

A Copula-Based Three-Dimensional Distribution to Study Pollutants Monthly Maxima: A Case Study Considering Three Regions in Mexico City

Juan A. Vazquez-Morales¹, Eliane R. Rodrigues^{2*}, Hortensia J. Reyes-Cervantes¹

¹Facultad de Ciencias Físico-Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

²Instituto de Matemáticas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

Email: *eliane@math.unam.mx

How to cite this paper: Vazquez-Morales, J.A., Rodrigues, E.R. and Reyes-Cervantes, H.J. (2026) A Copula-Based Three-Dimensional Distribution to Study Pollutants Monthly Maxima: A Case Study Considering Three Regions in Mexico City. *Journal of Environmental Protection*, 17, 594-609. <https://doi.org/10.4236/jep.2026.177030>

Received: May 29, 2026

Accepted: July 4, 2026

Published: July 7, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this work, we consider three-dimensional distributions obtained using copulas to jointly study pollutants monthly maxima collected in three regions of Mexico City. The copula models taken into account are the usual Gumbel-Hougaard and a nested copula with this copula as the base copula. Fréchet distributions are used as the marginal distributions in the copula formulations. Parameters of the models are estimated under the Bayesian point of view via Markov chain Monte Carlo algorithms. The models are applied to three pollutants obtained from the monitoring stations placed in three regions of the Mexico City metropolitan area.

Keywords

Copula, Extreme Value Distribution, Bayesian Inference, Air Pollution, Mexico City

1. Introduction

Mexico City is among many cities whose inhabitants suffer the consequences of high levels of air pollution. As many other countries and cities around the world, Mexico and, in particular, Mexico City have implemented actions aiming to reduce pollutants emission, as well as to prevent population exposure to high levels of pollution. As part of the actions taken are the large monitoring network located in Mexico City and its metropolitan area, and many regulations that have been

approved at local and country levels. Among them are [1]-[6]. In particular, [1] establishes the levels of ozone (O_3) and particulate matter with diameter smaller than 10 microns (PM_{10}) in order to declare emergency alerts in Mexico City. Depending on how many locations their corresponding environmental thresholds are exceeded a type of emergency alert is declared. These pollutants are also part of the set of pollutants used to classify the air quality in Mexico City [2]. In addition to these two pollutants, we also have particulate matter with diameter smaller than 2.5 microns ($PM_{2.5}$). The interest in studying these three pollutants is mainly because of their potential well known harmful impact on human health. For instance, if we have ozone levels above 0.11 parts per million (0.11 ppm), the ill, newborn, and elderly may experience serious health deterioration (see, for example, [7]-[10], among others). Additionally, exposure of pregnant women to PM_{10} and $PM_{2.5}$ may produce adverse effects on the newborn [11]-[13]. PM_{10} and $PM_{2.5}$ may also cause cardiovascular problems in the population in general and an increase in mortality of at-risk groups [14]-[17]. Therefore, to study the behavior of these and other pollutants, and in particular, the behavior of concentrations that present extreme values when compared to the remaining measurements are important issues. If extreme values are of interest, then extreme value models are one way of analyzing this problem.

Many issues in the environmental area are analysed taking into account the behavior of extreme values, see for instance [18]-[22] when the univariate approach is taken. When it comes to higher dimensional settings, [23] presents a bivariate extreme value model using a copula approach. The copula approach was also considered in the study of a variety of problems. For instance, [24]-[26] have applied this methodology to study multivariate flood frequency, and [27] have used it to analyze rainfall frequency. We also have [28] where a multivariate multi-parameter extreme value model is considered using a copula approach, as well as [29] using copula to study multidimensional association with an application to air pollution problems.

In the present study, we also consider extreme value and copula models, but now to study the association among pollutants obtained from three regions in the Mexico City metropolitan area. The pollutants considered are O_3 , PM_{10} , and $PM_{2.5}$ where their monthly maxima collected in regions northeast (NE), center (CE), and southwest (SW) are used as data. These three regions are taken into account because they are in the main prevailing wind path which goes from the northeast region of the city to the southwest. In a first instance, we will consider three three-dimensional vectors where in the first coordinate we have the pollutants measurements collected in region NE; in the second coordinate we have the same pollutant collected in region CE; and in the third coordinate are the data collected in region SW. Therefore, we consider the vectors $(O_3^{(NE)}, O_3^{(CE)}, O_3^{(SW)})$, $(PM_{10}^{(NE)}, PM_{10}^{(CE)}, PM_{10}^{(SW)})$, and $(PM_{2.5}^{(NE)}, PM_{2.5}^{(CE)}, PM_{2.5}^{(SW)})$, with pollutant^(x) meaning the pollutant measurements collected in region x . In a second instance, a nested copula [24] is taken into account. In this nested model, the order in which regions appear depends on the associated correlations between pairs of pollutants in the

corresponding region. The specific ordering will be given when the model is applied to the data sets. In all cases, the base copula used is the Gumbel-Hougaard (see [30]-[32]) using the Fréchet extreme value distribution.

The novelties here are the copula and dimension used, as well the data sets. The differences from previous works also using data from Mexico City is the observational period, the data taken into account, and the fact that we want to see the association of pollutants collected in different regions.

This work is organized as follows. In Section 2, we present the mathematical and the Bayesian formulations of the models. Section 3 gives an application to the case of Mexico City monthly maximum measurements of the three pollutants considered here. In Section 4, a discussion of the results, as well as some additional comments are presented. Two appendices, given after the list of references, present the explicit forms of the likelihood functions and some of the computational details.

2. The Stochastic Model

Denote by $Z_t^{(i)}$ the concentration of the pollutant in study at the i th site at time t ; $i = 1, 2, \dots, d$; $t = 1, 2, \dots, T$, ($T > 0$, $d \geq 1$). (In the present case a site will correspond to a region in the Mexico City metropolitan area). Let

$\mathbf{Z}^{(i)} = \{Z_t^{(i)} : t = 1, 2, \dots, T\}$ be the process recording the pollutant concentration values during the observational period $[0, T]$ at site i , $i = 1, 2, \dots, d$. We assume that

$Z_t^{(i)}$ has distribution and density functions $F_i(\cdot)$ and $f_i(\cdot)$, respectively, $i = 1, 2, \dots, d$; $t = 1, 2, \dots, T$. Denote by $F_{ijk}(\cdot, \cdot, \cdot)$ and $f_{ijk}(\cdot, \cdot, \cdot)$ the joint distribution and density functions of a given pollutant collected at sites i , j , and k , respectively; $i, j, k = 1, 2, \dots, d$; $i \neq j, k$; $j \neq k$. We assume that $F_{ijk}(\cdot, \cdot, \cdot)$ is given by a copula function $C_\theta(\cdot, \cdot, \cdot)$, i.e., $F_{ijk}(\cdot, \cdot, \cdot) = C_\theta(F_i(\cdot), F_j(\cdot), F_k(\cdot))$; $i \neq j, k$, $j \neq k$; $i, j, k = 1, 2, \dots, d$.

Due to the nature of the correlation between pairs of measurements ([23]), in the present case we assume a three-dimensional version of the Gumbel-Hougaard copula [30]-[32] for the triplets distribution functions, as well as for the base copula in the nested model. Therefore, two cases are taken into account. One in which we use the Gumbel-Hougaard (GH) copula in its three-dimensional version where only one association parameters is present, i.e., we consider the copula given by

$$C_\theta(u, v, w) = \exp \left[- \left([-\log(u)]^\theta + [-\log(v)]^\theta + [-\log(w)]^\theta \right)^{1/\theta} \right], \quad (1)$$

where $\theta \in [1, \infty)$ is the association parameter. (Note that when $\theta = 1$ we have independence of the three random variables whose joint distributions we are trying to model).

In another case the three-dimensional model is obtaining by nesting Gumbel-Hougaard (NGH) copulas in which two association parameters are present. This is a generalization of the Arquimedean copulas. Hence, when the nested copula [24] is taken into account, we consider the following formulation,

$$C(u, v, w) = C_1(w, C_2(v, u)), \quad (2)$$

where $C_1(\cdot, \cdot)$ and $C_2(\cdot, \cdot)$ are copulas with association parameters θ_1 and θ_2 , respectively, with $\theta_2 > \theta_1$.

Remarks. 1) If the copulas $C_1(\cdot, \cdot)$ and $C_2(\cdot, \cdot)$ are symmetric, i.e., $C_i(u, v) = C_i(v, u)$, $i = 1, 2$, then

$$C_1(w, C_2(v, u)) = C_1(C_2(u, v), w).$$

In the present case, we take $C_i(\cdot, \cdot)$ a Gumbel-Hougaard copula with association parameter θ_i , $i = 1, 2$; i.e., $C_{\theta_i}(u, v) = \exp\left[-\left([\log(u)]^{\theta_i} + [\log(v)]^{\theta_i}\right)^{1/\theta_i}\right]$

2) The Gumbel-Hougaard copula is used because the interest resides in working with max-stable copulas since the data under analysis correspond to monthly maxima. Another reason for this choice is that the Gumbel-Hougaard copula captures the possible dependence between measurements in different regions in the upper tail of each marginal distribution. Other copulas that satisfy these assumptions, such as the Galambos and Hüsler-Reiss copulas, have also been considered and comparative analyses show no significant differences with respect to the Gumbel-Hougaard copula, see, for instance, [33].

When $F_{ijk}(\cdot, \cdot, \cdot)$ is given by (1), the associated joint density function $f_{ijk}(\cdot, \cdot, \cdot)$ is

$$f_{ijk}(x, y, z) = f_i(x) f_j(y) f_k(z) c_{\theta}(F_i(x), F_j(y), F_k(z)), \quad (3)$$

where $c_{\theta}(u, v, w) = \frac{\partial^3 C_{\theta}(u, v, w)}{\partial u \partial v \partial w}$ is given by

$$c_{\theta}(u, v, w) = \frac{1}{uvw} (-\log u)^{\theta-1} (-\log v)^{\theta-1} (-\log w)^{\theta-1} \exp(-\kappa_1^{1/\theta}) \kappa_1^{-3+1/\theta} \\ \times \left[1 + 2\theta^2 + 3\theta(-1 + \kappa_1^{1/\theta}) - 3\kappa_1^{1/\theta} + \kappa_1^{2/\theta} \right],$$

with $\kappa_1 = (-\log u)^{\theta} + (-\log v)^{\theta} + (-\log w)^{\theta}$.

In the case where $F_{ijk}(\cdot, \cdot, \cdot)$ is given by (2) we have

$$f_{ijk}(x, y, z) = f_i(x) f_j(y) f_k(z) c_{\theta_1, \theta_2}(F_i(x), F_j(y), F_k(z)), \quad (4)$$

where $c_{\theta_1, \theta_2}(u, v, w) = \frac{\partial^3 C_{\theta_1, \theta_2}(u, v, w)}{\partial u \partial v \partial w}$ is given by

$$c_{\theta_1, \theta_2}(u, v, w) = \frac{1}{uvw} (-\log u)^{\theta_2-1} (-\log v)^{\theta_2-1} (-\log w)^{\theta_1-1} \kappa_2^{\frac{1}{\theta_1}-3} \kappa_3^{\frac{\theta_1}{\theta_2}-2} \\ \exp\left(-\kappa_2^{\frac{1}{\theta_1}}\right) \left\{ \kappa_3^{\frac{\theta_1}{\theta_2}} \left[1 - 3\kappa_2^{\frac{1}{\theta_1}} + \kappa_2^{\frac{2}{\theta_1}} \right] + \theta_1^2 \left(\kappa_3^{\frac{\theta_1}{\theta_2}} - (-\log w)^{\theta_1} \right) \right. \\ \left. + \theta_1 \left(-1 + \kappa_2^{\frac{1}{\theta_1}} \right) \left(2\kappa_3^{\frac{\theta_1}{\theta_2}} - (-\log w)^{\theta_1} \right) + \theta_1 \theta_2 \kappa_2 \right. \\ \left. + \kappa_2 \theta_2 \left(-1 + \kappa_2^{\frac{1}{\theta_1}} \right) \right\},$$

with $\kappa_2 = \left((-\log u)^{\theta_2} + (-\log v)^{\theta_2} \right)^{\frac{\theta_1}{\theta_2}} + (-\log w)^{\theta_1}$ and $\kappa_3 = (-\log u)^{\theta_2} + (-\log v)^{\theta_2}$

In the present case, we are studying monthly maximum measurements and, when using a copula in order to obtain the joint distributions of the pollutants measurements, we will use as marginal distributions Fréchet(α , σ) distributions which is given as follows ([34]). A random variable X is said to have Fréchet(α , σ) distribution if its distribution ($F_X(\cdot)$) and density ($f_X(\cdot)$) functions are of the following forms,

$$F_X(x) = \exp \left[- \left(\frac{x}{\sigma} \right)^{-\alpha} \right] \text{ and } f_X(x) = \frac{\alpha}{\sigma} \left(\frac{x}{\sigma} \right)^{-\alpha-1} \exp \left[- \left(\frac{x}{\sigma} \right)^{-\alpha} \right], \quad (5)$$

with $x, y, \alpha, \sigma > 0$.

In the present case, the Fréchet marginal distribution is considered based on the results reported by [23], where the marginal analysis shows that it provides a better fit compared to other extreme value distributions.

Denote by $(Z_t^{(i)}, Z_t^{(j)}, Z_t^{(k)})$ the random vector recording the pollutant concentration we are studying, $t = 1, 2, \dots, T$; $i, j, k = 1, 2, \dots, d$; $i \neq j, k$; $j \neq k$. (In the present work, each coordinate represent the monthly maximum concentrations of a given pollutant in one of the regions in the main wind corridor in Mexico City metropolitan area). Let (α_i, σ_i) , (α_j, σ_j) , and (α_k, σ_k) denote the vector of parameters of the distribution functions $F_i(\cdot)$, $F_j(\cdot)$, and $F_k(\cdot)$, respectively. We also assume that $F_{ijk}(\cdot, \cdot, \cdot)$ is given either by the copula (1) or by the copula (2) with $f_{ijk}(\cdot, \cdot, \cdot)$ the associated density function. Hence, in all cases the vector of parameters to be estimated is either $\boldsymbol{\phi}_{GH} = (\alpha_i, \sigma_j, \alpha_i, \sigma_j, \alpha_k, \sigma_k, \theta)$ or $\boldsymbol{\phi}_{NGH} = (\alpha_i, \sigma_j, \alpha_i, \sigma_j, \alpha_k, \sigma_k, \theta_1, \theta_2)$ where α and σ are the parameters of the particular Fréchet distribution, θ is the association parameter present in the copula given by (1), and θ_1 and θ_2 are the association parameters present in the nested copula (2).

Estimation of the parameters will be performed under the Bayesian point of view using information provided by the so-called posterior distribution of the vector of parameters. If $\boldsymbol{\phi}$ is the vector of parameters of a model describing a dataset $\mathbf{D}$, then the posterior distribution of $\boldsymbol{\phi}$, denoted by $P(\boldsymbol{\phi} | \mathbf{D})$, is such that $P(\boldsymbol{\phi} | \mathbf{D}) \propto L(\mathbf{D} | \boldsymbol{\phi}) P(\boldsymbol{\phi})$, where $L(\mathbf{D} | \boldsymbol{\phi})$ and $P(\boldsymbol{\phi})$ are, respectively, the likelihood function of the model and the so-called prior distribution of $\boldsymbol{\phi}$. The task is made easier by using the software OpenBugs ([35] [36]) where only the likelihood function of the model and the prior distributions of the parameters need to be specified.

In the present case, the observed data are given by the set

$\mathbf{D} = \left\{ (z_t^{(i)}, z_t^{(j)}, z_t^{(k)}), t = 1, 2, \dots, T; i, j, k = 1, 2, \dots, d; i \neq j, k; j \neq k \right\}$. The prior distributions of the parameters will be specified when the models are applied to the Mexico City data. In all cases we will assume their prior independence. The gen-

eral form of the likelihood function of the model is given by,

$$L(\mathbf{D}|\boldsymbol{\varphi}) = \prod_{t=1}^T f_{ijk}\left(z_t^{(i)}, z_t^{(j)}, z_t^{(k)}\right), \quad (6)$$

with $f_{ijk}(\cdot, \cdot, \cdot)$ the appropriate joint density function.

Remark. Note that we are assuming that the triplet containing the data from the different regions are temporally independent. This assumption presents a limitation of the present model. However, when we look at the application to the present data, this limitation seems not to pose a problem and the fit of the estimated distributions to the observed are good. In order to make the model more general we could consider a temporal Markovian dependence on the triplet or even some type of modification in order to include some seasonal dependence. However, this is not pursued here.

3. Application to Mexico City Data

Application will be made to O_3 , PM_{10} , and $PM_{2.5}$ measurements collected at Mexico City monitoring network

(<http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27>) in the three regions of Mexico City in the main wind path, *i.e.*, regions NE, CE, and SW. Pollutants concentrations are measured in parts per billion (ppb) in the case of O_3 and in micrograms per cubic meter ($\mu g/m^3$) when we consider PM_{10} and $PM_{2.5}$. Monitoring stations are place in several sites in these regions. Measurements in each station are obtained minute by minute and the averaged hourly result is reported at each station. The data actually used in the application are the monthly maximum measurements of these pollutants. The monthly maximum measurement in a given month in a given region is the maximum of all measurements collected in all stations located in that region at the given month.

Remarks. 1. In the cases where a monitoring station was removed or introduced in a given region during the observational period, the data from that station were included only for the time interval in which the station was operational. A comparative analysis of the maxima with and without these stations was conducted, and no significant differences were found. The resulting statistics were similar, no outliers were detected, and the interquartile ranges remained consistent.

2) Even though, in the case of ozone, the estimation of the parameters is made using the original unit of measure (*i.e.*, ppb) the discussion of the results is made in terms of ppm since legislations are stated in this unit of measure.

Depending on the pollutant we have an observational period. In the case of ozone the data used were collected from January 1990 to December 2021 giving a total of $T = 384$ monthly maxima; from January 1995 to December 2021 in the case of PM_{10} , giving a total of $T = 324$ measurements; and from August 2003 to December 2021 for the $PM_{2.5}$ data with a total of $T = 221$ values. Analysis will be performed for each pollutant separately. Hence, we have that each coordinate of the vector will represent a specific pollutant collect at a specific region.

The observation period for ozone data started in January 1990 because from that year on records are considered more reliable. In the case of PM_{10} and $PM_{2.5}$, the starting data of the observational periods correspond to the points in time where measurements of these pollutants became available in the regions taken into account in this study. In all cases, the observational period ends December 2021.

Since a Bayesian estimation of the parameters is performed and the software OpenBugs is used, we need to specify both the likelihood functions of the models and the prior distributions of the vectors of parameters. The explicit forms of the likelihood functions are given in **Appendix A**. The prior distributions, as well as the computational details such as burn-in periods, sampling gaps, and sample sizes are presented in **Appendix B**.

When the usual three-dimensional Gumbel-Hougaard copula (*i.e.*, (1)) is used we have that $(Z_t^{(i)}, Z_t^{(j)}, Z_t^{(k)})$ corresponds to measurements in regions NE, CE, and SW. In the case of the nested version (*i.e.*, the copula given by (2)), $(Z_t^{(i)}, Z_t^{(j)}, Z_t^{(k)})$ represents the pollutants O_3 and PM_{10} in the following order of regions (CE, SW, NE). In the case of $PM_{2.5}$, $(Z_t^{(i)}, Z_t^{(j)}, Z_t^{(k)})$ represents data from regions NE, SW, and CE. The selection of the order in which regions are placed in the case of the nested copula is a result of the values of the correlations between pollutants for the different pairs of regions. Hence, in the case of ozone we have that these values are 0.782, 0.942, and 0.732 if we consider the pairs (CE, NE), (CE, SW), and (NE, SW), respectively. The corresponding values in the case of PM_{10} are 0.676, 0.748, and 0.676, respectively, and in the case of $PM_{2.5}$ they are 0.619, 0.633, and 0.646.

The estimated parameters of interest for both versions of the three-dimensional copulas are given in **Table 1**.

Table 1. Estimated means, standard deviations (indicated by SD), and MC Error of the quantities of interest for all pollutants and copulas considered in the analysis. The symbol – is used to indicate that the specific parameter was not part of the model. In the case of the GH copula we represent θ by θ_1 .

		Mean		SD		MCErr	
		GH	NGH	GH	NGH	GH	NGH
O_3	α_1	4.72	3.764	0.112	0.0723	1.39E–3	1.69E–3
	α_2	3.759	3.644	8.82E–2	0.0696	1.14E–3	1.56E–3
	α_3	3.649	4.739	8.72E–2	0.11	1.112E–3	2.71E–3
	σ_1	131.3	151.3	0.896	1.042	1.13E–2	2.207E–2
	σ_2	152.6	166.1	1.242	1.178	1.58E–2	2.22E–2
	σ_3	167	132.8	1.421	0.904	1.75E–2	1.74E–2
	θ_1	2.485	2.21	8.38E–2	9.03E–2	1.052E–3	2.55E–3
	θ_2	–	4.694	–	0.163	–	3.78E–3

Continued

PM ₁₀	α_1	1.983	2.063	5.07E-2	0.045	5.87E-4	7.89E-4
	α_2	2.051	2.238	5.69E-2	0.0473	6.937E-4	7.897E-4
	α_3	2.218	1.984	5.96E-2	0.0488	7.035E-4	8.53E-4
	σ_1	296.1	172.1	5.493	2.455	7.098E-2	3.84E-3
	σ_2	173	147.4	3.01	1.962	3.71E-2	3.178E-2
	σ_3	147.9	298	2.386	5.339	2.93E-2	0.084
	θ_1	2.029	1.928	6.93E-2	0.0496	9E-4	8.492E-4
	θ_2	–	2.215	–	0.0655	–	1.1E-3
PM _{2.5}	α_1	2.243	2.326	8.29E-2	6.46E-2	1.02E-3	8.63E-4
	α_2	3.61	3.34	0.125	8.9E-2	1.472E-3	1.27E-3
	α_3	3.343	3.623	0.11	0.119	1.31E-3	1.53E-3
	σ_1	103.4	103.7	1.978	1.696	2.37E-2	2.45E-2
	σ_2	83.4	76.29	1.029	0.864	1.16E-2	1.1E-2
	σ_3	76.32	83.47	1.049	1.024	1.24E-2	1.36E-2
	θ_1	1.689	1.636	6.84E-2	2.92E-2	8.25E-4	4.25E-4
	θ_2	–	1.704	–	2.84E-2	–	3.38E-4

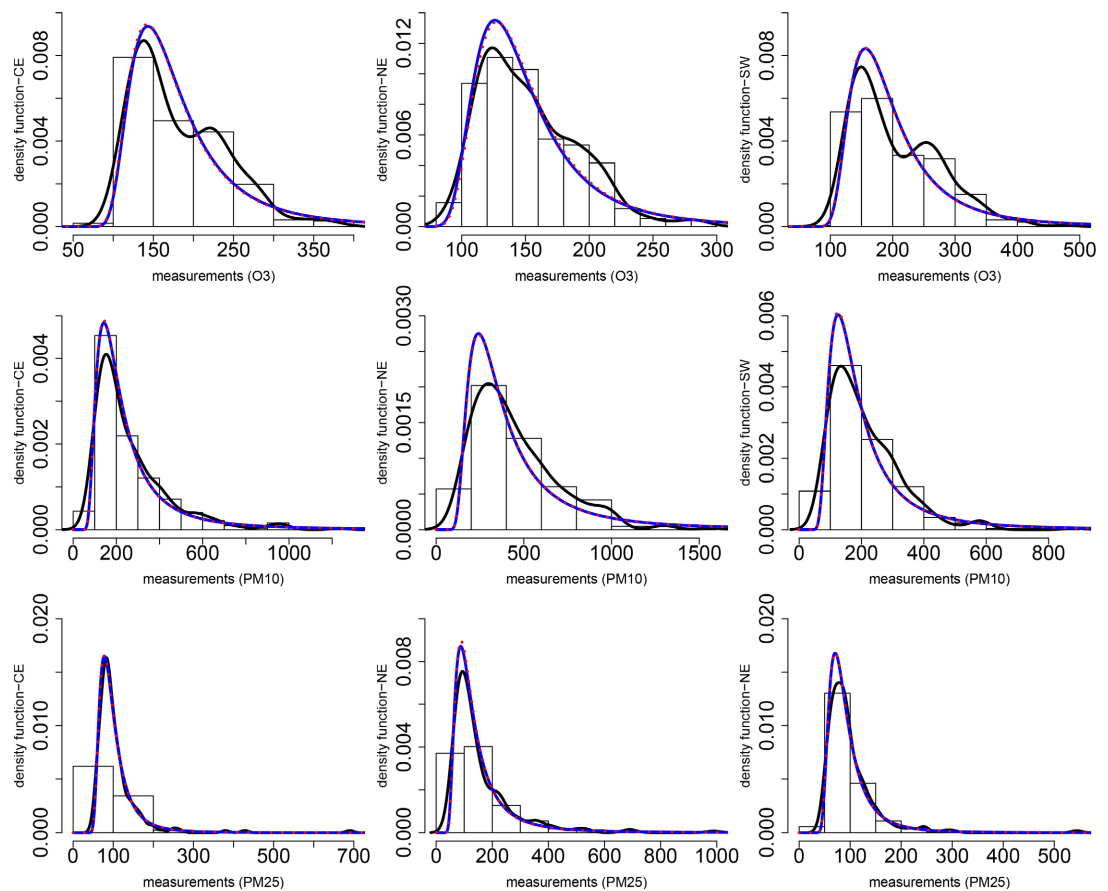


Figure 1. Blocks represent the empirical density, solid black line represents the smoothed empirical density, solid blue line corresponds to the GH fit, and red solid line represents the NGH fit.

Figure 1 shows that the Fréchet distributions estimated using the GH method (solid blue line) and the NGH method (red solid line) are very similar, to the point that they almost overlap. Additionally, the marginal distributions estimated in each case fit the data well, each displaying a single mode, in contrast to the smoothed empirical distribution (solid black line), which exhibits a more irregular shape.

Using the deviance information criterion (DIC) (see, for instance, [37]) to select the models that best explain the data, when the copula (1) is used, the values of the DIC are $3.072\text{E}+08$, $2.592\text{E}+08$, and $1.768\text{E}+08$ in the cases of ozone, PM_{10} , and $\text{PM}_{2.5}$, respectively. In the case of the copula (2) the corresponding values of the DIC are $3.84\text{E}+08$, $3.24\text{E}+08$, and $2.21\text{E}+08$. Hence, in all cases the selected model is the one where the copula (1) is considered since the smallest DIC is obtained when using that copula model. Therefore, we just have one association parameter.

4. Discussion and Comments

In this work, we have used three-dimensional copula models to study the joint behavior of the monthly maximum measurements of the pollutants ozone, PM_{10} , and $\text{PM}_{2.5}$ in three regions in Mexico City. We assume Fréchet distributions function as possible marginal distributions to be considered in the copulas. Two versions of a copula model are considered. In one of them we have the usual three-dimensional Gumbel-Hougaard copula with association parameter θ . In another we have a nested copula where the Gumbel-Hougaard copula is used as the base copula. In this nested version, two association parameters are present. One of them corresponds to the two-dimensional copula nested within the three-dimensional, and another is the outer copula association parameter.

Using the DIC to select the model that best fits the data, the usual three-dimensional Gumbel-Hougaard copula is the selected model for all three pollutants considered in this study. We see, looking at **Table 1**, that the largest association parameter in the selected model is related to the pollutant ozone followed by PM_{10} and $\text{PM}_{2.5}$. That is a reasonable result since ozone travels long distances and its precursors travel in the direction where these three regions are. Hence, it is more likely that the association parameter related to this pollutant is the largest. In the case of PM_{10} we have the presence of sources of this pollutant in the northeast region due to the large number of factories and also in the south region and northwest region due to the large presence of trucks. In the case of $\text{PM}_{2.5}$ there is a large presence of factories in the northeast region, but also a large number of vehicles in regions center and southwest. Hence, perhaps this may be the reason for the smallest association parameter produced.

In addition to the information provided by the values of the association parameters, another information that may be obtained is related to the probability of having a given pollutant in intervals of interest in each of the regions. For instance, if we consider the rule for declaring Phase I of an emergency alert in Mexico City due to ozone [1], then it is necessary to have ozone levels above the threshold

0.154 ppm in one of its regions. In the case of PM_{10} , it is necessary that the corresponding threshold of $214 \mu\text{g}/\text{m}^3$ is exceeded in at least two of its regions. We may also be interested in the case where we have Phase I, but no Phase II [1] declared due to ozone exceedances in these three regions of the main wind corridor. Hence, ozone measurements should be in a given interval of interest in all regions. Similarly, in the cases of PM_{10} and $PM_{2.5}$.

As an example, consider the cases of Phase I in the three regions of study. Assume we are interested in knowing the probability of having the threshold necessary to declare Phase I but not Phase II exceeded by the monthly maximum in all of the three regions. This means that in at least one day in a given month a Phase I emergency alert would be declared due to either ozone or PM_{10} exceedances of the thresholds 0.154 ppm (154 ppb) and $214 \mu\text{g}/\text{m}^3$, respectively. However, the thresholds 0.204 ppm (204 ppb) and $354 \mu\text{g}/\text{m}^3$ for Phase II are not exceeded in any of the three regions. In this case, we would calculate the following probabilities. If the pollutant of interest is ozone, then

$$P\left(\left(O_3^{(NE)}, O_3^{(CE)}, O_3^{(SW)}\right) \in [0.154, 0.204] \times [0.154, 0.204] \times [0.154, 0.204]\right) \\ = 4.36\text{E} - 2$$

and

$$P\left(\left(PM_{10}^{(NE)}, PM_{10}^{(CE)}, PM_{10}^{(SW)}\right) \in [214, 354] \times [214, 354] \times [214, 354]\right) \\ = 1.56\text{E} - 2,$$

if the pollutant of interest is PM_{10} . Hence, for both pollutants the probability of having Phase I but not Phase II declared in all regions of the main wind path is very small.

If we are interested in knowing the probability of having the Mexican ozone environmental threshold of 0.09 ppm (90 ppb) exceeded by the monthly maximum measurements but not the old threshold of 0.11 ppm (110 ppb) in all three regions in the main wind path, we would need to obtain,

$$P\left(\left(O_3^{(NE)}, O_3^{(CE)}, O_3^{(SW)}\right) \in [0.09, 0.11] \times [0.09, 0.11] \times [0.09, 0.11]\right) \\ = 2.83\text{E} - 3,$$

which is also very small.

Consider now the case where we want to know the probability of having Phase I threshold exceeded by the ozone monthly maxima in only one of the regions with no exceedances of the Phase II threshold in a given month. Hence, we would need to obtain,

$$P\left(\left(O_3^{(NE)}, O_3^{(CE)}, O_3^{(SW)}\right) \in [0.154, 0.204] \times [0, 0.154] \times [0, 0.154]\right) \\ + P\left(\left(O_3^{(NE)}, O_3^{(CE)}, O_3^{(SW)}\right) \in [0, 0.154] \times [0.154, 0.204] \times [0, 0.154]\right) \\ + P\left(\left(O_3^{(NE)}, O_3^{(CE)}, O_3^{(SW)}\right) \in [0, 0.154] \times [0, 0.154] \times [0.154, 0.204]\right) \\ \approx 0.006 + 0.042 + 0.132 \approx 0.18,$$

which is not so small.

If we consider the case of PM_{10} monthly maxima exceedances of Phase I threshold occurring in only two of the regions in the main wind corridor, with no exceedances of the Phase II threshold, we need to obtain

$$\begin{aligned} &P\left(\left(PM_{10}^{(NE)}, PM_{10}^{(CE)}, PM_{10}^{(SW)}\right) \in [214, 354] \times [214, 354] \times [0, 214]\right) \\ &+ P\left(\left(PM_{10}^{(NE)}, PM_{10}^{(CE)}, PM_{10}^{(SW)}\right) \in [214, 354] \times [0, 214] \times [214, 354]\right) \\ &+ P\left(\left(PM_{10}^{(NE)}, PM_{10}^{(CE)}, PM_{10}^{(SW)}\right) \in [0, 214] \times [214, 354] \times [214, 354]\right) \\ &\approx 0.0612 + 0.0254 + 0.0009 \approx 0.0875. \end{aligned}$$

Remarks. 1) We would like to call attention to the fact that these probabilities correspond to the probabilities that the monthly maxima fall in the specified intervals and not the probabilities of the daily or hourly measurements belong to them.

2) These probabilities were calculated using the well know relation

$$\begin{aligned} &P([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]) \\ &= F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ &\quad + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) - F(a_1, a_2, a_3), \end{aligned}$$

where the three-dimensional distribution functions are given by the selected copula model (1).

3) Note that the probability of having Phase I declared because the Phase I threshold is exceeded by ozone monthly maxima in one of the regions in the main wind path is not so negligible.

4) One could also consider a five dimensional model to jointly study the monthly maxima in all the five regions of the Mexico City metropolitan area. However, we do not pursue that here. It could be the topic of future analysis.

5) A copula model to jointly study the hourly/maximum daily measurements in the five regions could also be considered. We just need to change the type of distribution used as marginal distributions.

Acknowledgements

The authors thank an anonymous reviewer for the careful reading of the manuscript and for the comments that helped to improve the presentation of the results. This work is part of JAVM Ph.D. Thesis developed at the Benemérita Universidad Autónoma de Puebla, Puebla, México. JAVM thanks CONHACyT-Mexico.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] CDMX (2016) Gaceta oficial de la ciudad de México no. 230. Décima Novena Época. Administración Pública de la Ciudad de México. (In Spanish)

- [2] NADF (2018) Norma ambiental para el distrito federal NADF-009-AIRE-2017. Gaceta Oficial de la Ciudad de México. (In Spanish)
<http://www.aire.cdmx.gob.mx/descargas/monitoreo/normatividad/NADF-009-AIRE-2017.pdf>
- [3] NOM (2014) Norma oficial mexicana NOM-020-SSA1-2014. Diario Oficial de la Federación. (In Spanish)
- [4] NOM (2014) Norma oficial mexicana NOM-025-SSA1-2014. Diario Oficial de la Federación. (In Spanish)
- [5] NOM (2021) Norma oficial mexicana NOM-020-SSA1-2021. Diario Oficial de la Federación. (In Spanish)
- [6] NOM (2021) Norma oficial mexicana NOM-025-SSA1-2021. Diario Oficial de la Federación. (In Spanish)
- [7] Bell, M.L., McDermott, A., Zeger, S.L., Samet, J.M. and Dominici, F. (2004) Ozone and Short-Term Mortality in 95 US Urban Communities, 1987-2000. *JAMA*, **292**, 2372-2378. <https://doi.org/10.1001/jama.292.19.2372>
- [8] Dockery, D.W., Schwartz, J. and Spengler, J.D. (1992) Air Pollution and Daily Mortality: Associations with Particulates and Acid Aerosols. *Environmental Research*, **59**, 362-373. [https://doi.org/10.1016/s0013-9351\(05\)80042-8](https://doi.org/10.1016/s0013-9351(05)80042-8)
- [9] Cifuentes, L., Borja-Aburto, V.H., Gouveia, N., Thurston, G. and Davis, D.L. (2011) Assessing the Health Benefits of Urban Air Pollution Reductions Associated with Climate Change Mitigation (2000-2020): Santiago, São Paulo, México City, and New York City. *Environmental Health Perspectives*, **109**, 419-425. <https://doi.org/10.1289/ehp.01109s3419>
- [10] Loomis, D.P., Borja-Aburto, V.H., Bangdiwala, S.I. and Shy, C.M. (1996) Ozone Exposure and Daily Mortality in Mexico City: A Time Series Analysis. *Health Effects Institute Research Report*, **75**, 1-46.
- [11] Ritz, B. and Yu, F. (1999) The Effect of Ambient Carbon Monoxide on Low Birth Weight among Children Born in Southern California between 1989 and 1993. *Environmental Health Perspectives*, **107**, 17-25. <https://doi.org/10.1289/ehp.9910717>
- [12] Ritz, B., Yu, F., Chapa, G. and Fruin, S. (2000) Effect of Air Pollution on Preterm Birth among Children Born in Southern California between 1989 and 1993. *Epidemiology*, **11**, 502-511. <https://doi.org/10.1097/00001648-200009000-00004>
- [13] Ritz, B., Wilhelm, M., Hoggatt, K.J. and Ghosh, J.K.C. (2007) Ambient Air Pollution and Preterm Birth in the Environment and Pregnancy Outcomes Study at the University of California, Los Angeles. *American Journal of Epidemiology*, **166**, 1045-1052. <https://doi.org/10.1093/aje/kwm181>
- [14] Itô, K. and Thurston, G.D. (1996) Daily PM10/Mortality Associations: An Investigation of At-Risk Subpopulations. *Journal of Exposure Analysis and Environmental Epidemiology*, **6**, 79-95.
- [15] Mauderly, J.L. (1997) Relevance of Particle-Induced Rat Lung Tumors for Assessing Lung Carcinogenic Hazard and Human Lung Cancer Risk. *Environmental Health Perspectives*, **105**, 1337-1346. <https://doi.org/10.1289/ehp.97105s51337>
- [16] Mauderly, J. and Oberdörster, G. (1997) Current Understanding of the Health Effects of Particles and Characteristics that Determined Dose Effects. In: *Formation and Characterization of Particles*, Report of the 1996 Health Effects Institute Workshop, 1-5.
- [17] Thurston, G.D. (1996) A Critical Review of PM10 Mortality Time Series Studies. *Journal of Exposure Analysis and Environmental Epidemiology*, **6**, 3-21.

- [18] Rodríguez, S., Huerta, G. and Reyes, H. (2016) A Study of Trends for Mexico City Ozone Extremes: 2001-2014. *Atmósfera*, **29**, 107-120.
<https://doi.org/10.20937/atm.2016.29.02.01>
- [19] Tabari, H. (2021) Extreme Value Analysis Dilemma for Climate Change Impact Assessment on Global Flood and Extreme Precipitation. *Journal of Hydrology*, **593**, Article 125932. <https://doi.org/10.1016/j.jhydrol.2020.125932>
- [20] Surman, P.G., Boderó, J. and Simpson, R.W. (1987) The Prediction of the Numbers of Violations of Standards and the Frequency of Air Pollution Episodes Using Extreme Value Theory. *Atmospheric Environment* (1967), **21**, 1843-1848.
[https://doi.org/10.1016/0004-6981\(87\)90125-9](https://doi.org/10.1016/0004-6981(87)90125-9)
- [21] Ercelebi, S.G. and Toros, H. (2009) Extreme Value Analysis of Istanbul Air Pollution Data. *CLEAN—Soil, Air, Water*, **37**, 122-131. <https://doi.org/10.1002/clen.200800041>
- [22] Fang, J., Wahl, T., Zhang, Q., Muis, S., Hu, P., Fang, J., *et al.* (2021) Extreme Sea Levels Along Coastal China: Uncertainties and Implications. *Stochastic Environmental Research and Risk Assessment*, **35**, 405-418.
<https://doi.org/10.1007/s00477-020-01964-0>
- [23] Vazquez-Morales, J.A., Rodrigues, E.R. and Reyes-Cervantes, H.J. (2024) Bivariate Analysis of Pollutants Monthly Maxima in Mexico City Using Extreme Value Distributions and Copula. *Journal of Environmental Protection*, **15**, 796-826.
<https://doi.org/10.4236/jep.2024.157046>
- [24] Grimaldi, S. and Serinaldi, F. (2006) Asymmetric Copula in Multivariate Flood Frequency Analysis. *Advances in Water Resources*, **29**, 1155-1167.
<https://doi.org/10.1016/j.advwatres.2005.09.005>
- [25] Padhee, R.B., Kar, A.K. and Das, P.K. (2021) Gumbel-Hougaard Copula-Based Tetra-variate Flood Frequency Analysis for the Hirakud Reservoir Catchment. *ISH Journal of Hydraulic Engineering*, **28**, 430-437.
<https://doi.org/10.1080/09715010.2021.1974317>
- [26] Zhang, L. and Singh, V.P. (2007) Trivariate Flood Frequency Analysis Using the Gumbel-Hougaard Copula. *Journal of Hydrologic Engineering*, **12**, 431-439.
[https://doi.org/10.1061/\(asce\)1084-0699\(2007\)12:4\(431\)](https://doi.org/10.1061/(asce)1084-0699(2007)12:4(431))
- [27] Zhang, L., and Singh, V.P. (2007) Gumbel-Hougaard Copula for Trivariate Rainfall Frequency Analysis. *Journal of Hydrologic Engineering*, **12**, 409-419.
- [28] Salvadori, G. and De Michele, C. (2010) Multivariate Multiparameter Extreme Value Models and Return Periods: A Copula Approach. *Water Resources Research*, **46**, W10501. <https://doi.org/10.1029/2009wr009040>
- [29] Masseran, N. and Hussain, S.I. (2020) Copula Modelling on the Dynamic Dependence Structure of Multiple Air Pollutant Variables. *Mathematics*, **8**, Article 1910.
<https://doi.org/10.3390/math8111910>
- [30] Gumbel, E.J. (1960) Bivariate Exponential Distributions. *Journal of the American Statistical Association*, **55**, 698-707. <https://doi.org/10.1080/01621459.1960.10483368>
- [31] Hougaard, P. (1986) Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika*, **73**, 387-396.
<https://doi.org/10.1093/biomet/73.2.387>
- [32] Hougaard, P. (1986) A Class of Multivariate Failure Time Distributions. *Biometrika*, **73**, 671-678. <https://doi.org/10.1093/biomet/73.3.671>
- [33] Genest, C. and Favre, A. (2007) Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, **12**, 347-368. [https://doi.org/10.1061/\(asce\)1084-0699\(2007\)12:4\(347\)](https://doi.org/10.1061/(asce)1084-0699(2007)12:4(347))

- [34] Tsokos, C.P. (2011) Generalized Extreme Value Family of Probability Distributions. In: *International Encyclopedia of Statistical Science*, Springer Berlin Heidelberg, 585-589. https://doi.org/10.1007/978-3-642-04898-2_427
- [35] Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS Project: Evolution, Critique and Future Directions. *Statistics in Medicine*, **28**, 3049-3067. <https://doi.org/10.1002/sim.3680>
- [36] Spiegelhalter, D.J., Thomas, A. and Best, N.G. (1999) WinBUGS: Bayesian Inference Using Gibbs Sampling. MRC Biostatistics Unit.
- [37] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **64**, 583-639. <https://doi.org/10.1111/1467-9868.00353>

Appendices

Appendix A. The Particular Forms of the Likelihood Functions

In this appendix, we present the expressions for the likelihood functions when the marginal distributions used in the copula are Fréchet (α , σ) distributions. We present first the case where the three-dimensional distribution is constructed using the usual Gumbel-Hougaard copula (1) and then we present the expression in the case of the nested model (2).

Usual three-dimensional Gumbel-Hougaard copula

When we use the usual three-dimensional Gumbel-Hougaard copula given by (1) and Fréchet distributions as marginal distributions, the joint density function used in the likelihood function is

$$f_{ijk}(x, y, z) = \left(\frac{\alpha_i \alpha_j \alpha_k}{xyz} \right) \left(\frac{x}{\sigma_i} \right)^{2\alpha_i \theta} \left(\frac{y}{\sigma_j} \right)^{2\alpha_j \theta} \left(\frac{z}{\sigma_k} \right)^{2\alpha_k \theta} \exp \left(-\kappa_1^{\frac{1}{\theta}} \right) \left(\left(\frac{x}{\sigma_i} \right)^{\alpha_i \theta} \left(\frac{y}{\sigma_j} \right)^{\alpha_j \theta} + \left(\frac{x}{\sigma_i} \right)^{\alpha_i \theta} \left(\frac{z}{\sigma_k} \right)^{\alpha_k \theta} + \left(\frac{y}{\sigma_j} \right)^{\alpha_j \theta} \left(\frac{z}{\sigma_k} \right)^{\alpha_k \theta} \right)^{-3} \kappa_1^{\frac{1}{\theta}} \left(1 + 2\theta^2 - 3\kappa_1^{\frac{1}{\theta}} + \kappa_1^{\frac{2}{\theta}} + 3\theta \left(-1 + \kappa_1^{\frac{1}{\theta}} \right) \right)$$

where

$$\kappa_1 = \left(\frac{x}{\sigma_i} \right)^{-\alpha_i \theta} + \left(\frac{y}{\sigma_j} \right)^{-\alpha_j \theta} + \left(\frac{z}{\sigma_k} \right)^{-\alpha_k \theta},$$

Gumbel-Hougaard nested copula

When we use the nested three-dimensional Gumbel-Hougaard copula given by (2) and Fréchet distributions as the marginal distributions, the joint density function used in the likelihood function is

$$f_{ijk}(x, y, z) = \kappa_2^{\frac{1}{\theta_1}} \left(\frac{\alpha_i \alpha_j \alpha_k}{xyz} \right) \left(\frac{x}{\sigma_i} \right)^{\alpha_i \theta_2} \left(\frac{y}{\sigma_j} \right)^{\alpha_j \theta_2} \left(\frac{z}{\sigma_k} \right)^{\alpha_3 \theta_1} \left(\left(\frac{x}{\sigma_i} \right)^{\alpha_i \theta_2} + \left(\frac{y}{\sigma_j} \right)^{\alpha_j \theta_2} \right)^{-2} \left(1 + \kappa_3^{\frac{\theta_1}{\theta_2}} \left(\frac{z}{\sigma_k} \right)^{\alpha_3 \theta_1} \right)^{-3} \exp \left(-\kappa_2^{\frac{1}{\theta_1}} \right) \left\{ \theta_1^2 \left(-1 + \kappa_3^{\frac{\theta_1}{\theta_2}} \left(\frac{z}{\sigma_j} \right)^{\alpha_j \theta_1} \right) + \theta_2 \left(1 + \kappa_3^{\frac{\theta_1}{\theta_2}} \left(\frac{z}{\sigma_j} \right)^{\alpha_j \theta_1} \right) \left(-1 + \kappa_2^{\frac{1}{\theta_1}} \right) + \kappa_3^{\frac{\theta_1}{\theta_2}} \left(\frac{z}{\sigma_j} \right)^{\alpha_j \theta_1} \left[1 - 3\kappa_2^{\frac{1}{\theta_1}} + \kappa_2^{\frac{2}{\theta_1}} \right] + \theta_1 \left[\theta_2 + \theta_2 \kappa_3^{\frac{\theta_1}{\theta_2}} \left(\frac{z}{\sigma_j} \right)^{\alpha_j \theta_1} + \left(-1 + 2\kappa_3^{\frac{\theta_1}{\theta_2}} \left(\frac{z}{\sigma_j} \right)^{\alpha_j \theta_1} \right) \left(-1 + \kappa_2^{\frac{1}{\theta_1}} \right) \right] \right\},$$

with κ_2 and κ_3 given as in the main text with the appropriate substitution.

Appendix B. Computational Details

In this appendix we present the prior distributions used in each version of the model, as well as information related to the burn-in periods, sampling gaps, and sample sizes. We start by describing the information related to the three-dimensional Gumbel-Hougaard copula and move to the nested three-dimensional copula.

When we consider the Gumbel-Hougaard copula in three dimension and the Fréchet distributions as marginal distributions, the parameters α_i will have uniform prior distribution $U(0, 10)$ for all pollutants, $i = 1, 2, 3$. The parameters σ_i will have $U(0, 200)$ prior distribution, $i = 1, 2, 3$, in the case of ozone, and when we consider PM_{10} and σ_2 and σ_3 . In the case of $PM_{2.5}$ the parameters σ_2 and σ_3 will have $U(0, 100)$ prior distribution. When we consider the parameter σ_1 and the pollutants PM_{10} and $PM_{2.5}$, it will have $U(0, 350)$ and $U(0, 150)$ prior distributions, respectively. The parameter θ will have as prior distribution a $U(1, 15)$ in all cases. The burn-in period was of 20,000 steps and samples of size 8000 collected every 50th generated values were used to estimate the parameters in all cases.

Consider now the nested copula where the copulas involved in the nesting still are Gumbel-Hougaard with Fréchet distributions as marginal distributions. When this copula is considered, the parameters α_i will have $U(0, 10)$ prior distribution in all cases, $i = 1, 2, 3$. If we consider the parameters σ_i , $i = 1, 2, 3$, and ozone and σ_1 and σ_2 and PM_{10} , the prior distribution is $U(0, 200)$, and in the case of $PM_{2.5}$, σ_2 and σ_3 will have a $U(0, 100)$ prior distribution. If we take into account the parameter σ_3 and the pollutant PM_{10} and σ_2 and the pollutant $PM_{2.5}$, then the prior distributions are $U(0, 350)$ and $U(0, 150)$, respectively. In the cases of the parameters θ_1 and θ_2 we have the following. If we consider θ_1 , then it has prior distributions $U(1, 2.5)$, $U(1, 2)$, and $U(1, 1.67)$ in the cases of ozone, PM_{10} , and $PM_{2.5}$, respectively. The prior distributions in the case of θ_2 are $U(2.6, 15)$, $U(2, 15)$, and $U(1.671, 15)$ if the pollutants are ozone, PM_{10} , and $PM_{2.5}$, respectively. The burn-in period was of 20,000 steps in the cases of ozone and PM_{10} and was 30,000 steps in the case of $PM_{2.5}$. Samples of sizes 8000, 7000, and 12,000 collected every 50th generated values were used to estimate the parameters when the pollutants considered were O_3 , PM_{10} , and $PM_{2.5}$.