Fernández-Duque, B., Pérez, I. A., García, M. Á., Pardo, N., & Sánchez, M. L. (2019). Annual and seasonal cycles of CO2 and CH4 in a Mediterranean Spanish environment using different kernel functions. *Stochastic Environmental Research and Risk Assessment, 33*(3), 915-930. https://doi.org/10.1007/s00477-019-01655-5

Annual and seasonal cycles of CO_2 and CH_4 in a Mediterranean Spanish environment using different kernel functions

Concise title: CO_2 and CH_4 cycles in a Mediterranean environment

Beatriz Fernández-Duque (Corresponding author). Dept. of Applied Physics. Faculty of Sciences. University of Valladolid, Paseo de Belén, 7. 47011 Valladolid, Spain
phone: 34 983 424 187, beatriz.fernandez.duque@uva.es

Isidro A. Pérez
Dept. of Applied Physics. Faculty of Sciences. University of Valladolid
Paseo de Belén, 7. 47011 Valladolid, Spain, iaperez@fa1.uva.es

M. Ángeles García
Dept. of Applied Physics. Faculty of Sciences. University of Valladolid
Paseo de Belén, 7. 47011 Valladolid, Spain, magperez@fa1.uva.es

Nuria Pardo
Dept. of Applied Physics. Faculty of Sciences. University of Valladolid
Paseo de Belén, 7. 47011 Valladolid, Spain, npardo@fa1.uva.es

M. Luisa Sánchez
Dept. of Applied Physics. Faculty of Sciences. University of Valladolid
Paseo de Belén, 7. 47011 Valladolid, Spain, marisa@fa1.uva.es

Abstract

This paper is based on CO_2 and CH_4 semi-hourly mole fraction measurements obtained at the Low Atmosphere Research Centre (CIB) between 2010 and 2016 using a Picarro G1301 analyser. The main aims of the study were to examine the temporal variation of CO_2 and CH_4 by using six different kernel functions, and to study the suitability of these functions to the dataset. The method used for the current study was based on experimental contour plots of R^2 values in order to simultaneously determine the bandwidths of kernel functions for the long-term and short-term. An Epanechnikov, a Gaussian, a biweight, a triangular, a tricubic and a rectangular kernel function were applied to extract the salient features of both the long-term (trend) and the short-term (seasonality). The average linear increase growth rates found were mainly attributed to the terrestrial biosphere cycle and changes in the atmospheric circulation regime. The seasonal cycle exhibited a cyclical variation, revealing summer minima for both gases, which may be explained by a biological minimum. Kernel analysis showed two nocturnal CO_2 maxima, in spring and autumn, linked to an increase in rainfall. For CO_2 daytime records, only the spring peak was detected. As regards CH_4, the maximum was located in winter. The best fit for the trend was obtained by the biweight kernel. In contrast, the best adjustment for seasonality was achieved from the Gaussian and the triangular kernel. To sum up, optimal bandwidth selection is important when kernel regression functions are employed. Since no important differences were found between the kernels employed, those which involve least computational effort are recommended.

Key words: trend; seasonal cycle; rural; kernel functions; bandwidth; contour plot

1 Introduction

CO_2 and CH_4 are the two most important long-lived greenhouse gases in the atmosphere that are controlled by the Kyoto Protocol. According to the IPCC (2013) report, radiative forcing of 2.62 W m^{-2} is caused by the sum of CO_2 and CH_4 emissions. The rise in CO_2 and CH_4 mole fractions is contributing to increasing radiative forcing, a measure of the change in the total radiative flux at the tropopause due to a rise in the amount of greenhouse gases in the atmosphere (IPCC 2014). Estimating the distribution of both gases over time is crucial vis-à-vis understanding the processes that govern their cycles, forecasting future trends and, therefore, planning effective CO_2 and CH_4 mitigation policies and strategies (Artuso et al. 2009; Wu et al. 2012).

High-precision and continuous measurements of greenhouse gases in the atmosphere were initiated in 1958 by Charles Keeling, who began measuring atmospheric CO_2 at the Mauna Loa Observatory, Hawaii (Keeling 1960). Systematic worldwide measurements of tropospheric CH_4 began in 1978 (Blake et al.1982; Dlugokencky et al. 1994; Rasmussen et al. 1981). Greenhouse gas observations are currently supported on a global scale by international databases such as the World Data Centre for Greenhouse Gases (WDCGG) and the National Oceanic and Atmospheric Administration (NOAA), among others (Pérez et al. 2016; 2017). However, observations are still lacking in some areas of the world, especially in rural sites in the Mediterranean basin (Artuso et al. 2009). The present study is motivated by the importance of gaining better insights into CO_2 and CH_4 behaviour in a Mediterranean rural area of the Iberian Peninsula.

From a biogeochemical point of view, three processes dominate the turnover of atmospheric CO_2. These are the exchange of atmospheric CO_2 with terrestrial

ecosystems, the exchange of atmospheric CO_2 with the oceans and the emissions of CO_2 into the atmosphere from anthropogenic activities (Zhu and Yoshikawa-Inoue 2015) such as fossil fuel use and industry (9.4 ± 0.5 Gt C $year^{-1}$ during 2007-2016) (Le Quéré et al. 2018) or land use change (1.3 ± 0.7 Gt C $year^{-1}$ during 2007-2016) (Le Quéré et al. 2018). It is also important to point to photosynthesis (3.0 ± 0.86 Gt C $year^{-1}$ during 2007-2016) (Le Quéré et al. 2018) and oceans (2.4 ± 0.5 Gt C $year^{-1}$ during 2007-2016) (Le Quéré et al. 2018) as being mainly responsible for CO_2 removal.

The main CH_4 sources estimated for the period 2003-2012 by the Global Methane Budget (Saunois et al. 2016) were fossil fuel (121 Tg CH_4 $year^{-1}$), enteric fermentation and manure (106 Tg CH_4 $year^{-1}$), landfills and waste water management (59 Tg CH_4 $year^{-1}$) as well as biomass and biofuel burning (30 Tg CH_4 $year^{-1}$). In contrast, the major CH_4 sink is oxidation by OH radicals (454-617 Tg CH_4 $year^{-1}$) (Saunois et al. 2016), especially in the tropical mid-troposphere, according to Nisbet et al. (2014). Minor sinks include soil oxidation (9-47 Tg CH_4 $year^{-1}$), and reactions in the stratosphere (16-84 Tg CH_4 $year^{-1}$) (Saunois et al. 2016).

Over the last few decades, the scientific community has focused its attention on assessing temporal variability in an effort to quantify the magnitude and rate of greenhouse gas sources and sinks. Atmospheric time variability, which comprise a long-term (trend) and a short-term (seasonality) series, consist of regular observations made at discrete time intervals. Decomposing series into their constituent parts involves fitting mathematical functions to the data (Pickers and Manning 2015). This paper focuses on applying kernel regression to decompose a CO_2 and CH_4 dataset.

Kernel regression is a non-parametric technique that locates and separates precise peaks that are close to one other, while producing simple graphical output (Donnelly et al.

2011). Rather than assuming a parametric form, a smooth regression function is assumed and the relationship between two variables can be summarized, thus giving a more accurate estimation at the expense of substantially increased computational effort (Donnelly et al. 2011).

Many authors have employed kernel functions to locate (e.g., Henry et al. 2002), identify (e.g., Kim and Hopke 2004) and separate (e.g., Henry et al. 2011) sources of air pollution by relating wind direction to measured mole fractions. Yu et al. (2004) extended the approach adopted by Henry et al. (2002) to incorporate wind speed into the analysis, quantifying the impact of local sources. Donnelly et al. (2011; 2012) followed the analysis by Yu et al. (2004) to quantify the variation of NO_2 mole fractions at three background Irish sites with wind direction and wind speed. Henry (2008) and Henry et al. (2009) determined the impact of sources using short-time average air quality and wind data in an extension of the methodology proposed by Henry et al. (2002) and Yu et al. (2004). Kernel regressions have already been applied at the CIB station to provide the CO_2 daily cycle taking into account pairs of circular variables such as wind direction and time (Pérez et al. 2013). Said authors also explored the trajectories reaching the northern plateau of the Iberian Peninsula employing kernel regressions (Pérez et al. 2015) and investigated point and linear sources of air pollution in the area (Pérez et al. 2016). However, most of these studies involving the application of kernel functions on time series of trace gases rely on the two most common functions: the Gaussian and the Epanechnikov. Hence, it is not known whether other functions may prove more suitable for fitting the data and expressing the salient features of the time series. As stated by Nakazawa et al. (1997), assessing a time series using one kernel function might be different from that derived using the same data but a different kernel function.

115 Bearing this in mind, the current paper involves two complementary parts. The first
116 seeks to determine the suitable bandwidths given a kernel function by using a novel
117 methodology based on contour plots. The second is devoted to showing how six
118 different kernel functions, whose application has been more restricted in atmospheric
119 time series up to date, can be used to characterize the evolution of CO_2 and CH_4 in
120 terms of trend, average growth rate and seasonal cycle.

121

122 **2 Materials and methods**

123 **2.1 Experimental site**

124 The measurement campaign extended over five and a half years (15 October 2010 to 29
125 February 2016). Measurements were recorded at the Low Atmosphere Research Centre
126 (CIB) (lat: 41°48′49″N, long: 4°55′59″W, alt: 845 m). The monitoring station is located
127 at the centre of an extensive plateau in a Mediterranean rural area of Spain, near the
128 cities of Valladolid (24 km to the southeast) and Palencia (40 km to the northeast).

129 Croplands, coniferous trees and Mediterranean shrubs provide the main vegetation
130 covering the area. There are no important orographic reliefs near the site. Thus,
131 horizontal homogeneity can be assumed. A typical Mediterranean climate prevails in the
132 study site with a mean annual temperature of 12.6 ºC and a mean annual rainfall of
133 around 450 mm with an evenly distributed pattern throughout the year, peaking in
134 spring and autumn (Sánchez et al. 2014).

135 **2.2 Data acquisition**

136 CO_2 and CH_4 mole fractions were routinely collected by a Cavity Ring-Down
137 Spectrometer (Picarro G1301 analyser) and averaged in half-hour intervals. This
138 analyser is based on optical technology in which direct measurement of infrared
139 absorption loss in a sample cell is used to quantify the mole fraction of a gas. Laser light
140 is directed into an optical resonator (the optical cavity) consisting of three highly
141 reflective mirrors, serving as a compact flow cell with an effective optical path length of
142 around 20 km. This long path allows for high precision measurements (Rella et al.
143 2013). The technique is described in greater detail by Crosson (2008) and Rella et al.
144 (2013).

145 The analyser was calibrated fortnightly with three NOAA standards, and mole fractions
146 were corrected slightly. Measurements were taken at three heights, 1.8, 3.7 and 8.3 m.
147 Only the highest level was used in this paper, since mole fractions changes are more
148 stable (Pérez et al. 2012).

149 The data series was divided into two periods: daytime and night-time, defined by the
150 hourly evolution (GMT hour) obtained from the National Geographic Institute of Spain.
151 The availability of a large database with scant gaps is a key point in this paper.

152 **2.3 Kernel regression method: theoretical concepts**

153 To define the temporal patterns of CO_2 and CH_4, we used kernel regression to extract
154 the trend and the seasonal component. The following subsections introduce the method.

155 **2.3.1 Kernel functions**

156 The fitting function used to calculate the average smoothed mole fractions, $\bar{y}$, expressed
157 in ppm for CO_2 and ppb for CH_4, is defined by the following expression proposed by
158 Henry et al. (2002):

159 $$\bar{y}(t,h) = \frac{\sum_{i=1}^{N} K\left(\frac{t-t_i}{h}\right) y_i}{\sum_{i=1}^{N} K\left(\frac{t-t_i}{h}\right)} \quad (1)$$

160 where t is the time expressed in days, h is the bandwidth expressed in days, N is the
161 number of observations, K is the kernel function, and y_i is the mole fraction (expressed
162 in ppm for CO_2 and ppb for CH_4) in a time period t_i expressed in days.

163 Greater weight is given to mole fractions y_i observed at times t_i closer to time t where
164 the average mole fraction $\bar{y}$ (t,h) is calculated, and less weight is attached to distant
165 observations, following Eq. (1).

166 This paper seeks to analyse the six smoothing kernel functions listed in Table 1.

167 **Table 1** Kernel functions employed in the study.

Kernel function	[a]K(u)	Support [u for which K(u) > 0]	Efficiency
Epanechnikov	$(3/4)\ (1-u^2)$	$-1 < u < 1$	1.0000
Biweight	$(15/16)\ (1-u^2)^2$	$-1 < u < 1$	0.9939
Gaussian	$(2\pi)^{-1/2}\exp\ (-\ 0.5u^2)$	$-\infty < u < \infty$	0.9512
Rectangular	1/2	$-1 < u < 1$	0.9295
Triangular	$1 - \lvert u\rvert$	$-1 < u < 1$	0.9859
Tricubic	$(70/81)\ (1 - \lvert u\rvert^3)^3$	$-1 < u < 1$	0.9870

168 [a]$u = [(t-t_i)/h]$

169

170 **2.3.2 Bandwidth (*h*)**

171 The most important decision in kernel regression is choosing the bandwidth, *h*. The
172 different methods for determining its value aim to balance two counteracting measures,
173 the bias (increases for large *h* values) and the variance (increases for small *h* values) (de
174 Haan 1999). There are two main lines of research in the literature for determining
175 bandwidth: one focuses on mathematical procedures and the other is based on
176 experimentation, bearing in mind the characteristics of the dataset. For the first line of
177 research, there is abundant literature (e.g., Hall and Kang 2005; Härdle 1993; Harrold et
178 al. 2001; Scott 1992; Silverman 1998). The second line of research is, however, much
179 poorer in terms of the literature. Authors such as Nicolich and Jorgensen (2008) suggest
180 experimenting with the bandwidth as the best method to determine its value (Donnelly
181 et al. 2011). For exploratory analysis, a trial and error approach is also recommended
182 (Wilks 2006).

183 The current paper follows the second line of research since the mathematical procedures
184 found in the literature determine only one bandwidth while two different bandwidths are
185 needed to express the trend and seasonal evolution of the dataset. We provide a quick
186 and easy method by depicting contour plots to determine the suitable bandwidths of a
187 kernel function. This is an important strength of the proposed method given the
188 decrease in the required computational effort, and which may have applications in other
189 studies. The Epanechnikov function was used in this part of the study due to its
190 simplicity.

Graven et al. (2012) considered a cut-off period of 24 months to detrend the data, while Pérez et al. (2017) employed a bandwidth of 1000 days to detrend the data, producing graphical outputs without oscillations. Both references as well as other combinations of training sets were taken into account when establishing the suitable bandwidth for CO_2 and CH_4.

Two different bandwidths were used; *h1* to quantify the data trend and *h2* to express the seasonal component. Large *h1* values combined with small *h2* values and vice-versa, as well as intermediate values of both bandwidths were analysed. Bandwidth values ranged between 100 and 1000 days (with intervals of 100 days) for *h1* and in the range of 20 days to 160 days for *h2*, with intervals of 20 days from level to level. R^2 values between experimental mole fractions and model measurements were obtained for each bandwidth combination and depicted in contour maps, inspection of which would determine the selection of specific *h1* and *h2* values.

By using R^2 values as the statistic to determine the bandwidths, the arbitrariness of the method is substantially reduced. When the lines are close to one another, they reflect major changes in R^2 values. In contrast, when they are further apart, the change is more gradual, and R^2 values are more stable within this region. Hence, the *h1* and *h2* values were chosen depending on the trend change (i.e. inflection point), defined as the region in which lines change from being closer to one another to being further apart. Thus, the inflection point was established in the region where the contour lines are further apart, indicating stable R^2 values. According to this method, a region of possible bandwidth combination could be suggested rather than a single unique point. Other authors, such as Rodríguez-Cortés et al. (2015), also considered an interval of optimal bandwidth values after testing different bandwidth combinations. Due to the narrow interval calculation

for *h1* (100 days) and *h2* (20 days), no major changes would be expected for alternative surrounding *h1* and *h2* values. Thus, intermediate values for *h1* and *h2* within the region where R^2 values barely change were chosen. Likewise, Grange et al. (2016) reported that within a central range the final output is quite insensitive to the scaling values. According to de Haan (1999), when an optimal bandwidth has been chosen for one type of kernel function, the same bandwidth for other kernels will show similar smoothing characteristics, and it is possible to switch between kernels without reconsidering the optimal bandwidth. Thus, once the bandwidth was established, six different kernel functions were applied to the dataset in order to analyse the salient features of the trend and the seasonal component and to analyse the data adjustment of each kernel.

2.3.3 Calculation procedure

Our time series has few missing data. The procedure used was based on Pickers and Manning (2015). First, the trend of the experimental data is calculated by applying the kernel function with a wide bandwidth, *h1*. Second, observations were detrended by subtracting the previously calculated trend from the whole time series. Third, the seasonal component obtained from the second step was smoothed by a kernel function, *K*, using a narrow bandwidth, *h2*, expressing seasonal changes. Finally, the seasonal cycle was subtracted from the original observations in a deseasonalised process, and the resulting series was again smoothed with the initial bandwidth, *h1*. Additionally, the average growth rate was obtained, expressing the trend change over time.

The Gaussian kernel is computationally slower than the other kernel functions both because of the exponential function calls, and because its infinite support leads to all data values contributing to the smoothed estimate (Wilks 2006). For the Gaussian kernel, observations in the range of (-1,1) receive a weight between 0.4-0.2, while

observations in the range of (-3,3) decrease their statistical weight to 0.0039, and are virtually 0 for the remaining data observations (Casas 2010). For this reason, the interval calculation was shortened to (-3,3) and (-1,1) intervals, making the Gaussian function more flexible in order to decrease the computational effort without losing accuracy in data fitting. R^2 values were used as the goodness of fit indicator to compare and analyse the efficacy of the kernel functions.

2.4 Statistics

Some of the most important statistics were determined for the data trend, since they report valuable information about their evolution over the study period. The statistics used here were classified into location, spread, symmetry and peakedness measures.

2.4.1 Location

Location statistics report information concerning the central tendency (Wilks 2006). The mean (non-robust) and the median (robust) were used, since they are widely employed in other environmental studies.

2.4.2 Spread

Spread measures indicate the dispersion or degree of variation around the central value (Wilks 2006). The standard deviation, minimum and maximum value were used as non-robust statistics. As a robust statistic, the interquartile range, the difference between the upper ($Q_{0.75}$) and lower ($Q_{0.25}$) quartile, was calculated.

2.4.3 Symmetry

Symmetry measures describe data distribution around their centre (Wilks 2006). In this paper, the skewness (non-robust) and the Yule-Kendall index (robust) were used to describe data symmetry. The Yule-Kendall index is defined as follows (Wilks 2006):

$$Yule\ Kendall\ index = (Q_{0.25} - 2Q_{0.5} + Q_{0.75})/IQ \quad (2)$$

where $Q_{0.25}$, $Q_{0.5}$ and $Q_{0.75}$ are the lower quartile, the median, and the upper quartile respectively. IQ represents the interquartile range.

2.4.4 Peakedness

Peakedness expresses data concentration around the central value (Khurshid et al. 2007). The Kurtosis statistic has been widely employed to analyse the peakedness of a univariate probability distribution (Khurshid et al. 2007). Here we use the standard kurtosis (non-robust) and the robust kurtosis based on Sachs' (1978) definition, which is as follows:

$$Robust\ Kurtosis = (Q_{0.75} - Q_{0.25})/(2(DZ_9 - DZ_1)) \quad (3)$$

where $Q_{0.75}$ and $Q_{0.25}$ are the upper and lower quartile respectively, while DZ_9 and DZ_1 refer to the 9th and the 1st decile mole fraction value (expressed in ppm for CO_2 and ppb for CH_4).

3 Results and discussion

3.1 Bandwidth choice

Figure 1 presents the contour plots following the procedure described in Section 2.3.2. As a result, the inflexion point for our dataset was established by choosing a 500-day bandwidth for *h1* and 80 days for *h2* (Fig. 1). *The h1* and *h2* values proposed here seem to be reasonable since *h1* is around one and a half years which, according to Barlow et al. (2015), is a good indicator to express growth rate, and *h2* (80 days) virtually represents one whole season. These values agree fairly well with the results of Pérez et al. (2017), who considered a bandwidth of 500 days for *h1* and 60 days for *h2*. The

same bandwidth was proposed for the four scenarios so as to simplify subsequent subsections, allowing comparisons of the outcomes in similar conditions. Furthermore, large values of *h* result in smoother shapes suppressing details of the time series, while small values result in more irregular shapes dominated by noise or large peaks, making it difficult to locate true peaks.

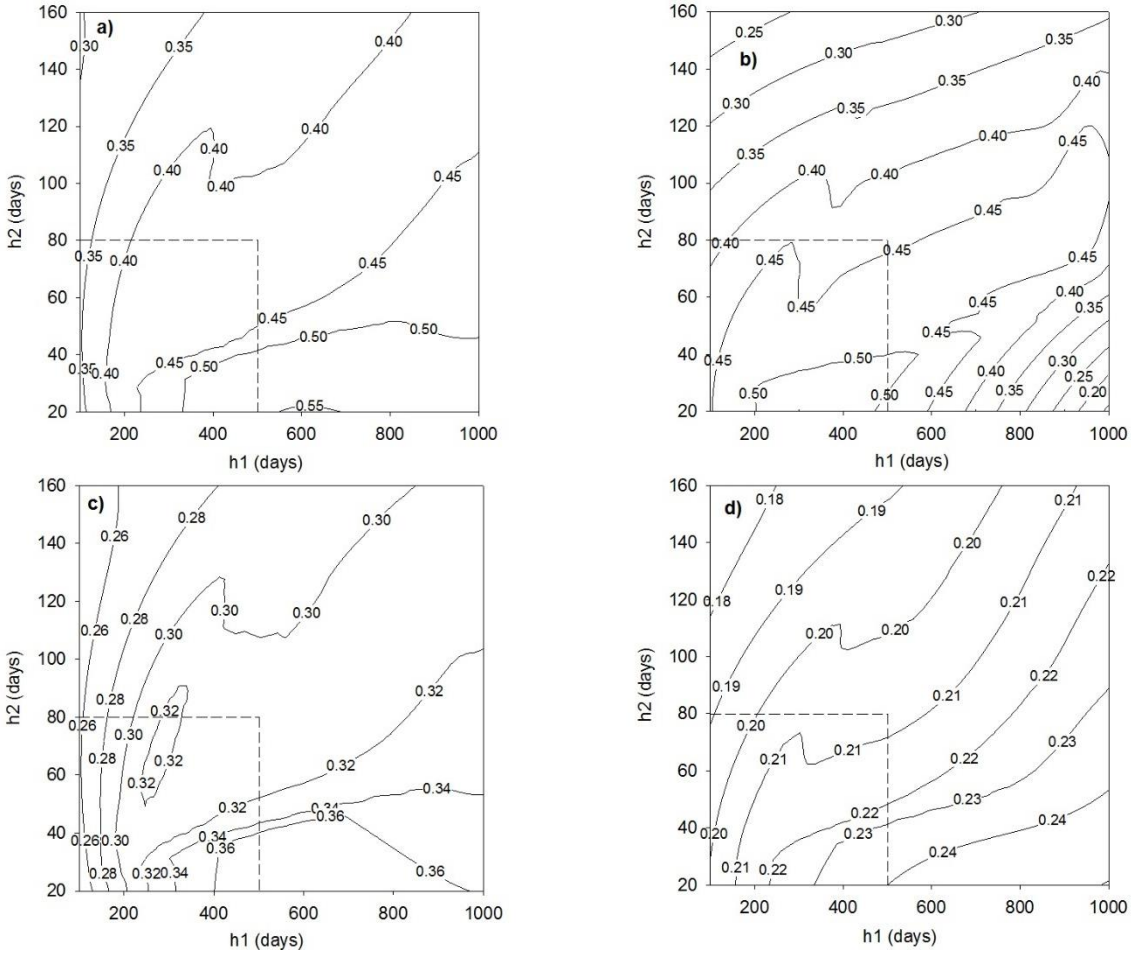


Fig. 1 Contour plots depicting the R^2 values of the average mole fractions for CO_2 daytime (a), CO_2 night-time (b), CH_4 daytime (c) and CH_4 night-time data (d). The dashed line corresponds to the bandwidth chosen for each scenario.

3.2 Analysis of temporal variations

3.2.1 Trend analysis

Figure 2 shows the CO_2 and CH_4 trend evolution (green line) and modelled results (red line). Monthly median observations (black solid dots) were considered instead of semi-hourly data in order to reduce the noise and influence of extreme values. This allows better data interpretation over time, without much information being lost. The 10th and 90th percentiles were also added in the graph in order to provide information about the monthly data range. Trend and modelled results were calculated by using the Epanechnikov Kernel due to its simplicity and widespread use (Rodríguez-Cortés et al. 2015).

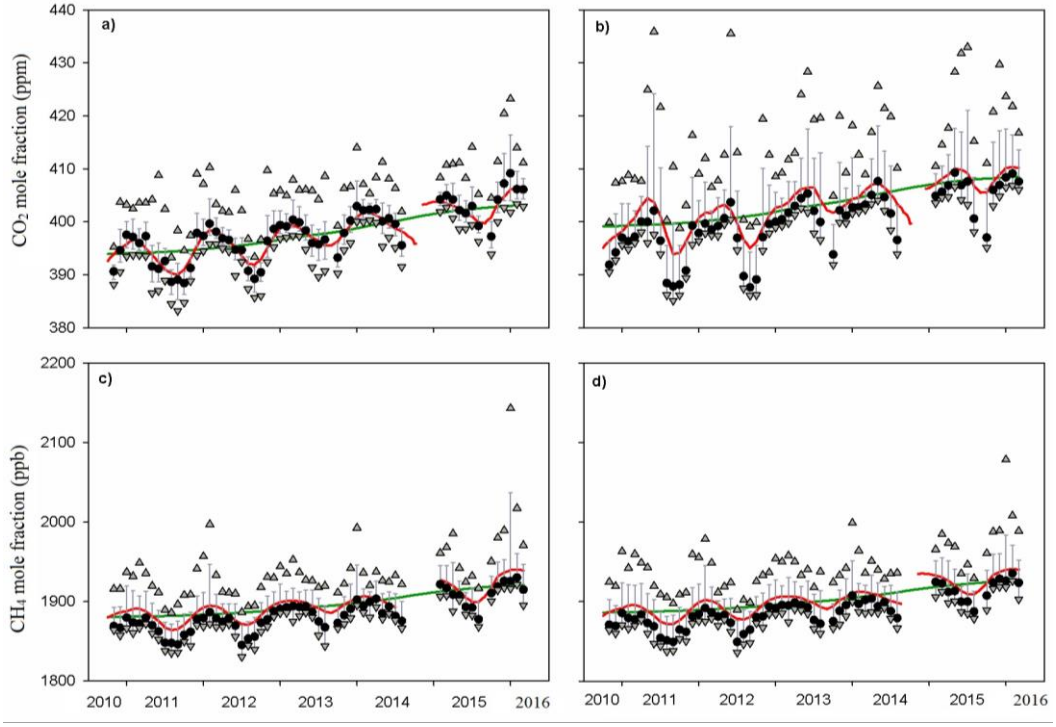


Fig. 2 CO_2 and CH_4 annual trend evolution and modelled results for daytime CO_2 (a), night-time CO_2 (b), daytime CH_4 (c) and night-time CH_4 data (d). Black solid dots represent monthly median observations and grey lines indicate the interquartile range. The upper triangles correspond to the 90th percentile and the lower triangles correspond to the 10th percentile values. Finally, the green line represents the annual trend that shows an upward pattern, whereas the red line depicts data distribution modelled by applying the Epanechnikov kernel considering all the observations.

312 Table 2 presents the main statistics for annual CO_2 and CH_4 trend values in order to

313 obtain a quantitative description of their evolution over the study period.

314 **Table 2** Summary statistics for CO_2 and CH_4 trend values over the study period.

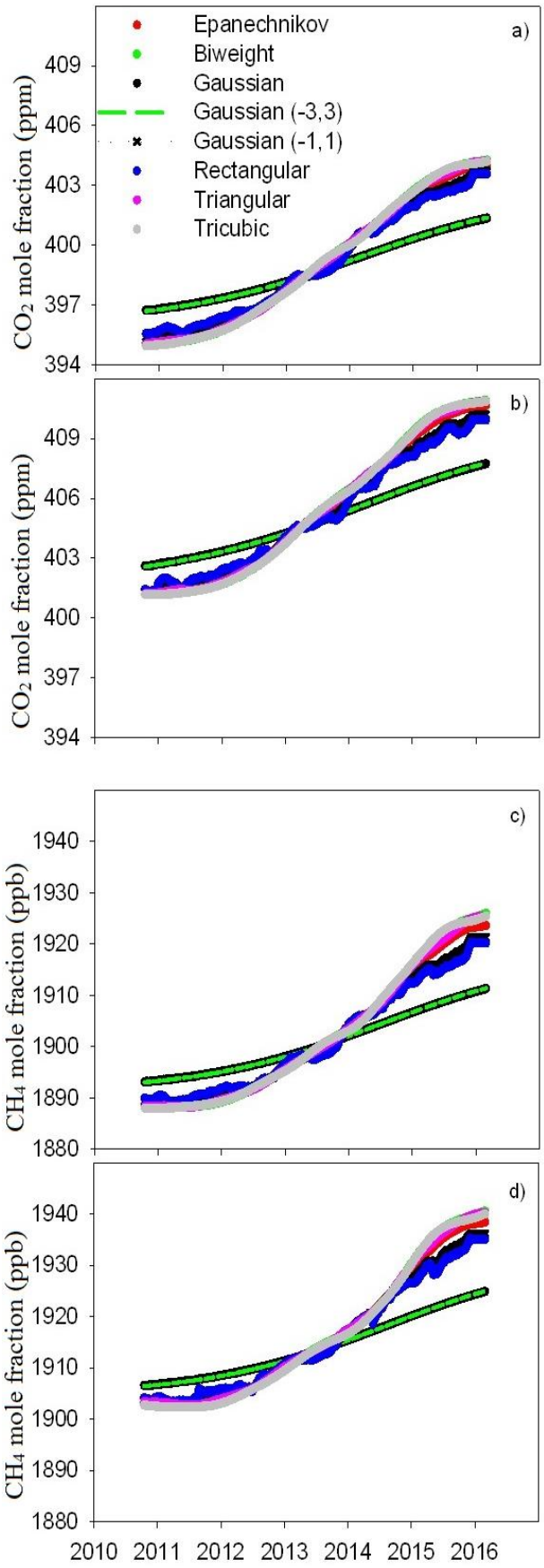
Year	2010*		2011		2012		2013		2014		2015		2016*	
CO_2 (ppm)	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**
Mean	395.16	401.20	395.51	401.50	396.74	402.75	398.86	405.07	401.26	407.62	403.35	409.99	403.99	410.63
Median	395.16	401.20	395.47	401.44	396.69	402.68	398.87	405.12	401.23	407.61	403.38	410.11	403.98	410.63
Stand. Dev.	0.02	0.00	0.20	0.18	0.52	0.54	0.63	0.64	0.75	0.79	0.43	0.46	0.01	0.02
Interq. range	0.04	0.01	0.30	0.26	0.87	0.86	1.04	1.03	1.33	1.34	0.77	0.77	0.02	0.03
Min	395.12	401.19	395.21	401.21	395.93	401.91	397.74	403.85	399.98	406.23	402.52	409.00	403.96	410.60
Max	395.21	401.22	395.94	401.91	397.74	403.85	399.98	406.23	402.52	409.01	403.96	410.60	404.01	410.67
Skewness	2.32	3.14	22.37	30.80	12.74	18.46	-0.84	-5.30	0.71	0.42	-11.01	-26.44	2.59	2.52
Yule-Kendall	0.05	-0.10	0.28	0.52	0.12	0.14	-0.04	-0.13	0.07	0.03	-0.05	-0.26	0.04	0.06
Stand. Kurt.	-14.12	6.69	-23.55	-19.55	-30.07	-26.25	-30.28	-27.76	-33.67	-31.16	-33.18	-26.36	-13.20	-12.85
Robust Kurt.	0.30	0.25	0.28	0.26	0.30	0.29	0.30	0.29	0.32	0.30	0.32	0.30	0.32	0.30
CH_4 (ppb)	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**
Mean	1888.40	1902.81	1889.01	1903.14	1892.40	1906.50	1899.15	1913.26	1908.93	1922.62	1919.90	1934.71	1923.45	1938.24
Median	1888.40	1902.81	1888.83	1902.91	1892.14	1906.16	1899.13	1913.30	1908.66	1922.25	1920.10	1935.29	1923.45	1938.23
Stand. Dev.	0.04	0.01	0.42	0.39	1.56	1.59	2.13	1.94	3.61	3.60	2.50	2.66	0.08	0.11
Interq. range	0.05	0.01	0.64	0.50	2.60	2.61	3.54	3.13	6.35	6.21	4.55	4.66	0.14	0.18
Min	1888.33	1902.78	1888.48	1902.82	1890.02	1904.16	1895.49	1909.77	1903.10	1916.94	1915.22	1929.26	1923.31	1938.06
Max	1888.49	1902.84	1890.03	1904.19	1895.50	1909.77	1903.11	1916.96	1915.22	1929.27	1923.32	1938.07	1923.60	1938.44
Skewness	8.23	20.70	45.97	67.80	18.64	23.87	4.18	2.10	6.54	11.25	-10.88	-23.21	2.54	2.45
Yule-Kendall	0.22	0.14	0.51	0.90	0.20	0.26	-0.01	-0.07	0.10	0.12	-0.05	-0.20	0.03	0.06
Stand. Kurt.	-7.67	19.14	-14.40	4.24	-28.71	-26.96	-30.29	-28.18	-33.38	-32.10	-34.40	-30.07	-13.04	-12.90
Robust Kurt.	0.26	0.22	0.30	0.26	0.30	0.30	0.30	0.29	0.32	0.31	0.33	0.32	0.31	0.30

*Refer to incomplete years measured. Year 2010 presents the results from 15th October to 31st December. Year 2016 presents the results from 1st January to 29th February.

315 According to the location measurements, both the mean and the median showed an
316 increased trend over time. Thus, an increase in the central data trend was inferred over
317 the study period. This upward trend can easily be explained on a global scale because of
318 the high annual increase in fossil fuel emissions in the northern temperate region (Piao
319 et al. 2017). On a smaller scale, the bulk of the Spanish economy is based on sectors
320 that emit large amounts of CO_2 to the atmosphere (Gutiérrez et al. 2008), which
321 partially explains the increasing CO_2 mean trend values presented in Table 2.
322 Furthermore, CH_4 emissions are the second main source of greenhouse gas emissions in
323 our country, as also reported Gutiérrez et al. (2008). Greater data dispersion was
324 observed from 2010 to 2014, since standard deviation and interquartile range were
325 higher during this time interval. Symmetry indicators, such as the Yule-Kendall index,
326 showed a general right-skewness distribution, which means that the distance between
327 the upper quartile and the median is greater than the distance between the median and
328 the lower quartile. Consequently, greater values than the median are predominant in the
329 dataset. As regards peakedness, almost negative standard kurtosis values were detected
330 for the data series, meaning that data distribution presents light tails and flatness when
331 compared to normal distribution (DeCarlo 1997). Robust kurtosis interpretation lies in
332 the same way. As pointed out by Sachs (1978), values above 0.263 were considered as
333 negative robust kurtosis. Negative robust kurtosis values were found (Table 2) for both
334 CO_2 and CH_4 in most years, indicating that data distribution for the study period
335 presents a wider peak (flatness) than normal distributions.

336 Analyses were performed by using the time as the independent variable, and mole
337 fractions as the dependent variable. The method depicts a smooth curve through the
338 points in a scatterplot by using Eq. (1). The kernel method produces smooth profiles
339 because the bandwidth plays the role of the smoothing parameter fitting the data (de

340 Haan 1999). The trend evolution of CO_2 and CH_4 was presented in Fig. 3. Smoother
341 curves were obtained with the Gaussian kernel, since all observations contribute to the
342 calculations. Furthermore, when it is not limited and when it is limited from (-3 to 3) the
343 Gaussian kernel presents the same graphical behaviour since the statistical weight is
344 nearly 0. Thus, both functions are virtually the same.


a)
b)
c)
d)
Epanechnikov
Biweight
Gaussian
Gaussian (-3,3)
Gaussian (-1,1)
Rectangular
Triangular
Tricubic
CO2 mole fraction (ppm)
CH4 mole fraction (ppb)
394
397
400
403
406
409
1880
1890
1900
1910
1920
1930
1940
2010
2011
2012
2013
2014
2015
2016

345

346 **Fig. 3** Trend fitting by different kernel functions. CO_2 daytime (a), CO_2 night-time (b),
347 CH_4 daytime (c) and CH_4 night-time (d) results.

348 The long-term trend of atmospheric CO_2 and CH_4 reflected an increased pattern over
349 time. From October 2010 to February 2016, CO_2 mean mole fractions at CIB were
350 around 399 ppm during the daytime and 405 ppm during the night-time. These values
351 were in consonance with global CO_2 values, which were around 395 ppm for the same
352 period (NOAA 2017a). As regards CH_4, daytime mole fractions were approximately
353 1.90 ppb and around 1.92 ppb during night-time. These results were in agreement with
354 CH_4 mole fractions values for latitudes higher than 30ºN, which were above 1.90 ppb in
355 2015 (Pérez et al. 2017), and with global CH_4 mole fraction values, which were around
356 1.8 ppb for the same study period (NOAA 2017b). These results were interpreted in
357 terms of average growth rate (section 3.2.2. Average growth rate analysis).

358 The accuracy of each kernel was judged by how close the estimated value, obtained
359 with kernel functions, is to the experimental data. R^2 values for the data trend were
360 calculated and presented in Table 3. It should be taken into account that all the r-critical
361 values (not presented here) were higher than the critical r-value (0.0003) for the dataset.
362 Therefore, r-values for all the kernels studied were statistically significant with a P-
363 value < 0.001. Thus, all the kernels are reliable and fit the data correctly, although some
364 of them fit better for the data trend according to the results in Table 3.

365 **Table 3** R^2 values for CO_2 and CH_4 data trend, applying a bandwidth of 500-days.

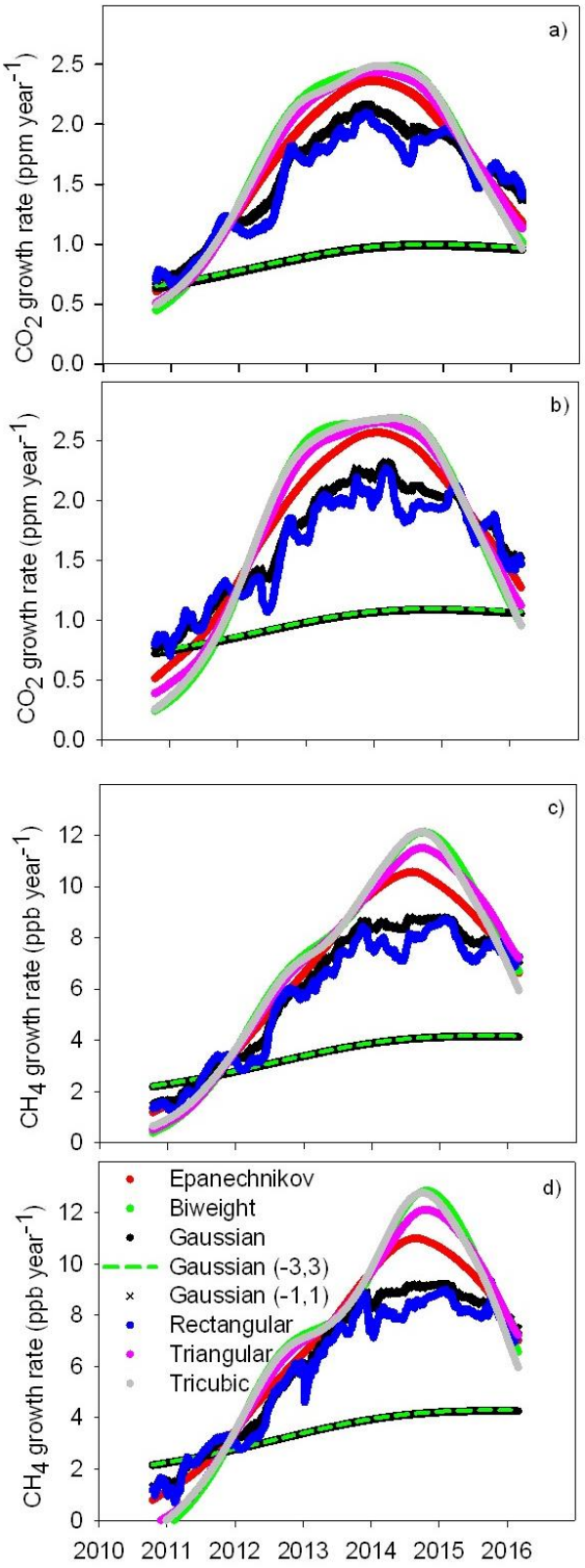
Kernel function	Daytime		Night-time	
	CO_2	CH_4	CO_2	CH_4
Epanechnikov	0.4360	0.2231	0.3708	0.1701

Biweight	0.4489	0.2329	0.3822	0.1760
Gaussian	0.3085	0.1554	0.2591	0.1180
Gaussian (-3,3)	0.3096	0.1559	0.2603	0.1185
Gaussian (-1,1)	0.4180	0.2113	0.3416	0.1628
Rectangular	0.4093	0.2059	0.3254	0.1599
Triangular	0.4461	0.2313	0.3803	0.1750
Tricubic	0.4473	0.2316	0.3812	0.1755

The average trend values agreed well for the six kernel functions employed (Table 3). The functions overlap their distributions in the graphical output. This evidence is in line with García-Portugués et al. (2014), since they consider minimal differences between different kernel functions used in distribution data. However, two groups could be inferred from Fig.3. A first group comprised the Gaussian and the Gaussian limited to (-3,3), and a second group was made up of the other kernel functions. Slightly better adjustments were obtained by applying the biweight kernel, and worse ones were obtained with the Gaussian function since it employs the total real number line. Similar R^2 values for the Gaussian and the Gaussian limited to (-3,3) could be inferred from Table 3. In contrast, R^2 values increase around 9% for CO_2 data and 5% for CH_4 when the Gaussian is limited to (-1,1) giving a better data fit, while computational time is cut by half. R^2 values were higher for the daytime (around 0.40 for CO_2 and 0.21 for CH_4) than during night-time (around 0.34 for CO_2 and 0.16 for CH_4). Moreover, CO_2 data fit is better than CH_4 data fit, in agreement with the R^2 values. These empirical results concur with those obtained by Rodríguez-Cortés et al. (2015), who advocate the importance of the kernel function chosen in the data fit.

3.2.2 Average growth rate analysis

383 The long-term trend is frequently discussed in terms of variations in growth rate. As
384 stated by Zhu and Yoshikawa-Inoue (2015), research into the growth rate requires a
385 long recording period. According to Barlow et al. (2015), periods of over a year and a
386 half are good indicators. In this case, a database of five and a half years has been
387 studied. Kernel smoothing average growth rates were presented in Fig. 4, showing three
388 different groups with regard to graphical behaviour. The same two groups as for the
389 data trend were found, although a third group formed by the rectangular and the
390 Gaussian limited to (-1,1) was also detected. The oscillatory behaviour of both functions
391 of this third group are partly caused by their lower efficiency (Table 1).



392

393 **Fig. 4** Average growth rate results for CO_2 daytime (a), CO_2 night-time (b), CH_4

394 daytime (c) and CH_4 night-time (d).

The overall CO_2 mean growth rate at CIB reached 1.72 ppm $year^{-1}$ (1.67 ppm $year^{-1}$ for the daytime and 1.77 ppm $year^{-1}$ for night-time). This value was in agreement with that found by Zhang et al. (2008) and Zhu and Yoshikawa-Inoue (2015), which were in the range of 1.7 – 3.6 ppm $year^{-1}$.

As regards CH_4, the mean growth rate was 6.68 ppb $year^{-1}$ (6.65 ppb $year^{-1}$ for the daytime and 6.71 ppb $year^{-1}$for night-time). This result was in line with others reported by Fang et al. (2016), Nisbet et al. (2014) and Vermeulen et al. (2011), which were in the range of 6-10 ppb $year^{-1}$, in line with an in-depth analysis carried out by Pérez et al. (2017).

Both gases presented an increasing linear rate with rapid growth from 2010 to 2014 and then a decreasing linear pattern until the end. The same behaviour was depicted for daytime and night-time records (Fig. 4). These increased long-term rates (trend and average growth rate) were mainly attributed to the rise in anthropogenic emissions from fossil fuel consumption that has increased globally from 1920 until the present (Le Quéré et al. 2018). Moreover, changes in the climatology perturb growth rate behaviour by shifts in the atmospheric circulation regime (Artuso et al. 2009). Climate affects vegetation carbon fluxes by modifying biosphere respiration and photosynthesis mechanisms (Artuso et al. 2009). Although global warming leads to a net positive feedback to the carbon balance (Artuso et al. 2009) its contribution would only explain a small part of the increasing trend pattern. Kim et al. (2015) also showed consistent changes in the global growth rate of annual CH_4 mole fractions, with ups and downs from the 1980s to 2010, although the reasons are not yet fully understood. The decreasing CH_4 pattern, mainly evidenced from 2015 to 2016, agrees with global CH_4 growth rate values, since a decrease of around 3 ppb $year^{-1}$ was obtained on a global

419 scale (NOAA 2017d). CO_2 global growth rate values (NOAA 2017c) also show a
420 decreasing pattern. However, the decreasing pattern reported on a global scale (NOAA
421 2017c) is less pronounced (around 0.1 ppm $year^{-1}$) than the CIB results. Local CO_2
422 emissions of anthropogenic activities in the surrounding area contribute to this value,
423 although the contribution is only small compared to global background values. The
424 main reason for this difference might stem from some border effect problems in the
425 kernel regression functions method.

426 Night-time results were slightly higher than daytime results. During the night, soil and
427 plant respiration together with stable stratification leads to higher values. By contrast,
428 during the day, CO_2 uptake by vegetation through photosynthesis, as well as turbulent
429 processes, cleans the lower atmosphere by convection.

430 Table 4 displays the average growth rate values for each kernel function used, showing
431 differences for the Gaussian kernel function.

432 **Table 4** Average growth rate values for CO_2 and CH_4 at CIB from October 2010 to
433 February 2016 applying a bandwidth of 500-days.

Kernel function	Daytime		Night-time	
	CO_2 (ppm $year^{-1}$)	CH_4 (ppb $year^{-1}$)	CO_2 (ppm $year^{-1}$)	CH_4 (ppb $year^{-1}$)
Epanechnikov	1.69	6.73	1.81	6.83
Biweight	1.75	7.18	1.83	7.13
Gaussian	0.88	3.45	0.97	3.48
Gaussian (-3,3)	0.89	3.48	0.99	3.51

Gaussian (-1,1)	1.58	6.09	1.69	6.25
Rectangular	1.52	5.76	1.62	5.92
Triangular	1.73	7.05	1.83	7.03
Tricubic	1.75	7.12	1.84	7.11

The range of average growth rate values, considering neither the Gaussian kernel, when limited to the interval calculation (-3, 3), nor when it is not limited, was narrow (Table 4). Both when limited to (-3,3) and when not limited, the Gaussian function showed very different results compared to the other kernel functions. However, when the Gaussian was limited to (-1,1), values doubled and approached those of the other functions, in addition to which the data fit improved.

3.3.3 Analysis of the seasonal cycle

The kernel estimation of the cycle provides detailed information about changes in the intra-annual evolution. Figures 5 and 6 present the seasonal evolution of CO_2 and CH_4, respectively. In this case, a more similar graphical output was detected (Fig. 5 and 6), which might be linked to smaller variability in the seasonal component.

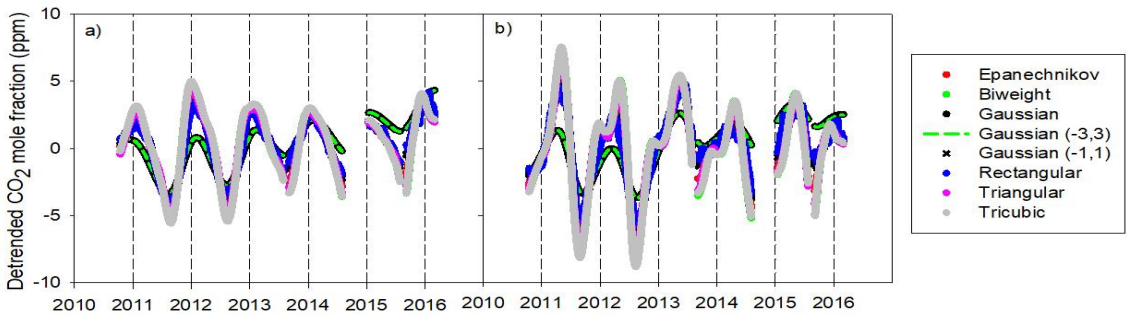
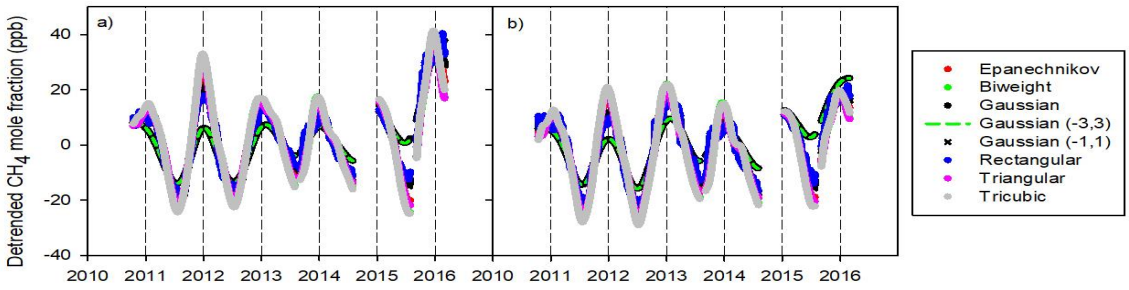


Fig. 5 Smoothed detrended CO_2 seasonal cycle for the six kernel functions: CO_2 daytime (a) and CO_2 night-time (b) results.

448 The CO_2 seasonal cycle (Fig. 5) is mainly dominated by the physiological activity of
449 plants and soil respiration. The CO_2 daytime cycle (Fig. 5a) was characterized by a
450 decline from December to August and a subsequent rise. CO_2 mole fractions decrease
451 more rapidly from spring due to more intense photosynthetic activity, reaching the
452 minimum in August (-5.57 ppm). The site is covered by agricultural crops from autumn
453 to spring. However, coniferous trees and Mediterranean shrubs surrounding the site are
454 particularly vigorous from May to July, promoting photosynthesis as the days grow
455 longer. The peak in December (4.97 ppm) might be due to the cumulative effects of
456 CO_2 emission via root respiration as suggested Wu et al. (2012). This pattern was
457 consistent with those of many northern hemispheric locations (e.g., Wu et al. 2012;
458 Zhang et al. 2008).

459 Nocturnal CO_2 mole fractions (Fig. 5b) increased from August to April, and then
460 strongly decreased from May to August. This cycle was in line with that reported by
461 Barichivich et al. (2013). From Fig. 5b, two nocturnal maxima could be inferred. The
462 first was detected in spring (late April: 7.52 ppm) and the second in autumn (late
463 November: 1.88 ppm). Both maxima were linked to an increase in rainfall during spring
464 and autumn contributing to vegetation growth at the site, thus increasing respiration
465 rates at night-time. A cycle with two maxima was also described by Hatakka et al.
466 (2003) for CO_2 mole fractions at Pallas, Finland, from 1997 to 2003 (Pérez et al. 2017).
467 The minimum was reached in August (-8.81 ppm) linked to a biological minimum
468 attributed to slight plant and soil agricultural activities, decreasing the respiration rate.

469 Figure 6 showed a simpler cycle for CH_4 data, revealing positive values from summer to
470 winter, and negative ones from winter to summer, both during daytime and night-time
471 records.



472

473 **Fig. 6** Smoothed detrended CH_4 seasonal cycle for the six kernel functions: CH_4

474 daytime (a) and CH_4 night-time (b) results.

475 Meteorological conditions play an important role in the CH_4 cycle, since higher

476 temperatures induce greater levels of OH radicals, which are the main CH_4 natural sink.

477 Thus, the CH_4 maximum was reached in December (41.30 ppb for the daytime and

478 24.20 ppb for night-time), while the CH_4 minimum was detected in July (-25.00 ppb for

479 the daytime and -29.00 ppb for night-time). This cycle was in line with that observed at

480 the Mauna Loa Observatory (MLO), although mole fractions at CIB were higher than

481 those at MLO due to the greater sensitivity from emission sources at CIB. Furthermore,

482 OH radical concentration decreases with latitude in the northern hemisphere (Kim et al.

483 2015). MLO is located closer to the equator, such that the photochemical effect is

484 stronger causing lower CH_4 mole fractions than at the CIB station.

485 CO_2 and CH_4 seasonal cycles reflected periodic behaviour and regular variations with

486 small interannual variability. Differences between day and night were observed for both

487 gases (Fig. 5 and 6), showing higher mole fractions during night-time. This was due to

488 many superimposed causes: firstly, because of the meteorological conditions that

489 favoured strong thermal temperature inversions at night, helping to trap CO_2 and CH_4

490 emissions between the ground and the top of the boundary layer, and increasing their

491 mole fraction values. Second, stable stratification at night together with the low

492 boundary layer height at the site, 405 m, dampens vertical dispersion and transport from

493 the nearest cities, favouring the accumulation of locally-emitted pollutants. In contrast,
494 during the daytime, thermal turbulence induces convective movements, as well as the
495 expansion of the boundary layer, increasing up to 809 m at CIB, contributing to the
496 dispersion of both gases due to vertical mixing.

497 Atmospheric transport flows should also be considered when seeking to understand
498 seasonal patterns. Artuso et al. (2009) found an important influence of industrialized
499 Western Europe at the Lampedusa station from 1992 to 2007 in winter, when the
500 vegetation sink is less effective. The same origin might be attributed to our sampling
501 site, since García et al. (2016) reported that the main pollutant sources lie in Europe,
502 explaining the higher CO_2 mole fractions during this season. Furthermore, the influence
503 of fossil fuel consumption is greater during the cold months, enhancing the mole
504 fraction values. By contrast, clean air masses approaching the monitoring station from
505 the Atlantic are more frequent in summer (García et al. 2016), explaining CO_2 and CH_4
506 troughs during this season. High values of both CO_2 and CH_4 may also be related with
507 transport from the nearby cities of Valladolid (SE) and Palencia (NE) (Pérez et al.
508 2013). Additionally, atmospheric CH_4 mole fractions increase when air masses
509 originating from the southeast sector, where an urban landfill is located, reach the CIB
510 station (García et al. 2016). Finally, it should not be forgotten that the CIB station is
511 located in a region where livestock plays an important role (García et al. 2016; Sánchez
512 et al. 2014). Sánchez et al. (2014) reported almost 200 kt $year^{-1}$ CH_4 emissions due to
513 livestock in the region, which partly explains the higher CH_4 values at the CIB station.
514 An irregular amplitude evolution is inferred from Fig.5. This irregular pattern is
515 explained by the mathematical fitting method used in the current paper. Kernel
516 regression functions are based on local calculations, since only data within the
517 established bandwidth are considered. In fact, when harmonic equations were applied

518 to the same dataset a more regular and upward trend over time was obtained
519 (Fernández-Duque et al. 2017). Furthermore, kernel regressions are more sensitive to
520 gaps since they are only based on the values inside the bandwidth whereas harmonic
521 equations consider the whole dataset for the calculations. From Fig. 5, summer minima
522 increase over time was inferred whereas maxima mole fractions decrease over time. The
523 CO_2 mean amplitude during the daytime was 7.81 ppm and was 11.45 ppm during
524 night-time. These results were in agreement with those reported by Zhang et al. (2008)
525 at the Waliguan (10.9 ppm) and Lin-an (11.5 ppm) stations. As regards CH_4 peak-to-
526 peak amplitude (Fig. 6), a decreasing pattern over time was noticed although no major
527 differences between daytime (41.54 ppb) and night-time (41.75 ppb) amplitude values
528 were detected.

529 As for R^2 values for the seasonal component, they showed slight differences among the
530 kernel functions employed (Table 5). As for the trend, r-values for all the kernels were
531 statistically significant with a P-value < 0.001, such that all the kernels fit the dataset
532 correctly. However, some differences in R^2 values were found among the kernels
533 employed (Table 5).

534 **Table 5** R^2 values for the CO_2 and CH_4 seasonal component applying a bandwidth of
535 80-days.

Kernel function	Daytime		Night-time	
	CO_2	CH_4	CO_2	CH_4
Epanechnikov	0.4236	0.2992	0.4473	0.2085
Biweight	0.4335	0.3053	0.4627	0.2128
Gaussian	0.4489	0.3041	0.3934	0.2195

Gaussian (-3,3)	0.4483	0.3037	0.3933	0.2192
Gaussian (-1,1)	0.4068	0.2874	0.4219	0.2002
Rectangular	0.3936	0.2789	0.3961	0.1932
Triangular	0.4328	0.3058	0.4671	0.2172
Tricubic	0.4314	0.3041	0.4558	0.2102

The Gaussian kernel provided the best description of the seasonal cycle for CO_2 daytime (0.45) and night-time CH_4 data (0.22), whereas a triangular function provided a better fit for daytime CH_4 (0.31) and night-time CO_2 (0.47) results. In contrast, the rectangular function provided the worst data adjustment, except in the case of the CO_2 nocturnal record, where the worst fit was obtained by using the Gaussian function limited to (-3,3). Slight differences (3.5% for CO_2 data and 2% for CH_4) were found when the Gaussian interval calculation was limited to (-1,1). This showed that limiting the interval calculation did not prove as effective as for the trend with regard to the seasonal component limit.

R^2 values fell to about 0.43 for CO_2 observations and to about 0.25 for CH_4 results, indicating a better fit for CO_2 data than for CH_4 data, in agreement with the results presented by Pérez et al. (2017).

No major differences were observed between daytime and night-time CO_2 results as regards R^2 (Table 5). However, CH_4 daytime results (0.30) were slightly better than night-time results (0.21).

4 Conclusions

553 This paper has demonstrated the usefulness of kernel regression for describing the
554 temporal variability of CO_2 and CH_4 in a Mediterranean rural area of the Spanish
555 plateau between 2010 and 2016.

556 We draw the following conclusions from our study:

557 (a) A novel procedure based on contour plots and its interpretation was proposed to
558 determine two optimal bandwidths at the same time. This methodology is easy,
559 quick and reproducible; thus its general use may be plausible.
560 (b) Temporal patterns were similar for the six kernel functions employed, obtaining
561 satisfactory R^2 values, since r-values for all the kernels used were higher than
562 the r-critical values for our specific dataset.
563 (c) As regards the data trend, the best fit was reached with the biweight kernel and
564 the worst with the Gaussian because its calculation comprises the whole real
565 line.
566 (d) For seasonal evolution, the triangular and the Gaussian kernel assumed the role
567 as best predictors. In contrast, the worst fits were obtained with the rectangular,
568 which could be attributed to a border effect.

569 (e) The differences between the kernels for the seasonal component were lower than
570 those for the trend, which could be due to slight seasonal data variability. The
571 Gaussian kernel presented a clearly different graphical behaviour because its
572 calculation comprises the whole real line.

573 (f) Additional run-time was needed when the Gaussian kernel was used since it is
574 involved in all observations. Shortening the interval calculation to (-3,3) did not
575 prove successful either for a decrease in calculation speed or for an increase in
576 R^2 values. However, when the interval calculation was limited to (-1,1), R^2

values increased substantially for the trend, and time calculation was cut by half. As regards the seasonal component, R^2 did not improve although average growth rate values performed far better and approached the values obtained with the other kernel functions. Thus, limiting the interval calculation to (-1,1) significantly increased the flexibility of the Gaussian kernel, making the calculation computationally feasible.

(g) Linear growth rates (1.72 ppm $year^{-1}$ for CO_2 and 6.68 ppb $year^{-1}$ for CH_4) were recorded in the research period. Anthropogenic emissions from fossil fuel consumption and other industrial activities on a global scale contributed to the notable increase.

(h) The combined effects of biological and physical processes resulted in differences between day and night-time for both CO_2 and CH_4 seasonal cycles. Lower mole fractions during the daytime were found due to thermal turbulence and higher planetary boundary layer height. Higher values at night were linked to stable stratification, lower boundary planetary height and thermal inversions. Focusing on CO_2, the influence of the terrestrial ecosystem was apparent, contributing to lower values during the daytime due to photosynthesis and higher values at night-time due to soil and plant respiration.

(i) Apart from the biological effects, during the winter, European airmasses might be linked to higher CO_2 mole fraction values in winter due to large industrial emissions. In summertime, the effect of masses from the Atlantic prevailed, which explains the lower CO_2 and CH_4 mole fractions. An urban influence in the SE direction was observed for the CH_4 results, accounting for the higher values.

600 Selecting the optimal bandwidth proves key in kernel regressions if good data
601 interpretation is to be achieved. Thus, we recommend choosing the bandwidth based on
602 the special characteristics of the time series, and not based on the typical values used in
603 the literature. Further inquiry is needed to improve the methodology described here for a
604 more precise bandwidth determination. As regards the kernel functions employed, no
605 major differences affecting temporal pattern interpretation were found between the
606 different kernels. Moreover, since R^2 values were very similar for the six kernels, we
607 recommend using those which involve least computational effort. All the kernels
608 employed have virtually the same computational effort, except the Gaussian when it is
609 not limited to (-1,1). As a result, we recommend shortening the calculation interval.
610 However, it could be interesting to use more than one kernel function so as to ensure
611 results are consistent, reproducible, and free from bias. Finally, we consider it important
612 to extend the use of these functions to other areas and to other important greenhouse
613 gases. Their application may improve current knowledge of temporal patterns and help
614 gain a better insight into how gases evolve in the low atmosphere.

615

616 **Acknowledgements**

617 The authors of this paper acknowledge financial support from the Spanish Ministry of
618 Economy and Competitiveness and ERDF funds (projects CGL2009-11979 and
619 CGL2014-53948-P). The authors sincerely thank the editor and the two anonymous
620 referees for their valuable comments and remarks.

621 **References**

622 Artuso F, Chamard P, Piacentino S, Sferlazzo DM, De Silvestri L, di Sarra A, Meloni
623 D, Monteleone F (2009) Influence of transport and trends in atmospheric CO_2 at
624 Lampedusa. Atmos Environ 43:3044-3051. doi: 10.1016/j.atmosenv.2009.03.027.

625 Barichivich J, Briffa KR, Myneni RB, Osborn TJ, Melvin TM, Ciais P, Piao S, Tucker
626 C (2013) Large-scale variations in the vegetation growing season and annual cycle of
627 atmospheric CO_2 at high northern latitudes from 1950 to 2011. Glob Change Biol
628 19:3167-3183. doi: 10.1111/gcb.12283.

629 Barlow JM, Palmer PI, Bruhwiler LM, Tans P (2015) Analysis of CO_2 mole fraction
630 data: first evidence of large-scale changes in CO_2 uptake at high northern latitudes.
631 Atmos Chem Phys 15:13739-13758. doi: 10.5194/acp-15-13739-2015.

632 Blake DR, Mayer EW, Tyler SC, Makide Y, Montague DC, Rowland FS (1982) Global
633 increase in atmospheric methane concentrations between 1978 and 1980. Geophys Res
634 Lett 9:477-480. doi: org/10.1029/GL009i004p00477.

635 Casas G (2010) Estimaciones vía kernel.
636 https://es.scribd.com/doc/57531637/Estimaciones-via-Kernel (accessed 5.5.17).

637 Crosson ER (2008) A cavity ring-down analyzer for measuring atmospheric levels of
638 methane, carbon dioxide, and water vapor. Appl Phys B 92:403-408. doi:
639 10.1007/s00340-008-3135-y.

640 DeCarlo LT (1997) On the meaning and use of kurtosis. Psychol Methods 2:292–307.

641 de Haan P (1999) On the use of density kernels for concentration estimations within
642 particle and puff dispersion models. Atmos Environ 33:2007-2021. doi: 10.1016/S1352-
643 2310(98)00424-5.

644 Donnelly A, Misstear B, Broderick B (2011) Application of nonparametric regression
645 methods to study the relationship between NO_2 concentrations and local wind direction
646 and speed at background sites. Sci Total Environ 409(6):1134-1144. doi:
647 10.1016/j.scitotenv.2010.12.001.

648 Donnelly A, Broderick B, Misstear B (2012) Relating background NO_2 concentrations
649 in air to air mass history using non-parametric regression methods: Application at two
650 background sites in Ireland. Environ Model Assess 17:363-373. doi: 10.1007/s10666-
651 011-9301-3.

652 Dlugokencky EJ, Steele LP, Lang PM, Masarie A (1994) The growth rate and
653 distribution of atmospheric methane. J Geophys Res 99:17021-17043. doi:
654 org/10.1029/94JD01245.

655 Fang SX, Tans PP, Dong F, Zhou H, Luan T (2016) Characteristics of atmospheric CO_2
656 and CH_4 at the Shangdianzi regional background station in China. Atmos Environ
657 131:1-8. doi: 10.1016/j.atmosenv.2016.01.044.

658 Fernández-Duque B, Pérez IA, Sánchez ML, García MÁ, Pardo N (2017) Temporal
659 patterns of CO_2 and CH_4 in a rural area in northern Spain described by a harmonic
660 equation over 2010-2016. Sci Total Environ 593-594:1-9. doi:
661 10.1016/j.scitotenv.2017.03.132.

662 García MÁ, Sánchez ML, Pérez IA, Ozores MI, Pardo N (2016) Influence of
663 atmospheric stability and transport on CH_4 concentrations in northern Spain. Sci Total
664 Environ 550:157-166. doi: 10.1016/j.scitotenv.2016.01.099.

García-Portugués E, Barros AMG, Crujeiras RM, González-Manteiga W, Pereira J (2014) A test for directional-linear independence, with applications to wildfire orientation and size. Stoch Environ Res Risk Assess 28:1261–1275. doi: 10.1007/s00477-013-0819-6.

Grange SK, Lewis AC, Carslaw DC (2016) Source apportionment advances using polar plots of bivariate correlation and regression statistics. Atmos Environ 145:128-134. doi: 10.1016/j.atmosenv.2016.09.016.

Graven HD, Guilderson TP, Keeling RF (2012) Observations of radiocarbon in CO_2 at La Jolla, California, USA 1992-2007: analysis of the long-term trend. J Geophys Res Atmos 117, D02302. doi: 10.1029/2011JD016533.

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A (2008) Trend analysis using nonhomogeneous stochastic diffusion processes. Emission of CO_2; Kyoto protocol in Spain. Stoch Environ Res Risk Assess 22:57–66. doi: 10.1007/s00477-006-0097-7.

Hall P, Kang KH (2005) Bandwidth choice for nonparametric classification. Ann Stat 33:284-306. doi: 10.1214/009053604000000959.

Härdle W (1993) Applied nonparametric regression. Cambridge University Press, Cambridge.

Harrold TI, Sharma A, Sheather S (2001) Selection of a kernel bandwidth for measuring dependence in hydrologic time series using the mutual information criterion. Stoch Environ Res Risk Assess 15:310–324. doi: 10.1007/s004770100073.

685 Hatakka J, Aalto T, Aaltonen V, Aurela M, Hakola H, Komppula M, Laurila T,
686 Lihavainen H, Paatero J, Salminen K, Viisanen Y (2003) Overview of the atmospheric
687 research activities and results at Pallas GAW station. Boreal Environ Res 8:365-383.

688 Henry RC, Chang YS, Spiegelman CH (2002) Locating nearby sources of air pollution
689 by nonparametric regression of atmospheric concentrations on wind direction. Atmos
690 Environ 36:2237-2244. doi: 10.1016/S1352-2310(02)00164-4.

691 Henry RC (2008) Locating and quantifying the impact of local sources of air pollution.
692 Atmos Environ 42:358-363. doi: 10.1016/j.atmosenv.2007.09.039.

693 Henry R, Norris GA, Vedantham R, Turner JR (2009) Source region identification
694 using kernel smoothing. Environ Sci Technol 43:4090-4097. doi: 10.1021/es8011723.

695 Henry RC, Vette A, Norris G, Vedantham R, Kimbrough S, Shores RC (2011)
696 Separating the air quality impact of a major highway and nearby sources by
697 nonparametric trajectory analysis. Environ Sci Technol 45:10471-10476. doi:
698 10.1021/es202070k.

699 IPCC, 2013. Summary for Policymakers. In: Climate Change 2013: The Physical
700 Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the
701 Intergovernmental Panel on Climate Change [Stocker TF, Qin D, Plattner GK, Tignor
702 M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.)]. Cambridge
703 University Press, Cambridge, United Kingdom and New York, NY, USA. pp. 29.

704 IPCC, 2014. Climate Change 2014: Mitigation of Climate Change. Contribution of
705 Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on
706 Climate Change [Edenhofer O, Pichs-Madruga R, Sokona Y, Farahani E, Kadner S,

Seyboth K, Adler A, Baum I, Brunner S, Eickemeier P, Kriemann B, Savolainen J, Schlömer S, von Stechow V, Zwickel T, Minx JC (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. pp. 1435.

Keeling CD (1960) The concentration and isotopic abundances of carbon dioxide in the atmosphere. Tellus 12:200-203. doi: org/10.1111/j.2153-3490.1960.tb01300.x.

Khurshid A, Hussain E, ul-Haq M (2007) A note on finding peakedness in bivariate normal distribution using Mathematica. Pak J Stat Oper Res 3:75-86. doi: 10.18187/pjsor.v3i2.61.

Kim E, Hopke PK (2004) Comparison between conditional probability function and nonparametric regression for fine particle source directions. Atmos Environ 38:4667-4673. doi: 10.1016/j.atmosenv.2004.05.035.

Kim HS, Chung YS, Tans PP, Dlugokencky EJ (2015) Decadal trends of atmospheric methane in East Asia from 1991 to 2013. Air Qual Atmos Health 8:293-298. doi: 10.1007/s11869-015-0331-x.

Le Quéré C, Andrew R M, Friedlingstein P, Sitch S, Pongratz J, Manning AC, Korsbakken JI, Peters GP, Canadell JG, Jackson RB, Boden TA, Tans PP, Andrews OD, Arora VK, Bakker DCE, Barbero L, Becker M, Betts RA, Bopp L, Chevallier F, Chini LP, Ciais P, Cosca CE, Cross J, Currie K, Gasser T, Harris I, Hauck J, Haverd V, Houghton RA, Hunt CW, Hurtt G, Ilyina T, Jain AK, Kato E, Kautz M, Keeling RF, Klein Goldewijk K, Körtzinger A, Landschützer P, Lefèvre N, Lenton A, Lienert S, Lima I, Lombardozzi D, Metzl N, Millero F, Monteiro PMS, Munro DR, Nabel JEMS, Nakaoka S-I, Nojiri Y, Padin XA, Peregon A, Pfeil B, Pierrot D, Poulter B, Rehder G, Reimer J, Rödenbeck C, Schwinger J, Séférian R, Skjelvan I, Stocker BD, Tian H,

730 Tilbrook B, Tubiello FN, van der Laan-Luijkx IT, van der Werf GR, van Heuven S,
731 Viovy N, Vuichard N, Walker AP, Watson AJ, Wiltshire AJ, Zaehle S, Zhu, D (2018)
732 Global Carbon Budget 2017. Earth Syst Sci Data 10:405-448. doi: 10.5194/essd-10-
733 405-2018.

734 Nakazawa T, Ishizawa M, Higuchi K, Trivett NBA (1997) Two curve fitting methods
735 applied to CO_2 flask data. Environmetrics 8:197-218. doi: 10.1002/(SICI)1099-
736 095X(199705)8:3<197::AID-ENV248>3.0.CO;2-C.

737 Nicolich M, Jorgensen G (2008) Graphical presentation of a nonparametric regression
738 with bootstrapped confidence intervals. http://103.28.21.22/Record/IOS10-
739 oai:CiteSeerX.psu:10.1.1.124.889. Accessed 30 April 2017.

740 Nisbet EG, Dlugokencky EJ, Bousquet P (2014) Methane on the Rise - Again. Atmos
741 Sci 343:493-495. doi: 10.1126/science.1247828.

742 NOAA (2017a) CO_2 globally averaged marine surface monthly mean data.
743 ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_gl.txt. Accessed 25 April 2017.

744 NOAA (2017b) CH_4 globally averaged marine surface monthly mean data.
745 ftp://aftp.cmdl.noaa.gov/products/trends/ch4/ch4_mm_gl.txt. Accessed 25 April 2017.

746 NOAA (2017c) Annual Mean Global Carbon Dioxide Growth Rates
747 https://www.esrl.noaa.gov/gmd/ccgg/trends/gl_gr.html. Accessed 15 May 2018.

748 NOAA (2017d) Annual Increase in Globally-Averaged Atmospheric Methane
749 https://www.esrl.noaa.gov/gmd/ccgg/trends_ch4/#global_growth. Accessed 15 May
750 2018.

751 Pérez IA, Sánchez ML, García MÁ, Pardo N (2012) Analysis of CO_2 daily cycle in the
752 low atmosphere at a rural site. Sci Total Environ 431:286-292. doi:
753 10.1016/j.scitotenv.2012.05.067.

754 Pérez IA, Sánchez ML, García MÁ, Pardo N (2013) Carbon dioxide at an unpolluted
755 site analysed with the smoothing kernel method and skewed distributions. Sci Total
756 Environ 456-457:239-245. doi: 10.1016/j.scitotenv.2013.03.075.

757 Pérez IA, Sánchez ML, García MÁ, Pardo N (2015) Daily patterns of CO_2 in the lower
758 atmosphere of a rural site. Theor Appl Climatol 122:195-205. doi: 10.1007/s00704-014-
759 1294-9.

760 Pérez IA, Sánchez ML, García MÁ, Pardo N (2016) Features of the annual evolution of
761 CO_2 and CH_4 in the atmosphere of a Mediterranean climate site studied using a
762 nonparametric and a harmonic function. Atmos Pollut Res 7:1013-1021. doi:
763 10.1016/j.apr.2016.06.006.

764 Pérez IA, Sánchez ML, García MÁ, Pardo N (2017) Trend analysis of CO_2 and CH_4
765 recorded at a semi-natural site in the northern plateau of the Iberian Peninsula. Atmos
766 Environ 151:24-33. doi: 10.1016/j.atmosenv.2016.11.068.

767 Piao S, Liu Z, Wang Y, Ciais P, Yao Y, Peng S, Chevallier F, Friedlingstein P, Janssens
768 IA, Peñuelas J, Sitch S, Wang T (2017) On the causes of trends in the seasonal
769 amplitude of atmospheric CO_2. Glob Chang Biol 24(2):608-616.
770 doi:10.1111/gcb.13909.

771 Pickers PA, Manning AC (2015) Investigating bias in the application of curve fitting
772 programs to atmospheric time series. Atmos Meas Tech 8:1469-1489. doi: 10.5194/amt-
773 8-1469-2015.

774 Rasmussen R, Khalil M (1981) Atmospheric methane (CH_4): Trends and seasonal
775 cycles. J Geophys Res 86: 9826-9832. doi: org/10.1029/JC086iC10p09826.

776 Rella CW, Chen H, Andrews AE, Filges A, Gerbig C, Hatakka J, Karion A, Miles NL,
777 Richardson SJ, Steinbacher M, Sweeney C, Wastine B, Zellweger C (2013) High
778 accuracy measurements of dry mole fractions of carbon dioxide and methane in humid
779 air. Atmos Meas Tech 6:837-860. doi: 10.5194/amt-6-837-2013.

780 Rodríguez-Cortés FJ, Mateu J (2015) Second-order smoothing of spatial point patterns
781 with small sample sizes: a family of kernels. Stoch Environ Res Risk Assess 29:295–
782 308. doi: 10.1007/s00477-014-0944-x.

783 Sachs L (1978) Applied Statistics: a handbook of techniques. Springer. Series in
784 Statistics. Fifth Edition, New York.

785 Sánchez ML, García MA, Pérez IA, Pardo N (2014) CH_4 continuous measurements in
786 the upper Spanish plateau. Environ Monit Assess 186:2823-2834. doi: 10.1007/s10661-
787 013-3583-7.

788 Saunois M, Bousquet P, Poulter B, Peregon A, Ciais P, Canadell JG , Dlugokencky EJ,
789 Etiope G, Bastviken D, Houweling S, Janssens-Maenhout G, Tubiello FN, Castaldi S,
790 Jackson RB, Alexe M, Arora VK, Beerling DJ, Bergamaschi P, Blake DR, Brailsford G,
791 Brovkin V, Bruhwiler L, Crevoisier C, Crill P, Covey K, Curry C, Frankenberg C,
792 Gedney N, Höglund-Isaksson L, Ishizawa M, Ito A, Joos F, Kim H-S, Kleinen T,

793 Krummel P, Lamarque J-F, Langenfelds R, Locatelli R, Machida T, Maksyutov S,
794 McDonald KC, Marshall J, Melton JR, Morino I, Naik V, O'Doherty S, Parmentier F-
795 JW, Patra PK, Peng C, Peng S, Peters GP, Pison I, Prigent C, Prinn R, Ramonet M,
796 Riley WJ, Saito M, Santini M, Schroeder R, Simpson IJ, Spahni R, Steele P, Takizawa
797 A, Thornton BF, Tian H, Tohjima Y, Viovy N, Voulgarakis A, van Weele , van der
798 Werf GR, Weiss R, Wiedinmyer C, Wilton DJ, Wiltshire A, Worthy D, Wunch D, Xu
799 X, Yoshida Y, Zhang B, Zhang Z, Zhu Q (2016) The global methane budget 2000-2012.
800 Earth Syst Sci Data 8: 697-751. doi: 10.5194/essd-8-697-2016.

801 Scott DW (1992) Multivariate density estimation: theory, practice, and visualization.
802 John Wiley & Sons, New York.

803 Silverman BW (1998) Density estimation for statistics and data analysis. Chapman &
804 Hall/ CRC, Boca Raton.

805 Vermeulen AT, Hensen A, Popa ME, van den Bulk WCM, Jongejan PAC (2011)
806 Greenhouse gas observations from Cabauw Tall Tower (1992-2010). Atmos Meas Tech
807 4:617-644. doi:10.5194/amt-4-617-2011.

808 Wilks DS (2006) Statistical methods in the atmospheric sciences. Elsevier. Second
809 Edition, Amsterdam.

810 Wu J, Guan D, Yuan F, Yang H, Wang A, Jin C (2012) Evolution of atmospheric
811 carbon dioxide concentration at different temporal scales recorded in a tall forest.
812 Atmos Environ 61:9-14. doi: 10.1016/j.atmosenv.2012.07.013.

Yu KN, Cheung YP, Cheung T, Henry RC (2004) Identifying the impact of large urban airports on local air quality by nonparametric regression. Atmos Environ 38:4501-4507. doi: 10.1016/j.atmosenv.2004.05.034.

Zhang D, Tang J, Shi G, Nakazawa T, Aoki S, Sugawara S, Wen M, Morimoto S, Patra PK, Hayasaka T, Saeki T (2008) Temporal and spatial variations of the atmospheric CO_2 concentration in China. Geophys Res Lett 35, L03801. doi: 10.1029/2007GL032531.

Zhang F, Zhou LX, Xu L (2013) Temporal variation of atmospheric CH_4 and the potential source regions at Waliguan, China. Sci China Earth Sci 56-727:736. doi: 10.1007/s11430-012-4577-y.

Zhu C, Yoshikawa-Inoue H (2015) Seven years of observational atmospheric CO_2 at a maritime site in northernmost Japan and its implications. Sci Total Environ 524-525:331-337. doi: 10.1016/j.scitotenv.2015.04.044.