



HAL
open science

Data quality assessment through a preference model

Julian Le Deunf, Arwa Khannoussi, Laurent Lecornu, Patrick Meyer, John Puentes

► **To cite this version:**

Julian Le Deunf, Arwa Khannoussi, Laurent Lecornu, Patrick Meyer, John Puentes. Data quality assessment through a preference model. *Journal of data and information quality*, 2023, 16 (1), pp.7. 10.1145/3632407. hal-04319806

HAL Id: hal-04319806

<https://imt-atlantique.hal.science/hal-04319806>

Submitted on 25 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

Data quality assessment through a preference model

JULIAN LE DEUNF, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, and Shom, France

ARWA KHANNOUSSI, IMT Atlantique, LS2N, UMR CNRS 6004, F-44300 Nantes, France

LAURENT LECORNU, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France

PATRICK MEYER, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France

JOHN PUENTES, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France

Evaluating the quality of data is a problem of a multi-dimensional nature and quite frequently depends on the perspective of an expected use or final purpose of the data. Numerous works have explored the well-known specification of data quality dimensions in various application domains, without addressing the inter-dependencies and aggregation of quality attributes for decision support. In this work we therefore propose a context-dependent formal process to evaluate the quality of data which integrates a preference model from the field of Multi-Criteria Decision Aiding. The parameters of this preference model are determined through interviews with work-domain experts. We show the interest of the proposal on a case study related to the evaluation of the quality of hydrographical survey data.

CCS Concepts: • **Mathematics of computing** → **Information theory**; • **Information systems** → **Data management systems**; **Information integration**; **Data analytics**; **Decision support systems**.

Additional Key Words and Phrases: Data quality assessment, multi-criteria decision aiding, preference modelling, quality parameters, assessment explanations

ACM Reference Format:

Julian Le Deunf, Arwa Khannoussi, Laurent Lecornu, Patrick Meyer, and John Puentes. 000. Data quality assessment through a preference model. *ACM J. Data Inform. Quality* 000, 000, Article 000 (000), 22 pages. <https://doi.org/000>

1 INTRODUCTION

From a conceptual point of view, the validity of collected data and extracted information should be consistently verified, by for instance comparing it with reference values, before it can be effectively used. Multiple hypotheses are nonetheless necessary to assess this quality when the ground truth is unavailable, restricted, inadequate, or estimated by theoretical models. Scenarios, on which various streaming sensors are used to evaluate physical conditions, fit into this particular kind of quality assessment. In such situations, regardless of quantitative quality measurements, some complementary qualitative evaluations are carried out by work-domain experts, to ensure a certain level of global assessment coherence. What's more, changing operational contexts, significant data volumes, different uses for a given dataset, and diverse user priorities and preferences, motivate a quality assessment by analyzing multi-dimensional and conflicting variable sensor streams, or attributes. Taking into

Authors' addresses: Julian Le Deunf, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, and Shom, Development bathymetry department, Brest, France, julian.le.deunf@shom.fr; Arwa Khannoussi, IMT Atlantique, LS2N, UMR CNRS 6004, F-44300 Nantes, France, arwa.khannoussi@imt-atlantique.fr; Laurent Lecornu, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France, laurent.lecornu@imt-atlantique.fr; Patrick Meyer, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France, patrick.meyer@imt-atlantique.fr; John Puentes, IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France, john.puentes@imt-atlantique.fr.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 000 Association for Computing Machinery.

1936-1955/000/000-ART000 \$15.00

<https://doi.org/000>

account simultaneously these conflicting attributes is thus necessary, to determine which is likely to be the best dataset for a particular use, depending on the deployed sensors types, their features, acquired data, characteristics of the datasets, and the purpose of the data. This is a complex decision problem, in which these datasets can be considered as decision alternatives, requiring formal tools to provide a response, i.e. the most appropriate alternatives for the user's need. The models and methods from the field of Multi-Criteria Decision Aiding (MCDA) are capable of evaluating such multidimensional alternatives, to suggest a decision recommendation when these attributes are conflicting or when the quality cannot be represented as a unique quantitative score. Yet, it is not clear if the models from this field are compatible with quality assessment, whether they can provide an adequate quality evaluation framework, and how they could be applied.

Our work therefore seeks to answer the question on how models and methods from MCDA can be used to determine the pertinence of a dataset for different uses, according to variable user priorities and preferences. In our case the inputs of the MCDA evaluation are estimated quality parameters of raw data streams from sensors, in specific acquisition conditions. Taking quality parameters pre-defined by work-domain experts, our work proposes a data quality assessment process which allows to define quality profiles according to user requirements, uses those parameters as input to an MCDA method, and evaluates the overall quality of a dataset through a preference model. A first argument in support of our approach is once the evaluation model is configured with the expert preferences it can be used even by a non expert. Then, it is also a context-related evaluation modeled through the expert's preferences. And finally, our approach can be applied on real data, illustrated through a real application, namely the quality of hydrographic surveys acquired by the French national hydrographic service (Shom).

1.1 Quality evaluation for decision support

Among the numerous possibilities to use data and information quality, its analysis for decision support has been rarely explored in the literature. Some works have approached the problem from the perspective of structuring information about quality measures within a system. For example, modules were monitored to provide quality estimations at local and global levels to illustrate the system behavior according to quality transfers [29]. This implies that quality estimations of the various involved variables in each module should be obtained and aggregated. Once the local quality of connected modules has been determined the global quality can be estimated in terms of quality propagation from the input(s) to the output(s) through the different modules, depending on the existing dependencies. Individual transfer functions assigned to each module make possible to follow quality propagation, reflecting the global status of a system at a given moment, along its miscellaneous connections. Another way to aggregate and process quality information for decision support has been to infer information quality as the combination of four facets [22]: main properties (accuracy, precision, uncertainty, validity, and suitability for use); assessment of the main properties (metadata and documentation); data management and archival; ease of use (search, access, understand, trust, and available expertise). In this way, multiple quality features are categorized and grouped in four sets, to deduce a global quality evaluation, enabling to decide whether or not a dataset is appropriate for a given use. Likewise, location information quality was described and organized depending on spatial (accuracy, precision, granularity, coverage, conflicts) and temporal (update rate and recency) aspects of sensors, along with adaptation strategies to moderate quality variations at the sensor, algorithm, and application levels [23]. However, the question about how to evaluate correctly the quality of location sensing systems, to determine for instance their adequacy to a user experience remains, given the difficulty to centralize and process comparatively the diverse aspects of sensors, algorithms, and applications. Also, to share openly datasets, quality information guidelines with practical recommendations were recently defined, integrated across commonly-used Earth science systems [13]. Those recommendations give indications about the representation, sharing, and reuse of quality information at the dataset level, to promote interdisciplinary expertise. Making

use of the guidelines, stake holders are expected to decide on their own, which quality information should be represented and for which purpose.

Even if all these works indicate that the proposed approaches to define data quality were intended to generate inputs for decision making, the corresponding decision support process was neither defined nor implemented. As a consequence, it is unknown how a particular quality evaluation scheme can be used for decision making. Studies integrating MCDA and information quality can also be found in [14] where hydrographic surveys are qualified according to various experts profiles. Still, these preliminary works do not propose a generic approach transferable to other domains and data types. In addition, the quality parameters were limited and extracted manually, resulting in an deficient workflow.

1.2 Multi-criteria decision aiding

As mentioned above, evaluating the quality of data is clearly a problem of a multi-dimensional nature, and potentially depends on the perspective of an expected use or purpose of the data. In this work we propose to use aggregation techniques from the field of MCDA, which allow to aggregate multiple dimensions, while taking into account the point of view of a work-domain expert, through what we call his/her preferences.

MCDA is a branch of operational research, and its objective is to help one or several decision-makers to prepare and make decisions on a finite set X of n decision objects (or alternatives), when several conflicting consequences (or criteria), represented by the finite set J , must be taken into account. This decision-maker can either be the person who takes the responsibility for the decision act, or he/she can be an *expert* whose priorities or preferences should be taken into account in the final assessment of the alternatives. Usually, three types of problems are considered in this decision aiding context [24]: the *choice problem* which aims to recommend a subset of alternatives, as restricted as possible, containing the “satisfactory” ones; the *sorting problem* which aims to assign each alternative into pre-defined categories or classes; and, the *ranking problem* which aims to order the alternatives by decreasing order of preferences.

The mathematical tools used in MCDA have their origins mainly in two very different methodological trends [6; 11; 24]. On the one hand, the European school of thought has been developed around the concept of the outranking relation, where the decision recommendation is constructed from pairwise comparisons of the alternatives. On the other hand, the Anglo-Saxon school is based on the concept of utility or value in the Multi-Attribute Value Theory in order to obtain, by aggregation, a total comparability of the alternatives. The main differences between these two methodological schools lie in the way the alternatives are compared and in the type of information which is required from the decision maker. Outranking methods might be preferable if some impreciseness about the preferences should be modeled. Value-based methods can be favored if a compensatory behavior in the assessment should be modeled, and if the overall evaluation should be a numerical score.

Regarding the link between MCDA and quality evaluation in general, several MCDA approaches were applied to assess software quality, even if relevant criteria to select those approaches reminded unclear [30]. On the other hand, considering that data quality parameters are analogous to a multivariate problem, the relative importance of those parameters was evaluated qualitatively by the answers of users given to a scoring questionnaire and then ranked applying MCDA techniques [2]. Moreover, depending on the data, information, and application types, grouping principles of quality elements can be defined as quality evaluation profiles [21]. Despite the fact that numerous works have explored the well-known specification of data quality dimensions in various application domains, the multi-dimensional nature of quality attributes, their inter-dependencies and aggregation, is complex to cope with when used for decision support. On the other hand, although multiple separate data and information quality measurements may be of interest to experts, a user does not commonly have the knowledge to interpret a dataset pertinence from those elements. Furthermore, quality analysis has been rarely used as input

to decision support approaches, because of the complexity to aggregate multiple indicators and their respective global assessment meaningfully [7].

1.3 Research questions and contributions

According the above literature review, the three major unsolved research questions addressed in this paper are the following :

- How to assess data quality, taking into account multiple quality parameters, often of a very heterogeneous nature?
- As the quality of data generally depends on the context in which it is used, how can this context be taken into account when assessing quality?
- Given that poor-quality data may, depending on the situation, be re-measured in the future, how can recommendations be made for future measurement campaigns, with the aim of improving quality?

The main contributions that distinguish this paper from the existing literature are therefore:

- A generic methodology that takes into account the context in which the data is used when assessing its quality, and which can be deployed in a wide range of use cases thanks to its genericity.
- The use of expert preference models to model the context in which data is used.
- The integration of multiple quality parameters with their business evaluation scales (deriving from international standards or expertise), without the need for normalization techniques.

2 MOTIVATING EXAMPLE

To motivate our discourse, we present in this section a fictive, but realistic example. It will also serve as illustrative example throughout the next section, where we detail our proposal.

Consider the example of a hobby meteorologist and an amateur airplane pilot who both want to predict the weather in the next few days in Gafsa, a city in Tunisia, based on satellite observations and local temperature readings. These local temperature readings are done with a certain amount of thermometers, spread throughout the territory. For the hobby meteorologist, this forecast will mainly serve to feed a weather forecast website, whereas for the pilot, the weather forecast is used to plan a flight over the area.

The weather prediction for the hobby meteorologist does not need to be very accurate, as the website is not followed many people. However, for the airplane pilot, the precision of the weather forecast is important because the flight route will strongly depend on it.

We can see that the data is exactly the same in both cases, but the final purposes of these data depend on each end user. Consequently, depending on the defined purpose, the data quality is perceived differently by the two individuals. Indeed, the hobby meteorologist considers that the density of the thermometers and their precision is sufficient to make a weather prediction for the website he or she is maintaining. However, the airplane pilot considers that the precision of the thermometers and their density in the area are not sufficient for a prediction in which he/she has enough confidence to plan the flight.

This argues for an assessment of data quality that depends on several elements, as the context of use of the data, the purpose of the data, and the perception of the quality of these data by an end user. Thus, evaluating the quality of the data of our example in an absolute way does not make a lot of sense, as for the hobby meteorologist, the quality of the data might be sufficient, whereas for the amateur pilot it might not be.

As we will show in the next section, our contribution takes into account these various elements, and proposes an assessment of the quality of data which is tuned for the purpose of the data.

3 PROPOSED APPROACH

The proposed data quality assessment approach is intended to evaluate the quality of a data item for a specific “use” or “purpose”. Two actors are concerned by this process : (i) an “end user” who might not necessarily be an expert of this purpose, but who wishes to know the quality of the data item, (ii) an “expert” whose knowledge of the purpose is used to configure once and for all the process by indicating the quality parameters that should be used to assess the quality of the data and expressing preferences regarding these parameters.

The details of this process is depicted on Fig. 1. The process starts on the left (Start) and ends on the right (End). The “end user” is also shown on the left. The classical rectangles represent activities, and the rectangles with a folded corner represent inputs and outputs of these activities. The continuous arrows represent the transitions between the various activities of this process, whereas the dotted arrows link the inputs and outputs to the activities (data, parameters, etc.).

As an illustration, recall the motivating example from Section 2. The hobby meteorologist and the amateur airplane pilot are two different end users. The data to be evaluated are the satellite observations and the local temperature readings. And the specific uses of these data are short term predictions to be published on a website for the meteorologist, and the planning of a flight for the pilot.

As one can see on Fig. 1, once the end user starts this process, the first activity (A) is to identify the use of the data. Then he/she needs to identify whether the process has already been configured for this purpose. If the answer to that question is “No”, then he/she follows the process to activity (B) in which an “expert” of the purpose of the data is asked to configure certain elements of the process. This activity is called “Preference elicitation” and consists of two tasks: first of all (B.1), the identification of the quality parameters, which are required to evaluate the quality of the data, and second (B.2), the preference parameters inference, which allow to model the preferences of the expert regarding the assessment of the data through the quality parameters. The output of (B.1) is the set of quality parameters which needs to be calculated, whereas the outputs of (B.2) are the values of the preference parameters of the multi-criteria assessment method that model the preferences of the expert.

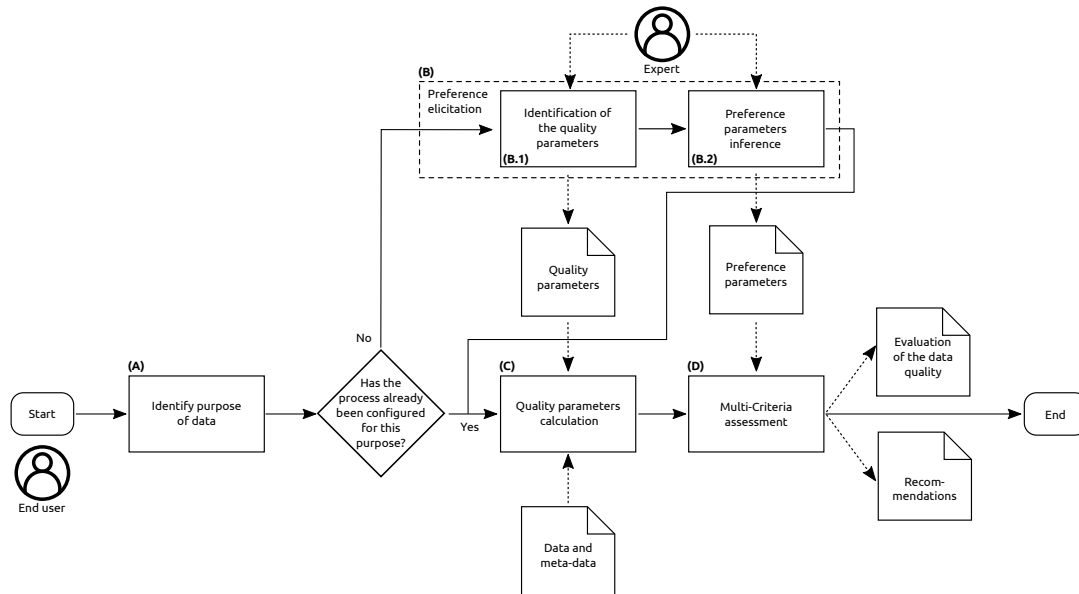


Fig. 1. Data quality assessment process

To illustrate these first steps, let us again consider the motivating example from Section 2, but let us focus only on the meteorologist. Imagine that this is the first time that he/she wishes to evaluate the quality of the data. The process is therefore not yet configured for his purpose, and we need to identify an expert, who will configure the process for future assessments. Let us suppose that such an expert has been found. This experienced hobby meteorologist has to identify in a first step the quality parameters (B.1) to be used to evaluate the satellite observations and the local temperature readings. For example, regarding the temperature readings, these quality parameters could be the precision of the thermometers and the frequency of the readings. Then, the experienced hobby meteorologist needs to provide the values of the preference parameters (B.2), as for example the importance of the various quality parameters in the overall evaluation of the data quality.

Once the input requirements for activities (C) and (D) are fulfilled, the process is configured for the given purpose, and activity (C) can start. This activity's goal is to calculate the quality parameters related to the piece of data the end user is evaluating for the identified purpose. The input of this activity are two elements: the data (and possibly some meta-data related to it), and the quality parameters which have to be used for this specific use. With these two inputs, activity (C) then comes down to calculating the values of these quality parameters for the given input data.

Once these parameters are calculated, the next step of the process (D) is to evaluate the overall quality of the piece of data using an MCDA technique. The preference parameters required to use the MCDA technique are given (as explained before). Once the overall evaluation is calculated, the output of this activity is twofold: first the evaluation as such, and second, some explanations which help to understand the origin of this specific evaluation, together with some recommendations indicating how this assessment could be improved. This ends the proposed process.

In our illustrative example, the output evaluation of this activity (D) could be that the given input data are of medium quality to predict the weather in Gafsa, even for a hobby meteorologist. Next to that, the process provides some recommendations to improve the quality of the data for future forecasts, like for example to use more precise thermometers, or to increase the frequency of the readings. In the following subsections we detail each of the steps of this process.

3.1 Identification of the purpose of the data (A)

The goal of this first activity of the process is to determine the purpose of the data, in order to check whether the process has already been configured for this use. If it is not the case, this will lead first to a sub-process, which is needed to configure the process for this specific use. It also requires in that case, the intervention of an expert user of this purpose, so that the end user can obtain an overall assessment of the data without having to be a work-domain expert.

3.2 Quality parameters calculation (C)

Under operational conditions, datasets acquired using sensors contain numerous imperfections, as uncertain, imprecise, unavailable, erroneous, or incomplete data. Furthermore, part of these data may be contradictory or redundant, as well as impacted by uncontrollable external factors, including sensors deterioration and functioning conditions. Other factors that also affect data quality are spatial and/or temporal coverage, variable acquisition parameters, and resolution of acquired values. Consequently, multiple causes can modify acquired data, impacting decision support in unknown ways, which could imply damaging outcomes. Quality assessment is therefore required to handle the variable data characteristics generated by many types of sensors for different applications, on varied operational conditions and contexts [12], [25]. However, being directly related to an application, conforming to specific conditions and a context, data quality estimation has diverse definitions and separate quality parameters, depending on the associated system [3], [32]. Quality parameters are variables characterized to

assess data factors like redundancy, accidental errors, data sources coherence, sensor attributes (type, uncertainty, precision, amount), duration of data acquisition, signal attributes (continuity, stability, noise, interval of accepted values) outliers, potential impact of errors, acceptable data subsets of interest, and data completeness, among others.

In our case, context emerges as part of the fitness for use evaluation at two stages. The first stage occurs when quality is assessed in alignment with field engineers' expert knowledge, to evaluate the global characteristics of acquired data. Afterward, the second stage emerges when specific uses of data – hydrographic, acoustic, and oceanographic – are determined by application experts that utilize a common dataset. Data acquisition circumstances, sensor models, weather conditions, and various uncertainty estimations analyzed by field engineers, indicate up to what point the acquisition context produced known global data issues, represented by metadata qualifications. On the other hand, work-domain experts deal with the resulting data quality and depending on the particular application context, decide of which variable data defaults can be tolerated to meet the corresponding requirements defined by the features of a given product.

Since collected datasets are much likely to encompass disparate imperfections, extracted information could be altered, depending on data acquisition circumstances. For this reason quality estimation methodologies and models are frequently being defined to determine whether collected data and extracted information are relevant to support decision making [16]. Furthermore, metadata analysis emerges as a complementary approach of data quality management, for specific application requirements [31]. Namely, metadata record operational aspects of data quality parameters, documenting key data features and dependencies [1]. On the other hand, it is necessary to specify data quality according to usability too, i.e. the conformance to a set of user requirements making a dataset quality suitable for a particular application.

3.3 Multi-Criteria assessment (D)

The quality parameters extracted in the previous step become the criteria in the MCDA context presented in Section 1.2. In this section we first motivate the choice of a specific MCDA preference model, before detailing its mathematical formulation.

The evaluation scales of the various quality parameters could potentially be quite heterogeneous. Some of them could be qualitative or ordinal, while others might be quantitative. This speaks in favor of the outranking school of thought mentioned in Section 1.2, which intrinsically manages well this diversity of scales, without requiring that all scales be brought down to quantitative scales. Next to that, to facilitate the adoption of our process by users, we seek to propose a solution in which each step is easily explainable, and where the final assessment can be easily interpreted, in order to generate recommendations for the improvement of the quality of the data. This interpretability can be achieved by solving the sorting problem from the outranking paradigm. Indeed, sorting alternatives in an MCDA framework comes down to creating an overall qualitative or ordinal assessment scale (based on so-called categories) to aggregate the multiple criteria. This is usually done through category limits or profiles, which can be seen as norms, against which the alternatives are compared, in order to decide to which category they belong. Among all the available sorting algorithms of the outranking paradigm, we propose to use the MR-Sort method, which has been characterized in [4; 5], and, has the advantage of generating highly interpretable results, while being a very expressive model. In the sequel we introduce the formalism of the MR-Sort method.

Let a be the data item whose quality needs to be assessed. Let J be a finite set of criteria indexes (corresponding to the quality parameters of the previous step). For each criterion, the possible evaluations are ordered according to the preferences of the expert, who defines the preference directions of the criteria (either a lower value is preferred to a higher value, or vice-versa). The overall evaluation being categorical, let $\{c_1, \dots, c_k\}$ be the k output categories (representing the k levels of the output ordinal assessment scale), ordered by their desirability, from c_1

being the worst category to c_k being the best one: $c_k > \dots > c_1$ ($>$ stands for “is strictly preferred to”). In our quality assessment context, these categories could for example be {very high quality $>$... $>$ medium quality $>$ low quality}. These categories are characterized by a set of separating profiles $B = \{b_1, \dots, b_{k-1}\}$. Each category c_h is thus defined through its upper limit, b_h , and its lower limit, b_{h-1} , with the exception of the worst and best categories, which have only one limit. Alternative a and each category limit can be represented by a vector of evaluations with respect to the criteria. Let $g_j(a)$ denote the evaluation of a on criterion j and $g_j(b_h)$ denote the evaluation of category limit $b_h, \forall h \in \{1, \dots, k-1\}$, on criterion j . In this presentation of MR-Sort, we assume, without loss of generality, that the performances are supposed to be such that a higher value denotes a better performance. It is obvious that in a real application this is not necessarily the case. Furthermore the performances of the category limits are non-decreasing, i.e. $\forall j \in J, 1 < h < k : g_j(b_{h-1}) \leq g_j(b_h)$.

MR-Sort uses two *assignment rules* for placing a into a category: the pessimistic and the optimistic assignment rules [6; 24]. The pessimistic rule assigns a to the highest possible category c_h so that a *outranks* the category’s lower frontier b_{h-1} . The optimistic rule assigns a to the lowest possible category c_h so that the category’s upper frontier b_h *outranks* a . We will use in this work the pessimistic rule, as it generates safer recommendations.

Therefore, a is said to *outrank* a frontier b_{h-1} if and only if there is a sufficient coalition of criteria supporting the assertion “ a is at least as good as b_{h-1} ” and no criterion strongly opposes (vetoes) that assertion. To measure if a sufficient coalition of criteria considers that a is at least as good as b_{h-1} , we first define for each criterion j a function C_j which assesses whether criterion j supports that statement or not:

$$\forall j \in J, 1 \leq h \leq k : \quad (1)$$

$$C_j(a, b_{h-1}) = \begin{cases} 1, & \text{if } g_j(a) \geq g_j(b_{h-1}), \\ 0, & \text{otherwise.} \end{cases}$$

To assess whether a coalition of criteria is in favor of the outranking or not, $\forall 1 \leq h \leq k$, we first define the overall concordance as:

$$C(a, b_{h-1}) = \sum_{j \in J} w_j C_j(a, b_{h-1}), \quad (2)$$

where w_j is the weight of criterion j . The weights are defined so that they are positive ($w_j \geq 0, \forall j \in J$) and sum up to one ($\sum_{j \in J} w_j = 1$). This overall concordance is then compared to a majority threshold $\lambda \in [0.5, 1]$ extracted from the expert’s preferences along with the weights.

Even when the coalition is strong enough, a criterion may veto the outranking situation. Alternative a is therefore in a veto relation (denoted with \vee) with a profile b_{h-1} when:

$$a \vee b_{h-1} \iff \exists j \in J : g_j(a) < g_j(b_{h-1}^v). \quad (3)$$

The veto profile b_{h-1}^v represents the minimum level of performance that an alternative needs to have in order to be allowed into category c_h via the weighted coalition of criteria in favor of this assignment. If on any criterion an alternative has a lower performance than the veto profile of c_h , then it is forbidden from being assigned to c_h or above.

To summarize, alternative a outranks frontier b_{h-1} (and therefore is assigned to at least the category c_h) if and only if:

$$a S b_{h-1} \iff C(a, b_{h-1}) > \lambda \text{ and not } (a \vee b_{h-1}). \quad (4)$$

The output of this activity is an overall assessment of the data item a on the one hand, and an interpretable explanation of its evaluation. This explanation comes as a set of rules, which can then be used to make recommendations on how to improve this assessment. These rules are deduced from Formulae 2, 3 and 4. They will be illustrated in the case study in Section 4.6.

3.4 Preference elicitation (B)

The preference elicitation activity requires that an expert of the purpose of the data has been identified. He or she is deemed to have mastered knowledge in the field of the purpose of the data. Ideally, he or she must have acquired experience that is recognized by his/her peers. In a first step, the expert should indicate which quality parameters have to be used, according to his/her expertise, to evaluate the quality of the data, with respect to the considered purpose (B.1).

The second task of this activity (B.2) consists in determining the values of the preference parameters of the MR-Sort model presented in Section 1.2. They can be inferred either *directly* or *indirectly* from the expert. In direct elicitation, the precise values of these parameters are determined by questioning the expert, usually in an interactive process, where the effect of the preference parameters on the overall evaluation is presented to the expert, in order to determine them as precisely as possible. In general, category boundaries and veto profiles can be considered as “norms” (pollution levels, budget limits, ...), while criteria weights are extracted through questions about “sufficient” or “insufficient” coalitions of criteria. In case the expert is unable to answer the questions in this direct process, an indirect approach can be used, in which the expert is asked to evaluate the overall quality of some data (called “learning examples” or “assignment examples”). From his/her answers, mathematical optimization models determine the values of the parameters of the MR-Sort model, which are compatible with these global evaluations (as in [15; 17–20; 27; 28]). This paradigm is very similar to what is done in machine learning.

4 EXPERIMENTAL EVALUATION: OCEAN OBSERVATION CASE STUDY

After presenting our proposal in the previous sections, we will apply it to a real case study: qualifying a hydrographic and oceanographic survey with data from the French national hydrographic service (Shom), according to various users’ needs. During a campaign at sea, the Shom sets up a set of sensors (such as sonars, tide gauges or current meters) consistent with the mission to be carried out, allowing the acquisition of a large set of data and quality associated with these data. These hydrographic and oceanographic surveys aim to measure and model the physics of the oceans. As an example, in order to model currents and tidal phenomena, it is critical to have an estimation of the water volume displaced by the model and thus the depth of the sea floor in the study area.

The objective of this case study is thus to define the quality level of a hydrographic survey according to certain quality parameters associated with the survey, as well as according to the context in which the survey was carried out and the final needs of the user associated with these ocean data. The process proposed in Fig. 1 is used to achieve this goal. We will therefore follow its various activities. As the process has not been tuned for this quality assessment task, some work-domain experts have to be interviewed in order to configure it.

4.1 Purpose of data acquisition at sea (A)

In the context of data measured during a campaign at sea, different purposes associated with the data have been identified. The most classical use of the data is for a nautical product (often marine cartography) and therefore the accuracy of the bathymetric measurement is fundamental. Another purpose of data is to acquire knowledge about the physics of the oceans and the water masses, which requires full understanding and measurement of the tide and current phenomena. Finally, an operating concept, specific to military needs, is the knowledge of underwater ambient acoustic noise, which is of particular interest for submarine operations. These three purposes of the data are called hydrographic, oceanographic, and acoustics purposes in the rest of the article.

4.2 Ocean quality parameters identification (B.1)

In order to know the most relevant quality parameters to extract from the data we conducted interviews with three work-domain experts in the fields of hydrography, oceanography, and underwater acoustics. These experts are familiar with the data acquisition conditions and therefore are able to attach quantitative and qualitative reliable qualifications to the respective datasets. In the proposed process of Fig. 1 this corresponds to activity B.1.

After the interviews with the experts and in the context of studying the quality of oceanographic and hydrographic surveys, ten different quality parameters of the main survey characteristics were selected among a set of numerous possibilities, namely, CATZOC, POSACC, SOUACC, hydrographic coverage, type of hydrographic sensors, sedimentology, ocean data, tide duration acquisition, tide gap, and tide outlier detection. These parameters are part of the classical parameters to express the quality of hydrographic and oceanographic data. CATZOC, POSACC and SOUACC are part of the S-57 [9] standard dealing with the transfer of digital hydrographic data, as quality metadata associated with bathymetric information, while the parameters on coverage and sedimentology are part of the S-44 [10] standard for hydrographic surveys. These standards are updated according to the evolution of sensor performances (uncertainties, resolution, sampling frequency...) used during hydro-oceanographic surveys. These updates are carried out by International Hydrographic Organization (IHO) working groups, which ensure that the technical capabilities of the sensors are always in line with the expected quality of the hydro-oceanographic data.

The retained quality parameters are described as follows, including the respective possible associated quality dimension according to [8]:

- CATZOC (for CATegory of Zone Of Confidence): categorical parameter (from the best to the worst qualification as: OS, A1, A2, B, C, D and U) indicating the level of confidence associated with the bathymetric data. The quality dimension is uncertainty (measures from several sensors with acquisition variations), related to the sensors known accuracy, completeness, and consistency.
- POSACC (for POSitional ACCuracy): quantitative parameter (in meters) representing the horizontal uncertainty of bathymetric data, represented by four depth intervals from the shortest to the longest as: $< 0.50\text{m}$, $< 2\text{m}$, $< 5\text{m} + 5\%$, $< 20 + 10\%$. The quality dimension is precision (variable position approximation), related to the sensors efficiency.
- SOUACC (for SOUding ACCuracy): quantitative parameter (in meters) representing the vertical uncertainty of bathymetric data, represented by six intervals from the shortest to the longest as: $< 0.25\text{m}$, $< 0.50\text{m}$, $< 1\text{m}$, $< 2\text{m}$, $< 10\text{m}$, $< 20\text{m}$. The quality dimension is precision (variable position approximation), related to the sensors efficiency.
- Hydrographic coverage: categorical parameter (composed of six values, 300, 200, 150, 100, 80, and 50) indicating the estimated degree of coverage, see [10]. The quality dimensions are completeness and precision.
- Type of hydrographic sensors: categorical parameter representing the four types of bathymetric sensors used at Shom (MultiBeam EchouSounder (MBES), bathymetric LiDAR (Light Detection And Ranging), SingleBeam EchoSounder (SBES) + side-scan sonar, and Satellite Derived Bathymetry (SDB)). The quality dimension is uncertainty related to the sensors accuracy and completeness.
- Sedimentology: quantitative parameter (from 5 to 0) indicating the number of different sensors used to describe the type of sedimentary seabed. The quality dimensions are accuracy, precision, and completeness.
- Ocean data: categorical parameter indicating the type of oceanographic data sensor used in a measurement campaign: tide, current, meteorology (MTO), water sound celerity (with CTD: Conductivity Temperature Depth) and depicted as four sets of grouped evaluations (MTO + CTD + tide + current, MTO + CTD + tide, MTO + CTD, and MTO). The quality dimensions are accuracy, precision, and completeness.

- Tide duration: quantitative parameter (times in days) indicating the duration of acquisition of the tide gauge sensor anchored during a hydrographic survey, represented by five intervals from the shortest to the highest as: no tide, < 30 days, < 1 year, < 18 years, and > 18 years. The quality dimensions are completeness and consistency.
- Tide gap: quantitative parameter (% of missing data) indicating the percentage of missing periods in the acquisition over the total duration of the tide gauge time series, represented by 4 value: no tide, $\geq 10\%$, < 10%, and 0%. The quality dimensions are accuracy, completeness, and consistency.
- Tide outliers detection: quantitative parameter (% of outlier data) indicating the percentage of outliers in the acquisition over the total duration of the tide gauge time series, represented by 4 value: no tide, $\geq 3\%$, < 3%, and 0%. The quality dimensions are accuracy, completeness, and consistency.

The quality parameters' values are included as part of the surveys' metadata (CATZOC, POSACC, SOUACC, type of hydrographic sensors, sedimentology, and ocean data) assigned by engineers responsible for the survey, or can be extracted automatically by processing the raw data (hydrographic coverage, tide duration, tide gap, and tide outliers detection). These values reflect some aspects of a survey quality estimation according to the actual acquisition context. In order to know if the resulting dataset is appropriate to generate the product of a particular profile, it is necessary to determine the conformity of these quality parameters with the end-use expert's preferences.

4.3 Marine expert profiles and their preferences (B.2)

As mentioned in section 4.2, we have identified three totally different work-domain expert profiles, in order to respond to three different purposes, presented in section 4.1, of the acquired data: the hydrographer, the oceanographer, and the acoustician.

The hydrographer measures the bathymetry of the oceans (measurement of marine depths). He/she therefore collects data that can have an impact on the measurement of depth and that are relevant for inclusion on nautical charts as tides and depths. At Shom, the hydrographer needs to ensure safe navigation bathymetric data for all nautical products, especially for nautical charts that have a legal validity in the event of an accident. The most important data to be accurate are those related to bathymetric information (as CATZOC, bathymetric coverage, and the hydrographic sensor type) as these are the ones that guarantee a just and safe nautical chart for the navigator. Furthermore, it is important to have a high quality depth information in the most critical areas for navigation: shallow areas and access channels to ports.

The oceanographer studies all activities related to the understanding and modeling of the physical parameters of the water column (temperature, salinity, transparency, etc.) and their evolution. It is important that the oceanographer has a good information quality of the ocean data and to a lesser extent of the sedimentology data. As the oceanographer builds the corresponding physics models from a volume of water, it is also important to have a precise knowledge of the bathymetric bottom and in particular, the real presence or absence of artifacts on it that can disturb the physics model (with complex boundary conditions to manage) such as wrecks or underwater obstruction given by the POSACC and SOUACC quality assessment.

The acoustician models underwater sound propagation. Acoustic sources are of different natures: biological, geophonic, and anthropogenic. Understanding the underwater soundscape requires to master all these parameters, which is a major challenge for the knowledge of the environment. In order to understand this underwater acoustic landscape it is therefore important to be able to model entirely what happens to acoustic waves in the water column, and to their reflection on the sea floor. The quality of the information needs to be appropriate in terms of sedimentology, ocean data, and with an MBES sensor with reliable coverage.

For each of the three experts, the goal of this activity is to elicit his/her preferences regarding the various quality parameters, which are used to assess the overall quality of surveys. This step is performed during interviews with the experts, and consists in determining the preference parameters of the MR-Sort model in a direct way.

In a first step, for each expert, we identify during the interview which quality parameters should be considered in the final assessment. The hydrographer considers all 10 parameters presented in Section 4.2 in the evaluation of the quality of a survey, whereas the acoustician and the oceanographer do not consider the CATZOC. Furthermore, the parameters taken into account by the oceanographer also depend on the average depth of the survey: if the survey is very deep (i.e. with an average depth above 1000 meters), the tide parameters are not considered. If however the survey is less deep (below 1000 meters), then all three tide parameters have to be considered. This leads to context-dependent preference parameters, where the context is defined by the average depth of the survey.

In a second step, we determine for each expert the granularity of the output assessment scale. For the acoustician, the two categories, “Good (G)” and “Bad (B)” are sufficient (obviously preferentially ordered as Good (G) > Bad (B)), whereas for the hydrographer and the oceanographer a supplementary intermediate “Acceptable (A)” category is needed (which leads to the preference order of Good (G) > Acceptable (A) > Bad (B)). The scales, for each expert, in the final assessment of a survey are derived directly from the way hydro-oceanographic surveys are evaluated today. Those preference scales rely on thoroughly tested and reviewed standards, defined consistently for practitioners, being as a result reliable, accurate, and effective.

Regarding the preference scales of the criteria, the three experts have unanimously defined them as follows (> denotes a strict preference between two successive levels of the scale, whereas ~ stands for equivalent levels:

- CATZOC: $A1 >_C A2 >_C B >_C C >_C D >_C U$
- POSACC: “< 0.5m” >_P “< 2m” >_P “< 5m + 5% depth” >_P “< 20m + 10% depth”
- SOUACC: “< 0.25m” >_{SA} “< 0.5m” >_{SA} “< 1m” >_{SA} “< 2m” >_{SA} “< 10m” >_{SA} “< 20m”)
- Hydrographic coverage “300%” >_H “200%” >_H “150%” >_H “100%” >_H “80%” >_H “50%”
- Hydrographic sensors: “MBES” >_{HS} “LiDAR” >_{HS} “SBES + Side-scan sonar” >_{HS} “SDB”
- Sedimentology (number of sensors) $5 >_S 4 >_S 3 >_S 2 >_S 1 >_S 0$
- Ocean data: “MTO + CTD + tide + current” >_{OD} “MTO + CTD + tide” ~_{OD} “MTO + CTD + current” >_{OD} “MTO + CTD” >_{OD} “MTO”
- Tide duration: “≥ 18 years” >_{TD} “≥ 1 year” >_{TD} “≥ 30 days” >_{TD} “< 30 days” >_{TD} “No tide”
- Tide gap: “0” >_{TG} “< 10%” >_{TG} “≥ 10%” >_{TG} “No tide”
- Tide outliers: “0” >_{TO} “< 5%” >_{TO} “≥ 5%” >_{TO} “No tide”

The remaining preference parameters identified during these three interviews are given in Tables 1 to 3. These preference parameters have been determined through a direct elicitation approach. The main reason for this choice is the poor availability of enough learning data. Next to that, the experts’ advanced knowledge eased the task of providing the separation profiles, the veto profiles, and the weights directly. For each expert we first present the separation profile(s) and the veto profile, before showing the criteria weights and the majority threshold. For the hydrographer and the oceanographer, the first column indicates the context of the survey, i.e. its average depth interval.

Table 1. Preference parameters for the acoustician end-use

Acoustician	POSACC	SOUACC	Cover.	Sensors	Sed.	Ocean data	Tide duration	Tide gap	Tide outl.
G / B (b_1)	$< 5m + 5\%$	$< 1m$	150%	MBES	4	MTO + CTD + tide + current	< 30 days	$\geq 10\%$	$\geq 3\%$
Veto (G) (b_1^v)	1
Weights (w_j)	2/46	1/46	7/46	9/46	9/46	6/46	5/46	4/46	3/46
Majority threshold (λ)	23/46								

Table 2. Preference parameters for the hydrographic end-use

Context	Hydrographer	CATZOC	POSACC	SOUACC	Cover.	Sensors	Sed.	Ocean data	Tide duration	Tide gap	Tide outl.
10m	G / A ($b_{2,10m}$)	A1	$< 0.5m$	$< 0.25m$	200%	MBES	2	MTO + CTD + tide + current	≥ 30 days	0	0
	A / B ($b_{1,10m}$)	B	$< 5m + 5\%$	$< 2m$	100%	SBES + Side-scan sonar	2	MTO + CTD + tide	< 30 days	$< 10\%$	$< 3\%$
100m	G / A ($b_{2,100m}$)	A2	$< 5m + 5\%$	$< 1m$	150%	MBES	2	MTO + CTD + tide + current	≥ 30 days	$< 10\%$	$< 3\%$
	A / B ($b_{1,100m}$)	D	$< 20m + 10\%$	$< 2m$	100%	SBES + Side-scan sonar	2	MTO + CTD + tide	< 30 days	$< 10\%$	$\geq 3\%$
1000m	G / A ($b_{2,1000m}$)	B	$< 20m + 10\%$	$< 20m$	100%	MBES	2	MTO + CTD	No tide	No tide	No tide
	A / B ($b_{1,1000m}$)	D	$< 20m + 10\%$	$< 20m$	100%	SBES + Side-scan sonar	2	MTO + CTD	No tide	No tide	No tide
All	Veto (G & A) (b_1^v & b_2^v)	.	.	.	100%
All	Weights (w_j)	10/55	6/55	7/55	9/55	8/55	3/55	5/55	4/55	2/55	1/55
All	Majority threshold (λ)	27.5/55									

4.4 Automatic parameter extraction (C)

From the data measured by the sensors or from the associated metadata it is possible to automatically extract the quality parameters that will feed our model. As mentioned in section 4.2, the quality parameters used in this process are either directly taken from the survey metadata (calculated/measured by the survey engineers) or extracted from the survey data under study. We observe in listing 1, which represents an extract from the metadata file of survey S2015001, some quality parameters chosen by the work-domain expert to evaluate the overall quality of a hydro-oceanographic survey.

Listing 1. Extract of the metadata file (.xml format) of the S2015001 survey.

```
<?xml version="1.0" encoding="UTF-8"?>
<BDB_Simple_Attributes>
  <Attribute name="CATZOC">
    <Value>zone of confidence B</Value>
  </Attribute>
  <Attribute name="OBJNAM">
```

```

    <Value>S201500100-1</Value>
  </Attribute>
  <Attribute name="POSACC">
    <Value>2.1</Value>
  </Attribute>
  <Attribute name="SOUACC">
    <Value>0.43</Value>
  </Attribute>
</BDB_Simple_Attributes>

```

In this manner, the CATZOC, POSACC, SOUACC, or Sensors are retrieved directly by parsing the metadata file in .xml format generated for each survey.

Other quality elements are extracted from the particular report (hydro-oceanographic campaign report) such as the Hydrographic coverage, Sedimentology, and Ocean data parameters. Finally, the remaining quality parameters, Tide duration, Tide gap and Tide outliers detection, can be directly computed from the raw water level data of the tide gauge sensors, anchored during the hydro-oceanographic campaigns.

To illustrate how quality parameters can be extracted automatically by processing raw data, below we describe various aspects of such parameters extraction, making use of the water level signal measured from a tide gauge at Pointe-des-Galets on Reunion Island (020° 55' 00.0" S, 055° 17' 00.0" E). The measured tidal time series of the acquisition during a tsunami event on August 13, 2021 is shown on the Fig. 2. This signal is significantly perturbed by the transit of the tsunami wave (at the beginning of the time series), but we can still automatically extract the quality parameters of the water level measurements presented in the previous subsection.

Table 3. Preference parameters for the oceanographic end-use

Context	Oceanographer	POSACC	SOUACC	Cover.	Sensors	Sed.	Ocean data	Tide duration	Tide gap	Tide outl.	
10/100m	G / A ($b_{2,10/100m}$)	< 2m	< 0.5m	200%	MBES	4	MTO + CTD + tide + current	≥ 1 year	0%	0%	
	A / B ($b_{1,10/100m}$)	< 5m + 5%	< 2m	100%	SBES + Side-scan sonar	2	MTO + CTD + tide	≥ 30 days	< 10%	< 3%	
1000m	G / A (1000m) ($b_{2,1000m}$)	< 2m	< 0.5m	200%	MBES	4	MTO + CTD + current	.	.	.	
	A / B (1000m) ($b_{1,1000m}$)	< 5m + 5%	< 2m	100%	SBES + Side-scan sonar	2	MTO + CTD	.	.	.	
10/100m	Veto (G & A) ($b_{1,10/100m}^v$ & $b_{2,10/100m}^v$)	< MTO + CTD	≤ <30 days	.	.	
	Veto (G & A) ($b_{1,1000m}^v$ & $b_{2,1000m}^v$)	< MTO + CTD	.	.	.	
10/100m	Weights	3/46	2/46	4/46	1/46	5/46	8/46	9/46	7/46	7/46	
1000m	Weights	3/21	2/21	4/21	1/21	5/21	6/21	.	.	.	
10/100m	Majority threshold						23/46				
1000m	Majority threshold						10.5/21				

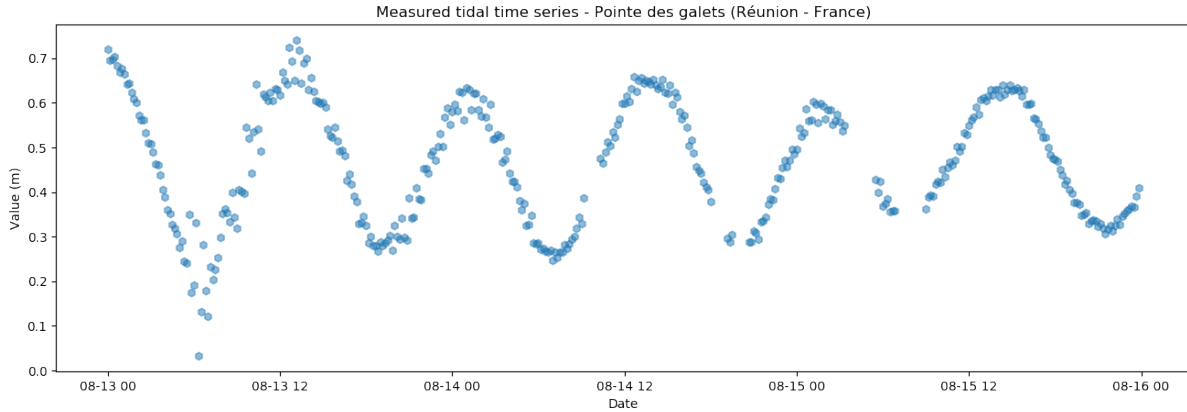


Fig. 2. Measured tidal time series at Pointe-des-Galets (Réunion - France) during tsunami event

The first of these quality parameters is the acquisition time ΔT . This quantity, in days, is obtained by the difference between the end date T_{end} of acquisition and the beginning date of acquisition T_{start} ($\Delta T = T_{end} - T_{start}$). For our example we have computed the value of three days.

The second quality parameter concerning the water level data is the percentage of missing periods in the time series. The tide gauge integration period ($\Delta_{integration}$), during which the tide gauge will average several measurements, in order to keep only one value for that period, being of 10 minutes, we consider that there is a missing period as soon as the temporal difference between two consecutive measurements t_i is strictly higher than 10 minutes.

Fig. 3 shows this quality parameter applied to the Pointe-des-Galets tide gauge data, a value of 1 indicating when data is present and value 0 when there is no data on the time series. In this case 5 missing periods should be noted for a total of 11% of missing data over the total duration of the time series. This quality parameter is very relevant to detect a possible sensor failure or a problem with the data link to the time series data storage servers.

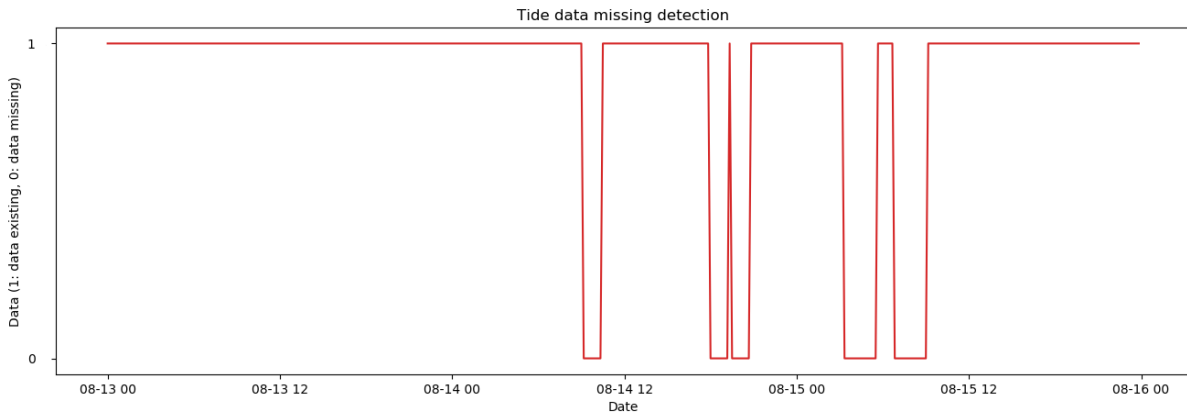


Fig. 3. Tide data missing for the Pointe-des-Galets (Réunion - France) tidal time series

Another quality parameter that can be automatically extracted from the tide gauge data is the presence of outliers or strong variations in the signal that could disturb future modeling. This parameter is crucial to detect extreme events such as tsunami or earthquakes, and it also ensures the quality of the data that will be used to model the tidal phenomenon over the long term. In order for this modeling to be as accurate as possible, it is necessary to remove these exceptional events that do not correspond to a normal tidal situation.

To compute the presence of an outlier, the difference between the absolute value of the local variation of the measurement and the absolute value of the local variation of the predicted tide is used. If this difference is above a certain threshold then the measurement is considered an outlier. This threshold is defined by the uncertainty related to the tide gauge measurement (5 cm) and a maximum acceptable atmospheric pressure variation of 10 hPa over one hour (i.e. a variation of 10 cm over one hour), which corresponds to strongly low-pressure weather events (a threshold at 7 cm in the case of our $\Delta_{integration}$).

Fig. 4 shows this quality parameter applied to the Pointe-des-Galets tide gauge data as well as the representation of the threshold by the blue horizontal line, in this case 9% of strong variations periods should be noted.

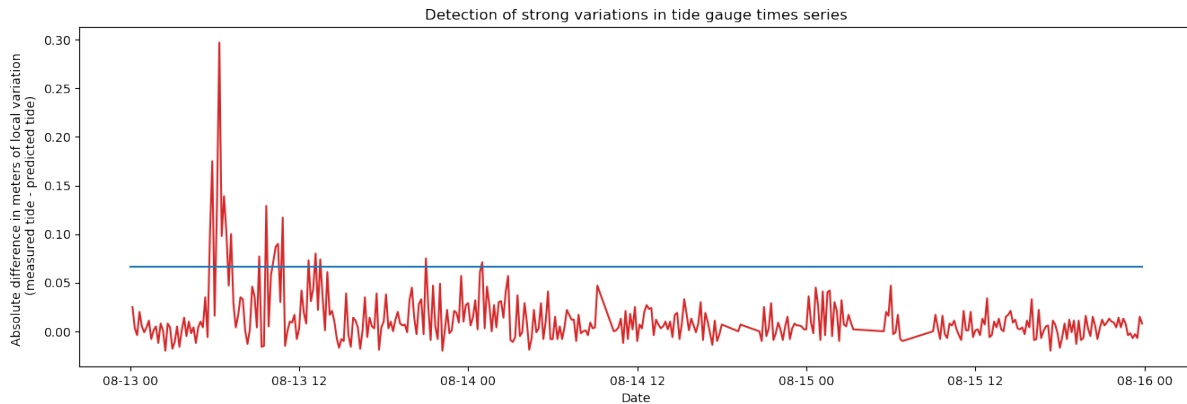


Fig. 4. Detection of strong variations for the Pointe-des-Galets (Réunion - France) tidal time series

We can now apply this automatic determination of quality parameters on real data from the Shom's bathymetric database.

4.5 Studied survey campaigns

The five hydrographic surveys considered in this application, shown in Fig. 5, have very different characteristics and geographic locations. They are thus representative of the types of surveys acquired at Shom and the kinds of data that ocean data experts may encounter. These data are available on the Shom diffusion portal, <https://www.data.shom.fr>, or will be once validated and integrated in the bathymetric database.

S2015001 [26] is an SBES (with side-scan sonar to perform a full obstruction detection coverage) exploration survey carried out on the Anse de Koumac (in New Caledonia, France). The average surveyed depth is 5 meters. The purpose of this survey was to measure a shoal in the bay that is a potential danger to navigation.

S2018057 is a classical MBES hydrographic survey which aims to improve the bathymetric knowledge of the traffic separation scheme off Ouessant Island (France). With an average depth of 100 meters, this survey allows for the updating of nautical works and more particularly nautical charts, but also to measure precisely the Ouessant ditch.

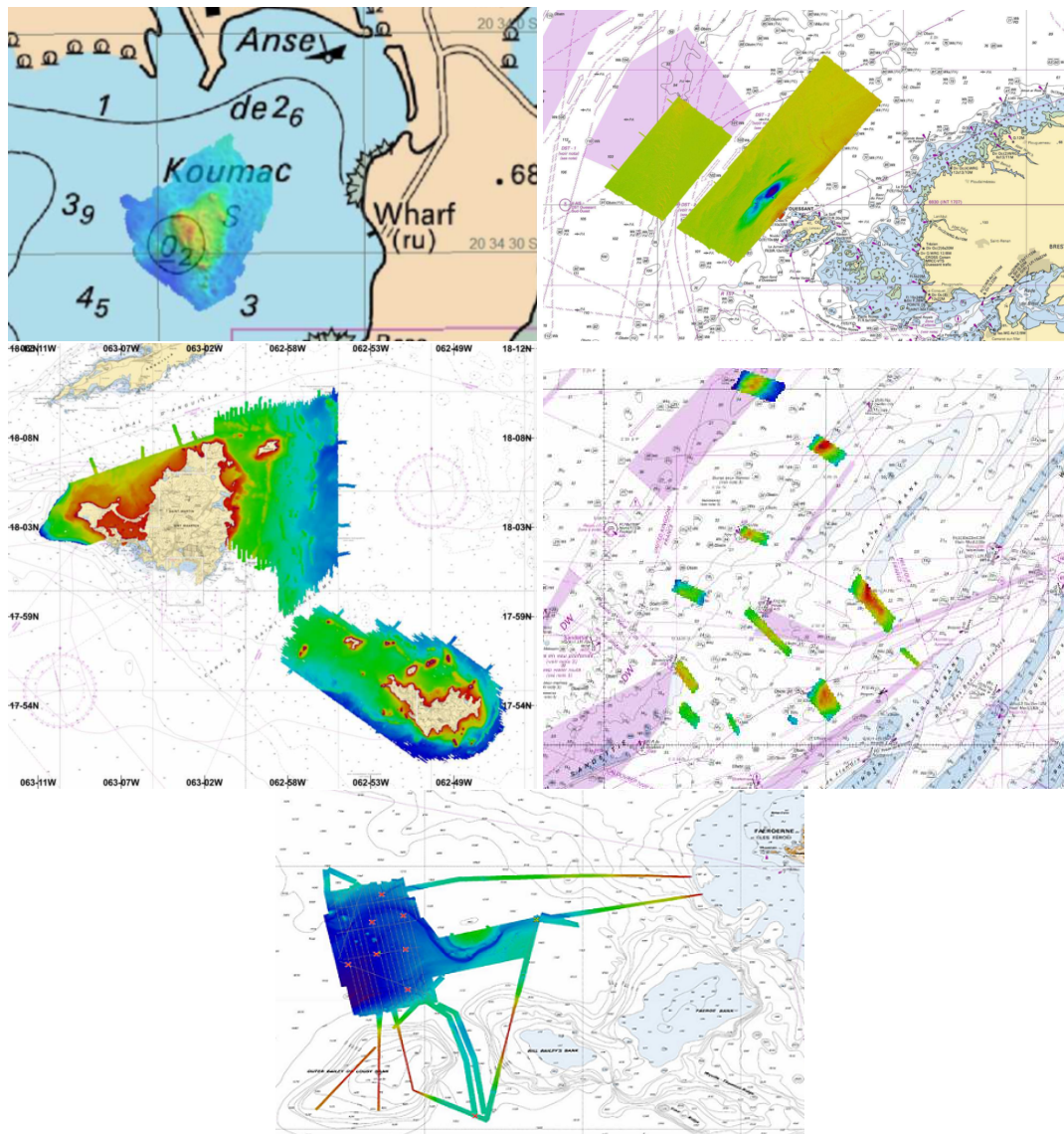


Fig. 5. Geographic extent of the five studied hydrographic surveys. From left to right and top to bottom: S2015001 (New Caledonia, France), S2018057 (Ouessant Island, France), S2019011 (Saint-Martin and Saint-Barthélemy, France), S2019029 (English Channel, France), and S2019026 (Faroe Islands).

S2019011 is an airborne LiDAR topo-bathymetric survey of Saint-Martin and Saint-Barthélemy Islands (France). The average depth is about 20 meters. This survey was designed to provide a new land-sea reference after Hurricane Irma in 2017.

The objective of the S2019029 survey is to follow the sediment dynamics in the English Channel near Dunkerque (France) and therefore the displacement of dunes that could be possible dangers for coastal navigation off Dunkirk, but also to improve knowledge about the hydro-oceanic dynamics of this region. Its average depth is 30 meters.

S2019026 is an offshore MBES survey which has been performed in the Atlantic Ocean off the coast of the Faroe Islands. Its average depth is 1400 meters. This survey was carried out in order to know the acoustic environment as precisely as possible, with a lot of oceanographic and sedimentological measurements.

For all the described surveys we indicate in the following tables in which context (average depth range) they are associated by the symbol \approx .

Each of the three experts has provided the list of quality parameters that need to be taken into account in the overall assessment of the surveys. This analysis is given as input for the quality parameters calculation (C) step of Fig. 1, together with the survey data, and meta-data. This step then generates the data of Table 4, necessary for the survey evaluation process.

Table 4. Performance table: output of the “Quality extraction / calculation” step

Survey (depth range)	CATZOC	POSACC	SOUACC	Cover.	Sensors	Sed.	Ocean data	Tide duration	Tide gap	Tide outl.
S2015001 ($\approx 10m$)	B	< 2m	< 0.50m	100%	SBES + Side-scan sonar	0	MTO + CTD	<30 days	0%	0%
S2018057 ($\approx 100m$)	B	< 5m + 5%	< 0.50m	200%	MBES	4	MTO + CTD + tide + current	≥ 30 days	0%	0%
S2019011 ($\approx 10m$)	B	< 5m + 5%	< 0.50m	200%	LiDAR	0	MTO	No tide	No tide	No tide
S2019029 ($\approx 10m$)	B	< 2m	< 0.50m	50%	MBES	3	MTO + CTD + tide	≥ 30 days	0%	0%
S2019026 ($\approx 1000m$)	B/C	< 20m + 10%	20m	200%	MBES	5	MTO + CTD + current	No tide	No tide	No tide

4.6 Interpretable global evaluation and explainable recommendations (D)

Table 4 serves as input for the “Multi-Criteria assessment” step, together with the previously identified preference parameters of the three experts. The MR-Sort overall assessment method generates for each end-use an overall quality evaluation, which depends on the configuration of the process. The output of this assessment step is shown in Table 5.

Table 5. Overall assessments: output of the “Multi-Criteria assessment” step

Survey (depth range)	Acoustic use	Hydrographic use	Oceanographic use
S2015001 ($\approx 10m$)	Bad	Acceptable	Bad
S2018057 ($\approx 100m$)	Good	Good	Good
S2019011 ($\approx 10m$)	Bad	Acceptable	Bad
S2019029 ($\approx 10m$)	Good	Bad	Acceptable
S2019026 ($\approx 1000m$)	Good	Good	Good

Fig. 6 shows the acceptable-bad separation profile (in black) on the left and the good-acceptable separation profile (in black) on the right for the hydrographic use and the 10m average depth context. Each criterion is represented by a vertical axis, for which the preferred values are towards the top of the figure. The 2 extreme (preferred and less preferred) values of the scales are labeled. The red line (with the triangles) represents the quality levels for the survey S2019029. It can be noted that in spite of very good quality level (as on Sensors or

SOUACC parameters), the survey will always be considered as a bad, because its hydrographic coverage quality parameter reaches a veto in this case.

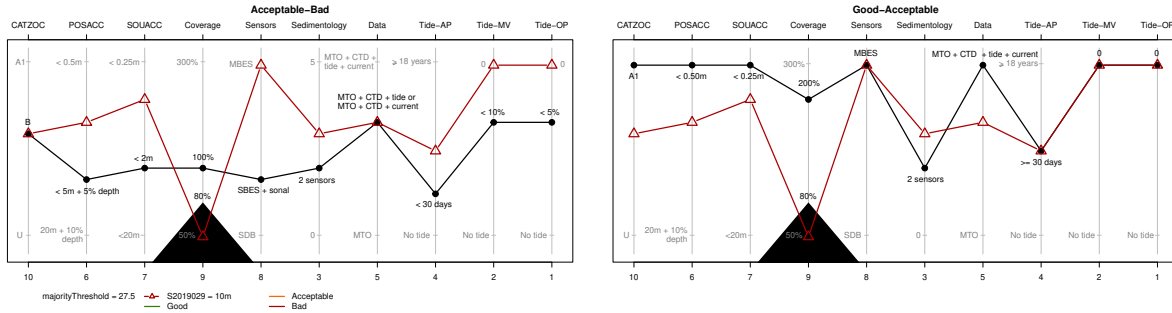


Fig. 6. S2019029 ($\approx 10m$) evaluated "Bad" for the hydrographic use

Fig. 7 represents the quality assessment for the S2019026 (green line with squares) and S2015001 (red line with triangles) surveys for acoustic use. Here also survey S2015001 has a bad quality level because its Sedimentology reaches a veto value. On the other hand, survey S2019026 has a good quality level because its quality parameters are above the separation profile for the most important criteria (Sedimentology with a weight of 9/46, Sensors with a weight of 9/46 and Coverage with a weight of 7/46). With these three parameters alone, the survey already reaches an overall concordance of 25/46, thus exceeding the majority threshold set at 23/46 for the acoustic profile.

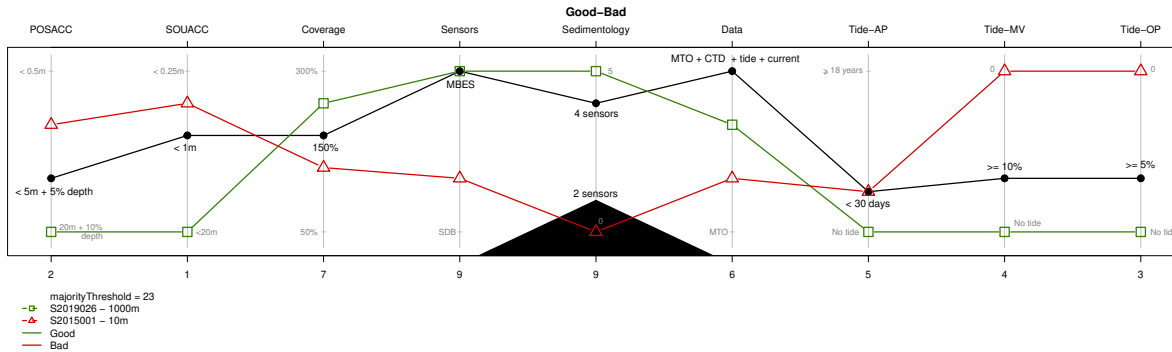


Fig. 7. S2019026 ($\approx 1000m$) (resp. S2015001 ($\approx 10m$)) evaluated "Good" (resp. "Bad") for the acoustic use

Finally, Fig. 8 shows the acceptable-bad (left) and good-acceptable separation profiles for oceanographic use for survey S2019029 (10m mean depth). We can thus see that, in regard to the quality parameters, the survey is quickly classified as acceptable, having only one quality criterion (hydrographic coverage) below the acceptable-bad separation profile. Contrarily, it doesn't have enough quality criteria with a strong enough weight for this survey to be classified as good as can be seen on the right figure (overall concordance of 20/46 for a majority threshold of 23/46).

As mentioned earlier, this process also allows recommendations to be made to improve the quality of future data acquisitions. Thus, still looking at Fig. 8, we can see that it would be enough to add a system to acquire

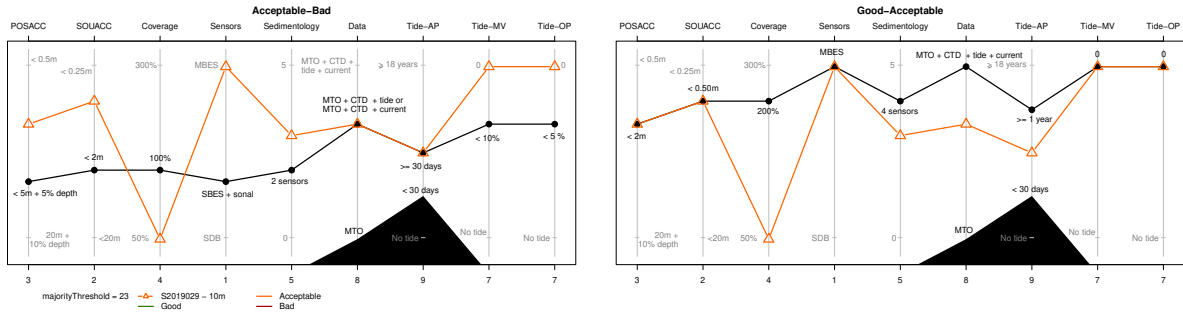


Fig. 8. S2019029 ($\approx 10m$) evaluated "Acceptable" oceanographic use

sedimentology data (the sedimentology criterion rising to 5) or to measure the current (the ocean data criterion being then at its maximum) for the coalition of quality criteria to exceed the majority threshold. We can see the strength of the interpretability of this type of method, which provides rapid access to recommendations for decision makers.

Applying multi-criteria decision methods to aggregate data quality assessments of multiple sensors is relatively novel. As a consequence, comparing obtained results with those of existing state of the art works, which do not examine that kind of problem is at the moment an exceptional challenge. Such current state of fact derives from the unavailability of comparable benchmarks and equivalent approaches in terms of data features, applied multi-criteria decision approach, and contextual inter-dependencies between quality parameters.

5 CONCLUSION AND PERSPECTIVES

In this work we propose a generic process to evaluate the quality of data or information taking into account the context of use and the preferences of work-domain experts. The process leads to a global evaluation of the quality of these data, and also allows to explain the evaluation, through simple rules understandable by the user. The proposal is experimented on a real case and real datasets, where the quality of hydrographic surveys for different uses has to be evaluated.

This process could be deployed to other domains, where the quality of datasets has to be evaluated for various uses. Since the proposed process is transparent and explainable, and the generated evaluations are interpretable, users will have a high level of confidence in the systems that implement it. The possibility of making recommendations for improvements of the data acquisition is also an important point of this proposal, as it allows, depending on the use, to arrive at qualitatively more interesting data, while minimizing the impact on the acquisition.

Our approach has assumed that preferences defined by work-domain experts are certain, i.e. quality and preference parameters are unique and precise. Nevertheless, this is not the case when imprecision should be taken into account and as a consequence, intervals of approximate validity have to be included to represent those parameters. Such a use case would require complementary analysis to include those variable elements.

We are aware of other limitations of our process, which are listed below:

- The need to have an expert at hand to configure the process.
- The risk that this expert cannot answer the questions during the preference elicitation phase, while there is not enough learning data available for a batch learning.
- The fact that missing quality evaluations lead to very pessimistic overall evaluations.

Different perspectives are possible for this work. Currently, as presented in the article in sections (section 3.1, section 3.2 and section 3.4), a number of operations are performed manually : by the domain expert in the process and in the explanation of the results provided by the process. So in order to automate the process, it would be necessary to experiment with incremental indirect preference elicitation approaches, i.e. where usage experts are confronted with examples of data that they evaluate according to their expertise. It would thus be possible to design a learning base where a domain expert has labelled numerous samples. This labelled base is then used by a learning algorithm, which determines automatically the parameters of the preference model. We could also automate the output of the process by providing a clear explanation for any user through rule generation algorithms. Another avenue of future research consists in integrating other notions of uncertainty (on the data in particular) into the MCDA model, as well as the handling of missing evaluations in the quality criteria. Finally, to show the scope of the proposed approach, it could be interesting to apply the process to other domains.

REFERENCES

- [1] M. Aljumaili, R. Karim, and P. Tretten. Metadata-based data quality assessment. *VINE Journal of Information and Knowledge Management Systems*, 2016.
- [2] M. Aljumaili, R. Karim, and P. Tretten. Data quality assessment using multi-attribute maintenance perspective. *International Journal of Information and Decision Sciences*, 10(2):147–161, 2018.
- [3] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3):1–52, 2009.
- [4] D. Bouyssou and T. Marchant. An axiomatic approach to noncompensatory sorting methods in MCDM, I: the case of two categories. *European Journal of Operational Research*, 178(1):217–245, April 2007.
- [5] D. Bouyssou and T. Marchant. An axiomatic approach to noncompensatory sorting methods in MCDM, II: more than two categories. *European Journal of Operational Research*, 178(1), April 2007.
- [6] D. Bouyssou, T. Marchant, M. Pirlot, A. Tsoukiàs, and P. Vincke. *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*. International Series in Operations Research and Management Science, Volume 86. Springer, Boston, 1st edition, 2006.
- [7] A. de Souza Barbosa and M. A. G. Shayani, R. A. and de Oliveira. A multi-criteria decision analysis method for regulatory evaluation of electricity distribution service quality. *Utilities Policy*, 53:38–48, 2018.
- [8] Fernando Gualo, Moisés Rodríguez, Javier Verdugo, Ismael Caballero, and Mario Piattini. Data quality certification using iso/iec 25012: Industrial experiences. *Journal of Systems and Software*, 176:110938, 2021.
- [9] IHO. Iho transfer standard for digital hydrographic data. Technical report, IHO, 2000. Special Publication N°57, 3 edition.
- [10] IHO. Iho standards for hydrographic surveys. Technical report, IHO, 2020. Special Publication N°44, 6 edition.
- [11] R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: Preferences and value tradeoffs*. J. Wiley, New York, 1976.
- [12] A. Klein and W. Lehner. Representing data quality in sensor data streaming environments. *Journal of Data and Information Quality (JDIQ)*, 1(2):1–28, 2009.
- [13] G. Lacagnina, C. and Peng, R. R. Downs, H. Ramapriyan, I. Ivánová, A. Ganske, D. Jones, L. Bastin, L. Wyborn, I. Bastrakova, et al. International community guidelines for sharing and reusing quality information of individual earth science datasets. *Earth and Space Science Open Archive*, 2021.
- [14] J. Le Deunf, A. Khannoussi, L. Lecornu, P. Meyer, and J. Puentes. Automatic data quality assessment of hydrographic surveys taking into account experts' preferences. In *OCEANS 2021: San Diego – Porto*, pages 1–10, 2021.
- [15] A. Leroy, V. Mousseau, and M. Pirlot. Learning the parameters of a multiple criteria sorting method. In R. Brafman, F. Roberts, and A. Tsoukiàs, editors, *Algorithmic Decision Theory*, volume 6992, pages 219–233. Springer, 2011.
- [16] D. McGilvray. *Executing data quality projects: Ten steps to quality data and trusted information (TM)*. Academic Press, 2021.
- [17] V. Mousseau, J. Figueira, and J.P. Naux. Using assignment examples to infer weights for ELECTRE TRI method: Some experimental results. *European Journal of Operational Research*, 130(2):263–275, April 2001.
- [18] V. Mousseau and R. Słowiński. Inferring an ELECTRE TRI model from assignment examples. *Journal of Global Optimization*, 12(2):157–174, 1998.
- [19] A. Ngo The and V. Mousseau. Using assignment examples to infer category limits for the ELECTRE TRI method. *JMCDA*, 11(1):29–43, November 2002.
- [20] A-L. Olteanu and P. Meyer. Inferring the parameters of a majority rule sorting model with vetoes on large datasets. In *DA2PL 2014: From multiple criteria Decision Aid to Preference Learning*, pages 87 – 94. Ecole Centrale Paris and Université de Mons, 2014.
- [21] J. Puentes, L. Lecornu, and B. Solaiman. Data and information quality in remote sensing. In *Information Quality in Information Fusion and Decision Making*, pages 401–421. Springer, 2019.

- [22] H. K. Ramapriyan, G. Peng, and C. Moroni, D.and Shie. Ensuring and improving information quality for earth science data and products. *D-Lib Magazine*, 23, 2017.
- [23] C. Ranasinghe and C. Kray. Location information quality: A review. *Sensors*, 18(11):3999, 2018.
- [24] B. Roy. *Multicriteria Methodology for Decision Aiding*. Kluwer Academic, Dordrecht, 1996.
- [25] K. Sha and S. Zeadally. Data quality challenges in cyber-physical systems. *Journal of Data and Information Quality (JDIQ)*, 6(2-3):1–4, 2015.
- [26] Shom. Lot bathymétrie s201500100-1. https://services.data.shom.fr/geonetwork/srv/api/records/LOTS_BATHY_S201500100-1, 2022. Accessed: 2022-09-27.
- [27] O. Sobrie, V. Mousseau, and M. Pirlot. Learning a majority rule model from large sets of assignment examples. In *Algorithmic Decision Theory*, pages 336–350. Springer Berlin Heidelberg, 2013.
- [28] O. Sobrie, V. Mousseau, and M. Pirlot. A population-based algorithm for learning a majority rule sorting model with coalitional veto. In H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecek, Y. Jin, and C. Grimme, editors, *Evolutionary Multi-Criterion Optimization*, pages 575–589, Cham, 2017. Springer International Publishing.
- [29] I. Todoran, L. Lecornu, A. Khenchaf, and J. Le Caillec. A methodology to evaluate important dimensions of information quality in systems. *Journal of Data and Information Quality (JDIQ)*, 6(2-3):1–23, 2015.
- [30] S. Trendowicz, A.and Kopczyńska. Adapting multi-criteria decision analysis for assessing the quality of software products. current approaches and future perspectives. *Advances in Computers*, 93:153–226, 2014.
- [31] L. Visengeriyeva and Z. Abedjan. Anatomy of metadata for data curation. *Journal of Data and Information Quality (JDIQ)*, 12(3):1–30, 2020.
- [32] N. K. Yeganeh, S. Sadiq, and M. A. Sharaf. A framework for data quality aware query systems. *Information Systems*, 46:24–44, 2014.