preprint of a paper published in:
Measurement, 72,77-87, 2015, doi:10.1016/j.measurement.2015.04.030

A conceptual framework for concept definition in measurement: the case of 'sensitivity'

Arianna Mencattini[1] and Luca Mari[2#]

[1] University of Rome Tor Vergata, Dept. of Electronic Engineering, Italy, mencattini@ing.uniroma2.it
[2] Università Cattaneo – LIUC, School of Industrial Engineering, Italy, lmari@liuc.it

Abstract
The concept 'sensitivity' has multiple and sometimes incompatible usages and definitions, as they can be found in the scientific and technical literature. A strategy is proposed towards a conceptual framework in which sensitivity is qualitatively intended as a feature of a black box behavior and quantitatively is defined according to specific evaluation types (interval / ratio, ordinal, nominal) for both deterministic and stochastic behaviors. The proposed formal definitions characterize stochastic sensitivity as constituted of "effective" and "confounding" components, that can be simultaneously present and contribute to a desirable and unwanted increment of global sensitivity respectively. Two examples taken from the context of imaging systems and image-based measuring systems, in which sensitivity is computed in presence of non-negligible uncertainty sources, provide some hints on the usefulness of the proposed framework.

Keywords: Terminology in metrology; Measuring systems; Instrumental uncertainty; Sensitivity

1. Introduction
In all non purely formal bodies of knowledge, and thus in experimental sciences and technology in particular, the relevant concepts are conveyed not only through mathematics but also by means of linguistic expressions, at least with the aim of interpreting mathematical constructs in terms of empirical entities. While maybe unavoidable, such usage of natural language is not exempt from problems, as it may lead to both synonyms (different terms designating the same concept) and polysemies (different concepts designated by the same term) [2]. Synonyms generate redundancy but their potentially negative effects can be easily prevented, for example by means of a thesaurus. On the other hand, polysemies are sources of ambiguities, whose presence reduces the chance or increases the cost of correct communication.
Of course, such issues are mainly due to traditions and consolidated usages of linguistic expressions and their relations to concepts, not to concepts as such. It is not amazing then that measurement, a process exploited in so many fields and since so long time, is particularly affected by an ambiguous lexicon. The plethora of specific measurement-related sublanguages hinders the inter-disciplinary communication and emphasizes an artificial separation between science(s) and society. Hence, contributing to bridge these gaps seems to be a promising, worthwhile effort.
It is the goal explicitly pursued by the Joint Committee for Guides in Metrology (JCGM) in the development of the *International vocabulary of metrology – Basic and general concepts and associated terms* (VIM) [3]. Building on the traditional ground of measurement of physical quantities, which was the basic scope of its first two editions, the current, third edition of the VIM explicitly aims at a broader reach, being "meant to be a common reference for scientists and engineers – including physicists, chemists, medical scientists – as well as for both teachers and practitioners involved in planning or performing measurements, irrespective of the level of measurement uncertainty and irrespective of the field of application. It is also meant to be a reference for governmental and inter-governmental bodies, trade associations, accreditation bodies, regulators, and professional societies." [3, Scope].
The VIM is then an important step towards a widespread, universally agreed concept system and lexicon for measurement science. On the other hand, the many existing sublanguages about measurement make the endeavor a complex and delicate one: whenever multiple concepts for the same term are found in the

---

Corresponding author: Luca Mari.
One of the authors is a member of the Joint Committee on Guides in Metrology (JCGM) Working Group 2 (VIM). The opinion expressed in this paper does not necessarily represent the view of this Working Group.
A preliminary version of this paper was presented at the IMEKO TC 7 Symposium, 4-6 September 2013, Genova, Italy, and then published as [1].

scientific and technical literature, a comparative study should be done to identify an appropriate, encompassing definition, or at least to emphasize the reasons and the nature of such multiplicity, including the possible differences in scope. An explicit strategy to produce consistent definitions / presentation should be adopted, particularly in the case the same concept is both qualitatively and quantitatively characterized and the two layers have to be properly coordinated.[1]

This paper is aimed at proposing a possible example of such a strategy in reference to the concept 'sensitivity', a case that at the same time is particularly difficult for its many, sometimes markedly different, characterizations and is relevant in many scientific and technical fields: the lessons learned might be then replicable to other measurement-related concepts. Hence this is in line with the attempt of the third edition of the VIM to widen the scientific and technical community to which the vocabulary is devoted.

2. Framing the concept 'sensitivity'

The concept 'sensitivity' is defined in so many contexts and with so many specifications that one might doubt that such a polysemy can ever be solved. A first problem is: sensitivity of what? Indeed, by means of a feature termed "sensitivity" different entities are characterized: measuring instruments and systems but also, e.g., methods, tests, and algorithms.[2]

Let us then adopt a black box (meta-)modeling strategy, by hypothesizing that sensitivity is a feature of an entity X under the only conditions that X is not static, i.e., it can produce different outputs $Y(t)$ depending on its input $U(t)$, and possibly a random contribution $N(t)$. The black box assumption implies that the analytical form of the input-output relation, generically denoted with f in the following, might not be known.[3] Furthermore, no constraints are given on the ranges and types of U and Y and on time domain, allowed to be either continuous or discrete.

On this basis some significant examples of the definition of 'sensitivity' follow, preliminarily classified in two general categories – let us denote them as α and β – in terms of this input-output characterization.

α. Sensitivity as $\Delta y/\Delta u$, where $u \in U$ and $y = f(u) \in Y$, possibly under specified conditions.

1. *Sensitivity of a measuring system*: "the quotient of the change in an indication of a measuring system and the corresponding change in a value of a quantity being measured"; VIM, 4.12 [3];
2. *Photocathode sensitivity*: "ratio of the photoelectric emission current from the photocathode to the incident luminous flux under specified conditions of illumination"; IEV, 394-38-11, Nuclear instrumentation – Instruments, systems, equipment and detectors / Characteristics of radiation detectors (all IEV definitions are taken from [5]);
3. *Sensitivity of a measuring assembly*: "for a given value of the measured quantity, ratio of the variation of the observed variable to the corresponding variation of the measured quantity"; IEV: 394-39-07, Nuclear instrumentation – Instruments, systems, equipment and detectors / Characteristics of radiation measuring assemblies;

---

1 The VIM3 offers some interesting examples of the qualitative vs quantitative distinction: "Measurement precision is usually expressed numerically by measures of imprecision..." [3, 2.15 N1]; "Measurement trueness is not a quantity and thus cannot be expressed numerically, but measures for closeness of agreement are given..." [3, 2.14 N1]; "The concept 'measurement accuracy' is not a quantity and is not given a numerical quantity value." [3, 2.13 N1]. It is the situation of three concepts with a qualitative definition and three different treatments as for their quantitative counterparts, without stated reasons justifying this difference.

2 In this paper the distinction between terms referring to objects, or concepts, or terms is critical. Hence we adopt the notational convention from ISO standards, e.g., [4]. A term, and more generally a linguistic expression, referring to:

– itself, i.e., a term, is *delimited by double quotes*;
– a concept, i.e., its meaning, is *delimited by single quotes*;
– an object, i.e., its referent, is *not delimited*.

Hence, (the concept) 'sensitivity' is expressed in English by (the term) "sensitivity" and is about (the object) sensitivity. The lack of delimiters around terms for objects (i.e., entities of the world) follows an economic principle: in everyday writing, we usually intend to refer to objects, and not to concepts or terms.

3 We will adopt the notational convention of denoting properties, typically modeled as random variables, by upper case characters and their values / occurrences by lower case characters. In general X is a multiple-input and multiple-output (MIMO) system, so that its input and output are vector properties. For the sake of simplicity, in the following we will assume that U and Y are scalars, then single components of these vectors.

4. *Dynamic sensitivity*: "under stated conditions of operation, the quotient of the variation of the photoelectric current of the device by the initiating small variation of the incident radiant power or luminous flux"; IEV, 531-44-25, Electronic tubes / General properties and quantities of photosensitive tubes;
5. *Analytical sensitivity of a method*: "the slope of the calibration curve and the ability of an analytical procedure to produce a change in the signal for a defined change of the quantity"; CLSI, Project ILA20 [6];
6. *Sensitivity analysis*: "a what-if type of analysis to determine the sensitivity of the outcomes to changes in parameters; if a small change in a parameter results in relatively large changes in the outcomes, the outcomes are said to be sensitive to that parameter"; OECD Statistics glossary [7].

Some comments on these definitions. α2 is the example of a specific version of α1. α3 emphasizes that the sensitivity may depend on the input value u, i.e., from $u_1 \neq u_2$ it generally follows that $\Delta f(u_1)/\Delta u_1 \neq \Delta f(u_2)/\Delta u_2$. α4 emphasizes that the sensitivity may be time-variant (i.e., generally $\Delta f(U(t_1))/\Delta U(t_1) \neq \Delta f(U(t_2))/\Delta U(t_2)$), thus possibly distinguishing between transient and steady-state sensitivity. In synthesis, sensitivity of X is then generally sensitivity of X with respect to U and Y, and relatively to given u and t. α5 points out that for measuring instruments the relation between U and Y is formalized by the calibration function, so that in this case the sensitivity is the first derivative of such function. Finally α6 shows that X is not required to be a physical device: it can be a formal procedure, e.g., an algorithm.[4] Note that these definitions are dimensionally coherent: $\dim(sensitivity_\alpha) = \dim(Y) / \dim(U)$.

β. Sensitivity as a given function $g(u)$ of $u \in U$, such that $y = f(u) \in Y$, satisfies a given condition.

1. *Intrinsic sensitivity*: "the minimum level of signal at the input of a radio receiver, which produces an output signal of specified quality, when the gain is sufficient and assuming no external noise"; IEV, 713-10-55, Radiocommunications: transmitters, receivers, networks and operation / Radio reception and receivers;
2. *Sensitivity*: "the input luminous flux necessary to obtain a stated signal-to-noise ratio in the output"; IEV, 531-45-13, Electronic tubes / Camera tubes;
3. *Knee sensitivity*: "the luminous flux at the knee point of the light-signal transfer characteristic"; IEV, 531-45-10, Electronic tubes / Camera tubes;
4. *Electrical sensitivity*: "reciprocal of the threshold for electrical stimulation"; IEV, 891-02-35, Electrobiology / Electrophysiology;
5. *Diagnostic sensitivity*: "the proportion of patients with a well-defined clinical disorder (or condition of interest) whose test values are positive or exceed a defined decision limit (i.e., a positive result and identification of the patients who have a disease)"; CLSI, Project MM07 [6].

Some comments on these definitions. β1 and β2 are examples where the sensitivity is $\min(u)$ such that $f(u) \geq y'$, where y' is a given threshold. β3 defines the sensitivity as u such that $f(u) = y'$, where y' is a given value. β4 defines the sensitivity as u^{-1} where $\min(u)$ such that $f(u) \geq y'$, where y' is a given value. β5 is a more complex case, and can be interpreted as follows. Let $\{u_i\}$ be a finite set of entities having a given property (e.g., a clinical disorder), being u the cardinality of the set, $u = \#\{u_i\}$. For each u_i, let $y_i = f(u_i)$ be the outcome of a test X applied to u_i and such that $y_i = 1$ if and only if $y_i \geq y'$, where y' is a given threshold, where $u = \#\{y_i: y_i = 1\}$. Sensitivity is then defined here as y/u. Note that, from the dimensional viewpoint, $\dim(sensitivity_{\beta 1\text{-}3}) = \dim(U)$; $\dim(sensitivity_{\beta 4}) = \dim(U)^{-1}$; $\dim(sensitivity_{\beta 5}) = 1$.

As a first synthesis, it appears that there is a coherent conceptual network around sensitivity$_\alpha$, whose elements can be characterized as different specializations and operationalizations of a basic definition such as "the derivative $\partial Y_j / \partial X_i$ of an output Y_j versus an input U_i can be thought of as a mathematical definition of

4 The possibility to consider sensitivity of both empirical and formal entities X arises an interesting issue: if X is empirical, its sensitivity cannot be evaluated on the involved properties U and Y, but *on a model* of them, and therefore in reference to their values.

the sensitivity of Y_j versus U_i" [8] (for consistency symbols have been changed from the original text). At the same time several other definitions can be found, gathered above under the term "sensitivity$_\beta$" but so mutually inconsistent that they do not even share the same input-output characterization. The fact is also remarkable that all α definitions assume that sensitivity can be quantified in black box conditions, i.e., even with no knowledge of the internals of X, whereas less clear is whether this can be done also in the case of β definitions. Furthermore, while all α definitions are given in purely differential terms, i.e., an experiment to quantify sensitivity$_\alpha$ in principle only requires the ability to generate variations in the input of X, at least some of the definitions of sensitivity$_\beta$, and β5 in particular, imply the knowledge of specific input values ("patients having a well-defined clinical disorder"). Finally, all quoted definitions but β5 implicitly assume a deterministic modeling of X behavior, that is instead statistic in β5.
On the basis of these considerations our strategy has been designed: in the following we will propose a conceptual framework in which sensitivity, generally intended as sensitivity$_\alpha$, is at first defined qualitatively and then the definition is specified quantitatively in reference to given scale types. This will allow us then to propose a generalization to the case where uncertainty cannot be neglected.

3. A formal concept of sensitivity

A conceptual framework in which 'sensitivity' is given both a general definition and some type-related specific quantitative definitions provides users with two levels of access to information and reading of documentation:
– the general definition offers a wide accessibility to heterogeneous scientific groups and easy interpretation of the concept;
– the specific definitions are oriented to focused targets and can be customized according to specific theoretical assumptions or application needs.
The hypotheses we propose to ground this framework are as follows.

Hypothesis 1 – Sensitivity is a property of an entity X modeled as a black box and under the only condition that the output $Y(t_2)$ of X at the time t_2 causally depends on its input $U(t_1)$, $t_1 \leq t_2$.

X could be a measuring system but also a test, an algorithm, ... As the quoted VIM definition (α1) shows, Hypothesis 1 is immediately applicable to measuring systems, by just assuming that Y = indication, and U = quantity being measured. This hypothesis admits that the analytical form of the input-output relation, i.e., the entity behavior, is not known, and does not impose any constraints on the types of the properties U and Y.

Hypothesis 2 – The sensitivity of X is a feature related to the ability of the entity X to detect variations of its input U.

These hypotheses lead to the following definitions, that have been first proposed in [1] and provide a conceptual baseline for the rest of the paper.

General ("qualitative") definition – *Sensitivity is a property of an entity, possibly modeled as a black box, and characterizes the ability of the entity to produce a variation of its output in response to a variation of the received input*.

This definition, that does not explicitly provide a way to assign a value to the sensitivity of a given entity X, is based on a generic concept of variation, that has to be specified in reference to specific evaluation types [9], thus leading to multiple quantitative concepts. Despite this generality, the process of sensitivity assessment that is assumed here has a definite operational structure: two distinct inputs must be applied to X, so to produce a variation of the received input, and the corresponding outputs must be recorded and compared.
In the following u and y denote input and output property values respectively, where the time dependence of the properties U and Y is omitted for simplicity. By firstly assuming that all uncertainties are negligible, the previous, qualitative definition can be specified in quantitative terms as follows.

General ("quantitative") definition – Sensitivity at the input value u is:

$$S(u) = \frac{var(y(u), y(u))}{var(u,u')} \qquad (1)$$

where var is a generic function of variation and u' is an input property value that minimizes var(u, u') under the condition that var(u, u') > 0.

The concept of variation, as formalized by the function var, has then a crucial role for this definition of sensitivity. In the case of interval or ratio evaluations, variation is intrinsically part of the scale, and therefore it can be exploited as is, corresponding numerically to the difference of distinct values v_i in the scale:

$$\mathrm{var}(v_i, v_j) = |v_i - v_j|$$

Not so univocal is the concept for ordinal quantities, where the only information meaningfully available on the quantities is ranking and therefore their algebraic difference is not invariant for scale transformation. On the other hand, variations in a totally ordered scale can be simply evaluated, e.g., by counting the number of values in the scale within the values corresponding to the quantities under consideration. If it is supposed that the scale index i runs monotonically with the scale order, so that the scale values are $v_1 < v_2 < v_3 <$..., then trivially:

$$\mathrm{var}(v_i, v_j) = |i - j|$$

where also in this case var(v_i, v_j) is a metric, and the minimum non-null variation is obtained in the case var(v_i, v_{i+1}) = 1.

When properties are evaluated in a nominal scale even order cannot be exploited: two properties are either in the same class or in distinct classes, and therefore they are either associated to the same value or to distinct values, but neither distance nor order are invariant among such values. This suggests a concept of variation such as:

$$\mathrm{var}(v_i, v_j) = 1 - \delta_{i,j}$$

where the function:

$$\delta_{i,j} = 1 \text{ if } i = j \text{ and } = 0 \text{ otherwise}$$

is the Kronecker delta. It can be easily shown that also in this case var(v_i, v_j) is a metric, and that the minimum (actually: the only) non-null variation is when var(v_i, v_j) = 1.

Hence, for each property evaluation type at least one variation function var(v_i, v_j) is available that is applied to property values and represents information on the empirical variation of the corresponding properties. The fact that, independently of the type, var(v_i, v_j) is a metric, i.e., a function ranging in R^+, allows us to exploit it to define a concept of sensitivity which is at the same time qualitatively encompassing and quantitatively specifiable to each type. Furthermore, it is implied here that the input to X is not constant, so that for a generic u the element u' such that var(u, u') > 0 generally exists. On the other hand, at least in the weakest case of nominal evaluation the uniqueness of u' is not guaranteed.

From these general definitions some type-aware specific definitions of sensitivity can be proposed [1] as follows.

Specific definition 1 – Sensitivity at the input value u for an *interval or ratio* evaluation is:

$$S(u) = \frac{\Delta y(u)}{\Delta u} \qquad (2)$$

Of course, this is the customary definition of sensitivity$_\alpha$ (see in particular α1) mentioned above.

Specific definition 2 – Sensitivity at the input value u_i for an *ordinal* evaluation is:

$$S(u_i) = \frac{var(y(u_i), y(u_{i+1}))}{var(u_i, u_{i+1})} = |k_0 - k_1| \qquad (3)$$

where k_0 e k_1 are the indexes of $y(u_i)$ and $y(u_{i+1})$ respectively in the order specified by the evaluation (note that by construction var(u_i, u_{i+1}) = 1).

Specific definition 3 – Sensitivity at the input value u_i and in reference to the input value u_j, $u_i \neq u_j$, for a *nominal* evaluation is:

$$S_{u_j}(u_i) = \frac{var(y(u_i), y(u_j))}{var(u_i, u_j)} = 1 - \delta_{k_0, k_1} \quad (4)$$

where k_0 e k_1 are the indexes of $y(u_i)$ and $y(u_j)$ respectively in the classification specified by the evaluation, and δ_{k_0, k_1} is the Kronecker delta. Due to the lack of algebraic structure on the sets of values for a nominal evaluation, sensitivity in this definition is in fact a bi-argumental function: as mentioned above, for any given u_i the value u_j such that var(u_i, u_j) > 0 is minimum is not unique, and therefore must be explicitly specified.

Despite the differences in the implied algebraic structure, these three specific definitions assume exactly the same procedure to evaluation of sensitivity:
– two distinct inputs u and u' such that their variation is (positive but) minimum are independently applied to X;
– the corresponding outputs $y(u)$ and $y(u')$ are registered;
– the sensitivity of X at u is evaluated by the ratio: variation of outputs / variation of inputs.
Sensitivity assessment is then the example of an operative activity of metrological characterization of X,[5] together with, e.g., the assessment of its resolution, selectivity with respect to specific influence properties, and repeatability. Such activities are typically performed by system manufacturers, with the aim of specifying the technical data sheet of X, and therefore in reference conditions such as those available in laboratory. Hence, the hypothesis of assuming deterministic inputs seems to be acceptable. On the other hand, even laboratory conditions cannot guarantee that the behavior of X is deterministic: the next Section is devoted to exploring the possibility to give a generalized, quantitative definition of sensitivity, applicable for example in the case of non-negligible instrumental uncertainty [3].

4. Towards a concept of sensitivity for entities with non-deterministic behavior

If the behavior of X is not deterministic, its output is uncertain. Still maintaining the black box hypothesis, and then in particular independently of the analysis of the possible sources of uncertainty, the behavior of X can be then modeled in terms of "the distribution of values that could reasonably be attributed" to the output property Y [10], so that in this case assessing the sensitivity of X requires estimating the probabilities of the outputs.
Let us then suppose that the output Y of X depends not only on the input U but also on a random process N, assumed to formalize the non-deterministic behavior of X, and in particular its instrumental uncertainty, all other sources of uncertainty being considered negligible (i.e., definitional uncertainty) or included in N (i.e., reading uncertainty). Moreover, let us suppose that an explicit model of N is not available, so that a black box model is maintained, in which an uncertainty analysis cannot be analytically performed. The new procedure is then analogous to the previous one: X is fed with two distinct input values u_1 and u_2 and the corresponding output values y_1 and y_2 are registered. The process is assumed now to be repeatable, given the same values u_1 and u_2 (without the assumption of repeatability, in principle X could be admitted changing its sensitivity between process instances; without the assumption of input constancy, sensitivity would be evaluated at different values). For each j-th instance of the n performed repetitions the output values y_{1j} and y_{2j} are registered, thus obtaining two samples $\langle y_{11}, ..., y_{1n}\rangle$ and $\langle y_{21}, ..., y_{2n}\rangle$, that can be formally intended as outcomes of two output random processes Y_1 and Y_2 [11]. The target is to estimate the stochastic variation in output of X under the repeatability conditions. Two alternative strategies for defining 'sensitivity' in this generalized context can be followed:

1. sensitivity as a generalized ratio of variations of random variables (the denominator has been assumed to be a deterministic variation, but it could be more generally a random variation in case of not negligible uncertainty on the input property), so that sensitivity would be itself a random variable and its quantitative analysis would involve moments estimation;
2. sensitivity as the ratio of the root second non-central moment of the variable $Y_2 - Y_1$ divided by the deterministic variation $u_2 - u_1$, so that sensitivity remains a scalar entity as in the deterministic case.

5 Of course, eq. (2) is easily extended to the continuous case, as $S(u) = dy(u)/du$, that however can be operatively assessed only if the black box hypothesis is removed and the function $y(u)$ is supposed to be analytically known.

In this paper, we develop the second strategy, thus considering sensitivity as a scalar quantity and not a random variable. This choice is in continuity with the deterministic representation, which might increase the acceptability of the generalization, and is consistent with the black-box assumption on both the entity *X* and the random process *N*. Indeed, the alternative strategy of considering sensitivity as a random variable would require the acquisition of additional knowledge about the entity even in cases in which it is modeled by a general computational, explicit or not, procedure (e.g, iterative, differential equation, etc).[6]
According to the second strategy, and in line with the type-aware characterization presented above in the deterministic case, the following definitions are proposed.

Stochastic specific definition 1 – Stochastic sensitivity at the input value *u* for an *interval or ratio* evaluation with an instrumental uncertainty represented by a random process *N* is:

$$S(u) = \frac{\sqrt{E\left[(Y_2 - Y_1)^2\right]}}{\Delta u} \tag{5}$$

where $E\left[(Y_2 - Y_1)^2\right]$ is the expected value of the random process $(Y_2 - Y_1)^2$, being Y_i the output random process obtained by the repeated deterministic application of input u_i, affected by the random process *N* (the square is needed to take into account the statistical correlation among output variables and the individual variation).

Some considerations on this definition.
First, it is well known that:

$$E\left[(Y_2 - Y_1)^2\right] = \sigma^2_{Y_2 - Y_1} + \left(\mu_{Y_1} - \mu_{Y_2}\right)^2 \tag{6}$$

where $\sigma^2_{Y_2 - Y_1}$ is the variance of the random process $(Y_2 - Y_1)$ and μ_{Y_1} and μ_{Y_2} are the expected values of variables Y_1 and Y_2 respectively. Hence sensitivity is here the superposition of two components, the square difference of the expected values of output processes Y_1 and Y_2 and the variance of the process $Y_2 - Y_1$, thus highlighting the twofold nature of sensitivity in the non-deterministic case: an "effective" sensitivity, given by the first term in eq. (6), such that when the difference of the expected values of the two output random processes increases a higher capability to discriminate outputs is expected; and a "confounding" sensitivity, given by the second term in eq. (6), such that an increased variance produces larger variations on average, that however cannot be related to an increased amount of differential information. A high sensitivity of this second kind can be the symptom of low repeatability, hence not necessarily an optimal working condition.
Second, the equality $\sigma^2_{Y_2 - Y_1} = \sigma^2_{Y_1} + \sigma^2_{Y_2} - 2cov_{Y_1 Y_2}$ shows that when uncertainty decreases, i.e., variables Y_1 and Y_2 tend to their expected value, then $\sqrt{E\left[(Y_2 - Y_1)^2\right]} = \mu_{Y_1} - \mu_{Y_2} = var(Y_1, Y_2)$ in analogy with the deterministic case in eq. (2). Moreover, if Y_1 and Y_2 are uncorrelated the numerator in eq. (5) can be rewritten as $\sqrt{E\left[(Y_2 - Y_1)^2\right]} = \sqrt{\left(\mu_{Y_1} - \mu_{Y_2}\right)^2 + \left(\sigma_{Y_1} - \sigma_{Y_2}\right)^2}$, underlying the fact that stochastic sensitivity is zero when the expected values of the two output random processes and their standard deviations are simultaneously equal.
Finally, it should be noted that the proposed definition requires the estimation of the first and second moment of the random process $(Y_2 - Y_1)$. An operative definition can be obtained by estimation, under ergodicity assumption, by considering the two samples $\langle y_{11}, ..., y_{1n}\rangle$ and $\langle y_{21}, ..., y_{2n}\rangle$, and estimating $S(u)$ by:

6 The first strategy, that considers sensitivity as a random variable, has some possible advantages. First of all, in the case uncertainty contributions are extended to the inputs, the ratio of generalized variations is a random variable: assuming sensitivity as a function of random variable, and therefore a random variable itself, can offer a solution for a more general discussion. Secondly, when defining sensitivity as a random variable its expected value assumes the more immediate meaning of an expected output variation, while its variance can be seen as the quantification of the amount of uncertainty that propagates on the sensitivity of the entity. We plan to develop the first strategy in future works, possibly based on a less general framework with restricted assumptions (in particular, requiring calibration), then allowing propagation rules for the sensitivity analysis.

$$\hat{S}(u) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{1i} - y_{2i})^2}}{\Delta u} \tag{7}$$

Stochastic specific definition 2 – Stochastic sensitivity at the input value u_i for an *ordinal* evaluation with an instrumental uncertainty represented by a random process N is:

$$S(u_i) = \frac{\sqrt{E\left[(Y(u_i) - Y(u_{i+1}))^2\right]}}{var(u_i, u_{i+1})} = \sqrt{E\left[(Y(u_i) - Y(u_{i+1}))^2\right]} = \sqrt{E\left[(K_0 - K_1)^2\right]} \tag{8}$$

where K_0 and K_1 are discrete ordinal random processes representing the indexes of $Y(u_i)$ and $Y(u_{i+1})$ respectively.

Ordinal quantities are not invariant under standard algebraic operations, required to compute an operative estimation of sensitivity. To solve this problem the expected value for ordinal quantity can be reformulated, by simply substituting the empirical mean with the median, and the empirical standard deviation with the square median absolute deviation (MAD), so that:

$$\hat{S}(u_i) = \sqrt{1.4826 \cdot MAD^2(Y_2 - Y_1) + (median_j(y_{2j}) - median_j(y_{1j}))^2}$$

where:

$$MAD(Y_2 - Y_1) \equiv median_l(|(y_{2l} - y_{1l}) - median_j(y_{2j} - y_{1j})|)$$

and where y_{1j} and y_{2j}, $j = 1,\ldots, n$, are the samples obtained by repeated application of input values u_i and u_{i+1}.[7]
It can be noted that by setting $n = 1$ in the previous estimation we get:

$$\hat{S}(u_i) = |y_2 - y_1| = |Y(u_{i+1}) - Y(u_i)| = |k_1 - k_0|$$

in analogy to the deterministic case in eq. (3).

Stochastic specific definition 3 – Stochastic sensitivity at the input value u_i, and in reference to the input value u_j, $u_i \neq u_j$, for a *nominal* evaluation with an instrumental uncertainty represented by a random process N is:

$$S_{u_j}(u_i) = \frac{\sqrt{E\left[(Y(u_i) - Y(u_j))^2\right]}}{var(u_i, u_j)} \tag{9}$$

being $var(u_i, u_j) = 1$ for $i \neq j$, and where Y may only assume values in {0, 1} and therefore is a Bernoulli discrete random process.

The probability density function of process Y is given by:

$$f_Y(Y) = q \cdot \delta(Y) + p \cdot \delta(Y - 1), \quad p + q = 1$$

being p and q the probabilities of the two values, 0 and 1. If it is assumed, in absence of further information, that $p = q = 0.5$, then:

$$f_Y(Y) = 0.5 \cdot [\delta(Y) + \delta(Y - 1)]$$

that corresponds to a pair of impulses.
Under the assumption that $Y(u_i)$ and $Y(u_j)$ are independent, it can be proved that the process:

$$D_{ij}^2 = (Y(u_i) - Y(u_j))^2$$

is still a Bernoulli random process with values in {0,1}, such that sensitivity is the root expected value of D_{ij}^2.
The operative estimation of sensitivity is given by:

7 The value 1.4826 originates from the reciprocal of the standardized normal inverse cumulative distribution function evaluated at 3/4. The demonstration of the formula is beyond the object of the present paper. For further mathematical details see [12].

$$\hat{S}_{u_j}(u_i) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_{ki} - y_{kj})^2} = \sqrt{1 - \frac{1}{n}\sum_{k=1}^{n}\delta(y_{ki} - y_{kj})} \tag{10}$$

where δ is the Kronecker delta. Sensitivity is computed as the square root of one minus the average number of pairs (y_{ki}, y_{kj}) such that $y_{ki} = y_{kj}$, and therefore is directly related to the cardinality of the differences that are not recognized.

Finally, by setting $n = 1$ in the previous estimation we get:

$$\hat{S}_{u_j}(u_i) = \sqrt{1 - \delta(y_i - y_j)} = 1 - \delta(y_i - y_j)$$

since the term in the square root can only assume values in {0,1}. This is in analogy to the deterministic case in eq. (4).

5. Numerical examples

The proposed definitions extend the concept of sensitivity to the non-deterministic case and are specialized to the different (interval or ratio, ordinal, and nominal) evaluation types. To show their usefulness in this Section we present two numerical examples taken from the context of imaging systems and image-based measuring systems, in which sensitivity is computed in presence of non-negligible uncertainty sources.

5.1. Sensitivity in X-ray imaging system

A typical imaging system is constituted of an image formation system, a detector, and a recorder [13]. For example, an imaging medical device such as an analog mammographic unit, consists of a mammography X-ray equipment, a device to impress a photographic film, and a specific high performing scanner to digitize the film and produce a high resolution digital mammographic image (usually 12-16 bit per pixel). The scanner is the critical device in this kind of applications, since it introduces an unwanted amount of noise and degradation effects especially in very dark regions of the scanned image. A general mathematical model for such a system is [13]:

$$v(x,y) = g(w(x,y)) + \eta(x,y) \tag{11}$$

$$\eta(x,y) = f(w(x,y)) \cdot \eta_1(x,y) \tag{12}$$

where $w(x,y)$ is the input gray level intensity, $v(x,y)$ is the output gray level intensity, $g(\ )$ is a nonlinear function that represents the input-output response of the scanner, $f(\ )$ contributes to the signal-dependent noise standard deviation, and $\eta_1(x,y)$ commonly represents white noise. Consider an M2100 ImageClear DBA scanner used for digitizing X-ray mammographic images. From its datasheet, the output response is modeled as:

$$g = 10^{\frac{w(x,y)+d_0}{\gamma}}$$

with $\gamma = -0.92977$ and $d_0 = 5.16966$.

Assuming a negligible noise term and by evaluating the derivative of the output response, we get:

$$S_0(w) = \frac{1}{\gamma} \cdot ln(10) \cdot 10^{\frac{w(x,y)+d_0}{\gamma}} \tag{13}$$

However, due to the presence of uncertainty sources related to noise, the previous procedure is not correct and robust. Noise is characterized in the datasheet as reported in Fig. 1.

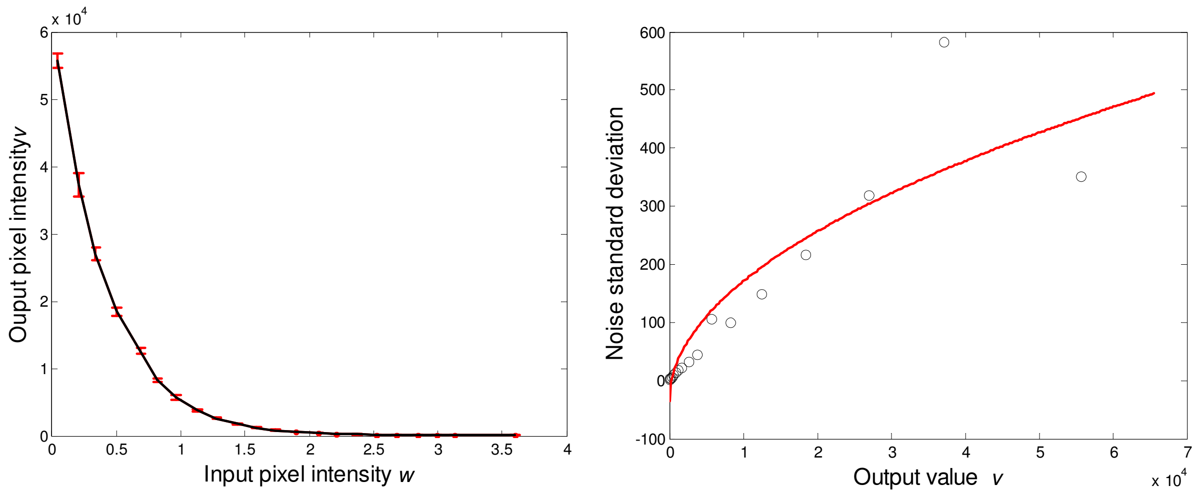

Figure 1. (left) Input-output characteristic curve. The red bars represent the 3σ confidence intervals. (right) The red curve represents the noise model of the relation between the output values and the noise standard deviation, estimated by a nonlinear least square approach.

In particular, Fig. 1 (left) illustrates the calibration curve of the scanner and Fig. 1 (right) illustrates the noise model represented by the relation between the output response v and the corresponding standard deviation measured in different uniform regions of a calibration wedge. According to this, and considering the previous definition of stochastic sensitivity for ratio or interval properties, an experiment based on Monte Carlo simulation can be developed as follows. At each of the $n = 10^4$ runs a value for the two output random variables v_1 and v_2 is generated, by repeatedly feeding the device with two distinct input gray levels w_1 and w_2 and generating the random terms accordingly (i.e., using the noise model described above). Due to the signal dependence of the noise model, the analysis has to be performed for different pairs of values in the whole input range. Given that, in logarithmic scanners, the input variable is in a nominal range of [0, 4], for each w_0 in this range, we derive the pair w_1 and w_2 as $w_1 = w_0 - Dw$, $w_2 = w_0 + Dw$, with $Dw = 0.01$ and w_1, w_2 in the range $[Dw, 4 - Dw]$. In order to highlight the relevance of the proposed solution for different noise levels, let us vary the amount of noise by multiplying by a factor c in [1, 3] the noise standard deviation. For each c value, we compare stochastic sensitivity $\hat{S}(w)$ with nominal sensitivity $S_0(w)$ given by eq. (11), by computing the relative root mean square error (RRMSE) of the two characteristics as follows:

$$RRMSE^c\left(\hat{S}(w), S_0(w)\right) = \sqrt{\frac{1}{n}\sum\left(\frac{\hat{S}^c(w) - S_0(w)}{S_0(w)}\right)^2} \qquad (14)$$

where $w = w_0$ means that the sum is evaluated for each pair w_1, w_2 around all the considered values w_0, and c indicates the dependence of stochastic sensitivity from the amount of noise. Fig. 2 illustrates the obtained RRMSE (figure top right) along with some examples of comparison (figures top-left and bottom left/right) of the obtained stochastic sensitivity (black marked curves) and the nominal sensitivity (green curves). Note that when c (and hence noise) increases there is a considerable discrepancy among the two characteristics, especially for high input values. This is due to the fact that for high input values nominal sensitivity goes to zero. So, from the discussion of the previous Section, we have a small “effective” term and a more relevant “confounding” term related to the noise variance captured by stochastic sensitivity. A further consideration is needed in this case. Due to the input-output inverse relation manifested in films recording and due to X-ray absorption phenomena, brighter regions in the original film image (i.e., fat tissue or film itself) are mapped into dark regions and vice versa for dark regions (i.e., dense tissue) mapped into brighter areas of the final image. Consequently, such scanners have a sensitivity that is smaller for regions that are not important for diagnostic using X-rays. Analogously, it is expected that noise has higher degradation effects in the same non-medically interesting regions.

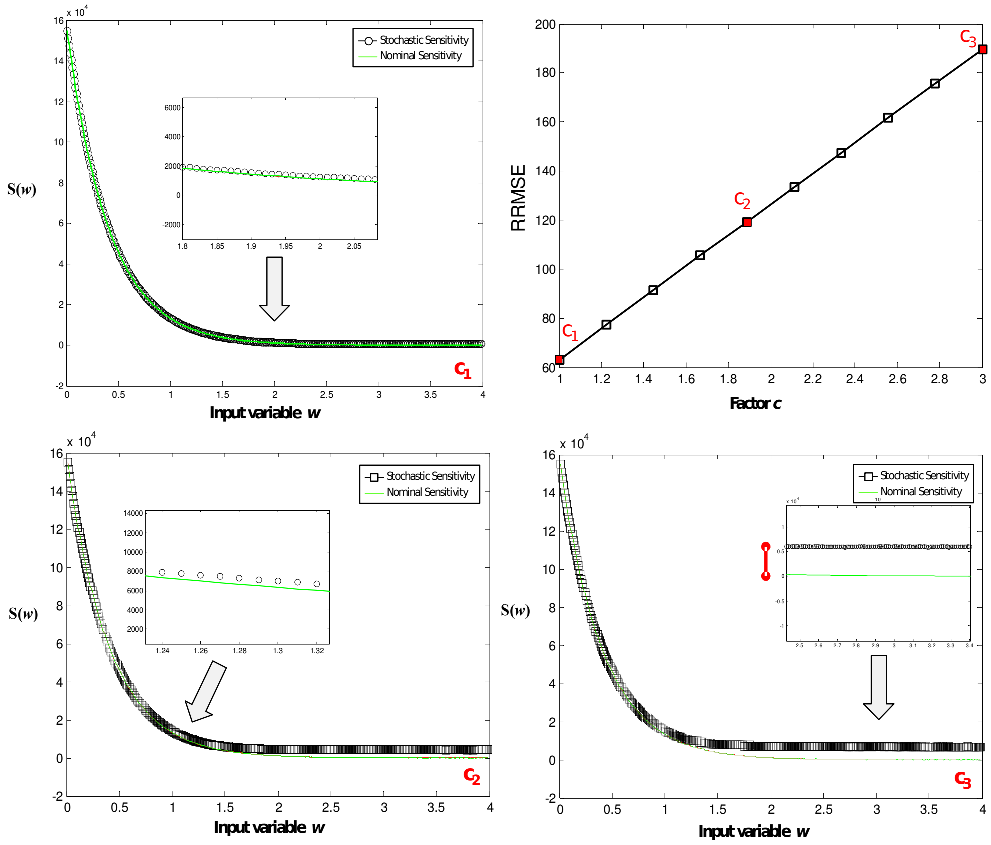

Figure 2. (top-right) RRMSE obtained for different values of c. Three particular choices, c_1=1, c_2=2, and c_3=3 are highlighted by the red squares and illustrated by the three graphs of sensitivity: top-left graph corresponds to c_1, bottom-left graph corresponds to c_2, and bottom-right graph corresponds to c_3. The black marked curve represents the stochastic sensitivity computed in each case, while the green curve represents the nominal sensitivity computed as the derivative of the input-output response.

5.2. Sensitivity in computerized diagnosis of breast calcifications

According to the BI-RADs lexicon [14], microcalcifications (MC) are small deposits of calcium (0.1-1 *mm*): they are usually the result of a genetic mutations somewhere in the breast tissue, but can also be due to other conditions. The size, distribution, form, and density of MCs are thought to give clues as to the potentially malignant nature of their origin. Due to the spatial resolution, mammography, screened by expert radiologist, is the unique imaging modality to find MCs. Previous works provided demonstration that malignant microcalcifications group in clusters and are more numerous and smaller than benign calcifications. Nowadays, computerized diagnostic systems are used in the clinical practice as second reader to improve the radiologist interpretation of the mammograms.

Based on previous data collected on mammographic images taken from the largest public dataset of mammographic images, the Digital Database for Screening Mammography [15], we provide here an estimation of sensitivity of an automatic diagnosis system for calcifications. We collected 122 regions of interest (ROIs) containing malignant microcalcifications and 115 ROIs with benign calcifications. Two examples of ROIs are illustrated in Fig. 3 (a)-(c) for malignant and (d)-(f) for benign. In particular Fig. 3 (a), (d) illustrate the original ROIs with malignant and benign calcifications, respectively, Fig. 3 (b),(e) represent the automatically enhanced calcifications, and Fig. 3 (c),(f) show the automatically identified (i.e., segmented) calcifications. Therefore, the entity X is here a complex algorithm able to autonomously enhance and segment the calcifications in a cropped region of the mammogram [16], compute the number of disjoint bright areas, and provide a category for the cluster, i.e., malignant vs. benign.

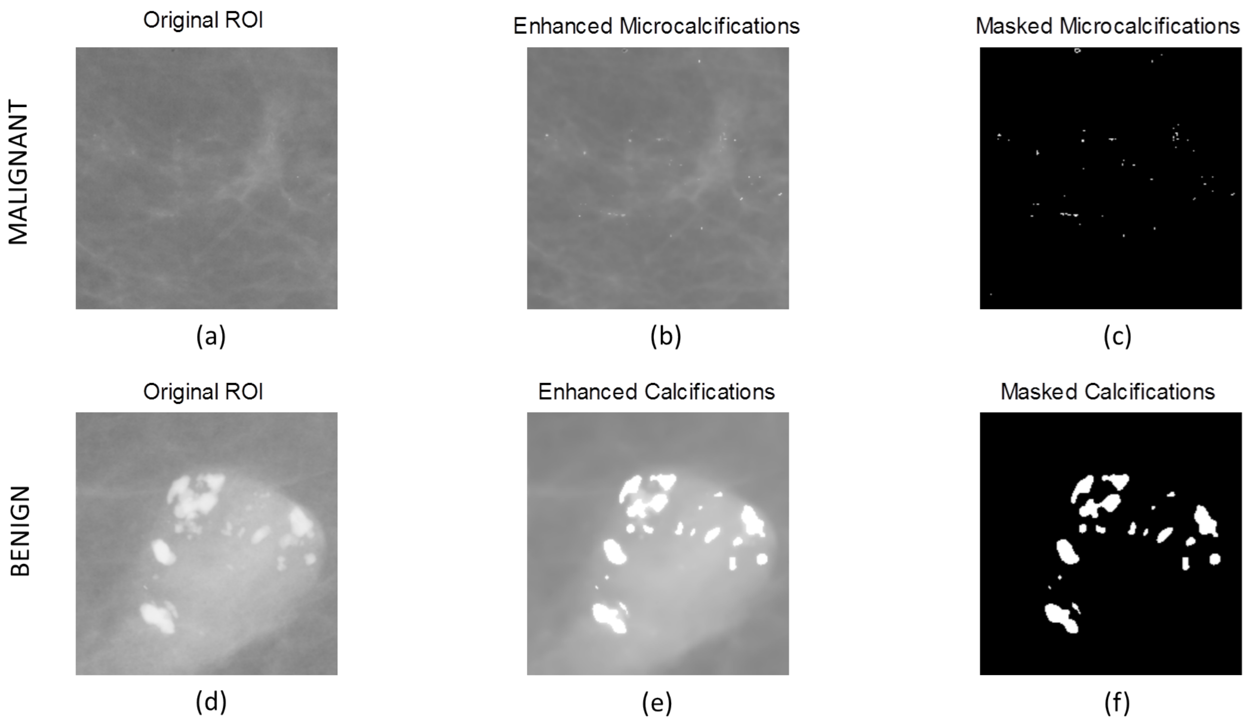

Figure 3. (a)-(c) malignant cluster, (d)-(f) benign cluster. (a),(d) Original region of interest, (b),(e) enhanced region of interest, (c),(f) identified clusters.

Starting from the images in Fig. 3 (c),(f), binary operators are applied to estimate the number of disjoint regions (N_{MC}) that appears in the binary image. As an example, in the two cases shown in Fig. 3 the values assumed by the variable N_{MC} are equal to 62 and 19, respectively. When computed on the entire set of 237 ROIs the values of the variable N_{MC} range in [1, 241], with a mean value of 38 on the malignant cases and of 8 on benign cases. By estimating a logistic regression model and then applying thresholding, the computerized diagnosis system is modeled by the function:

$$y(N_{MC}) = \begin{cases} 1 & if \dfrac{1}{1+\exp(-(\beta_0+\beta_1 \cdot N_{MC}))} > th \\ 0 & otherwise \end{cases}$$

where β_0, β_1, and th are parameters to be estimated. An output value equal to 1 means that the cluster has been assigned to the category *malignant* while a value of 0 means that the cluster has been assigned to the category *benign*. Experiments run on the 237 ROIs provided the following estimated values for the parameters along with their standard deviations: average value for β_0 equal to –4.73, standard deviation equal to 0.64, average value for β_1 equal to 0.38, standard deviation equal to 0.05. The value of threshold th has been fixed to 0.46 using Receiving Operating Characteristic (ROC) analysis. The uncertainty source that has been considered as the most relevant is that related to the selection of model parameters. Hence, in order to compute the stochastic sensitivity of such diagnostic system corresponding to a one-variation in the input variable N_{MC}, we refer to eq. (10).

Let us assume that coefficients β_0 and β_1 are affected by independent additive Gaussian uncertainty sources with zero mean and standard deviation provided by the experiments. So, for each pair N_{MCi}, N_{MCi+1}, $N_{MCi+1} - N_{MCi} = 1$, being N_{MC} a counting variable, and considering the two random processes $\beta_0 = -4.73 + \boldsymbol{N}(0; 0.64^2)$ and $\beta_1 = 0.38 + \boldsymbol{N}(0; 0.05^2)$, we ran repeated experiments collecting the corresponding pair $y_i = y(N_{MCi})$, $y_j = y(N_{MCi+1}) = y(N_{MCi}+1)$. Inserting the output values in eq. (10) and repeating for all the values of N_{MCi} in the range [1, 241–1] we obtain the estimate $\hat{S}_{N_{MCi+1}}(N_{MCi})$ reported in Fig. 4. Red and blue markers identify the actual category corresponding to each input value as reported in the histological report.

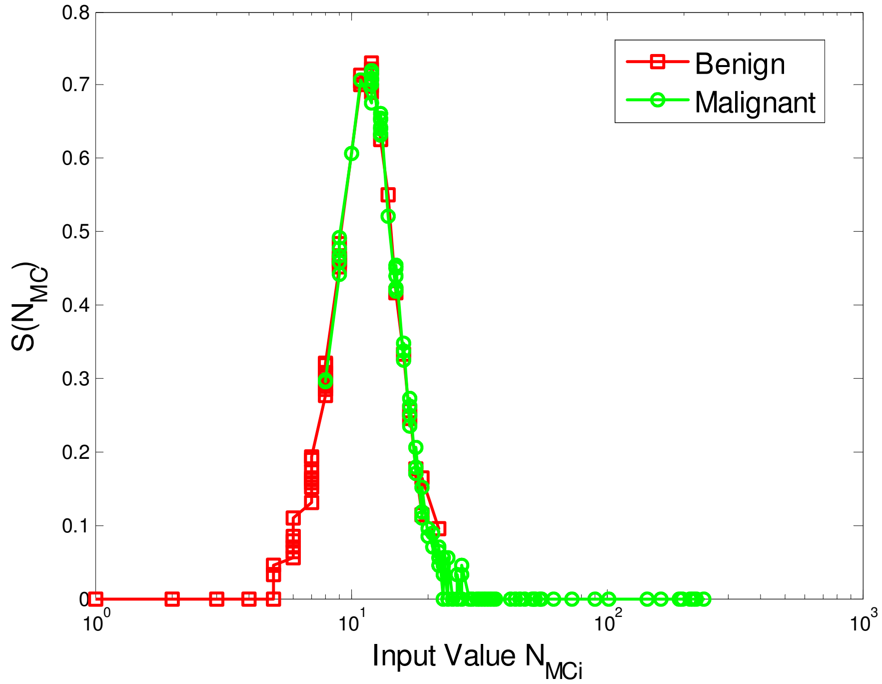

Figure 4 Estimated stochastic sensitivity for one-variation in N_{MC} in a computerized diagnostic system for breast calcifications. Red and blue markers identify benign and malignant clusters, respectively.

Results indicate that sensitivity to one-variation in N_{MC} is negligible for input values below 5 and greater than 24. On the contrary, sensitivity is remarkably higher for input values in the range [10, 15]. This is due to the fact that the most subtle clusters fall in this range, namely benign clusters with a high number of calcifications or malignant clusters with a low (probably difficult to detect) number of MCs. The analysis demonstrates that also for non-differentiable entity models with categorical output variables, sensitivity can be estimated using the procedure described in this paper.

In the specific context, the sensitivity behavior also suggests the need for the implementation of a multivariate diagnostic system receiving in input not only the number of calcifications, but also the related statistics on the MC area distribution, shape of the clusters, average distance among calcifications in the clusters, to name a few [17]. In light of these considerations, further works will certainly investigate the case of MISO systems, accounting for effects like input variables inter-correlation and cross-sensitivity (e.g., selectivity).

6. Conclusions

'Sensitivity' is a multifaceted concept, even when considered in the specific context of metrology, as its many definitions that can be retrieved from the scientific and technical literature witness. As our analysis has shown, it is unfortunate that this multiplicity produces at least partially inconsistent meanings: sorting out the concept would increase the possibility of non-ambiguous inter-subjective communication, as related in particular to the specifications of measuring instruments and systems. The strategy we have proposed and adopted here assumes the possibility and suggests the appropriateness to develop a conceptual framework in which sensitivity is qualitatively a feature of a black box behavior, quantitatively specified according to specific evaluation types. Formal definitions have been proposed and commented for both deterministic and stochastic behaviors in reference to the three basic types (interval / ratio; ordinal; nominal), in such a way that for each type 'deterministic sensitivity' is a specific case of 'stochastic sensitivity'.

When uncertainty is not negligible, the proposed definitions significantly embed a distinction between "effective" and "confounding" components, that can be simultaneously present in stochastic sensitivity and contribute to a desirable and unwanted increment of global sensitivity respectively. Two examples, about the characterization of imaging systems such as scanners for medical applications and of a computerized diagnostic system for breast calcifications in mammography, provide some hints on the usefulness of the proposed framework for an understanding of different working conditions of such devices or for the improvement of the effectiveness of such diagnostic systems in screening mammography. Future studies could be devoted both to extending this strategy to more complex systems and situations whose sensitivity could be evaluated, and to applying the strategy itself, in its purely structural components, to the construction of structured definitions of other metrologically relevant concepts such as selectivity, resolution, and repeatability.

References

[1] L. Mari, A. Mencattini, A structured strategy of concept definition in measurement: the case of 'sensitivity', Journal of Physics: Conference Series, Vol. 459, No. 1, IOP Publishing, 2013.

[2] ISO 1087-1:2000, Terminology work – Vocabulary – Part 1: Theory and application, International Organization for Standardization, 2000.

[3] JGCM 200:2012, International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM), Joint Committee for Guides in Metrology, 2012 (2008 edition with minor corrections) [http://www.bipm.org/en/publications/guides/vim.html].

[4] ISO 704:2009, Terminology work – Principles and methods, International Organization for Standardization (ISO), 3rd ed., 2009.

[5] IEC 60050, International Electrotechnical Vocabulary (IEV), International Electrotechnical Commission, different publication dates for the various parts [http://www.electropedia.org].

[6] CLSI, Harmonized terminology database, Clinical and Laboratory Standard Institute [http://www.clsi.org/AM/Template.cfm?Section=Harmonized_Terminology_Database].

[7] OECD, Glossary of statistical terms, Organisation for Economic Co-operation and Development [http://stats.oecd.org/glossary].

[8] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global sensitivity analysis. The primer, Wiley, 2008.

[9] A. Giordani, L. Mari, Property evaluation types, Measurement, 45, 437–452, 2012.

[10] JGCM 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, 2008 (GUM, originally published in 1993) [http://www.bipm.org/en/publications/guides/gum.html].

[11] A. Papoulis, S. Unnikrishna Pillai, Probability, random variables, and stochastic processes, McGraw-Hill, 2002.

[12] P.J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, J. American Statistical Association, 88, 424, 1273–1283, 1993.

[13] A.K. Jain, Fundamentals of digital image processing, Prentice Hall, 1989.

[14] C.J. D'Orsi, L.W. Bassett, W.A. Berg, S.A. Feig, V.P. Jackson, D.B. Kopans, BI-RADS: Mammography, American College of Radiology, 4^{th} edition, 2003.

[15] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The Digital database for screening mammography, proc. 5th Int. Workshop on Digital Mammography, pp.212–218, Medical Physics Publishing, 2001.

[16] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio, F. Caselli, Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing, IEEE Trans. Instrumentation and Measurement, 57, 7, 1422–1430, 2008.

[17] A. Ferrero, S. Salicone, A. Mencattini, G. Rabottino, M. Salmeri, Uncertainty evaluation in a fuzzy classifier for microcalcifications in digital mammography, IEEE Instrumentation and Measurement Technology Conference (I^2MTC), pp.1250–1255, 2010.