16th International Conference on Information Fusion
Istanbul, Turkey, July 9-12, 2013

Information quality evaluation in fusion systems

Ion-George Todoran†, Laurent Lecornu†, Ali Khenchaf‡, Jean-Marc Le Caillec†
† Institut Mines Telecom - Telecom Bretagne, CNRS UMR 6285 Lab-STICC,
Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3 - France
‡ ENSTA Bretagne, CNRS UMR 6285 Lab-STICC
2 rue François Verny, 29806 BREST Cedex 9
Email: {iongeorge.todoran, laurent.lecornu, jm.lecaillec}@telecom-bretagne.eu ; ali.khenchaf@ensta-bretagne.fr

***Abstract*—Advances in new information and communication technology have considerably increased the quantity of available data and permitted the implementation of more and more complex processing techniques. As a result of these changes it appeared the need of evaluating the information fusion system quality and to propose to the user the information accompanied with confidence coefficients. The information fusion systems used nowadays are complex systems that are difficult to be evaluated. To overcome this problem we propose a new quality evaluation methodology based on the decomposition of the information fusion system in its elementary modules that allows the evaluation of the quality at two levels. The first one (global) that describes the entire information fusion system and the second one (local) for each elementary module. As data and information change over time, this decomposition allows to directly evaluate the global quality of the information fusion system using the local quality evaluation.**

***Index Terms*—Information quality, quality evaluation, information fusion system, quality measures**

I. Introduction

With the evolution of the information and communications technology, institutions are confronted to some new problems and questions about the quantity and the quality of the data, information and knowledge that they possess. An information system has the role of collecting, processing, storing, analysing and disseminating information for a specific purpose [1]. The large amount of available data, together with the complex techniques of processing it, has in many cases a negative impact on the quality of information, mainly because of a misconception of the information system [2].

The actual information systems do not offer to the user a degree of confidence for the information proposed to him by the system. In this paper we consider that the degree of confidence has the same importance as the proposed information and so, the notions of information and information quality coexists. Based on this assumption we propose a new methodology of quality evaluation as detailed below. Furthermore, to assess the quality of the information system we insist on the need of knowing the information quality transformations through the system for being able to present to the user a reliable estimation of the information quality.

The quality evaluation is by its nature a multidimensional problem [3], [4]. Data and information quality evaluation has become a central problem in domains of research like organisations management [3], [4], web information systems [5] or information fusion [6], [7]. An important amount of research has been done in evaluating the quality of the data used by the information system as an input [4]. In contrast with a few years ago, when only one database was used, containing homogeneous data, in the present an information system uses distributed databases containing a huge volume of complex and heterogeneous data. Because an information system helps the user in his tasks, one of the most important problems is how to evaluate the quality of the information system proposed to him. Thus, it is not so important to optimise individually each part of an information system (databases, processing algorithms, human-machine interface, etc.) but to optimise the couple {information system, user} with respect to the task that has to be done. From users perspective we can observe the need of having pertinent, accurate, complete, consistent, up-to-date information, presented in an easy to understand manner [7]. Unfortunately the passage from data to information is not evident and the use of a large quantity of data having a good quality does not guarantee that the user will have at his disposal good quality information.

At this moment a methodology that permits the evaluation of a general information system is not available. In practice, the performance of an information system is usually given in terms of a precision-recall curve or a ROC graph [8]. Even though this type of evaluation is very useful for classification problems, in the case of a complex information system it is not adapted mainly because it does not allow an instantaneous performance evaluation. A methodology for evaluating the quality of a complex information system needs to be generally enough for being used in all the cases and in all domains of application (medical, financial, defence, management, etc.).

To overcome this limitation we propose a new methodology of quality evaluation in the case of an information fusion system. We build our methodology based on the internal structure of the information fusion system. We adapt the quality criteria and measure proposed in the literature to the case of a complex system. We introduce two different levels of quality evaluation: local - measuring the quality of a module (as part of the information fusion system) and global - measuring the quality of the information fusion system. The idea behind this decomposition is to calculate the global quality from the local one. This allows to update the output information quality based on the information quality changes in the system. Moreover, we show that this methodology is adapted to the systems working in a dynamic context.

978-605-86311-1-3 ©2013 ISIF

When this quality evaluation is rigorously done it will help two entities : the information system constructor by choosing the processing modules and the final user by allowing him to better understand the used system and the information delivered to him.

This paper is structured as follows, the second section presents a review of the information quality evaluation in three fields of research: organisations management, web information systems and information fusion. Our methodology of information quality evaluation in the case of an information fusion system is presented in section 3. This methodology is developed in three steps : the first one is the information quality evaluation in a punctual place of the information fusion system, the second one is the development of a strategy of information quality evaluation in the output of a processing module knowing the information quality in its input and finally, the third step is the evaluation of the information fusion system quality, named global information quality evaluation. Finally, conclusions will be drawn and the future works are presented.

II. Review of information quality evaluation

Evaluating the quality of an information system is a difficult problem. This difficulty is principally due to the generation and processing techniques that deliver complex data and information.

Data and information quality evaluation at different stages of an information system is not an easy task. Also, there is the question of the final user (or more generally of the final users). Depending on the application domain and his expertise, different levels of information quality can be considered as satisfying.

The decisions proposed to the user can be seen as information and for being able to qualify them, we have to study their quality. There are different fields of research interested on data and information quality evaluation: organisations management, web information systems, information fusion, etc.

A large quantity of research was oriented towards the data quality evaluation [4]. Recent studies are trying to evaluate data quality by guiding the user for building an information system. For example the tool *PaREn Automatic System Construction Wizard*[1] included in Rapid Miner™helps the user to set up his own classification system based on the characteristics of the database.

In most of the cases data is seen as a product and the management of the data quality is done following the case of product manufacturing management [3], [4]. Moreover, data and information are considered synonymous and the data/information quality dimensions are considered to be the same. Data can be seen as a static resource, but the notion of information is context dependent. The context depends on the domain of application and on the user who will use this information. Because of this dependence on the context, the information quality has its characteristics depending on the domain of application. A brief introduction to information quality in 3 domains of research will be further considered: organisations management, web information systems and information fusion.

[1]http://rapid-i.com/content/view/240/1/lang,en/

TABLE I
Quality dimensions proposed by Wang and Strong [3]

Intrinsec	Contextual	Representational	Accessibility
Accuracy	Value-Added	Interpretability	Access
Believability	Relevancy	Ease Understanding	Security
Objectivity	Timeliness	Repr. consistency	
Reputation	Completeness	Repr. conciseness	
	Data amount	Manipulability	

TABLE II
Web information systems data and information quality [5]

Content	Technical	Intellectual	Instantiation
Accuracy	Availability	Believability	Data amount
Completeness	Latency	Objectivity	Repr. conciseness
Customer support	Price	Reputation	Repr. consistency
Documentation	QoS		Understandability
Interpretability	Response		Verifiability
Relevancy	Security		
Value-added	Timeliness		

A. The case of organisations management (Wang and Strong model)

Wang and Strong [3] proposed in 1996 one of the first frameworks for data/information quality evaluation. At the end of their research they arrived at 16 dimensions, classified in 4 categories, for characterising data/information quality. In table I are presented these 16 dimensions of data and information quality.

Although the framework of Wang and Strong appears to be generally usable, it is more adapted for the domain of organisations management. As they consider the information system to be a black box, the quality evaluation is done only at the output of the system.

B. The case of web information systems

Nowadays, this field of research has the most part of applications in the domain of Web search of information. With the increasing volume of data available on the Internet and with the technology advancements that allows high-speed connection to every user, the quality of information delivered by the web information system is of great importance. In [5] general criteria for information quality evaluation were presented. These criteria (compiled from multiple sources) were classified in 4 categories and they are presented in table II.

As it was mentioned in [5] not all of these criteria are independent and not all of them should be used at the same time. Compared with the case of organisations management, in this case the quality evaluation is oriented towards the service

TABLE III
QUALITY OF INFORMATION FOR AN INFORMATION FUSION SYSTEM [7]

Content	Sources	Presentation
Availability	Reliability	Understandability
Accessibility	Credibility	Completeness
Timeliness	Relevancy	Timeliness
Integrity	Objectivity	Interpretability
Relevancy	Level of expertise	
	Truthfulness	
	Reputation	

offered by the web information system. As a consequence the user can only use the web information system without having a deep understanding of it and without the possibility of influencing it.

C. The case of information fusion

The information fusion domain is a very active field of research and has as objective the combination of heterogeneous and conflicting information (the case of complex information systems) in order to obtain reliable and complete information. This final information usually is needed for a complex situation assessment (e.g. battle field). The problem of situation assessment is by its nature dynamic and because of this, the quality evaluation of an information fusion system needs to be a dynamic process. In time, the different information sources propose information of different quality levels that need to be fused together to produce a final information. In this field, recent work was carried out addressing the problem of fusing information having different qualities [6], [7].

A methodology of information quality evaluation for an information fusion system was proposed in [7]. Three types of information were depicted in this study, following the same reasoning as in the case of the Wang and Strong model. For each of these types, quality dimensions were proposed, as shown in table III.

Other studies tried to present the information quality impact on situation assessment [9], [10]. In [9] the proposed information quality criteria were: *Timeliness*, *Confidence*, *Accuracy*, *Throughput*, *Cost*, *Security* and *Parsimony*. Again this quality dimensions follow the Wang and Strong model.

D. Critics of the quality evaluation models

The information quality frameworks developed previously follow the model proposed by Wang and Strong in 1996. One of the most important drawback of this model is the supposition that the entire information system can be seen as a black box. This view is adapted only for the case when the final user is interested on the decisions proposed to him by the information system and not on the process of information production. In this way the quality evaluation is done at the beginning of the information system, a data quality evaluation, and at the end of the information system, the quality of the information proposed to the user. Because of the complexity of the information system, this type of information quality evaluation is usually done by using forms and by asking users to give their opinion [11].

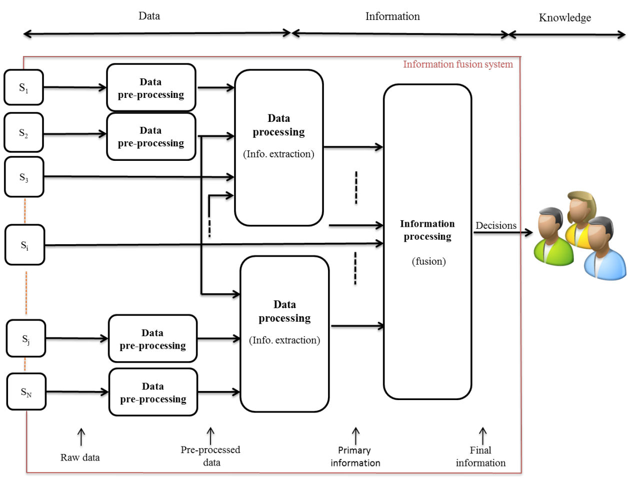

Fig. 1. Data, information and knowledge in an information fusion system

Another observation is that due to their ambiguity of defining information quality criteria, these frameworks cannot be easily applied. This ambiguity is mainly due to the confusion between data and information.

A major limitation of these models is the fact that they are evaluating the mean quality of an information system. As the input data are changing because of a dynamic context, we argue that the user needs information delivered with the instantaneous quality.

To overcome these problems, in the next section we will define a methodology on how to use these quality criteria in the case of an information fusion system. Quality measures will be associated in order to be able to quantify the quality.

III. LOCAL VERSUS GLOBAL INFORMATION QUALITY EVALUATION

First a general information fusion system will be described and afterwards our methodology is presented.

For developing the framework of information quality evaluation we propose the architecture of a general information fusion system, as shown in figure 1. Data, information and knowledge notions will be defined as in [12] and [13] :

- *Data*: are discrete and objective representations of events and entities under observation;
- *Information*: is processed data in order to be put into a meaningful context;
- *Knowledge*: is a synthesis obtained from information combination and evaluation by the user.

Multiple data sources (databases, sensors, experts opinions) that contain raw data are presented as input for the system. Some of this data has to be preprocessed in order to be adapted to the process of information extraction. In this type of information fusion system architecture, the fusion process is done at different levels : data fusion, information fusion. The last fusion process needs to be constructed to deliver final

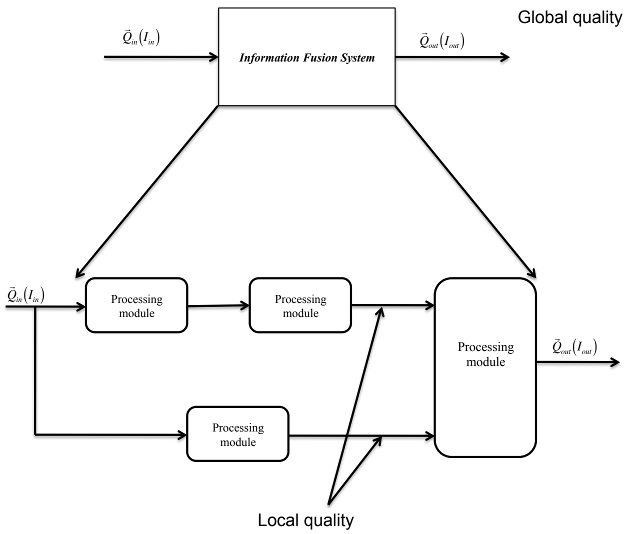

Fig. 2. Local and global quality in an information fusion system

TABLE IV
DATA QUALITY DIMENSIONS WITH THEIR MEASURES

	Quality criteria	Quality measures
Static data quality	Accuracy	standard deviation
	Completeness	proportion of missing values, population completeness
	Accessibility	Time to access, failure rate, time to recovery
	Up to date	Refresh time
	Security	Level of security
	Data amount	Nb of attributes, nb. of entities, volume (bits)
	Consistency	Respect to the format type, proportion of redundancies
Context dependent data quality	Relevancy	User assessment
	Confidence	User assessment, standards
	Objectivity	Expert assessment
	Ease of manipulation	User assessment
	Interpretability	User assessment

information that supports the user(s) in the decision process. Different types of data and information can be observed inside the information fusion system.

As we consider that the information evolves over the time, a natural assumption is to consider that its quality also changes. The information evolution can be caused by new available data or by the modification of some processing modules (e.g. changing some functioning parameters). The quality of the information fusion system is directly dependant on these changes and we want to have an instantaneous estimation of its quality. As the system is too complex, the characterisation of the system cannot be done based only on its inputs and outputs. This leads to an inaccurate evaluation of the system. To overcome this problem we propose the decomposition of the information fusion system in its elementary modules and the quality study after each module, as shown in figure 2. Therefore, we can evaluate the quality changes induced by the modules and the information quality delivered to the user can be assessed. As we are supposing that the quality evolves over time, we can propagate the quality changes from the local point to the output of the system.

We consider that we have access to the processing modules for obtaining the necessary knowledge about the behaviour and the functioning of each of them. This supposition is verified in practice because the processing modules are already built.

A. Local information quality evaluation

The decomposition of the information fusion system in its elementary modules shows that the definition of the quality is different depending on the place in the information fusion system where the quality evaluation has to be done. This definition change is visible when we study the quality at the beginning of the information fusion system chain, where we are talking about data quality and at the end of the information fusion system chain, where we are talking about information quality. As data is different from information at a semantic level, the quality of these two entities is also different. In order to exemplify, the quality dimension *Completeness* is considered : at data level it refers to the degree of values presence or at information level it refers to the degree to which it represents all the real world characteristics needed by the user. Therefore, these quality dimensions should be regarded as quality criteria and for each of this criteria quality measures should be proposed.

The goal is to construct an exhaustive list of quality criteria that will permit the local quality analysis. These criteria are based on the models presented in the previous section. From these models we take the list of criteria and we split it into criteria adapted for evaluating data and information quality. Moreover, we merge the 4 categories of quality dimensions into 2 categories depending on the context dependence. Depending on the processing level, only some criteria will be used. Moreover, associated with these criteria, quality measures will be also presented.

Two points of view for characterising the data quality can be identified: static and context-dependent. The static one refers to the data in general, while the second one puts data in context and describes its usefulness for the current application. In table IV are presented the data quality criteria together with their quality measures.

The information quality criteria can be separated in two categories: objective and subjective. The objective criteria describe information independently on the context of application and on the user, while the subjective dimensions describe the fitness of the information proposed to the user. In table V are presented the information quality criteria for the two categories and the measures for the information quality criteria.

As a general observation, just as in the case of the quality evaluation models presented in the previous section, there are correlations between the quality criteria. For example in the case of data quality criteria, *Security* and *Accessibility* are

TABLE V
INFORMATION QUALITY CRITERIA WITH THEIR MEASURES

	Quality criteria	Quality measures
Objective quality criteria	Correctness	Validity degree of the information
	Currency	Age of information
	Consistency	Degree to which the information is not contradictory
Subjective quality criteria	Completeness	Degree to which all the elementary info. are contained in the final decision
	Relevancy	Degree to which information is applicable and helpful for the task
	Timeliness	Degree to which the currency of information is suitable for use

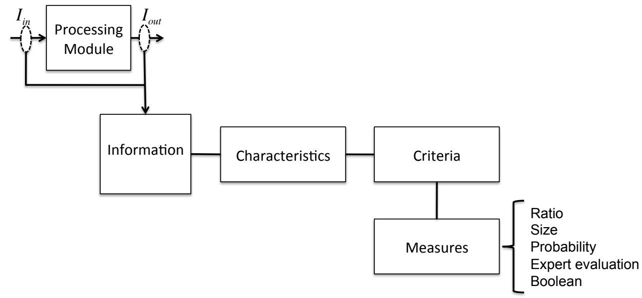

Fig. 3. Information quality formal representation

criteria characterised by a negative correlation meaning that a high level of security implies a difficulty of having a fast accessibility. Another observation is that the quality evaluation is done independently for each criteria.

For being able to use the quality criteria in practice, we need to associate to each of them quality measures in order to obtain quantitative values of the quality level. Usually the quality measures encountered in practice are expressed by real numbers in the unit interval, for example probability measures, evidence measures, etc. For example, in the case of an image the quality criteria *Data amount* can posses as quality measures: the image size (lines; columns), the size in number of bits.

The numerical values are not necessarily adapted for all quality criteria and for the context-dependent criteria, in the most cases linguistic representations are more adapted. This is because these quality values are usually assessed by an expert and for this he uses natural language. For this reason the mathematical framework for developing quality measures will be considered the generalized information theory. In this general framework probability measures, possibility measures and evidence functions coexists. The description of the mathematical framework is presented in [14].

We propose a formal representation of the information quality in figure 3. Depending on the module characteristics and the context of application not all quality criteria will be used to describe the information quality. These characteristics are expressed by the information value delivered by the module and by the meaning of that information. For example consider the context of image processing with a module implementing an image de-noising and another implementing an image segmentation. The two of them will deliver an image in the output, but the semantic content of the two images will be different implying the use of different quality criteria.

The evaluation of this criteria will be made using quality measures. To exemplify, take the case of a processing module implementing the position estimation of a vehicle and delivering this information directly to the user. In the output the information value will have 3 real numbers corresponding to the 3 spatial dimensions. The quality of this output will be evaluated using the following criteria with their specific measures :

- *Correctness*: the precision for each spatial dimension;
- *Currency*: the age of the information.

In the case when the user needs other information as the identity of the vehicle, additional quality criteria are necessary as *Completeness*.

By considering the case of a processing module, some question can be raised regarding the quality evaluation. The first question is how to define the quality criteria for describing the input quality, Q_{in}, and the output quality, Q_{out} of the processing module. Can the input and output information quality be evaluated using the same quality criteria? And if so, are these criteria defined in the same way? The second question is how can the quality evaluation be obtained at the output of a module, knowing the quality in its entry and having a characterisation of the module? For answering these two questions, we propose in the next subsection to model the processing module using a $\mathbf{Q_f}$ "function" that captures the information quality transfer through the system taking into account the module's intrinsic quality.

B. Processing module influence on the quality

As our interest is to model the influence of a processing module on the quality, we will study not the information exchanges but the quality exchanges. In figure 4 is represented a processing module of an information fusion system, having as input the information I_{in} and as output the information I_{out}. Together with the information flow is also represented the quality flow. The quality evaluation, represented by the $\mathbf{Q_{eval}}$, can be assessed by using the quality criteria described in table IV or in table V. The aim is to investigate the possibility of evaluating the output information quality based on the input information quality and not using a separate evaluation of the type $\mathbf{Q_{eval}}$. For doing this we introduce a function ($\mathbf{Q_f}$) characterising the module influence on the quality. We name this function: quality transfer function.

As the notion of quality is multidimensional, the function $\mathbf{Q_f}$ is multidimensional. As an example we assume that at the entry of the module the quality criteria used to characterise the input information quality are $\{Cr_1, Cr_2, Cr_4\}$. For each criteria, measures to express the quality values are associated, see figure 5.

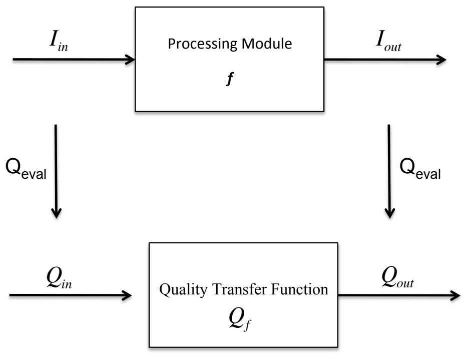

Fig. 4. Information quality transfer through a processing module

We assume that $\{Cr_2, Cr_3\}$ are the criteria that characterize the output information quality. Some of these criteria were also used in the input $\{Cr_2\}$ and some of them were not used $\{Cr_3\}$. Moreover, depending on the information to describe, these quality criteria could be described by different quality measures in the input and in the output of the module.

$$Q_{in} = \begin{pmatrix} Cr_1 : \{m_1^1, m_1^2\} \\ Cr_2 : \{m_2^1\} \\ Cr_4 : \{m_4^1\} \end{pmatrix} \xrightarrow{Q_f} Q_{out} = \begin{pmatrix} Cr_2 : \{m_2^1\} \\ Cr_3 : \{m_3^1\} \end{pmatrix}$$

Fig. 5. Example of a quality transformation function

The importance of this quality transformation function is to characterise the influence of the processing module on the information quality. We assume that we have a complete knowledge of each of the processing modules. This knowledge needs to be expressed by the quality criteria used at the input and at the output of the module. The use of binary vectors tacking unitary values for each criteria used and zero value for the unused criteria, is an example of representation. Taking the example presented in figure 5, in the case of an information quality evaluation (see table V), we have: $Q_{in} = [1\ 1\ 0\ 1\ 0\ 0]$ and $Q_{out} = [0\ 1\ 1\ 0\ 0\ 0]$. In this example the input quality criteria are *Correctness*, *Currency* and *Completeness*, while the output quality criteria are *Correctness* and *Consistency*. Moreover, for each of the input/output criteria, we need to have the quality measures adapted to be used.

1) Analytical evaluation of $\mathbf{Q_f}$*:* At this point we have the input-output quality relation, but we do not have the relation between the quality values, measured by the quality measures. We begin with an example of a processing module implementing a two-class Bayesian classification indicating the presence or the absence of a signal. For the simplicity of the example let us suppose that each input observation $\mathbf{X}$ has a gaussian probability distribution given by the relation: $\mathbf{X} = \mathbf{S} + \mathbf{N}$, with $\mathbf{S}$ a binary variable tacking the values 0 or A with equal probability and $\mathbf{N}$ the noise described by a standard gaussian probability distribution. In figure 6 is represented this classification problem. The classification of the input is done by comparing it to a threshold: if the value is inferior to the threshold value, the observation belongs to class 0 (no signal detected), otherwise class 1 is designated (signal detected).

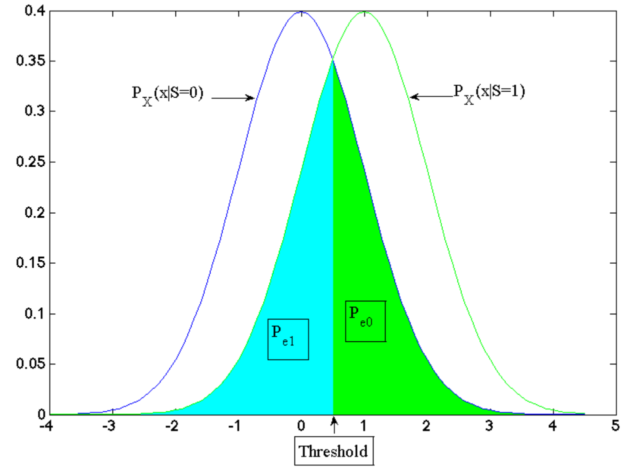

Fig. 6. Two-class Bayesian classifier

In the case of this example the input quality is given by the noise level and the output quality by the error probability which has an analytical formula : $P_{err} = \frac{1}{2}(P_{e0} + P_{e1}) = Q(\frac{A}{2})$, where $Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{+\infty} \exp(-\frac{u^2}{2}) du$. In conclusion we have an analytical formula to calculate the output information quality based on the input information quality and having the knowledge on the processing module: two-class Bayesian classifier, the value of the threshold.

The general expression of the analytical relation between the input quality and the output quality will be: $\vec{Q}_{out} = \mathbf{Q_f}(\vec{Q}_{in}, I_{in})$. In this equation we express the output quality as depending on the input quality and on the input information value. The dependence on the input information value is necessary to be included because the processing module is directly influenced by it and by consequence, also its quality.

2) Non-analytical evaluation of $\mathbf{Q_f}$*:* In some cases analytical formula are not available. This is the situation when the input and the output quality have an important number of dimensions, with dependencies between them. Another situation is when we do not have a complete knowledge of the module functioning behaviour. In each case we state that at least the knowledge on the input-output characteristics and the module general functioning behaviour can be obtained. By means of this knowledge, the qualitative behaviour of the processing module can be studied varying the input quality and measuring the corresponding output quality. In this way we obtain pairs of the type $(\vec{Q}_{in},\ \vec{Q}_{out})$ for each processing module. These pairs allow to determine relations between the input and the output quality by using statistical methods like regression.

We give an illustrative example of obtaining the function $\mathbf{Q_f}$ from measuring the pairs $(\vec{Q}_{in},\ \vec{Q}_{out})$ in figure 7. For the simplicity of presentation we consider the unidimensional case for the quality evaluation, taking the same module example as the one in the previous subsection. The knowledge about the module is: it functions as a two-class Bayesian classifier; it receives in its input a unidimensional signal susceptible of being affected by a additive gaussian noise; it delivers in its output the signal estimated class. Without the knowledge on the value of the threshold, the analytically determination of the $\mathbf{Q_f}$ is not possible. The input quality in this case is represented by the noise level and the output quality by the probability of correct classification (the detection probability). The output quality evaluation was carried out using a Monte Carlo simulation consisting in the generation of 10000 data samples taken from a Bernoulli distribution with success probability $p = 0.5$. The value of the signal parameter A was 5 and the noise $N \sim \mathcal{N}(0, \sigma^2)$ with σ^2 varying from 0.25 to 100. With red points in figure 7 is represented the probability of correct detection. Small biases can be observed between the simulated $\mathbf{Q_f}$ and the theoretical one. With the $\mathbf{Q_f}$ determined, for each new input data, the output quality can be directly evaluated using an interpolation between the two closest points to the input noise level.

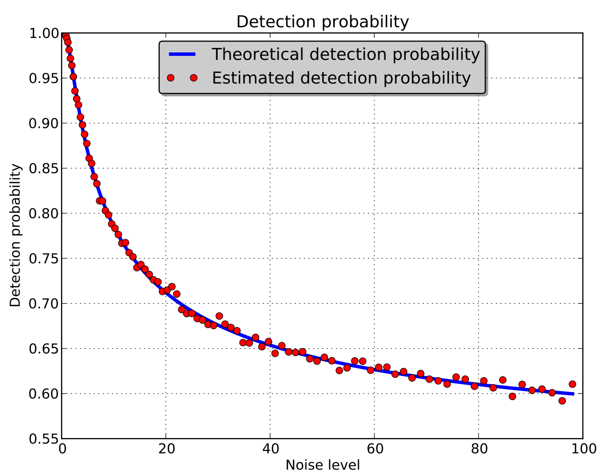

Fig. 7. Non-analytical evaluation of a two-class Bayesian classifier transfer function $\mathbf{Q_f}$ (detection probability)

For the multidimensional case, we propose to evaluate independently each dimension of the output quality, based on all of the input quality dimensions. In this way each of the output quality dimension will be given by a n-dimensional surface (with n the input quality dimension).

As at each level a number of quality criteria are used for evaluating the information quality, we can be put in the situation of expressing this quality in only one dimension: one value. This can be the case if the final user wants to have an idea of the quality evolution through the system and he does not have the time to integrate all the quality dimensions. As the different quality criteria are evaluated using incommensurable measures, not having the same domain of definition, meaning or importance, the aggregation process is not so trivial. A solution to this problem is to use a fuzzy integral, like Choquet's or Sugeno's integral [15]. The main characteristics of the fuzzy integrals are the non-linearity of the aggregation and the possibility of capturing interactions between the measures to combine.

C. Global information quality evaluation

In this section we will present how the local quality evaluation can help in evaluating the information fusion system quality. Until now we defined how to evaluate the information quality in the case of a single processing module. Now we present how this quality evaluation can be extended to the concatenation of processing modules. In figure 8 are presented two successive modules that need to be concatenated. To be capable of working together the output information of the first module needs to be adapted to the input of the second module. This adaptation is transferred to the quality domain by the fact that the output information quality of the first module is the same as the input information quality of the second module: $Q^i_{out} \equiv Q^{i+1}_{in}$.

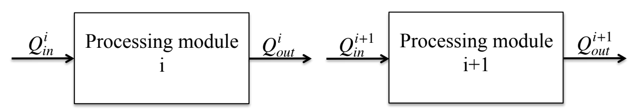

Fig. 8. Concatenation of two modules

Once defined the methodology of evaluating the quality locally, we need to extent it to the evaluation of the information fusion system quality, which is the final purpose of our work. The starting point is the observation that a quality variation in the input of a module implies a quality variation in the output of that module. Using this reasoning, we propose the principle of quality changes propagation. It states that the local variations of the quality propagate through the information system to its output as shown in figure 9. Applying this principle, the information quality evaluation can be done in cascade, after each module, to finally arrive to the evaluation of the information proposed to the user.

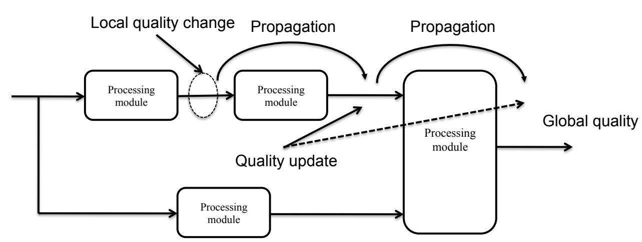

Fig. 9. Dynamic global information quality evaluation

We want that the evolution of the input data to be captured by our methodology. The quality of these databases/sensors can be evaluated by simply applying the methodology presented in the section III-A. For capturing the quality evolution we propose the use of a probe measuring the quality at regular intervals of time determined by the specificity of the application. If the database/sensor is considered as a module, the quality in its output will be described as a list of data quality criteria $Cr_i,\ i \in \{i, \ldots, n\}$ with values given by quality measures $m_i^j,\ i \in \{i,\ldots,n\} \quad j \in \{i,\ldots,N_i\}$, as presented in figure 10.

$$Q_{out} = \begin{pmatrix} Cr_1 : \{m_1^1, m_1^2, \cdots, m_1^{N_1}\} \\ Cr_2 : \{m_2^1, m_2^2, \cdots, m_2^{N_2}\} \\ \vdots \\ Cr_n : \{m_n^1, m_n^2, \cdots, m_n^{N_n}\} \end{pmatrix}$$

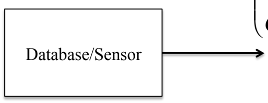

Fig. 10. Quality of a sensor/database measured by a probe

IV. Conclusions and future work

In this paper we propose a new information quality methodology. Based on reference works [3], [6], [11] we define a new method of evaluating data and information quality. Unlike these references we make a distinction between data and information, theirs definition differences implying a different quality definition.

We placed our study in the context of a complex information system. To overcome the evaluation problem of a such system quality, we proposed a decomposition of the information system in its elementary modules. This decomposition allowed us to have a local vision of the information system. In this vision we defined a methodology for evaluating the influence of a processing module on the information quality. We modelled the output information quality by defining a quality transfer function, taking as arguments the input information quality and the value of the input information. Moreover, we showed how to determine such a function in single and multi dimensional cases.

With the local quality evaluation, the information fusion system evaluation is done by propagating the quality through the system. Using this principle of quality propagation we answered another question about the instantaneous quality evaluation.

Our methodology can be used by the users of an information system who want to understand the information system quality evaluation and the various quality exchanges inside it. We are convinced that this information system quality evaluation will help users to act better in the process of decision making. Also, the local quality evaluation allows an information system analyst to check the processing modules performances depending on the application context. Having at his disposal a methodology of evaluating not only the processing modules performances but also the entire information system quality, makes possible to compare different types of information systems architectures.

In future research, we will continue the development of this methodology and we will employ it in a real world application such as Hospital Information System

Acknowledgment

This research was granted by Direction Générale de l'Armement (French MoD) and by Brittany Council, France.

References

[1] E. Turban, J. Aronson, and T.P.. Liang, *Decision support systems and intelligent systems 7ed.*, Pearson Education, Upper Saddle River, New Jersey, 2005.

[2] T. Redman, and R. Pautke, "The organisation's most important data issues", In *Information and database quality*, Eds. Piattini, M., Calero, C. and Genero, M., Kluwer Academic Publishers 2002.

[3] R. Wang and D. Strong, *Beyond accuracy: what data quality means to data consumers*, Journal of Management Information Systems, 12(4), 1996.

[4] S. Madnick, R. Wang, Y. Lee and H. Zhu *Overview and Framework for Data and Information Quality Research*, Journal of Data and Information Quality, 1(1), 2009.

[5] F. Naumann and D. Strong, *Quality-Driven Query Answering*, LNCS 2261, Springer-Verlag, 2002.

[6] G. Rogova and E. Bossé, *Information Quality effects of Information Fusion*, DRDC Valcartier TR 2005-270.

[7] G. Rogova and E. Bossé, *Information Quality in Information Fusion*, 13th Conf. on Information Fusion, 2010.

[8] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters 27, 2006.

[9] E. Blasch and S. Plano, *DFIG Level 5 (User Refinement) issues supporting Situational Assessment Reasoning*, 8th Conf. on Information Fusion, 2005.

[10] E. Blasch, I. Kadar, J. Salerno, M. Kokar, G. Powell, D. Corkill and E. Ruspini, *Issues and Challenges in Situation Assessment (Level 2 Fusion)*, Journal of Advances in Information Fusion, Vol. 1, No. 2, 2006.

[11] L. English, *Information Quality Applied*, Wiley Publishing, Indianapolis, Indiana, 2009.

[12] A. Aamodt and M. Nygård, *Different roles and mutual dependencies of data, information, and knowledge - an AI perspective on their integration*, Data and knowledge engineering, vol. 16, 1995.

[13] P. Gustavsson and T. Planstedt, *The road towards multi-hypothesis intention simulation agents architecture - fractal information fusion modeling*, Proceedings of the 2005 Winter Simulation Conference.

[14] G. Klir and T. Planstedt, *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley, Hoboken, NJ, 2005

[15] M. Grabisch and C. Labreuche *A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid*, 4OR: A Quarterly Journal of Operations Research, 6(1), 2008