11}$$

The temporal averages of $\widehat{\mathbf{ECO}}$ for January (left) and July 2010 (right) are depicted in figure 11. In the two maps, it can be seen, that as already explained in section 2.1, the terrestrial ecosystem seasonally changes from a CO_2 source from late autumn to early spring into a sink from late spring to early autumn. The sink behavior in summer is thereby more pronounced in the north-east of Europe and in coastal areas, whereas the source activity of the ecosystem is stronger in central and south-

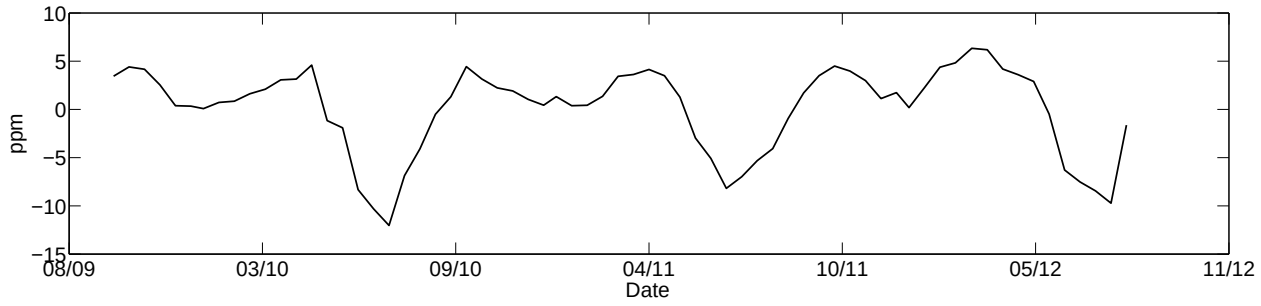

Figure 12: Spatial average of ecosystem influence on surface CO_2 concentrations over time, Unit: ppm

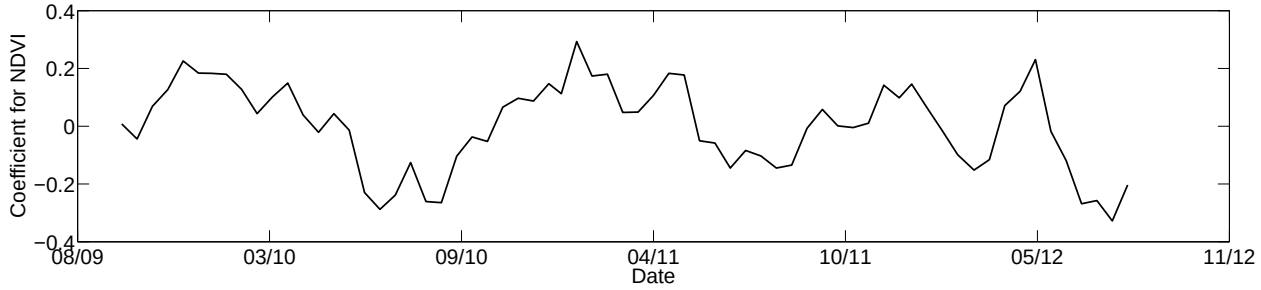

Figure 13: Spatio-temporal coefficient for NDVI over time

west Europe and also at the coast. Both source and sink intensity is higher in regions with much vegetation, as expected. Regions with elevated source activity in winter lead to an increase in the local surface CO_2 concentration between 5 and $10ppm$ and the strong north-eastern sink in summer causes CO_2 concentrations to be reduced by 10 to $15ppm$. In figure 12 the spatial average of the predicted ecosystem influence is shown over time. By comparing the figure with the spatial average of the CO_2 observation data in figure 5, it can be stated, that the shape of both plots are very similar with the two peaks in winter, which results from the two shifted seasonal cycles of the GPP and the heterotrophic respiration originating from the ecosystem. The model was able to describe the space-time varying source and sink activity of the ecosystem and is able to quantify the Net Ecosystem Production and its seasonal behavior, as already described in Jiang et al. [2012]. On average, the ecosystem reduces surface CO_2 concentrations up to $10ppm$ between May and September and leads to an increase in CO_2 concentrations up to $5ppm$ between October and April. The annual averages of the years 2010 and 2011 are close to zero and amounted to $-0.4ppm$ and $0.13ppm$, which means that the ecosystem was a small net sink in 2010 and a small CO_2 source in 2011. In figure 13 the space-time varying coefficients for the $NDVI$ are depicted. In comparison with figure 8, it can be stated, that the model was able to replicate the space-time varying dependence structure between the NDVI and the surface CO_2 concentrations.

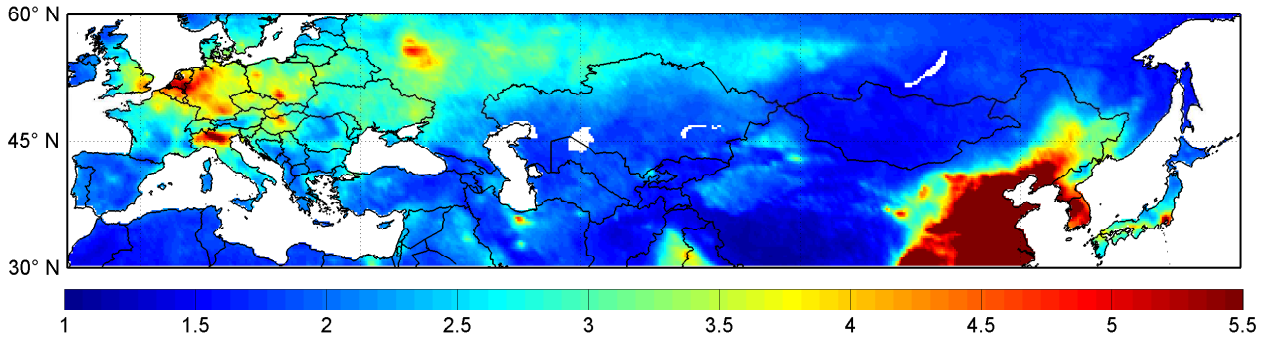

Figure 14: Anthropogenic influence on surface CO_2 concentration in January 2010

Between October and April, the spatial average coefficient (like the linear correlation between $NDVI$ and CO_2) tends to be positive and it is mostly negative between May and September.

The anthropogenic influence on surface CO_2 concentration ($\widehat{\mathbf{ANTHRO}}$) is now analyzed by collecting the fixed and random effects related to CO and NO_2:

$$\begin{aligned}\widehat{\mathbf{ANTHRO}}(\mathcal{S}_0,t) &= \mathbf{CO}(\mathcal{S}_0,t)\hat{\beta}_{CO} + \mathbf{NO_2}(\mathcal{S}_0,t)\hat{\beta}_{NO_2} + \hat{\alpha}_{CO}\cdot\mathbf{CO}(\mathcal{S}_0,t)\odot\mathbf{w}_{CO}^T(\mathcal{S}_0,t) \\ &\quad +\hat{\alpha}_{NO_2}\cdot\mathbf{NO_2}(\mathcal{S}_0,t)\odot\mathbf{w}_{NO_2}^T(\mathcal{S}_0,t).\end{aligned} \quad (12)$$

The temporal average prediction of the anthropogenic influence on surface CO_2 concentrations $\widehat{\mathbf{ANTHRO}}$ for January 2010 is shown in figure 14. The spatial pattern of anthropogenic CO_2 emission activity is closely related to urbanization. Emissions are high in strongly industrialized and populated regions, as for example in Germany, the Benelux countries, the Lombardi region and around most big cites, like Moscow, London, Warsaw, Teheran, Tokio and New-Delhi. The biggest anthropogenic CO_2 source can be identified in China. In the urbanized areas, the anthropogenic activity led to an elevated level of surface CO_2 concentration between 4 and 5.5ppm in January 2010, whereas in China the maximum was 8ppm. The anthropogenic emissions also follow a seasonal cycle, as can be seen in figure 15. The same seasonal pattern was already identified in Rotty [1987], where monthly fossil fuel consumption data was used in order to describe the seasonal CO_2 emission cycle. The anthropogenic influence on surface CO_2 concentrations seasonally varies between 2 and 2.6ppm in the spatial average. Likewise the peak-to-valley distance is around 25% of the mean level, which is also confirmed by the study of [Rotty, 1987][p. 195]. Due to the increased electricity and fuel demand in winter months in the northern hemisphere, which result from the low amount of daylight and the cold temperatures, the CO_2 emissions peak in January-February and are the lowest in June-July, where the electricity and fossil fuel consumption is the lowest. It becomes obvious, that CO_2 emissions and its seasonal cycle strongly depends on climate conditions, for example a warm winter leads to lower fossil fuel consumption and consequently to lower CO_2 emissions. This relationship becomes important, while evaluating national

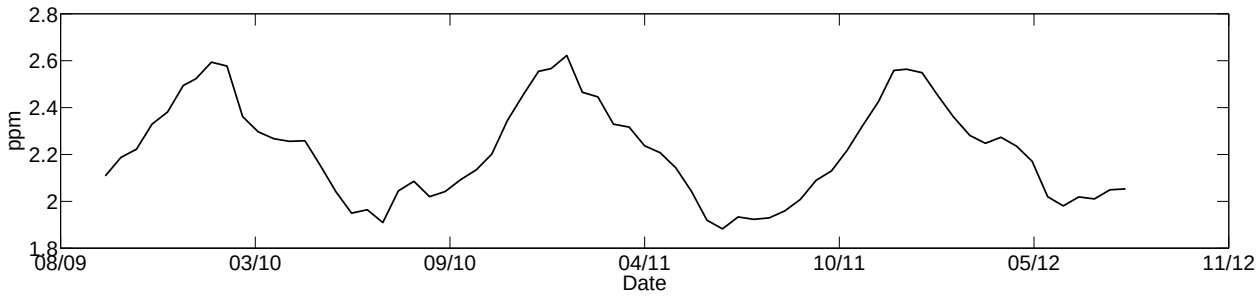

Figure 15: Spatial average of anthropogenic influence on surface CO_2 concentrations over time, Unit: ppm

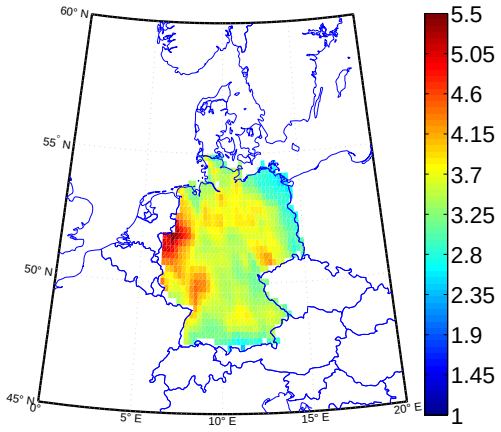

Figure 16: Anthropogenic influence on surface CO_2 concentrations in Germany on the 16^{th} of October 2009

CO_2 balances, because inter-annual emission reductions may not solely result from reduction efforts, but also from climate anomalies.

4 Validation of CO_2 emissions reports

The predictions of the anthropogenic influence on surface CO_2 concentrations from equation 12 can now be used to monitor relative annual changes in national CO_2 emissions. Therefore land masks for several northern hemisphere countries have been created using shapefiles, which can be freely downloaded at `http://gadm.org/`. In that way, the grid cells, which are located within the respective country borders, can be selected, as can be seen for Germany for example in figure 16 for the 16^{th} of October 2009. The next step is to sum up the anthropogenic CO_2 concentration anomalies over all

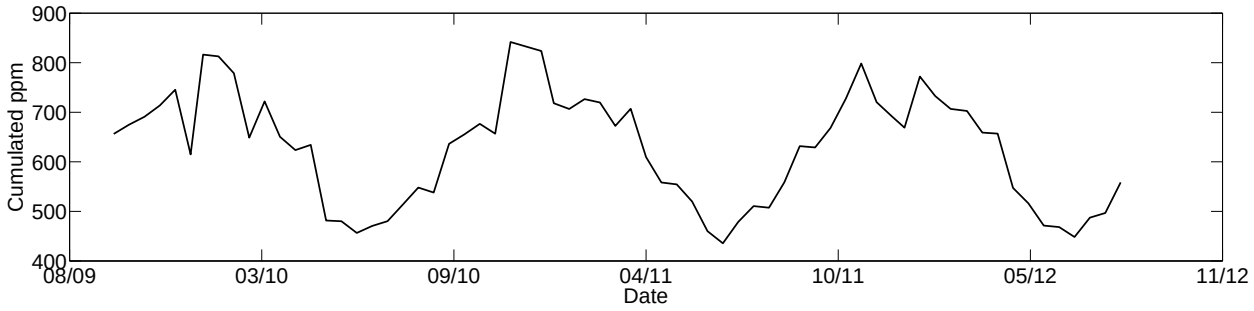

Figure 17: Anthropogenic influence over time in Germany accumulated over space

grid cells belonging to the country. The corresponding time series of sums for the Germany example can be seen in figure 17. The cumulated anthropogenic CO_2 concentration anomalies are used as an approximation for national CO_2 emissions. The absolute values are not comparable with emissions, since the anthropogenic CO_2 concentration anomalies are expressed in ppm. But relative changes should be directly comparable with changes in CO_2 emissions. In this study annual change rates are considered. In effect, the change in the cumulated sum over the year is computed for 2010 and 2011, since these are the only years with full record. These rate of changes are used for a comparison with official reports on national greenhouse gas inventories within the UNFCCC (UNFCCC [2015]) and a more scientifically derived estimate from the Emission Database for Global Atmospheric Research (EDGAR, documented in Olivier et al. [2015]), where aggregated country emissions, using emission factors for different sectors, are spatially gridded using proxy variables, such as population density and road network. In table 3 the annual rate of change in CO_2 emissions of the UNFCCC reports and the EDGAR emissions data are compared with the rate of change of the anthropogenic influence on surface CO_2 concentrations, obtained from the model predictions. The first thing to notice is, that CO_2 emissions have experienced a strong decline in central European countries from 2010 to 2011. However, it has to be noted, that a large proportion of this decline is attributed to the differing climate conditions in both years. Compared to 2011, the winter months in 2010 were much colder, resulting in an increased demand in heating, which led to a fossil fuel consumption and increased CO_2 emissions in winter. For example, according to Wetterdienst [14.07.2016], the average temperature in Germany for the months January, February and December 2010 was $-0.75°C$, compared to an average of $2.58°C$ in 2011. This strengthens the need for accounting for climate anomalies in the inter-annual comparison of CO_2 emissions. Nevertheless, strong CO_2 reduction were realized in France, the UK, Hungary and the Netherlands, whereas increases can be found in Japan, South Korea and Spain.

By comparing the estimated rate of changes from the three methods, it can be stated, that the rates obtained from the model output of equation 12 are close to the reported values from EDGAR

	Model	EDGARv4.3	UNFCCC
Germany	-2.27%	-2.31%	-2.28%
France	-4.62%	-7.42%	-6.71%
Poland	-1.74%	-1.50%	-0.57%
Italy	-3.46%	-2.45%	-2.74%
UK	-7.88%	-7.67%	-8.11%
Spain	0.20%	0.25%	0.19%
Hungary	-3.92%	-3.65%	-3.50%
Czech Republic	-2.25%	-1.05%	-1.77%
Netherlands	-5.36%	-6.23%	-7.33%
Japan	3.01%	3.04%	4.16%
South Korea	2.82%	2.27%	4.96%

Table 3: Estimated changes of national CO_2 emissions from 2010 to 2011

and the UNFCCC, in terms of direction and magnitude. However, there is no clear tendency, whether the results from the model are closer to that of the UNFCCC or the EDGAR. The advantage of using an observation based statistical model is that it is more objective, it requires less assumptions and less conventions have to be made in order to clarify reporting standards and definitions of emissions factors. In that way, it saves a lot of costs for maintaining a monitoring and reporting infrastructure. In effect, it is much easier to apply and less prone to errors and inaccuracies, resulting from emission factors not accounting properly for differences in technology. Furthermore, it can be applied globally and not only for countries, who agreed for the reporting and monitoring standards. It can also be used as an instrument of verification and as a device for trust building, since it cancels out the opportunity to manipulate reports. Consequently, it potentially stabilizes climate negotiation processes through building trust in the reported CO_2 emissions.

5 Concluding Remarks

The present paper offers a helpful perspective on the analysis of sources and sinks of CO_2. It provides a spatio-temporal statistical methodology, that is purely data-driven, which is the main advantage over existing approaches from the natural science and from the reporting infrastructure, since it does not rely on critical assumptions on physical dynamics and emission factors and many more. Based on satellite data on surface CO_2 concentrations and on covariate data, such as the NDVI and the co-polluters

NO_2 and CO, it is possible to derive estimates of the Net Ecosystem Production and anthropogenic emissions. Especially, objective and reliable estimates of anthropogenic sources are of importance in a political framework. Beginning in 2016, all countries to the United Nations Framework Convention on Climate Change (UNFCCC) will report their national emission plans as Nationally Determined Contributions (NDCs), which are the basis of the Paris agreement on climate change. All NDCs are based on inventory data, although they take very diverse forms, e.g. based on different sectors, different baseline years, make different hypotheses of future economic development, and different accounting approaches. In this context, it is critical that emissions baselines and future reduction efforts can be compared among nations on the basis of transparent, consistent, verifiable and up-to-date information. In the past two decades, evidence of sustained climate change and its negative impacts on human welfare has fostered a consensus among nations to act collectively to reduce emissions in the future. During the same period, fossil CO_2 emissions have continued to increase, and in 2015, they were 54% above the levels of 1990. The recent increase in emissions since the late 1990s has occurred overwhelmingly in developing nations, which contributed 69% of the global emissions in 2014, as compared to 32% in 1990, the baseline year of the Kyoto Protocol. China, which accounted for almost half of this share, has recently put a considerable effort in revising its energy statistics in years since 2000. While this has definitely resulted in better estimates of the real fossil-fuel consumption in this period, the uncertainty levels have not decreased by much and rumors of under-reporting are repeatedly circulating (New York Times, 4.11.2015, Nature Climate Change 2012). Because emission inventories in developing nations are overall less accurate than they are in developed countries, the uncertainty associated with fossil CO_2 emissions and their trends has increased up to the point where it could undermine the credibility and the stability of future climate agreements. The NDCs of the national emission reduction plans, that form the basis of the Paris agreement, require robust emissions mitigation efforts, as well as the monitoring of energy- and fossil-fuel intensive national activities to establish a global MRV infrastructure. In that way, a data-driven spatio-temporal statistical methodology could be a promising MRV-tool. Likewise it is useful for analyzing the carbon sequestration potential of the terrestrial vegetation, since the method allows for a global monitoring of the Net Ecosystem Production in narrow time intervals. Finally, it provides a further step for demystifying the global distribution of CO_2 sources and sinks.

References

K Chance. OMI algorithm theoretical basis document, volume IV: OMI trace gas algorithms, August 2002. URL `http://ozoneaq.gsfc.nasa.gov/media/docs/ATBD-OMI-04.pdf`.

N. Cressie and C. K. Wikle. Space-time Kalman filter. *Encyclopedia of environmetrics*, 2002.

N. Cressie and C. K. Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2011.

Noel Cressie, T Shi, and E.L. Kang. Fixed Rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, 19:724–745, 2010.

B. F. Farrell and P. J. Ioannou. State estimation using a reduced-order Kalman filter. *Journal of the Atmospheric Sciences*, 58(23):3666–3680, 2001.

A. Fassò and F. Finazzi, editors. *Statistical mapping of air quality by remote sensing*, 2010. Proceedings of the Accuracy 2010 conference.

A. Fassò and F. Finazzi. Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics*, 22(6):735–748, 2011.

A. Fassò and F. Finazzi. A varying coefficients space-time model for ground and satellite air quality data over Europe. *Statistica & Applicazioni - Special Issue 2013*, pages 45–56, 2013.

A. Fassò, F. Finazzi, and C. D'Ariano. Integrating satellite and ground level data for air quality monitoring and dynamical mapping. *Working papers GRASPA*, 34, 2009.

R. Furrer, M. G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, September 2006.

Wenceslao González-Manteiga, Rosa M. Crujeiras, Matthias Katzfuss, and Noel Cressie. Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics*, 23(1):94–107, 2012.

A. Huete, C. Justice, and W. van Leeuwen. MODIS vegetation index (MOD13) algorithm theoretical basis document. Technical report, NASA Goddard Space Flight Center, 1999.

Xun Jiang, Moustafa T Chahine, Qinbin Li, Maochang Liang, Edward T Olsen, Luke L Chen, Jingqian Wang, and Yuk L Yung. CO2 semiannual oscillation in the middle troposphere and at the surface. *Global Biogeochemical Cycles*, 26(3), 2012.

Matthias Katzfuss and Noel Cressie. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32:430–446, 2011.

K. V. Mardia, C. Goodall, E. J. Redfern, and F. J. Alonso. The kriged Kalman filter. *Test*, 7(2): 217–282, 1998.

C. W. O'Dell, B. Connor, H. Bösch, D. O'Brien, C. Frankenberg, R. Castano, M. Christi, D. Eldering, B. Fisher, M. Gunson, J. McDuffie, C. E. Miller, V. Natraj, F. Oyafuso, I. Polonsky, M. Smyth, T. Taylor, G. C. Toon, P. O. Wennberg, and D. Wunch. The ACOS CO 2 retrieval algorithm - Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Technique*, 5(1):99–121, 2012.

OECD/IEA. *Key world energy statistics*. IEA Publishing, 2015.

Jos GI Olivier, Jeroen AHW Peters, Greet Janssens-Maenhout, and M Muntean. Trends in global CO2 emissions 2015 report, 2015.

G. Osterman, A. Eldering, C. Avis, C. O'Dell, E. Martinez, D. Crisp, C. Frankenberg, B. Fisher, and D. Wunch. ACOS Level 2 Standard Product Data User's Guide, v.3.3. Technical report, Goddard Earth Science Data Information and Services Center (GES DISC), 2013.

Iain Colin Prentice, GD Farquhar, MJR Fasham, ML Goulden, Martin Heimann, VJ Jaramillo, HS Kheshgi, C LeQuéré, RJ Scholes, and Douglas WR Wallace. The carbon cycle and atmospheric carbon dioxide. In JT Houghton, Y Ding, DJ Griggs, M Noguer, PJ van der Linden, and D Xiaosu, editors, *Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 3, pages 183–239. Cambridge University Press, 2001.

M. Reichstein, E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havránkova, H. Ilvesniemi, D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, M. Ourcival, J. Pumpanen, S. Rambal, E. Rotenberg, M. Sanz, J. Tenhunen, G. Seufert, F. Vaccari, T. Vesala, D. Yakir, and R. Valentini. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, 11(9):1424–1439, 2005.

Ralph M Rotty. Estimates of seasonal variation in fossil fuel CO 2 emissions. *Tellus B*, 39(1-2), 1987.

N. Saigusa, S. Yamamoto, S. Murayama, H. Kondo, and N. Nishimura. Gross primary production and net ecosystem exchange of a cool-temperate deciduous forest estimated by the eddy covariance method. *Agricultural and Forest Meteorology*, 112(3):203–215, 2002.

P. C. Stoy, G. G. Katul, M. Siqueira, J. Juang, K. A. Novick, J. M. Uebelherr, and R. Oren. An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration. *Agricultural and Forest Meteorology*, 141(1):2–18, 2006.

A. E. Suyker and S. B. Verma. Year-round observations of the net ecosystem exchange of carbon dioxide in a native tallgrass prairie. *Global Change Biology*, 7(3):279–289, 2001.

UNFCCC, editor. *National greenhouse gas inventory data for the period 1990-2013. Note by the secretariat*, November 2015. United Nations Office at Geneva.

F. Veroustraete, J. Patyn, and R. B. Myneni. Estimating net ecosystem exchange of carbon using the normalized difference vegetation index and an ecosystem model. *Remote Sensing of Environment*, 58(1):115–130, 1996.

P. Vetter, W. Schmid, and R. Schwarze. Efficient Approximation of the Spatial Covariance Function for Large Datasets - Analysis of Atmospheric CO_2 Concentrations. *Journal of Environmental Statistics*, 6(3):1–36, 2014.

Patrick Vetter, Wolfgang Schmid, and Reimund Schwarze. Spatio-temporal statistical analysis of the carbon budget of the terrestrial ecosystem. *Statistical Methods & Applications*, 25(1):143–161, 2015. ISSN 1613-981X. doi: 10.1007/s10260-015-0342-7. URL `http://dx.doi.org/10.1007/s10260-015-0342-7`.

A. Voutilainen, T. Pyhälahti, K. Y. Kallio, J. Pulliainen, H. Haario, and J. P. Kaipio. A filtering approach for estimating lake water quality from remote sensing data. *International journal of applied earth observation and geoinformation*, 9(1):50–64, 2007.

R. T. Watson, I. R. Noble, B. Bolin, N. H. Ravindranath, D. J. Verardo, and D. J. Dokken. Land Use, Land-Use Change and Forestry. *IPCC - Special report of the Intergovernmental Panel on Climate Change*, 2000.

Deutscher Wetterdienst. Zeitreihe der Lufttemperatur in Deutschland. `http://www.dwd.de/DE/leistungen/zeitreihenundtrends/`, 14.07.2016.

C. K. Wikle and N. Cressie. A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829, 1999.

C. K. Wikle and M. B. Hooten. Hierarchical Bayesian spatio-temporal models for population spread. *Applications of computational statistics in the environmental sciences: hierarchical Bayes and MCMC methods*, pages 145–169, 2006.

T. Yokota, Y. Yoshida, N. Eguchi, Y. Ota, T. Tanaka, H. Watanabe, , and S. Maksyutov. Global concentrations of CO2 and CH4 retrieved from GOSAT: First preliminary results. *Sola*, 5:160–163, 2009.

Bohai Zhang, Huiyan Sang, and Jianhua Z Huang. Full-Scale Approximations of Spatio-Temporal Covariance Models for Large Datasets. *Statistica Sinica*, 2013.