Science of the Total Environment 819 (2022) 153129

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Trend analysis and outlier distribution of CO_2 and CH_4: A case study at a rural site in northern Spain

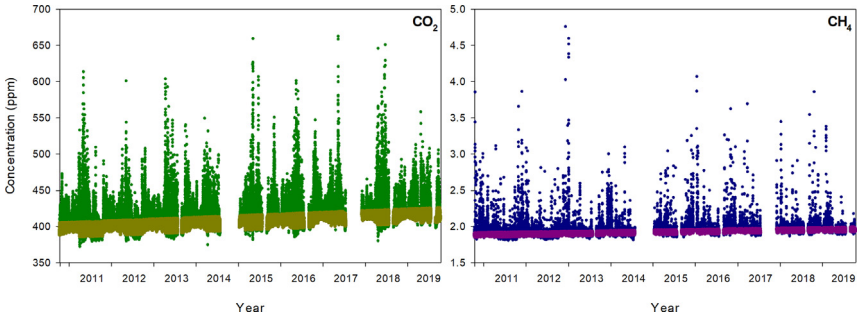
Isidro A. Pérez *, M. Ángeles García, M. Luisa Sánchez, Nuria Pardo

Department of Applied Physics, Faculty of Sciences, University of Valladolid, Paseo de Belén, 7, 47011 Valladolid, Spain

HIGHLIGHTS

- A fraction of outliers should be excluded to determine the concentration trend.
- Outliers can be described by skewed distributions.
- The distributions for outliers above or below the trend line may be different.
- The modified Nash-Sutcliffe efficiency is a sensitive goodness-of-fit statistic.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 20 October 2021
Received in revised form 10 January 2022
Accepted 11 January 2022
Available online 15 January 2022

Editor: Pavlos Kassomenos

Keywords:
Skewed distributions
Greenhouse gases
Efficiency estimators
Distribution fitting
Outlier distribution
Emissions statistical control

ABSTRACT

CO_2 and CH_4 outliers may have a noticeable impact on the trend of both gases. Nine years of measurements since 2010 recorded at a rural site in northern Spain were used to investigate these outliers. Their influence on the trend was presented and two limits were established. No more than 23.5% of outliers should be excluded from the measurement series in order to obtain representative trends, which were 2.349 ± 0.012 ppm year^{-1} for CO_2 and 0.00879 ± 0.00004 ppm year^{-1} for CH_4. Two types of outliers were distinguished. Those above the trend line and the rest below the trend line. Outliers were described by skewed distributions where the Weibull distribution figures prominently in most cases. A qualitative procedure was presented to exclude the worst fits, although five statistics were considered to select the best fit. In this case, the modified Nash-Sutcliffe efficiency is prominent. Finally, three symmetrical distributions were added to fit the observations when outliers are excluded, with the Gaussian and beta distributions providing the best fits. As a result, certain skewed functions, such as the lognormal distribution, whose use is frequent for air pollutants, could be questioned in certain applications.

1. Introduction

Outliers appear in measurement series with unequal frequency. The reasons for this may vary substantially. Scattered outliers of high CO_2 and CH_4 values recorded at rural sites were linked with local sources and/or particular weather conditions, such as temperature inversions and low wind speed, whereas low concentration events may be determined by wind direction or local sinks (Belikov et al., 2019; Yang et al., 2019). At urban sites, these outliers are more prominent than in rural environments due to pollution sources (Yang et al., 2021).

Observation selection is necessary for specific applications. For instance, data must be filtered when the aim is to determine background concentrations or large-scale features where certain local influences such as ecosystem emissions or nearby traffic must be isolated (Wang et al., 2020). In the same line, Fang et al. (2016b) excluded high CO_2 concentration night-time episodes, linked with biospheric respiration in a poorly

* Corresponding author.
E-mail address: iaperez@fa1.uva.es (I.A. Pérez).

http://dx.doi.org/10.1016/j.scitotenv.2022.153129
0048-9697/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

mixed atmosphere, and local sources or sinks associated with wind speeds lower than 1.5 m s^{-1}, from regional events.

Following Ben-Gal (2005), procedures for outlier detection may be divided between parametric and non-parametric methods. Parametric methods either consider a known data distribution or are based on calculations of unknown distribution parameters (van Zoest et al., 2018). In contrast, within non-parametric methods, those which are distance-based stand out since they are suitable for large databases (Kontaki et al., 2011). In some cases, the average value is locally computed in a given neighbourhood, and outliers are observations that differ significantly from this average. In this situation, one additional problem is the definition of neighbourhood (Carrilho et al., 2018). When a distance is required, a threshold, which considers the standard deviation of observations, must be established (Martínez et al., 2014).

Procedures used to treat atmospheric concentrations do not usually include distribution function fitting (Fang et al., 2014; Kilkki et al., 2015). However, some examples demonstrate the usefulness of such analyses. Karaca et al. (2005) used ten distribution functions to study particulate matter in the municipality of Istanbul. Air quality in Delhi for SO_2, NO_2, and particulate matter was studied by Sharma et al. (2013), and the statistical analysis included 12 distribution functions to identify and estimate the best-fit distribution. A generalised form of the Gumbel distribution was introduced by Korkmaz (2015) and applied to air pollution data to demonstrate its modeling potential. Statistics such as skewness and kurtosis as well as the generalised extreme value distribution were used by Battista et al. (2016) to investigate air pollution in Rome, Italy. Martins et al. (2017) used the generalised extreme value and the generalised Pareto distributions to analyse air pollution in the two largest urban areas in Brazil. Finally, the contrast between CO_2 concentrations in Seoul city centre, South Korea, and a mountain in its surroundings was described by the probability density function, which was quite flat in winter, but more peaked in summer (Park et al., 2021).

CO_2 and CH_4 evolutions are measured in various environments due to their link to climate change. This objective is achieved by station networks such as stations under the Global Atmosphere Watch (GAW) programme of the World Meteorological Organization (WMO) (WDCGG, 2021), or by the cooperative air sampling network of the Global Monitoring Laboratory (NOAA, 2021). These networks focus on measurements to determine trend evolution and its accompanying cycles. However, particular analyses are carried out at polluted sites, which are considered sources of both gases (Lian et al., 2021; Nyasulu et al., 2021), when transport plays a key role (Jain et al., 2021), or at certain sites of specific interest (Kurbatova et al., 2020).

The measurement site considered in the current analysis is a place where the evolution of both gases has already been investigated (Fernández-Duque et al., 2020; Pérez et al., 2020). However, the database used in this paper was increased vis-à-vis that employed in previous studies. The first objective is to study outlier influence on the CO_2 and CH_4 trend measured at this rural site. Consequently, this trend is calculated and compared with values determined in various environments. The influence of the number of observations on the trend of both gases has been calculated by Pérez et al. (2019). However, outlier influence remains unexplored. Moreover, compared to time series analysis (Bianchi et al., 2020), which could be seen as quite complex, the procedure for isolating outliers presented in this study is relatively easy, since it is based on linear regressions.

The second objective is to describe these outliers by some distribution functions. Since the studies cited above focus on skewed distributions, some distributions of this type will be used. In order to simplify the calculation procedure, iterative calculation methods were excluded. For this objective, the key point is that distributions are only used for outliers and not for all observations. A second key point lies in the statistics used to establish the goodness of the fit. This paper considers statistics that have scarcely been investigated to date.

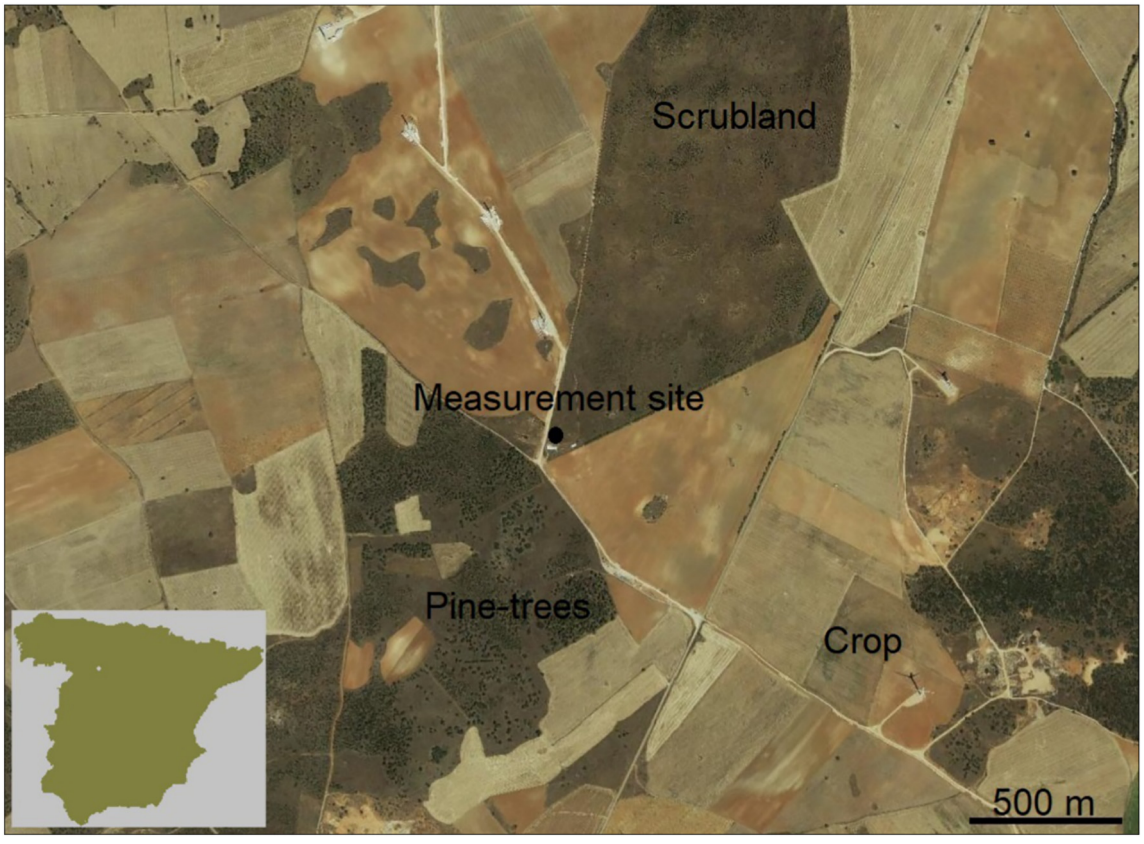


Fig. 1. Image courtesy of © ign.es showing the measurement site and its surrounding area.

This analysis assumes that the observations present a defined trend and noticeable outliers are a fraction of observations that follow a skewed distribution. Consequently, fairly regular observations, where outliers are marginal or are clusters of values, fall outside the scope of this study.

In the materials and methods section, the measurement site is described, followed by the procedure employed to isolate outliers and the distribution functions used together with the goodness-of-fit statistics. The results section presents the trends under different outlier percentages, and the selection of specific percentages is investigated. Moreover, the fits of ten skewed distributions are presented as a function of outlier percentage in order to compare the values of the goodness-of-fit statistic. Finally, a similar treatment is applied to observations excluding outliers by adding some symmetrical distributions to the skewed ones previously used.

2. Materials and methods

2.1. Observations

Dry concentrations of CO_2 and CH_4 were measured over nine years, commencing in October 2010, at the Low Atmosphere Research Centre (CIBA station, 41° 48′ 50” N, 4° 55′ 59” W, 852 m a.s.l., Fig. 1). The site is nearly flat and the vegetation is formed by scrublands surrounded by rainfed crops and sparse trees. The climate of the site determines both the vegetation and its annual evolution. Following the Köppen classification, the climate is Cfb, i.e. temperate without a dry season and with a temperate summer. The nearest urban site is the city of Valladolid (around 300,000 inhabitants), located some 25 km SE of the measurement site.

Observations were obtained from a Picarro G1301 analyser where three levels were considered; at heights of 1.8 m, 3.7 m, and 8.3 m. However, only half-hourly averages of measurements at the lowest level were used in this analysis.

Calibrations were periodically performed with three NOAA standards, which were 452.56 ppm, 399.27 ppm and 348.55 ppm for CO_2, and 1.9904 ppm, 1.8420 ppm and 1.6310 ppm for CH_4, i.e. above, around and below ambient concentrations. However, due to a technical problem, there were two distinct measurement periods; the first until 19 July 2017, and the second from 6 December 2017. The corresponding equations for correcting the measurements were:

$$\left.\begin{aligned} C_{CO_2C} &= 1.00355\, C_{CO_2} - 0.18482 \\ C_{CH_4C} &= 0.99264\, C_{CH_4} + 0.01092 \end{aligned}\right\} \text{for the first period.}$$
$$\left.\begin{aligned} C_{CO_2C} &= 1.00572\, C_{CO_2} - 1.17899 \\ C_{CH_4C} &= 0.99017\, C_{CH_4} + 0.00884 \end{aligned}\right\} \text{for the second period.} \quad (1)$$

In these equations, subscript *C* corresponds to the corrected concentration.

2.2. Procedure for obtaining outliers

Outliers are determined by an iterative procedure where the starting point is a straight line calculated with initial observations. Once the first outlier from this line is determined, it is excluded from the observations and the procedure continues with the remaining observations (Fig. 2), i.e., a new straight line is calculated, the second outlier is determined and then excluded, and so on. Absolute values of outlier residuals against their corresponding linear fits were calculated and their medians were also obtained. Since outliers may be located above the trend line or below this line, their residuals were denoted as top and bottom, respectively. An exponential expression was also used for outlier selection. Some robust statistics were calculated. The median was used as a location indicator, the interquartile range indicates measurement spread, symmetry is calculated by the Yule-Kendall index, and flatness is expressed by the robust kurtosis, which is the quotient between the interquartile range and the double distance between the 9th and 1st deciles.

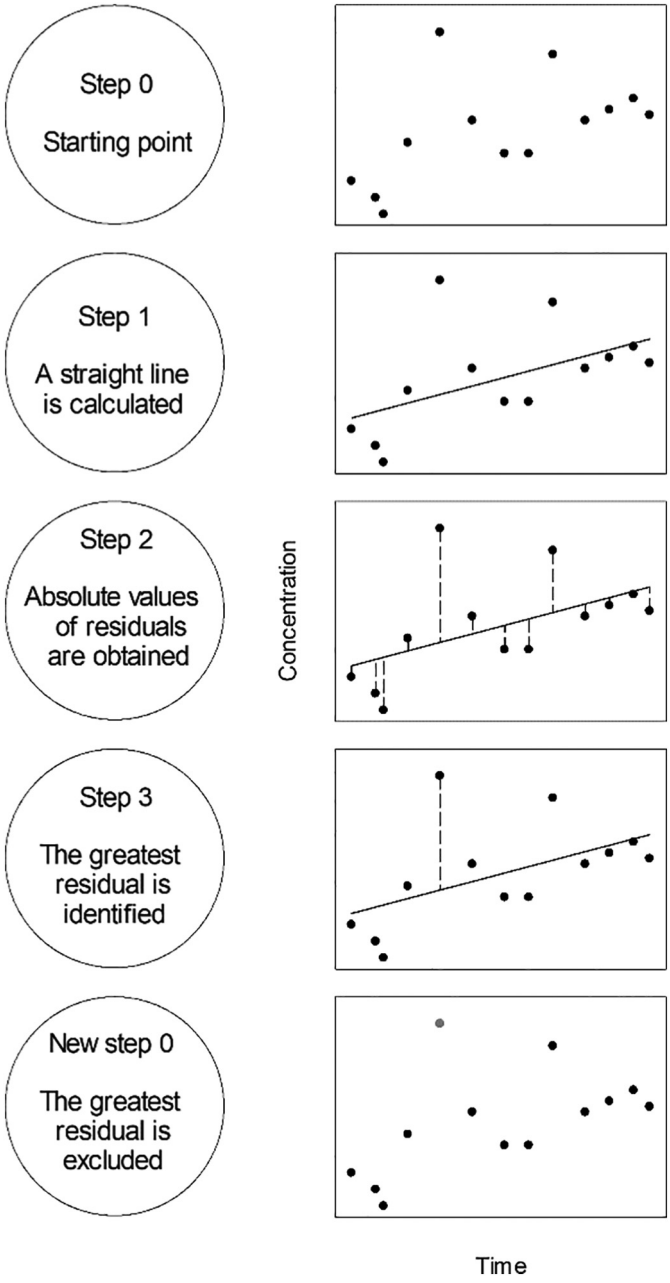


Fig. 2. Schematic representation of the outlier determination method.

2.3. Distribution fitting

Thirteen distribution functions presented in Table 1 were employed. Ten of them were skewed and three were symmetrically distributed.

Theoretical and experimental cumulative distribution functions were calculated with programs in Fortran developed by the authors and the goodness of fit was determined with five efficiency statistics (Krause et al., 2005), which are presented in Table 2, where *O* refers to observed values and *C* to calculated values.

3. Results

3.1. Trend calculation

Fig. 3(a, b) presents the results for both - lineal and exponential - fits following the percentage of outliers suppressed. Two values were considered

Table 1
Distribution functions used to fit the outlier residuals (only the skewed ones), and the central residuals (all the functions)*.

Type	Distribution	Probability density function	Parameter calculation
Skewed	Beta [a]	$f(x)=\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]x^{\alpha-1}(1-x)^{\beta-1};0\le x\le 1;\alpha,\beta>0$	$\alpha=\frac{\overline{x}^2(1-\overline{x})}{s^2}-\overline{x};\beta=\frac{\alpha(1-\overline{x})}{\overline{x}};\overline{x}=\frac{\overline{y}-a}{b-a};$ $s_x^2=\frac{s_y^2}{(b-a)^2}$
	Exponential [b]	$f(x)=\theta\ exp\ (-\theta x);x>0;\theta>0$	$\theta=1/\overline{x}$
	Frechet [c]	$f(x)=\frac{\lambda}{s}\left(\frac{s}{x}\right)^{\lambda+1}exp\left\{-\left(\frac{s}{x}\right)^{\lambda}\right\};x\ge 0;\sigma,\lambda>0$	$-\ ln\ [-\ ln\ \{F(x)\}]\ =\ -\ \lambda\ \ln\ s\ +\ \lambda\ \ln\ (x)$
	Gamma [a]	$f(x)=\frac{(x/\beta)^{\alpha-1}\ exp\ (-x/\beta)}{\beta\ \Gamma(\alpha)};x,\alpha,\beta>0$	$D=\ \ln\ (\overline{x})-\frac{1}{n}\sum_{i=1}^{n}\ln\ (x_i);\alpha=\frac{1+\sqrt{1+\frac{4D}{3}}}{4D};$ $\beta=\overline{x}/\alpha$
	Gumbel [a]	$f(x)=\frac{1}{\beta}\ exp\left\{-exp\left[-\frac{(x-\zeta)}{\beta}\right]-\frac{(x-\zeta)}{\beta}\right\}$	$\beta=\frac{s\sqrt{6}}{\pi};\zeta=\overline{x}-\gamma\beta;\gamma=0.57721$
	Lindley [b]	$f(x)=\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x};x>0,\theta>0$	$\theta=\frac{-(\overline{x}-1)+\sqrt{(\overline{x}-1)^2+8\overline{x}}}{2\overline{x}}$
	Lognormal [d]	$f(x)=\frac{1}{xs\sqrt{2\pi}}\ exp\left(-\frac{(\ln x-\mu)^2}{2s^2}\right)$	$\mu=\overline{\ln x};\ s\ =\ s_{\ln x}$
	Triangular [d]	$f(x)=\begin{cases}\frac{2(x-a)}{[(b-a)(c-a)]},if\ a\le x\le c\\ \frac{2(b-x)}{[(b-a)(b-c)]}.if\ c\le x\le b\end{cases}$	$c=3\overline{x}-a-b$
	Wald [d]	$f(x)=\sqrt{\frac{\lambda}{2\pi x^3}}\ exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right)$	$\mu=\overline{x},\lambda=\frac{\mu^3}{\sigma^2}$
	Weibull [e]	$f(x)=\left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1}\ exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right];x,\alpha,\beta>0$	$ln(-\ ln\ (1\ -\ F(x)))\ =\ \alpha\ ln\ (x)\ -\ \alpha\ ln\ (\beta)$
Symmetrical	Gaussian [d]	$f(x)=\frac{1}{s\sqrt{2\pi}}\ exp\left(\frac{-(x-\mu)^2}{2s^2}\right)$	$\mu=\overline{x}$
	Laplace [d]	$f(x)=\frac{1}{2b}\ exp\left(-\frac{\lvert x-a\rvert}{b}\right)$	$a=\overline{x},b=\sqrt{\frac{\sigma^2}{2}}$
	Logistic [d]	$f(x)=\frac{exp\ [(x-a)/b]}{b\{1+exp\ [(x-a)/b]\}^2}$	$a=\overline{x},b=\sqrt{3}s/\pi$

*$\overline{x}$ is the average and s the standard deviation.
[a] (Wilks, 2019).
[b] (Ghitany et al., 2008).
[c] (Bury, 1999).
[d] (Forbes et al., 2011).
[e] (Akdağ and Dinler, 2009).

for the exponential fit since trends are different for the beginning and end of the measurement period. A first section including around 4% of outliers is associated with a noticeable change in the trend. The second outlier border may be better defined for CO_2 than for CH_4. The outlier frontier proposed was around 23.5% of all the outliers. This interval between both limits is featured by a slow decrease in the CO_2 trend and is followed by a relatively steady trend. However, the CH_4 trend increased in this interval and was followed by a slight decrease and a subsequent increase. The trend values calculated with the linear fit for this outlier percentage are 2.349 ± 0.012 ppm year^{-1} for CO_2, and 0.00879 ± 0.00004 ppm year^{-1} for CH_4. Fig. 3(c) shows the coefficient of determination, r^2, for the linear fit as a function of the outlier percentage. It is slightly better for CH_4 than for CO_2. Moreover, the increase is noticeable at low percentages for CH_4, revealing the few outliers that have a pronounced impact on the trend.

Since the trend for the linear fit is around the average of the trend values for the exponential fit, the linear fit was used in the rest of the current study. Fig. 4 presents the medians of absolute values of residuals for both groups together with their outlier percentages. A noticeable drop in the median is observed for the first outliers. However, once this initial drop is passed, the decrease in the outlier residual median with the outlier percentage is slight. This result indicates that the number of conspicuous outliers is low. Moreover, a higher number of outliers is obtained above the trend line and their concentrations are higher.

Fig. 5 presents the box-plot for the three types of observations when outliers account for 23.5% of data. In this figure, the median is the line in the box representing the interquartile range. Whiskers extend from the 10th to the 90th percentile and isolated dots correspond to the 5th and the 95th percentiles. Differences between pairs of group means are statistically significant following Fisher's least significant difference method. Observation dispersion is similar for central and bottom CO_2 values and for the three CH_4 groups.

Concentrations of absolute values of CO_2 and CH_4 residuals are quite dissimilar. However, they may be scaled in an interval between 0 and 1 to make them comparable. The second frontier, 23.5% of all outliers, was selected to make additional calculations of these scaled values. Table 3 presents some robust statistics. The dispersion of CO_2 scaled residuals is similar for both residual types. However, a noticeable contrast is obtained for CH_4, since the dispersion of the top residuals is low, but higher for the bottom residuals. Finally, the robust kurtosis is similar to that of a Gaussian distribution, 0.262, for the bottom residuals. However, it is lower for the top residuals, revealing that the top residuals present peaked distributions.

These features may be observed in Fig. 6, where the cumulative distribution function (CDF) of scaled residuals is presented. The four curves are featured by a sharp increase in CDF values, although these are smoother for the CH_4 bottom concentrations. This rapid increase is attributed to the high agglomeration of observations and is followed by a slow rise due to the relatively few large observations. The concave downward shape is associated with positively skewed data (Wilks, 2019), and the range between bottom and top scaled residuals is greater for CH_4 than for CO_2.

Table 2
Efficiency statistics used in this study.

Name	Equation
Coefficient of determination	$r^2=\left(\frac{\sum_{i=1}^{n}(O_i-\overline{O})(C_i-\overline{C})}{\sqrt{\sum_{i=1}^{n}(O_i-\overline{O})^2}\sqrt{\sum_{i=1}^{n}(C_i-\overline{C})^2}}\right)^2$
Willmott index of agreement	$d=1-\frac{\sum_{i=1}^{n}(O_i-C_i)^2}{\sum_{i=1}^{N}\left(\lvert C_i-\overline{O}\rvert+\lvert O_i-\overline{O}\rvert\right)^2}$
Modified index of agreement	$d_{mod}=1-\frac{\sum_{i=1}^{n}\lvert O_i-C_i\rvert}{\sum_{i=1}^{N}\left(\lvert C_i-\overline{O}\rvert+\lvert O_i-\overline{O}\rvert\right)}$
Nash-Sutcliffe efficiency	$E=1-\frac{\sum_{i=1}^{n}(O_i-C_i)^2}{\sum_{i=1}^{n}(O_i-\overline{O})^2}$
Modified Nash-Sutcliffe efficiency	$E_{mod}=1-\frac{\sum_{i=1}^{n}\lvert O_i-C_i\rvert}{\sum_{i=1}^{n}\lvert O_i-\overline{O}\rvert}$

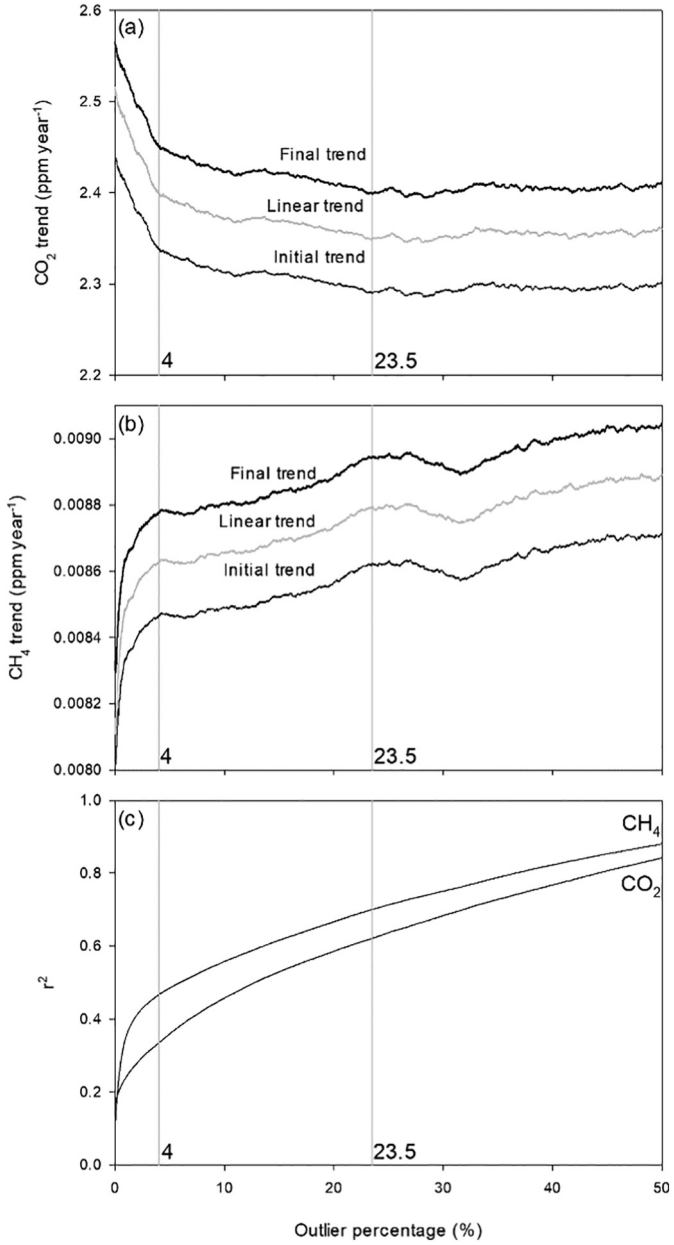


Fig. 3. CO_2 (a) and CH_4 (b) trends following the percentage of outliers supressed with linear and exponential fits. Initial and final refer to the trends calculated at the beginning and end of the measurement period, respectively, with the exponential fit. (c) r^2 values for the linear fit. Vertical lines correspond to the frontiers suggested.

3.2. Distribution fitting

The absolute value of residuals was considered for the second frontier, 23.5%, of all the outliers. Ten skewed distribution functions were used whose experimental, F_1, and theoretical CDF, F_2, were compared in Fig. 7. The better the agreement the closer the corresponding line is to the diagonal. This plot provides a qualitative representation of the goodness of fit, since the fits of CH_4 top outliers for the beta, Gumbel, and triangular functions must be discarded. However, selecting the best satisfactory agreement may prove difficult since some representations look similar. Although every one of these fits could be chosen, a precise choice can be made with the numerical efficiency criteria presented in Table 2. Fig. 8 presents the values of these efficiency criteria. The high values of r^2 contrast with the low values of the modified expression of the index of agreement and the

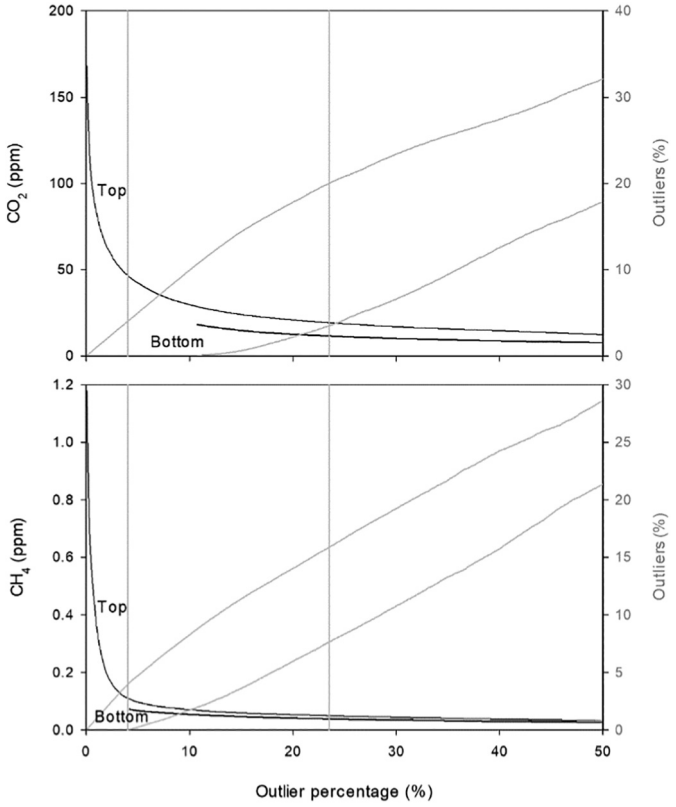


Fig. 4. Medians of the absolute values of residuals in the linear fit and the outlier percentages following the percentage of all outliers. Vertical grey lines correspond to the 4% and 23.5% borders.

lowest values of the modified Nash-Sutcliffe efficiency, which was the most sensitive statistic among those used in this paper. These results determined that the modified Nash-Sutcliffe efficiency was chosen as the reference estimator in the rest of this study. The range covered was narrow for the CH_4 bottom residuals and wide for the CH_4 top residuals. This figure indicates that similar results were obtained for the exponential and Lindley

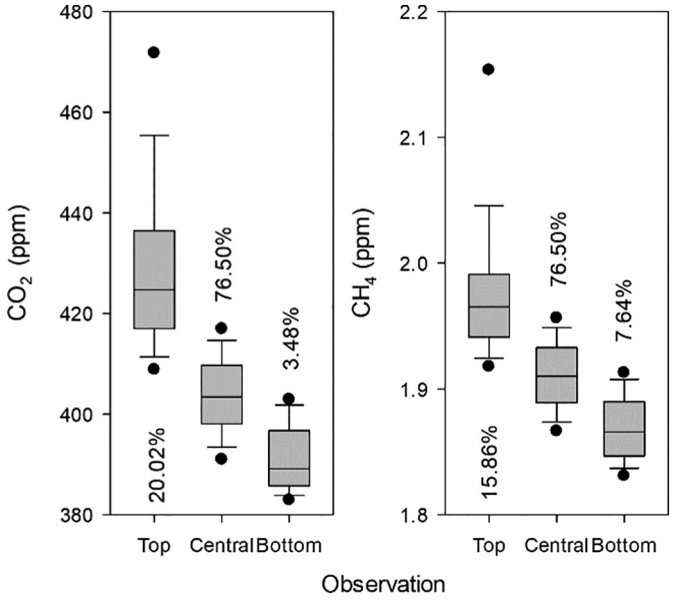


Fig. 5. Box-plot of observations when outliers are 23.5%. Three types of data are established, i.e. values linked with top and bottom residuals and the rest, which are the central observations.

Table 3
Robust statistics calculated for the scaled absolute values of residuals for the 23.5% of outliers.

Residual	Median	Interquartile range	Yule-Kendall index	Robust kurtosis
CO_2 Top	0.038	0.070	0.340	0.226
CO_2 Bottom	0.069	0.094	0.189	0.266
CH_4 Top	0.007	0.012	0.341	0.174
CH_4 Bottom	0.121	0.177	0.229	0.263

distributions applied to the CH_4 residuals. Moreover, the function that provides the best agreement is that with the highest value of the modified Nash-Sutcliffe efficiency.

Since the distribution fitting was made for a specific percentage of all the outliers (23.5%), the response of residuals to skewed distributions must be investigated in a wide range of outlier percentage. Fig. 9 presents such a response. Efficiencies are similar for every distribution function for high outlier percentages, whereas major changes are observed for low percentages. Values are gradually distributed for CO_2 top residuals and the best fit is for the Weibull distribution. The best result for the CH_4 top residuals is reached with the lognormal distribution, although the Wald distribution provides similar values at around 11% of outliers. Moreover, values of the Weibull distribution are similar to those of the lognormal distribution at high outlier percentages. Efficiencies for the bottom outliers appear mixed. Beta, Weibull and gamma distributions provide satisfactory fits in a wide range of percentages (the Linley distribution must also be considered for CO_2) whereas the Frechet distribution stands out due to its low values.

3.3. Central value fitting

Skewed distributions used to fit the absolute value of residuals for outliers were also used to fit the residuals of central values when outliers are excluded. Table 4 presents some robust statistics for location, spread, symmetry and flatness for central residuals when removing 23.5% of outliers. The median is almost zero, the Yule-Kendall index reveals that the distribution is nearly symmetrical and the robust kurtosis is close to that for a Gaussian distribution, which is 0.262. Taking into account these values, three symmetrical distributions are included; the Gaussian, Laplace, and logistic distributions. The results of these fits calculated by the modified Nash-Sutcliffe efficiency are presented in Fig. 10. Lower outlier percentages are featured by noticeable changes of the statistic used. Low values of the statistic were obtained for the triangular function at low outlier percentages. Most of the distributions provided statistic values of above 0.8 in a wide interval of outlier percentage. Satisfactory fits were observed for the Gaussian distribution, although only in a narrow interval of outlier percentage. When this percentage increases, the best fits were reached by the beta distribution, which was flexible enough to successfully describe both skewed and symmetric distributions. In contrast, the exponential function was the least suitable to fit these residuals, together with the Lindley function for CH_4. The values of the poor fits are lower than those presented in Fig. 9, revealing that certain skewed distributions are unsuitable for these data.

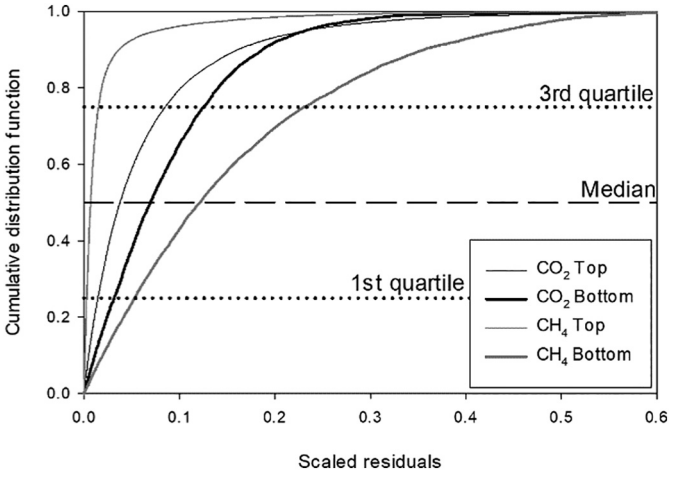


Fig. 6. Cumulative distribution function of scaled residuals.

4. Discussion

4.1. Trend evolution

Belikov et al. (2019) analysed CO_2 and CH_4 evolution at nine sites in Siberia using a model developed by Harvey and Peters (1990), later used by Taylor and Letham (2018), and which considers three contributions. The first is a trend function that models no periodic changes. The second refers to periodic seasonal processes, and the third contribution is related with irregular, undetermined transformations. Several equations were used to explain the concentration trend. The most complex equation considers a fourth order polynomial (Inoue et al., 2006). Artuso et al. (2009) used an exponential function to describe the CO_2 trend at Lampedusa, Italy. However, the linear fit is the simplest and most frequent approximation to describe CO_2 and CH_4 time evolution, proving to be successful in varied environments, such as Finland (Aalto et al., 2002), central Siberia (Timokhina et al., 2015), the Netherlands (Vermeulen et al., 2011), or northeast China (Wu et al., 2012).

WMO (2020) presented atmospheric CH_4 in the period 1984–2020 where four intervals may be established. The first extended until 1992 with a noticeable increase that slowed down until 1999. Concentrations then remained nearly steady until 2007, when a new period of increases was noticed. CO_2 evolution may be divided into several periods with constant increases, although the interval from 1995 to 2011 stands out, followed by a second period up to the present day. Consequently, the evolution may be considered almost linear during the period analysed in the current paper. The linear fit of concentrations was successfully used by Pérez et al. (2020) to investigate concentration trend and annual cycle at the site over a six-year period.

The trend values calculated in this paper are in agreement with those presented in Table 5 for various sites, where some places in China figure prominently due to the high CO_2 values, which are above 3 ppm $year^{-1}$, contrasting with the 1.3 ppm $year^{-1}$ in Antarctica. Similarly, the CH_4 trend in this study was close to that measured at Cabaw, the Netherlands, and is situated between the low value found in Norway and the high value obtained in Hungary. Table 5 also presents the measurement procedure. Since the methods are different, the comparison between the outlier number and the influence of these outliers on the trend obtained by these procedures is a matter that is still open to study.

4.2. Outliers

Observations are frequently included in some studies by both scatter or line plots, with the number of outliers and their contribution to all the measurements not usually being the objective of the research. However, these graphs illustrate specific features linked with the site where observations were obtained. For instance, CO_2 outliers are extremely infrequent in Minamitorishima, a remote island in the western North Pacific (Wada et al., 2007). A different pattern is formed by noticeable outliers during the night linked with the stable stratification of the nocturnal boundary layer accompanied by low values during the day, caused by photosynthesis and dilution in the expanding mixing layer. Such is the case of Fraserdale, Canada (Higuchi et al., 2003) or Rishiri, Japan (Zhu and Yoshikawa-Inoue, 2015). Another pattern is observed in Sammaltunturi, Finland, above the Arctic Circle, where a noticeable contrast between summer and winter is obtained (Lohila et al., 2015). The high concentrations are usually the main outliers, such as at Cape Point, South Africa (Labuschagne et al., 2018). Air transport from polluted sites, such as cities, may be a noticeable

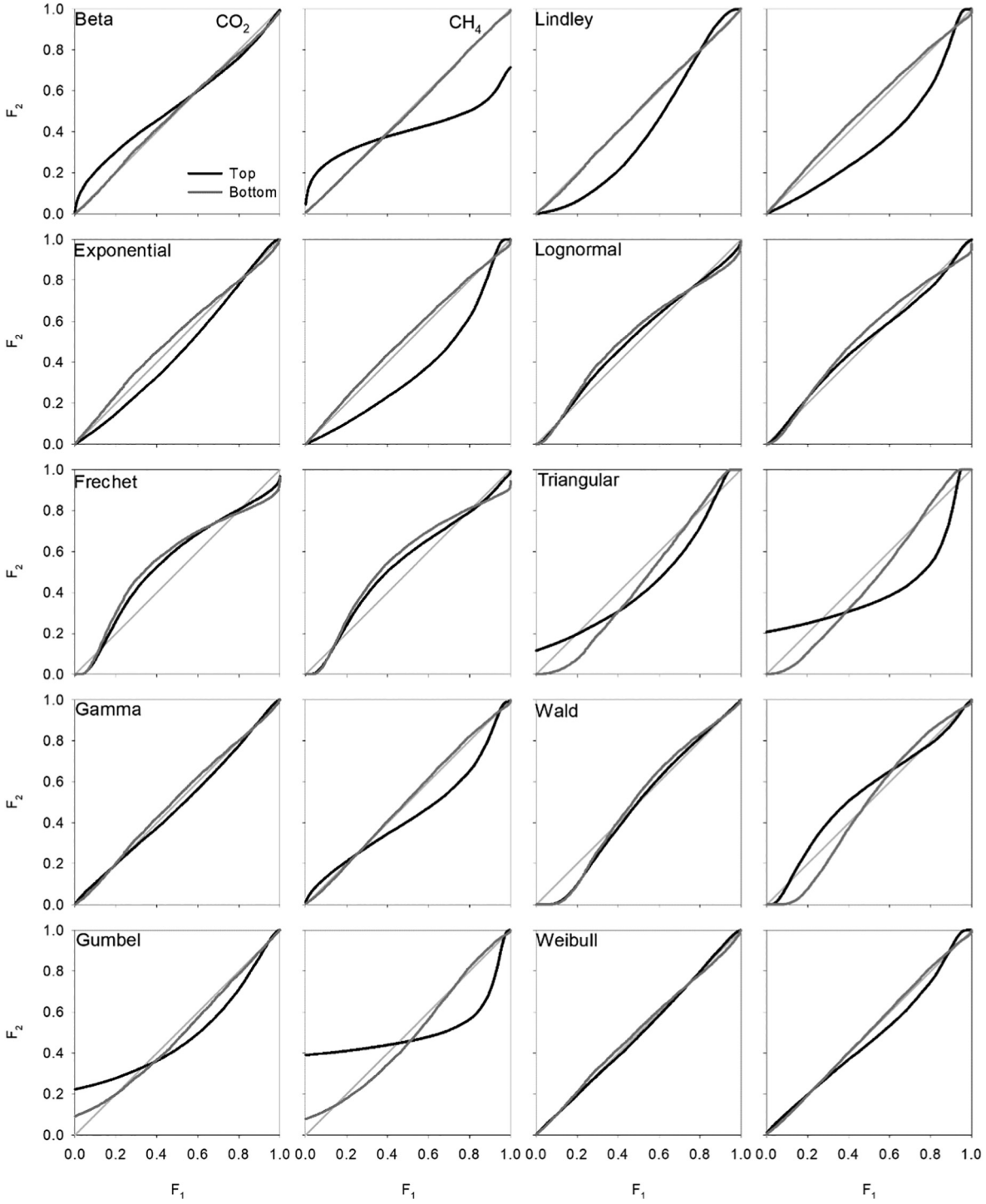


Fig. 7. Representation of the theoretical cumulative distribution function, F_2, against the cumulative experimental distribution function, F_1, for the absolute values of residuals of the linear fit.

source of outliers at remote sites that could be considered reference places of background air, such as Jungfraujoch, Switzerland (Affolter et al., 2021). However, extremely low concentrations may be a relevant feature in some places such as Takayama, in central Japan (Murayama et al., 2003).

Outlier description has sometimes been presented qualitatively. Wei et al. (2020) studied the relationship between air pollutants and meteorological conditions in Shanghai, China and considered some violin plots where concentrations are occasionally distributed symmetrically, for example, CO_2 on a clean day. However, asymmetrical distributions prevailed with tails on the right. In fact, the shape of the distribution function was that of a symmetrical distribution without most of the left tail.

Previous analyses at the site revealed that outliers have a twofold origin: the first is the stability of the low atmosphere during the night (Sánchez et al., 2010), and the second is the Valladolid urban plume (Pérez et al., 2012). This study explores certain features of these outliers in greater depth, such as the asymmetric shape of their distribution, the contrast between the outliers below and above the trend line, and the contrast between the outliers of both gases.

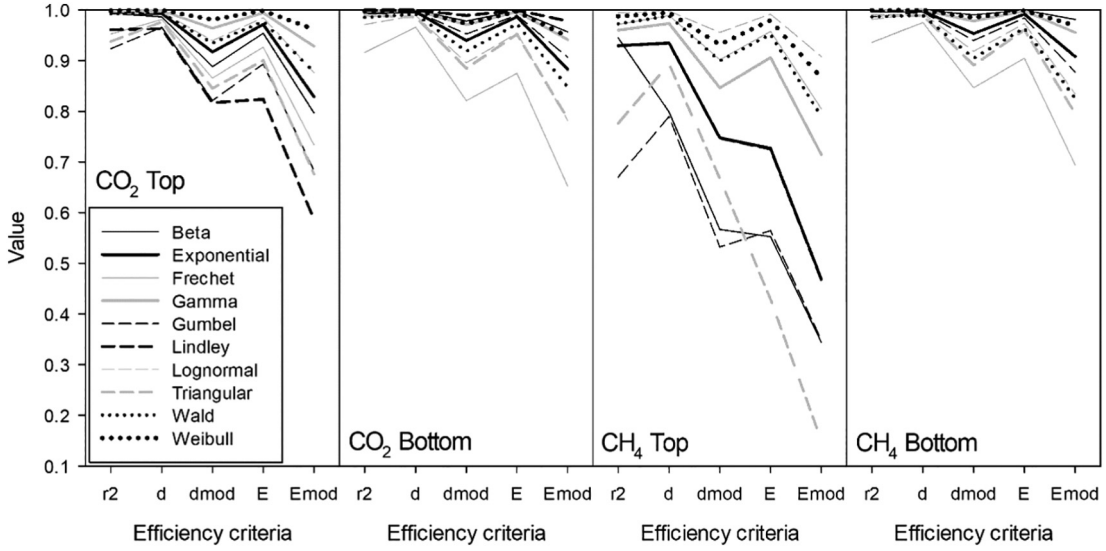


Fig. 8. Efficiency criteria used in the current study.

4.3. Concentration distribution

CO_2 concentration skewness has already been investigated (Pérez et al., 2014). Skewed distributions have been employed to describe these concentrations (Pérez et al., 2013). Recently, Pérez et al. (2021) fitted both greenhouse gas concentrations to seven skewed distributions following nine wind speed intervals. The current study expands this research, since 10 skewed distributions are used for outliers, with this outlier fit

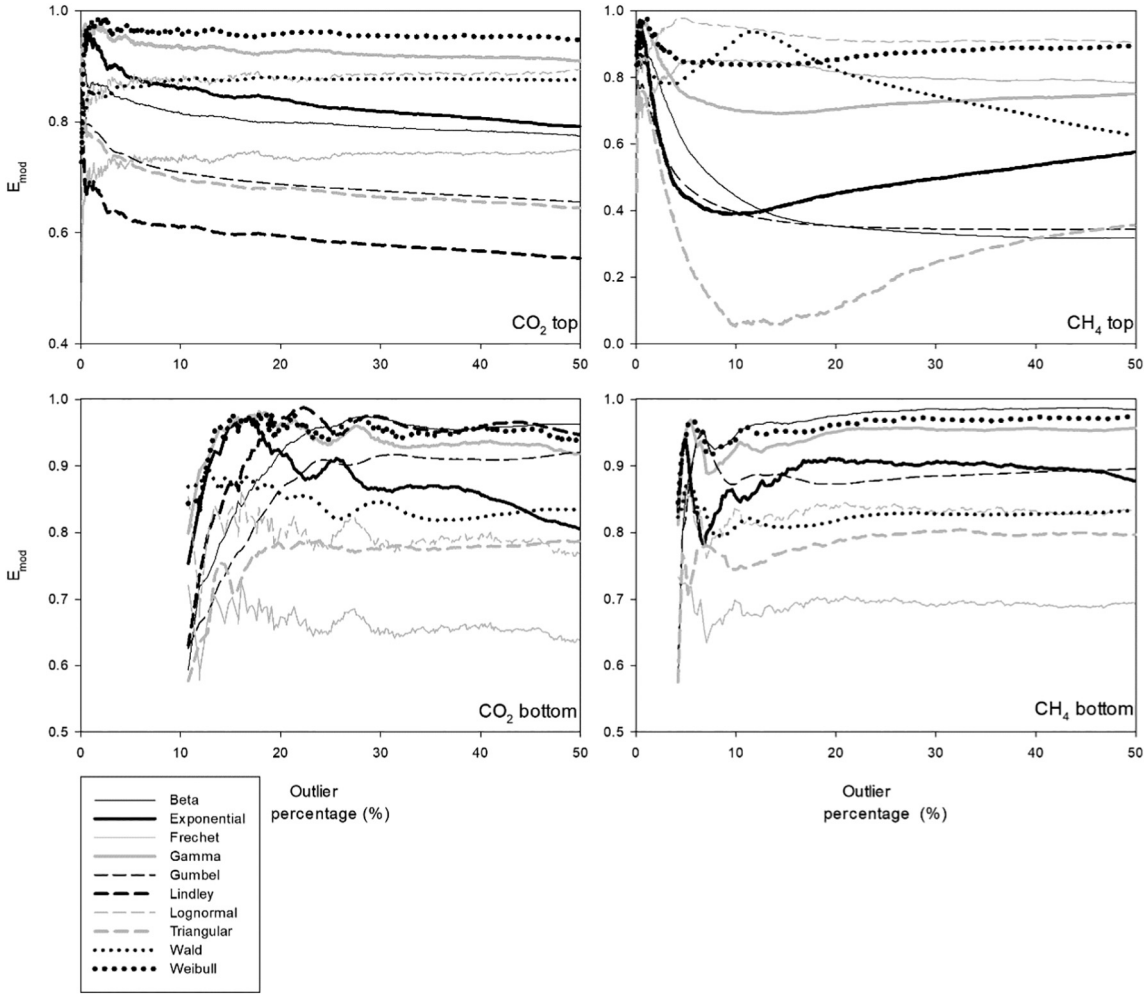


Fig. 9. The modified Nash-Sutcliffe efficiency for the outlier fitting with varied skewed distributions.

Table 4
Robust statistics for residuals of central values once 23.5% of outliers were removed.

Gas	Median (ppm)	Interquartile range (ppm)	Yule-Kendall index	Robust kurtosis
CO_2	0.01	7.17	−0.0018	0.2662
CH_4	0	0.0231	−0.0025	0.2713

being a specific contribution of this paper. In particular, this study presents two outlier types that must be treated separately and indicates that the Weibull distribution stands out as a function that could be considered by air pollution control agencies to analyse concentration limits for air quality. Moreover, three symmetrical distributions are also used for non-outlier observations.

Krause et al. (2005) presented five efficiency criteria to compare simulated and observed variables. These criteria were used by Pérez et al. (2021) to contrast experimental and theoretical cumulative distribution functions. The current analysis considers the same criteria and, in agreement with the previous study, the modified Nash-Sutcliffe efficiency proved to be the best

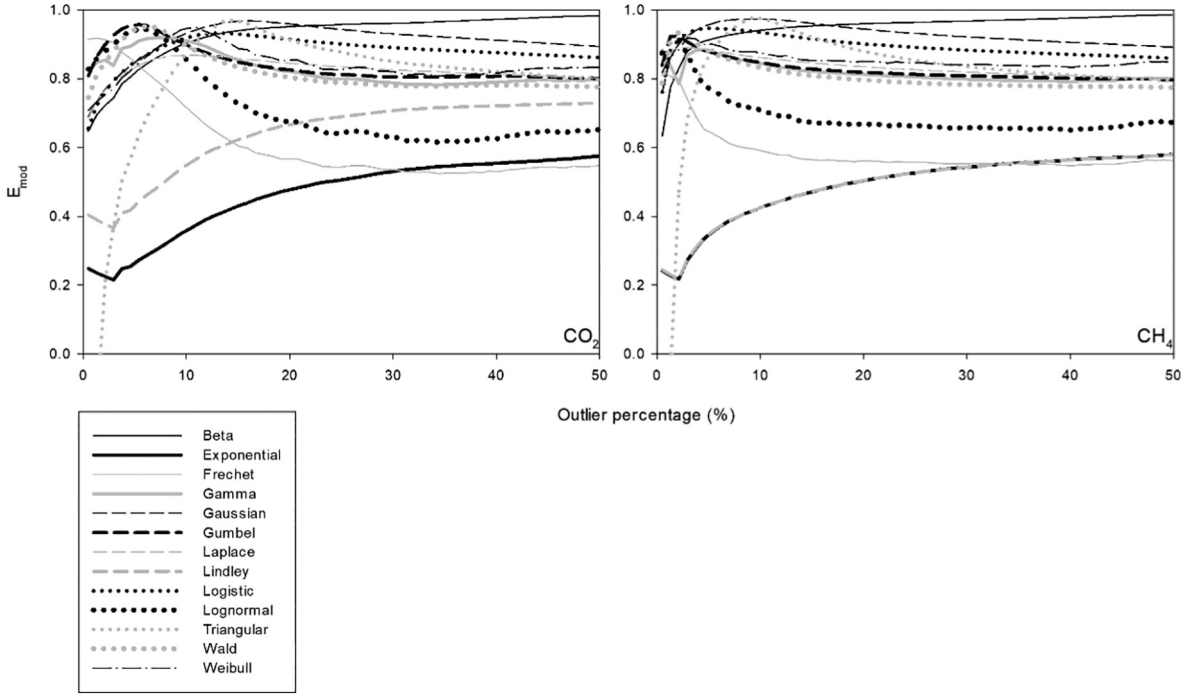


Fig. 10. The modified Nash-Sutcliffe efficiency for the central residual fitting with varied distributions (Gaussian, Laplace, and logistic are symmetrical) when outliers are excluded.

Table 5
Trends determined at different sites.

Gas	Reference	Site	Trend (ppm $year^{-1}$)	Period	Measurement procedure
CO_2	Aalto et al. (2002)	Pallas, Finland	2.5	1996–2000	Infrared analyser
	Artuso et al. (2009)	Lampedusa, Italy	1.9	1992–2007	Infrared analyser
	Cundari et al. (1995)	Mt. Cimone, Italy	1.66	1979–1991	Infrared analyser
	Fang et al. (2016a)	Shangdianzi, China	2.7–3.8	2009–2013	Cavity ring-down spectroscopy
	Guo et al. (2020)	Mt. Waliguan, China	2.45	2010–2016	Cavity ring-down spectroscopy
	Hernández-Paniagua et al. (2015)	Egham, UK	2.45	2000–2012	Infrared analyser and cavity ring-down spectroscopy
		Mace Head, Ireland	1.9	2000–2011	
	Inoue et al. (2006)	Tsukuba, Japan	2	1992–2003	Infrared analyser
	Jain et al. (2005)	Maitri, Antarctica	1.3	2002–2003	Gas chromatography
	Jain et al. (2021)	Gadanki, India	2.5	2016–2019	Cavity ring-down spectroscopy
	Labuschagne et al. (2018)	Cape Point, South Africa	1.65–2.80	1993–2016	Gas chromatography and cavity ring-down spectroscopy
	Liu et al. (2015)	Different sites in the Northern Hemisphere	2.04	1997–2006	
	McClure et al. (2016)	Mt. Bachelor, Oregon	1.48	2012–2014	Cavity ring-down spectroscopy
	Tans et al. (1989)	Point Barrow, Alaska	1.44	1983–1985	Infrared analyser
	Timokhina et al. (2015)	Central Siberia, Russia	2.02	2006–2013	Infrared analyser and cavity ring-down spectroscopy
	Vermeulen et al. (2011)	Cabaw, The Netherlands	2	2005–2009	Infrared analyser
	Wu et al. (2012)	Northeast China	1.7	2003–2010	Infrared analyser
	Zhang et al. (2008)	Seven sites in China	1.7–3.6	2003–2006	Infrared analyser
CH_4	Fang et al. (2016a)	Shangdianzi, China	0.006–0.010	2009–2013	Cavity ring-down spectroscopy
	Guo et al. (2020)	Mt. Waliguan, China	0.0082	2010–2016	Cavity ring-down spectroscopy
	Haszpra et al. (2011)	Hegyhátsál, Hungary	0.017	2007–2009	Gas chromatography
	Jain et al. (2021)	Gadanki, India	0.0111	2016–2019	Cavity ring-down spectroscopy
	Nisbet et al. (2014)	Globally averaged	0.006	2007–2013	
	Pedersen et al. (2005)	Mt. Zeppelin, Norway	0.00334–0.00363	1998–2004	Gas chromatography
	Vermeulen et al. (2011)	Cabaw, The Netherlands	0.0074	2005–2010	Gas chromatography

estimator for establishing the fit between observations and theoretical values. Consequently, using this statistic should be recommended in analyses that deal with comparisons of values.

5. Conclusions

CO_2 and CH_4 concentrations recorded at a rural site in northern Spain over nine years presented a linear evolution. Outlier analysis revealed that 4% of them make a noticeable contribution to the trend. An outlier limit of 23.5% was suggested to obtain a representative trend, which was 2.349 ± 0.012 ppm $year^{-1}$ for CO_2 and 0.00879 ± 0.00004 ppm $year^{-1}$ for CH_4.

Most outliers were obtained for observations above the trend line and their concentrations were higher than those below the trend. However, a surprising response was obtained for scaled outliers. In particular, CH_4 distribution presented contrasting shapes for scaled outliers above the trend line, where saturation is reached quickly, and below the trend line, where saturation is reached slowly.

The qualitative fit between experimental and theoretical cumulative distribution functions for ten skewed distributions used with outliers at the 23.5% limit evidenced the functions that may be discarded, such as the Gumbel, for outliers above the trend line. However, quantitative estimators are required to select the range of the best fit. The modified Nash-Sutcliffe efficiency was the most sensitive statistic for this objective. The fits for outliers above the trend were the worst, especially for CH_4.

The skewed distribution functions determined a noticeable range of the modified Nash-Sutcliffe efficiency values. Among these functions, the Weibull distribution stands out due to its satisfactory agreement.

For central values, the Gaussian distribution stands out when the number of outliers excluded is low and it is replaced by the beta distribution with a high number of outliers excluded.

Once outliers have been isolated and described by distribution functions, the next step is to establish the reasons that determine these outliers. For higher concentrations, they may be due to transport from sources or meteorological factors, such as boundary layer depth. However, determining which reasons cause low concentrations must be investigated. Moreover, this study may be the starting point for more detailed analyses where additional variables are involved, such as wind speed or wind direction. By way of an illustration, outliers and trends may be analysed by wind sectors or by wind speed intervals to explore the contrast between polluted and clean air masses. These research lines remain open with regard to securing a better insight into the evolution of the two gases.

CRediT authorship contribution statement

Isidro A. Pérez: Conceptualization, Formal analysis, Writing – original draft. **M. Ángeles García:** Data curation, Writing – review & editing. **M. Luisa Sánchez:** Funding acquisition. **Nuria Pardo:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the Ministry of Economy and Competitiveness and ERDF funds, project numbers CGL-2009-11979 and CGL2014-53948-P.

References

Aalto, T., Hatakka, J., Paatero, I., Tuovinen, J.P., Aurela, M., Laurila, T., Holmén, K., Trivett, N., Viisanen, Y., 2002. Tropospheric carbon dioxide concentrations at a northern boreal site in Finland: basic variations and source areas. Tellus B 54, 110–126. https://doi.org/10.1034/j.1600-0889.2002.00297.x.

Affolter, S., Schibig, M., Berhanu, T., Bukowiecki, N., Steinbacher, M., Nyfeler, P., Hervo, M., Lauper, J., Leuenberger, M., 2021. Assessing local CO_2 contamination revealed by two near-by high altitude records at Jungfraujoch, Switzerland. Environ. Res. Lett. 16, 044037. https://doi.org/10.1088/1748-9326/abe74a.

Akdağ, S.A., Dinler, A., 2009. A new method to estimate Weibull parameters for wind energy applications. Energy Conv. Manag. 50, 1761–1766. https://doi.org/10.1016/j.enconman.2009.03.020.

Artuso, F., Chamard, P., Piacentino, S., Sferlazzo, D.M., De Silvestri, L., di Sarra, A., Meloni, D., Monteleone, F., 2009. Influence of transport and trends in atmospheric CO_2 at Lampedusa. Atmos. Environ. 43, 3044–3051. https://doi.org/10.1016/j.atmosenv.2009.03.027.

Battista, G., Pagliaroli, T., Mauri, L., Basilicata, C., De Lieto Vollaro, R., 2016. Assessment of the air pollution level in the city of Rome (Italy). Sustainability 8 (838). https://doi.org/10.3390/su8090838.

Belikov, D., Arshinov, M., Belan, B., Davydov, D., Fofonov, A., Sasakawa, M., Machida, T., 2019. Analysis of the diurnal, weekly, and seasonal cycles and annual trends in atmospheric CO_2 and CH_4 at tower network in Siberia from 2005 to 2016. Atmosphere 10, 689. https://doi.org/10.3390/atmos10110689.

Ben-Gal, I., 2005. Outlier detection. In: Maimon, O., Rockach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Kluwer Academic Publishers, Netherlands, pp. 131–146.

Bianchi, S., Plastino, W., di Sarra, A.G., Piacentino, S., Sferlazzo, D., 2020. Carbon dioxide time series analysis: A new methodological approach for event screening categorization. In: Cannarsa, P., Mansutti, D., Provenzale, A. (Eds.), Mathematical Approach to Climate Change and its Impacts. 38. Springer, pp. 201–209. https://doi.org/10.1007/978-3-030-38669-6_7.

Bury, K., 1999. Statistical distributions in engineering. Cambridge University Press, Cambridge, UK, pp. 294–310.

Carrilho, A.C., Galo, M., Dos Santos, R.C., 2018. Statistical outlier detection method for airborne LiDAR data. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 42–1, 87–92. https://doi.org/10.5194/isprs-archives-XLII-1-87-2018.

Cundari, V., Colombo, T., Ciattaglia, L., 1995. Thirteen years of atmospheric carbon dioxide measurements at Mt. Cimone station, Italy. Nuovo Cim. C 18 (1), 33–47. https://doi.org/10.1007/BF02561457.

Fang, S., Tans, P.P., Steinbacher, M., Zhou, L., Luan, T., Li, Z., 2016b. Observation of atmospheric CO_2 and CO at Shangri-La station: Results from the only regional station located at southwestern China. Tellus B 68, 28506. https://doi.org/10.3402/tellusb.v68.28506.

Fang, S.X., Zhou, L.X., Tans, P.P., Ciais, P., Steinbacher, M., Xu, L., Luan, T., 2014. In situ measurement of atmospheric CO_2 at the four WMO/GAW stations in China. Atmos. Chem. Phys. 14, 2541–2554. https://doi.org/10.5194/acp-14-2541-2014.

Fang, S.X., Tans, P.P., Dong, F., Zhou, H., Luan, T., 2016a. Characteristics of atmospheric CO_2 and CH_4 at the Shangdianzi regional background station in China. Atmos. Environ. 131, 1–8. https://doi.org/10.1016/j.atmosenv.2016.01.044.

Fernández-Duque, B., Pérez, I.A., García, M.A., Pardo, N., Sánchez, M.L., 2020. Local regressions for decomposing CO_2 and CH_4 time-series in a semi-arid ecosystem. Atmos. Pollut. Res. 11, 213–223. https://doi.org/10.1016/j.apr.2019.10.012.

Forbes, C., Evans, M., Hastings, N., Peacock, B., 2011. Statistical Distributions. Fourth edition. John Wiley & Sons, New Jersey https://doi.org/10.1002/9780470627242.

Ghitany, M.E., Atieh, B., Nadarajah, S., 2008. Lindley distribution and its application. Math. Comput. Simulat. 78, 493–506. https://doi.org/10.1016/j.matcom.2007.06.007.

Guo, M., Fang, S., Liu, S., Liang, M., Wu, H., Yang, L., Li, Z., Liu, P., Zhang, F., 2020. Comparison of atmospheric CO_2, CH_4, and CO at two stations in the Tibetan Plateau of China. Earth Space Sci. 7, e2019EA001051. https://doi.org/10.1029/2019EA001051.

Harvey, A.C., Peters, S., 1990. Estimation procedures for structural time series models. J. Forecast. 9, 89–108. https://doi.org/10.1002/for.3980090203.

Haszpra, L., Barcza, Z., Szilágyi, I., Dlugokencky, E., Tans, P., 2011. Trends and temporal variations of major greenhouse gases at a rural site in central Europe. In: Haszpra, L. (Ed.), Atmospheric Greenhouse Gases: The Hungarian Perspective. Springer, pp. 29–47.

Hernández-Paniagua, I.Y., Lowry, D., Clemitshaw, K.C., Fisher, R.E., France, J.L., Lanoisellé, M., Ramonet, M., Nisbet, E.G., 2015. Diurnal, seasonal, and annual trends in atmospheric CO_2 at southwest London during 2000-2012: Wind sector analysis and comparison with Mace Head, Ireland. Atmos. Environ. 105, 138–147. https://doi.org/10.1016/j.atmosenv.2015.01.021.

Higuchi, K., Worthy, D., Chan, D., Shashkov, A., 2003. Regional source/sink impact on the diurnal, seasonal and inter-annual variations in atmospheric CO_2 at a boreal forest site in Canada. Tellus B 55, 115–125. https://doi.org/10.1034/j.1600-0889.2003.00062.x.

Inoue, H.Y., Matsueda, H., Igarashi, Y., Sawa, Y., Wada, A., Nemoto, K., Sartorius, H., Schlosser, C., 2006. Seasonal and long-term variations in atmospheric CO_2 and 85Kr in Tsukuba, Central Japan. J. Meteorol. Soc. Jpn. 84, 959–968. https://doi.org/10.2151/jmsj.84.959.

Jain, S.L., Ghude, S.D., Kumar, A., Arya, B.C., Kulkarni, P.S., 2005. Continuous observations of surface air concentration of carbon dioxide and methane at Maitri, Antarctica. Curr. Sci. 88, 1941–1948.

Jain, C.D., Singh, V., Akhil Raj, S.T., Madhavan, B.L., Ratnam, M.V., 2021. Local emission and long-range transport impacts on the CO, CO_2, and CH_4 concentrations at a tropical rural site. Atmos. Environ. 254, 118397. https://doi.org/10.1016/j.atmosenv.2021.118397.

Karaca, F., Alagha, O., Ertürk, F., 2005. Statistical characterization of atmospheric PM10 and PM2.5 concentrations at a non-impacted suburban site of Istanbul, Turkey. Chemosphere 59, 1183–1190. https://doi.org/10.1016/j.chemosphere.2004.11.062.

Kilkki, J., Aalto, T., Hatakka, J., Portin, H., Laurila, T., 2015. Atmospheric CO_2 observations at Finnish urban and rural sites. Boreal Environ. Res. 20, 227–242.

Kontaki, M., Gounaris, A., Papadopoulos, A.N., Tsichlas, K., Manolopoulos, Y., 2011. Continuous monitoring of distance-based outliers over data streams. IEEE 27th International Conference on Data Engineering, Art. No. 5767923, pp. 135–146 https://doi.org/10.1109/ICDE.2011.5767923.

Korkmaz, M.Ç., 2015. Two-sided generalized Gumbel distribution with application to air pollution data. Int. J. Stat. Distrib. Appl. 1, 19–26. https://doi.org/10.11648/j.ijsd.20150101.14.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97. https://doi.org/10.5194/adgeo-5-89-2005.

Kurbatova, J.A., Aleshnovskij, V.S., Kuricheva, O.A., Avilov, V.K., Bezrukova, A.V., Gazaryan, V.A., Chulichkov, A.I., Shapkina, N.E., 2020. Seasonal and interannual variability of CO_2 above the moist tropical forest of southern Vietnam. IOP Conf. Ser. Earth Environ. Sci. 606, 012027. https://doi.org/10.1088/1755-1315/606/1/012027.

Labuschagne, C., Kuyper, B., Brunke, E.-G., Mokolo, T., van der Spuy, D., Martin, L., Mbambalala, E., Parker, B., Khan, M.A.H., Davies-Coleman, M.T., Shallcross, D.E., Joubert, W., 2018. A review of four decades of atmospheric trace gas measurements at cape point, South Africa. Trans. Roy. Soc. S. Afr. 73, 113–132. https://doi.org/10.1080/0035919X.2018.1477854.

Lian, J., Bréon, F.-M., Broquet, G., Lauvaux, T., Zheng, B., Ramonet, M., Xueref-Remy, I., Kotthaus, S., Haeffelin, M., Ciais, P., 2021. Sensitivity to the sources of uncertainties in the modeling of atmospheric CO_2 concentration within and in the vicinity of Paris. Atmos. Chem. Phys. 21, 10707–10726. https://doi.org/10.5194/acp-21-10707-2021.

Liu, M., Wu, J., Zhu, X., He, H., Jia, W., Xiang, W., 2015. Evolution and variation of atmospheric carbon dioxide concentration over terrestrial ecosystems as derived from eddy covariance measurements. Atmos. Environ. 114, 75–82. https://doi.org/10.1016/j.atmosenv.2015.05.026.

Lohila, A., Penttilä, T., Jortikka, S., Aalto, T., Anttila, P., Asmi, E., Aurela, M., Hatakka, J., Hellén, H., Henttonen, H., Hänninen, P., Kilkki, J., Kyllönen, K., Laurila, T., Lepistö, A., Lihavainen, H., Makkonen, U., Paatero, J., Rask, M., Sutinen, R., Tuovinen, J.P., Vuorenmaa, J., Viisanen, Y., 2015. Preface to the special issue on integrated research of atmosphere, ecosystems and environment at Pallas. Boreal Environ. Res. 20, 431–454.

Martínez, J., Saavedra, Á., García-Nieto, P.J., Piñeiro, J.I., Iglesias, C., Taboada, J., Sancho, J., Pastor, J., 2014. Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain). Appl. Math. Comput. 241, 1–10. https://doi.org/10.1016/j.amc.2014.05.004.

Martins, L.D., Wikuats, C.F.H., Capucim, M.N., de Almeida, D.S., da Costa, S.C., Albuquerque, T., Barreto Carvalho, V.S., de Freitas, E.D., de Fátima Andrade, M., Martins, J.A., 2017. Extreme value analysis of air pollution data and their comparison between two large urban regions of South America. Weather Clim. Extremes 18, 44–54. https://doi.org/10.1016/j.wace.2017.10.004.

McClure, C.D., Jaffe, D.A., Gao, H., 2016. Carbon dioxide in the free troposphere and boundary layer at the Mt. Bachelor observatory. Aerosol Air Qual. Res. 16, 717–728. https://doi.org/10.4209/aaqr.2015.05.0323.

Murayama, S., Saigusa, N., Chan, D., Yamamoto, S., Kondo, H., Eguchi, Y., 2003. Temporal variations of atmospheric CO_2 concentration in a temperate deciduous forest in Central Japan. Tellus B 55, 232–243. https://doi.org/10.1034/j.1600-0889.2003.00061.x.

Nisbet, E.G., Dlugokencky, E.J., Bousquet, P., 2014. Methane on the rise – again. Science 343, 493–495. https://doi.org/10.1126/science.1247828.

NOAA, 2021. https://gml.noaa.gov/ccgg/flask.html. (Accessed 10 July 2021).

Nyasulu, M., Haque, M.M., Kumar, K.R., Banda, N., Ayugi, B., Uddin, M.J., 2021. Temporal patterns of remote-sensed tropospheric carbon dioxide and methane over an urban site in Malawi, Southeast Africa: implications for climate effects. Atmos. Pollut. Res. 12, 125–135. https://doi.org/10.1016/j.apr.2021.02.005.

Park, C., Jeong, S., Park, H., Woo, J.-H., Sim, S., Kim, J., Son, J., Park, H., Shin, Y., Shin, J.H., Kwon, S.M., Lee, W.Y., 2021. Challenges in monitoring atmospheric CO_2 concentrations in Seoul using low-cost sensors. Asia-Pac. J. Atmos. Sci. 57, 547–553. https://doi.org/10.1007/s13143-020-00213-2.

Pedersen, I.-T., Holmén, K., Hermansen, O., 2005. Atmospheric methane at Zeppelin Station in ny-Ålesund: presentation and analysis of in situ measurements. J. Environ. Monit. 7, 488–492. https://doi.org/10.1039/b416934d.

Pérez, I.A., Sánchez, M.L., García, M.Á., Pardo, N., 2012. Analysis and fit of surface CO_2 concentrations at a rural site. Environ. Sci. Pollut. Res. 19, 3015–3027. https://doi.org/10.1007/s11356-012-0813-4.

Pérez, I.A., Sánchez, M.L., García, M.A., Pardo, N., 2013. Carbon dioxide at an unpolluted site analysed with the smoothing kernel method and skewed distributions. Sci. Total Environ. 456–457, 239–245. https://doi.org/10.1016/j.scitotenv.2013.03.075.

Pérez, I.A., Sánchez, M.L., García, M.A., Ozores, M., Pardo, N., 2014. Analysis of carbon dioxide concentration skewness at a rural site. Sci. Total Environ. 476–477, 158–164. https://doi.org/10.1016/j.scitotenv.2014.01.019.

Pérez, I.A., Sánchez, M.L., García, M.A., Pardo, N., Fernández-Duque, B., 2019. Influence of dataset density on CO_2 and CH_4 trend calculation. Air Qual. Atmos. Health 12, 613–625. https://doi.org/10.1007/s11869-019-00681-0.

Pérez, I.A., Sánchez, M.L., García, M.A., Pardo, N., Fernández-Duque, B., 2020. Statistical analysis of the CO_2 and CH_4 annual cycle on the northern plateau of the Iberian Peninsula. Atmosphere 11, 769. https://doi.org/10.3390/ATMOS11070769.

Pérez, I.A., García, M.A., Sánchez, M.L., Pardo, N., 2021. Influence of wind speed on CO_2 and CH_4 concentrations at a rural site. Int. J. Environ. Res. Public Health 18, 8397. https://doi.org/10.3390/ijerph18168397.

Sánchez, M.L., Pérez, I.A., García, M.A., 2010. Study of CO_2 variability at different temporal scales recorded in a rural Spanish site. Agric. For. Meteorol. 150, 1168–1173. https://doi.org/10.1016/j.agrformet.2010.04.018.

Sharma, P., Sharma, P., Jain, S., Kumar, P., 2013. An integrated statistical approach for evaluating the exceedence of criteria pollutants in the ambient air of megacity Delhi. Atmos. Environ. 70, 7–17. https://doi.org/10.1016/j.atmosenv.2013.01.004.

Tans, P.P., Thoning, K.W., Elliott, W.P., Conway, T.J., 1989. Background atmospheric CO_2 patterns from weekly flask samples at Barrow, Alaska: Optimal signal recovery and error estimates. NOAA Tech. Mem. ERL ARL-173, pp. 112–123.

Taylor, S.J., Letham, B., 2018. Forecasting at scale. Am. Stat. 72, 37–45. https://doi.org/10.1080/00031305.2017.1380080.

Timokhina, A.V., Prokushkin, A.S., Onuchin, A.A., Panov, A.V., Kofman, G.B., Verkhovets, S.V., Heimann, M., 2015. Long-term trend in CO_2 concentration in the surface atmosphere over Central Siberia. Russ. Meteorol. Hydrol. 40, 186–190. https://doi.org/10.3103/S106837391503005X.

Vermeulen, A.T., Hensen, A., Popa, M.E., Van Den Bulk, W.C.M., Jongejan, P.A.C., 2011. Greenhouse gas observations from Cabauw Tall Tower (1992–2010). Atmos. Meas. Tech. 4, 617–644. https://doi.org/10.5194/amt-4-617-2011.

Wada, A., Sawa, Y., Matsueda, H., Taguchi, S., Murayama, S., Okubo, S., Tsutsumi, Y., 2007. Influence of continental air mass transport on atmospheric CO_2 in the western North Pacific. J. Geophys. Res.-Atmos. 112, D07311. https://doi.org/10.1029/2006JD007552.

Wang, S., Zhu, L., Yan, S., Li, Y., Wang, W., Gao, X., Ma, Z., Liu, P., Liang, M., 2020. Atmospheric CO_2 data filtering method and characteristics of the mole fractions at Wutaishan station in Shanxi of China. Aerosol Air Qual. Res. 20, 2953–2962. https://doi.org/10.4209/aaqr.2020.01.0026.

WDCGG, 2021. https://gaw.kishou.go.jp/. (Accessed 10 July 2021).

Wei, C., Wang, M., Fu, Q., Dai, C., Huang, R., Bao, Q., 2020. Temporal characteristics of greenhouse gases (CO_2 and CH_4) in the megacity Shanghai, China: Association with air pollutants and meteorological conditions. Atmos. Res. 235, 104759. https://doi.org/10.1016/j.atmosres.2019.104759.

Wilks, D.S., 2019. Statistical methods in the atmospheric sciences. Fourth edition. Elsevier, Amsterdam.

WMO, 2020. WMO Greenhouse Gas Bulletin, No. 16. https://library.wmo.int/index.php?id=21795&lvl=notice_display#.YV9A09pBwuU. (Accessed 10 July 2021).

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

Yang, Y., Wang, T., Wang, P., Zhou, M., Yao, B., 2019. In-situ measurement of CO_2 at the Xinglong regional background station over North China. Atmos. Ocean. Sci. Lett. 12, 385–391. https://doi.org/10.1080/16742834.2019.1644949.

Yang, Y., Zhou, M., Wang, T., Yao, B., Han, P., Ji, D., Zhou, W., Sun, Y., Wang, G., Wang, P., 2021. Spatial and temporal variations of CO_2 mole fractions observed at Beijing, Xianghe, and Xinglong in North China. Atmos. Chem. Phys. 21, 11741–11757. https://doi.org/10.5194/acp-21-11741-2021.

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

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. https://doi.org/10.1016/j.scitotenv.2015.04.044.

van Zoest, V.M., Stein, A., Hoek, G., 2018. Outlier detection in urban air quality sensor networks. Water Air Soil Pollut. 229, 111. https://doi.org/10.1007/s11270-018-3756-7.