Iranian Journal of Fuzzy Systems

Volume 19, Number 5, (2022), pp. 1-15

Original paper

Possibilistic max-mean dispersion problem

R. Saltos[1] and W. Pedrycz[2]

[1] *Facultad de Innovación y Tecnología, Universidad Del Pacífico, Km. 7.5 Vía a la Costa, Guayaquil - Ecuador.*
[2] *Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6R 2V4, Canada.*

ramiro.saltos@upacifico.edu.ec, wpedrycz@ualberta.ca

Abstract

The Max-Mean Dispersion Problem comprises selecting a subset of elements from a set while maximising the average of a measure of dispersion. We usually compute this measure of dispersion based on some distance metric between the elements of the candidate set. However, in real-world applications, this measure of dispersion could be ill-defined or vague. For example, consider the problem of building a team to execute some kinds of project. We can see the dispersion between the members of the team as the team chemistry, so the coach is interested in maximising this chemistry. In this example, it would be very difficult to compute exactly a dispersion measure for each candidate. This is due to the lack of information about how well the candidates worked together in the past. To cope with imprecise or vague information, in this paper, we propose three mixed integer linear programming models based on possibility, necessity, and credibility measures. To the best of our knowledge, this is the first approach which explicitly considers this type of uncertainty in this optimisation problem.

Keywords: Fuzzy sets, possibility theory, uncertainty, max-mean dispersion problem.

1 Introduction

Consider the problem of a marketing manager who wishes to carry out a survey to get information about a new product that is under development. The manager would like to have a representative sample of individuals with a wide spectrum of characteristics. This sample must be as diverse as possible to reflect the preferences of the entire population. The manager has a set of potential candidates to be selected as part of the sample. This problem is known in the literature as the maximum diversity problem [17]. Mathematically, it could be described as the problem of selecting a subset of a predetermined number of elements that is as diverse as possible. According to [14], some of the most popular applications of diversity problems arise in the academic context, where forming diverse groups of students in business schools, or peer reviewers in scientific journals, has a big relevance to get better scholarly results. Other range of real-world applications are plant breeding, product design, workforce management, among others [23].

For this problem, different mathematical models and solutions approaches have been proposed in the literature [16, 17, 23, 27–30, 32, 33, 35]. All of them require a measure of diversity which is typically computed as the distance between the elements belonging to the reference set and assume it to be precisely known [34]. However, real-world problems are affected by a variety of uncertainties, like the variability of stock markets or vagueness in measurements because of the lack of precise information about the object that is observed. The first type of uncertainty had been deeply studied by the theory of probability, while the second one had been handled by the fuzzy sets and possibility theory [12].

In most settings modelled by the maximum diversity problem, it is difficult to get an accurate diversity measure between two individuals. In the example of building a group of students, it will be necessary historical data about how well the students worked together in similarly past environments. However, this data could be not available. For these

Corresponding Author: R. Saltos
Received: August 2021; Revised: July 2022; Accepted: August 2022.
https://doi.org/10.22111/IJFS.2022.7153

scenarios, possibility theory together with fuzzy set theory [42,44] provides a complete set of sound mathematical tools that allows incorporating imprecise or vague information into mathematical models, see for example references [13,21,31].

Two of the most popular models in diversity maximisation are the Max-Sum Dispersion Problem [23] and the Max-Mean Dispersion Problem [35]. The former looks to maximise the sum of the dispersions between the elements selected from the candidate set, while the latter aims to maximise the average of this dispersion. The major difference between them is in the first one the number of elements to be selected is fixed while in the second one not. Since the Max-Mean Dispersion Problem is more general than the Max-Sum Dispersion Problem, we will focus attention on the first problem, although the approaches developed can be easily adapted for the Max-Sum Problem.

To the best of our knowledge, no prior attempt has been carried out to consider either probabilistic or possibilistic uncertainty in diversity maximisation problems, in particular, for the max-mean dispersion problem, see for example [4,6,8,15,18,19,24,25,39]. In this paper, we propose to represent the measure of diversity between two elements of the reference set by the means of fuzzy numbers. Next, using their membership functions as possibility distributions, we develop three possibilistic programming models based on possibility, necessity and credibility measures to include the vague or imprecise data into the classical model. Finally, we run some computational experiments to make comparisons between the solutions got from these models and discuss their differences. We summarize the main contributions of the paper as follows:

- We incorporate uncertainty into the max-mean dispersion problem using three possibilistic integer programming models: possibility-based, necessity-based, and credibility-based.
- We extend the max-mean dispersion problem to the possibilistic version using a chance-constrained similar approach without losing its linearity.
- The decision-maker can choose their attitude towards the risk by varying the parameter α in the possibilistic models.
- The possibilistic chance-constrained approach is relatively easy to follow. Consequently, practitioners can apply it with similar optimisation problems.

The rest of the paper is arranged as: In Section 2, we review the classical Max-Mean Dispersion model together with the fundamentals of possibility theory and possibilistic mathematical programming. Section 3 develops the possibilistic versions of the max-mean dispersion problem while in Section 4 the computational experiments are presented. Finally, Section 5 concludes this paper and presents hints for future developments.

2 Related works

In this section, we first present the mathematical models related to the max-mean dispersion problem, then the fundamentals of possibility theory are introduced. Finally, we explain the chance constrained possibilistic approach for mathematical programming problems and the recent advances in the max-mean dispersion problem.

2.1 The max-mean dispersion problem

Suppose given is a set N of n elements, and $d_{i,j}$ is the inter-element distance between any two elements i and j. The *Max-Mean Dispersion Problem (Max-Mean DP)* is to select a subset $M \subseteq N$ of cardinality $m = |M|$ such that the following equity function of $d_{i,j}$ is maximised [35]:

$$z(M) = \frac{\sum_{i,j \in M} d_{i,j}}{m}. \tag{1}$$

Note the value of m in (1) is not fixed a priori. It must be determined from the solution of the optimisation model, i.e., m is a decision variable. The model can be formulated as the following fractional programming problem [35]:

$$Max \quad z = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{i,j} x_i x_j}{\sum_{i=1}^{n} x_i}$$

$$s.t.$$

$$\sum_{i=1}^{n} x_i \geq 1$$

$$x_i \in \{0,1\} \qquad \forall i = 1,\ldots,n$$

where x_i takes the value of 1 if the element $i \in N$ is included in the subset M, 0 otherwise. The linear mixed 0-1 programming reformulation is [35]:

$$Max \quad z = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{i,j}t_{i,j} \tag{2}$$

s.t.

$$y - v_i \leq 1 - x_i \qquad \forall i = 1,\ldots,n \tag{3}$$
$$v_i \leq y \qquad \forall i = 1,\ldots,n \tag{4}$$
$$v_i \leq x_i \qquad \forall i = 1,\ldots,n \tag{5}$$
$$y - t_{i,j} \leq 2 - x_i - x_j \qquad \forall 1 \leq i < j \leq n \tag{6}$$
$$t_{i,j} \leq y \qquad \forall 1 \leq i < j \leq n \tag{7}$$
$$t_{i,j} \leq x_i \qquad \forall 1 \leq i < j \leq n \tag{8}$$
$$t_{i,j} \leq x_j \qquad \forall 1 \leq i < j \leq n \tag{9}$$
$$\sum_{i=1}^{n} x_i \geq 1 \tag{10}$$
$$\sum_{i=1}^{n} v_i = 1 \tag{11}$$
$$v_i, t_{i,j}, y \geq 0 \qquad \forall i,j = 1,\ldots,n \tag{12}$$
$$x_i \in \{0,1\} \qquad \forall i = 1,\ldots,n \tag{13}$$

Prokopyev [35] proved that the Max-Mean DP problem is strongly NP-Hard if the distances (diversity measure) take both positive and negative values. To solve this optimisation problem, many heuristics and meta-heuristics had been proposed in the literature [3,4,18,19,29,35]. All these approaches assume the diversity measure between the candidate elements is exactly known. In Section 3 we drop this assumption and explicitly introduce uncertainty in the values of the parameter $d_{i,j}$ by the means of fuzzy set theory.

2.2 Fundamentals of possibility theory

Possibility theory is one of the current uncertainty theories devoted to the handling of incomplete information. It is like probability theory because it is based on set functions, but differs from the latter by the use of a pair of dual set functions called possibility and necessity measures instead of only one [10]. These core functions are defined:

Definition 2.1 (Possibility Measure [11])**.** *A possibility measure on a reference set U is a set function Π from $\mathcal{P}(U)$, the set of subsets of U, to the unit interval $[0,1]$, such that:*

(1) $\Pi(\emptyset) = 0$,

(2) $\Pi(U) = 1$,

(3) $\forall A, B \in \mathcal{P}(U), \quad \Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$.

Definition 2.2 (Necessity Measure [11,38])**.** *A necessity measure is a set function $\mathcal{N} : \mathcal{P}(U) \to [0,1]$ such that:*

(1) $\mathcal{N}(\emptyset) = 0$,

(2) $\mathcal{N}(U) = 1$,

(3) $\forall A, B \in \mathcal{P}(U), \quad \mathcal{N}(A \cap B) = \min\{\mathcal{N}(A), \mathcal{N}(B)\}$.

Definition 2.3 (Possibility Distribution [11])**.** *Given a normalised fuzzy set F, the quantity $\Pi_F(A)$ derived from the membership function μ_F by:*

$$\Pi_F(A) = \sup_{u \in A}\{\mu_F(u)\}, \qquad \forall A \subseteq U, \tag{14}$$

defines a possibility measure. μ_F is the possibility distribution underlying Π_F and is often denoted as π_F.

Equation (14) can be interpreted as the possibility of realising event A when the possibility of elementary events is known [11, 12]. The relation between possibility and necessity measures is given by Proposition 2.4 [11].

Proposition 2.4. *Let $\bar{A}$ be the complementary set of A, and Π be a possibility measure. Then the set function $\mathcal{N}$ defined by:*

$$\mathcal{N}(A) = 1 - \Pi(\bar{A}), \tag{15}$$

a necessity measure for all $A \subseteq U$. Similarly, given a necessity measure $\mathcal{N}$, then:

$$\Pi(A) = 1 - \mathcal{N}(\bar{A}), \tag{16}$$

defines a possibility measure for all $A \subseteq U$.

Equations (15) and (16) motivate the name "necessity measure" (also called certainty measure [43]): the grade of necessity of an event A is the grade of impossibility of the opposite event.

An important question in fuzzy numbers is how a real number may be placed in relation to a fuzzy number. [11], based on possibility theory, derived a complete set of comparison indices to solve this issue. The indices are given:

Let X be a real fuzzy number with membership function $\mu(x) : \mathcal{R} \to [0, 1]$. Let b be any real number. Then:

$$Pos(X \leq b) = \Pi\left[(-\infty, b]\right] = \sup_{x \leq b}\{\mu(x)\}, \tag{17}$$

$$Pos(X > b) = \Pi\left[(b, +\infty)\right] = \sup_{x > b}\{\mu(x)\}, \tag{18}$$

$$Nec(X \leq b) = \mathcal{N}\left[(-\infty, b]\right] = \inf_{x \leq b}\{1 - \mu(x)\}, \tag{19}$$

$$Nec(X > b) = \mathcal{N}\left[(b, +\infty)\right] = \inf_{x > b}\{1 - \mu(x)\}. \tag{20}$$

Equations (17) and (18) describe the grade of possibility of the events $x \leq b$ and $x > b$ when x is restricted by the fuzzy number X, respectively. Similarly, equations (19) and (20) measure the grade of necessity of the events $x \leq b$ and $x > b$ when x is restricted by the fuzzy number X, respectively. When dealing with two or more fuzzy numbers, the above indices can be extended as proved in reference [11]. Possibility and necessity measures have a lot of other interesting properties which can be easily addressed in references [9, 11, 12, 43].

2.3 Possibilistic linear programming

One of the extensive number of applications that fuzzy set theory has is mathematical programming. Since the introduction of fuzzy mathematical programming by [1], it has been continuously growing as the main alternative to handle the uncertainties derived by the ambiguity and vagueness of parameters and targets in optimisation problems [13,20,31]. Basically, there are two major branches in fuzzy mathematical programming:

- *Flexible programming* treats the decision making problem under fuzzy goals and constraints. These fuzzy goals and constraints represents the flexibility of the target values of objectives functions and the elasticity of the constraints [20, 37].
- *Possibilistic programming* deals with ambiguous coefficients in objective functions and constraints but does not handle the flexibility in goals and constraints [20]. Usually, these ambiguous parameters in optimisation models are represented by fuzzy numbers. Since each fuzzy number has associated a membership function which can be used as a possibility distribution [12, 43, 44], the ambiguous parameters in mathematical programming problems are usually called possibilistic variables [20].

Based on the idea of chance constraint programming [5, 26], in a possibilistic programming model we assume that the objective function or constraints with possibilistic parameters will hold with at least some possibility/necessity confidence level α. In general, the problem:

$$\begin{aligned} &Max \quad \tilde{Z} = \tilde{c}'x \\ &s.t. \\ &\qquad\quad x \in X, \end{aligned} \tag{21}$$

where X is the feasible region, could be interpreted as:

$$Max \quad z^*$$

$$s.t. \tag{22}$$
$$Pos/Nec\left(\tilde{Z} \geq z^*\right) \geq \alpha$$
$$x \in X.$$

The auxiliary deterministic mathematical model to solve the above problem depends on the form of the membership functions. For a complete set of detailed examples, we refer the reader to [20].

2.4 Recent advances for max-mean dispersion problem

Since its introduction by Prokopyev2009 [35], the max-mean dispersion problem has attracted a lot of attention from the scientific community given its wide range of applications including: sentiment analysis, pollution control, capital investment, genetic engineering, web pages ranking, community mining in a signed social network, trusting networks, among others [4, 25]. Given the max-mean dispersion problem is a NP-Hard combinatorial optimisation problem [35], the main focus of academia has been the development of efficient solution approaches for the problem, mainly by using heuristic or metaheuristics algorithms. The Table 1 shows the recent advances made to the problem and the type of uncertainty considered.

Table 1: Recent Advances for Max-Mean DP

Author	**Year**	**Method**	**Uncertainty**
Prokopyev [35]	2009	Heuristics	Deterministic
Marti [29]	2013	GRASP-PR	Deterministic
Carrasco [4]	2015	Tabu Search	Deterministic
Gortazar [18]	2015	VNS	Deterministic
Della [8]	2016	MIP + PR	Deterministic
Lai [24]	2016	Tabu Search - Memetic Algorith	Deterministic
Brimberg [3]	2017	Reinforcement Learning + VNS	Deterministic
Garraffa [15]	2017	Semidefinite Programming	Deterministic
Cura [6]	2019	Ant System Algorithm	Deterministic
Nijimbere [33]	2019	Distribution Algorithm	Deterministic
Song [39]	2019	Multi-wave Algorithm	Deterministic
Lai [25]	2020	Evolutionary Computation	Deterministic
Nijimbere [32]	2020	Tabu Search + Reinforcement Learning	Deterministic
Gu [19]	2021	Reinforcement Learning	Deterministic
Martinez [30]	2021	GRASP + Tabu Search	Deterministic

As could be appreciated in Table 1, almost all recent advances made for max-mean dispersion problem focus on the development of heuristic and metaheuristics algorithms to efficiently solve the problem. All of them have in common the parameters of the problem, i.e., the dispersion measure $d_{i,j}$, is known with certainty. To the best of our knowledge, no effort has been carried to explicity introduce uncertainty in the parameters of the problem. In this paper, we cope with this issue by using the fuzzy set theory and the possibility theory.

3 Possibilistic max-mean dispersion problem

As pointed out in Section 1, real-world problems are affected by a variety of uncertainties. In particular, in the context of diversity maximisation, the problem of how to compute the grade of diversity between two individuals is an ill-posed one since it is vague and ambiguous. We say the problem is vague because it is difficult to clearly define the meaning of "diversity" according to the context being modeled. Assuming this issue was solved by a consensus of experts, the ambiguity appears when the measurement of this diversity is carried out and will be reflected in the values of the parameters $d_{i,j}$. To explicitly include the uncertainty affecting the parameters $d_{i,j}$ in the max-mean dispersion problem[1], we treat them as possibilistic variables represented by real fuzzy numbers, and develop three fuzzy chance constrained models based on possibility, necessity and credibility measures.

[1]The same approach can be used for the weighted max-mean dispersion problem.

For sake of simplicity, and without lost of generality, assume that the diversity measure $d_{i,j}$ between individual i and j is a real normalized triangular fuzzy number $\tilde{d}_{i,j} = \left(d^m_{i,j}, l_{i,j}, r_{i,j}\right)$ where $l_{i,j}, r_{i,j} > 0$ are the left and right spread from the most plausible value $d^m_{i,j}$, respectively. We want to solve the following optimisation problem:

$$
\begin{aligned}
&Max \quad \tilde{Z} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \tilde{d}_{i,j} t_{i,j}, \\
&s.t. \\
&\qquad (3)-(13)
\end{aligned}
\tag{23}
$$

Since decision variables $t_{i,j} \geq 0$, $\tilde{Z}$ is also a triangular fuzzy number defined by $\tilde{Z} = (Z_m, Z_l, Z_r)$ where:

$$
\begin{aligned}
Z_m &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d^m_{i,j} t_{i,j}, \\
Z_l &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} l_{i,j} t_{i,j}, \\
Z_r &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} r_{i,j} t_{i,j}.
\end{aligned}
$$

Similarly to stochastic programming approaches, the direct optimisation of the objective function (23) makes no sense since its coefficients are ambiguous. In this case, it is necessary to introduce a specific interpretation to clarify it [20]. We propose three approaches to solve this issue based on possibility theory and chance constrained programming [5,26].

3.1 Possibility-based model

Similarly to stochastic chance constraint programming, assume the decision maker wants a solution that achieves the possibility of $\tilde{Z}$ being greater or equal than z^* is at least some confidence level α. This implies to introduce the following possibility-based chance constraint into the model:

$$
\begin{aligned}
&Max \quad z^* \\
&s.t. \\
&\qquad Pos\left(\tilde{Z} \geq z^*\right) \geq \alpha. \\
&\qquad (3)-(13)
\end{aligned}
\tag{24}
$$

where $\alpha \in [0,1]$ is the confidence level that the decision maker wants to achieve. Based on this interpretation and using the identities reviewed in section 2.2, it is easy to obtain that:

$$
Pos\left(\tilde{Z} \geq z^*\right) = \begin{cases} 1 & \text{if } z^* \leq Z_m \\ \frac{Z_m + Z_r - z^*}{Z_r} & \text{if } z^* \in (Z_m, Z_m + Z_r] \\ 0 & \text{if } z^* > Z_m + Z_r \end{cases}
\tag{25}
$$

From equation (47) we have that:

$$
\begin{aligned}
&\frac{Z_m + Z_r - z^*}{Z_r} \geq \alpha, \\
&Z_m + Z_r - z^* \geq \alpha Z_r, \\
&Z_m + (1-\alpha) Z_r \geq z^*.
\end{aligned}
$$

So, model (24) is equivalent to:

$$
\begin{aligned}
&Max \quad z^* \\
&s.t. \\
&\qquad Z_m + (1-\alpha) Z_r \geq z^*.
\end{aligned}
\tag{26}
$$

$$(3) - (13)$$

Finally, based on these preliminary steps, we got the following auxiliary deterministic mixed integer programming model for the possibility-based max-mean dispersion problem (PB):

$$\begin{aligned} Max \quad & \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[d_{i,j}^{m} + (1-\alpha) r_{i,j} \right] t_{i,j}, \\ s.t. \quad & \\ & (3) - (13) \end{aligned} \tag{27}$$

The possibility-based max-mean dispersion model is ideal for risk-loving decision-makers, i.e., this model represents an optimistic scenario where the selected individual will cooperate fruitfully. However, since we do not know the exact diversity measure between the individuals, the decision-maker takes some risks with this model because it will select more individuals for the group, as we will see in Section 4. The decision-maker can adjust their risk attitude by varying the α parameter.

3.2 Necessity-based model

Consider now the decision maker needs to achieve some level of certainty about the event $\tilde{Z} \geq z^*$. For this case, the necessity measure must be introduced in the model instead of the possibility measure. The necessity model is:

$$\begin{aligned} Max \quad & z^* \\ s.t. \quad & \\ & Nec\left(\tilde{Z} \geq z^*\right) \geq \alpha, \\ & (3) - (13) \end{aligned} \tag{28}$$

where $\alpha \in [0, 1]$ is the certainty level that the decision maker wants to achieve. Recall the fact that $\mathcal{N}(A) = 1 - \Pi(\bar{A})$, then:

$$\begin{aligned} Nec\left(\tilde{Z} \geq z^*\right) &\geq \alpha, \\ 1 - Pos\left(\tilde{Z} < z^*\right) &\geq \alpha, \\ Pos\left(\tilde{Z} < z^*\right) &\leq 1 - \alpha. \end{aligned}$$

Using the indices defined in section 2.2, we obtain:

$$Pos\left(\tilde{Z} < z^*\right) = \begin{cases} 1 & \text{if } z^* \geq Z_m \\ \frac{z^* - Z_m + Z_l}{Z_l} & \text{if } z^* \in (Z_m, Z_m - Z_l] \\ 0 & \text{if } z^* < Z_m - Z_l \end{cases} \tag{29}$$

From equation (29) we have:

$$\begin{aligned} \frac{z^* - Z_m + Z_l}{Z_l} &\leq 1 - \alpha, \\ z^* - Z_m + Z_l &\leq (1-\alpha) Z_l, \\ Z_m - \alpha Z_l &\geq z^*. \end{aligned}$$

Next, the model (28) is equivalent to:

$$\begin{aligned} Max \quad & z^* \\ s.t. \quad & \\ & Z_m - \alpha Z_l \geq z^*, \\ & (3) - (13) \end{aligned} \tag{30}$$

Finally, the auxiliary deterministic mixed integer programming model for the necessity-based max-mean dispersion problem (NB) is:

$$\begin{aligned} &Max \quad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^{m}-\alpha l_{i,j}\right]t_{i,j},\\ &s.t.\\ &\qquad (3)-(13) \end{aligned} \tag{31}$$

The necessity-based max-mean dispersion model is ideal for risk-averse decision-makers since this model represents a pessimistic scenario where the selected individual will probably not cooperate fruitfully. For this reason, this model will select fewer individuals for the group, as we will see in Section 4. The decision-maker can adjust their risk attitude by varying the α parameter.

3.3 Credibilistic-based model

Consider the decision maker wants to achieve some trade-off between the solutions obtained from the possibilistic and necessity models [41]. This can be done by the means of a credibilistic model which is given as follows:

$$\begin{aligned} &Max \quad z^{*}\\ &s.t.\\ &\qquad Cr\left(\tilde{Z}\geq z^{*}\right)\geq\alpha,\\ &\qquad (3)-(13) \end{aligned} \tag{32}$$

where $\alpha\in[0,1]$ is the certainty level that the decision maker wants to achieve. Recall the fact that $Cr(A)=\frac{1}{2}\left[\Pi(A)+\mathcal{N}(A)\right]$, then:

$$\begin{aligned} Cr\left(\tilde{Z}\geq z^{*}\right)&\geq\alpha,\\ \frac{1}{2}\left[Pos\left(\tilde{Z}\geq z^{*}\right)+Nec\left(\tilde{Z}\geq z^{*}\right)\right]&\geq\alpha,\\ Pos\left(\tilde{Z}\geq z^{*}\right)+Nec\left(\tilde{Z}\geq z^{*}\right)&\geq 2\alpha. \end{aligned}$$

Combining equations (47) and (29) we can obtain the following formula:

$$2\cdot Cr\left(\tilde{Z}\geq z^{*}\right)=\begin{cases} 2 & \text{if } z^{*}\leq Z_m+Z_r,\\ 2-\frac{z^{*}-Z_m+Z_l}{Z_l} & \text{if } z^{*}\in(Z_m-Z_l,Z_m],\\ \frac{Z_m+Z_r-z^{*}}{Z_r} & \text{if } z^{*}\in(Z_m,Z_m+Z_r],\\ 0 & \text{if } z^{*}>Z_m+Z_r. \end{cases} \tag{33}$$

From equation (33) we have:

$$\begin{aligned} 2-\frac{z^{*}-Z_m+Z_l}{Z_l}&\geq 2\alpha,\\ 2Z_l-z^{*}+Z_m-Z_l&\geq 2\alpha Z_l,\\ Z_m+(1-2\alpha)Z_l&\geq z^{*}, \end{aligned}$$

and

$$\begin{aligned} \frac{Z_m+Z_r-z^{*}}{Z_r}&\geq 2\alpha,\\ Z_m+Z_r-z^{*}&\geq 2\alpha Z_r,\\ Z_m+(1-2\alpha)Z_r&\geq z^{*}. \end{aligned}$$

Next, the model (32) is equivalent to:

$$Max \quad z^{*},$$

$$
\begin{aligned}
&s.t. \\
&\qquad Z_m + (1-2\alpha)Z_l \geq z^*, \\
&\qquad Z_m + (1-2\alpha)Z_r \geq z^*, \\
&\qquad (3)-(13)
\end{aligned}
\tag{34}
$$

Finally, the auxiliary deterministic mixed integer programming model for the credibilistic-based max-mean dispersion problem (CB) is:

$$
\begin{aligned}
&Max \quad z^*, \\
&s.t. \\
&\qquad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^m + (1-2\alpha)l_{i,j}\right]t_{i,j} \geq z^*, \\
&\qquad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^m + (1-2\alpha)r_{i,j}\right]t_{i,j} \geq z^*. \\
&\qquad (3)-(13)
\end{aligned}
\tag{35}
$$

The credibility-based max-mean dispersion model is ideal for risk-neutral decision-makers since this model allows a trade-off between the pessimistic and the optimistic scenarios. If $\alpha > 0.5$, we have an optimistic solution, otherwise we have a pessimistic one. The decision-maker can adjust their risk attitude by varying the α parameter.

4 Computational experiments and discussion

To test the performance of the models proposed in this paper, we used benchmark data sets available at `tinyurl.com/55hmcrcf`. There are two type of instance data sets with 20 individuals each of them: Type I data sets were randomly generated from a continuous uniform distribution between -10 and 10, while Type II data sets were randomly generated from a non-convex continuous uniform distribution in $[-10,-5] \cup [5,10]$.

The latter instances reflect the polarization that occurs when people get together to work in groups and there is no room for indifference [29]. Additionally, three fuzzy models were derived to contrast their solutions with those proposed in this paper. These models can be obtained using some common ranking indices [2, 7, 22, 36, 40].

In this setting, we used Yager First Index, Yager Third Index and Center of Gravity. The comparison with other ranking methods or defuzzification strategies is out of the scope of this paper.

Table 2 and Table 3 display the optimal solution of all models tested. The first column shows the instance number, columns $\boldsymbol{Z_c}, \boldsymbol{Z_{fi}}, \boldsymbol{Z_{ti}}, \boldsymbol{Z_{cg}}, \boldsymbol{Z_p}, \boldsymbol{Z_n}, \boldsymbol{Z_{cr}}$ show the optimal value of the objective function for the deterministic model, the Yager First Index (YFI) model, the Yager Third Index (YTI) model, the Center of Gravity (CG) model, the possibility-based (PB) model, the necessity-based (NB) model and the credibilistic-based (CB) model, respectively.

Similarly, columns $\boldsymbol{m_c}, \boldsymbol{m_{fi}}, \boldsymbol{m_{ti}}, \boldsymbol{m_{cg}}$, $\boldsymbol{m_p}, \boldsymbol{m_n}, \boldsymbol{m_{cr}}$ output the optimal number of individuals to be selected to maximise the dispersion (diversity) between them for all the models in the same order mentioned above. For the possibility-based model, the necessity-based model and the credibilistic-based model the decision maker's confidence level was set at 0.9. Other values of the confidence level α were also tried and the results are available in the supplementary material.

Table 2: Results for Type I Instances

Inst.	Z_c	m_c	Defuzzification Approach Z_{fi}	m_{fi}	Z_{ti}	m_{ti}	Z_{cg}	m_{cg}	Possibilistic Approach Z_p	m_p	Z_n	m_n	Z_{cr}	m_{cr}
1	13.9	7	14.7	7	14.5	7	10.7	4	14.7	7	9.0	7	20.7	11
2	13.6	5	13.8	5	13.8	5	12.0	4	14.2	6	9.4	4	19.1	6
3	11.8	7	11.8	7	11.8	7	10.9	4	12.5	7	7.8	4	19.9	9
4	17.5	8	17.4	8	17.5	8	12.7	6	18.5	8	9.7	6	26.1	15
5	16.0	8	16.5	8	16.4	8	13.4	5	16.9	8	9.5	5	23.3	15
6	14.6	11	14.4	11	14.5	11	12.9	3	15.8	11	9.2	5	26.5	18
7	14.9	9	14.2	8	14.3	10	11.9	4	15.9	11	8.4	4	24.2	11
8	14.5	7	14.6	7	14.6	7	12.0	5	15.4	12	9.1	5	26.1	15
9	14.0	6	14.3	6	14.2	6	13.2	5	14.6	6	10.0	5	24.3	15
10	13.4	6	13.4	6	13.4	6	10.7	3	14.2	7	7.8	5	21.0	12
Mean	14.4	7.4	14.5	7.3	14.5	7.5	12.0	4.3	15.3	8.3	9.0	5.0	23.1	12.7
SD	1.5	1.7	1.5	1.6	1.5	1.8	1.0	0.9	1.7	2.2	0.8	0.9	2.8	3.6

In Table 2 we can observe that YFI, YTI and PB models tends to select the same number of individuals as in the deterministic model. This makes sense since both, YFI and YTI models give more relevance to the most plausible value of the fuzzy number, i.e, the value its membership degree is one, while the PB model reflects an optimistic scenario for the decision maker's preferences which is equivalent to assume the most plausible value of the fuzzy number will occur. On the other hand, the CG and NB models tend to select a lower number of individuals than the deterministic model. The latter model reflects a pessimistic scenario and the decision maker's necessity of having a high degree of certainty in the quality of this selection. Finally, the worst performance is for the CB model. It tends to select almost 75% of the whole data set which in real practice will imply a high cost solution, although the performance of this model could be improved decreasing the confidence level α. When $\alpha = 0.5$, the CB model outputs the solution of the deterministic model.

Table 3: Results for Type II Instances

Inst.	Z_c	m_c	Defuzzification Approach Z_{fi}	m_{fi}	Z_{ti}	m_{ti}	Z_{cg}	m_{cg}	Possibilistic Approach Z_p	m_p	Z_n	m_n	Z_{cr}	m_{cr}
1	18.9	8	18.7	8	18.7	8	16.7	6	19.9	10	13.3	5	29.9	12
2	17.8	7	17.9	7	17.8	7	15.9	6	18.6	7	12.0	6	26.2	13
3	18.1	7	18.1	7	18.1	7	16.0	4	18.8	7	11.8	7	25.3	12
4	17.8	10	17.4	8	17.5	10	17.0	5	18.8	10	12.1	5	25.9	10
5	16.3	5	16.6	7	16.5	7	17.4	5	16.8	7	12.5	5	24.2	13
6	17.6	6	17.7	6	17.7	6	15.9	6	18.3	6	12.1	6	23.6	11
7	18.9	6	19.1	6	19.1	6	17.5	6	19.6	6	13.4	6	26.2	11
8	21.9	8	21.6	9	21.7	9	17.6	8	22.9	9	13.3	8	30.7	14
9	19.8	8	19.3	10	19.5	10	18.5	6	20.8	10	12.5	7	28.8	11
10	22.6	10	21.8	10	22.0	10	18.1	8	23.6	10	13.2	8	31.1	11
Mean	19.0	7.5	18.8	7.8	18.9	8.0	17.1	6.0	19.8	8.2	12.6	6.3	27.2	11.8
SD	2.0	1.6	1.7	1.5	1.8	1.6	0.9	1.2	2.1	1.8	0.6	1.2	2.7	1.2

Type II instances do not allow room for indifference between individuals, i.e., individuals can not state there is not importance in working together with other particular individual. In this setting, YFI, YTI, CG and NB models perform similarly as in Type I instances. However, the PB models tends to select more individuals than the deterministic model, reflecting a more plausible optimistic scenario. On the other hand, the CB model improves its performance decreasing the percentage of selected individual to almost 55% of the whole data set.

To contrast there is a statistical significant difference between the solutions of the models tested, we perform a one-way analysis of variance for the value of the objective function (Z) and the number of individuals chosen (m) using the type of the model as a categorical factor. The p-value of all hypothesis tests were less than the standard signigicant level 0.05. Therefore, there is a significant difference between the solutions of all models for both variables (Table 4 and

Table 4: Analysis of Variance for Type I Instances

	Variable	**df**	**SumSq**	**MeanSq**	**F-value**	**p-value**
Type of Model	Z	6	959.0	159.83	71.11	$2e^{-16}$
Residuals		59	132.6	2.25		
Type of Model	m	6	409.4	68.23	26.89	$3.47e^{-15}$
Residuals		59	149.7	2.54		

Table 5: Analysis of Variance for Type II Instances

	Variable	**df**	**SumSq**	**MeanSq**	**F-value**	**p-value**
Type of Model	Z	6	1120.1	186.68	56.77	$2e^{-16}$
Residuals		63	207.2	3.29		
Type of Model	m	6	216.4	36.06	16.78	$1.86e^{-11}$
Residuals		63	135.4	2.15		

5). This imply models provide different solutions that will fit better according to the environment faced by the decision maker. However, no matter the environment, uncertainty should be included in these mathematical models to reflect closely the reality.

5 Conclusions

In this paper three novel fuzzy mixed integer linear programming models for the max-mean dispersion problem were introduced. The numerical results indicate the potential these models has to incorporate vague or imprecise data when modeling real-world problems. Each of them represent an optimistic, pessimistic and a trade-off scenario, respectively, in which the decision maker will have a set of different options to deal with uncertainty based on his risk profile and past experience. In summary, a list of the main contributions of this paper is presented.

The three models of max-mean dispersion problem incorporate epistemic uncertainty in the selection of a set of individuals which maximises a diversity measure. To the best of our knowledge, this is the first attempt which explicitly takes uncertainty into account in this kind of problems. At the same time, the use of triangular or gaussian fuzzy numbers maintain the linear nature of the classical model, which in fact, is a NP-Hard combinatorial optimisation problem, so our approach does not increase the computational complexity of the problem. Since all three models need the confidence level α to be fixed a priori, for each value of this parameter, the decision maker will possibly obtain a different solution which probably fit better for the scenario faced. This fact gives to the decision maker some degree of flexibility than classical approaches.

More research has to be done in order to better understand the potential provided by fuzzy set theory to model the uncertainty that naturally appears in some real-world optimisation problems, specially those in which the human factor is the main source of the data. As future work, a multi-objective version of this model will be proposed which improves the performance of the credibilistic-based model.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62076182 (W. Pedrycz). The authors are grateful to the anonymous reviewers who contributed to improving the quality of the original paper.

Appendix 1: Derivatives models of max-mean dispersion problem

In case the possibilistic parameters $d_{i,j}$ will be represented by gaussian fuzzy numbers, it can be easily check that the auxiliary deterministic counterpart of models (24), (28) and (32) are:

$$\begin{aligned} &Max \quad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^{m}+\sqrt{-\ln(\alpha)}\sigma_{i,j}\right]t_{i,j},\\ &s.t.\\ &\qquad (3)-(13) \end{aligned} \tag{36}$$

$$\begin{aligned} &Max \quad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^{m}-\sqrt{-\ln(1-\alpha)}\sigma_{i,j}\right]t_{i,j},\\ &s.t.\\ &\qquad (3)-(13) \end{aligned} \tag{37}$$

$$\begin{aligned} &Max \quad z^{*}\\ &s.t.\\ &\qquad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^{m}+\sqrt{-\ln(2\alpha)}\sigma_{i,j}\right]t_{i,j}\geq z^{*},\\ &\qquad \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[d_{i,j}^{m}+\sqrt{-\ln(2-2\alpha)}\sigma_{i,j}\right]t_{i,j}\leq z^{*},\\ &\qquad (3)-(13) \end{aligned} \tag{38}$$

where $d_{i,j}^{m}$ is the center of the gaussian fuzzy number while $\sigma_{i,j}>0$ is the desviation from its center. It is important to note that α must be in the interval $(0, 0.5]$ in model (38) to avoid non-real numbers. One of the main advantages of the possibilistic chance constrained model over the stochastic one is the former does not loose its linearity as usually happens with the latter [20].

Appendix 2: Aplication of the methodology to generic binary optimization models

Consider the next generic binary optimization model:

$$\max \ \tilde{Z}=\sum_{i=1}^{n}\tilde{c}_i x_i, \tag{39}$$

$$s.t. \tag{40}$$

$$\sum_{i=1}^{n}a_{i,j}x_i\leq b_j \qquad \forall j=1,2,\ldots,m, \tag{41}$$

$$x_i\in\{0,1\} \qquad \forall i=1,2,\ldots,n. \tag{42}$$

where $a_{i,j}\in\mathcal{R}$ and $b_j\in\mathcal{R}$. Without lost of generality, let $\tilde{c}_i=(c_i^m,l_i,r_i)$ be a triangular fuzzy number where $l_i,r_i>0$ are the left and right spread from the most plausible value c_i^m, respectively. Since the decision variables x_i are binary, $\tilde{Z}$ is also a triangular fuzzy number defined by $\tilde{Z}=(Z_m,Z_l,Z_r)$ where:

$$Z_m=\sum_{i=1}^{n}c_i^m x_i, \tag{43}$$

$$Z_l = \sum_{i=1}^{n} l_i x_i, \tag{44}$$

$$Z_r = \sum_{i=1}^{n} r_i x_i. \tag{45}$$

Let suppose the decision maker wants a solution where the possibility of $\tilde{Z}$ being greater or equal than z^* is at least some confidence level α. This implies to introduce the following possibility-based chance constraint into the generic model (39)-(42):

$$\begin{aligned} & \max \ z^*, && (46)\\ & s.t. \\ & Pos\left(\tilde{Z} \geq z^*\right) \geq \alpha, \\ & (41) - (42) \end{aligned}$$

where $\alpha \in [0, 1]$ is the confidence level that the decision maker wants to achieve. Using the identities presented in section 2.2 of the manuscript, we got:

$$Pos\left(\tilde{Z} \geq z^*\right) = \begin{cases} 1 & \text{if } z^* \leq Z_m \\ \frac{Z_m + Z_r - z^*}{Z_r} & \text{if } z^* \in (Z_m, Z_m + Z_r] \\ 0 & \text{if } z^* > Z_m + Z_r \end{cases} \tag{47}$$

From equation (47) we have that:

$$\begin{aligned} & \frac{Z_m + Z_r - z^*}{Z_r} \geq \alpha, \\ & Z_m + Z_r - z^* \geq \alpha Z_r, \\ & Z_m + (1 - \alpha) Z_r \geq z^*. \end{aligned}$$

Then, the model (46) is equivalent to:

$$\begin{aligned} \max \quad & z^*, && (48)\\ s.t. \quad & \\ & Z_m + (1 - \alpha) Z_r \geq z^*, \\ & (41) - (42) \end{aligned}$$

Finally, the auxiliary deterministic binary programming model is:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \left[c_i^m + (1 - \alpha) r_i\right] x_i, && (49)\\ s.t. \quad & && (50) \end{aligned}$$

The same approach can be used to derive the remaining models. As can be seen, we can apply the methodology to any binary optimization problem since the uncertainty does not affect the feasible region but only the coefficients of the objective function.

References

[1] R. E. Bellman, L. A. Zadeh, *Decision-making in a fuzzy environment*, Management Science, **17**(4) (1970), B-141.

[2] G. Bortolan, R. Degani, *A review of some methods for ranking fuzzy subsets*, Fuzzy Sets and Systems, **15**(1) (1985), 1-19.

[3] J. Brimberg, N. Mladenović, R. Todosijević, D. Urošević, *Less is more: Solving the max-mean diversity problem with variable neighborhood search*, Information Sciences, **382** (2017), 179-200.

[4] R. Carrasco, A. Pham, M. Gallego, F. Gortázar, R. Martí, A. Duarte, *Tabu search for the max-mean dispersion problem*, Knowledge-Based Systems, **85** (2015), 256-264.

[5] A. Charnes, W. W. Cooper, *Chance-constrained programming*, Management Science, **6**(1) (1959), 73-79.

[6] T. Cura, M. Ozdemir, *A new use of the ant system algorithm for the max-mean dispersion problem*, Computers and Industrial Engineering, **135** (2019), 628-642.

[7] S. K. De, I. Beg, *Triangular dense fuzzy sets and new defuzzification methods*, Journal of Intelligent and Fuzzy Systems, **31**(1) (2016), 469-477.

[8] F. Della Croce, M. Garraffa, F. Salassa, *A hybrid three-phase approach for the max-mean dispersion problem*, Computers and Operations Research, **71** (2016), 16-22.

[9] D. Dubois, *Possibility theory and statistical reasoning*, Computational Statistics and Data Analysis, **51**(1) (2006), 47-69.

[10] D. J. Dubois, H. Prade, *Fuzzy sets and systems: Theory and applications*, **144**, Academic Press, 1980.

[11] D. Dubois, H. Prade, *Ranking fuzzy numbers in the setting of possibility theory*, Information Sciences, **30**(3) (1983), 183-224.

[12] D. Dubois, H. Prade, *Possibility theory: An approach to computerized processing of uncertainty*, Springer Science and Business Media, 1988.

[13] A. Ebrahimnejad, J. L. Verdegay, *A survey on models and methods for solving fuzzy linear programming problems*, In Fuzzy Logic in Its 50th Year, Springer, 2016, pages 327-368.

[14] M. Gallego, M. Laguna, R. Martí, A. Duarte, *Tabu search with strategic oscillation for the maximally diverse grouping problem*, Journal of the Operational Research Society, **64**(5) (2013), 724-734.

[15] M. Garraffa, F. Della Croce, F. Salassa, *An exact semidefinite programming approach for the max-mean dispersion problem*, Journal of Combinatorial Optimization, **34**(1) (2017), 71-93.

[16] J. B. Ghosh, *Computational aspects of the maximum diversity problem*, Operations Research Letters, **19**(4) (1996), 175-181.

[17] F. Glover, C. C. Kuo, K. S. Dhir, *Heuristic algorithms for the maximum diversity problem*, Journal of Information and Optimization Sciences, **19**(1) (1998), 109-132.

[18] F. Gortázar, R. Carrasco, A. P. Trinh, M. Gallego, A. Duarte, *VNS variants for the max-mean dispersion problem*, Electronic Notes in Discrete Mathematics, **47** (2015), 253-260.

[19] X. Gu, S. Zhao, Y. Wang, *Reinforcement learning enhanced multi-neighborhood tabu search for the max-mean dispersion problem*, Discrete Optimization, (2021), page 100625.

[20] M. Inuiguchi, J. Ramık, *Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem*, Fuzzy Sets and Systems, **111**(1) (2000), 3-28.

[21] S. Kikuchi, P. Chakroborty, *Place of possibility theory in transportation analysis*, Transportation Research Part B: Methodological, **40**(8) (2006), 595-615.

[22] H. Kumar, *Some recent defuzzification methods*, In Theoretical and Practical Advancements for Fuzzy System Integration, (2017), 31-48.

[23] C. C. Kuo, F. Glover, K. S. Dhir, *Analyzing and modeling the maximum diversity problem by zero-one programming*, Decision Sciences, **24**(6) (1993), 1171-1185.

[24] X. Lai, J. K. Hao, *A tabu search based memetic algorithm for the max-mean dispersion problem*, Computers and Operations Research, **72** (2016), 118-127.

[25] X. Lai, J. K. Hao, F. Glover, *A study of two evolutionary/tabu search approaches for the generalized max-mean dispersion problem*, Expert Systems with Applications, **139** (2020), 112856.

[26] B. Liu, K. Iwamura, *Chance constrained programming with fuzzy parameters*, Fuzzy Sets and Systems, **94**(2) (1998), 227-237.

[27] R. Martí, M. Gallego, A. Duarte, *A branch and bound algorithm for the maximum diversity problem*, European Journal of Operational Research, **200**(1) (2010), 36-44.

[28] R. Martí, M. Gallego, A. Duarte, E. G. Pardo, *Heuristics and metaheuristics for the maximum diversity problem*, Journal of Heuristics, **19**(4) (2013), 591-615.

[29] R. Martí, F. Sandoya, *GRASP and path relinking for the equitable dispersion problem*, Computers and Operations Research, **40**(12) (2013), 3091-3099.

[30] A. Martínez-Gavara, T. Corberán, R. Martí, *Grasp and tabu search for the generalized dispersion problem*, Expert Systems with Applications, **173** (2021), 114703.

[31] M. J. Naderi, M. S. Pishvaee, S. A. Torabi, *Applications of fuzzy mathematical programming approaches in supply chain planning problems*, In Fuzzy Logic in Its 50th Year, Springer, (2016), 369-402.

[32] D. Nijimbere, S. Zhao, X. Gu, M. O. Esangbedo, N. Dominique, *Tabu search guided by reinforcement learning for the max-mean dispersion problem*, Journal of Industrial and Management Optimization, (2020), 3223-3246.

[33] D. Nijimbere, S. Zhao, H. Liu, B. Peng, A. Zhang, *A hybrid metaheuristic of integrating estimation of distribution algorithm with tabu search for the max-mean dispersion problem*, Mathematical Problems in Engineering, **2019** (2019), 1-16.

[34] F. Parreño, R. Alvarez-Valdés, R. Martí, *Measuring diversity. A review and an empirical analysis*, European Journal of Operational Research, **289**(2) (2021), 515-532.

[35] O. A. Prokopyev, N. Kong, D. L. Martinez-Torres, *The equitable dispersion problem*, European Journal of Operational Research, **197**(1) (2009), 59-67.

[36] S. Roychowdhury, W. Pedrycz, *A survey of defuzzification strategies*, International Journal of Intelligent Systems, **16**(6) (2001), 679-695.

[37] M. Safi, H. Maleki, E. Zaeimazad, *A note on the zimmermann method for solving fuzzy linear programming problems*, Iranian Journal of Fuzzy Systems, **4**(2) (2007), 31-45.

[38] G. Shafer, et al, *A mathematical theory of evidence*, Princeton University Press Princeton, **1**, 1976.

[39] J. Song, Y. Wang, H. Wang, Q. Wu, A. P. Punnen, *An effective multi-wave algorithm for solving the max-mean dispersion problem*, Journal of Heuristics, **25**(4-5) (2019), 731-752.

[40] A. Talon, C. Curt, *Selection of appropriate defuzzification methods: Application to the assessment of damperformance*, Expert Systems with Applications, **70** (2017), 160-174.

[41] X. Yang, *Some properties for fuzzy chance constrained programming*, Iranian Journal of Fuzzy Systems, **8**(4) (2011), 1-8.

[42] L. A. Zadeh, *Fuzzy sets*, Information and Control, **8**(3) (1965), 338-353.

[43] L. A. Zadeh, *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets and Systems, **100** (1999), 9-34.

[44] H. J. Zimmermann, *Fuzzy set theory and its applications*, Kluwer Academic Publishers, 4 Edition, 2001.