A procedure to estimate variances and covariances on GHG emissions and inventories

Ernesto C. Marujo\textsuperscript{a}, Gleice G. Rodrigues\textsuperscript{b}, Weber A. N. Amaral\textsuperscript{c}, Fernanda Leonardis\textsuperscript{b} and Arthur Covatti\textsuperscript{b}

\textsuperscript{a}Department of Fundamental Sciences, Instituto Tecnologico de Aeronautica, São José dos Campos, São Paulo, Brazil; \textsuperscript{b}Deep ESG, São José dos Campos, São Paulo, Brazil; \textsuperscript{c}Department of Forest Sciences, Escola de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil

Abstract

This study presents a method for estimating the mean and variance of total CO\textsubscript{2} emission from multiple sources used by a company. The procedure is also readily applicable to estimate these parameters for other greenhouse gases (GHG) inventories and to determine a reliable confidence interval for the total emissions of GHG of a company. Our method represents an improvement over the existing methods that assume independence between emissions from different sources. The foundation of the proposed method is an iterative decomposition process applied to analyze the emissions correlations among activities, raw materials and other inputs used in a company’s operations. From these correlations and the individual estimates of means and variances of emission factors, we show how to generate a confidence interval for the total GHG emission of a company. The application of the method is illustrated for a hypothetical manufacturing plant of bicycles and car toys, whose total CO\textsubscript{2} emission is estimated within a precise confidence interval.

Keywords

GHG emission inventory; correlation of emissions; confidence interval for emissions; emission errors; ESG accounting; carbon accounting

Introduction

Since the 1990s, companies have been encouraged, in some cases required, to measure and report their greenhouse gases (GHG) emissions\textsuperscript{[1]} for different purposes and objectives. Properly accounting of GHG emissions by companies is fundamental to their management, since it enables investors and other stakeholders to assess how companies are mitigating environmental risks while working towards carbon emissions reductions and potentially generating additional sources of revenues and corporate prestige\textsuperscript{[2,3]}. The Intergovernmental Panel on Climate Change (IPCC) leads an effort to develop internationally agreed guidelines for calculating and supporting reports on GHG emissions and inventories\textsuperscript{[4]}. Although the procedures are based on statistical data and parameters\textsuperscript{[5]}, the IPCC\textsuperscript{[6]} stresses that there may be uncertainties in emission factor estimations, particularly because these factors have been established using references from few geographical sources which are not necessarily widely applicable. For companies reporting projects and progress associated with carbon emission reductions, the accuracy and reliability of emission estimates could help stakeholders and markets better evaluate carbon stocks and offsets and, potentially, trade these offsets with reduced risks to sellers and buyers of these credits.

More recently, at the COP 26 (Conference of the Parties for the United Nations Framework Convention on Climate Change, 2021) in Glasgow, countries, companies, and NGOs alike addressed climate goals and commitments aiming the reduction of GHG emissions and mitigation of their impacts.

International bodies and organizations, as well as companies, are adopting ESG (environmental, social and governance) metrics and indicators to report their commitments towards sustainable development. Faulty, loose, or inaccurate measurements or indicators can generate criticism and associate a company with “green washing,” even if it is dedicated to responsible investments\textsuperscript{[7]}. Proper measurement and reporting of indicators...
could remove barriers to the widespread integration of ESG criteria in the management of companies, reducing investment risk. In addition, it could enable necessary links to global standards regimes, such as the ISO (International Organization for Standardization), particularly ISO 14.064-2, which defines rules and processes for greenhouse gases (GHG) emission reduction projects, and requires proper quantification, monitoring procedures, emissions reductions and removal enhancements reporting, and the reduction of errors in the estimation of the total emission of a company.

In this paper, we propose a method to estimate the mean and variance of inventories of GHG of a company with multiple sources of emissions. We address possible causes for emission factor variability and how emissions from different sources could be correlated. Our method is unique in its approach, because it adopts a robust statistical approach applied to GHG inventories and addresses very sensitive risk management issues for carbon projects worldwide. This problem has been recognized as a difficult one. Carotenuto et al. [8] states that: “Uncertainties associated with each step in compiling emission inventories typically sum up, though it is complicated to estimate the total effect due to the difficulty in ascertaining the uncertainty at each step and how uncertainties interact with each other.” Moreover, uncertainties may change over the years with the improvement of emission-producing activities and source characterization [9,10].

Uncertainty analysis is an important aspect of Life Cycle Assessment (LCA), a method widely used to assess the energy efficiency and environmental impact of products, services, and processes, encompassing parts or all their life cycle stages, from raw material production throughout the end-of-life management [11–14]. According to Groen et al. [15], in LCA, uncertainty analysis is usually performed using Monte Carlo or other sampling schemes (latin hypercube, quasi-Monte Carlo, analytical uncertainty propagation, and fuzzy interval arithmetic). Park et al. [16] used the error propagation method instead of the probabilistic approach for identifying key input parameters that affect the uncertainty of the carbon footprint result for cattle dairy farms in Korea. In our understanding, these methods do not address the covariance of emissions and, as Igos et al. [17], indicate, emission studies should discuss the uncertainties associated with the results, trace the sources of uncertainties, including correlations, and estimate errors.

Whitaker et al. [18], in a survey of 44 LCA studies of first-and second-generation biofuels identified the key sources of variability in published emission factors: (1) “real” variability in parameters; (2) “methodological” variability; and (3) “uncertainty” due to rarely included and poorly quantified parameters. The first category may be exemplified in the work of Miranda [19], that showed that electric power generation in Brazil presents a wide variability of GHG emission factors caused by the different characteristics of the power plants, even for the same power generation technology.

Any GHG emission assessment, be it based on Monte Carlo or analytical methods, must include the effect of covariances. This work offers a suggestion on how to treat this issue, avoiding the usual simplistic assumption of independence of emissions from different sources. The correlation between GHG emissions from different sources may be negligible when the scope is restricted to direct emissions only (Scope 1). For all other cases, the correlation of emissions plays an important role in determining the confidence interval for the total emission estimated.

In the next section, we will propose a method to improve the estimation of total GHG emission of a company considering many possible sources of emission and their uncertainties. In comparison to the methods we found in the literature, the novel procedure has the advantage of incorporating the correlations of emissions in the analysis.

Following this introduction, we define the problem in mathematical terms and show the proposed model to deal with variances and covariances in carbon inventory for corporations. We first develop the main idea using a simple case with two inputs and then generalize it to multiple inputs. The method is summarized in terms of the steps of an algorithm. The section ends by stating the parameters and notation that we consider when referring to confidence intervals. We then present the results of applying the method to a manufacturing plant. The paper also discusses the results that show significant improvements in the precision of the confidence interval, using the proposed method. Finally, we elaborate on conclusions regarding the importance of variance and covariance in determining the emission of a plant, in particular of a corporate inventory of GHG.
Materials and methods

Problem definition

Suppose the company of our interest uses $x_i$ units of raw material $i$, for $i = 1..I$, and we want to determine the mean and variance of the combined carbon dioxide (CO$_2$) emissions from these inputs.

We formulate a fundamental underlying assumption to the development of this work. We call it the “total sum assumption,” i.e. the total CO$_2$ emission of the company equals the sum of the individual emissions of each input.

We recognize, however, that there are some important cases where this assumption would be violated. For example, a possible violation of the “total sum assumption” arises when, in a company, two processes share the same source of heat and the emission for both processes are not the sum of the emission of each process in isolation. Nonetheless, the total sum assumption seems to be reasonable in most cases since the reported emission factors already consider the common practice of using, or not, a shared resource. This possibility and the uncertainty represented is an example of the intrinsic variability in the emission factors reported in the literature. Further investigation of conditions where the total sum assumption is not valid is left as a suggestion for future works.

Note that the “total sum assumption” does not imply that the CO$_2$ emissions of all inputs are independent or uncorrelated. In cases where the emission from each source is statistically independent of the emission from every other source, the variance of the emission from all sources is the sum of the variances of the emission of each source. In this case, the estimation of the confidence interval for the total emission is facilitated. However, if the emissions are correlated, we must first determine the correlation between emissions to be able to estimate the variance of total emission.

In the case of CO$_2$ emissions, we observe that the literature presents estimates of emission factors, in terms of mean and variance, for many raw materials [20,21]. Therefore, in this work, we focus on the challenge of determining the covariances between emission factors from different sources. We will concentrate the analysis on the emission of a single GHG, in particular, CO$_2$ or CO$_2$ equivalent, denoted CO$_2$e.

Correlation between emissions may depend on the season of the year, location of the plant, and particularities of its supply chain. Therefore, in principle, the correlation coefficients would have to be specifically determined for each plant.

The proposed method

Consider a company that uses a number of different inputs $i$. Each input may represent a source of energy, a raw material or an activity. Call $p_i$ the amount of input $i$ that the company uses in a certain time interval. The value of $p_i$ should be measured in kWh, kg, m$^3$, or any other appropriate measure for that input (including a monetary measurement, although this should be avoided in preference to physical and stoichiometric measurements). Let us denote by capital letter $P_i$ the emission associated with the use of $p_i$ units of input $i$.

Suppose each input can be characterized as a linear combination of components so that $p_i$ units of input $i$ uses $x_{ij}$ units of component $j$. Let $A_j$ represents the emission associated with the use of $a_j$ units of component $j$.

We consider that these emissions are random variables with means (expected values) $E(P_i)$ and $E(A_j)$ and variances $\text{Var}(P_i)$ and $\text{Var}(A_j)$.

Under the “total sum assumption,” there is no interaction amongst the CO$_2$ emissions of each input. Thus, the emission of a linear combination of inputs is the same linear combination of their individual emissions (Eq. (1)).

$$P_i = \sum_{j=1}^J x_{ij}A_j$$

In some cases, $A_j$ may not be related to a raw material or a physical ingredient but, rather, to a process. Therefore, the assumption includes the possibility of a combination of emissions from different ingredients and/or processes.

The case of two inputs

Let us first consider the simple case where a company uses only two inputs. Their emissions are $P_1$ and $P_2$ and we want to determine the mean and variance of the random variable $P_t$ that represents the total emission of the company in that period (Eq. (2)):

$$P_t = P_1 + P_2$$

From basic probability theory, the mean (Eq. (3)) and variance (Eq. (4)) of this sum of two random variables are:

$$E(P_t) = E(P_1 + P_2) = E(P_1) + E(P_2)$$
\[ \text{Var}(P_i) = \text{Var}(P_1 + P_2) = \text{Var}(P_1) + \text{Var}(P_2) + 2\text{Cov}(P_1, P_2) \quad (4) \]

The difficulty in the computation of the variance formula is that we seldom have an indication of covariances between different inputs. Thus, we propose a procedure to compute \( \text{Cov}(P_1, P_2) \). The procedure is based on the idea that if we decompose \( P_1 \) and \( P_2 \), we will eventually identify common components for which the covariance is obvious and other components that do not present any relationship and, thus, have zero covariance. We decompose \( P_1 \) and \( P_2 \) and compute the covariance according to Eq. (5).

\[ \text{Cov}(P_1, P_2) = \text{Cov} \left( \sum_{i=1}^{j} x_{ij} A_j, \sum_{k=1}^{K} x_{jk} A_k \right) = \sum_{i=1}^{j} \sum_{k=1}^{K} x_{ij} x_{jk} \text{Cov}(A_j; A_k) \quad (5) \]

Therefore, the computation of the covariance of emissions from different inputs would be written in terms of the covariances of emissions of the components of these inputs.

**Generalization for the case of multiple inputs**

We have shown the formula to compute the variance of the sum of only two random variables. The general formula to compute the variance of a linear combination of any number of random variables is facilitated using matrix notation. Let us denote:

- \( X \) = vector of elements \( x_i \) representing the amount of each input \( i \) used by the company
- \( A \) = vector of elements \( A_i \), the random variable representing the emission of each input
- \( M \) = vector of means of \( A_i \)
- \( S \) = vector of standard deviations of \( A_i \)
- \( S_i = \text{StdDev}(A_i) = S(A_i) \)
- \( V \) = covariance matrix with elements \( V_{ij} = \text{Cov}(A_i; A_j) \)
- \( R \) = Rho-matrix made of correlations \( \rho_{ij} = \rho(A_i; A_j) \).

With this notation, the mean (Eq. (6)) and variance (Eq. (7)) of the total emission from the company are:

\[ E(P_i) = E \left( \sum_{i=1}^{j} x_{ij} A_j \right) = E(X' A) = X'M \quad (6) \]

\[ \text{Var}(P_i) = \text{Var} \left( \sum_{i=1}^{j} x_{ij} A_j \right) = X'VX = X'S'R'SX \quad (7) \]

In many cases, the value of the elements of \( R \) might be easily determined because \( \rho_{ij} \) is clearly 0 or 1. In other cases, the estimation of the covariances \( \text{Cov}(A_i; A_m) \) might depend on a decomposition analysis and could be written as (Eq. (8)):

\[ \text{Cov}(A_i; A_m) = \text{Cov} \left( \sum_{j=1}^{J} x_{ij} A_j, \sum_{k=1}^{K} x_{mk} A_{mk} \right) = \sum_{j=1}^{J} \sum_{k=1}^{K} x_{ij} x_{mk} \text{Cov}(A_j; A_{mk}) \quad (8) \]

Equation (8) is a generalization of Eq. (5) where index 1 and 2 were substituted by \( i \) and \( m \), respectively.

**The algorithm**

The following steps summarize the proposed algorithm to estimate the covariance between the emissions from two or more sources and with various levels of decomposition:

1. Determine the composition of the emission of source \( i \), for \( i = 1 \): \( A_i = \sum_{j=1}^{J} x_{ij} A_j \).
2. Estimate the covariances \( \text{Cov}(A_j; A_{mk}) \) by first estimating matrix \( R \) whose elements \( \rho(A_j; A_{mk}) \), represent the correlations between emissions of components \( j \) and \( mk \).
   a. For the cases where \( A_{ij} \) represents the emission of the same component as \( A_{mk} \), the correlation between these is simply \( \rho(A_j; A_{mk}) = 1 \).
   b. For the cases where \( A_{ij} \) and \( A_{mk} \) represent independent emission factors, then \( \rho(A_j; A_{mk}) = 0 \). The independence of the emission from two components could be determined using a set of criteria such as: “Do the components originate from unrelated processes?”; “Are their upstream activities unrelated?”; “Are the emission estimates of these components estimated by independent authors/studies?”; “Do the components not share the same sub-components?”; etc.
   c. For the cases where \( A_{ij} \) and \( A_{mk} \), present variations in opposition, \( \rho(A_j; A_{mk}) = -1 \).
   d. Otherwise, further decompose \( A_{ij} \) and, similarly, \( A_{mk} \), another level. If \( A_{ij} \) is formed by the combination \( A_{ij} = \sum_{n=1}^{N} y_{ijn} A_{jn} \) we substitute the column \( ij \) of matrix \( A \) by \( N \) new columns with the values of \( A_{ijn} \). We
also substitute column \(ij\) of matrix \(S\) by \(N\) new columns with the values of \(S(A_{ijn})\) and column \(ij\) of vector \(X'\) by \(N\) new columns with corresponding values \(x_{ij}y_{ijn}\). Finally, we substitute the row \(ij\) of matrix \(A\) by \(N\) new rows with the values of \(A_{ijn}\); substitute row \(ij\) of matrix \(S\) by \(N\) new rows with the values of \(S(A_{ijn})\).

3. Estimate each element of matrix \(R\) using the same logic as in step 2.
4. Calculate \(\text{Var}(P_3)\) using Eq. (7).

We note that if the company uses a single input and we do not know its emission variance, but know its composition, then we could use the procedure outlined in the above algorithm to estimate the variance of its emission. We also note that the decomposition steps of the procedure might be unnecessary if we begin with a list of inputs already decomposed in such a granularity that allows the assessment of the correlation between these components. In future research we aim to present a list of factors that can cause two emissions to correlate with each other, to complement the findings presented here.

**Confidence interval**

Once we have the estimators for the mean and variance of total emission, we can determine the confidence interval for the true mean emission. We use the following definitions and conventions:

Random variable \(F\) represents the possible values for the emission of a certain company during a certain period.

- \(\mu\) is the true, but unknown, expected value, also called mean, of \(F\).
- \(m\) is a point estimator of \(\mu\).
- \(\sigma^2\) is the true, but unknown, variance of \(F\).
- \(\sigma\) is the square root of the variance of \(F\) and is called standard deviation of \(F\).

\(S^2\) is a point estimator of \(\sigma^2\)

Under usual assumptions of normality, the confidence interval for the mean emission is limited by its Lower Confidence Limit (LCL) and Upper Confidence Limit (UCL) (Eq. (9)):

\[
LCL = \mu - Z_{\alpha/2} \sigma, \quad UCL = \mu + Z_{\alpha/2} \sigma.
\]

We have used the normal probability factor \(Z_{\alpha/2}\) assuming that we know \(\sigma\) or that the variance of \(S\) is small. If that is not the case, it would be necessary to use Student’s \(T\) distribution or even other distributions to compute the confidence interval. For \(\alpha = 5\%\), the Normal probability distribution tabulates \(Z_{0.025} = 1.96\).

In scientific notation, it is usual to express the estimators for the mean and standard deviation separated by a \(\pm\) sign. If a normal distribution is considered, this is equivalent to a confidence interval with \(\alpha = 32\%\), and \(Z_{0.025} = 1.00\).

We stress that confidence intervals differ from prediction intervals which is a wider interval that, with \((1 - \alpha)\) probability, would contain not the mean, but any individual emission value under these circumstances. In order to determine prediction intervals, we need to know (or estimate) the variability of individual values of emission in relation to the mean emission. We will leave the issue of prediction intervals to be investigated in future works and will adhere to the concept of confidence intervals for the cases presented in this document.

**Results**

**A hypothetical case**

In this section, we show an application of the proposed procedure for estimating the confidence interval of total emission of a company, using the hypothetical case of a company with two lines of products: bicycles and toy cars. Tables 1 and 2 below present the composition and emission factor of each element that makes up the bicycle and the toy car, respectively. Emissions could be

---

**Table 1. Components of a bicycle and their emission factors.**

| Components                        | Code | Component’s quantity | Mean emission factor | HWCI 95%** | Std. dev. | Reference |
|----------------------------------|------|----------------------|----------------------|------------|-----------|-----------|
| Aluminum tubes for frame         | A    | 0.73 kg              | 1.7 kg/kg            | (+10%)     | –         | [22]      |
| Cassette and wheels (steel)      | B    | 0.15 kg              | 1.72 kg/kg           | (+25%)     | –         | [23]      |
| Transport-Shipping (diesel)      | C    | 0.251                | 2.88 kg/l            | (–2%)      | –         | [24]      |
| Axles, pedal and ball bearings (steel) | D   | 0.04 kg              | 1.72 kg/kg           | (+25%)     | –         | [23]      |
| Rubber parts                     | E    | 0.08 kg              | 0.2026 kg/kg         | –          | 0.0306    | [25,26]   |
| Energy (all production)          | F    | 0.0045 kwh           | 0.3049 kg/kwh        | –          | 0.2003    | [27–30]   |

Note: Employee travel was negligible since the employees use bicycles to move from home to the plant location.

*The quantities in the table refer to emissions per kg of bicycle. We will consider that the company produces bicycles that weight 19.9 kg each [31] and the use of energy to produce a bicycle is 319 kJ [32].

**HWCI** refers to half width of the 95% confidence interval.

**For Diesel, the relative standard deviation is between 2% and 1%. We considered 2%.**
Table 2. Components of a toy car and their emission factors.

| Components Code | Component's quantity* | Mean emission factor | HWCI 95%** | Std. dev. |
|-----------------|-----------------------|----------------------|------------|-----------|
| Aluminum tubes for frame | G | 0.33 kg | 1.7 kg/kg | (+10%) | 0.206 kg/kg |
| Cassette and wheels (steel) | H | 0.20 kg | 2.86 kg/l | (−2%) | 0.200 kg/l |
| Transport-Shipping (diesel) | I | 0.25 kg | – | – | – |
| Axles, pedal and ball bearings (steel) | J | 0.37 kg | – | – | – |
| Rubber parts | K | 0.10 kg | – | – | – |
| Energy (all production) | L | 0.0059 kwh | 0.3049 kg/kwh | – | 0.0306 kg/kwh |

Note: Employee travel was negligible since the employees use bicycles to move from home to the plant location.

*The quantities in Table 2 refer to emissions per kg of toy car. We will consider that the company produces toy cars that weight 3.0 kg each [33]. **HWCI refers to half width of the 95% confidence interval. ***For Diesel, the relative standard deviation is between 2% and 1%. We considered 2%.

If divided into emission scopes/categories, which is not being done in this work to simplify the example of the procedure.

For simplicity’s sake, we assume that all emissions in the example are of the same scope and belong to the same categories.

The existence of clearly unrelated components was analyzed and, for these cases, a 0 (zero) was assigned to the Rho matrix. This was the case for many pairs of components. For example, A and B are made from different materials and use independent suppliers. Thus, \( \rho(A, B) = 0 \) was assigned.

For cases where the components are correlated, such as the pair (E, K), which denote rubber parts, the value 1 (one) was assigned in the Rho-matrix, presented by Figure 1.

For some pairs of components, assigning an appropriate Rho was not an obvious task and the elements were decomposed until an expert could identify subcomponents that are either common, unrelated or need to be further decomposed. In some cases where the composition or manufacturing process of some elements is not evident, it was pragmatically (or empirically) assigned 0 or 1 based on knowledge of similar products and/or processes by experts. The effects of such approximations were assessed using sensitivity analysis as we comment later in this section.

The algorithm helps to determine correlation coefficients and does not require the experts to assign only zeroes or ones. In fact, it is expected that the correlation coefficients between the emission of two sources to be an intermediate value between 0 and 1 in the common case where some subcomponents are common and some subcomponents are unrelated. The use of the algorithm may establish partial correlation coefficients between components even though the subcomponents are either fully correlated or not correlated at all.

It is interesting to comment on the importance of the algorithm in cases of partial correlations. For instance, suppose a production process that uses two components: A and B. In turn, each of these is made of subcomponents, respectively A₁ and A₂ and B₁ and B₂. Also, consider that the emissions associated with each subcomponent are, on average, approximately equal to each other but they are subject to variability. Finally, assume the emissions associated with A and B are partially correlated and \( \rho(A, B) = 0.5 \) because half of the emissions of A correlates with half of emissions of B and the other halves do not correlate. If the case is more complex, for instance, if the average emissions of subcomponents are not equal, the assessment of \( \rho(A, B) \) is not trivial and the algorithm is advisable. Using the algorithm, the correlation of emissions from A and B would be computed automatically after the values for the correlations \( \rho(A₁, A₂), \rho(A₁, B₁), \rho(A₂, B₁), \rho(A₁, B₂), \rho(A₂, B₂) \) and \( \rho(B₁, B₂) \) are inputted. Of course, there is a range (from −1 to +1) of possible correlations between two emissions, but once the correlations have been estimated, possibly with the use of the algorithm, the computation of the total emission, in terms of its estimated mean, variance and confidence interval, would be a simple computation using the formulas Eqs. (6), (7), and (9).

From the Rho-matrix the covariances are determined using the standard deviations of the emissions of the components. Standard deviations, determined in kg of CO₂, were obtained from the information of confidence intervals that are presented in percentage terms in Tables 1 and 2, in which it was considered that the confidence intervals indicated in refer to intervals of \( (1 - \alpha) = 0.95 \) probability.

It was assumed that symmetrical intervals refer to normal random variables and the standard deviations, used to determine those intervals, are precisely estimated so that the confidence interval is taken inverting the normal distribution and not the Student’s t, for example. Therefore, it was considered that the length of the confidence interval to be \( 2ZₜStdDev \). Also, \( Zₜ=1.96 \), is approximated to
be 2 and the resulting length of the confidence intervals equals 4 times the standard deviation.

For the case of transport (Diesel) the Confidence Interval, given in Table 2, is asymmetrical. Assuming the probability distribution resembles a gamma with \( \alpha = 2 \) and \( \beta = 1 \), and using “Student’s t test tables” of cumulative probability distributions, the 95% confidence interval was found to cover 3.8 standard deviations and, for simplicity, it was assumed to equal 4 times the standard deviation.

With the Rho matrix and the standard deviations, we obtain the covariance matrix \( V \) (Figure 2) of the variances and covariances between emission factors of the components of the bicycles and toy cars.

Finally, the estimated variation of the total CO2 emission was calculated using Eq. (11):

\[
Var_{\text{total}} = X'VX
\]  

(11)

where \( X \) is the vector of components participation.

The estimated standard deviation for the company’s total emission is determined from Eq. (12):

\[
\text{StdDev}_{\text{total}} = \sqrt{Var_{\text{total}}}
\]  

(12)

Thus, considering that in a period of one day the company produces 100 units of the bicycle (Product 1) and 1000 units of the toy car (Product 2). This production corresponds to 1990 kg of Product 1 and 3000 kg of Product 2.

Using the notation “mean ± standard deviation,” the total CO2 emission by the company during this period is estimated to be 11,448 ± 500 kg of CO2.

In comparison, if we assumed the independence of emissions of different components, it would have been assigned \( \rho = 1 \) for all pairs \((i,j)\) where \( i = j \) and \( \rho = 0 \) for all pairs \((i,j)\) where \( i \) do not equal \( j \), and the resulting estimative would be 11,448 ± 317 kg of CO2.

On the other hand, if we considered that all emissions are perfectly correlated, that is, \( \rho = 1 \) for all pairs \((i,j)\), the result would be 11,448 ± 710 kg of CO2. The careful estimation of standard deviation resulted in 500 kg of CO2, that is the intermediate value between the equivalent measure under the hypothesis of complete independence (317 kg of CO2) and under the hypothesis of complete dependence (710 kg of CO2).

In general, sensitivity analysis may help the analyst reach an adequate trade-off between accuracy and simplicity. In this example, the standard deviations were of the same order of magnitude. This robustness is a desirable property when one is not sure about the estimate of some correlation coefficients. The sensitivity analysis for this example is summarized in Figure 3. The figure shows the increase in the standard deviation of total emission of CO2 as a function of the values of \( \rho \) for all pairs of components \((i,j)\) where \( i \) is not equal to \( j \).

In this example, the standard deviation increases noticeably even for small values of correlations. This fact leads to the conclusion that the “total independence of CO2 emissions” is a very strong hypothesis. On the other end, assuming...
that the emissions of all components are pairwise correlated has a relatively minor effect.

We also tested the case in which the emission variances for products $P_1$ and $P_2$ are considered adequate but the covariance between their emissions is approximated to zero. The resulting estimated total emission presented a standard variation (409 kg of CO$_2$) that falls between the two extremes and close to the most accurately estimated value (500 kg of CO$_2$).

**Discussion**

The problem of estimating variance in GHG emissions inventories is of a complex nature due to different sources of errors, spatial and temporal variation of these emissions, difficulty in establishing system's boundaries and confirmation biases, among others. However, different attempts are being made by scientists, stakeholders, and policy-makers to reduce these errors and risks. In this paper, we addressed the errors of statistical nature. In general, the variance of total emission is not equal to the sum of the variances of the emissions from individual sources. The usual procedure is to consider that all of these emissions are uncorrelated. This is a strong assumption, especially if we are using the concept of product carbon footprint (PCF).

We proposed a method to estimate the variance of total emission of a company. Our method is an interactive procedure that requires the identification of components or subcomponents of emission sources until an expert can identify the existence

![Figure 2. Covariance Matrix $V$ between emission factors of the example.](image)

![Figure 3. Changes in standard deviation for total CO$_2$ emission in the example when we vary the correlation coefficients for all pairs of components with non-zero correlation.](image)
or non-existence of correlation of emissions between these components.

We tested the method using data from a fictitious company that produced 1000 bicycles and 100 toy cars in a certain period. Each of these products was decomposed into 6 components. For each of these components a realistic assessment of CO₂ emission, in terms of mean and standard deviation, was obtained from the IPCC database and reports found in the specialized literature.

The elements $q_{ij}$ of the correlation matrix Rho, were evaluated by experts. We used the expertise of 4 professionals who work in carbon emission estimating and accounting for corporations. Variance and covariances are not usually investigated in depth in studies of GHG emissions for corporations but the experts have a very solid academic background in concepts of variance, correlations and covariances and in estimating emissions at a corporate level. The correlation coefficient matrices were estimated first by each expert separately and then in groups. The final result emerged in only one meeting because there was no controversy in the assignment of the correlations. The important task for the success of the exercise was the characterization of the company, its production process and its supplier’s location, size and age.

There were 144 elements of matrix Rho to be evaluated. Of those, all 12 $q_{ij}$ where $i=j$ are necessarily 1. Because of symmetry, only 66 of the remaining $q_{ij}$ need to be estimated. Of these, a total of 19 $q_{ij}$ were considered to be 1 (indicating perfect correlation between emission of elements i and j). This was a trivial task for most cases because they involve i and j that represent the same element, for instance, rubber materials. The only cases where an assumption was used without justification are the remaining 47 $q_{ij}$’s that were considered to be 0.

Using this procedure, the variance of total emission was estimated and the 67% confidence interval for the total emission of CO₂ was estimated as: 11,448 ± 500 kg of CO₂. To our knowledge, the common practice is to disregard the covariances of emissions of different components. The result of this simplistic assumption would be the interval 11,448 ± 317 kg of CO₂.

Notice that the amplitude of the confidence interval using the simplistic method is smaller in comparison with the confidence interval obtained by the method that we have proposed and advocated in this article as accurate.

The total variance is an additive function of individual variances and covariances. Therefore, unless we are facing unusual cases where there are negative correlations in the emissions, the variability of total CO₂ emission is underestimated when we use the simplistic method of neglecting correlations.

Conclusion

The computation of the variance of total emission depends on estimating correlations. The existing methods, however, usually assume independence among emissions from different sources, thus avoiding the burden of estimating correlations, but this approximation deteriorates the accuracy of total emission estimates. Monte Carlo methods would also be more precise if they incorporated correlations into the models.

Using the method proposed in this work, the task of correlations estimation is facilitated. The method is based on decomposing activities until an expert is eventually able to identify identical components that have perfectly correlated emissions; components that are unrelated or intermediate cases, where the correlation is estimated based on experience or other methods. As a suggestion to future works, researchers could investigate factors that influence emission correlations and criteria that would determine non-correlation, for example, between fugitive emission and energy.

We also have to recognize that it is not reasonable to assume that the objective of GHG management of a company would be to form a portfolio of activities with small or negative correlated emissions. Contrary to finance portfolios, the important issue regarding GHG emission is not so much the variance of emission in a short time interval, but the cumulative emission over a long period of time. For GHG emissions, the important management actions would be directed to curbing the upper tail of emissions. Therefore, the variance may not be the preferred measure for management to control.

Thus, future works may prefer to substitute confidence intervals by other measures such as “maximum possible value” or “peak month emission” or other measurements of environmental impact. In these cases, an accurate estimate of the variance of emissions is crucial and the issues discussed in this paper would be even more important.

In this work, we have assumed that we could neglect the cross-interference of CO₂ emissions
from different sources. Using this assumption, the CO₂ emission from source A and from source B would be computed in isolation, regardless if A and B are processed in conjunction or not. Future works should investigate under what conditions this is valid. If the assumption of no cross-interference is proven inadequate, not only would it be necessary to reform the developments for the computation of the variance of the emission but, more importantly, it would also invalidate the computation of the mean of emission which would no longer be the sum of expectations of individual emissions.

The method outlined in this paper represents a contribution to improving the quality of a company’s GHG emission estimates. Better accuracy and confidence in emissions reporting might contribute to adequate certification and feasibility analysis of carbon projects worldwide, reducing uncertainties and risks of these projects, including financial and reputational ones.

Data availability statement
Data sharing not applicable—the data used are fictitious.

Disclosure statement
No potential competing interest was reported by the authors.

References
1. Kauffmann C, Less CT, Teichmann D. Corporate greenhouse gas emission reporting: a stocktaking of government schemes. Paris: OECD Working Papers on International Investment, 2012/01, OECD Publishing; 2012. p. 74.
2. Inakollu S, Morin R, Keefe R. Carbon footprint estimation in fiber optics industry: a case study of OFS estufa da geração de energia elétrica no Brasil: implicações da aplicação da avaliação do ciclo de vida. [Greenhouse gas emission factor from electricity generation in Brazil: implications from life cycle assessment adoption] [dissertation]. São Carlos (SP): Universidade de São Paulo; 2012.
3. Carbon Disclosure Project (CDP). Global 500 report. London: CDP; 2009. p. 49.
4. Yona L, Cashore B, Jackson RB, et al. Refining national greenhouse gas inventories. Ambio. 2020;49(10):1581–1586. doi:10.1007/s13280-019-01312-9.
5. Tonin S, La Notte A, Nocera S. A use-chain model to deal with uncertainties. A focus on GHG emission inventories. Carbon Manag. 2016;7(5–6):347–359. doi:10.1080/17583004.2016.1249217.
6. Intergovernmental Panel on Climate Change (IPCC). IPCC guidelines for national greenhouse gas inventories. Volume 1: General guidance and reporting. Kanagawa: National Greenhouse Gas Inventories Programme, IPCC; 2006.
7. Brandon RG, Krueger P, Schmidt PS. ESG rating disagreement and stock returns. Finance Working Paper N° 651/2020. European Corporate Governance Institute; 2021. p. 57. doi:10.2139/ssrn.3433728.
8. Carotenuto F, Guaitieri G, Miglietta F, et al. Industrial point source CO₂ emission strength estimation with aircraft measurements and dispersion modelling. Environ Monit Assess. 2018;190(3):165. doi:10.1007/s10661-018-6531-8.
9. Jonas M, Marland G, Krey V, et al. Uncertainty in an emissions-constrained world. Clim Change. 2014;124(3):459–476. doi:10.1007/s10584-014-1103-6.
10. Lesiv M, Bun A, Jonas M. Analysis of change in relative uncertainty in GHG emissions from stationary sources for the EU 15. Clim. Change. 2014;124(3):505–518. doi:10.1007/s10584-014-1075-6.
11. Keoleian GA, Spitzley DV. Chapter 7. Life cycle based sustainability metrics. In: Sustainability science and engineering. Vol. 1(127). 2006. p. 159. doi:10.1016/S1871-2711(06)80014-0.
12. Meinrenken CJ, Chen D, Esparza RA, et al. Carbon emissions embodied in product value chains and the role of Life Cycle Assessment in curbing them. Sci Rep. 2020;10(1):6184. https://www.nature.com/articles/s41598-020-62030-x doi:10.1038/s41598-020-62030-x.
13. Zabed H, Sahu JN, Suely A, et al. Bioethanol production from renewable sources: current perspectives and technological progress. Renew Sust Energy Rev. 2017;71:475–501. doi:10.1016/j.rser.2016.12.076.
14. Lima RS, Caldeira-Pires AA, Cardoso AN. Uncertainty analysis in life cycle assessment applied to bioenergies systems: a critical review of the literature. Process Integr Optim Sustain. 2020;4(1):1–13. doi:10.1007/s41660-019-00103-9.
15. Groen EA, Heijungs R, Bokkers EAM, et al. Methods for uncertainty propagation in life cycle assessment. Environ Model Softw. 2014;62:216–325. doi:10.1016/j.envsoft.2014.10.006.
16. Park Y-S, Yeon S-M, Lee G-Y, et al. Proposed consecutive uncertainty analysis procedure of the greenhouse gas emission model output for products. Sustainability. 2019;11(9):2712. doi:10.3390/su11092712.
17. Igos E, Benetto E, Meyer R, et al. How to treat uncertainties in life cycle assessment studies? Int J Life Cycle Assess. 2019;24(4):794–807. doi:10.1007/s11367-018-1477-1.
18. Whitaker J, Ludley KE, Rowe R, et al. Sources of variability in greenhouse gas and energy balances for biofuel production: a systematic review. GCB Bioenergy. 2010;9(9):99–112. doi:10.1111/j.1757-1707.2010.01047.x.
19. Miranda MM. Fator de emissão de gases de efeito estufa da geração de energia elétrica no Brasil: implicações da aplicação da avaliação do ciclo de vida. [Greenhouse gas emission factor from electricity generation in Brazil: implications from life cycle assessment adoption] [dissertation]. São Carlos (SP): Universidade de São Paulo; 2012.
20. Losekann L, Marrero GA, Ramos-Real FG, et al. Efficient power generating portfolio in Brazil: conciliating cost, emissions and risk. Energy Policy. 2013;62:301–314. doi:10.1016/j.enpol.2013.07.049.
21. Buchwitz M, Reuter M, Schneising O, et al. Computation and analysis of atmospheric carbon dioxide annual mean growth rates from satellite observations during 2003–2016. Atmos Chem Phys. 2018;18(23):17355–17370. doi:10.5194/acp-18-17355-2018.
22. Intergovernmental Panel on Climate Change (IPCC). CO₂ emission factor for primary aluminium production. EFDB emission factor database (ID 214133). [cited 2021. Sept 17]. Available from: https://www.ipcc-nggip.iges.or.jp/EFDB/ef_detail.php.
23. Intergovernmental Panel on Climate Change (IPCC). CO₂ emission factor for iron & steel production. EFDB emission factor database (ID 214096). [cited 2021. Sept 17]. Available from: https://www.ipcc-nggip.iges.or.jp/EFDB/ef_detail.php.
24. Intergovernmental Panel on Climate Change (IPCC). Mobile Combustion. Chapter 3. [cited 2021 Sept 14]. Available from: https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/2_Volume2/V2_3_Ch3_Mobile_Combustion.pdf.
25. Prefeitura Municipal de Campinas. Produto 4 - Cálculo das emissões de GEE da região metropolitana de campinas [product 4 - Calculation of GHG emissions in the metropolitan region of campinas]. Campinas: Way Carbon; 2018. p. 89.
26. Environmental Declaration Office (KEITI). [Internet]. LCI DB Raw materials and Energy. [cited 2021. Oct 5]. Available from: http://www.epd.or.kr/eng/ldi/lciCo200.do.
27. Sistema de Registro Nacional de Emissões (SIRENE). [Internet]. Fatores de emissão (MDL/SIN). [cited 2021. Oct 2]. Available from: https://sirene.mctic.gov.br/portal/openems/textoGeral/2020/fatores-de-emissao.html.
28. Sanquetta CR, Flizikowski LC, Corte AP, et al. Estimativa das emissões de gases de efeito estufa em uma obra de construção civil com a metodologia GHG protocol [estimating GHG emissions for a civil construction building by using GHG protocol methodology]. Encicl Biosf. 2013;9:1088–1106.
29. Calili RF. Políticas de eficiência energética no Brasil: uma abordagem em um ambiente econômico sob incerteza. [Energy efficiency policies in Brazil: an economic environment under uncertainty approach] [thesis]. Rio de Janeiro (SP): Pontifícia Universidade Católica do Rio de Janeiro; 2013.
30. Rehm Thermal Systems GmbH. [Internet]. Top soldering results with a low CO₂ footprint. [cited 2021. Oct 22]. Available from: https://www.rehm-group.com/en/news/blog/details/top-soldering-results-with-a-low-co2-footprint.html.
31. European Cyclists Federation (ECF). Quantifying CO₂ savings of cycling. Brussels: European Cyclists’ Federation ASBL; 2016. p. 28.
32. Lacap J, Barney R. Calculating changes in CO₂ emissions as a result of increased cycling. Davis: University of California; 2015. p. 51.
33. Design decisions Wiki. [Internet]. Scooter. [cited 2021. Nov 01]. Available from: https://wiki.ece.cmu.edu/ddi/index.php/Scooter#Manufacturing.