Scientific
Research
Publishing

Extreme Values Approach in Food Risk Modeling

Komla Elom Adedje, Diakarya Barro*

Université Thomas Sankara, Ouagadougou, Burkina Faso
Email: keadedje@gmail.com, *dbarro2@gmail.com

Abstract

In this paper we estimate the incubation period of a possible pathology following exposure to dioxins during a poor diet. The tools developed for this purpose include the probabilistic extremal model and the stochastic behavior of the distribution tails of the contamination. We propose a cumulative distribution function for a random variable that follows both a Gaussian distribution and a GPD. A global optimization method is also explored for the efficient estimation of parameters of GPD.

Keywords

Extreme Values, Tail, Dioxin, Dietary Risk, Incubation Period

1. Introduction

It happens that food can have the unexpected opposite effect on the body when it is full of dangerous toxic substances. Cardiovascular diseases, diabetes, immune system degradation, cancer, etc. are among the consequences of consuming these foods. These events are qualified to be "extreme" insofar as the primary objective of the food is to maintain the health of the consumer. A good diet is, therefore, an essential element in the management of our health. It must be noted that most products or foods consumed today present a danger or risk of contamination, whether related to their composition or their conservation. This poses a public health problem. Faced with this observation, it noticed that in this work we set up a risk indicator that allows decision-makers to take reliable risk management measures for food and human safety. Increasingly, researchers are taking a particular interest in the modeling of extreme events in many fields, such that hydrology for flood analysis, meteorology, physics, finance, etc. However, we have developed a joint probability density of a random variable follow-

ing both the generalized Pareto distribution (GPD) and the normal distribution.

In medicine, decision-makers are throwing a lot of money into clinical trials to determine the delay between contamination and the onset of symptoms of a disease in order to circumscribe the pandemic and take appropriate measures. The new estimate of the average incubation period of the SARS-CoV-2 coronavirus, *i.e.* the time between infection with the virus and the onset of the first symptoms of COVID-19, is based on data from patients who contracted the virus in China or through contact with an infected person. This update of clinical data, proposed in the Annals of Internal Medicine [1] and reported in the New England Journal of Medicine Journal Watch, confirms that most patients who become symptomatic develop the first symptoms within 11 or 12 days and the vast majority within 14 days of infection. This allows health authorities to quarantine infected people to break the chain of contamination. Hence this indicator in health management and safety is needed. Also, a team of researchers from Cornell University (New York) was able to show in the journal eLife that the incubation periods of infected patients are normally distributed. Their study contributed to the understanding of this variability in that many diseases take much longer to develop in some affected individuals than in others using the formalism of evolutionary graph theory.

The paper is organized as follows. Section 2 reminds the materials and methods used for the study within the framework of the extreme values theory and the concept of Optimization Preserving Operator (OPO). Section 3 presents the main results of the study. It provides the estimation of the extreme quantile, the incubation period, and a joint probability density function. Specifically the GPD parameters are estimated using the maximum likelihood method (MLM) and the numerical Alienor method (AM).

2. Materials and Methods

In this section, we remind an overview of extreme value tools, threshold determination, extremal index, and then a survey of the Alienor approach.

2.1. An Overview of Extreme Value Tools

The Peaks-Over-Threshold (POT) approach has the advantage to take into account all the observations that exceed a given threshold u large enough in the sample and not only the maximums per block (BM), see [2] for more details.

Let $\{X_i, i = 1, 2, \cdots, n\}$ be a sequence of random variables independent and identically distributed (i.i.d) of common law F and density f. According to Gnedenko [3], for $X_{n,n} = \max_{1 \le i \le n}(X_i)$ if there exist two real normalizing sequences $(a_n)_{n \ge 1} > 0$ and $(b_n)_{n \ge 1} \in \mathbb{R}$ and a non-degenerated law $\mathbb{H}_\gamma$, such that:

$$\lim_{n \to \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) = \lim_{n \to \infty} F^n(a_n x + b_n) = \mathbb{H}_\gamma(x); \quad (1)$$

then, $\mathbb{H}$ is one of the three limits which are the Fréchet type laws, the Weibull type laws, and the Gumbel type laws, for more see in [4]. A parametrization of

these three distributions into a single formula called Generalized Extreme Values (GEV) distribution is given by

$$\mathbb{H}_\gamma(x) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\} & \text{si} \quad \gamma \neq 0 \\[2em] \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right\} & \text{si} \quad \gamma = 0. \end{cases} \tag{2}$$

where $\{\mu \in \mathbb{R}\}, \{\sigma > 0\}$ and $\{\gamma \in \mathbb{R}\}$ are respectively parameters of location, scale, and shape. Notice that analytically the three extreme distributions are different from a modeling point of view, more details in [5]. In this paper, the POT approach is considered for the adequate estimation of quantiles in the sense that it is easier to obtain an excess sample. If F belongs to the catchment area of the distribution of extreme values (Weibull, Fréchet, Gumbel), then there is a function $\sigma(u)$ strictly positive such as:

$$\lim_{u \nearrow x^F} \sup_{0 < Y < x^F - u} \left| F_u(y) - \mathbb{H}_{\mu,\sigma(u),\gamma}(y) \right| = 0; \tag{3}$$

where x^F and $\mathbb{H}_{\mu,\sigma(u),\gamma}(x)$ are the right terminal point and the GPD function, respectively. In particular, in toxicology, the threshold of toxicity noted $e(u)$ is derived from medical studies beyond which there may be danger to health. In the framework of our study, the value proposed by FAO/WHO is used for the efficient result. Generally, it is determined graphically by the Mean Excess Plot function under the R software. We call the mean function of the excesses of the random variable X over the threshold $u < x^F$ and note $e(u)$, the expectation function of the random variable Y excess of X over the threshold u defined by:

$$e(u) = E(X - u \mid X > u). \tag{4}$$

The estimation of the GEV parameters, more precisely the estimation of the tail index γ plays a key role in the behavior of the law of extremes. There are two possible approaches for estimating γ. The BM approach uses the asymptotic GEV law and the POT one pointed out by the GPD models. Making the assumption of overestimation of the food risk, we are interested in the GPD approach. We use the maximum likelihood method to estimate the parameter γ of the GPD of excesses, see more (3.4). Most estimators of the tail parameters by GPD approach rely on the use of the order statistic, the i-th value of a sample of size n. The Hill estimator is the most used because it has a smaller asymptotic variance than the others. It is defined by the following statistic:

$$\hat{\gamma}_{k,n}^H = \frac{1}{k-1} \sum_{j=1}^{k-1} \ln\left(\frac{X_{j,n}}{X_{k,n}}\right); \tag{5}$$

and it converges in law to the true value of the parameter for k large enough. A second estimator, the Pickands one is the simple and oldest of the tail estimators. It is defined by the statistic:

$$\hat{\gamma}_{k,n}^{P} = \frac{1}{\ln 2} \ln\left(\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}\right); \tag{6}$$

where $\hat{\gamma}_{k,n}^{P}$ is the estimation of the tail index which plays an essential role in the behaviour of the law of extremes (refer to [6] for more details).

2.2. A Survey of the Alienor's Algorithm

The algorithm of the Alienor approach is as follows: the basic idea is to transform n-variable likelihood function into a single variable function θ, and the OPO, then, be used to obtain the global optimum (see [7] for more details). Let us consider n independent random variables $X_1, X_2, \cdots, X_n$ that are assumed to be identically distributed according to a given distribution $F(x,\theta)$, where θ is the vector of parameters, with density function $f(x,\theta)$. Let $x_1, x_2, \cdots, x_n$ be the realizations of these random variables which form the observed sample.

Step 1: definition of the objective function.

Let L be the likelihood function defined by

$$L : \mathbb{R}^{n+3} \to [0,1]$$
$$(x_1, x_2, \cdots, x_n, u, \gamma, \sigma) \mapsto \prod_i f(x_i, u, \gamma, \sigma)$$

we look for the values of u, γ and σ that maximize L. Given x_i, we consider the likelihood L as a function of three variables u, γ and σ. By applying the following reductive transformation on the likelihood function L;

$$\begin{cases} h_1(\theta) = u(\theta) = (\min x_i - a) \times \theta + a; \\ h_2(\theta) = \gamma(\theta) = \frac{1}{2}(1 - \cos(m \times \pi\theta)); \\ h_3(\theta) = \sigma(\theta) = \frac{1}{2}(1 - \cos(m^2 \times \pi\theta)); m = 2;3, \end{cases} \tag{7}$$

where $u \in [a, \min x_i]$, $\gamma, \sigma \in [0, +\infty)$ and $a = \frac{1}{m}\sqrt{2}$, x denotes observations,

The problem becomes

$$Glob. \max_{u, \gamma, \sigma} L^*(u, \gamma, \sigma).$$

$$u \in [0, \min x_i], \gamma, \sigma \in [0, +\infty);$$

where the notation Glob. max means the maximum global. The problem of estimating parameters of the GPD distribution is considered as a global optimization problem.

We have a maximization problem that can be transformed into a minimization problem as follows

$$\max_{\theta \in [0,1]} L^*(\theta) = -\min_{\theta \in [0,1]}\left(-L^*(\theta)\right),$$

with

$$-L^* : [0,1] \to [-1,0]$$

$$\theta \mapsto -L^*(\theta).$$

The sequence of steps in the algorithm is as follows:

Initialization

$\rightharpoonup$ one randomly chooses θ_0 in $[0, \theta_{\max}]$ then,

$\rightharpoonup$ We build the function

$$T_{L^*}(\theta) = \frac{1}{2}\left[L^*(\theta) - L^*(\theta_0) + \left| L^*(\theta) - L^*(\theta_0) \right| \right]. \tag{8}$$

Step 2: Solving algebraic equation

$\rightharpoonup$ one solves $T_{L^*}(\theta) = 0$ then,

$\rightharpoonup$ The solution set is built

$$S_L = \left\{ \theta \in I, T_{L^*}(\theta) = 0 \right\} \tag{9}$$

If $S_L = \left\{ \theta^* \right\}$, then

θ^* is the global minimizer of L;

Otherwise go to the next step.

Step 3: modification of step i

$\rightharpoonup$ one chooses θ_i in $S_L = \left\{ \theta \in I, T_{L^*}(\theta) = 0 \text{ and } L(\theta_i) < L(\theta_{i-1}) \right\}$;

Then

$\rightharpoonup$ the function $T_{L^*}(\theta) = \frac{1}{2}\left[L^*(\theta) - L^*(\theta_i) + \left| L^*(\theta) - L^*(\theta_i) \right| \right]$,

Then,

$\rightharpoonup$ one starts again at step 1;

The following section deals with our results.

3. Main Results

In estimating the parameters of the GPD and GEV distribution, the most popular and conditionally most efficient method is the maximum likelihood method. However, this method is biased towards small sample sizes. To overcome this drawback we propose an approach based on a global optimization method so called Alienor method.

3.1. Estimation of Extreme Quantile and Incubation Period

In this subsection, the estimation of extreme quantiles and the incubation period for chronic dioxin exposure are discussed. Let $X_1, X_2, \cdots, X_n$ be a sample of i.i.d random variables with distribution function F and x_F a terminal point. Then, for a fixed threshold $u < x_F$, consider the n observations $x_{i,1}, \cdots, x_{i,n}$ exceeding the threshold u large enough. Let $Y_j = X_{ij} - u$, $j = 1, \cdots, n$, be the excess variables above the threshold u. Recall that the law of excesses beyond a sufficiently high u is:

$$F_u(y) = P(X - u \le y / X > u) = \frac{F(u+y) - F(u)}{\overline{F}(u)}; \tag{10}$$

where $\overline{F}$ is the survival function of F given by $\overline{F}(u) = 1 - F(u)$.

Proposition 1. Let F be the common distribution of a sequence of i.i.d ran-

dom $(X_i)_{i \in \mathbb{N}}$. Then, for all $y, u > 0$

$$\overline{F}(u+y) = \overline{F}(u)\overline{F}_u(y); \tag{11}$$

where $\overline{F}$ is the GEV distribution of $(X_i)_{i \in \mathbb{N}}$

Proof. It is easy to see that

$$
\begin{aligned}
F_u(y)\overline{F}(u) &= F(u+y) - F(u) \\
&= -1 + F(u+y) + 1 - F(u) \\
&= -[1 - F(u+y)] + [1 - F(u)] \\
&= -\overline{F}(u+y) + \overline{F}(u).
\end{aligned}
$$

Furthermore, it follows that

$$\overline{F}(u+y) = \overline{F}(u) - F_u(y)\overline{F}(u) = \overline{F}(u)[1 - F_u(y)] = \overline{F}(u)\overline{F}_u(y). \tag{12}$$

Application 1. *By using* (3) *it comes that*

$$F_u(y) \approx \mathbb{H}_{\mu, \sigma(u), \gamma}(y) \quad \text{and} \quad \overline{F}_u(y) \approx \overline{\mathbb{H}}_{\mu, \sigma(u), \gamma}(y). \tag{13}$$

For u large enough and $y \geq 0$. Therefore, an empirical estimator of $\overline{F}(u)$ is

$$\hat{\overline{F}}(u) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i > u\}} = \frac{N_u}{n}; \tag{14}$$

where $I_{\{X_i > u\}}$ is an indicator function defined as

$$I_{\{X_i > u\}} = \begin{cases} 1 & \text{if } X_i > u \text{ for all } i = 1, \cdots, n \\ 0 & \text{else} \end{cases} \tag{15}$$

and N_u is the number of excesses above the threshold u. Notice that the law of F_u excesses can be uniformly approximated by a generalized Pareto distribution defined as follows,

$$\forall x > 0 \text{ and } 1 + \gamma\frac{x}{\sigma} > 0, F_u(x) = 1 - \left(1 + \gamma\frac{x}{\sigma}\right)^{-\frac{1}{\gamma}} \text{ for all } \gamma \neq 0. \tag{16}$$

By using (11) and by posing $\overline{F}(u+y) = \overline{F}(x)$, and considering that $\overline{F}_u = 1 - F_u$, the form of the tail estimator is

$$\hat{\overline{F}}(x) = \frac{N_u}{n}\left(1 + \gamma\frac{x}{\sigma}\right)^{-\frac{1}{\gamma}}. \tag{17}$$

Since the upper part of the tail is estimated of $\overline{F}$, in addition the restriction of the domain of Fréchet and by inversion, the extreme quantile estimator x_{pt} is obtained by the following Formula (see [8] for a new approach)

$$\hat{x}_{pt}^{GPD} = u + \frac{\hat{\sigma}}{\hat{\gamma}}\left[\left(\frac{n}{N_u}(1-pt)\right)^{-\hat{\gamma}} - 1\right], \text{ for all } \hat{\gamma} \neq 0. \tag{18}$$

$1 - pt$ is equivalent to a confidence level of typically 95% or 99%, n is the sample size; N_u is the number of observations above the threshold u; $\hat{\gamma}$ and $\hat{\sigma}$ are the estimators of the parameters of the GPD distribution. We, therefore, estimate the incubation period for the extreme value distribution GPD by:

$$\hat{T}^{GPD} = \frac{n}{N_d} \frac{1}{1 - \mathbb{H}_{\hat{\gamma},\hat{\sigma}}(x_{pt})}, \quad \text{for all } \hat{\gamma} \in \mathbb{R}, \hat{\sigma} > 0. \tag{19}$$

3.2. Estimation of a Joint Probability Density Function

The calibration of the female contamination data shows that the distribution of the observations simultaneously follows the Laplace-Gauss law and the $\mathbb{GPD}$ law. It is therefore necessary to reflect on the search for a joint law allowing a better appreciation of the phenomena. A joint probability density function based on the composite density function is a new way in modeling of the dependence of a random vector.

Proposition 2. Suppose that X and Y be two random variables from $\mathbb{GPD}(\gamma, \sigma_1)$ and $\mathbb{N}(m, \sigma_2)$, respectively. Under some convergence conditions, the joint random variable U converges almost surely to the law whose probability density function is defined by:

$$g(u) = \frac{1}{\sqrt{u}} e^{-\pi u} \quad \text{for all } u > 0. \tag{20}$$

where $u = \left(\frac{y - m}{\sigma_2}\right)^2$, m the mean and σ_2 the standard deviation of the random variables Y.

Proof. Let X be a GPD random variable whose bi-parametric density [9] is given by

$$f_X(x) = \frac{1}{\sigma_1} e^{\left(-\frac{1}{\gamma} - 1\right) \ln\left(1 + \frac{\gamma}{\sigma_1} x\right)} \quad \text{for all } x \geq 0. \tag{21}$$

According to the relation (21), we assume that

$$f_X \circ f_Y(y) = f_X \left[f_Y(y) \right];$$

where $f_Y(y) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y - m}{\sigma_2}\right)^2}$ is probability density function of the normal distribution. It follows that, for all $y \in \mathbb{R}$

$$f_X \circ f_Y(y) = \frac{1}{\sigma_1} e^{\left(-\frac{1}{\gamma} - 1\right) \ln\left(1 + \frac{\gamma}{\sigma_1} \left(\frac{1}{\sigma_1 \sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y - m}{\sigma_2}\right)^2}\right)\right)}.$$

By setting $z = \frac{y - m}{\sigma_2}$, it comes that

$$f_Z(z) = \frac{1}{\sigma_1 \left(1 + \frac{\gamma}{\sigma_1 \sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2}z^2}\right)^{1 + \frac{1}{\gamma}}}.$$

Let's set now $\gamma = \sigma_1 \sigma_2 \sqrt{2\pi}$, it is easy to find that,

$$1 + e^{-\frac{1}{2}z^2} = 2 e^{-\frac{1}{4}z^2} \cosh\left(\frac{1}{4}z^2\right).$$

Finally, on obtains, for all $z \in \mathbb{R}$

$$f_Z(z) = \left[\overbrace{\frac{1}{2\sigma_1 \cosh\left(\frac{1}{4}z^2\right)}} \, e^{-\frac{1}{4}z^2} \right]^{1+\frac{1}{\gamma}} \overset{p.s}{\rightsquigarrow} e^{-\frac{1}{4}\left(1+\frac{1}{\gamma}\right)z^2}.$$

Then, notice that for all $z \in \mathbb{R}$,

$$\cosh\left(\frac{1}{4}z^2\right) \geq 1 \Leftrightarrow \frac{1}{2\sigma_1 \cosh\left(\frac{1}{4}z^2\right)} \leq \frac{1}{2\sigma_1} < 1.$$

By setting $\xi = \frac{1}{2}\left(1+\frac{1}{\gamma}\right)$, one has $f(z) = e^{-\frac{1}{2}\xi z^2}$. Let's now, show for which

value of ξ, $f(z)$ is a probability density function.

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\xi z^2} \, dz = 2\int_0^{+\infty} e^{-\frac{1}{2}\xi z^2} \, dz. \tag{22}$$

Let's perform the following change of variable, by posing

$$t = \frac{1}{2}\xi z^2 \Rightarrow z = \sqrt{\frac{2}{\xi}t} \Rightarrow dz = \frac{dt}{\sqrt{2\xi t}}.$$

The relation (22) becomes

$$2\int_0^{+\infty} e^{-t} \frac{dt}{\sqrt{2\xi t}} = \frac{2}{\sqrt{2\xi}} \int_0^{+\infty} t^{-\frac{1}{2}} e^{-t} dt = \frac{2}{\sqrt{2\xi}} \Gamma\left(\frac{1}{2}\right) = \frac{2}{\sqrt{2\xi}} \sqrt{\pi};$$

Γ is gamma function. So, f is a probability density function when

$$\xi = 2\pi \quad \text{et} \quad f_Z(z) = e^{-\pi z^2}.$$

The proof is completed by substituting $u = z^2$, the probability density function of u is obtained by the following

If $u > 0$,

$$F(u) = \mathbb{P}(U \leq u) = \mathbb{P}(Z^2 \leq u).$$

F is the cumulative distribution function. Which yields

$$F(u) = \mathbb{P}\left(-\sqrt{u} \leq Z \leq \sqrt{u}\right) = \int_{-\sqrt{u}}^{\sqrt{u}} e^{-\pi z^2} \, dz.$$

Thus,

$$F(u) = 2\int_0^{\sqrt{u}} e^{-\pi z^2} \, dz \quad \text{and} \quad F(u) = 0 \quad \text{if} \quad u < 0.$$

The derivative of the cumulative distribution function is given by the probability density function as follows

$$g(u) = F'(u) = \frac{\partial F(u)}{\partial u}.$$

Recalling the following result

$$\frac{\partial}{\partial u}\left(\int_0^{w(u)} g(x) \, dx\right) = w'(u) g\left[w(u)\right].$$

This leads to the desired result.

The previous proposition is interesting in practice because it allows to obtain the joint probability density resulting from a convex combination of $\mathbb{GPD}$ and Laplace-Gauss law. This gives us the following statement.

Proposition 3. Let X be a random variable with a joint probability density from GPD and a Gaussian distribution. Then, the expectation and the variance are given respectively

$$\mathbb{E}(X) = \frac{1}{2\pi} \quad \text{and} \quad \mathbb{V}(X) = \frac{1}{2\pi^2}. \tag{23}$$

Proof. According to the above proposition 2, one has,

$$g(x) = \begin{cases} \dfrac{1}{\sqrt{x}} e^{-\pi x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases} \tag{24}$$

$$E(x) = \int_{-\infty}^{+\infty} x g(x) \, dx = \int_0^{+\infty} x \times \frac{1}{\sqrt{x}} e^{-\pi x} \, dx$$

By setting $u = \pi x$, which leads to the following result

$$E(x) = \frac{\sqrt{\pi}}{\pi^2} \int_0^{+\infty} u \times \frac{1}{\sqrt{u}} e^{-u} \, du = \frac{\sqrt{\pi}}{\pi^2} \int_0^{+\infty} u^{\frac{3}{2}-1} e^{-u} \, du$$

We finally obtain

$$E(x) = \frac{\sqrt{\pi}}{\pi^2} \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{\pi^2} \times \frac{1}{2} \times \sqrt{\pi} = \frac{1}{2\pi}$$

where Γ is the gamma function. Knowing that

$$V(X) = E(X^2) - (E(X))^2, \tag{25}$$

one has

$$E(X^2) = \int_0^{+\infty} x^2 \times \frac{1}{\sqrt{x}} e^{-\pi x} \, dx.$$

Proceeding as above, one obtains

$$E(X^2) = \frac{\pi}{\pi^3} \int_0^{+\infty} u^2 \times \frac{1}{\sqrt{u}} e^{-u} \, du = \frac{\pi}{\pi^3} \int_0^{+\infty} u^{\frac{5}{2}-1} e^{-u} \, du$$

One finally obtains

$$E(X^2) = \frac{\pi}{\pi^3} \Gamma\left(\frac{5}{2}\right) = \frac{\pi}{\pi^3} \times \frac{3}{2} \times \frac{1}{2} \times \sqrt{\pi} = \frac{3}{4\pi^2}.$$

Finally, an elementary calculation allows us to obtain the following result

$$\mathbb{V}(X) = \frac{1}{2\pi^2}.$$

Remark. The Fisher dispersion index of the resulting density function is less than 1. Such a variation is referred to as underdispersion.

Based on this remark, we propose a density of probability that fits under-dispersed data.

3.3. The Poisson Mixture Law

The occurrence of certain phenomena may follow a Poisson law with parameter λ considered as a random variation. We will show that the Poisson law mixed with the joint law obtained previously gives an over-dispersed law. Since the prevalence of a condition differs from one stratum to another in a population, we arrive at the following definition.

Definition 1. Let X_λ be a Poisson random variable of parameter λ which is the realization of a random variable u of continuous density g. We call the resulting law of a Poisson mixture by X_λ, the law of a random variable Y of probability density function, for all $y \in \mathbb{N}; \lambda > 0, \theta > 0, u > 0$

$$p(y,\theta,\lambda) = \int_0^{+\infty} P(X_\lambda = y) g(u,\lambda) \mathrm{d}\lambda, \ \text{with } \theta = \frac{\lambda}{u} \qquad (26)$$

where $P(X_\lambda = y) = \dfrac{\lambda^y}{y!} e^{-\lambda}$ and $g(u,\lambda) = \dfrac{1}{\sqrt{u}} e^{-\frac{1}{2}\lambda u}$ is a probability density called a mixing law.

It may be noted that similar work can be found in [10]. The following result may now be formulated.

Proposition 4. Let Y be a positive random variable with the probability mass function p resulting from the Poisson mixing law and joint probability density from GPD and Laplace-Gauss. Then, for any positive constant θ, one has

$$p(y,\theta,\lambda) = \frac{\Gamma(y+1)}{y!\sqrt{\dfrac{\lambda}{\theta}}\left(1+\dfrac{\lambda}{2\theta}\right)^y} \qquad (27)$$

with a known parameter λ.

Proof. Indeed, from (26), it is easy to show that

$$p(y,\theta,\lambda) = \frac{1}{y!\sqrt{u}} \int_0^{+\infty} \lambda^y e^{-\lambda\left(1+\frac{1}{2}u\right)} \mathrm{d}\lambda. \qquad (28)$$

At this point, observe that by setting

$$t = \lambda\left(1+\frac{1}{2}u\right) \Rightarrow \lambda = \frac{t}{1+\dfrac{1}{2}u};$$

we find the probability mass function as follows

$$p(y,\theta,\lambda) = \frac{1}{y!\sqrt{u}} \int_0^{+\infty} \left(\frac{t}{1+\dfrac{1}{2}u}\right)^y e^{-t} \mathrm{d}t = \frac{1}{y!\sqrt{u}} \frac{1}{\left(1+\dfrac{1}{2}u\right)^y} \int_0^{+\infty} t^y e^{-t} \mathrm{d}t.$$

By posing $u = \dfrac{\lambda}{\theta}$, this leads to the following form

$$p(y,\theta,\lambda) = \frac{\Gamma(y+1)}{y!\sqrt{\dfrac{\lambda}{\theta}}\left(1+\dfrac{\lambda}{2\theta}\right)^y}.$$

3.4. Estimation of the GPD Parameters by Maximum Likelihood Approach

The generalized Pareto distribution (GPD) is a two-parameter family of distributions that can be used to model exceedances over a threshold. Maximum likelihood estimators of the parameters are used, since they are asymptotically normal and asymptotically efficient. Numerical methods are required for maximizing the log-likelihood. The density of GPD function being [11]

$$f(x) = \frac{1}{\sigma}\left(1 + \gamma\left(\frac{x}{\sigma}\right)\right)^{-\frac{1}{\gamma}-1}, \text{ for } x > 0 \text{ if } \gamma > 0.$$

We then deduce the log-likelihood function of the GPD distribution

$$\log L(x_1, \cdots, x_n, \gamma, \sigma) = -n\ln\sigma - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{n}\ln\left(1 + \gamma\frac{x_i}{\sigma}\right).$$

By deriving the log-likelihood function with respect to each parameter, one obtains the following system

$$\begin{cases} -\dfrac{n}{\sigma} + \left(\dfrac{1}{\gamma}+1\right)\sum_{i=1}^{n}\dfrac{\gamma x_i}{\sigma^2 + \sigma\gamma x_i} = 0 \\ \sum_{i=1}^{n}\ln\left(1 + \dfrac{\gamma}{\sigma}x_i\right) - \gamma(\gamma+1)\sum_{i=1}^{n}\dfrac{x_i}{\sigma + \gamma x_i} = 0 \end{cases} \quad (29)$$

By numerical algorithms, of the GPD function of the R software, one obtains the parameters in **Table 1** based on data from dietary intake of dioxins of men exposure, such as (note that data is taken from **Table 3**)

$$Male = (181.8, 137.7, 230.7, 178.2, 181.5, 192.6, 113.1, 75.6, 59.4, 45, 38.1)$$

3.5. Estimation of the GPD Parameters by Alienor's Method

We now investigate the Alienor approach for efficient estimation of the parameters. Considering the bi-parametric probability density function of GPD law whose analytical expression is given by

$$f(x) = \frac{1}{\sigma}\left(1 + \gamma\left(\frac{x}{\sigma}\right)\right)^{-\frac{1}{\gamma}-1};$$

where γ, σ, x are the shape parameter, the scale parameter, and the excess variable, respectively. By section (2.2), one has the steps as follows:

Step 1. Definition of the objective function.

The likelihood function associated with probability density function of GPD gives

Table 1. The parameters MLM.

$\hat{\gamma}$	$\hat{\sigma}$
0.3576522	106.5762338

$$L(x_1,\cdots,x_n,\gamma,\sigma) = \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^n \left(1 + \frac{\gamma}{\sigma}x_i\right)^{-\frac{1}{\gamma}-1}.$$

The objective function noted

$$L^* = L \circ h : [0,1] \rightarrow [0,1]$$

$$\theta \mapsto L \circ h(\theta).$$

Knowing that

$$h(\theta) = \gamma(\theta) = \frac{1}{2}\left(1 - \cos(m \times \pi\theta)\right)$$

$$h(\theta) = \sigma(\theta) = \frac{1}{2}\left(1 - \cos(m^2 \times \pi\theta)\right).$$

Thus,

$$L^*(\theta) = \left[\frac{1}{\frac{1}{2}\left(1 - \cos(m^2 \times \pi\theta)\right)}\right]^n \times \left[\prod_{i=1}^n \left(1 + \frac{1 - \cos(m \times \pi\theta)}{1 - \cos(m^2 \times \pi\theta)}x_i\right)^{-\frac{1}{1-\cos(m \times \pi\theta)}-1}\right]. \quad (30)$$

We have a maximization problem that we can transform into a minimization problem as follows:

$$\max_{\theta \in [0,1]} L^*(\theta) = -\min_{\theta \in [0,1]}\left(-L^*(\theta)\right).$$

$$-L^* : [0,1] \rightarrow [-1,0]$$

$$\theta \mapsto -L^*(\theta).$$

Step 2. Application of $O.P.O^*$

Let's arbitrarily choose $y_0 = -L^*(\theta_o) \in [-1,0]$ and let's apply $O.P.O^*_{T_L}$ to the point y_0. By Choosing $\theta_0 = \frac{1}{4}$, $m = 2$ and for $n = 8$ distribution of contaminant observations (32) above the 70 kg/body weight threshold, by using (30), one obtains

$$L^*(\theta_0) = \left[\frac{1}{\frac{1}{2}\left(1 - \cos\left(4\pi \times \frac{1}{4}\right)\right)}\right]^8$$

$$\times \left[1 + \frac{1 - \cos\left(2\pi \times \frac{1}{4}\right)}{1 - \cos\left(4\pi \times \frac{1}{4}\right)} \times 181.8\right]^{-\frac{1}{1-\cos\left(2\pi \times \frac{1}{4}\right)}-1}$$

$$\times \left[1 + \frac{1 - \cos\left(2\pi \times \frac{1}{4}\right)}{1 - \cos\left(4\pi \times \frac{1}{4}\right)} \times 137.7\right]^{-\frac{1}{1-\cos\left(2\pi \times \frac{1}{4}\right)}-1}$$

$$\times\left[1+\frac{1-\cos\left(2\pi\times\frac{1}{4}\right)}{1-\cos\left(4\pi\times\frac{1}{4}\right)}\times230.7\right]^{-\frac{1}{1-\cos\left(2\pi\times\frac{1}{4}\right)}-1}$$

$$\times\left[1+\frac{1-\cos\left(2\pi\times\frac{1}{4}\right)}{1-\cos\left(4\pi\times\frac{1}{4}\right)}\times178.2\right]^{-\frac{1}{1-\cos\left(2\pi\times\frac{1}{4}\right)}-1}$$

$$\times\left[1+\frac{1-\cos\left(2\pi\times\frac{1}{4}\right)}{1-\cos\left(4\pi\times\frac{1}{4}\right)}\times181.5\right]^{-\frac{1}{1-\cos\left(2\pi\times\frac{1}{4}\right)}-1}$$

$$\times\left[1+\frac{1-\cos\left(2\pi\times\frac{1}{4}\right)}{1-\cos\left(4\pi\times\frac{1}{4}\right)}\times192.6\right]^{-\frac{1}{1-\cos\left(2\pi\times\frac{1}{4}\right)}-1}$$

$$\times\left[1+\frac{1-\cos\left(2\pi\times\frac{1}{4}\right)}{1-\cos\left(4\pi\times\frac{1}{4}\right)}\times113.1\right]^{-\frac{1}{1-\cos\left(2\pi\times\frac{1}{4}\right)}-1}$$

$$\times\left[1+\frac{1-\cos\left(2\pi\times\frac{1}{4}\right)}{1-\cos\left(4\pi\times\frac{1}{4}\right)}\times75.6\right]^{-\frac{1}{1-\cos\left(2\pi\times\frac{1}{4}\right)}-1}$$

we obtain after the previous calculation, the value of the objective function of the sample

$$L^{*}\left(\theta_{0}\right)=2.069063\times10^{-23} \tag{31}$$

By applying the operator (8) and (9) of (2.2), one has $\theta_{0}=7.749788e^{-10}$, the one that maximizes likelihood, this gives the following estimators (**Table 2**)

4. Some Technical Results

Using the method of maximum likelihood estimation encounters difficulties including lack of solutions of the likelihood. Moreover, in the case where solutions exist, the numerical resolution of the equations requires an iterative method which is often not convergent. Thus, to solve these difficulties, the concept of OPO is introduced.

Table 2. The parameters obtained by AM.

$\hat{\gamma}$	$\hat{\sigma}$
0.143527	98.717793

4.1. Description of the Data

The data were provided by Health Canada, which is an active participant in the international assessments of chemical contaminants in foods conducted by the Joint FAO/WHO Expert Committee on Food Additives. Dioxins are currently on the CODEX priority list for evaluation by JECFA, which is driving our interest in this area of research. The tolerable threshold for dioxins in humans, the recent JECFA evaluations is 70 pg/kg/month and the dioxin intake by age and sex is shown in the table which represents the extreme values observed in males and females. Our study determines the incubation period and the probability of occurrence of the event if we assume that exceeding the tolerable threshold for the organism (70 pg/kg/month) is dangerous for the health. Of course, the application is done on two cohorts. The first investigation points out the evolution of the quantity of the contaminant in the man's body and the second is based on the woman's one. Table 3 shows that women of childbearing age between 15 and 45 years old are daily exposed to dioxins between 1.21 and 1.78 pg/kg/day.

Table 3 shows that the newborn of 4 - 6 months is much more exposed to the effects of dioxins followed by those who are 1 - 4 years old. According to Figure 1, some foods contain large amount of dioxin (pg/g body weight), namely butter, poultry, chicken, milk then ground beef, egg, etc. And an excessive

Table 3. Daily and Monthly intake of dioxin.

Age-Gender Groups	DDI	MDI
0 - 1 month, boys and girls	6.06	181.8
2 - 3 months, boys and girls	4.59	137.7
4 - 6 months, boys and girls	7.69	230.7
7 - 9 months, boys and girls	5.94	178.2
9 - 12 months, boys and girls	6.05	181.5
1 - 4 years, boys and girls	6.42	192.6
5 - 11 years, boys and girls	3.77	113.1
12 - 19 years, boys and girls	2.52	75.6
20 - 39 years, men	1.98	59.4
40 - 64 years, men	1.5	45
65+ years, men	1.27	38.1
12 - 19 years, girls	1.78	53.4
20 - 39 years, women	1.53	45.9
40 - 64 years, women	1.21	36.3
65+ years women	0.99	29.7

DDI: is Daily Dioxin Intake (pg/kg bw/day) and MDI: is Monthly Dioxin Intake (pg/kg bw/month). bw: body weight.
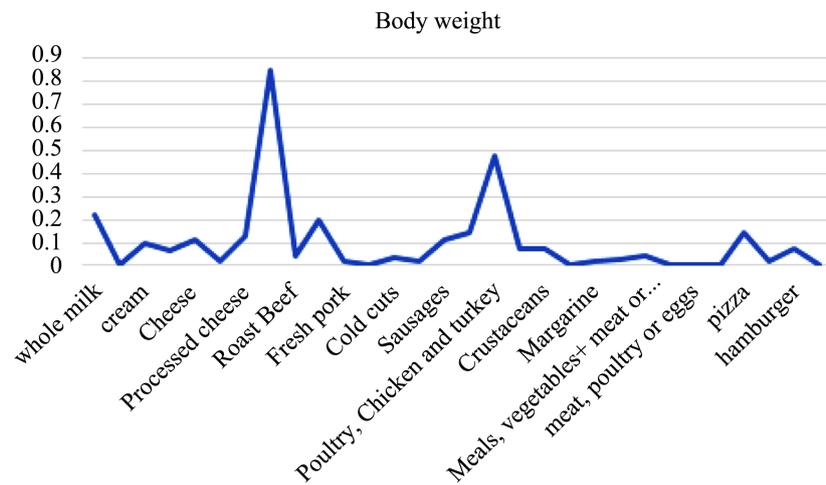
Figure 1. Toxin levels in food.

consumption of these foods can be dangerous for your health (the data presented in this **Figure 1** are available on request from the authors). The objective of this section is to assess exposure by means of two indicators, the probability that the exposure process exceeds a threshold based on the extreme value theory and the incubation period of the event. The ruin probability is obtained by GPD model.

4.2. Digital Application

One has used the intake of dioxin in the diet, an extreme distribution by age group and gender. Let us point out that the investigation was conducted based on two cohorts, male and female.

4.2.1. Male Gender

Cancers are diseases that are usually detected at the end of a person's life. However, carcinogenic pathologies (what we consider extreme events) are the body's response to contamination processes. Our approach is a probabilistic approach in modeling the risks incurred by a population in the exposure to food risks, telling when a person is at risk and developing a pathology following regular consumption of food contaminated by dioxins. We have the dietary intake of dioxin in males from 0 to 65 years of age and over. According to **Table 3**, we have the following data: (**Figure 2**)

$$Male = (181.8, 137.7, 230.7, 178.2, 181.5, 192.6, 113.1, 75.6, 59.4, 45, 38.1) \quad (32)$$

The data on the evolution of the contamination of men show according to the visual diagnosis that from 0 to 100 pg/kg/month the distributions of the observations are stationary and from 100 to 250 pg/kg/month the behavior is Gaussian. This brings us to a normality test to see if the observations converge to a normal distribution in order to generate a sample size of 1000. We adjust our data by normal law and follow Shapiro's test. The adjustment of the data of our sample by the normal distribution gives 0.91263 representing a high rate of
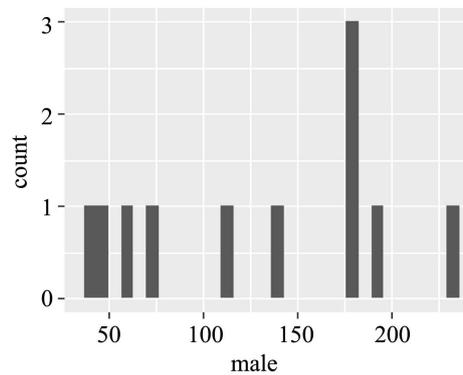
Figure 2. Evolution of dioxin contamination.

compatibility between our data and the normal distribution and a p-value equal to 0.263. This last value is thus higher than the significance threshold $\alpha = 0.05$, so the sample follows a normal distribution. The choice of the threshold is very important because it induces a great variability in the estimation of extreme quantiles and parameters of the law of excesses. There are different approaches for the choice of the threshold of the POT method. Usually, the threshold is determined graphically by the Mean Excess Plot (meplot function) in R software (**Figure 3**).

According to **Figure 4**, the threshold at value $u = 75$. But the threshold 70 recommended by FAO/WHO is used for the estimation of the parameters. It can be seen that the distribution of excesses converges towards a Generalized Pareto Distribution (GPD). Recall that the upward trend in 4 shows heavy tail behavior. In particular, a straight line with a positive gradient above a certain threshold is a sign of Pareto tail behavior. The survival function of the GPD law depends on the following parameters from **Table 1**. In the simulated sample size 1000, there are 811 observations that exceed the threshold 70.

$$\mathbb{H}_{\hat{\gamma},\hat{\sigma}}\left(x_{PT}\right) = 1 - \left(1 + \frac{0.3576522 x_{pt}}{106.5762338}\right)^{\frac{-1}{0.3576522}} \quad \text{si } \gamma \neq 0.$$

We used (18) an approach based on the excess method by the approximation of GPD (knowing that $N_u = 811$ and 1-pt is equivalent to 95%)

$$\hat{x}_{pt}^{GPD} = 401.9712377$$

The incubation period obtained using the Formula (19) based on the maximum likelihood method is:

$$\hat{T}_r^{GPD} = 13.42598873$$

Using the result in **Table 2** based on Alienor approach, we obtain:

$$\hat{T}_r^{GPD} = 2.12058$$

We applied the POT approach to male dietary dioxin intake data. We estimated the parameters of the GPD distribution by using the maximum likelihood method, then the extreme quantile and the incubation period. The estimation of
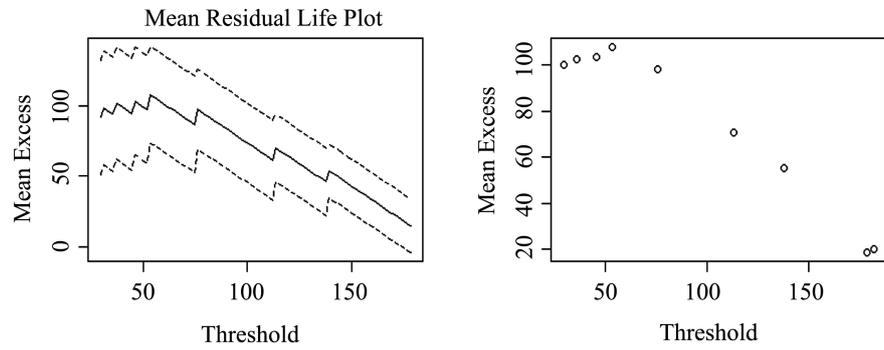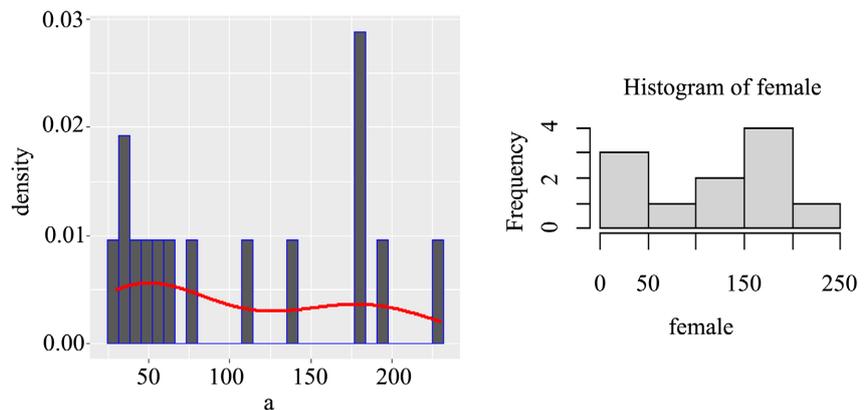
Figure 3. Diagnostic of threshold.

Figure 4. Evolution of dioxin contamination.

the incubation period allows us to conclude that male exposure to dioxins can cause serious damage to the organism approximately 13.4259 months after dioxin contamination, *i.e.* 13 months 12 days 18 hours after the man is considered at risk and can manifest a pathology with a probability of 0.811 but approximately 2 months based on Alienor approach.

4.2.2. Female Gender

We need to use the dietary intake of dioxin for females from 0 to 65 years of age and older according to **Table 3**.

$$Female = \left(181.8, 137.7, 230.7, 178.2, 181.5, 192.6, 113.1, 53.4, 45.9, 36.3, 29.7\right). \quad (33)$$

After diagnosis of **Figure 4**, we notice Pareto shape at the beginning of the histogram and Gaussian behavior thereafter. This has motivated the hypothesis of proposition 2. **Figure 5** illustrates the trajectory of the joint probability law, by numerical resolution.

The distribution obtained previously is used in the model. The GPD was fitted by the maximum likelihood method from 469 values exceeding the threshold, among the 1000 simulated values. We can estimate that the ruin probability is 0.469. The estimated parameters are summarized in **Table 4**.

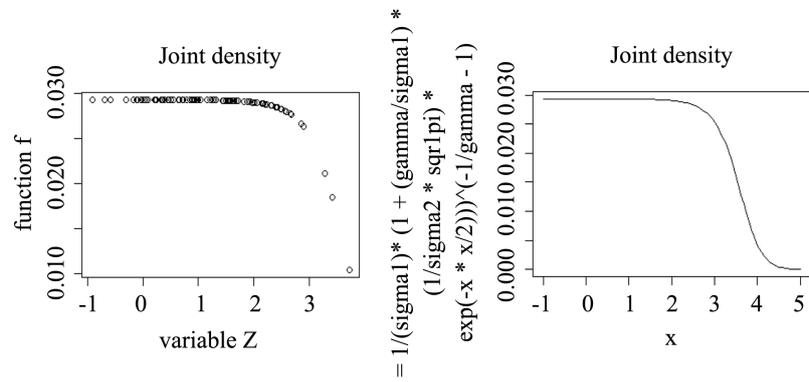The adequacy of the data to the GPD law allows the estimation of an extreme quantile by **Table 4**.

Figure 5. The trajectory of the joined probability density.

Table 4. The parameters of scale and shape.

$\hat{\gamma}$	$\hat{\sigma}$
0.09576862	119.45236039

$$\hat{x}_{pt}^{GPD} = 377.4350556$$

And using **Table 4** and the survival function gives.

$$\mathbb{H}_{\hat{\gamma}\hat{\sigma}}\left(x_{pt}\right) = 0.9367358776$$

The incubation period of the event obtained after calculus is

$$\hat{T}_{sim}^{GPD} = 15.83842421$$

Estimating the incubation period allows us to conclude that exposure of women to dioxins can cause serious damage to the body approximately 15.83842421 months after exposure. That is to say that a woman who maintains a steady consumption pattern of dioxin-contaminated foods such as fish, poultry, and butter… can develop pathology in 15 months 25 days 3 hours after infection.

In summary, one obtains that the incubation period for dioxin exposure is defined as follows

Male: 13 months 12 days 18 hours;

Woman: 15 months 25 days 3 hours.

5. Conclusion and Discussions

With this paper, we have contributed to estimating the incubation period following exposure to dioxins, a risk indicator that will allow decision-makers to take measures to reduce the exposure of consumers, and waste incinerators (resulting from respiratory contamination). This paper developed also instead of the existing law a new probability density function for a phenomenon that follows both normal and GPD distributions. In the future, it will be taken into account the different aspects of contamination which are the interaction of contaminants with others once they enter the body and the different compartments in which they are stored.

Acknowledgements

The authors would like to thank Health Canada through **Jose Roy**, Substance Management Information Line Officer and **Cynthia Bainbridge**, Unit Head, Risk Management of Persistent Organic Pollutants for the data made available and we would also like to thank the **referee** for his details comments. His comments have led to a much clearer and better presentation of our results.

Data Availability

The data used to support the findings of this study were supplied by **Health Canada**.

Conflicts of Interest

The authors declare that there are no conflicts of interest for the publication of this paper.

References

[1] Lauer, S.A., Grantz, K.H., Bi, Q.F., Jones, F.K., Zheng, Q.L., Meredith, H.R., Azman, A.S., Reich, N.G. and Lessler, J. (2020) The Incubation Period of Coronavirus Disease 2019 (COVID-19) from Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, **172**, 577-582. https://doi.org/10.7326/M20-0504

[2] Tressou, J., Crepet, A., Bertail, P., Feinberg, M.H. and Leblanc, J.Ch. (2004) Probabilistic Exposure Assessment to Food Chemicals Based on Extreme Value Theory. Application to Heavy Metals from Fish and Sea Products. *Food and Chemical Toxicology*, **42**, 1349-1358. https://doi.org/10.1016/j.fct.2004.03.016

[3] Moncoulon, D. and Quantin, A. (2013) Modélisation des événements extrêmes d'inondation en France métropolitaine. *La Houille Blanche*, No. 1 , 22-26. https://doi.org/10.1051/lhb/2013004

[4] Diakarya, B., Blami, K. and Moussa, S. (2012) Spatial Stochastic Framework for Sampling Time Parametric Max-Stable Processes. *International Journal of Statistics and Probability*, **1**, 203. https://doi.org/10.5539/ijsp.v1n2p203

[5] Nguyen, T.-H., El Outayek, S., Lim, S.H. and Nguyen, V.T.V. (2017) A Systematic Approach to Selecting the Best Probability Models for Annual Maximum Rainfalls—A Case Study Using Data in Ontario (Canada). *Journal of Hydrology*, **553**, 49-58. https://doi.org/10.1016/j.jhydrol.2017.07.052

[6] Diamoutene, A., Kamsu-Foguem, B., Noureddine, F. and Barro, D. (2018) Prediction of U.S. General Aviation Fatalities from Extreme Value Approach. *Transportation Research Part A*: *Policy and Practice*, **109**, 65-75. https://doi.org/10.1016/j.tra.2018.01.022

[7] Konfe, B.O., Cherruault, Y., Some, B. and Benneouala, T. (2005) Alienor Method Applied to Operational Research. *Kybernetes*, **34**, 1211-1222. https://doi.org/10.1108/03684920510605984

[8] Adedje, K.E. and Barro, D. (2022) A Stochastic Approach to Modeling Food Pattern. *International Journal of Mathematics and Mathematical Sciences*, **2022**, Article ID: 9011873. https://doi.org/10.1155/2022/9011873

[9] Velázquez, J.C., Van Der Weide, J.A.M, Hernández, E. and Hernández, H.H. (2014)

Statistical Modelling of Pitting Corrosion: Extrapolation of the Maximum Pit Depth-Growth. *International Journal of Electrochemical Science*, **9**, 4129-4143.
http://www.electrochemsci.org/papers/vol9/90804129.pdf

[10] Jongbloed, G. and Koole, G. (2014) Managing Uncertainty in Call Centres Using Poisson Mixtures, *Applied Stochastic Models in Business and Industry*, **17**, 307-318.
https://doi.org/10.1002/asmb.444

[11] Barro, D., Diallo, M. and Bagré, R.G. (2016) Spatial Tail Dependence and Survival Stability in a Class of Archimedean Copulas. *International Journal of Mathematics and Mathematical Sciences*, **2016**, Article ID: 8927248.
https://doi.org/10.1155/2016/8927248