Research Article

Resolution of the Min-Max Optimization Problem Applied in the Agricultural Sector with the Estimation of Yields by Nonparametric Statistical Approaches

Ghizlane Kouaiba and Driss Mentagui

Faculty of Sciences, Department of Mathematics, Ibn Tofail University, Kenitra, Morocco

Correspondence should be addressed to Ghizlane Kouaiba; kouaiba.ghizlane@gmail.com

Received 2 October 2020; Accepted 3 March 2021; Published 15 April 2021

Academic Editor: Paul Elie

Copyright © 2021 Ghizlane Kouaiba and Driss Mentagui. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The ultimate objective of the problem under study is to apply the min-max tool, thus making it possible to optimize the default risks linked to several areas: the agricultural sector, for example, which requires the optimization of the default risk using the following elements: silage crops, annual consumption requirements, and crops produced for a given year. To minimize the default risk in the future, we start, in the first step, by forecasting the total budget of agriculture investment for the next 20 years, then distribute this budget efficiently between the irrigation and construction of silos. To do this, Bangladesh was chosen as an empirical case study given the availability of its data on the FAO website; it is considered a large agricultural country in South Asia. In this article, we give a detailed and original in-depth study of the agricultural planning model through a calculating algorithm suggested to be coded on the R software thereafter. Our approach is based on an original statistical modeling using nonparametric statistics and considering an example of a simulation involving agricultural data from the country of Bangladesh. We also consider a new pollution model, which leads to a vector optimization problem. Graphs illustrate our quantitative analysis.

1. Introduction

The study is based on the quantification of the risk of having the need exceeding the production and the quantity of the production already ensiled; the same quantification will be applied in the case where the production of a given year exceeds both the need as well as the capacity of the silos for the same year; the idea is to calculate these risks of faults over 20 years in the future, based on the total investment amounts allocated for irrigation and the construction of the silos planned over these 20 years, via 3 calculation methods simulated in N iterations (N distributions of the amounts of irrigation and construction of the silos), and optimize them via a vector Pareto optimization algorithm for faults calculated by considering the pollution and those calculated without the assumption of the pollution; the strategies (belonging to the Pareto front) said optimal strategies will be linked to their amount of investment in irrigation and construction of silos affected and based on the total of these planned investment

amounts. Therefore, it is up to the decision-maker to choose a strategy for allocating these investment amounts, among all optimal strategies retained (strategies belong to the Pareto front).

2. The Problem Schematization

2.1. The General Idea of the Work. Minimize the difference between the production $P(n)$, the requirement $\varphi(n)$, and the quantity to be removed or ensiled $Q(n)$ after having maximized the risk through 3 scenarios that will be well explained after.

The variables

$$\begin{aligned} S_{\text{tot}} &= S(n) + s(n), \\ P(n) &= p(n) * S(n) + q(n) * s(n), \end{aligned} \quad (1)$$

with S_{tot} as the total cultivated area in year n , $S(n)$ the surface area of irrigated land with yield p in year n , $s(n)$ the surface

of nonirrigated land with yields q in year n , $P(n)$ the total production in year n , $p(n)$ the yield of irrigated land, and $q(n)$ the yield of nonirrigated land.

Knowing the requirement $\varphi(n)$ of year n , the difference between the production $P(n)$ of year n and the requirements can be explained as follows:

$$\begin{aligned} \text{If : } P(n) - \varphi(n) > 0 : \text{ quantity to ensile} \\ \text{Otherwise: the silo stock is to be consumed} \end{aligned} \quad (2)$$

However, the quantities to ensile or to consume are not often equal, respectively, to the capacity of the silo or to the harvest stock for a given year, which requires an optimization of resources. To do this, we need to calculate the default Δ , as follows:

$$\Delta(n) = (P(n) - \varphi(n)) - Q(n), \quad (3)$$

with

$$Q(n) = \begin{cases} \min(P(n) - \varphi(n), G(n) - R(n-1)), & \text{if } P(n) \geq \varphi(n), \\ \max(P(n) - \varphi(n), -R(n-1)), & \text{if } P(n) < \varphi(n), \end{cases} \quad (4)$$

Such as

- (i) $G(n)$: the silo capacity in year n
- (ii) $R(n-1) = R(n-2) + Q(n-1)$: the quantity ensiled in the silo in $n-1$:

$$G(n) = G(n-1) + \frac{P_{\text{cons}}(n-1)}{c_x} \quad (5)$$

- (iii) $P_{\text{cons}}(n-1)$: the budget allocated to the construction of the silo for year n

- (iv) c_x : the cost price of a unit of silo capacity:

$$S(n) = S(n-1) + \frac{P_{\text{irrig}}(n-1)}{c_y}, \quad (6)$$

- (v) $P_{\text{irrig}}(n-1)$: the budget allocated to irrigation for year n
- (vi) c_y : the expenditure per unit of irrigated land
- (vii) $\varphi(n)$: the product requirements in year n

Therefore, the total budget $P_{\text{tot}}(n)$ is given by

$$P_{\text{tot}}(n) = P_{\text{irrig}}(n) + P_{\text{cons}}(n). \quad (7)$$

3. Fundamental Elements for the Calculation

- (1) Initial conditions: G (2001), R (2000), S (2001), p (2001), q (2001), s (2001), P_{irrig} (2000), P_{cons} (2000), and φ (2001): to fix by assumptions

- (2) Total agricultural areas, areas of irrigated and non-irrigated land

The data was collected via the FAO site, selected country: Bangladesh (total agricultural land = 8.549 million ha and irrigated land areas) [1, 2].

The surface area of nonirrigated lands has been deducted as shown in Table 1.

- (3) Irrigation and silo construction budgets: based on FAO data (if we consider that the unit costs of irrigation and silo construction are, respectively $c_y = 622$ \$/ha [3] and $c_x = 25$ \$/tonne [4] (for the case of tank stores, source FAO)), and based on the total production of the crops and the irrigated land area, we will have Table 2

It is also assumed that 75% of the total production is devoted to cereals; therefore, the total production can be approximated to the production of cereals.

- (4) Irrigated and nonirrigated lands yields: based on FAO publications [5], the yields of irrigated land are recorded between 5 tonnes/ha and 13 tonnes/ha from an irrigation of 4500 m³/ha and from 0 tonne/ha to 5 tonnes/ha for nonirrigated lands; therefore, we go through a numerical simulation of the respective yield values of irrigated and nonirrigated lands to deduce the simulated productions, while assuming these yields, do not take into account the factor of pollution (the yields of the supposedly polluted land is to be deduced later) (see Table 3)

In order to determine the distribution function of the returns p and q in order to deduce the expectations, nonparametric methods will be the subject of the next part of this study.

4. Nonparametric Estimation of Yield Density: Theoretical Study

4.1. The Histogram Method. The basic idea is to segment the observations belonging to the interval $[a, b]$ into m intervals with length $h = (b - a)/m$.

In addition, for $\forall x_i \in C_k = [L_{k-1}, L_k[$ / $i \in \{1, \dots, n\}$, $k \in \{1, \dots, m\}$ and $L_k = L_{k-1} + h$.

n is the number of the observed values in the interval $[a, b]$, and x is a continued value that belongs to the same interval that we are looking for its probability formula.

The density function [6] associated to x is in the following form:

$$\hat{f}_H(x) = \sum_{i=1}^n \frac{1}{nh} I_{[L_{k-1}, L_k[}(x_i), x \in C_k = \frac{\hat{p}_k}{h}, \quad (8)$$

Such that $\hat{p}_k(x) = \sum_{i=1}^n (1/n) I_{[L_{k-1}, L_k[}(x_i), x \in C_k$.

TABLE 1: Total agricultural areas, areas of irrigated and nonirrigated lands.

Year	Irrigated lands (1000 ha)	Nonirrigated lands (1000 ha)
2000	3850	4699
2001	3985	4564
2002	4187	4362
2003	4421	4128
2004	4597	3952
2005	4725	3824
2006	4817	3732
2007	4875	3674
2008	4965	3584
2009	5007	3542
2010	5050	3499
2011	5100	3449
2012	5180	3369
2013	5264	3285
2014	5300	3249
2015	5500	3049
2016	5500	3049

TABLE 2: Irrigation and silo construction budgets and simulated requirements.

Year	Irrigation expenses (million USD)	Cost for the construction of silos (million USD)	Simulated requirements (million tonnes)
2000	35,76589	20,3	24,36247
2001	37,18973	12,7	8,25524
2002	38,22525	21,6	19,56358
2003	38,96953	13,8	78,19966
2004	39,43875	39,9	28,2052
2005	40,16685	17,4	13,24554
2006	40,50663	13,2	104,13378
2007	40,8545	33,5	5,73262
2008	41,259	24,7	26,96004
2009	41,9062	19,3	2,48501
2010	42,58576	37,8	14,62258
2011	42,877	15,8	26,44989
2012	44,495	11,9	17,04362
2013	44,495	13,1	5,39411
2014	44,495	7,8	2,74671
2015	44,495	22,6	13,35862

Therefore, the problem comes down to estimating the probability vector:

$$\hat{p} = \{\hat{p}_1, \dots, \hat{p}_k, \dots, \hat{p}_m\}. \quad (9)$$

While the question that arises is how to choose h , this comes down to choosing the number of intervals to have a well-smoothed distribution (do not fall in the case of a histo-

TABLE 3: Irrigated and nonirrigated land yields.

Year	Irrigated lands yield (tonnes/ha)	Non-irrigated lands yield (tonnes/ha)
1986	6,61	2,93
1987	10,56	1,81
1988	13,06	1,19
1989	8,21	2,4
1990	7,59	2,75
1991	9,38	1,82
1992	7,49	2,91
1993	10,63	0,86
1994	7,02	3,03
1995	10,85	0,47
1996	8,79	1,93
1997	5,7	4,28
1998	9,61	2,54
1999	11,77	1,37
2000	7,71	4
2001	11,69	0,47
2002	10	2
2003	8,75	2,75
2004	12,32	0,15
2005	11,7	0,91
2006	9,95	3,5
2007	12,38	1,76
2008	6,61	2,93
2009	10,56	1,81
2010	13,06	1,19
2011	8,21	2,4
2012	7,59	2,75
2013	9,38	1,82
2014	7,49	2,91
2015	10,63	0,86
2016	7,02	3,03

gram called “oversmoothing” when h is larger or the opposite case of a histogram called “undersmoothing” when h tends towards zero).

(1) The quadratic risk of $\hat{f}_H(x)$

The solution proposed in this sense is to establish a risk function $R(h)$ and minimize it as a function of the window h ; the function adequate to this problem is that expressing the expectation of the quadratic deviation between $\hat{f}_H(x)$ and $f(x)$:

$$\begin{aligned} \text{EQM}(x, h) &= E\left(\hat{f}_H(x) - f(x)\right)^2 \\ &= \left(E\left(\hat{f}_H(x)\right) - f(x)\right)^2 + E\left(\hat{f}_H(x) - E\left(\hat{f}_H(x)\right)\right)^2 \\ &= \text{bias}^2\left(\hat{f}_H(x)\right) + \text{variance}\left(\hat{f}_H(x)\right), \end{aligned} \quad (10)$$

such that

$$E(\hat{f}_H(x)) = E\left(\sum_{i=1}^n \frac{1}{nh} I_{[L_{k-1}, L_k]}(x_i)\right) = \frac{1}{nh} E\left(\sum_{i=1}^n I_{[L_{k-1}, L_k]}(x_i)\right). \quad (11)$$

However, $\sum_{i=1}^n I_{[L_{k-1}, L_k]}(x_i) = n_k$ and $\sum n_k = n$; this says that $n_k \sim \mathcal{B}(p_k, n)$: the law of succeeding an experiment with a probability $p_k = n_k/n$, repeating n times.

Therefore, we will have

$$E(\hat{f}_H(x)) = \frac{np_k}{nh} = \frac{p_k}{h}, \quad (12)$$

which implies

$$\text{bias}^2(\hat{f}_H(x)) = \left(E(\hat{f}_H(x)) - f(x)\right)^2 = \left(\frac{p_k}{h} - f(x)\right)^2, \quad (13)$$

and that

$$\begin{aligned} \text{variance}(\hat{f}_H(x)) &= \frac{1}{(nh)^2} \text{variance}\left(\sum_{i=1}^n I_{[L_{k-1}, L_k]}(x_i)\right) \\ &= \frac{np_k(1-p_k)}{(nh)^2} = \frac{p_k(1-p_k)}{nh^2}. \end{aligned} \quad (14)$$

Finally, the expectation of the quadratic deviation of $\hat{f}_H(x)$ is written as follows:

$$\text{EQM}(x, h) = \left(\frac{p_k}{h} - f(x)\right)^2 + \frac{p_k(1-p_k)}{nh^2}. \quad (15)$$

(2) The quadratic risk of $\hat{f}_H(x)$ integrated

For the function to be as a function of h , the EQM(h, x) just mentioned, it will be convenient to integrate the function over the interval $[a, b]$:

$$\begin{aligned} \text{IntegratedRisk}(h) &= R(h) = \int_a^b \text{EQM}(x) dx \\ &= \int_a^b \left(\frac{p_k}{h} - f(x)\right)^2 dx + \int_a^b \frac{p_k(1-p_k)}{nh^2} dx, \\ R(h) &= \int_a^b \text{EQM}(x) dx \\ &= \sum_k \int_{c_k} \left(\frac{p_k}{h} - f(x)\right)^2 dx + \sum_k \int_{c_k} \frac{p_k(1-p_k)}{nh^2} dx, \\ R(h) &= -\frac{\sum_k p_k^2}{h} + \int_a^b f^2(x) dx + \frac{1}{nh} - \frac{\sum_k p_k^2}{nh}, \\ R(h) &= -\frac{n+1}{nh} \sum_k p_k^2 + \int_a^b f^2(x) dx + \frac{1}{nh}. \end{aligned} \quad (16)$$

It seems that the speed of convergence of $R(h)$ depends on nh , which leads us to reformulate the $R(h)$ and going through the limited expansion of $f(x)$:

$$\begin{aligned} E(\hat{f}_H(x)) - f(x) &= \frac{p_k}{h} - f(x) = \frac{\int_{c_k} f(t) dt}{h} - f(x) \\ &= \frac{1}{h} \int_{c_k} (f(t) - f(x)) dt. \end{aligned} \quad (17)$$

Passing through the limited development of the first Taylor degree at point x , we will have

$$f(t) - f(x) = (t-x)f'(x) + o(t-x). \quad (18)$$

Therefore,

$$\int_{c_k} (f(t) - f(x)) dt = \int_{c_k} ((t-x)f'(x) + o(h^2)) dt. \quad (19)$$

Proof of Equation 19. We consider that $v(t) = (f(t) - f(x) - (t-x)f'(x))$.

Moreover, based on the definition of a negligible function [7] on a point that says the following: let a function $f: I \rightarrow \mathbb{R}$ and f not vanish at $I \setminus \{a\}$, f is said to be negligible in the neighborhood of $\{a\}$ in front of g if and only if

$$\forall \varepsilon > 0 \exists \alpha > 0, |x-a| < \alpha \implies |f(x)| < \varepsilon |g(x)|. \quad (20)$$

Since $v(t)$ is negligible in front of $(t-x)$, then

$$\begin{aligned} \forall \varepsilon > 0 \exists \alpha > 0, \\ |t-a| < \alpha \implies |v(t)| < \varepsilon |t-x|. \end{aligned} \quad (21)$$

$\forall \varepsilon > 0 \exists \alpha > 0$ and $|t-a| < \alpha \implies |v(t)| < \varepsilon |(t-x)| < \varepsilon h$ because $(t, x) \in c_k$.

Therefore,

$$\begin{aligned} \int_{c_k} |v(t)| dt &< \int_{c_k} \varepsilon h dt = \frac{\varepsilon h^2}{2} = \varepsilon' h^2. \\ &< \varepsilon |h^2| \end{aligned} \quad (22)$$

However, we have

$$\left| \int_{c_k} v(x) dt \right| < \int_{c_k} |v(t)| dt. \quad (23)$$

Therefore,

$$\begin{aligned} \forall \varepsilon' > 0 \exists \alpha > 0, \\ |t-a| < \alpha \implies \left| \int_{c_k} v(t) dt \right| < \varepsilon h^2. \end{aligned} \quad (24)$$

Finally,

$$\int_{c_k} \nu(t) dt = \int_{c_k} \left(f(t) - f(x) - (t-x)f'(x) \right) dt = o(h^2), \quad (25)$$

which implies

$$\int_{c_k} (f(t) - f(x)) dt = \int_{c_k} (t-x)f'(x) dt + o(h^2). \quad (26)$$

After the calculation is done, we will have

$$\begin{aligned} \int_{c_k} (t-x)f'(x) dt &= \left(h^2 \left(k - \frac{1}{2} \right) - xh \right) f'(x) \\ &= h \left(h \left(k - \frac{1}{2} \right) - x \right) f'(x), \\ \int_{c_k} (t-x)f'(x) dt &\approx h^2 f'(x). \end{aligned} \quad (27)$$

Finally, we will have

$$E(\hat{f}_H(x)) - f(x) = \frac{1}{h} \int_{c_k} (f(t) - f(x)) dt \approx \frac{1}{h} \left(h^2 f'(x) + o(h^2) \right),$$

$$E(\hat{f}_H(x)) - f(x) \approx \left(hf'(x) + o(h) \right). \quad (28)$$

For its part, the variance is given by

$$\begin{aligned} \text{variance}(\hat{f}_H(x)) &= \frac{p_k(1-p_k)}{nh^2}, \\ \text{variance}(\hat{f}_H(x)) &= \frac{1}{nh^2} \int_{c_k} f(t) dt \left(1 - \int_{c_k} f(t) dt \right). \end{aligned} \quad (29)$$

If we approach the quantity $(1 - \int_{c_k} f(t) dt)$ to 1, the variance can be expressed in the following form:

$$\begin{aligned} \text{variance}(\hat{f}_H(x)) &\approx \frac{1}{nh^2} \left(hf(x) + h^2 f'(x) + o(h^2) \right) \\ &= \frac{f(x)}{nh} + \frac{f'(x)}{n} + o\left(\frac{1}{n}\right). \end{aligned} \quad (30)$$

Finally,

$$\begin{aligned} \text{EQM}(x, h) &= \left(hf'(x) + o(h) \right)^2 + \frac{f(x)}{nh} + \frac{f'(x)}{n} + o\left(\frac{1}{n}\right) \\ &= \left(hf'(x) \right)^2 + 2f'(x)o(h^2) + o(h^2) + \frac{f(x)}{nh} \\ &\quad + \frac{f'(x)}{n} + o\left(\frac{1}{n}\right) = \left(hf'(x) \right)^2 + \frac{f(x)}{nh} + \frac{f'(x)}{n} \\ &\quad + o\left(\frac{1}{n}\right) + o(h^2) + 2f'(x)o(h^2). \end{aligned} \quad (31)$$

However, the quantification of this risk function, in other words, the determination of optimal h , must be carried out over the entire interval $[a, b]$; therefore, the integral is applied only to the nonnegligible part of $\text{EQM}(x, h)$:

$$\begin{aligned} \text{IntegratedRisk}(h) &= R(h) = \int_a^b \text{EQM}(x, h) dx \\ &= \int_a^b \left(\left(hf'(x) \right)^2 + \frac{f(x)}{nh} + \frac{f'(x)}{n} \right) dx \\ &= \int_a^b \left(hf'(x) \right)^2 dx + \frac{1}{nh} + \frac{f(b) - f(a)}{n}. \end{aligned} \quad (32)$$

The optimal h corresponds to finding

$$\min(R(h)). \quad (33)$$

This leads us to look for the argument of the null derivative of $R(h)$.

$$\begin{aligned} \frac{\partial(R(h))}{\partial h} &= \frac{\partial \left(\int_a^b \left(hf'(x) \right)^2 dx + (1/nh) + (f(b) - f(a)/n) \right)}{\partial h}, \\ \frac{\partial(R)}{\partial h}(h_{\text{opt}}) &= 2h_{\text{opt}} \int_a^b \left(f'(x) \right)^2 dx - \frac{1}{nh_{\text{opt}}^2} = 0, \\ &= 2h_{\text{opt}}^3 \int_a^b \left(f'(x) \right)^2 dx - \frac{1}{n} = 0. \end{aligned} \quad (34)$$

Therefore,

$$h_{\text{opt}} = \left(\frac{1}{2n \int_a^b \left(f'(x) \right)^2 dx} \right)^{1/3}. \quad (35)$$

However, we do not have the distribution of f , to deduce $\int_a^b \left(f'(x) \right)^2 dx$.

To do this, it will be necessary to go through the “cross-validation” estimator.

(3) The “cross-validation” formula of the estimation

It seems that, according to the formulation of Integrate-Risk, the calculation of $\int_a^b (f'(x))^2 dx$ turns out to be difficult in practice; hence, the idea is to calculate $de R(h) - \int f^2(t)dt$; this amounts to calculating the formula “cross-validation” (see Figure 1):

$$\begin{aligned}
 CV(h) &= IR(h) - \int f^2(t)dt = \int E(\hat{f}_H(x) - f(x))^2 dx - \int f^2(x)dx \\
 &= \int E(\hat{f}_H^2(x))dx - 2 \int E(\hat{f}_H(x))f(x)dx \\
 &\quad + \int f^2(x)dx - \int f^2(x)dx \\
 &= \int (E^2(\hat{f}_H(x)) + \text{var}(\hat{f}_H(x)))dx \\
 &\quad - 2 \int E(\hat{f}_H(x))f(x)dx \\
 &= \sum_K \left(\int_{c_k} (E^2(\hat{f}_H(x)) + \text{var}(\hat{f}_H(x)))dx \right. \\
 &\quad \left. - 2 \int_{c_k} E(\hat{f}_H(x))f(x)dx \right). \tag{36}
 \end{aligned}$$

If we estimate $E(\hat{f}_H(x)) = p_k/h$, we will have

$$\begin{aligned}
 \check{CV}(h) &= \sum_K \left(\int_{c_k} \left(\left(\frac{p_k}{h} \right)^2 + \frac{p_k(1-p_k)}{nh^2} \right) dx - 2 \int_{c_k} \frac{p_k}{h} f(x) dx \right) \\
 &= \sum_K \left(\frac{p_k^2}{h} + \frac{p_k(1-p_k)}{nh} - 2 \frac{p_k^2}{h} \right) = \sum_K \left(\frac{p_k(1-p_k)}{nh} - \frac{p_k^2}{h} \right) \\
 &= \sum_K \left(\frac{p_k(1-p_k)}{nh} - \frac{p_k^2}{h} \right) = \frac{1}{nh} - \frac{\sum_K p_k^2(1+n)}{nh}. \tag{37}
 \end{aligned}$$

Finally, we have

$$h_{CV} = \text{argmin}(\check{CV}(h)). \tag{38}$$

4.2. The Kernel Method. The kernel method is applied if the continuity of the distribution function is ensured (the distribution function belongs to class functions C^1). Let us consider in a first step the following estimate: $F_M(x) = 1/M \sum_{i=1}^M I(X_i \leq x)$, $\forall X_i \in [a, b]$ and the X_i are $i, i, d\forall i = 1, \dots, M$.

Let $X'_i = I(X_i \leq x)$.

The strong law of large numbers indicates that $(1/M) \sum$

$X'_i \xrightarrow{M \rightarrow +\infty} E(X')$ almost certainly/ $E(X')$ calculation:

$$\begin{aligned}
 E(X') &= E(I(X \leq x)) = \int_{-\infty}^x 1 * f(t)dt = \int_{-\infty}^x f(t)dt \tag{39} \\
 &= P(X \leq x) = F(x).
 \end{aligned}$$

And finally, $F_M(x) \xrightarrow{M \rightarrow +\infty} F(x)$ almost certainly.

The derivative of the function $F(x)$ is given by

$$f(x) = \frac{F(x+h) - F(x-h)}{2h}, \quad \text{with } h \rightarrow 0^+. \tag{40}$$

This implies

$$f_M^R(x) = \frac{F_M(x+h) - F_M(x-h)}{2h}. \tag{41}$$

Such an estimator is known as the Rosenblatt estimator.

$$\begin{aligned}
 f_M^R(x) &= \frac{F_M(x+h) - F_M(x-h)}{2h} = \frac{F_M(x-h < X < x-h)}{2h} \\
 &= \frac{\sum_{i=1}^M I(x-h < X_i < x-h)}{2hM} \\
 &= \frac{1}{2hM} \sum_{i=1}^M I(x-h < X_i < x-h) = \frac{1}{hM} \sum_{i=1}^M K_0\left(\frac{X_i - x}{h}\right). \tag{42}
 \end{aligned}$$

Generally, we put [8]

$$\hat{f}_K(x) = \frac{1}{hM} \sum_{i=1}^M K\left(\frac{X_i - x}{h}\right). \tag{43}$$

K is called the kernel of the estimate; generally, this kernel can be illustrated following different paces [9]:

- (i) The triangular kernel: $K(u) = (u+1)I_{[-1,0]} + (1-u)I_{[0,1]}(u)$
- (ii) The parabolic kernel: $K(u) = (3/4)(1-u^2)I_{[-1,1]}(u)$
- (iii) The Gaussian kernel: $K(u) = (1/\sqrt{2\pi}) \exp(-(u^2/2))$

(1) The quadratic risk of $\hat{f}_K(x)$

The mean square deviation as mentioned in the histogram approach is defined as follows:

$$\text{EQM}(x, h) = \text{bias}^2(\hat{f}_K(x)) + \text{variance}(\hat{f}_K(x)). \tag{44}$$

Definition 1. Let T be an interval in R and the pair $(\beta, L) \in R^{+2}$,

 FIGURE 1: Algorithm of the calculation of $\check{C}\check{V}$.

the class Hölder $\mathcal{H}(\beta, L)$ defined on T is the set of differentiable functions $l = \lfloor \beta \rfloor$ which satisfy

$$|f^l(x) - f^l(x')| \leq L|x - x'|^{\beta-l} \forall x, x' \in T. \quad (45)$$

Definition 2. Let $l \geq 1$ be an integer; we would say that $K : R \rightarrow R$ is a kernel function of order l [10] if the functions $v \rightarrow v^j K(u)$, $j = 0, \dots, l$, are integrable and which satisfy

$$\begin{aligned} \int K(v) dv &= 1, \\ \int v^j K(v) dv &= K_j, \forall j = 1, \dots, l. \end{aligned} \quad (46)$$

In the following, we will focus on the kernel function of order 2.

Definition 3 (Cauchy-Schwarz inequality). If f and g are two integral functions on T , then the absolute value of the integral of their product satisfies the following inequality:

$$\int_T f(x) * g(x) dx \leq \sqrt{\int_T f^2(x) dx * \int_T g^2(x) dx}. \quad (47)$$

Let us calculate the upper bound of the bias [11, 12]:

$$\begin{aligned} \text{bias}(\hat{f}_K(x)) &= E(\hat{f}_K(x)) - f(x) = \frac{1}{h} \int K\left(\frac{y-x}{h}\right) f(y) dy - f(x) \\ &= \int K(u) f(uh+x) du - f(x). \end{aligned} \quad (48)$$

Let $\wp(\beta, L) = \{f \in \mathcal{H}(\beta, L) / f \geq 0, \int f(y) dy = 1 \forall y \in R\}$.

Going through a Taylor expansion of order 2 of $p(uh+x)$, we have

$$f(uh+x) = f(x) + uh * f'(x) + \frac{(uh)^2}{2} f''(x + \tau uh), \quad (49)$$

with $0 < \tau < 1$.

$$\begin{aligned} \int K(u) f(uh+x) du &= \int K(u) f(x) du + \int K(u) uh * f'(x) du \\ &\quad + \int K(u) \frac{(uh)^2}{2} f''(x + \tau uh) du. \end{aligned} \quad (50)$$

However, we previously assumed that the kernel function

is of order 2, and therefore,

$$\begin{aligned}
 \int K(u)f(uh+x)du - p(x) &= \int K(u)\frac{(uh)^2}{2}f''(x+\tau uh)du \\
 &= \int K(u)\frac{(uh)^2}{2}f''(x+\tau uh)du \\
 &\quad - \int K(u)\frac{(uh)^2}{2}f''(x)du \\
 &\quad + \int K(u)\frac{(uh)^2}{2}f''(x)du \\
 &= \int K(u)\frac{(uh)^2}{2}(f''(x+\tau uh) - f''(x))du \\
 &\quad + \int K(u)\frac{(uh)^2}{2}p''(x)du.
 \end{aligned} \tag{51}$$

The hypothesis of $p \in \mathcal{P}(\beta, L)$ implies

$$\begin{aligned}
 K(u)(f''(x+\tau uh) - f''(x)) &\leq |K(u)| |f''(x+\tau uh) - f''(x)| \\
 &\leq L|K(u)| |\tau uh|^{\beta-2}.
 \end{aligned} \tag{52}$$

So

$$\text{bias}^2(\hat{f}_K(x)) \leq \left(\int |K(u)|L|\tau uh|^{\beta-2}du \right)^2 + \int K(u)\frac{(uh)^2}{2}f''(x)du. \tag{53}$$

Following Cauchy-Schwartz's theorem, we will have the following inequality:

$$\begin{aligned}
 \left(\int |K(u)|L|\tau uh|^{\beta-2}du \right)^2 &\leq \int |K(u)|^2du * \int (L|\tau uh|^{\beta-2})^2du \\
 &\leq \int |K(u)|^2du * L^2(h)^{2\beta-4} \int |u|^{2\beta-4}du.
 \end{aligned} \tag{54}$$

Knowing that the calculation of the expectation is done on the interval $[a, b]$: $u > 0$ if $y > x$ and vice versa.

$$\begin{aligned}
 \int_{\left[\frac{a-x}{h}, \frac{b-x}{h}\right]} |u|^{2\beta-4}du &= \int_{\left[\frac{a-x}{h}, 0\right]} (-u)^{2\beta-4}du + \int_{\left[0, \frac{b-x}{h}\right]} (u)^{2\beta-4}du, \\
 \int_{\left[\frac{a-x}{h}, \frac{b-x}{h}\right]} |u|^{2\beta-4}du &= -\frac{1}{2\beta-5} \left(\left(-\frac{a-x}{h} \right)^{2\beta-5} \right) \\
 &\quad + \frac{1}{2\beta-5} \left(\left(\frac{b-x}{h} \right)^{2\beta-5} \right).
 \end{aligned} \tag{55}$$

Finally,

$$\begin{aligned}
 \text{bias}^2(\hat{f}_K(x)) &\leq \int |K(u)|^2du * L^2(h)^{2\beta-4} \\
 &\quad * \left(-\frac{1}{2\beta-5} \left(\left(-\frac{a-x}{h} \right)^{2\beta-5} \right) + \frac{1}{2\beta-5} \right. \\
 &\quad \cdot \left. \left(\left(\frac{b-x}{h} \right)^{2\beta-5} \right) \right) + \int K(u)\frac{(uh)^2}{2}f''(x)du.
 \end{aligned} \tag{56}$$

Likewise, the variance of $\hat{f}_K(x)$ is calculated as follows:

$$\begin{aligned}
 \text{variance}(\hat{f}_K(x)) &= E(\hat{f}_K^2(x)) - E^2(\hat{f}_K(x)). \\
 \text{variance}(\hat{f}_K(x)) &\leq E(\hat{f}_K^2(x)) \leq \frac{1}{nh} \int K^2(u)f(x+hu)du.
 \end{aligned} \tag{57}$$

The maximum of a probability function = 1:

$$\text{variance}(\hat{f}_K(x)) \leq \frac{1}{nh} \int K^2(u)du. \tag{58}$$

Put $\int K^2(u)du = c_1$.

Finally,

$$\begin{aligned}
 \text{EQM}(x, h) &\leq c_1 \left(\frac{1}{nh} + L^2(h)^{2\beta-4} * \left(-\frac{1}{2\beta-5} \left(\left(-\frac{a-x}{h} \right)^{2\beta-5} \right) \right. \right. \\
 &\quad \left. \left. + \frac{1}{2\beta-5} \left(\left(\frac{b-x}{h} \right)^{2\beta-5} \right) \right) \right) + \int K(u)\frac{(uh)^2}{2}f''(x)du.
 \end{aligned} \tag{59}$$

Computing h_{opt} amounts to minimizing the upper bound of $\text{EQM}(x, h)$.

(2) The quadratic risk of $\hat{f}_K(x)$ integrated

Computing h_{opt} from $R(h)$ amounts to deducing it from the following integral:

$$\begin{aligned}
 \int \left[c_1 \left(\frac{1}{nh} + L^2(h)^{2\beta-4} * \left(-\frac{1}{2\beta-5} \left(\left(-\frac{a-x}{h} \right)^{2\beta-5} \right) \right. \right. \right. \right. \\
 \left. \left. \left. + \frac{1}{2\beta-5} \left(\left(\frac{b-x}{h} \right)^{2\beta-5} \right) \right) \right) + \int K(u)\frac{(uh)^2}{2}f''(x)du \right] dx,
 \end{aligned} \tag{60}$$

FIGURE 2: Algorithm of the calculation of strategies (J_i, J^*_i) .

FIGURE 3: Algorithm of the Pareto front.

because

$$\begin{aligned}
 R(h) &= \int \text{EQM}(x, h) dx \\
 &\leq \int \left[c_1 \left(\frac{1}{nh} + L^2(h)^{2\beta-4} * \left(-\frac{1}{2\beta-5} \left(\left(-\frac{a-x}{h} \right)^{2\beta-5} \right) \right. \right. \right. \\
 &\quad \left. \left. \left. + \frac{1}{2\beta-5} \left(\left(\frac{b-x}{h} \right)^{2\beta-5} \right) \right) \right) \right] + \int K(u) \frac{(uh)^2}{2} f''(x) du dx.
 \end{aligned} \tag{61}$$

(3) The “cross-validation” formula of the estimation

As mentioned previously, the cross-validation formula “CV(h)” of the estimation of distributions is defined by

$$CV(h) = IR(h) - \int f^2(x) dx, \tag{62}$$

```

> Shapiro.test(irrig1$irrig)

Shapiro-Wilk normality test

data:  irrig1$irrig
W = 0.95955, pvalue = 0.2839

> Shapiro.test(irrig1$non.irrig)

Shapiro-Wilk normality test

data:  irrig1$non.irrig
W = 0.97134, pvalue = 0.5566

```

FIGURE 4: Shapiro-Wilk test of the Gaussian law for the yields of irrigated and nonirrigated land follows a Gaussian law.

```

> summary(irrig1$non.irrig)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.150  1.280   2.000   2.114   2.910   4.280
> kernel

Call:
  density.default(x = irrig1$non.irrig, kernel = "Gaussian", window = kernel)

Data: irrig1$non.irrig (31 obs.);    Bandwidth 'bw' = 0.4725

      x              y
Min.  :-1.2674  Min.  :0.0003502
1st Qu.: 0.4738  1st Qu.:0.0229044
Median : 2.2150  Median :0.1170662
Mean   : 2.2150  Mean   :0.1434246
3rd Qu.: 3.9562  3rd Qu.:0.2622122
Max.   : 5.6974  Max.   :0.3396516

```

FIGURE 5: Calculation of the mean by the arithmetic mean method and by the Kernel method for the nonirrigated land yield $q(n)$.

```

> summary(irrig1$irrig)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.70  7.59    9.38    9.43  10.74   13.06
> kernel

Call:
  density.default(x = irrig1$irrig, kernel = "gaussian", window = kernel)

Data: irrig1$irrig (31 obs.);    Bandwidth 'bw' = 0.9419

      x              y
Min.   : 2.874  Min.   :0.0001661
1st Qu.: 6.127  1st Qu.:0.0124796
Median : 9.380  Median :0.0817346
Mean    : 9.380  Mean    :0.0767713
3rd Qu.:12.633  3rd Qu.:0.1398747
Max.    :15.886  Max.    :0.1515159

```

FIGURE 6: Calculation of the mean by the arithmetic mean method and by the Kernel method for the irrigated land yield $p(n)$.

```

> adf.test(budget, alternative = "stationary")

Augmented Dickey-Fuller Test

data: budget
Dickey-Fuller = -2.1996, Lag order = 2, pvalue = 0.4963
alternative hypothesis: stationary

> budgdiff = diff(budget, differences =2)
> adf.test(budgdiff, alternative = "stationary")

Augmented Dickey-Fuller Test

data: budgdiff
Dickey-Fuller = -4.0391, Lag order = 2, pvalue = 0.02189
alternative hypothesis: stationary

```

FIGURE 7: The increased Dickey-Fuller test before and after differentiations.

which implies that

Let us say

$$CV(h) = E\left(\int \hat{f}_K^2(x) dx\right) - 2E\left(\int \hat{f}_K(x)f(x) dx\right). \quad (63)$$

$$\hat{f}_K(x) = \hat{p}_{M,-i} = \frac{1}{(M-1)h} \sum_{j \neq i}^M K\left(\frac{X_j - x}{h}\right), \quad (64)$$

FIGURE 8: The partial autocorrelation function PACF of the differentiated series.

FIGURE 9: The autocorrelation function ACF of the differentiated series.

and show that

$$\hat{\beta} = \frac{1}{M} \sum \hat{p}_{M,-i} \quad (65)$$

is an unbiased estimator of $E(\int \hat{f}_K(x)f(x)dx)$.

This amounts to demonstrating that

$$\beta = E\left(\int \hat{f}_K(x)f(x)dx\right) = E\left(\frac{1}{M} \sum \hat{p}_{M,-i}\right) = E(\hat{\beta}). \quad (66)$$

Under the assumption that X_j are i, i, d , therefore we have

$$\begin{aligned} E\left(\frac{1}{M} \sum \hat{p}_{M,-i}\right) &= E(\hat{\beta}) = E(\hat{p}_{M,-1}) \\ &= E\left(\int \frac{1}{(M-1)h} \sum_{j \neq 1}^M K\left(\frac{X_1 - z}{h}\right) f(z) dz\right) \\ &= \int \left(\int \frac{1}{h} K\left(\frac{x-z}{h}\right) f(z) dz\right) f(x) dx. \end{aligned} \quad (67)$$

Likewise,

$$\begin{aligned} \beta &= E\left(\int \hat{f}_K(x)f(x)dx\right) \\ &= E\left(\int \frac{1}{(M-1)h} \sum_{j \neq 1}^M K\left(\frac{X_1 - z}{h}\right) f(z) dz\right) \\ &= \int \left(\int \frac{1}{h} K\left(\frac{x-z}{h}\right) f(z) dz\right) f(x) dx. \end{aligned}$$

```
> AICS(budget)
      [,1]      [,2]      [,3]
[1,] 236.3139 233.0066 233.8537
[2,] 231.9620 233.2609 232.4389
[3,] 233.6878 234.7520 234.2710
> min(AICS(budget))
[1] 231.962
```

FIGURE 10: AIC value matrix.

$$\widehat{CV}(h) = E\left(\int \hat{f}_K^2(x)dx\right) - 2\frac{1}{M} \sum \hat{p}_{M,-i}. \quad (68)$$

5. Theoretical Calculation of Strategies

Recall that the default risk $\Delta(n)$, for a given year n , “as discussed by Moiseev [13],” is calculated as follows:

$$\begin{aligned} \Delta(n) &= (p(n) * S(n) + q(n) * s(n) - \varphi(n)) - Q(n), \\ Q(k) &= \begin{cases} \min(P(n) - \varphi(n), G(n) - R(n-1)), & \text{if } P(n) \geq \varphi(n), \\ \max(P(n) - \varphi(n), -R(n-1)), & \text{if } P(n) < \varphi(n). \end{cases} \quad (\text{see (4)}) \end{aligned} \quad (69)$$

The nonparametric estimate of yield expectations will allow the calculation of strategies for the distribution of budgets between irrigation and the construction of silos over the next N years; these strategies are given by

$$J_1 = \max |E(\Delta(n))|_{1 \leq n \leq N}. \quad (70)$$

FIGURE 11: Forecast of the time series of investment budgets over the period (2017-2036).

	X	rendirrig	surfirrig	rendnonirrig	surfnonirrig	budgdirrig	budgsilo.n.1	besoin.n.	G.n..	R.n.1.
17	2016	7.020000	5.500000	3.0300000	3.049000	3421.000	905.0900	88.557360	550.1120	139.631640
18	2017	8.412624	5.548807	2.6936142	3.000193	3451.358	730.7370	137.133351	586.3156	98.922750
19	2018	10.557841	5.680749	1.9172333	2.868251	3533.426	1156.7117	12.684032	615.5451	16.550790
20	2019	8.550525	5.687998	3.4120471	2.861002	3537.934	983.8069	26.805830	661.8135	69.342310
21	2020	7.145503	5.767823	3.3655217	2.781177	3587.586	880.5051	29.947657	701.1658	100.933723
22	2021	6.925932	5.893405	2.9405817	2.655595	3665.698	1085.8495	185.631160	736.3860	121.560177
23	2022	8.360919	6.040980	1.9131831	2.508020	3757.489	943.2487	23.839571	779.8200	0.000000
24	2023	10.566845	6.062344	3.4687855	2.486656	3770.778	926.4644	87.208585	817.5500	31.466872
25	2024	8.689837	6.098096	0.2317701	2.450904	3793.016	1074.6093	43.385019	854.6085	16.943811
26	2025	8.134480	6.114380	3.7167466	2.434620	3803.145	1069.9316	81.830829	897.5929	27.118301
27	2026	9.045773	6.159555	3.8377105	2.389445	3831.243	1065.8745	118.955982	940.3902	4.073639
28	2027	8.785541	6.242225	1.9681452	2.306775	3882.664	1127.9842	41.040587	983.0252	0.000000
29	2028	7.077784	6.310201	1.7718580	2.238799	3924.945	1117.9796	17.367674	1028.1445	18.340809
30	2029	7.461217	6.627985	1.7224669	1.921015	4122.607	959.3411	98.222787	1072.8637	49.602208
31	2030	8.753841	6.654806	1.1396739	1.894194	4139.289	1027.6431	65.792778	1111.2373	4.141141
32	2031	9.618824	6.713577	3.7536811	1.835423	4175.845	1036.1047	14.720141	1152.3431	0.000000
33	2032	9.855054	6.858074	1.3627798	1.690926	4265.722	993.3387	25.847309	1193.7873	56.746168
34	2033	10.986247	6.997220	0.5807868	1.551780	4352.271	977.4605	51.118700	1233.5208	100.789906
35	2034	9.619804	7.061795	3.5567055	1.487205	4392.436	988.3384	47.008399	1272.6192	127.445640
36	2035	6.561602	7.242686	1.2938402	1.306314	4504.951	927.2726	78.358275	1312.1528	153.659871
37	2036	8.247729	7.252756	0.8594025	1.296244	4511.214	984.5115	8.296786	1349.2437	124.515380

Q.n.	delta	totalprod	totalbudg	irrig_pol2	nonirrig_pol2	prod_pollutionR(n-1)	pollution	Q_pollutiondelta_pollution
-40.708890	0.000000	47.84847	4326.090	6.177600	2.4240000	41.36758	111.745062	-47.189784
-82.371961	0.000000	54.76139	4182.095	6.730099	2.1548914	43.80911	64.555278	-64.555278
52.791520	0.000000	65.47555	4690.138	8.446273	1.5337866	52.38044	0.000000	39.696410
31.591413	0.000000	58.39724	4521.741	6.840420	2.7296377	46.71779	39.696410	19.911964
20.626454	0.000000	50.57411	4468.091	5.716403	2.6924173	40.45929	59.608374	10.511632
121.560177	-15.444664	48.62632	4751.547	5.540746	2.3524654	38.90106	70.120006	-70.120006
31.466872	0.000000	55.30644	4700.738	6.688735	1.5305465	44.24515	0.000000	20.405583
-14.523061	0.000000	72.68552	4697.242	8.453476	2.7750284	58.14842	20.405583	-20.405583
10.174490	0.000000	53.55951	4867.625	6.951870	0.1854161	42.84761	0.000000	0.000000
-23.044662	0.000000	58.78617	4873.076	6.507584	2.9733973	47.02893	0.000000	0.000000
-4.073639	-49.994408	64.88794	4897.118	7.236619	3.0701684	51.91035	0.000000	0.000000
18.340809	0.000000	59.38140	5010.648	7.028433	1.5745162	47.50512	0.000000	6.464529
31.261400	0.000000	48.62907	5042.924	5.662227	1.4174864	38.90326	6.464529	21.535585
-45.461067	0.000000	52.76172	5081.948	5.968974	1.3779736	42.20938	28.000114	-28.000114
-4.141141	-1.237762	60.41388	5166.932	7.003073	0.9117391	48.33110	0.000000	0.000000
56.746168	0.000000	71.46631	5211.950	7.695059	3.0029449	57.17305	0.000000	42.452906
44.043738	0.000000	69.89105	5259.061	7.884043	1.0902238	55.91284	42.452906	30.065529
26.655734	0.000000	77.77443	5329.731	8.788997	0.4646294	62.21955	72.518435	11.100847
26.214231	0.000000	73.22263	5380.775	7.695843	2.8453644	58.57810	83.619282	11.569705
-29.144491	0.000000	49.21378	5432.223	5.249281	1.0350722	39.37103	95.188987	-38.987248
52.635976	0.000000	60.93276	5495.726	6.598183	0.6875220	48.74621	56.201739	40.449424

FIGURE 12: An illustration (of Figure 1) of calculating the risk of default Δ^* "delta" and Δ^* "delta_pollution" for as mentioned in Figure 1 under the R software: the last iteration of simulation.

The project strategy can be calculated in another way:

$$J_2 = \sum_{n=1}^N |E(\Delta(n))| \text{ with } \Delta(n) \neq 0. \quad (71)$$

Strictly positive or strictly negative defaults do not have the same meaning. This leads us to calculate the budget allocation strategy using the following formula:

$$J_3 = \sum_{n=1}^N |E(\Delta(n))| \text{ with } \Delta(n) < 0. \quad (72)$$

The problem of optimizing the min-max comes down to choosing the optimal strategy for the allocation of budgets that minimizes the risk of default; this amounts to finding the following scalar minimum:

$$\min (J_1, J_2, J_3). \quad (73)$$

In the case where the pollution factor is considered, the new default risk values will be obtained from the formula below:

$$\Delta^*(n) = (p^*(n) * S(n) + q^*(n) * s(n) - \varphi(n)) - Q^*(n),$$

[illegible]

FIGURE 14: The outcome of Pareto optimization Figure 3 under the R software.

function of $\Delta(n)$ and that calculated as a function of $\Delta^*(n)$; this leads us to calculate:

$$\begin{aligned} J_1^* &= \max |E(\Delta^*(n))|_{1 \leq n \leq N}, \\ J_2^* &= \sum_{n=1}^N |E(\Delta^*(n))| \text{ with } \Delta^*(n) \neq 0, \\ J_3^* &= \sum_{n=1}^N |E(\Delta^*(n))| \text{ with } \Delta^*(n) < 0. \end{aligned} \quad (75)$$

Then, minimize the following vector:

$$\min (J_i, J_i^*) \quad \text{with } i = 1, 2, 3. \quad (76)$$

6. Multiobjective Pareto Optimization (Pareto Front)

A multiobjective optimization problem presents itself, in general, as follows [14, 15]:

$$P = \begin{cases} \min (F(X) = (F_1(X), F_2(X), \dots, F_n(X))), \\ \text{under constraints } g_j(x) \geq 0 \quad j = 1, \dots, m, \\ h_k(x), k = 1, \dots, l. \end{cases} \quad (77)$$

6.1. Dominance Principle and Pareto Optimum. For a given order, relation $<_p$ in a space of dimension p and let $Y = (Y_1, Y_2, \dots, Y_p)$ and let $Z = (Z_1, Z_2, \dots, Z_p)$ be two vectors in the decision space and $U(Y) = (U_1, U_2, \dots, U_m)$, $U(Z) = (U'_1, U'_2, \dots, U'_m)$. We will say that Y dominates [16] Z , and we write $Y <_p Z$, if and only if [17]

$$\begin{cases} \forall i \in 1, \dots, p, & U_i \leq U'_i, \\ \exists! j \in 1, \dots, p, & U_j < U'_j. \end{cases} \quad (78)$$

A solution $X' = (X'_1, X'_2, \dots, X'_n)$ of the multiobjective problem P is said to be Pareto optimal [18] if there is no point of the decision space $X^* = (X^*_1, X^*_2, \dots, X^*_n)$ which dominates X' such that $F_i(X^*) \leq F_i(X')$ for each i and j which verifies that $F_j(X^*) < F_j(X')$. All of these non-dominated points in the Pareto sense constitute the Pareto front [19]. In our example, the vector of irrigation prices over the next 20 years will be the vector of decisions and the vector of objectives is the vector (J_i, J_i^*) .

Notice that “yes” means the corresponding identity holds and “no” means the identity does not hold.

Once the strategies are simulated, we will have a data frame of two columns and of length $N_{\text{rows}_{\text{tab}}}$ of (J_i, J_i^*) (see Figure 2). Therefore and by applying the principle of dominance, the efficient strategies (Pareto Front strategies) are calculated by the algorithm of the Figure 3.

```
> Pareto=DF_sans[, -c(1,2)]
> colnames (stimule)[1]<-"X3"
> colnames (stimule)[2]<-"X4"
> library(dplyr)
> DF_sans=DF %>%
+   na.omit()
> DF_sans

> Pareto
      X3      X4
13  34.45495 80.11261
25  32.78667 80.78384
28  31.99396 81.57723
46  42.75440 75.66432
58  31.42880 81.91839
61  41.46395 75.77376
64  37.76186 77.79991
112 37.62379 79.04978
127 28.88922 83.48608
133 37.47183 79.51963
151 24.02549 89.05216
157 18.46402 89.35771
166 28.33349 85.04257
169 44.42350 75.28163
172 56.08805 74.49381
184 39.45260 76.59739
223 35.98916 79.57679
226 32.83340 80.47168
232 45.23063 75.12231
```

FIGURE 15: The optimal Pareto values of (J_i, J_i^*) simulated.

7. Empirical Part

Following a Shapiro-Wilk test [20] of the Gaussian law, the p value of the test ($>5\%$) indicates that the yields of irrigated and nonirrigated land follow a Gaussian law (Figure 4).

Under the assumption of the normal distribution of these returns, we choose to calculate their expectations through a nonparametric estimate of the density by the kernel method by the choice of a Gaussian kernel, through the following program in Figures 5 and 6.

Since the hypothesis of the normality of yields of nonirrigated land and those of irrigated land is verified, the median and the mean must be equal, given the Gaussian law is a symmetrical law.

Through the “summary” command, we notice that the arithmetic mean is not significant for estimating the expectation because it is different from the median.

By using the kernel estimate, the average obtained corresponds perfectly to the median; this means that the estimate of the density by this method turns out to be robust for the calculation of the expected returns (mean = median = $E(q) = 2,215$ tonnes/ha for nonirrigated land and mean = median = $E(p) = 9,38$ tonne/ha for irrigated land).

We assume that the pollution factor can reduce the yields of irrigated and nonirrigated land by 12% and 20%, respectively. The average of the returns by the kernel

FIGURE 16: Set of strategies (J_i, J_i^*) simulated 249 times and the Pareto front of the min-max problem; the x-axis represents J_i ; the y-axis represents J_i^* .

method makes it possible to deduce that $E(q^*(n)) = 1,77$ tonnes/ha and $E(p^*(n)) = 8,25$ tonnes/ha.

The yields $q(n)$ and $p(n)$ were simulated by the numerical “Box-Müller” method taking into account the intervals of the values of $q(n)$ and $p(n)$ according to the FAO site (from 0 to 5 tonnes/ha for nonirrigated land and more than 5 tonnes/ha for irrigated land).

The nonsimulated data series cover the period (area of irrigated land) 2001-2016 (source: FAO site); thereafter, the areas of irrigated land will be forecast over the period 2017-2036.

Total agricultural land = 8,549 million ha (to deduct the surface areas of nonirrigated land).

The total budget to be invested will be planned based on the ARIMA time series (p, d, q) [21], then will be distributed between irrigation and construction of silos; the total budget to be invested will be forecast [22] based on the ARIMA time series (p, d, q) , then will be distributed between irrigation and construction of silos.

If a process X_t follows an ARIMA (p, d, q) model, then it is represented by the following relation:

$$\phi(L)(1-L)^d X_t = \phi(L) Y_t = \Theta(L) \varepsilon_t, \quad (79)$$

such that

- (i) ε_t is white noise
- (ii) d is order of differentiation of the process X_t
- (iii) L is the delay operator
- (iv) $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$
- (v) $\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$

If X_t follows an ARIMA (p, d, q) process, so Y_t follows an ARIMA (p, q) process.

In practice, the series of budgets was not stationary following the increased Dickey-Fuller test (p value $> 5\%$) and therefore, the alternative hypothesis of stationarity is rejected.

After two differentiations, the series of investment budgets became stationary (p value $< 5\%$) as shown in the code in Figure 7.

To determine the degrees p and q of the series, we use the partial autocorrelation function PACF and the autocorrelation function ACF, respectively (Figures 8 and 9, respectively).

From the two graphs, we can first assume that $p = 2$ and $q = 1$ (after these values, the PACF and AFC functions practically cancel each other out).

To ensure the choice of the appropriate model, we proceed to build the AIC matrix (Akaike Information Criterion) of the ARIMA models $(i, 2, j)$ with i ranging from 1 to 3 and j ranging from 1 to 3 and retain the model which corresponds to the minimum AIC index (Figure 10).

The AIC matrix confirms that the model to be retained is ARIMA $(2, 2, 1)$.

The forecast of this series under the software over the period 2017-2036 is represented by the graph in Figure 11.

The construction of the budgets allocated to irrigation and the construction of silos over the period 2001-2016 is based, respectively, on the unit prices $c_y = 622\$/ha$ and $c_x = 25\$/tonne$; then, their forecasts (2017-2036) will be a distribution simulation of the series of investment budgets between irrigation and the construction of silos by digital simulation methods (we go through 4 distribution simulations of the total investment budget).

The product quantity requirements for the year over the period 2017-2036 will be simulated; after the calculation of the “delta” faults under the software, we move on to calculate the (J_i, J_i^*) with $i = 1, 2, 3$ over the period 2017-2036.

An example of calculation of Δ^* and Δ for the $N_{rows_{tab}} = 249$; the number of simulation of (J_i, J_i^*) with $i = 1, 2, 3$ is

$N_{\text{simule}} = 249/3 = 83$ (tab, $N_{\text{rows}_{\text{tab}}}$ and N_{simule} are mentioned in Figure 1), (Figure 12).

The table of the values (tab) of the strategies $(J_i, J^*_i) = (X3, X4)$ simulated (in each iteration of simulation, we obtain a distribution of the total investment budget between the construction of the silos and the irrigation over 20 future years) follows Figure 2 (Figure 13).

The strategies belonging to the so-called efficient Pareto front calculated following Figure 1 are presented as follows (the strategies which do not belong to the front register a missing value NA as indicating the algorithm) (Figure 14).

After eliminating the missing values, NA was obtained by the Pareto front algorithm (Figure 15).

By going through the program encoded on R [23], we can plot the graphs the strategies obtained following Figure 1 and the Pareto front following Figure 3 (Figure 16).

8. Conclusion

Therefore, each point belonging to the Pareto front corresponds to its vector of irrigation and construction of silos amounts and it is up to the decision-maker to choose the distribution of the total investment budget between irrigation and the construction of silos, which gives a strategy belonging to the Pareto front and minimizing the agricultural risk defaults in the case of Bangladesh in future. In this article, the min-max tool was suggested to be an appropriate formula to determine strategies said “optimal” especially when we use the dimension rather than $d = 1$ and to give an exhaustive study, other mathematics principles are invited to apply non-parametric statistics, ARIMA time series, etc.

Data Availability

All data are available on the FAO website.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Aquastat (FAO), “Paper notes of Bangladesh,” http://www.fao.org/nr/water/aquastat/countries_regions/bgd/BGD-CP_eng.pdf.
- [2] “Data imported from the FAO irrigated land site,” <http://www.fao.org/faostat/en/#data/RL>.
- [3] “Unit prices for irrigation in Bangladesh,” http://www.fao.org/nr/water/aquastat/countries_regions/BGD/.
- [4] “Unit prices for the construction of silos,” <http://www.fao.org/3/t1838e/T1838E1c.htm>.
- [5] “The yield ranges of irrigated and non-irrigated land,” <http://www.fao.org/3/Y3918E/y3918e10.htm>.
- [6] A.-S. Dalalyan, *Statistique Avancée: Méthodes Non-Paramétriques*, Ecole Centrale de Paris, 2016.
- [7] M. Péchaud, “Comparaisons de fonctions,” 2008-2009, <http://mickaelpechaud.free.fr/cours/comparaison.pdf>.
- [8] A.-B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Science+Media Business, New York, USA, 2009.
- [9] B.-E. Hansen, *Lecture Notes on Nonparametrics*, Springer, University of Wisconsin, 2009.
- [10] F. Ferraty and P. Vieu, *Non parametric functional data analysis: theory and practice*, Springer Science+Media Business, New York, USA, 2006.
- [11] M.-P. Wand and M. C. Jones, *Kernel Smoothing, Monographs on Statistics and Applied Probability 60*, Springer Science+Media Business, New York, USA, 1995.
- [12] A. G. Bors and N. Nasios, “Kernel bandwidth estimation for nonparametric modeling,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 6, pp. 1543–1555, 2009.
- [13] N. N. Moiseev, “Mathematical Problems of System Analysis,” Nauka, Moscow, 1981.
- [14] B. Huang, P. Fery, L. Xue, and Y. Wang, “Seeking the Pareto front for multi-objective spatial optimization problems,” *International Journal of Geographical Information Science*, vol. 22, no. 5, pp. 507–526, 2008.
- [15] T. Grabener and A. Berro, *Optimisation multiobjectif discrète par propagation de contrainte*, Toulouse University, Toulouse, France, 2008.
- [16] A. Chinchuluun, P. M. Pardalos, and L. Pitsoulis, “Pareto Optimality, Game Theory and Equilibria,” in *Springer Optimization and Its Applications*, New York, USA, Springer, 2008.
- [17] A. Benki, *Méthodes efficaces de capture de front de Pareto en conception mécanique multicritère: applications industrielles*, Université Nice Sophia Antipolis, 2014, HAL archives-ouvertes.fr.
- [18] J. Dipama, *Optimisation Multi-Objectif Des Systèmes Énergétiques*, de Montréal University, Montreal, Canada, 2010.
- [19] G. Guillopé, *Optimisation sous contrainte*, Laboratoire de mathématiques Jean Leray, Département de mathématiques, Nantes University, Nantes, France, 2015-2016.
- [20] R. Rakotomalala, *Tests de normalité Techniques empiriques et tests statistiques, Version 2,0*, Lumière Lyon 2 University, Lyon, France, 2011.
- [21] A. Charpentier, *Cours de Séries Temporelles Théorie et Applications (volume 2)*, Paris Dauphine University, DESS Actuariat & DESS Mathématiques de la Décision, 2018.
- [22] B. El Griny, G. Kouaiba, M. Imegri, Y. El Qalli, and D. Mentagui, “Les Modèles ARCH et GARCH: Application au CAC40,” *Iejmae*, vol. 8, no. 3, pp. 36–51, 2017.
- [23] “R project website,” <http://www.r-project.org>.