Decision support system using flexible query and reliability assessment - Application to biodegradable and biosourced packaging design

Sébastien Destercke, Patrice Buche, Brigitte Charnomordic, Valérie V. Guillard

Decision Support System using Flexible Query and Reliability Assessment - Application to Biodegradable and Biosourced Packaging Design

Sébastien Destercke
CNRS, JRU Heudiasyc
Rue Personne de Roberval
60200 Compiègne, France
Email: destercke@hds.utc.fr

Patrice Buche
INRA, JRU IATE, LIRMM
2 Place Viala
34060 Montpellier, France
Email: buche@supagro.inra.fr

Brigitte Charnomordic
INRA, JRU MISTEA
2 Place Viala
34060 Montpellier, France
Email: bch@supagro.inra.fr

Valérie Guillard
UM2, JRU IATE
2 Place Viala
34060 Montpellier, France
Email:guillard@univ-montp2.fr

***Abstract*—In decision tasks, imprecision and uncertainty can arise from many sources such as data uncertainty, data reliability, or the necessity to use intermediate (non-fully reliable) models. User preferences may also be included into the decision process, making it even more complex. This is particularly true for processes involving biological materials. Soft computing methods have the potential to be the kingpin of specialized software that can be integrated in decision support systems intended to solve the mentioned issues.**

This work presents three such methods and their implementation: *(i)*-simulation module with uncertain inputs and interval analysis, *(ii)*-reliability module allowing to rank candidates according to expert criteria, and *(iii)*-bipolar flexible querying module, that makes the difference between constraints and wishes in a data base query. A decision support system architecture is proposed and illustrated on a case study, dealing with biodegradable and biosourced packaging design.

keywords: software, data reliability, flexible querying, uncertainty simulation

I. Introduction

Product design is a complex task that usually requires several steps and the examination of numerous factors. In practice, design problems have to face multiple sources of uncertainty and imprecision. Those arise from the need to cope with flexible preferences, uncertain and non-totally reliable data, conflicting information.... Decision support systems should therefore be based on generic methods and reusable software modules to face such uncertainties, while being instantiated to meet the particular needs of a particular application.

In this paper, three aspects in which flexible or soft modelling is needed will be considered: *(i)*-uncertainty arising from models and uncertain data, *(ii)*-uncertainty arising from non-total reliability of data and *(iii)*-graduality needed to model user preferences. We will focus on a case of biosourced product design.

When using mathematical models or data, there is a need to take account of the uncertainty in data measurements and/or in the model. The problem is even more severe when dealing with cases where Biology is involved, such as biosourced food packaging design for fresh respiring produces, because of the product intrinsic high variability. Furthermore, when mathematical models are available in Biology, they are often partial models validated in specific experimental conditions at a given scale. It is difficult to evaluate the model sensitivity to biological parameters, as simultaneous variability of many biological parameters needs to be taken into account. Simulation steps can be necessary to evaluate the model output for imprecise input data, and the imprecision needs to be properly propagated through the model [1].

In nowadays applications, it also often happens that data are gathered through the web or imported from (electronic) documents. In this case, estimating data reliability may be a critical issue (especially in scientific environment), especially if data are collected automatically. There are few works considering this issue with experimental or technical data. Yet some work [2] proposes to consider the estimation of data reliability from a set of customizable criteria by the means of evidence theory and introduces a way to rank the query results in function of their reliability.

Finally, once uncertainty coming from the data has been modelled, another challenge is to model user preferences to interrogate the decision support system. User preferences can possess many different aspects, such as being gradual or bipolar, that classical crisp preferences cannot capture properly. It is therefore important, to help the product designers, to model these different aspects. In particular, recent works have emphasized the need to differentiate negative preferences (corresponding to constraints) from positive ones (corresponding to wishes). The notions of bipolar preferences and of bipolar information, which have received increasing attention in the past few years [3], [4], can help to model negative and positive preferences in order to better model user preferences.

For all these reasons, when dealing with imprecise data, multiple sources of information of various reliability, and flexible preferences, it is important to take all these elements into account, in order to have more informed and better decisions. Soft computing techniques are yet largely ignored in such domains as product design, even if they can be useful in complicated cases of multi-criteria decision making (MCDM). To make soft computing methods more widely known, it is important to develop reusable software using innovative methods, and especially free software to help dissemination. In this paper, we propose a software architecture based on several soft computing-based modules, most of them open source.

The first module handles simulation and interval analysis, the second one deals with the topic of data reliability, and the third one consider flexible query with a bipolar approach. The paper is organized as follows. The next section discusses the state of the art for the various topics involved: flexible query, simulation with imprecise variables, and source reliability. The proposed architecture is presented in Section III, and illustrated in Section IV with a case study on biodegradable and biosourced packaging design. Finally, Section V gives some conclusions and perspectives.

II. State of the art and need for specialized software

This section reviews previous works concerning the different aspects of the decision support system. We focus on aspects that are related to imprecision, uncertainty or flexibility, as these are the ones in which fuzzy sets and related models are instrumented. Although ideas coming from previous works can be adapted to the problem treated in this paper, we will argue that most of the time this adaptation is not straightforward, and requires to tailor the system and the methods to our current need.

a) Uncertainty simulation: parameters used in mathematical models are often only known up to some uncertainty, and this is particularly true for models in Life Sciences, where variability is high and measurements costly. In this case, it is necessary to propagate this uncertainty through the models, the outputs being then uncertain. Uncertainty propagation is a well-known problem, and many different techniques exists in the literature. Monte-carlo Sampling is a classical choice when probabilistic models can be assumed [5], while a mixture of interval analysis and Monte-Carlo sampling is usually used when uncertainty cannot be properly modelled by probabilities and is modelled by fuzzy sets, random sets, … [6].

b) Reliability analysis: when data are not produced by models or by experiments but are extracted from external documents, their reliability should be taken into account. Although there are many works proposing to deal with the problem of how to integrate and model information reliability in systems (e.g., [7] and Refs. therein), few of them actually deal with the problem of estimating this reliability. The existing proposals mainly compare information sources between them [8] or compare provided information to some reference value [7], two approaches that cannot be applied in our systems, as we are not in a learning process and have access to very few sources.

Another approach is to use meta-information to evaluate the data reliability. To our knowledge, there are no approach proposing to use such meta-information when analysing scientific data, but such ideas can be found in problems where the trust we can have in an agent has to be assessed. Such problems typically arise in web-based or network-based applications [9], [10], in which cases results interpretation is of lesser interest (as long as reliability assessment are relevant). There is therefore a need for methods to analyse scientific data reliability coming from few (sometimes one) sources and in absence of reference value. Such an approach is summarised in Section III-B.

c) Flexible querying: once uncertainty in data and models have been processed, the next step is to interrogate those data using user preferences in order to indicate what are the possible best designs. The advantage of using conjointly fuzzy sets to model gradual preferences and possibility distributions to model uncertainty in data has already been discussed by many authors [11], [12]. More recently, several authors [13], [14] have discussed the interest of using bipolar preferences in the fuzzy setting, differentiating between negative preferences (i.e., constraints, in which case objects not satisfying them are rejected) and positive preferences (i.e., wishes, in which cases objects not satisfying them are deemed less desirable but are not rejected). To our knowledge, the only other work mixing possibilistic modelling with bipolar fuzzy queries is [15], however they treat positive and negative preferences under a common umbrella, while we treat them in a separate way.

III. Proposed architecture

The architecture of the proposed decision support system (called in the following FQR-DSS for Flexible Query and Reliability assessment Decision Support System) is summarised in Fig. 1. First, (possibly uncertain) data are collected from the database and/or the web. Some of these data can be used to simulate mathematical models (Section III-A), and when necessary, the reliability of the available data is evaluated (Section III-B). The output of these processes are respectively fuzzy sets over the model output variables and bounds over reliability estimation. Second, user preferences on criteria are collected and shaped into a bipolar flexible query (Section III-C) that is used on the previously obtained data. Note that most of these tasks can be made in parallel, and that only the final querying requires all data.

The main originality of the proposal is to give practical examples and implementations of all of the proposed methods and to set forth a sequence of use cases. Wherever possible, the use of open source software or libraries has been preferred.

Fuzzy sets will be used to model preferences in queries and possibility distributions to model uncertainty. We denote a (normalised) fuzzy set membership function over a variable X taking values in $\mathscr{X}$ (here a finite set or a subset of the real line) by $\mu : \mathscr{X} \to [0,1]$. Possibility distributions are formally equivalent to fuzzy sets, but are used here to model uncertainty rather then preferences. To differentiate the two uses of membership functions, we will denote possibility distribution by π. This means that $\pi(x) = 1$ ($\pi(x) = 0$) means that x is totally plausible (not possible), while $\mu(x) = 1$

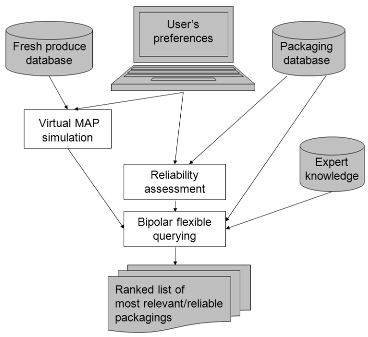

Fig. 1. Architecture of the decision support system (FQR-DSS).

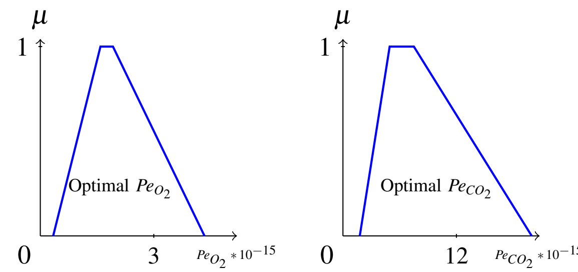

Fig. 2. Result of simulation module .

($\mu(x) = 0$) means that x is in one of the most preferred value (least preferred).

In the next sections, we will illustrate the modules on a sustainable package design problem for fresh fruits and vegetable. In such cases, the designer typically starts by specifying a fruit/vegetable, and expresses its preferences on the packaging material in terms of biodegradability, fruit/vegetable shelf life, transparency, ... Optimal technical features of the package (O2 and CO2 permeances, that measure gas fluxes through the package) are typically assessed from fruit/vegetable characteristic and gas exchange mathematical models. For each module, we will first briefly recall the method, and then give some elements about the software implementation.

A. Simulation module and interval analysis

In product design, it happens very often that some criteria of interest are not directly given but can be computed from a mathematical model. That is, given values on V variables $Y_1, \ldots, Y_V$, models have to be used to derive optimal values on a design variable X using a model $f(Y_1, \ldots, Y_V) = X$. When variable values are uncertain, different models can be used, for instance Monte-Carlo simulations, bootstrapping or random forests.

Uncertainty propagation must be performed to obtain uncertain optimal values of X. Two basic choices to perform such a propagation are Monte-Carlo simulations with probabilistic models and interval analysis methods [6]. The first one provides informative estimations of X but requires using probability distributions, while the second one yields less informative but more robust models of X and requires less information.

In biological environment and applications, variables $Y_1, \ldots, Y_V$ can display a great variability and data/information about them can be very scarce (sometimes one or two measurements over different populations). It is also often the case that models f are complex (differential equation systems, stochastic models, ...).

1) Outline of the method: In our case, output variables of interest are the O2 and CO2 permeances (Pe_{O_2} and Pe_{CO_2}) described by the following differential equations:

$$\dot{p}_{O_2}^{pkg} = \frac{Pe_{O_2} \cdot S}{e}(p_{O_2}^{ext} - p_{O_2}^{pkg}) - RR_{O_2} \cdot m = f_1 \quad (1)$$

$$\dot{p}_{CO_2}^{pkg} = \frac{Pe_{CO_2} \cdot S}{e}(p_{CO_2}^{ext} - p_{CO_2}^{pkg}) + RR_{O_2} \cdot m \cdot QR = f_2 \quad (2)$$

with

$$RR_{O_2} = \frac{RR_{O_2max} \cdot p_{O_2}^{pkg}}{(Km_{appO_2} + p_{O_2}^{pkg}) \cdot \left(1 + \frac{p_{CO_2}^{pkg}}{Ki_{CO_2}}\right)}. \quad (3)$$

These equations describe the evolution of the partial pressures ($p_{O_2}^{pkg}$ $p_{CO_2}^{pkg}$) in the packaging. They depend on several parameters usually difficult to identify, and for which only a few measurements are typically available (more details can be found in Ref. [1]). Interval analysis therefore seems the best option to evaluate optimal values of Pe_{O_2} and Pe_{CO_2}.

In [1], we have shown that Eq. (1) and (2) could be solved efficiently by using interval analysis (it comes down to solve, for each equation, a pair of classical systems, using its monotonicity properties [16]), and we have proposed a method that takes intervals on $S, e, m, RR_{O_2max}, Km_{appO_2}, Ki_{CO_2}, QR$ as inputs, and outputs fuzzy sets of optimal values over Pe_{O_2} and Pe_{CO_2}. Those fuzzy sets can then be used in the bipolar flexible querying module. Fig. 2 illustrates the output result of this module, computations being done by the Virtual MAP module (see [1]).

2) Software implementation: The database is a relational database, managed with the MySQL database management system. The user interface is written in Java. The interval analysis simulation is available as a freely available *Matlab*[17] package, that uses the optimisation toolbox to solve nonlinear least-squares problems. This is the only dependence to a commercial product. The use of that toolbox motivated the choice of *Matlab*, which requires a commercial license, as the simulation environment. A port of the simulation to an open source environment, such as R[18], or *Scilab*[19] is feasible with some effort. The calls to *Matlab* are done using the JMatLink[1] open source library to connect *Matlab* and Java.

B. Reliability module

Given an object o collected from some (electronic) document or the web, the role of the reliability is to affect

[1]http://jmatlink.sourceforge.net/

an interval-valued score $[\underline{\mathbb{E}}_o, \overline{\mathbb{E}}_o]$ that reflect the *a priori* reliability of information o. The interval is obtained through an expert system using meta-information, and the length or imprecision of $[\underline{\mathbb{E}}_o, \overline{\mathbb{E}}_o]$ reflects to which extent the various pieces of meta-information are consistent. The system is built as follows.

1) Principle: First, an ordered finite reliability space $\Theta = \{\theta_1, \dots, \theta_R\}$ is built, θ_1 being the lowest reliability value, θ_R the highest. Usually, $R = 5$ (this paper) or $R = 7$ to ensure a good compromise between complexity and expressiveness. A non-decreasing score function f on Θ is then defined, in our case $f(\theta_i) = i$.

Second, S groups $A_1, \dots, A_S$ of *meta-information* that will be used to assess reliability are defined, a group A_i taking C_i values $a_{i1}, \dots, a_{iC_i}$. Various different types of meta-information, summarized in Table I can be considered:

- meta-information on the data source itself: for instance the source type (e.g., scientific publication, technical report, ...), the source reputation, citation data;
- meta-information related to means used to collect data. Such information is typically included in a section called *material and method* in papers based on experiments in Life Science, which thoroughly describes the experimental protocol and material. Some methods may be known to be less accurate than others, but still be chosen for practical considerations;
- meta-information related to statistical procedures: presence of repetitions, uncertainty quantification (variance, confidence interval), elaboration of an experimental design.

In practice, the groups are made so that their impact on reliability can be estimated independently, which can lead to make groups A_i containing multiple criteria (e.g., number of citation and publication date).

TABLE I
RELIABILITY CRITERIA.

Source	Production	Statistics
Type	Protocol	Repetitions
Reputation	Material	Uncertainty quantification
Citation count		Experimental design
Publication date		

After groups have been formed, for each value $a_{i,j}$, $i = 1, \dots, S$, $j = 1, \dots, C_i$, an expert of the field from which data are collected gives his/her opinion about how reliable is a data whose meta-information is $a_{i,j}$. This opinion is expressed linguistically, chosen from a set of limited modalities (or combinations of them), e.g., *very unreliable*, *slightly unreliable*, *neutral*, *slightly reliable*, *very reliable* and *unknown*. Each modalities is then transformed into a fuzzy set on Θ (Fig. 3 illustrates such a fuzzy set).

To an object o are then associated S fuzzy sets $\mu_{a_1^o}, \dots, \mu_{a_S^o}$ defined on Θ corresponding to the *meta-information* associated to it. Those fuzzy sets are then merged together using evidential theory and a maximal coherent subset approach

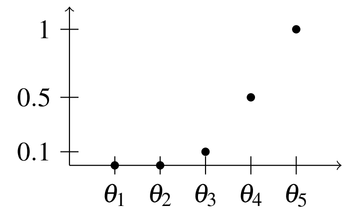

Fig. 3. Fuzzy set corresponding to the term *very reliable* defined on Θ with $R = 5$.

that allows us to deal with conflicting evidences (i.e., assessment of high reliability for an aspect but of low reliability for another). The result of this merging is a mass distribution $m_o : 2^\Theta \to [0,1]$ that reflect the global reliability of o (due to lack of space, we refer to [2] for details). Final score $\underline{\mathbb{E}}_o$ is then computed using the formula

$$\underline{\mathbb{E}}_o = \sum_{E \subseteq \Theta} m(E) \inf_{\theta_i \in E} f(\theta_i) \tag{4}$$

and $\overline{\mathbb{E}}_o$ is obtained by the same formula, replacing inf with sup. These data can then be used in the querying system to order objects of the data-base according to their reliability. In [2] are also discussed various means to analyse the result of the reliability, such as the reasons that have led to an imprecise assessments and the detection of subgroups of agreeing/disagreeing meta-information.

2) Software implementation: All belief functions-related computations have been implemented as a R [18] package. The package is called *belief* [20], and it includes basic functions to manipulate belief functions and associated mass assignments (currently on finite spaces only). It is aimed at providing a basic skeleton for belief function applications using R. It also contains an example of a script for using belief function in reliability computations, according to the method described above.

C. Bipolar flexible querying module

The final step is the querying module. For an attribute X of an object o, we consider a possibility π_D describing our knowledge of X value, and a fuzzy set μ_P expressing the user preferences about X values. Our knowledge about the imprecise evaluation of P given uncertainty D is summarised by the following upper and lower values [21]:

$$\begin{aligned} \Pi(P;D) &= \sup_{x \in \mathscr{X}} \min(\mu_P(x), \pi_D(x)), \\ N(P;D) &= \inf_{x \in \mathscr{X}} \max(\mu_P(x), 1 - \pi_D(x)). \end{aligned} \tag{5}$$

In the following, we will speak of evaluations of a fuzzy preference when talking about the interval $[N(P;D), \Pi(P;D)]$.

1) Method: We assume that we have a database consisting in a set $\mathscr{T}$ of T objects o_t, $t = 1, \dots, T$, with each object taking its values on the Cartesian product $\times_{i=1}^N \mathscr{X}_i$ of N domains $\mathscr{X}_1, \dots, \mathscr{X}_N$. An object o_t is here described by a set of N possibility distributions π_t^i, $i = 1, \dots, N$, where $\pi_t^i : \mathscr{X}_i \to [0,1]$ is the possibility distribution describing our knowledge about the value of the i^{th} attribute of object t. Such distributions can be directly inferred from (uncertain) data or be produced by a simulation (Section III-A).

In a query, we assume that the user provides the following information:

- a set $\mathscr{C} = \{C_1^{i_1}, \dots, C_{N_c}^{i_{N_c}}\}$ of N_c constraints to be satisfied by the retrieved objects with $C_j^{i_j} : \mathscr{X}_{i_j} \to [0,1]$ a normalised fuzzy set defined on the attribute i_j.
- a set $\mathscr{W} = \{W_1^{i_1}, \dots, W_{N_w}^{i_{N_w}}\}$ of N_w wishes that the retrieved objects should satisfy if possible, with $W_j^{i_j} : \mathscr{X}_{i_j} \to [0,1]$ a normalised fuzzy set defined on the attribute i_j.
- complete (pre-)orderings $\leq_c$ and $\leq_w$ between the constraints to be satisfied and between the wishes, respectively. These (pre)-orderings model the fact that some constraints are considered as more important to satisfy than others (and similarly for wishes). We denote by $\mathscr{C}_{(i)}$ (resp. $\mathscr{W}_{(i)}$) the constraints (resp. the wishes) that have rank i w.r.t. to the (pre-)ordering[2] $\leq_c$ (resp. $\leq_w$). We denote by $|\leq_c|$ and $|\leq_w|$ the total number of ranks induced by the two orderings.

To summarize how an object o_t satisfies constraints $\mathscr{C}_{(i)}$ of rank i, we aggregate them in an interval $[N_t^{(i)}, \Pi_t^{(i)}]_c$ such that:

$$N_t^{(i)} = \top_{C_k^{j_k} \in \mathscr{C}_{(i)}} N(C_k^{j_k}; \pi_t^{j_k}); \Pi_t^{(i)} = \top_{C_k^{j_k} \in \mathscr{C}_{(i)}} \Pi(C_k^{j_k}; \pi_t^{j_k}) \quad (6)$$

with $N(C_k^{j_k}; \pi_t^{j_k})$, $\Pi(C_k^{j_k}; \pi_t^{j_k})$ given by Eq. (5), and $\top$ a conjunctive t-norm [22], chosen here for the reason that ALL constraints have to be satisfied. Here, we take $\top = \min$, the minimum operator.

Similarly, we build, for each $\mathscr{W}_{(i)}$ and object o_t satisfying the constraints (other objects being rejected), an interval $[N_t^{(i)}, \Pi_t^{(i)}]_w$

$$N_t^{(i)} = \oplus_{W_k^{j_k} \in \mathscr{W}_{(i)}} N(W_k^{j_k}; \pi_t^{j_k}); \Pi_t^{(i)} = \oplus_{W_k^{j_k} \in \mathscr{W}_{(i)}} \Pi(W_k^{j_k}; \pi_t^{j_k}), \quad (7)$$

where $\oplus$ is an aggregation operator that can be a t-norm, an averaging operator such as an OWA [23] operator or a t-conorm, depending the behaviour we want to adopt w.r.t. the satisfaction of wishes (as wishes can be treated with more flexibility [24]). In this paper, we will use the min in our case study.

The set $\mathscr{T}$ is then partitioned into increasing equivalent classes $\{\mathscr{T}_0, \dots, \mathscr{T}_M\}$, $\mathscr{T}_0$ containing the rejected objects (not satisfying the constraints, i.e., $\Pi_t^{(i)} = 0$ for some $C_{(i)}$), $\mathscr{T}_1$ and $\mathscr{T}_M$ the least and most relevant objects, respectively. The partitioning is achieved iteratively by using a lexicographic ordering: the complete pre-order is refined iteratively, first using constraints and penalizing imprecision (large $[N_t^{(i)}, \Pi_t^{(i)}]_c$) as associated preferences are negative, then using wishes and not penalizing imprecision as associated preferences are positive. Details can be found in [4], and we only provide an example of the process here.

[2]As $\leq_c$ and $\leq_w$ are complete pre-orderings, the rank is well-defined.

TABLE II
EXAMPLE 1 EVALUATIONS FOR CONSTRAINTS AND WISHES.

	$[N_t^{(1)}, \Pi_t^{(1)}]_c$	$[N_t^{(2)}, \Pi_t^{(2)}]_c$	$[N_t^{(1)}, \Pi_t^{(1)}]_w$
o_1	$[0.1, 0.4]$	$[0.8, 1]$	$[1, 1]$
o_2	$[0.5, 0.8]$	$[0.5, 0.6]$	$[0.6, 0.9]$
o_3	$[0.3, 1]$	$[0.4, 0.8]$	$[0, 1]$
o_4	$[0.8, 1]$	$[0, 0]$	$[0.5, 0.7]$
o_5	$[1, 1]$	$[0.2, 0.4]$	$[0, 0]$
o_6	$[0, 1]$	$[0.6, 0.9]$	$[0.3, 0.7]$

Example 1. *Let us consider a set $\mathscr{T}$ of six objects $o_1, \dots, o_6$, two ranks of constraints and only one rank of wish. The intervals $[N_t^{(i)}, \Pi_t^{(i)}]_c$ ($i = \{1,2\}$) and $[N_t^{(1)}, \Pi_t^{(1)}]_w$ are summarized in table II.*

First, object o_4 is rejected $\mathscr{T}_0 = \{o_4\}$, being the only one that necessarily does not satisfy constraints ($\Pi_4^{(2)} = 0$). Partitioning on first constraint give $\mathscr{T}_0 = \{o_4\} < \mathscr{T}_1 = \{o_1, o_6\} < \mathscr{T}_2 = \{o_2, o_3\} < \mathscr{T}_3 = \{o_5\}$, where o_5 is the best as it fully satisfies the most important constraint, and o_6 is at the end since we adopt a penalizing attitude towards imprecision in constraints. The refinement on the second constraint rank gives $\mathscr{T}_0 = \{o_4\} < \mathscr{T}_1 = \{o_6\} < \mathscr{T}_2 = \{o_1\} < \mathscr{T}_3 = \{o_2, o_3\} < \mathscr{T}_4 = \{o_5\}$, differentiating o_1 and o_6. The use of lexicographic ordering avoids o_5 to be penalized. Finally, using the only rank of wish we get the total $\mathscr{T}_0 = \{o_4\} < \mathscr{T}_1 = \{o_6\} < \mathscr{T}_2 = \{o_1\} < \mathscr{T}_3 = \{o_2\} < \mathscr{T}_4 = \{o_3\} < \mathscr{T}_5 = \{o_5\}$, where o_3 is preferred to o_2 because it potentially satisfy the wish better (we are not penalizing imprecision).

Reliability information can easily be integrated into the process, by considering intervals $[\underline{\mathbb{E}}_{o_i}, \overline{\mathbb{E}}_{o_i}]$ as an additional rank of wishes and or constraints, depending on how important is the reliability issue. In product design, reliability of data is usually checked experimentally or ensured by the manufacturer, hence it seems reasonable to consider it as an additional wish (either of first or last rank).

2) Software implementation: The flexible query functionalities are available in a Java engine, with a web-like interface. The open source Google Web Toolkit (GWT[3]), which is an open source set of tools that allows web developers to create and maintain complex JavaScript front-end applications in Java, is used to develop the user interface, as well as the Sencha GXT[4], that provides high performance widgets and is free of use for non commercial developments.

IV. CASE STUDY

In this section, we present a use case of the DSS concerning the choice of a packaging for endive. The user has to specify a set of parameters needed by the DSS to determine the optimal O2 permeance of the targeted packagings required for endive preservation. The mass of vegetable chosen is 500 g in a package volume of 0.002 m^3 with a surface of 0.14 m^2. The film thickness was considered

[3]https://developers.google.com/web-toolkit/doc/1.6/DevGuideUserInterface
[4]http://www.sencha.com/products/gxt

as 50 microns for calculations and the targeted shelf life is 7 days. Simulations were performed at 20 $^{\circ}C$ because interest of using virtual MAP simulation for this kind of fresh produce is to avoid intensive use of the chilled chain, an energy-consuming process. Using the respiration parameters retrieved from the Fresh produce database (see Fig. 1), an optimal window of O2 permeance of [2.92E-11, 4.38E-11] $mol.m^{-2}.s^{-1}.Pa^{-1}$ obtained taking into all uncertainty propagation during virtual MAP simulation as presented in section III-A.

This optimal permeance will be considered as one of the criteria used to scan the Packaging database in the next step of multi-criteria querying. Then, the range of optimal permeances was enlarged taking a 60% of variation around the mean value of 3.65E-11 $mol.m^{-2}.s^{-1}.Pa^{-1}$ as described above. Result is the fuzzy graph presented in Fig. 5 illustrating that, from the aforementioned targeted O2 permeance, a [1.46E-11 - 5.85E-11] min-max span for O2 permeance was considered as lower and upper limits. The graphical user interface which permits to compute the optimal O2 permeance is presented in Fig. 4. We consider in this use case that the user is also interested by two additional criteria: the biodegradability and the transparency. Transparency may be an asset for fresh produce because consumer could see the product and thus judge of its quality (or not).

An excerpt of the Packaging database content is presented in Tables III and IV. It is composed of 14 packagings, characterized by their permeance values obtained at a given temperature (see Table III), their transparency and their biodegradability (see Table IV). It will be used in the following to illustrate the process of multi-criteria querying using the procedure described in Section III-C. The following query will be considered: "I want firstly guaranteed product quality all long its shelf life at room temperature and secondly a transparent packaging material and if possible firstly a biodegradable material and secondly using reliable packaging data". The fuzzy sets associated with the permeance and temperature preferences, corresponding to a guaranteed product quality all long its shelf life at room temperature, are presented in Fig. 5.

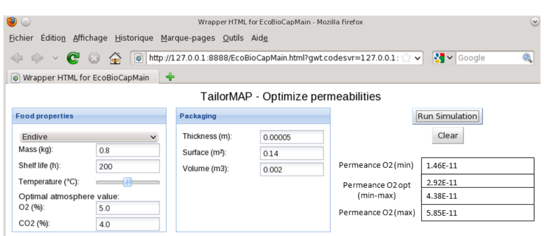

Fig. 4. DSS graphical user interface which permits to compute the optimal O2 permeance.

The fuzzy set associated with the transparency (resp. biodegradability) criterion is[5]:

[5]Here, we adopt the usual notation (x,y) for specifying fuzzy sets over symbolic variables, where (x,y) means that modality x has membership value y.

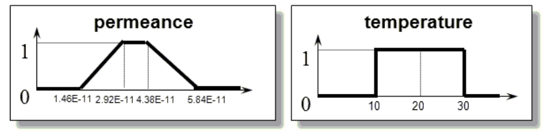

Fig. 5. Preferences for permeance and temperature.

TABLE III
PERMEANCE AT A GIVEN TEMPERATURE FOR A EXCERPT OF THE PACKAGING DATABASE.

o_{id}	*PackagingType*	*Permeance* ($mol.m^{-2}.s^{-1}.Pa^{-1}$)	*Temperature* ($^{\circ}C$)
o_1	*Polyolefin*	$1.29E-13$	23
o_2	*Polyolefin*	$4.05E-11$	23
o_3	*Cellophane*	$1.55E-14$	23
o_4	*Polyolefin*	$[1.96E-11, 2.39E-11]$	20
o_5	*Cellulose*	$1.55E-14$	23
o_6	*Polyester*	$4.46E-12$	23
o_7	*Polyolefin*	$1.50E-11$	23
o_8	*Polyester*	$1.55E-13$	23
o_9	*Polystyrene*	$1.03E-12$	23
o_{10}	*Polyester*	$6.23E-12$	23
o_{11}	*Wheatgluten*	$[1.55E-11, 1.67E-11]$	25
o_{12}	*PolyVinylChloride*	$7.47E-11$	25
o_{13}	*Polysaccharides*	$[2.95E-11, 3.00E-11]$	20
o_{14}	*PolyVinylChloride*	$3.99E-11$	20

$Pref_{transparency}$={(transparent,1),(translucent,0),(opaque,0)} (resp. $Pref_{biodegradability}$={(yes,1),(no,0)}). They correspond to crisp requirements provided by the user, as the concept of graded biodegradability made little sense to the user, while translucency is not graded in our current data. In the following, we firstly present the reliability assessment of the packaging data in section IV-A and the bipolar query evaluation in section IV-B.

A. Reliability assessment

The reliability assessment is applied, using the procedure described in Section III-B, on the excerpt of the Packaging database content presented in tables III and IV. We first give

TABLE IV
TRANSPARENCY AND BIODEGRADABILITY FOR THE SAME EXCERPT OF THE PACKAGING DATABASE.

o_{id}	*PackagingType*	*Transparency*	*Biodegradability*
o_1	*Polyolefin*	*transparent*	*no*
o_2	*Polyolefin*	*transparent*	*no*
o_3	*Cellophane*	*transparent*	*yes*
o_4	*Polyolefin*	*transparent*	*no*
o_5	*Cellulose*	*transparent*	*yes*
o_6	*Polyester*	*transparent*	*yes*
o_7	*Polyolefin*	*transparent*	*no*
o_8	*Polyester*	*translucent*	*yes*
o_9	*Polystyrene*	*translucent*	*no*
o_{10}	*Polyester*	*translucent*	*yes*
o_{11}	*Wheatgluten*	*translucent*	*yes*
o_{12}	*PolyVinylChloride*	*transparent*	*no*
o_{13}	*Polysaccharides*	*transparent*	*yes*
o_{14}	*PolyVinylChloride*	*transparent*	*no*

the criteria suited to this field, as well as the corresponding expert opinions and fuzzy sets. We then detail the results.

1) Customized criteria: We define criteria corresponding to the three kinds discussed in Section III-B: source, production and statistics. They have been determined with the expert, and their values are given in Table V, together with the labels.

TABLE VI
FUZZY SETS ON Θ CORRESPONDING TO LINGUISTIC MODALITIES.

	$\mu(\theta_1)$	$\mu(\theta_2)$	$\mu(\theta_3)$	$\mu(\theta_4)$	$\mu(\theta_5)$
Unknown	1	1	1	1	1
Very unrel.	1	0.5	0.1	0	0
Slightly unrel.	0.9	1	0.5	0	0
Neutral	0	0.5	1	0.5	0
Reliable	0	0	0.5	1	0.9
Very reliable	0.	0	0.1	0.5	1

2) Fuzzy sets definition: Associated fuzzy sets defined on Θ are shown in Table VI. Let us point out the difference between the *Unknown* modality, meaning a lack of information, and the *Neutral* one, associated to a neutral opinion.

3) Expert opinions: Experts provided opinions about reliability values for the different criteria labels, according to the linguistic terms defined in Table VI. They are summarised in Table VII for $\mathscr{A}_1$ and $\mathscr{A}_2$ and in Table VIII for the A_3 group (*citation number and publication age*).

TABLE VII
EXPERT OPINIONS ABOUT RELIABILITY FOR $\mathscr{A}_1$ AND $\mathscr{A}_2$.

$\mu_{a_{11}}$	very reliable
$\mu_{a_{12}}$	very reliable or reliable
$\mu_{a_{13}}$	neutral
$\mu_{a_{14}}$	slightly unreliable
$\mu_{a_{21}}$	reliable
$\mu_{a_{22}}$	slightly unreliable
$\mu_{a_{41}}$	very unreliable
$\mu_{a_{42}}$	reliable
$\mu_{a_{43}}$	very reliable

TABLE VIII
EXPERT OPINIONS ABOUT RELIABILITY FOR $\mathscr{A}_3$ - CITATION NUMBER AND PUBLICATION AGE INTERDEPENDENT CRITERIA.

#cit. \ age	0-2	3-8	+8
0-5	unknown	slightly unrel.	very unrel.
6-10	unknown	neutral	slightly unrel.
11-20	slightly rel.	slightly rel.	slightly unrel.
21-40	very rel.	very rel.	slightly rel.
+40	very rel.	very rel.	very rel.

4) Results: The citation number has been determined using Google Scholar. Reliability features and results are given in Table IX for the packagings presented in tables III and IV.

The relatively small size of this example (three meta-information groups, fourteen data) allows us to illustrate the behaviour of the reliability module. Roughly speaking, we can distinguish three groups of objects. Objects $\{o_1, o_5, o_6, o_7, o_8, o_9, o_{10}\}$ receive imprecise reliability scores (large $[\underline{\mathbb{E}}_{o_i}(f_\Theta), \overline{\mathbb{E}}_{o_i}(f_\Theta)]$) due to the fact of two conflicting information, namely the reliable source (technical sheet) but the relatively unreliable statistics concerning them (no repetitions). On the contrary, objects $\{o_2, o_3, o_{11}, o_{13}\}$ are deemed highly reliable, coming from often cited sources in which statistical repetition were performed. Object $\{o_4\}$ and objects $\{o_{12}, o_{14}\}$ receive respectively slightly less reliable estimate, due to their number of citations, but remains nevertheless quite reliable.

TABLE IX
RELIABILITY CRITERIA VALUES AND RESULTS ASSOCIATED WITH THE PACKAGINGS PRESENTED IN TABLES III AND IV.

o_{id}	$\mathscr{A}_1$ (#cit vs age)	$\mathscr{A}_2$ repetitions	$\mathscr{A}_3$ source type	$[\underline{\mathbb{E}}_{o_i}(f_\Theta), \overline{\mathbb{E}}_{o_i}(f_\Theta)]$
o_1	-	no	technical sheet	[1.40,4.66]
o_2	(79,3)	yes	journal paper	[4.67,4.97]
o_3	(79,3)	yes	journal paper	[4.67,4.97]
o_4	(33,9)	yes	journal paper	[4.38,4.90]
o_5	-	no	technical sheet	[1.40,4.66]
o_6	(-,6)	no	technical sheet	[1.40,4.66]
o_7	(-,6)	no	technical sheet	[1.40,4.66]
o_8	(-,6)	no	technical sheet	[1.40,4.66]
o_9	(-,6)	no	technical sheet	[1.40,4.66]
o_{10}	(-,6)	no	technical sheet	[1.40,4.66]
o_{11}	(130,16)	yes	journal paper	[4.67,4.97]
o_{12}	(9,8)	yes	journal paper	[3.35,4.67]
o_{13}	(151,18)	yes	journal paper	[4.67,4.97]
o_{14}	(9,8)	yes	journal paper	[3.35,4.67]

B. Bipolar query evaluation

Using the notations introduced in the Section III-C, the query is built as follows: $\mathscr{C}_{(1)} = \{Pref_{permeance}, Pref_{temperature}\}$, $\mathscr{C}_{(2)} = \{Pref_{transparency}\}$ and $\mathscr{W}_{(1)} = \{Pref_{biodegradability}\}$, $\mathscr{W}_{(2)} = \{Pref_{datareliability}\}$

Let us consider the set $\mathscr{T} = \{o_1, \ldots, o_{14}\}$ of the fourteen packages whose characteristics are given in tables III and IV and whose evaluations for the constraint and wishes of the example of query are computed according to Eq. (7)). After the run of the flexible querying procedure described in Section III-C, we obtain the following partition:

$$\mathscr{T}_0 = \{o_1, o_3, o_5, o_6, o_8, o_9, o_{10}, o_{11}, o_{12}\} < \mathscr{T}_1 = \{o_7\} <$$
$$\mathscr{T}_2 = \{o_4\} < \mathscr{T}_3 = \{o_{14}\} < \mathscr{T}_4 = \{o_2\} < \mathscr{T}_5 = \{o_{13}\}.$$

It must be noticed that the better reliability of the data associated with object $\{o_2\}$ compared to object $\{o_{14}\}$ allowed us to refine their relative ranking as they were in the same partition before this ultimate refinement.

V. CONCLUSION

In this paper, we have presented a decision support system architecture FQR-DSS, that can be applied to different problems. Although we have illustrated its application on a packaging material design problem, the used methods are generic and can be used in other case-studies with little effort (i.e., adapting their characteristics to the particular situation). Most of the methods are developed using open-source code and software, for a better dissemination. The

TABLE V
RELIABILITY CRITERIA.

Source				*Statistics*		*Production*
$\mathscr{A}_1$ source type				$\mathscr{A}_2$ experiment repetition		$\mathscr{A}_3$ citation number × publication age
journal paper a_{11}	international report a_{12}	project report a_{13}	technical sheet a_{14}	Yes a_{21}	No a_{22}	{[0,5],[6,10],[11,20],[21,40],[40+]} (number) × {[0,2],[3,8],[8+]} (age)

system is modular, and each part (uncertainty propagation, reliability assessment, flexible querying) can be implemented and customized separately. A demonstration is available online [6]. Undergoing developments are currently being done to remove the dependence to commercial software (*Matlab*). There are a few desirable evolutions to the current system that we can think of :

- first, it would be desirable to integrate in the methods the preferences of multiple-users or the opinions of multiple experts. This requires to address two different but related problems, namely expert opinion [25] and preference aggregation [26];
- second, it would be necessary to check that the use of a lexicographic ordering is supported by user preferences, and possibility to even learn such an ordering from user preferences observations [27];
- third, make a more important use of structured knowledge, and in particular of ontological knowledge concerning the different attributes of objects [28].

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement no FP7-265669-EcoBioCAP project.

REFERENCES

[1] S. Destercke and V. Guillard, "Interval analysis on non-linear monotonic systems as an efficient tool to optimise fresh food packaging," *Computers and Electronics in Agriculture*, vol. 79, pp. 116–124, 2011.

[2] S. Destercke, P. Buche, and B. Charnomordic, "Evaluating Data Reliability: An Evidential Answer with Application to a Web-Enabled Data Warehouse," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2012.

[3] D. Dubois and H. Prade, "An overview of the asymmetric bipolar representation of positive and negative information in possibility theory," *Fuzzy Sets and Systems*, vol. 160, no. 10, pp. 1355–1366, 2009.

[4] S. Destercke, P. Buche, and V. Guillard, "A flexible bipolar querying approach with imprecise data and guaranteed results," *Fuzzy Sets and Systems*, vol. 169, no. 1, pp. 51–64, 2011.

[5] T. Bedford and R. Cooke, *Probabilistic Risk Analysis. Foundations and Methods*. UK: Cambridge University Press, 2001.

[6] C. Baudrit, D. Dubois, and N. Perrot, "Representing parametric probabilistic models tainted with imprecision," *Fuzzy Sets and Systems*, no. 15, pp. 1913–1928, 2008.

[7] D. Mercier, B. Quost, and T. Denoeux, "Refined modeling of sensor reliability in the bellief function framework using contextual discounting," *Information Fusion*, vol. 9, pp. 246–258, 2008.

[8] F. Delmotte, "Detection of defective sources in the setting of possibility theory," *Fuzzy Sets and Systems*, vol. 158, pp. 555–571, 2007.

[9] A. Denguir-Rekik, J. Montmain, and G. Mauris, "A possibilistic-valued multi-criteria decision-making support for marketing activities in e-commerce: Feedback based diagnosis system," *European Journal of Operational Research*, vol. 195, no. 3, pp. 876–888, 2009.

[10] S. Ramchurn, D. Huynh, and N. Jennings, "Trust in multi-agent systems," *The Knowledge Engineering Review*, vol. 19, pp. 1–25, 2004.

[11] P. Bosc, L. Lietard, and O. Pivert, *Fuzziness in Database Management Systems*. Academic Press, 1999, ch. Fuzzy theory techniques and applications in data-base management systems, pp. 666–671.

[12] P. Buche and O. Haemmerlé, "Towards a unified querying system of both structured and semi-structured imprecise data using fuzzy views," in *Proceedings of the 8th ICCS, LNAI #1867*, August 2000, pp. 207–220.

[13] D. Dubois and H. Prade, "An overview of the asymetric bipolar representation of positive and negative information in possibility theory," *Fuzzy Sets and System*, vol. 160, pp. 1355–1366, 2008.

[14] P. Bosc, O. Pivert, A. Mokhtari, and L. Lietard, "Extending relational algebra to handle bipolarity," in *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC), Sierre, Switzerland, March 22-26, 2010*. ACM, 2010, pp. 1718–1722.

[15] G. D. Tré, S. Zadrozny, and A. Bronselaer, "Handling bipolarity in elementary queries to possibilistic databases," *IEEE Trans. on Fuzzy Systems*, vol. 18, pp. 599–612, 2010.

[16] N. Ramdani, N. Meslem, and Y. Candau, "Computing reachable sets for uncertain nonlinear monotone systems," *Nonlinear Analysis: Hybrid Systems*, vol. 4, pp. 263 – 278, 2010.

[17] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[18] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[19] Scilab Enterprises, *Scilab: Le logiciel open source gratuit de calcul numérique*, Scilab Enterprises, Orsay, France, 2012. [Online]. Available: http://www.scilab.org

[20] N. Maillet, B. Charnomordic, and S. Destercke, *belief: Contains basic functions to manipulate belief functions and associated mass assignments*, 2010, r package version 1.0. [Online]. Available: http://CRAN.R-project.org/package=belief

[21] D. Dubois and H. Prade, *Fuzziness in Database Management Systems*. Physica-Verlag, 1995., ch. Tolerant fuzzy pattern matching: An introduction.

[22] E. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. Dordrecht: Kluwer Academic Publisher, 2000.

[23] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decision making," *IEEE Trans. on Syst., Man, and Cybern.*, vol. 18, pp. 183–190, 1988.

[24] S. Benferhat, D. Dubois, S. Kaci, and H. Prade, "Bipolar possibility theory in preference modelling: representation, fusion and optimal solutions," *Information Fusion*, vol. 7, pp. 135–150, 2006.

[25] S. Sandri, D. Dubois, and H. Kalfsbeek, "Elicitation, assessment, and pooling of expert judgments using possibility theory," *Fuzzy Systems, IEEE Transactions on*, vol. 3, no. 3, pp. 313–335, 1995.

[26] M. Croitoru, J. Fortin, and N. Oren, "Arguing with preferences in ecobiocap," in *COMMA*, 2012, pp. 51–58.

[27] F. Yaman, T. J. Walsh, M. L. Littman, and M. desJardins, "Learning lexicographic preference models," in *Preference Learning*, J. Fürnkranz and E. Hüllermeier, Eds. Springer-Verlag, 2010, pp. 251–272.

[28] O. Haemmerlé, P. Buche, and R. Thomopoulos, "The miel system: Uniform interrogation of structured and weakly-structured imprecise data," *J. Intell. Inf. Syst.*, vol. 29, no. 3, pp. 279–304, 2007.

[6] http://web.supagro.inra.fr/partage/buche/demoApricot.zip