Novel mathematical and statistical approaches to uncertainty evaluation

A guide to decision-making and conformity assessment

Deliverable 3.3.1

Work Package WP3: Decision-making and conformity assessment

A report of the EMRP joint research project
NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"

1. Work package WP3	**2. Deliverable number** D3.3.1	**3. Reporting date** May 2015
4. Title A guide to decision-making and conformity assessment		
5. Author(s) SP, JV, INRIM, LGC, LNE, NPL, PTB, SCA		**6. Lead author (e-mail)** Leslie.Pendrill@sp.se
7. Contributing researchers (institutes) H Karlsson (JV), N Fischer (LNE), S Demeyer (LNE), A Allard (LNE)		
8. Other contributing work packages WP1 and WP2		**9. Researchers in other WPs**
10. Supplementary notes		

11. Abstract
Decisions based on measurement, such as tests of the significance of differences and of conformity, are currently made in many important application areas without a clear and harmonised basis for assessing the impact and sharing the risks that arise from measurement uncertainty. The need for improved insight into setting a 'fit-for-purpose' level of measurement effort *prior* to performing a given task is increasingly appreciated on all sides – from industrial, academic & societal end users as well as national metrology institutes. Better mutual understanding between the metrologist and those ordering measurement tasks about the import and limitations of the measurements when making decisions of conformance *after* a measurement is also needed. Generic guidance including some original material presented in this document, addresses the role of measurement uncertainty in decision-making and conformity assessment for multivariate cases, regression and computationally expensive models illustrated for a number of case studies such as fire engineering, healthcare and electricity energy metering. The guide is available electronically via the EMRP NEW04 (2012 – 5) "Novel mathematical and statistical approaches to uncertainty evaluation" JRP website (http://www.ptb.de/emrp/1371.html).

12. Keywords: Measurement - Uncertainty - Multivariate – Decision theory – Conformity assessment - Risks - Costs

Contents

List of Tables 4

List of Figures 4

1 Decision-making and conformity assessment 6
1.1 Introduction 6

2 Appraisal of existing guides 7
2.1 General role of measurement uncertainty in decision-making and conformity assessment 7

2.1.1 Conformity assessment 8

2.1.2 Assuring measurement quality in conformity assessment 8

2.2 Separating production & measurement errors 9

2.2.1 Separating production & measurement errors - conceptually 9

2.2.2 Separating production & measurement errors - numerically 9

2.2.3 Measurement conformity assessment. Limits on measurement capability factors 10

2.3 Deciding if entity is within specification 11

3 Conformity assessment for computationally expensive systems 14
3.1 Computationally expensive systems 14

3.2 Fire engineering 14

3.2.1 Results of sequential co-kriging for 2D fire case study 16

4 Conformance probabilities for multivariate cases 21
4.1 Post Office (PO) parcel problem 22

4.1.1 PO Problem: Specifications 22

4.1.2 Measurements: PO Problem 23

4.1.3 Conformance probabilities: PO Problem 23

4.1.4 Decision rules: PO Problem 24

4.2 Software for conformity assessment in multivariate cases 26

4.2.1 Installation of software 26

4.2.2 Functionalities 27

4.3 Multivariate case study of healthcare products 31

4.3.1 Conformance probabilities in multivariate cases: healthcare study 31

4.3.2 Specific observations X assessment: healthcare study 32

4.3.3 Specific PC assessment: healthcare study 34

4.3.4 Specific PC Region: Healthcare study 34

4.3.5 Results of PCR: healthcare study 37

5 Costs and Economic Risks in Conformity Assessment 39
5.1.1 Introducing cost into Conformity Assessment risks 40

5.1.2 Balancing test costs against consequence costs 41

5.1.3 Optimised uncertainties: by variable and by attribute 43

6 Concluding remarks 46
Acknowledgments 46
References 47

List of Tables

Table 3.1 Recommendations for the choice of statistical methods to compute the probability of conformity……14

Table 3.2 Summary results of the probability of non-conformity $\hat{p}_f$, its associated uncertainty $u(\hat{p}_f)$ and its coefficient of variation (relative uncertainty $cv(\hat{p}_f)$ obtained with various kriging based methods……17

Table 3.3 Summary results of the probability of non-conformity $\hat{p}_f$, its associated uncertainty $u(\hat{p}_f)$ and its coefficient of variation (relative uncertainty) $cv(\hat{p}_f)$ obtained at each iteration of the sequential procedure. Iteration 0 gives the baseline results obtained with the initial database (9 points)……...18

Table 4.1 Summary of calculated bivariate X conformance probabilities in the healthcare study: specific conformity……33

Table 4.2 Summary of calculated bivariate PC conformance probabilities in the healthcare study: specific conformity……36

List of Figures

Figure 2.1 Decision-making: Significant difference between repeated measurements?.....7

Figure 2.2 Conformity assessment of test result with respect to a lower specification limit, L_{SL}…….....11

Figure 2.3 Decision table……12

Figure 3.1 Flowchart of the sequential co-kriging procedure……..16

Figure 3.2 Initial database and initial level plot of the probability of excursion function. In green, the points that are only evaluated by CFAST; in red, the points that are evaluated by CFAST and FDS……..17

Figure 3.3 Contour plot (zoom) of the initial probability of excursion (a) and the updated probability of excursion after 2 points have been added at the first iteration of the sequential co-kriging procedure (b)……19

Figure 3.4 Contour plot of the updated probability of excursion at the second iteration of the sequential co-kriging procedure (a) and at the third iteration (b)……19

Figure 3.5 Link to D3.2.2 Conformity assessment for computationally expensive systems: application to fire engineering safety……...20

Figure 4.1 Conforming and non-conforming regions for a bivariate Normal distribution [1]……...21

Figure 4.2 Wedge-shaped region denoting tolerance region for Post Office parcel problem……22

Figure 4.3 Screenshot of main menu of software "Conformance probability"……..26

Figure 4.4 Different versions of the MCR. R2012b (8.0) for Windows (32-bit or 64-bit) should be downloaded and installed before the first use of the software…….27

Figure 4.5 Main menu of the software for conformity assessment……27

Figure 4.6 Example of computation of conformance probability based on joint probability distribution…….28

Figure 4.7 Example of computation of conformance probability based on marginal probability distributions……..29

Figure 4.8 Example of computation of conformance probability based on conditional probability distributions…….30

Figure 4.9 Parameters for Gibbs algorithm…….30

Figure 4.10 Correlation plot between a pair of explanatory variables X of six topical formulations [physical material properties: (x-axis) (friction) and (y-axis) (adhesion)] and corresponding specification limits for the healthcare case studies [2] …….33

Figure 4.11 Correlation plot between a pair of explanatory variables **X'** of six topical formulations [physical material properties: (x-axis) $\mu_{BS}FB_{0min}$ (friction) and (y-axis) $A_{BS}FB_{0min}$ (adhesion)] and PC corresponding specification limits for the healthcare case studies [2]. Dispersion intervals on each point are based on correlation matrix (measurement uncertainty neglected)……..36

Figures 4.12 Relations in the healthcare study between sensory response (Slipperiness) (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples, measured (y-axis) and predicted (x-axis) from explanatory variables (ForceBoard, BIOSKIN), respectively from PCR analyses [2]……37

Figure 4.13 Link to D3.2.4 Case study of health care product…….38

Figure 5.1 Different costs and income of conformity-assessed entity from the point of view of the supplier. Green represents a profit, while red represents a loss [adapted from [3]]……40

Figure 5.2 Including cost and impact in operating characteristics and optimised dispersion…….42

Figure 5.3 Costs as function of attribute sampling size for electricity meters (red line: eq. 5.2; blue line testing costs)……44

Figure 5.4 Link to Conformance assessment of electrical energy meters investigated by risk analysis – a case study ……45

1 Decision-making and conformity assessment

1.1 Introduction

Decisions based on measurement, such as tests of the significance of differences and of conformity, are currently made in many important application areas without a clear and harmonised basis for assessing impact and sharing the risks that arise from measurement uncertainty. The need is increasingly appreciated on all sides for improved insight into setting a 'fit-for-purpose' level of measurement effort *prior* to performing a given task; and better mutual understanding between the metrologist and those ordering measurement tasks about the import and limitations of the measurements when making decisions of conformance *after* a measurement.

The present document is intended as a Guide to the state-of-the art of *Uncertainty Evaluation in Decision Making and Conformity Assessment*, and accompanies the other Guides – to *Bayesian Inference for Regression Problems* [4] and to *Uncertainty Evaluation for Computationally Expensive Models* [5] – as part of the EMRP NEW04 project (2012 – 5) "Novel mathematical and statistical approaches to uncertainty evaluation" which belongs to the European Metrology Research Programme (EMRP, FP7 Art. 185), jointly funded by the EMRP participating countries within EURAMET [6] and the European Union.

Generic guidance including some original material presented in the present document - intended for industrial, academic & societal end users as well as national metrology institutes - addresses the role of measurement uncertainty in decision-making and conformity assessment for multivariate cases, regression and computationally expensive models illustrated for a number of case studies such as fire engineering, healthcare and electricity energy metering.

2 Appraisal of existing guides

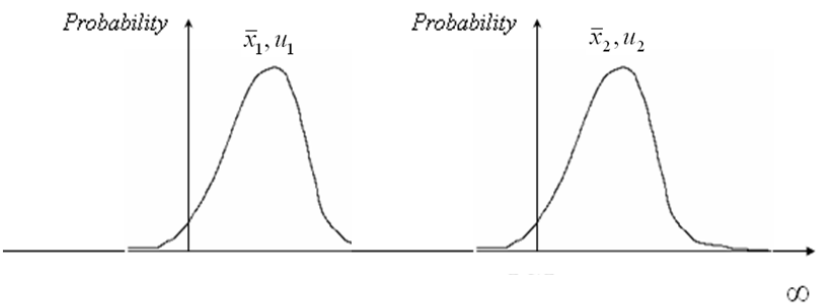
2.1 General role of measurement uncertainty in decision-making and conformity assessment

> "No single method used to integrate uncertainty of measurement into decision making process. Decision rules differ between products, fields of measurement, profession and countries" [7]

Measurement uncertainty in a test result on inspection of an entity (measurement object) can:

- lead to **incorrect estimates of consequences of entity error** – e.g. in the flow of a liquid monitored by a flowmeter (as studied in NEW04 WP1, §4, [4])
- aid in assessment of **significance of apparent differences** between repeated test results [Figure 2.1]
- increase **risk of making incorrect decisions**, such as failing conforming entity or passing non-conforming entity when test result close to tolerance limit [Figure 2.2]

Decisions of conformity are currently made in many important application areas, such as environmental monitoring, the health sector and product safety testing, but without a clear and harmonised basis for sharing the risks that arise from measurement uncertainty between the consumer and the supplier.



$$P_{success} = 1 - \frac{1}{2} \cdot erfc\left[\frac{|\bar{x}_1 - \bar{x}_2|}{k \cdot \sqrt{(u_1^2 + u_1^2)}}\right]; erfc[z] = \frac{2}{\sqrt{\pi}} \cdot \int_z^{\infty} e^{-t^2} \cdot dt$$

Significant difference?

Figure 2.1 Decision-making: Significant difference between repeated measurements?

2.1.1 Conformity assessment

Not all guides are so clear about why conformity assessment is important.

"Conformity assessment involves a set of processes that show your product, service or system meets the requirements of a standard [8]:

- provides confidence for consumer that requirements on products and services are met
- provides the producer and supplier with useful tools to ensure product quality
- often essential for reasons of public interest, public health, safety & order, protection of environment and consumer and of fair trading."

Major recent guides concerning the role of *measurement uncertainty* in Conformity Assessment, which is the main subject of the current work, give different emphasis about the motivation for considering Conformity Assessment:

- In the introduction to the ISO standard 10576-1 [9], reference is made to "supplier's declaration,… third party certification,…, health-related characteristics…".
- The French standard FD x07-022 [7] states in its Foreword: "We are continually faced with situations in which we have to take decisions or make choices…: is the product… 'conforming'?". That standard usefully provides a number of diverse examples of Conformity Assessment in its Table 1, including of a batch of products using a sampling plan; type acceptance; comparison of a given characteristic of two products; validation that a given method is suited to a given need, etc.
- The Scope of the EURACHEM guide [10] is described as "applicable to decisions on compliance with regulatory or manufacturing limits".
- In the Introduction to the recent JCGM 106 guide [11] mention is made of "accreditation of Conformity Assessment bodies and the use of conformity in facilitating trade".

2.1.2 Assuring measurement quality in conformity assessment

The role of measurement uncertainty as a measure of measurement quality in decision-making and conformity assessment naturally follows the essential steps of any quality assurance loop [Deming, ISO 9001]:

a)	Define entity and its quality characteristics to be assessed for conformity with specified requirements (e.g. U_{SL} – upper specification limit; L_{SL} – lower specification limit; *MPE* – maximum permissible (entity) error).
b)	Set corresponding specifications on the measurement methods and their quality characteristics (such as maximum permissible uncertainty and minimum measurement capability) required by the entity assessment at hand
c)	Produce test results by performing measurements of the quality characteristics together with expressions of measurement uncertainty.
d)	Decide if test results indicate that the entity, as well as the measurements themselves, is within specified requirements or not.
e)	Assess risks of incorrect decisions of conformity.
f)	Final assessment of conformity of entity to specified requirements in terms of impact.

2.2 Separating production & measurement errors

Since neither the production nor measurement processes are perfect, there will always be some dispersion in the observed product value either for repeated measurements of one item or for measurements of a series of items. Practically in cases where measurement dispersion is of comparable size to actual product variation, it can be difficult to separate these [12], [13]. This separation is however essential:

- If one overestimates measurement dispersion, then actual dispersion in product values will not be detected and thus lead to poorer product quality. Additional costs will also be incurred if it is decided, on the basis of estimated poor measurement quality, to spend more on (unnecessarily) improving measurement resources.
- An underestimation of measurement dispersion will lead to unnecessary adjustment of the production process and thus to increased production costs together with poorer product quality, where spurious measurement dispersion is transferred to product dispersion.

Confidence in the measurements performed in the conformity assessment of any entity (product, process, system, person or body) can be considered sufficiently important that the measurements themselves will be subject to the steps a) to f) above [§2.1.2] as a kind of metrological conformity assessment 'embedded' in any product conformity assessment.

2.2.1 Separating production & measurement errors - conceptually

Some of the difficulty in separating production and measurement dispersion has to do with a lack of clarity in concepts, definitions and nomenclature which arises at the interface where two disciplines – Metrology and Conformity Assessment – meet.

Two principally distinct, but closely related and easily confounded concepts coming from the two disciplines are, respectively:

- A 'measurand' is a quantity intended to be measured [14] [VIM]
- A 'quality characteristic' is a quantity intended to be assessed

Clear definitions of key Conformity Assessment concepts such as 'entity' and 'quality characteristic' can be found for instance in ISO 10576-1:2003 [9] and their systematic use in the metrological literature can reduce confusion in the field.

The entity subject to Conformity Assessment may be a single item, a collective sample of items or maybe not even a physical object but a service. Irrespective of which entities are subject to Conformity Assessment, it is important to specify the assessment target as clearly as possible: 'Global' conformity denotes the assessment of populations of typical entities, while 'specific' Conformity Assessment refers to inspection of single items or individuals, as defined by Rossi and Crenna [13].

2.2.2 Separating production & measurement errors - numerically

There are several ways of ensuring that limited measurement quality does not adversely affect Conformity Assessment. Risks and the consequences of incorrect decision-making in Conformity Assessment can be minimised with the following three steps:

A. set limits on maximum permissible measurement uncertainties (equivalently, minimum measurement capability) and on maximum permissible consequence costs at the specification stage of any task [§2.2.3];
B. agree on acceptable locations of the uncertainty interval with respect to a specification limit [§2.3];
C. optimise measurement uncertainty proactively, ***ahead*** of a series of measurements, by designing experiments so that the sum of costs of measurement and of incorrect decisions of conformity is at a minimum [§5].

2.2.3 Measurement conformity assessment. Limits on measurement capability factors

No harmonised limits on measurement capability [15]

Corresponding specified requirements on measurement are, respectively:

- limits on **maximum permissible measurement uncertainty** (or, equivalently, minimum measurement capability) when testing product.
- limits on **maximum permissible error** in the indication of the measurement equipment/system intended to be used in the measurements when testing product;

A first step to minimising the effects of less than perfect measurement quality on Conformity Assessment is simply to set a limit ***proactively***, before starting the measurements, on how large measurement uncertainty is allowed to be [16]. This limit is often expressed as a maximum permissible uncertainty or 'target uncertainty', $MPU = 1/C_{m,min}$ in terms of a corresponding minimum measurement capability.

In a manner analogous to process control, a measurement capability index, C_m, can be defined in terms of estimated measurement variations as:

$$C_m = \frac{U_{SL} - L_{SL}}{M \cdot u_m} \tag{2.1}$$

– with standard measurement uncertainty u_m and typically $M = 4$ (corresponding to a *coverage* factor, $k = 2$ and 95% confidence). [Note that it is more usual in Metrology to use a coverage factor of 2, while in SPC the corresponding coverage factor for process capability is often 3.]

In various sectors of Conformity Assessment, different limits on measurement capability have become established, with $C_{m,min}$ ranging from typically 3 to 10. A common limit to ensure that measurement quality variations are small is $\frac{u_m}{s_p} < 30\%$, as in Measurement System Analysis (MSA) in the automobile industry [17], for instance. Even in qualitative measurement, such as made on ordinal scales and using questionnaires, reference is made to minimum values of a reliability coefficient, given by: $R_\theta = \frac{\text{True variance}}{\text{Observed variance}} = \frac{\text{var}(\vartheta')}{\text{var}(\theta)} = \frac{\text{var}(\theta) - \text{var}(\varepsilon_\theta)}{\text{var}(\theta)}$ where $\theta = \vartheta' + \varepsilon_\theta$ for an attribute θ of an entity

(object or item) with a value ϑ' when the measurement error ε_θ is zero. In the literature [18], a recommended value of a reliability coefficient is 0.8, corresponding to a 'separation' of G = 2, or in other words, a measurement uncertainty $\sqrt{\mathrm{var}(\varepsilon_\theta)}$ not larger than half the object standard deviation $\sqrt{\mathrm{var}(\theta)}$.

2.3 Deciding if entity is within specification

No consensus about agreeing on acceptable locations of uncertainty interval with respect to specification limit [7]

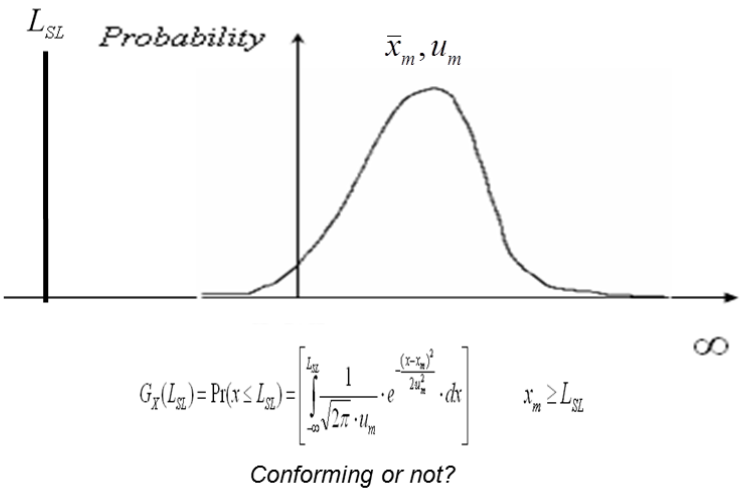


Figure 2.2 Conformity assessment of test result with respect to a lower specification limit, L_{SL}

The JCGM 106:2012 [11] document for instance addresses the technical problem of calculating the conformance probability and the probabilities of the two types of incorrect decisions – that is, supplier (β) and consumer (α) risks expressed in percentages, – given a probability density function (PDF) for the measurand, the tolerance limits and the limits of the acceptance interval. The decision matrix, **P**, in this simplest, dichotomous case is:

$$\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \tag{2.2}$$

where the diagonal elements give the probabilities of making the correct decisions and the off-diagonal elements, the risks of incorrect decisions, as illustrated in the decision table shown in Figure 2.3.

		Conformity	
		CONFORM/PASS	NON-CONFORM/PASS
Inspection		CONFORM/FAIL	NON-CONFORM/FAIL

Figure 2.3 Decision table

In the simplest case where the prior state of an entity (the measurement object) being observed is known to be in a particular category M, $p_k = \begin{cases} 1; k = M \\ 0; k \neq M \end{cases}$, for the stimulus variable, then the probability, q_c, of classifying the response (R) of the measurement system in a category c, is given by the accumulation of probabilities [19] ,

$$q_c = \sum_{k=1}^{2} p_k \cdot P_{c,k} = 1 - \alpha = P_{success}. \tag{2.3}$$

Equation (2.3) links the response variable (probability of successful categorization, $P_{success}$, in the decision-making) to the risks, α, of incorrect decisions arising from uncertainty in measurement of the explanatory variable.

These decision risks can be modelled for two kinds of human-based perception, identification (figure 2.2) and choice (figure 2.1) as dealt with in psychophysics [20],[21]. **Identification** involves in the dichotomous case a yes-no detection – is the stimulus within tolerance or outside a region of permissible values specification limit? The decision ('consumer') risk, α, is in this case estimated as the cumulative distribution function (CDF) beyond the specification limit (U_{SL}, say) on the explanatory variable, x, of the initial set of observations [15],

$$\alpha = G_X(U_{SL}) = \Pr(x \geq U_{SL}) = \left[\int_{U_{SL}}^{\infty} \frac{1}{\sqrt{2\pi} \cdot u_m} \cdot e^{-\frac{(x - x_m)^2}{2 \cdot u^2_m}} \cdot dx \right]; x_m \leq U_{SL} \tag{2.4}$$

This discussion can be extended to the multinominal, polytomous case, as required.

The cumulative probability of a 'correct' decision can in turn be related to the change of entropy on information transmission [22]. This can be interpreted as the combined process of observation (measurement) of an explanatory variable and response (decision-making), linearized using logistic regression. The logistic regression link function:

$$z = \log\left[\frac{P_{success}}{1 - P_{success}}\right] \tag{2.5}$$

can in fact be derived [23], [24], [25] in terms of the distance, *z*, between two stimuli along the explanatory variable dimension. Examples of this distance include the difference between (i) the intrinsic quality of a product and the leniency of a customer; (ii) the level of challenge of a certain task and the ability of a person to tackle the challenge; (iii) the ability of a indenter and the hardness of a material; (iv) the ability of a bullet to penetrate and the degree of resistance to penetration offered by a bullet-proof vest, to name just a few examples. This approach is applied in the multivariate case study of healthcare products given in §4.3.

This generalisation connects the measurement uncertainty of qualitative observations [26] to decision risks and covers not only explicitly non-linear responses but also poorly known responses to the explanatory variable, perhaps allowing only less quantitative appraisals, such as on an ordinal or even a merely nominal scale.

In an important sense, this approach to measurement uncertainty puts more emphasis on the risks of incorrect decisions arising from uncertainty and how these vary with the explanatory variable, in contrast to the traditional metrological evaluation of the measurement result (output of the measurement system) accompanied by an estimation (traditionally a standard deviation) of an 'uncertainty' which only subsequently is used to derive decision risks. As is well-known, the logistic regression approach does not rely on assuming a particular probability density function but instead can be expressed in terms of the more general concept of information entropy, which can handle even expert elicitation and qualitative observations.

3 Conformity assessment for computationally expensive systems

In a study of statistical methods addressing conformity assessment for **computationally expensive systems**, recent work has focussed on the probability of conformity that output quantities exceed a regulatory threshold using in particular Monte Carlo methods, as well as importance sampling and sequential sampling. A method to combine expensive runs with fast approximations has also been addressed. The strengths, limitations and applicability of each method are stressed and key papers cited. A case study on fire engineering employs these new methods.

3.1 Computationally expensive systems

When dealing with a computational code, one may be interested in propagating the uncertainties related to the input variables to estimate the uncertainty associated with the output variable.

In the decision theoretical framework pertaining to conformity assessment, one is interested also in the position of the output variable with respect to a given threshold (regulatory threshold...). The problem of knowing whether a computationally expensive model exceeds a given threshold is very common for reliability analysis and safety-critical applications such as aerospace, nuclear power stations and civil engineering (models of bridges and buildings etc). The threshold can be interpreted as a lower specification limit defining a one-sided tolerance interval.

The statistical methods that are usually used to deal with the propagation of uncertainties may need to be adapted to compute the probability of exceeding a threshold. This is particularly the case when the probability is small or when the code is computationally expensive. D3.1.4 reports statistical methods from the literature to address conformity assessment for computationally expensive systems. These methods were initially developed for reliability assessment based on computer experiments. In D3.1.4 (as in the following deliverables) the probability of conformity is assessed as the probability that output quantities exceed a regulatory threshold. D3.1.4 focuses on Monte Carlo methods based on a large number of calls of, either the code or a metamodel of the code. Other methods addressed are importance sampling which is a variance reduction alternative to Monte Carlo sampling of the code and sequential sampling which builds sequentially a learning database for metamodelling by sampling points in the (unknown) conformity domain. A method to combine expensive runs with fast approximations is also addressed.

Table 3.1 Recommendations for the choice of statistical methods to compute the probability of conformity [Table 2, D3.1.4]

	direct sampling	**surrogate sampling**
target region sampling	importance sampling	metamodel, optimized learning design
whole domain sampling	Monte Carlo sampling	metamodel, LHS learning design

3.2 Fire engineering

D3.2.5 reports probability of conformity estimates for the fire engineering case study described in D3.1.3 obtained with two statistical methods for conformity assessment from D3.1.4. The direct Monte Carlo method provides baseline results with CFAST code, obtained with a high computational cost. The metamodel-based importance sampling method (on CFAST code) has been implemented to reduce the computation time, while correcting for the metamodel approximation. Several

configurations are studied to find a good compromise between the accuracy of the estimates and the computation time.

D3.2.5 aims at computing the probability that the upper layer temperature exceeds the 200°C threshold for the fire engineering case study. Two statistical methods from deliverable D3.1.4 have been implemented in this report. The direct Monte Carlo method is used to provide baseline results. For large fires, the non-conformity probability estimate is 36,8%. The metamodel-based importance sampling method shows promising results for the fire engineering case study. For well-chosen combinations of the number of training points and simulation points, the method is able to produce an estimate of the target probability at a much lower cost than the direct Monte Carlo method with the same accuracy. This relies on an optimization of the number of calls to the code CFAST (done manually in this study). It is important to note that this method corrects for the metamodel approximation by importance sampling. The correction is a multiplicative factor whose computation requires predictions from a metamodel, calls to the code and smart samples from the input domain obtained with importance sampling.

D3.2.6 extends the scope of D3.2.5 to assess the conformity of computationally expensive systems (FDS) using cheaper approximations (CFAST) based on modelling the relationship between the two codes (CFAST and FDS) and the behaviour of the codes viewed as black boxes evaluated at input points using co-kriging [see [27] and [28]]. Here the kriging-based [see [29] and [30]] results obtained with numerous CFAST simulations are sequentially corrected by more reliable but very costly and parsimonious FDS simulations sampled in the (unknown) target region. This so-called sequential co-kriging method is applied to a 2D version of the case study (inputs: fire area Af and characteristic heat release rate Qdotsec) with the remaining 5 variables fixed at their mean values.

Sequential co-kriging consists in sequentially adding points to the learning databases D_1 and D_2 (with $D_2 \subseteq D_1$ for computational convenience) that will be evaluated by the two code levels, respectively CFAST and FDS, to decrease the relative uncertainty (coefficient of variation $cv(\hat{p}_f) = \frac{u(\hat{p}_f)}{\hat{p}_f}$) of the estimated probability of non-conformity $\hat{p}_f$ with associated uncertainty $u(\hat{p}_f)$. The procedure is repeated until a targeted relative uncertainty is reached (cv.seuil).

Sequentially adding points to a learning database (also called sequential planning) according to a space filling design brings a global reduction of uncertainty [31]. However this approach is not necessarily pertaining to the problem of determining the failure area and its probability accurately. Indeed when the target area has a low probability, points have a very low probability to fall in. This is particularly critical in high dimensional problems. Specific sequential methods targeting the failure area are required. Such methods have already been developed to estimate a probability of failure based on a kriging model of a computer code [32],[33]. (Since the posterior predictive distribution of FDS outputs denoted $\tilde{y}_2(.)|y,\hat{\gamma}$ given observations $\boldsymbol{y}$ and estimated parameters $\hat{\gamma}$ is a Gaussian process such methods can be applied to $\tilde{y}_2(.)|y,\hat{\gamma}$.

The sequential co-kriging procedure returns an estimate of the probability of failure (non conformity) of FDS bases of a very limited number of FDS runs thanks to numerous cheap approximations provided a valid relationship between the two codes. A flowchart of the complete sequential co-kriging procedure is displayed at Figure 3.1:

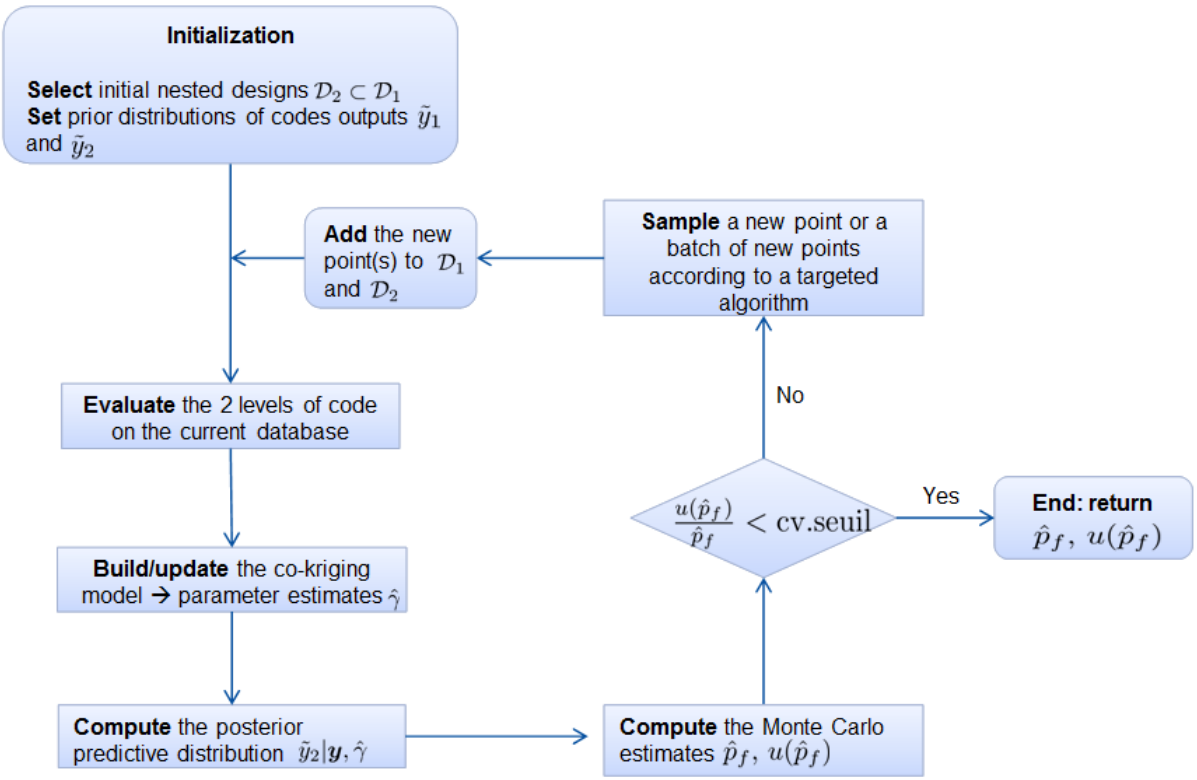


Figure 3.1 Flowchart of the sequential co-kriging procedure

3.2.1 Results of sequential co-kriging for 2D fire case study

Preliminary remark: In this report, the additional FDS points required to carry out the sequential co-kriging method have been replaced by the predicted mean of the current metamodel of FDS at each iteration. Actual additional FDS points will be displayed in the paper following this deliverable. Thus, the following interpretation of the results should not lead to an overvalued efficiency of the sequential co-kriging method.

The initial database evaluated by CFAST comprises 9 points displayed in Figure 3.2, among them 5 points (in red) are evaluated by FDS. Voluntarily, no FDS points exceed the threshold. The contour lines of the probability of excursion (function of the inputs giving the pointwise probability of exceeding the threshold) are also plotted. Their dispersion represents the uncertainty on the location of the frontier between conformity and non-conformity given the observations and the co-kriging model. For instance the predicted 0.5 line indicates that there is 50% chance that the non-conformity domain lies above this line and the predicted 0.9 line indicates that there is 10% chance that the non-conformity domain lies above this line.

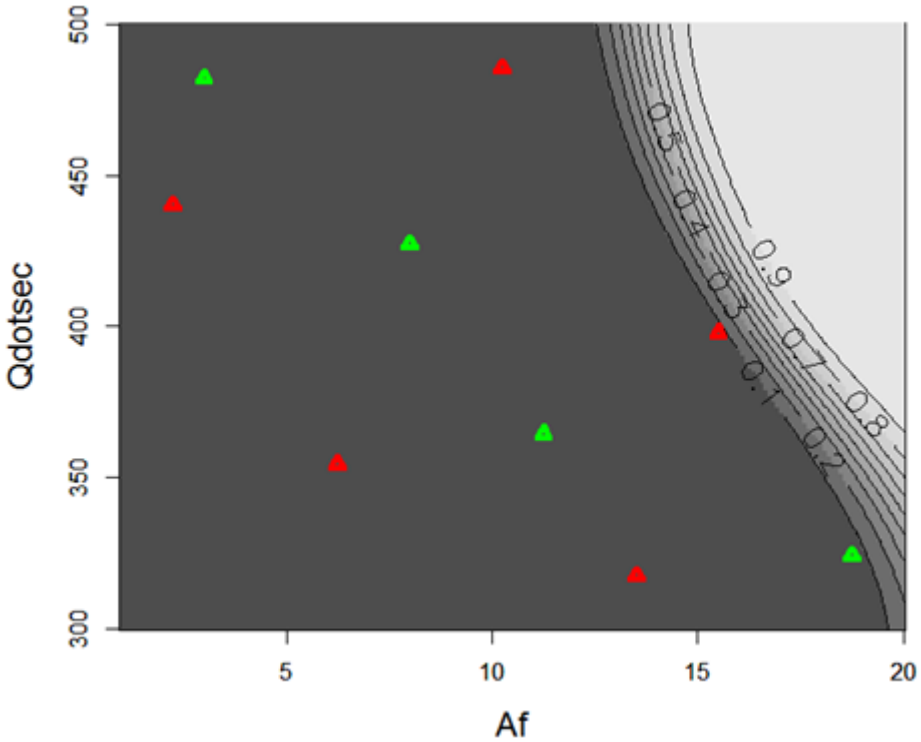


Figure 3.2 Initial database and initial level plot of the probability of excursion function. In green, the points that are only evaluated by CFAST; in red, the points that are evaluated by CFAST and FDS.

A comparison of various kriging based methods has been carried out on the fire engineering case study to show the influence of the number of points in the database and the influence of their location on the estimates of the probability of conformity and its accuracy (relative uncertainty cv). Results are displayed in Table 3.2.

number of FDS runs	type of analysis	$\hat{p}_f$	$u(\hat{p}_f)$	$cv(\hat{p}_f)$
5	kriging	0.1634	0.0648	0.3968
5	co-kriging	0.172	0.022	0.13
100	kriging	0.1349	0.0056	0.0415
100	Monte Carlo	0.13	0.033	0.25
$5 + 3 \times 2$	sequential co-kriging	0.1710	0.0013	0.0078

Table 3.2 Summary results of the probability of non-conformity $\hat{p}_f$, its associated uncertainty $u(\hat{p}_f)$ and its coefficient of variation (relative uncertainty $cv(\hat{p}_f)$ obtained with various kriging based methods

The point estimate produced by the kriging model ($\hat{p}_f = 0.1349$) built on 100 points is very close to the point estimate computed with the Monte Carlo method on the same points ($\hat{p}_f = 0.13$). However,

the kriging based estimate is much more accurate ($cv(\hat{p}_f) = 0.0416$ versus $cv(\hat{p}_f) = 0.25$). The reason is that kriging based estimates involve a lot of cheap simulations from the metamodel that increase the accuracy of the estimation. It is important to notice that both estimations seem to be biased (much lower values than the sequential co-kriging method) but that the kriging model tends to produce a higher value of the estimate.

For the same number of FDS runs (5 runs) co-kriging reduces the relative uncertainty of the point estimate. A relative uncertainty divided by 3 ($cv(\hat{p}_f) = 0.13$ versus $cv(\hat{p}_f) = 0.3968$) was obtained at the cost of 4 cheap CFAST simulations. This shows the positive impact of combining expensive runs with cheap runs.

Final results of the sequential co-kriging show a dramatic decrease of the relative uncertainty with respect to co-kriging results ($cv(\hat{p}_f) = 0.0078$ versus $cv(\hat{p}_f) = 0.13$) at a cost of only 6 smartly chosen new expensive runs. This relative uncertainty is also much lower than the one obtained with kriging based on 100 expensive runs ($cv(\hat{p}_f) = 0.0078$ versus $cv(\hat{p}_f) = 0.0415$). Besides these results are twice as accurate as the results obtained with 100 expensive FDS simulations ($cv(\hat{p}_f) = 0.13$ versus $cv(\hat{p}_f) = 0.25$).

Results obtained at each iteration of the sequential co-kriging procedure are displayed in Table 3.3. The first line (iteration 0) gives the co-kriging based Monte Carlo estimates obtained with the initial database (9 points including 5 expensive FDS simulations) displayed at line 2 of Table 2.3.

iteration	number of added points	$\hat{p}_f$	$u(\hat{p}_f)$	$cv(\hat{p}_f)$
0	-	0.172	0.022	0.13
1	2	0.1707	0.0076	0.0446
2	2	0.1709	0.0023	0.0132
3	2	0.1710	0.0013	0.0078

Table 3.3 Summary results of the probability of non-conformity $\hat{p}_f$, its associated uncertainty $u(\hat{p}_f)$ and its coefficient of variation (relative uncertainty) $cv(\hat{p}_f)$ obtained at each iteration of the sequential procedure. Iteration 0 gives the baseline results obtained with the initial database (9 points).

As points are sequentially added to the database, the uncertainty exhibited in Figure 3.2 and reproduced in Figure 3.3-a (zoom) is reduced. The first two points added (see Figure 3.3-b) (red crosses) have a high predicted variance in the target area, as they are far from FDS points (red triangles). Their predicted mean is close to the threshold so that the level lines interpolate the points. The lines are also narrowed in between, close to an initial FDS point, which shows the effect of this point given the additional knowledge.

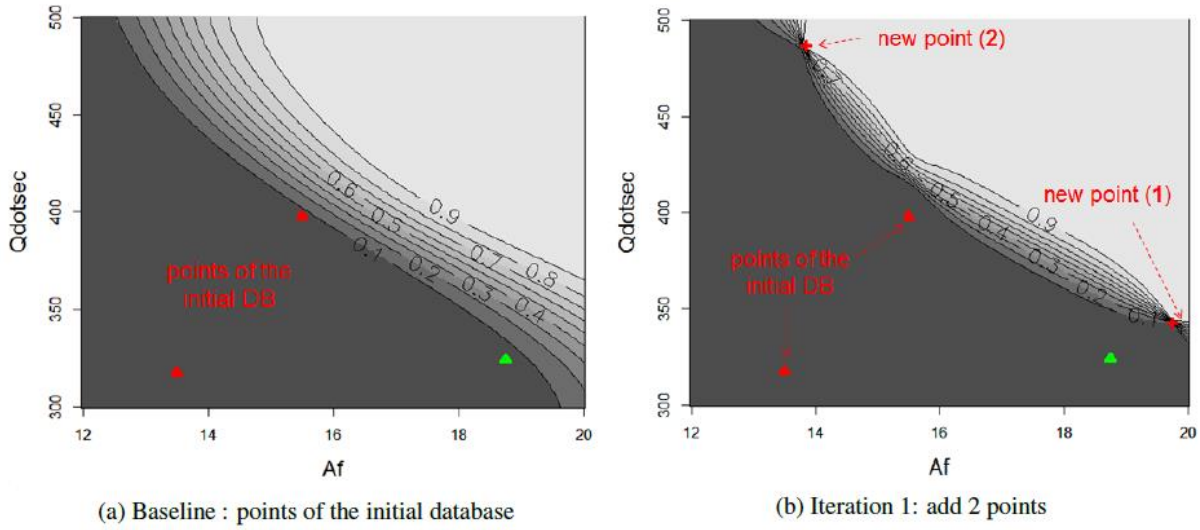


Figure 3.3 Contour plot (zoom) of the initial probability of excursion (a) and the updated probability of excursion after 2 points have been added at the first iteration of the sequential co-kriging procedure (b).

The next two points added (see Figure 3.4-a) have also a predicted mean close to 200°C and yield reduced uncertainties so that 1-probability of non-conformity zones appear. Finally, two more points are needed (see Figure 3.4-b) to reach cv<=1%, with the effect of creating a 1-probability of non-conformity zone, which means that fires starting with coordinates lying in this zone will lead to non-conformity (UL>200°C).

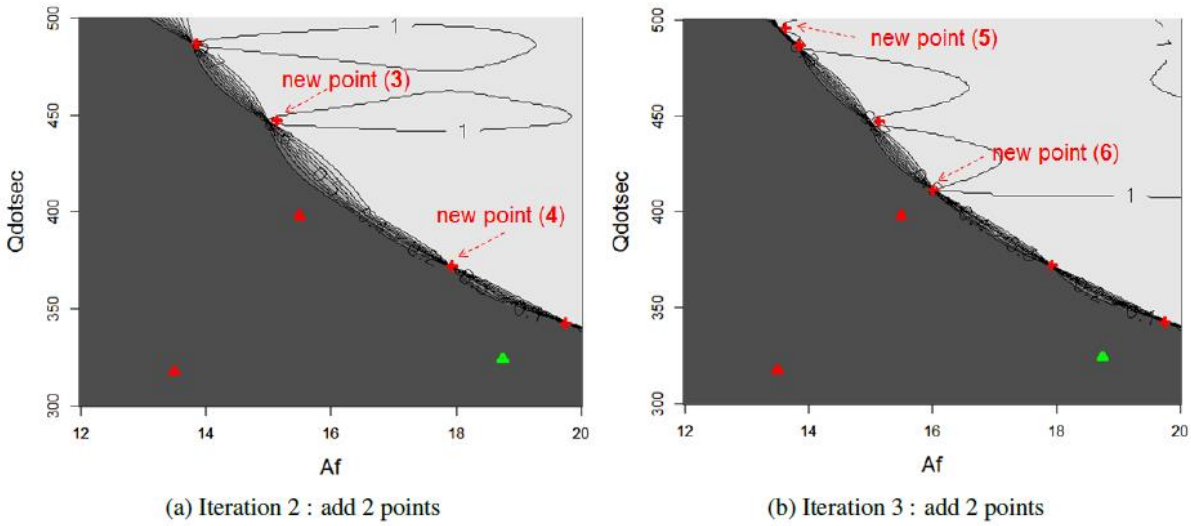
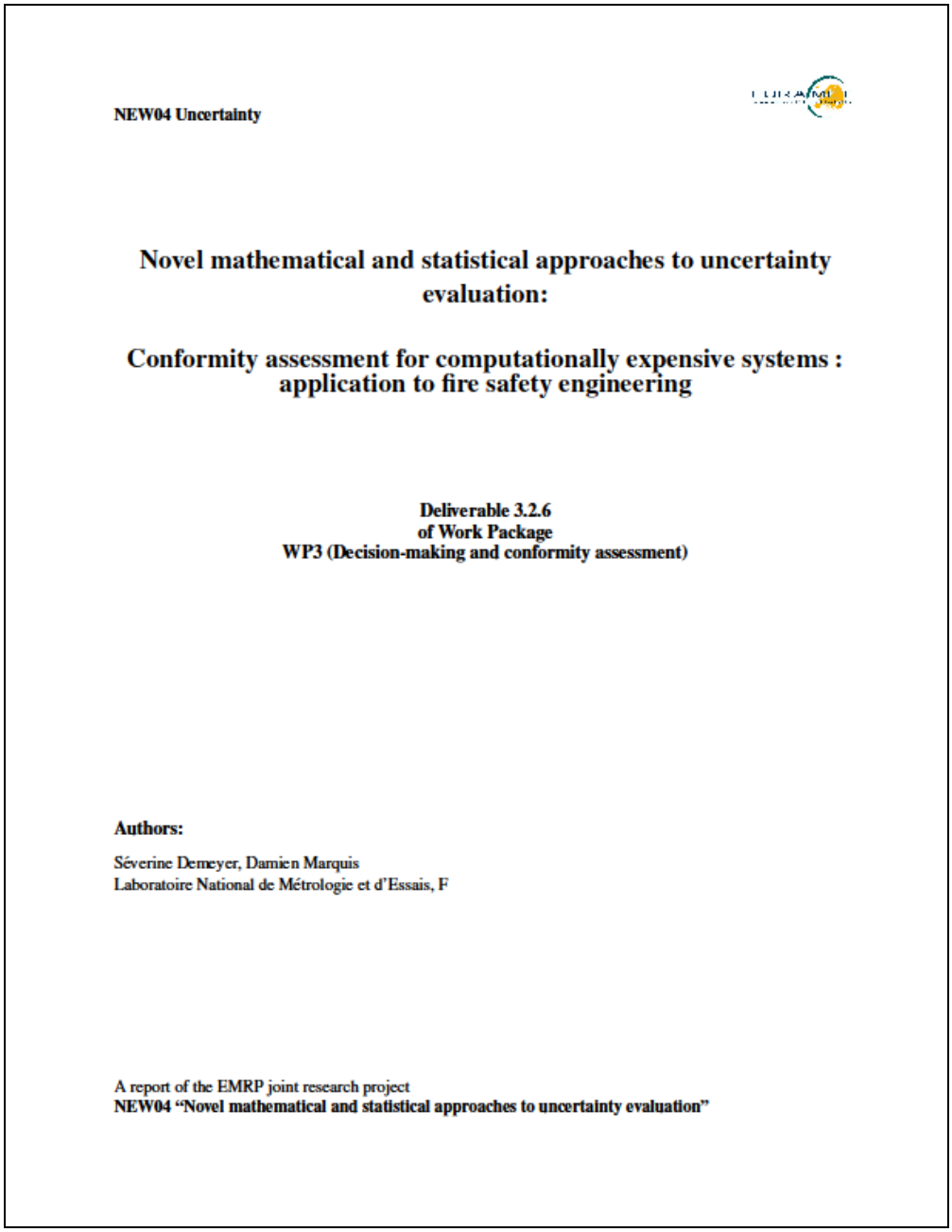


Figure 3.4 Contour plot of the updated probability of excursion at the second iteration of the sequential co-kriging procedure (a) and at the third iteration (b).

Conclusion: The methodology presented in this section to assess the conformity of an expensive computational code has proved efficient to overcome too parsimonious evaluations from an expensive code when fast approximations are available. This method builds on the kriging models usually used

to model code outputs, to improve the predictions of the kriging model of the expensive code. Another desirable feature is that the method allows smart sampling of new points targeting the non-conformity domain in order to further reduce the uncertainty of the probability of non-conformity. Although demonstrated in a simple but realistic case study, the method is flexible to take into account more complex relationships between the two codes in higher dimensional problems.



NEW04 Uncertainty

Novel mathematical and statistical approaches to uncertainty evaluation:

Conformity assessment for computationally expensive systems : application to fire safety engineering

Deliverable 3.2.6
of Work Package
WP3 (Decision-making and conformity assessment)

Authors:

Séverine Demeyer, Damien Marquis
Laboratoire National de Métrologie et d'Essais, F

A report of the EMRP joint research project
NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"

Figure 3.5 Link to D3.2.2 Conformity assessment for computationally expensive systems: application to fire engineering safety

4 Conformance probabilities for multivariate cases

New software for assessing the **conformance probability for bivariate cases** has been developed, implemented, and validated with a number of case studies in NEW04 [§4.2]. One such case study is on **health care products** [§4.3], which has reported new analyses of decisions of conformity based on multivariate (surface topography, friction and hardness) and qualitative data (perceived smoothness of material surfaces (skin)). The introduction of psychometric (Rasch) and generalised linear modelling is considered in multivariate data analysis.

In a metrological study of conformity assessment in a multivariate situation, the clear resolution of the multivariability into components associated with production and measurement errors presents a challenge (at least as significant as in the univariate case [Chapter 2]). In traditional pragmatic approaches to multivariate analysis, measurement uncertainty is assumed to be negligibly small. As we look to include explicitly measurement uncertainty in multivariate cases, useful analogies can be drawn with the reasonably established treatment of multivariate entity dispersion, as will be described below.

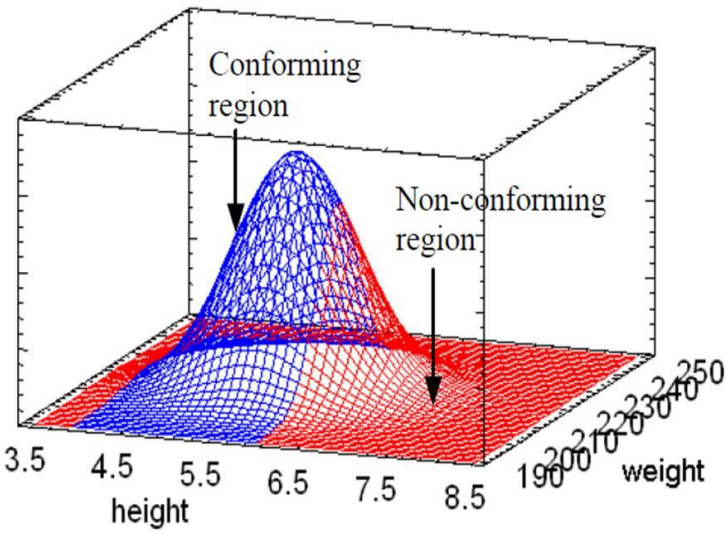


Figure 4.1 Conforming and non-conforming regions for a bivariate Normal distribution [1]

A simple illustration of bivariate conformance regions and probability density functions (PDF) is shown in Figure 4.1 (taken from multivariate process control where entity variation was assumed much larger than measurement uncertainty [1]. A value drawn at random from this PDF that lies within the tolerance region is a conforming value (**blue**); otherwise it is judged as a case of non-conformity (**red**). For many draws from the PDF, the fraction of conforming values is the conformance probability [see [34] for multivariate terminology].

Unfortunately there are very few guidelines which treat the multivariate case about handling the risks of incorrect decisions caused by measurement uncertainty when attempting to answer questions about

an entity, for example, whether the test result is inside or outside specifications. Multivariate guidelines which specifically include cost modelling are even fewer.

4.1 Post Office (PO) parcel problem

To fix ideas (but the concepts are general), consider a simple bivariate problem, which we term the "Post Office parcel problem".

4.1.1 PO Problem: Specifications

To qualify as a "standard parcel" in the UK, a package must satisfy the following conditions:

Length	$L \leq 1.5$ m,
Length + Girth	$L + G \leq 3.0$ m,
Mass	$M \leq 20$ kg.

This is actually a trivariate problem (in the quantities L, G and M), but for illustrative purposes we assume the mass condition to be satisfied and therefore consider the bivariate problem given by the first two inequalities.

The bivariate tolerance region is the wedge-shaped or trapezoidal region shown in figure 4.2 bounded by the inequalities:

$$\begin{aligned} 0 \leq L \leq 1.5, \\ 0 \leq L + G \leq 3.0. \end{aligned} \tag{4.1}$$

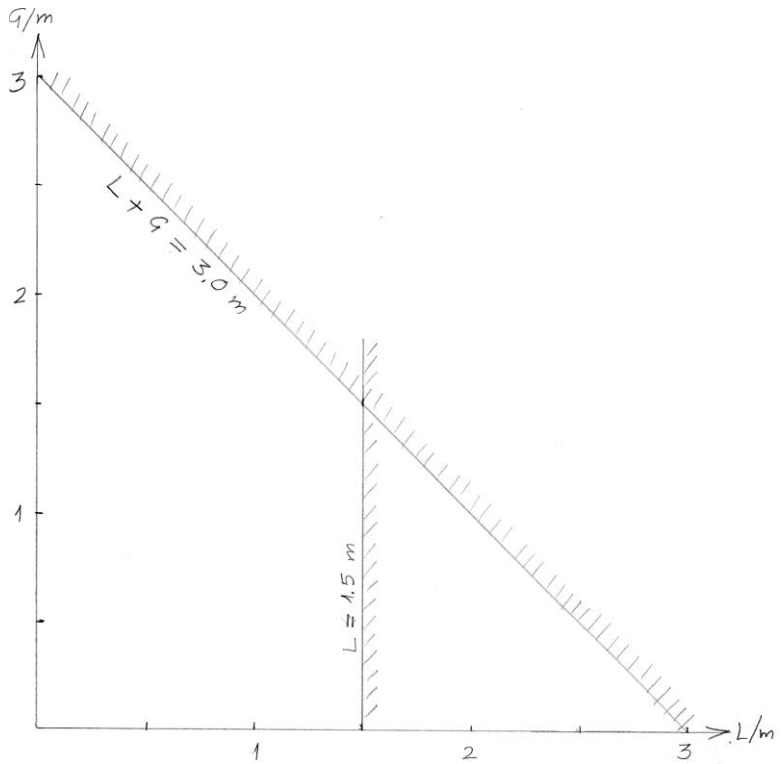


Figure 4.2 Wedge-shaped region denoting tolerance region for Post Office parcel problem

4.1.2 Measurements: PO Problem

Two separate measurements are thus required (L and G) and the acceptance tests first consider L on its own and $L + G$. In addition, there will be correlation between L and G if they are both measured with the same instrument, so that both L and G inherit uncertainties that are associated with the calibration of the measuring instrument. In general in conformance testing, one aims to ensure that uncertainties associated with the measuring instrument itself are small (and perhaps negligible) compared with the uncertainties associated with the estimate of the value of the quality characteristics. In that case, decisions concerning entity results that fall close to the boundary of an acceptance interval will not be complicated by the need to consider the quality of the measurement system (such as the question of whether it is both true and precise enough for the conformance task in question).

Suppose we have a PDF for the quantities L and G, which will generally be a joint PDF for L and G, but may be the product of a PDF for L and a PDF for G when these quantities are mutually independent. The PDF for L and G would have often been obtained by a Monte Carlo method or by some other means (say, in WP2) to produce a point cloud taken as constituting a representative set of points.

4.1.3 Conformance probabilities: PO Problem

We divide the points in the cloud into feasible points in this instance points satisfying inequalities (4.1), and those that are unfeasible. The conformance probability is the probability that the possible pairs of values of L and G lie in the tolerance region. For a PDF characterized by a point cloud, the conformance probability is (approximately) equal to the fraction of those points that are feasible. In general, the conformance probability is the probability that a value of the quantity is feasible.

In order to compute the conformance probability for the particular measurement results at hand, we integrate the measurement PDF, g, over the interval given by the two specifications [4.1] of the tolerance region.

In the framework of the present work, calculations have been made assuming that a bivariate Normal distribution with the zero vector as mean and the identity matrix as covariance can be assigned to the quantities L and G. Additionally we introduce a correlation between L and G equal to 0.5. So the couple (L,G) follows a Normal distribution centred in zero and with $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ as covariance matrix, Σ.

A general expression for the collective probability that all of q object variables, $\mathbf{X}$, lie in conformance within their multivariate tolerance region is [35]:

$$\Pr\left(\mathbf{LSL}_{X,1} \le X_1 \le \mathbf{USL}_{X,1} \ldots \mathbf{LSL}_{X,q} \le X_q \le \mathbf{USL}_{X,q}\right) = \int_{\mathbf{LSL}_{X,1}}^{\mathbf{USL}_{X,1}} \ldots \int_{\mathbf{LSL}_{X,q}}^{\mathbf{USL}_{X,q}} g\left(X_1,\ldots,X_q\right) dX_1 \ldots dX_q \qquad (4.2)$$

In the current bivariate case we have:

$$p_C = \iint_{\{D(L,G)\}} g_{(L,G)}(l,g)dl \cdot dg = \iint_{\{D(L,G)\}} \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}\begin{pmatrix} l-0 \\ g-0 \end{pmatrix}^T \cdot \Sigma^{-1} \cdot \begin{pmatrix} l-0 \\ g-0 \end{pmatrix}} dl \cdot dg$$

where $|\Sigma|$ is the determinant of the covariance matrix and *D(L,G)* is the domain of integration. Specifically

$$p_C = \iint_{\{0 \le l \le 1.5, 0 \le l+g \le 3\}} \frac{1}{2\pi\sqrt{0.75}} e^{-\frac{1}{2*0.75}\left(l^2+g^2-l\,g\right)} dldg\,.$$

Using a diffeomorphism, we simplify the double integral and obtain:

$$p_C = \iint_{\{0 \le x \le 1.5, 0 \le y \le 3\}} \frac{1}{2\pi\sqrt{0.75}} e^{-\frac{1}{2 \cdot 0.75}\left(3x^2+y^2-3xy\right)} dl \cdot dg$$

$$= \iint_{\{0 \le x \le 1.5, 0 \le y \le 3\}} \frac{1}{2\pi\sqrt{0.75}} e^{-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^T \cdot \begin{pmatrix} 1 & 1.5 \\ 1.5 & 3 \end{pmatrix}^{-1} \cdot \begin{pmatrix} x \\ y \end{pmatrix}} dl \cdot dg$$

$$= Pr\left[0 \le X \le 1.5, 0 \le Y \le 3\right]$$

where (*X*,*Y*) follows a bivariate Normal distribution with the zero vector as mean and $\begin{pmatrix} 1 & 1{,}5 \\ 1{,}5 & 3 \end{pmatrix}$ as covariance matrix. The use of the package mvtnorm on R [36] yields 0.3383063 as a theoretical result for the conformance probability, p_C.

Supposing that we have *N* measures of *L* and *G*, which follow this kind of distribution and adding up all the measures contained in the previous interval, we can directly calculate, as a ratio, an estimation of the conformance probability. For example, we find, for a random set of 10000 measures, a conformance probability of 0.334.

Reference [35] provides examples of calculations of how the bivariate conformance probability varies as a function of $\Sigma_{i,j,}$ the correlation coefficient between components *i* and *j*, at various distances of the test result from the specification limit.

4.1.4 Decision rules: PO Problem

To determine if any one point lies within the tolerance region, a conformity check is required. It is straightforward to carry out the check for the Post Office parcel problem (values of *L* and *G* that satisfy the inequalities). In general it does not matter how complicated it is to test whether a point lies in the tolerance region. The only consideration is that a test exists that can be implemented.

Once the conformance probability has been established an appropriate decision rule is applied.

It may be useful to have further knowledge than just the conformance probability, particularly as regards formulating decision rules. A procedure in clause 7.7.4 of [34] can be applied to provide

a 100p % coverage region for the measurand (here L and G) according to stipulated coverage probability p. Given a point cloud that procedure can be outlined thus:

1	Construct a (hyper-)rectangular region in the space of the output quantities;
2	Subdivide this rectangular region into a mesh of small rectangles;
3	Assign each point in the cloud to the small rectangle containing it;
4	Use the fraction of points assigned to each rectangle as the approximate probability that the quantity lies in that rectangle;
5	List the rectangles in terms of decreasing probability;
6	Form the cumulative sum of probabilities for these listed rectangles, stopping when this sum is no smaller than p;
7	Take the corresponding set of rectangles as defining the *smallest coverage region*

For a given coverage probability p, the 100p % coverage region is not unique. The procedure provides (an approximation) to the smallest coverage region. The quality of the approximation depends on the level of discretization (size of the small rectangles) in step 2 and the number of points in the point cloud.

In cases where it is not possible to obtain a representative set of points characterizing the PDF, we suggest two options:

Option I (make a parametric assumption) Provide as large a point cloud as economically possible. Calculate the average and associated covariance matrix for these points. Take as the PDF for the measurand the multivariate Gaussian PDF having these parameters. For details see note 5 of clause 7.7.2 of [34]

Option II (invoke a pessimistically large — but safe — coverage region) Use a multivariate counterpart of the Chebyshev or Gauss inequality [section 4.4.4, [37] to provide a coverage region for the measurand. The use of this option would necessitate different considerations such as determining the intersection of two regions.

A procedure that takes into account that a point may be unfeasible is as above except there is a further additional initial step and step 3 is modified:

0 Initially mark each point as being feasible or unfeasible:

3 Assign each feasible point in the cloud to the small rectangle containing it.

4.2 Software for conformity assessment in multivariate cases

Software for the calculation of the conformance probability in bivariate cases has been developed and validated with different cases studies in WP3, such as the one related to health care products. It is a Matlab executable, which means that it doesn't require any licence purchase from the end-user. It offers three possibilities, available through a main menu (Figure 4.3), depending on the available information:

- "Conditional distributions" : Only the conditional distributions are available,

- "Marginal distributions" : Only the marginal distributions are available,

- "Joint probability distribution" : The joint probability distribution is available.

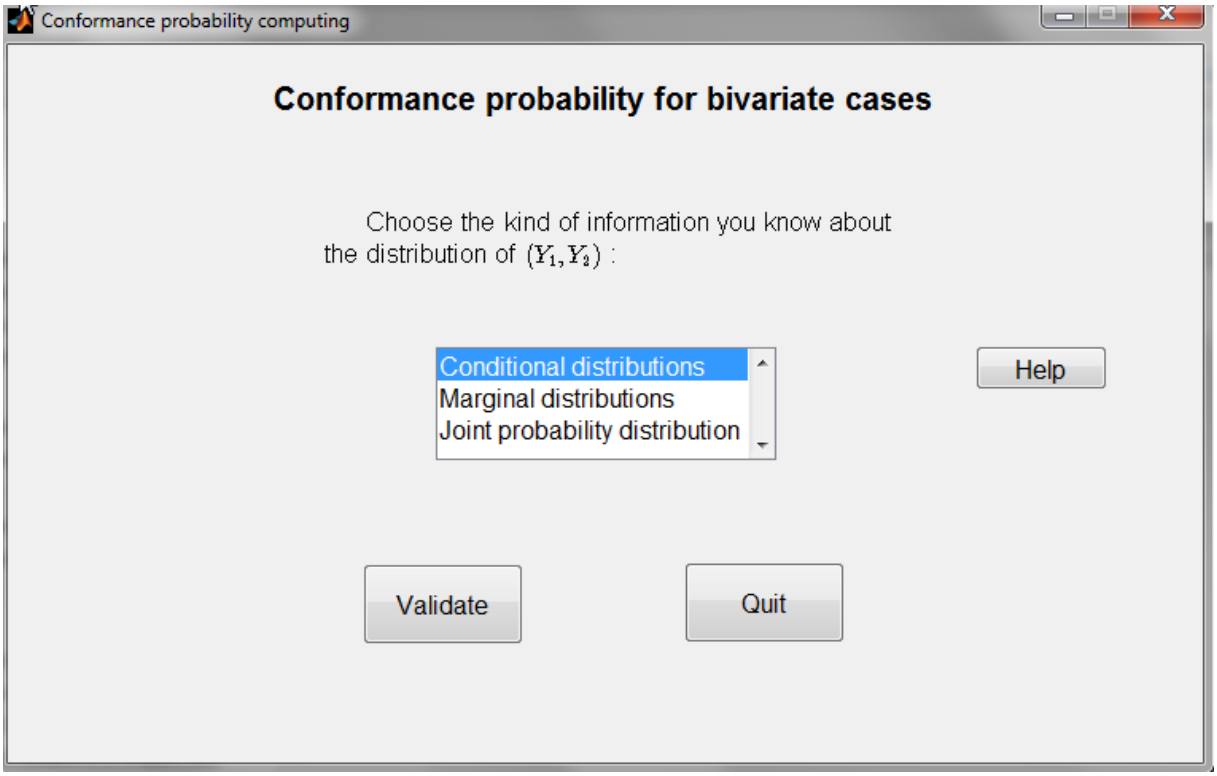


Figure 4.3 Screenshot of main menu of software "Conformance probability"

4.2.1 Installation of software

The software is distributed for Windows environment 32-bit or 64-bit. Prior to the first launch of the software, the user should ensure that the MATLAB Compiler Runtime (MCR), version 8.0 (R2012b), is installed on the computer. If not, it is possible to freely download the Windows 32-bit (or 64-bit) version of the MCR for R2012b from the MathWorks Website (Figure 4.4).
http://www.mathworks.com/products/compiler/mcr/index.html

Release (MATLAB Runtime Version#)	Windows	Linux	Mac
R2015a (8.5)	32-bit / 64-bit	64-bit	Intel 64-bit
R2014b (8.4)	32-bit / 64-bit	64-bit	Intel 64-bit
R2014a (8.3)	32-bit / 64-bit	64-bit	Intel 64-bit
R2013b (8.2)	32-bit / 64-bit	64-bit	Intel 64-bit
R2013a (8.1)	32-bit / 64-bit	64-bit	Intel 64-bit
R2012b (8.0)	32-bit / 64-bit	64-bit	Intel 64-bit
R2012a (7.17)	32-bit / 64-bit	32-bit / 64-bit	Intel 64-bit

Figure 4.4 Different versions of the MCR. R2012b (8.0) for Windows (32-bit or 64-bit) should be downloaded and installed before the first use of the software

4.2.2 Functionalities

The software offers three possibilities to compute the conformance probability, depending on the kind of available information. The user should choose in a main menu the kind of information available according to Figure 4.5.

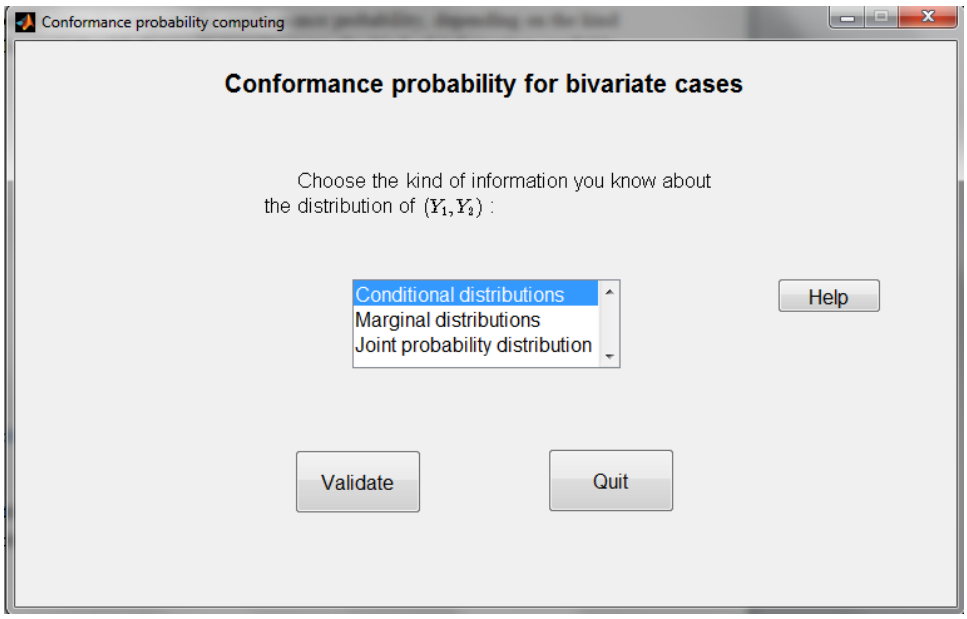


Figure 4.5 Main menu of the software for conformity assessment

4.2.2.1 *Joint probability distribution*

Let $f_{(Y_1;Y_2)}$ be the joint distribution for $(Y_1; Y_2)$ and $[LSL_{Y_1}; USL_{Y_1}] \times [LSL_{Y_2}; USL_{Y_2}]$ be the tolerance region, the conformance probability is then obtained as:

$$P_C = \int_{LSL_{Y_1}}^{USL_{Y_1}} \int_{LSL_{Y_2}}^{USL_{Y_2}} f_{(Y_1;Y_2)}(y_1;y_2)dy_2\, dy_1$$

In the software, such an integral is estimated using the trapezoidal method. Available distributions are Multivariate Gaussian distribution and Multivariate *t*-distribution. An example is given in Figure 4.6 where:

- $Y = (Y_1, Y_1)^T$ is Normally distributed with independent covariance matrix.
- The conformance region is given as : $[1499.8; 1500.2] \times [1499.8; 1500.2]$

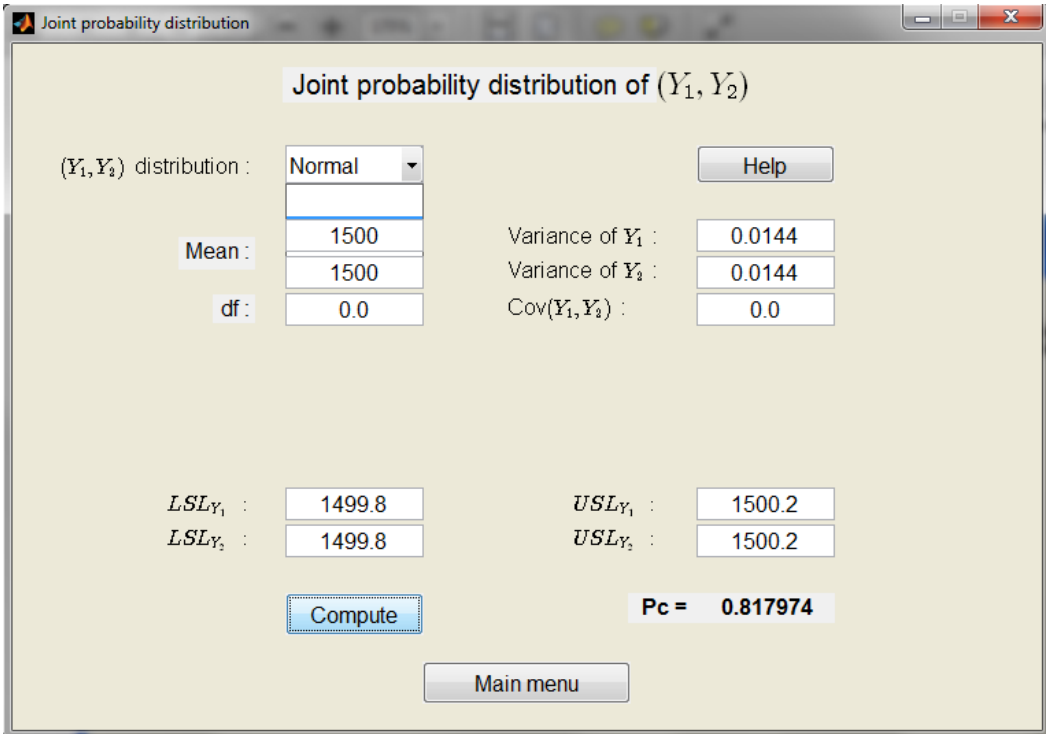


Figure 4.6 Example of computation of conformance probability based on joint probability distribution

4.2.2.2 Marginal probability distributions

In the case of independent quantities, the joint probability distribution is simply the product of the marginal distributions and the two-dimensional integral above can be simplified as the product of two one-dimensional integrals. Then, if f_{Y_1} and f_{Y_2} denote respectively the marginal probability distributions for Y_1 and Y_2, then the conformance probability can be obtained as :

$$P_C = \int_{LSL_{Y_1}}^{USL_{Y_1}} f_{Y_1}(y_1)dy_1 \times \int_{LSL_{Y_2}}^{USL_{Y_2}} f_{Y_2}(y_2)dy_2$$

Both integrals are estimated in the software using the trapezoidal method. Available distributions are Beta, Gamma, log-normal, Gaussian, t-Student and Uniform. An example is given in Figure 3.7 where:

- Y_1 and Y_2 are normally distributed,
- The conformance region is given as : $[1499.8; 1500.2] \times [1499.8; 1500.2]$

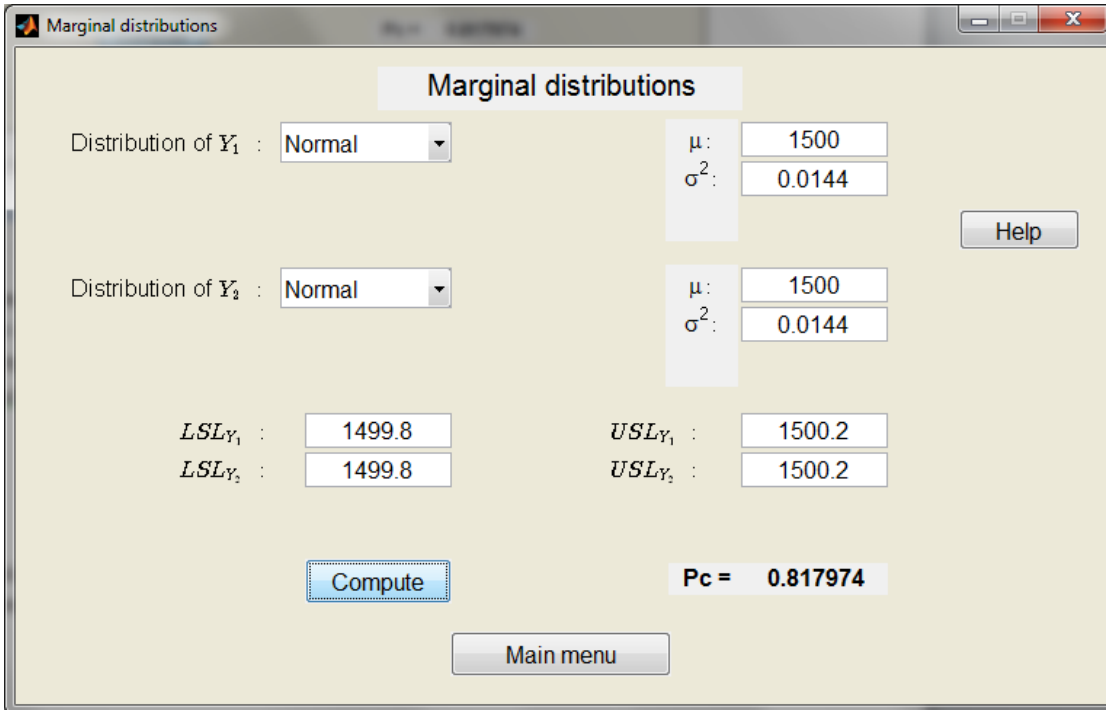


Figure 4.7 Example of computation of conformance probability based on marginal probability distributions

4.2.2.3 Conditional probability distributions

In some cases, only conditional distributions may be known to the user. The conformance probability may then be obtained as :

$$P_C = \int_{LSL_{Y_1}}^{USL_{Y_1}} \int_{LSL_{Y_2}}^{USL_{Y_2}} f_{Y_1}(y_1|y_2) f_{Y_2}(y_2) dy_1\, dy_2$$

The computation of this integral is performed in the software using the Gibb's algorithm. The user is expected to specify the related parameters (initial values of the Markov chains, number of loops and Burn-In), according to Figure 4.9. The available probability distributions in this section are Normal, Beta, Binomial, Gamma and Inverse Gamma.

An example of the computation of the conformance probability is given in Figure 4.8.

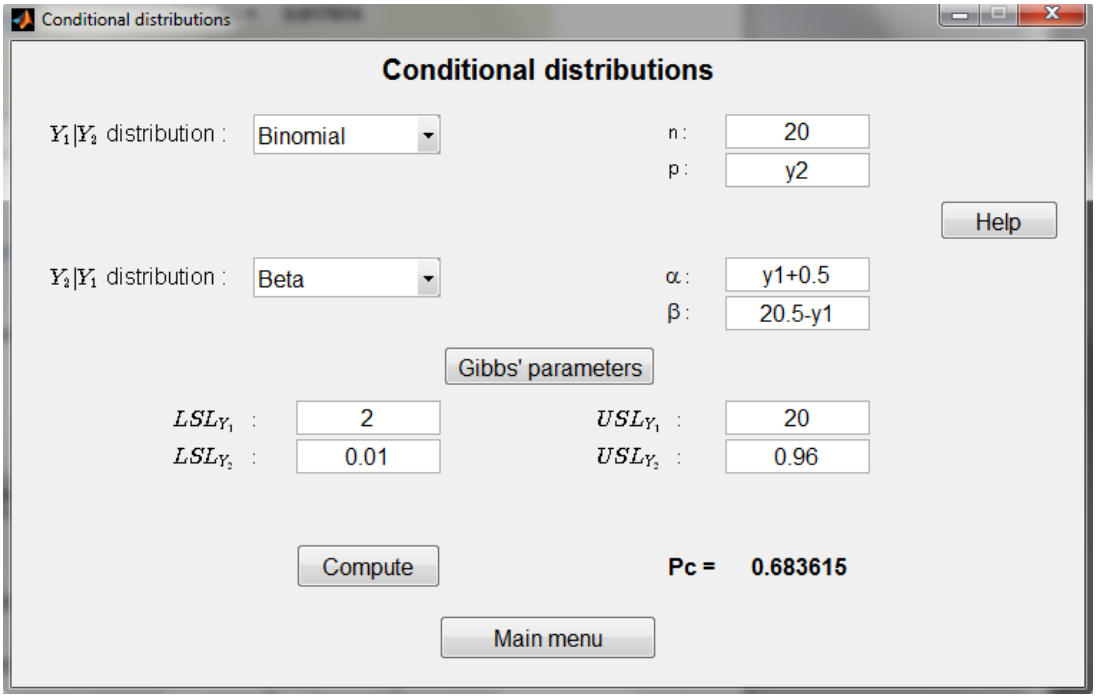


Figure 4.8 Example of computation of conformance probability based on conditional probability distributions

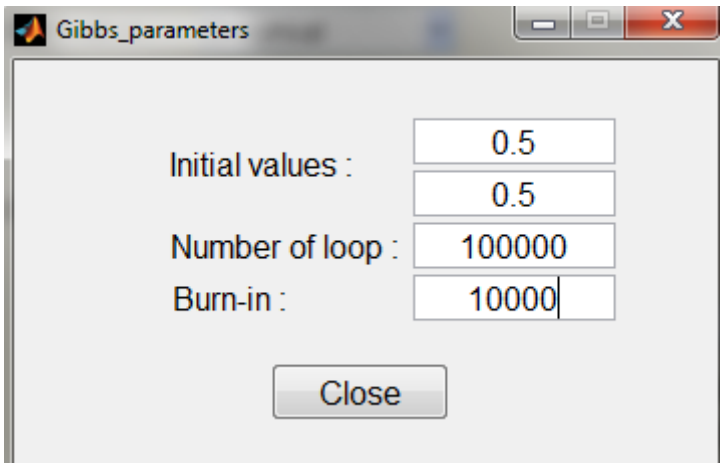


Figure 4.9 Parameters for Gibbs algorithm

4.3 Multivariate case study of healthcare products

Various healthcare studies [[38], [39], [2]] are chosen here to illustrate novel approaches to multivariate conformity assessment and decision-making, as reported in an intermediate report on case study on health care products in the earlier NEW04 deliverable D3.2.4. Such cases are characterised as dealing with properties important for consumers, such as smoothness of material surfaces (skin) as perceived in measurements by human panellists, which in turn depend on the surface topography, friction and hardness, of interest to the manufacturer wishing to fashion the product to the consumer's satisfaction.

Mathematical and statistical approaches to uncertainty evaluation are introduced in studies where typically there is no simple theory of how the perceived ('response') properties **Z** depend on the physical material ('explanatory') properties **X** of the material, that is, how the expression $\mathbf{Z} = f(\mathbf{X})$ is to be formulated.

Material and perceived quantities may be correlated in various ways and, in addition, limited measurement quality may enter on both sides of the expression. Principal component regression [Annex A of D3.2.4], consisting of an initial principal component analysis (PCA), can tackle correlation amongst the explanatory variables (**X**), thereby enabling both conformity assessment against explanatory variable **X** specification limits and multivariate regression including **Z**. A second challenge is that some measurements will be qualitative – such as the perceived 'slipperiness' in a current healthcare study – and require analysis on an ordinal scale, where many of the traditional tools of statistical analysis cannot be applied [Annex D of D3.2.4].

Procedures for setting multivariate tolerance and specification regions and for calculating multivariate conformance properties are reviewed and applied. This includes validation of new software for bivariate conformance probability calculations developed in earlier parts of this NEW04 project [§4.2]. Methods for treatment of measurement uncertainty, reliability and decision risks in multivariate conformity assessment are developed, often drawing analogies with the treatment of corresponding product dispersion in traditional multivariate statistical process control.

4.3.1 Conformance probabilities in multivariate cases: healthcare study

An account of conformance probabilities for multivariate cases is given in Annex A of D3.2.4. An overall multivariate specification region has been proposed by [35] as the intersection of the specification regions $S_X = \{\mathbf{x} | \mathbf{LSL}_X \le \mathbf{x} \le \mathbf{USL}_X\}$ for the individual multiple variables (assumed independent) of the observation matrix **X**:

$$\bigcap_{i=1}^{q} \{X_i | \mathbf{LSL}_{X,i} \le \mathbf{X} \le \mathbf{USL}_{X,i}\} \qquad \text{(A.1)}$$

where lower respectively upper specification limits are set for **X** as follows in terms of the univariate specification limits:

$$\mathbf{LSL}_X{}^T = (LSL_1, \ldots, LSL_q); \qquad \mathbf{USL}_X{}^T = (USL_1, \ldots, USL_q)$$

There remains in the international literature a number of different approaches to defining this multivariate specification region, see D3.2.4.

The collective probability that all of the object variables lie in conformance within their multivariate tolerance region can be calculated with:

$$\Pr\left(\mathbf{LSL}_{X,1} \le X_1 \le \mathbf{USL}_{X,1} \ldots \mathbf{LSL}_{X,q} \le X_q \le \mathbf{USL}_{X,q}\right) = \int_{\mathbf{LSL}_{X,1}}^{\mathbf{USL}_{X,1}} \ldots \int_{\mathbf{LSL}_{X,q}}^{\mathbf{USL}_{X,q}} g\left(X_1, \ldots, X_q\right) dX_1 \ldots dX_q \qquad \text{(A.2)}$$

When the different components, q, are independent:

$$\Pr_{conform} = \prod_{i=1}^{q} \int_{\mathbf{LSL}_{X,i}}^{\mathbf{USL}_{X,i}} g\left(X_i, \overline{X}_i, u_i\right) dX_i \qquad \text{(A.3)}$$

Conformance probabilities in multivariate cases needed in the present healthcare studies can be evaluated using available software, such as in [36] for multivariate Normal and *t* distributions. MathCad contains a number of multivariate functions which we employ. Further, software has been developed in the present NEW04 project and is available specifically for calculation of conformance probabilities for bivariate cases, as reported above [§4.2].

As we proceed with calculations of multivariate conformance probabilities in these healthcare studies, the target of each conformity assessment will be clearly specified: 'Global' conformity denotes the assessment of populations of typical entities, while 'specific' conformity assessment refers to inspection of single items or individuals, as defined by [13].

4.3.2 Specific observations X assessment: healthcare study

As a typical case to be studied in healthcare, reference [2] has recently investigated how, for a set of topical formulations (skin creams), perception of various consumer-related quantities, **Z**, such as slipperiness and smoothness, when each cream is applied on human skin, is related to a number of physiochemical properties, **X**, measured when the same creams are applied on artificial skin substrates.

We introduce the study by making firstly a multivariate analysis of the measured physiochemical properties, **X**, such as $\mu_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (friction) and $A_{\mathrm{BS}}\mathrm{FB}_{0\mathrm{min}}$ (adhesion), plotted in Figure 4.10 together with corresponding upper and lower specification limits (*USL* and *LSL*), respectively, as set by the manufacturer with due consideration to both what is feasible to manufacture but ultimately what is required by the consumer.

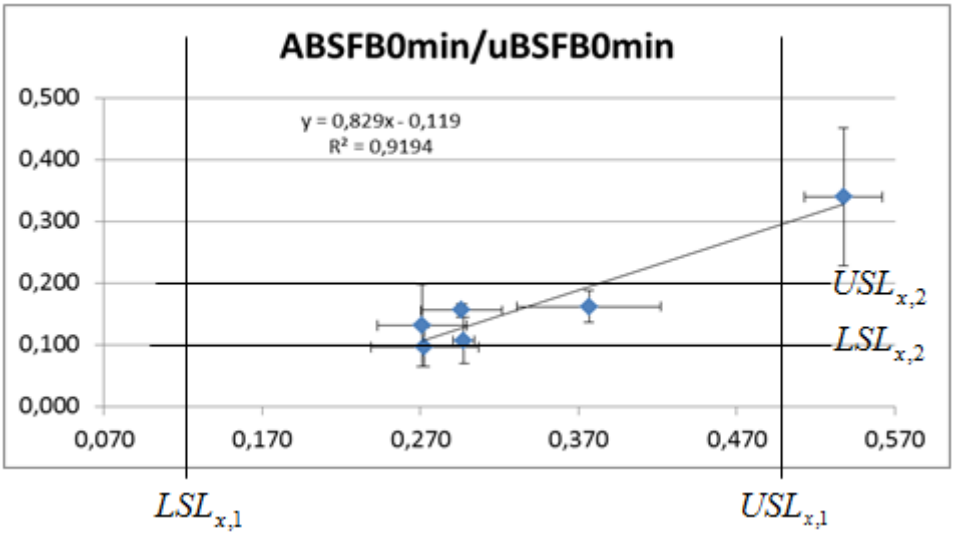


Figure 4.10 Correlation plot between a pair of explanatory variables **X** of six topical formulations [physical material properties: (x-axis) $\mu_{BS}FB_{0min}$ (friction) and (y-axis) $A_{BS}FB_{0min}$ (adhesion)] and corresponding specification limits for the healthcare case studies [2]

An initial study of <u>specific</u> conformity assessment - inspection of single items - was made in this bivariate healthcare case, where conformance probabilities associated with measurement uncertainty, u_m, were calculated using the software *MathCad*, assuming Normal distributions $N\left(x, u_m^2\right)$; no measurement correlation; and no correlation between the two physiochemical properties **X** (i) $\mu_{BS}FB_{0min}$ (friction) and (ii) $A_{BS}FB_{0min}$ (adhesion), for two samples (B and F):

Table 4.1 Summary of calculated bivariate **X** conformance probabilities in the healthcare study: specific conformity

	Specific conformity assessment $(\hat{x}_i, u_m)$	Sample B $\mu_{BS}FB_{0min}$	Sample B $A_{BS}FB_{0min}$	Sample F $\mu_{BS}FB_{0min}$	Sample F $A_{BS}FB_{0min}$
		0,538(24)	0,34(11)	0,377(46)	0,162(25)
	USL	0,5	0,2	0,5	0,2
	LSL	0,1	0,1	0,1	0,1
Univariate conformance probabilities	$P_{conf,1,M,P}(\hat{x}_1, u_m) = \Pr(LSL_1 < x < USL_1 \mid \hat{x}_1, u_m) = \int_{LSL_x}^{USL_x} \frac{1}{\sqrt{2\pi} \cdot u_m} \cdot e^{-\frac{(x-\hat{x})^2}{2 \cdot u_m^2}} \cdot dx; LSL_1 < \hat{x}_1 < USL_1$	5,667%	8,959%	99,625%	92,918%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,M,P}(\hat{x}_1, u_m) \cdot P_{conf,2,M,P}(\hat{x}_1, u_m)$ [eq. A.3]	0,5077066%		92,56923%	
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	0,507707%		92,5692%	

There is excellent agreement from this validation study between bivariate conformance probabilities calculated with the MathCad and MatLab softwares.

4.3.3 Specific PC assessment: healthcare study

It is evident from Figure 4.10 that there is some degree of correlation amongst the pair of material property values plotted.

Covariance and correlation matrices, respectively, for the pair of explanatory variables **X** [physical material properties, $\mu_{BS}FB_{0min}$ (friction) and $A_{BS}FB_{0min}$ (adhesion) in the present healthcare case for a set of topical formulations (skin creams), plotted in Figure 4.10] were deduced from the experimental data with a MathCad program for PCA developed in the project:

$$\mathbf{Cov(X)} = \begin{bmatrix} 0{,}011 & 8{,}846\times10^{-3} \\ 8{,}846\times10^{-3} & 7{,}957\times10^{-3} \end{bmatrix};\ \mathbf{Cor(X)} = \begin{bmatrix} 1 & 0{,}959 \\ 0{,}959 & 1 \end{bmatrix} \tag{4.3}$$

using the MathCad function

$$\mathrm{Covar}(X) := \left| \begin{array}{l} \text{for } i \in 0..\,\mathrm{cols}(X) - 1 \\ \quad X^{\langle i \rangle} \leftarrow X^{\langle i \rangle} - \mathrm{mean}\left(X^{\langle i \rangle}\right) \\ \dfrac{X^{T}\cdot X}{\mathrm{rows}(X) - 1} \end{array} \right.$$

where the non-zero off-diagonal elements of **Cor(X)** indicate the degree of correlation.

Note that the appreciable correlation evident from the off-diagonal elements of the matrices of eq. (4.3) is not a measurement covariance but rather an actual correlation between the different material properties, e.g. one might expect there would be a relation between the adhesion and friction when specimens are formulated, for instance. Such correlation is not unusual, particularly in investigations such as the present where one deliberately chooses to measure a number of explanatory properties on which the consumer response is expected to depend in various ways.

Calculation of conformance probabilities will need to take account of this correlation, as is demonstrated with a principal component analysis [§A.2.3 of D3.2.4], as an example of PCA multivariate conformity assessment [§A.2.4 of D3.2.4] against principal component specification limits. According to the new bivariate conformance probabilities P_c software (NEW04 D3.1.5, §4.2), allowing for the covariances in **X** (from eq. (4.3)), conformance probabilities for one sample (e.g. item F) of 36,9% assuming zero covariance, increase to 42,1% including covariance.

4.3.4 Specific PC Region: Healthcare study

To eliminate this covariance, which confounds the calculation of conformance probabilities, a PCA analysis [§A.2.2 of Annex A of D3.2.2] in the current healthcare case is now made.

For this PCA, a linear model is assumed with model errors $\boldsymbol{\varepsilon}_{\mathbf{x}}$:

$$\mathbf{X} = \mathbf{T}\cdot\mathbf{P} + \boldsymbol{\varepsilon}_{\mathbf{x}} \tag{A.6}$$

and the data matrix (**X**) is first centred. The PCs are the eigenvectors of the covariance matrix of **X**, that is, the vectors $\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_k}$ satisfying the equations

$$\mathbf{Cov(X)} \cdot \mathbf{p_n} = \lambda_n \cdot \mathbf{p_n} \tag{A.7}$$

where λ_n is the nth eigenvalue of $\mathbf{Cov}[\mathbf{X}]$.

In the present case, two (k = 2) principal components $\mathbf{p_1}, \mathbf{p_2}$ of variation amongst the physiochemical properties were identified:

$$\mathbf{P} = \begin{bmatrix} 0{,}759 & -0{,}651 \\ 0{,}651 & 0{,}759 \end{bmatrix} \tag{4.4}$$

using the MathCad function $\text{eigenvals}(\text{Covar}(\text{X}))$ to calculate the eigenvalues, λ_k, followed by $\text{P}^{\langle k \rangle} := \text{eigenvec}(\text{Covar}(\text{X}), \lambda_k)$ as an evaluation of eq. (A.7) in Annex A of D3.2.4.

The resulting covariance matrix:

$$\mathbf{Cov(X')} = \begin{bmatrix} 0{,}018 & 0 \\ 0 & 3{,}761 \times 10^{-4} \end{bmatrix} \tag{4.5}$$

for the transformed data set $\mathbf{X'} = \mathbf{X} \cdot \mathbf{P}$ clearly shows, with the zero off-diagonal elements, the orthogonality of the two PCs.

Adopting the approach of [40] a multivariate PC specification region $S_{PC1} = \{ y_i \in \mathbf{X'} | \mathbf{LSL}_{PC} \le y_i \le \mathbf{USL}_{PC} \}$ [§A.2.4 (A.8) of D3.2.4] in the current healthcare case is bounded by:

$$\mathbf{USL}_{pc} = \mathbf{p}^T \cdot \mathbf{USL} = \begin{bmatrix} 0{,}759 & 0{,}651 \\ -0{,}651 & 0{,}759 \end{bmatrix} \cdot \begin{bmatrix} 0{,}5 \\ 0{,}2 \end{bmatrix} = \begin{bmatrix} 0{,}510 \\ -0{,}174 \end{bmatrix}$$

$$\mathbf{LSL}_{pc} = \mathbf{p}^T \cdot \mathbf{LSL} = \begin{bmatrix} 0{,}759 & 0{,}651 \\ -0{,}651 & 0{,}759 \end{bmatrix} \cdot \begin{bmatrix} 0{,}1 \\ 0{,}1 \end{bmatrix} = \begin{bmatrix} 0{,}141 \\ 0{,}011 \end{bmatrix}$$

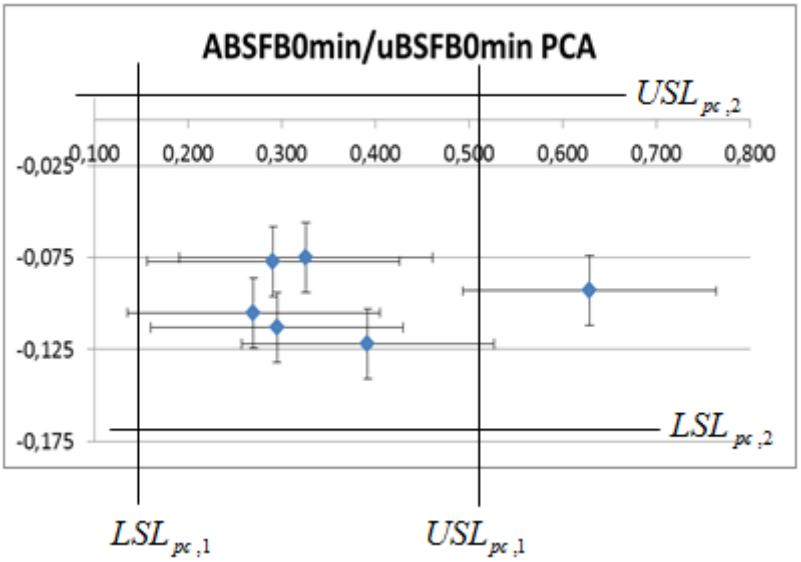


Figure 4.11 Correlation plot between a pair of explanatory variables **X'** of six topical formulations [physical material properties: (x-axis) $\mu_{BS}FB_{0min}$ (friction) and (y-axis) $A_{BS}FB_{0min}$ (adhesion)] and PC corresponding specification limits for the healthcare case studies [2]. Dispersion intervals on each point are based on correlation matrix (measurement uncertainty neglected)

Figure 4.11 show the corresponding bivariate conformity assessment in the space of the principal components (PC1 and PC2) and indicates that most of the variability is concentrated in the first principal component, which of course reflects – as observed in Figure 4.10 – that the observations amongst the six samples are clustered in essentially only two distinct points, thus providing a challenge to reliable correlation estimation.

Table 4.2 Summary of calculated bivariate PC conformance probabilities in the healthcare study: specific conformity

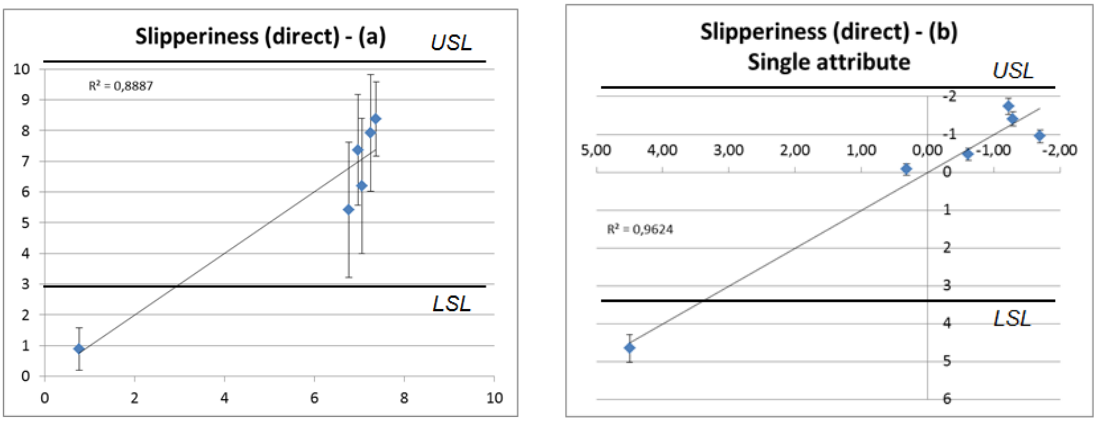
	Specific conformity assessment $(\hat{x}_i, u_m)$	Sample B $\mu_{BS}FB_{0min}$	Sample B $A_{BS}FB_{0min}$	Sample F $\mu_{BS}FB_{0min}$	Sample F $A_{BS}FB_{0min}$
		0,629(24)	-0,093(11)	0,392(46)	-0,122(25)
	USL	0,510	0,011	0,510	0,011
	LSL	0,141	-0,173	0,141	-0,173
Univariate conformance probabilities	$P_{conf,1,M,P}(\hat{x}_1, u_m) = \Pr(LSL_1 < x < USL_1 \mid \hat{x}_1, u_m) = \int_{LSL_x}^{USL_x} \frac{1}{\sqrt{2\pi}\cdot u_m} \cdot e^{-\frac{(x-\hat{x})^2}{2\cdot u_m^2}} \cdot dx;\ LSL_1 < \hat{x}_1 < USL_1$	18,863%	99,998%	77,700%	99,584%
Bivariate conformance probabilities (assuming zero covariance)	$P_{conf,1,M,P}(\hat{x}_1, u_m) \cdot P_{conf,2,M,P}(\hat{x}_1, u_m)$ [eq. A.3]	18,863%		77,377%	
Validation, Bivariate conformance probabilities *Pc* (NEW04 D3.1.5)	Joint probability distributions: Normal	18,7409%		77,9757%	

Even in this PC assessment case, there is evidently excellent agreement from this validation study between bivariate conformance probabilities calculated with the *MathCad* and *MatLab* software.

Note finally that the above PC conformity assessment has neglected measurement uncertainty. In §6 of D3.2.4 an account is given of how to include the (smaller) additional effects of measurement uncertainty on decision risks.

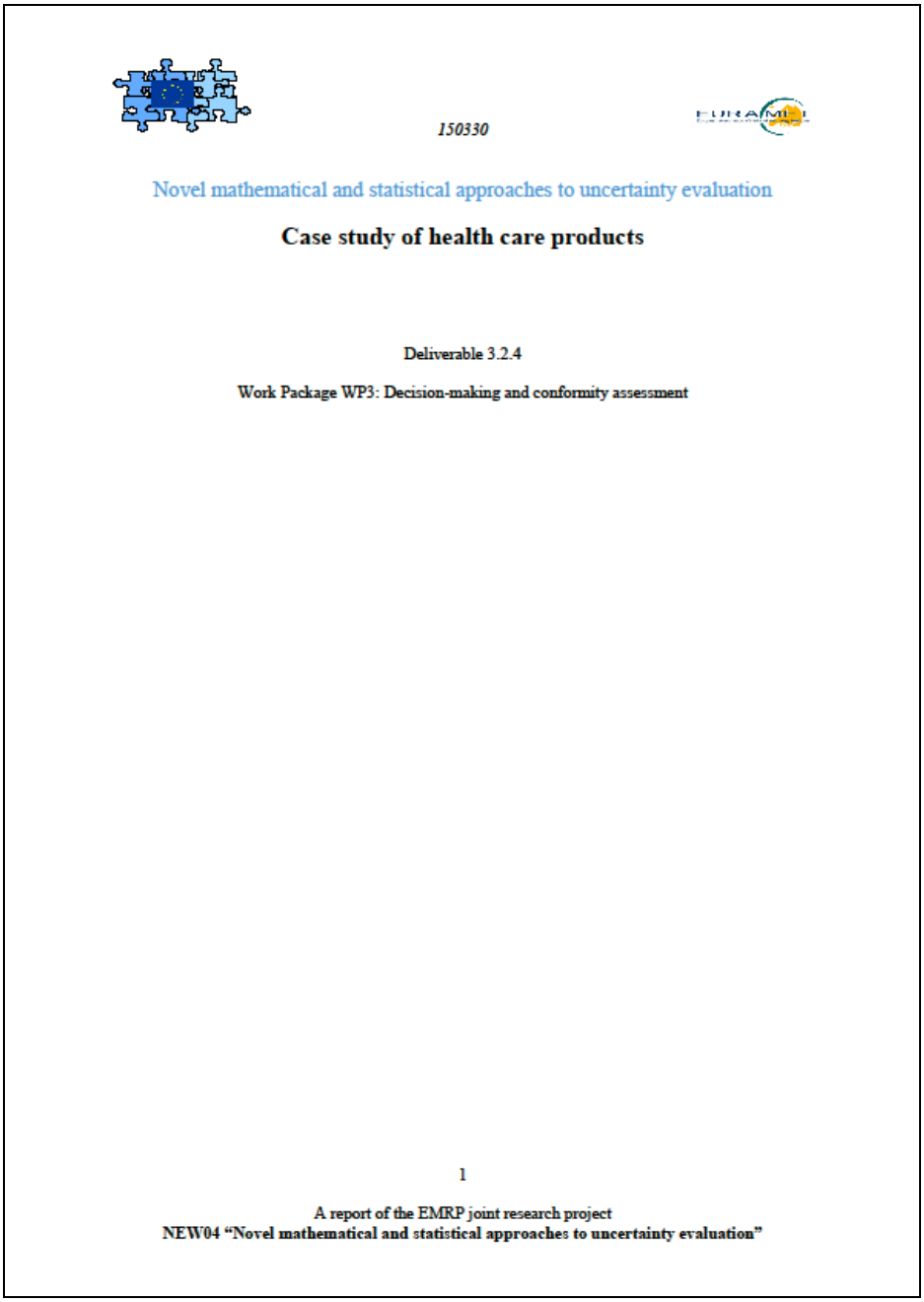
4.3.5 Results of PCR: healthcare study

In Figures 4.12 are shown for comparison typical results of principal component regressions (PCR) for (a) traditional test theory and (b) Rasch analysis [Annex D of D3.2.4], providing explicit relations between sensory ('Slipperiness') and explanatory variables.



Figures 4.12 Relations in the healthcare study between sensory response (*Slipperiness*) (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples, measured (*y*-axis) and predicted (*x*-axis) from explanatory variables (*ForceBoard*, *BIOSKIN*) from PCR analyses [2]

It is evident from Figure 4.12 that the Rasch analysis of the sensory data yields lower measurement uncertainties than conventional test theory, thanks to Rasch analysis' ability to treat ordinal data properly as well as to provide separate estimates of item (sample cream) and person (sensory panellist) attribute values.



150330

Novel mathematical and statistical approaches to uncertainty evaluation

Case study of health care products

Deliverable 3.2.4

Work Package WP3: Decision-making and conformity assessment

1

A report of the EMRP joint research project
NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"

Figure 4.13 Link to D3.2.4 Case study of health care products

5 Costs and Economic Risks in Conformity Assessment

None of the major recent guides about the role of measurement uncertainty in Conformity Assessment deal with costs and impact into risk assessment in Conformity Assessment.

Ultimately when appropriate [or fit-for-purpose [41]] levels of uncertainty and associated risks of incorrect decisions are to be set, reference will need to be made to measures of impact for these various stakeholder groups. The arbitrariness of various rule-of-thumb limits [§2.2.3] on measurement uncertainty may be resolved by economic consideration.

There are examples of successful implementation in areas such as the legal metrological control of measurement instruments in a range of societal important sectors such as utility, commodity and environmental monitoring. Examples of loss function implementation in legal metrology cover instrument categories such as:

- electricity meters [3]
- exhaust gas analysers [42]
- fuel meters [43]

Additionally an analysis of loss models in pre-packaged goods provides a prototype to general product conformity assessment [44],[15].

Economic costs for both analysis and consequence are admittedly sometimes difficult to estimate and not always the best measure of impact. It is better, however, to attempt an economic analysis – however rough and ready – than to set arbitrarily impact costs of incorrect decisions from measurement uncertainty to zero. None of the major recent guides about the role of measurement uncertainty in Conformity Assessment deal in depth with costs and impact into risk-assessment in Conformity Assessment and the current literature survey has provided a summary of recent literature in this area.

This section of the Guide gives reference to literature where loss functions and economic decision theory have started to be introduced when considering this role.

In Fig. 5.1 an overall picture of the classic decision table (figure 2.3) but including costs explicitly is given of sources of both profit and loss from the point of view of the supplier when assessing the conformity of a particular value of an important characteristic of the entity of interest.

Irrespective of the result of product (or entity) conformity, there will always be the costs of production and testing of product (at the centre of Fig. 5.1). Then, for each specimen of product, the actual true value, μ (although unknowable exactly), of the characteristic will either conform or not conform depending on whether the value is inside or outside, respectively, of a specification limit (U_{SL}, upper specification limit in the current example). Correct decisions of conformity relate to both the profit made on selling product which has been correctly assessed to be conforming (top, left of Fig. 5.1), and the losses made on product correctly assessed to be non-conforming (bottom, right of Fig. 5.1).

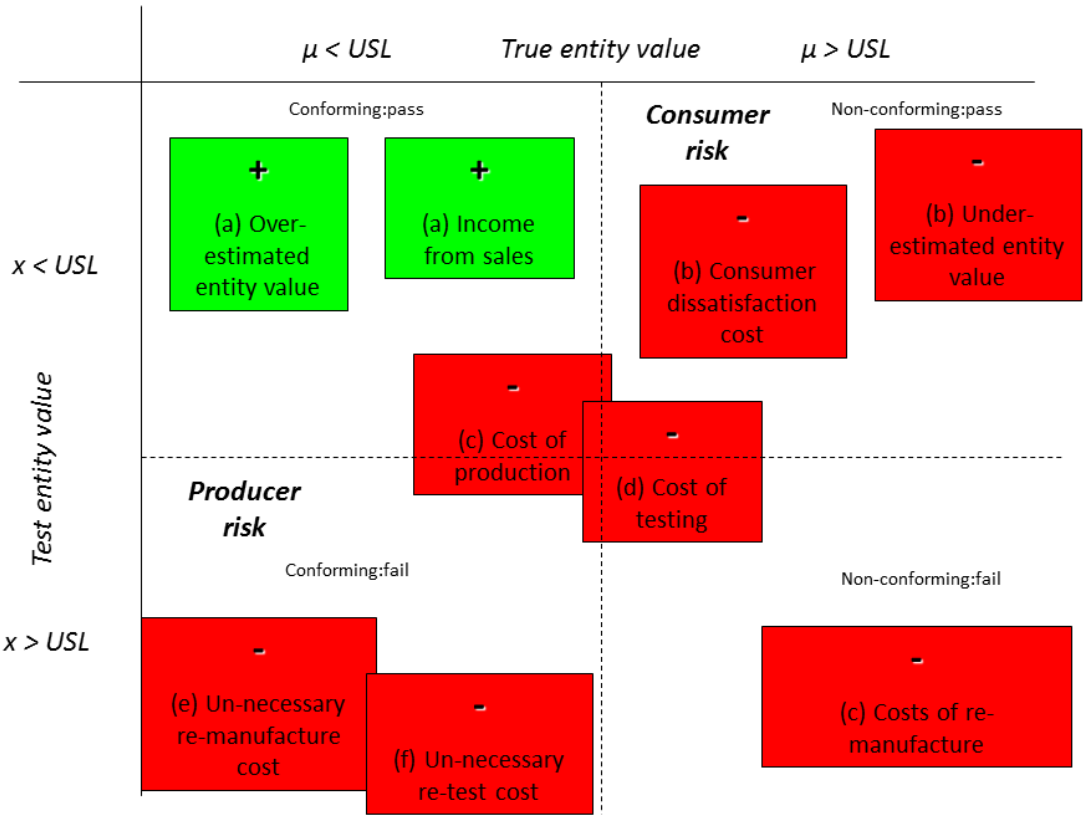


Figure 5.1 Different costs and income of conformity-assessed entity from the point of view of the supplier. Green represents a profit, while red represents a loss [adapted from [3]]

A general formulation of the overall profit [45] would be a sum of the various incomes and losses shown in Fig. 5.1, including: (a) Income from sales of passed, conforming product; (b) Loss associated with customer risk (passed, non-conforming product); (c) Cost of (all) manufactured product (exclusive test); (d) Cost of testing (all) product; (e) Loss associated with re-manufacturing with supplier risk (failed, conforming product); and (f) Loss associated with re-testing with supplier risk (failed, conforming product).

5.1.1 Introducing cost into Conformity Assessment risks

In general, the impact of a wrong decision in Conformity Assessment is expressed as a risk R, defined as the probability p of the wrong decision occurring multiplied by the cost C of the consequences of the incorrect decision [7]:

$$R = p \cdot C$$

5.1.2 Balancing test costs against consequence costs

An in-correct accept on inspection of a non-conforming object will lead to customer costs associated with out-of-tolerance product. Overall costs, consisting of a sum of testing costs, *D*, and the costs, *C*, associated with customer risk, can be calculated with the expression:

$$E(\eta,\sigma) = D(\eta,\sigma) + C(\eta) = \frac{D}{\sigma^2} + \int_{\eta \notin R_{PV}} C(\eta) \cdot g(\eta|\hat{\eta}) \cdot d\eta \qquad (5.1)$$

associated with a finite spread in object values described with a probability density function *g* with dispersion σ and mean $\hat{\eta} \in R_{PV}$, where R_{PV} denotes the region of permissible values - an expression which can be applied to both specific and global conformity assessment – see [15] for details.

Examples of explicitly including a cost function in the integrand when evaluating expression (5.1) for optimized uncertainty by variable can be found in [3] and references therein.

Eq. (5.1) can be evaluated in two complementary ways, as summarised in Figure 5.2:

- a range of quantity values of $L_{SL} - h \cdot u_m \le y_m \le L_{SL} + h \cdot u_m$ for a given test uncertainty, u_m , and 'guard-band' factor *h* – yielding an "operating cost characteristic" analogous to the traditional, probability-based operating characteristic
- a range of test uncertainties, u_m , for a given quantity value $y_m \ge L_{SL}$, the so-called "optimised uncertainty curve".

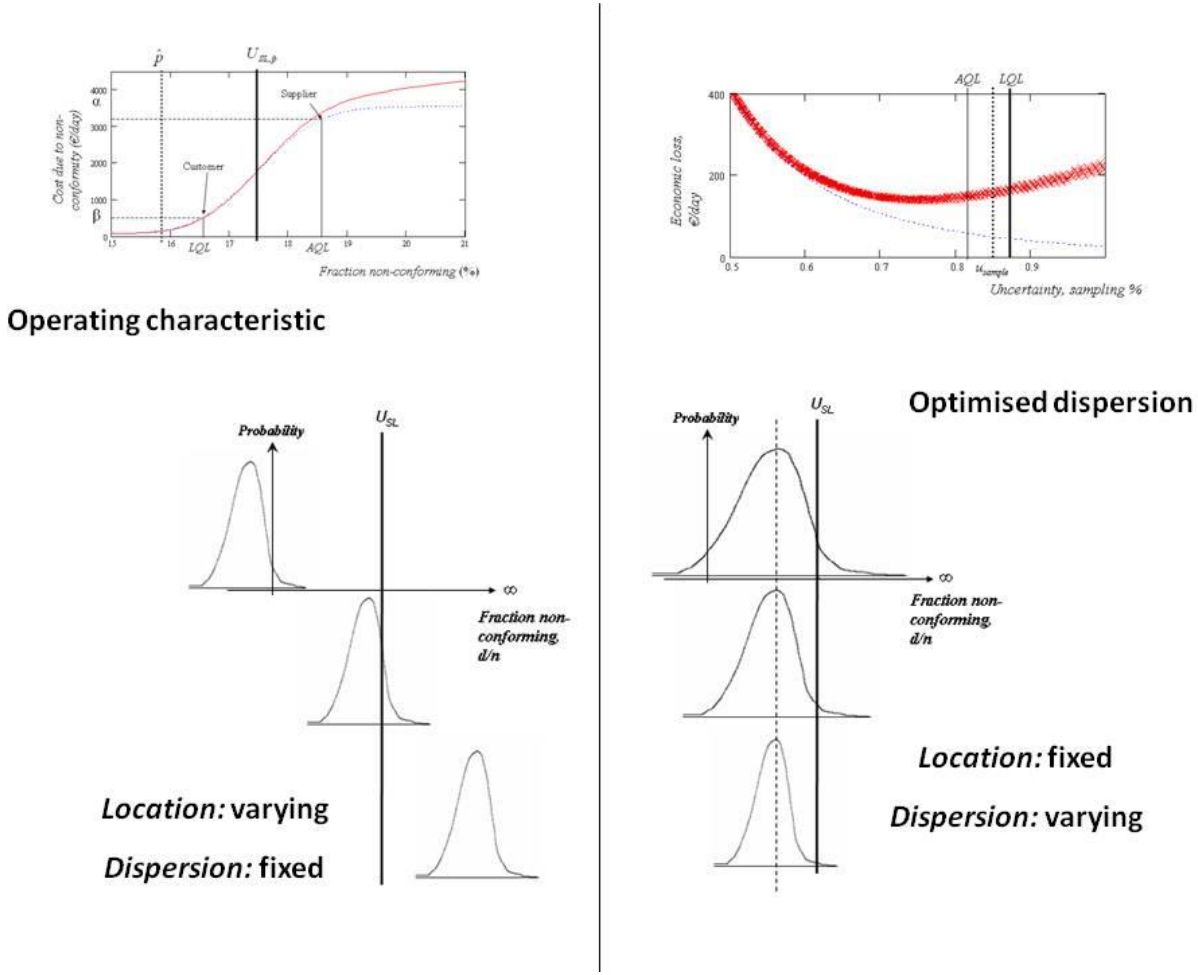


Figure 5.2 Including cost and impact in operating characteristics and optimised dispersion

5.1.3 Optimised uncertainties: by variable and by attribute

Estimation of optimised sampling uncertainties - where measurement costs are balanced against consequence costs of incorrect decisions of compliance according to the expression (5.1) – is good practice ahead of any measurement task.

As an example, the case study of conformance assessment of electrical energy meters investigated by risk analysis reported in the next section [§5.2] can usefully design experiments in advance by optimising measurement uncertainties when:

- testing errors in meter display by variable

The consequence cost of incorrectly accepting non-conforming meters can be calculated from the product of overall cost times the probability of incorrect decisions with respect to metrological (by variable) upper and lower specification limits, USL_z and LSL_z, on meter reading error (MPE = 3.5%). For a specific sample of electricity meters, typically with a mean error of -1.2% and standard deviation 2.5%, and weighted linearly with z:

$$C_{USL} = Cost \cdot \left[\int_{USL_z}^{\infty} z \cdot \frac{1}{\sqrt{2\pi} \cdot s_{instrument}} \cdot e^{-\frac{(z-\hat{z}_{instrument})^2}{2 \cdot s_{instrument}^2}} \cdot dz \right] = 1.4 M€ / year$$

$$C_{LSL} = Cost \cdot \left[\int_{-\infty}^{LSL_z} z \cdot \frac{1}{\sqrt{2\pi} \cdot s_{instrument}} \cdot e^{-\frac{(z-\hat{z}_{instrument})^2}{2 \cdot s_{instrument}^2}} \cdot dz \right] = -9.8 M€ / year$$

In these expressions '*Cost*' is the average loss of revenue: the national annual tax cost, estimated to be 47.5%·0.06€/kWh·40·10^9kWh = 1.1 G€/year, based, respectively, on a 47.5% tax rate; an overall cost of energy per kWh (including price and grid rent); and the national annual electrical energy consumption in kWh [46]. (Similar calculations can be done for other losses shown in the decision matrix of Figure 5.1.)

- sampling fraction non-conforming meters by attribute

A main source of uncertainty in attribute sampling is the finite number of meters in any batch. This is a statistical uncertainty, in the number of non-conforming units, where the non-conforming fraction $\hat{p} = \frac{d}{n}$ can be well represented as a random variable following a binomial distribution, where the probability of selecting $M = d$ non-conforming and $(n - d)$ conforming entities when a sample of size n if taken from an infinitely large population is given by: $\hat{P}_{j,k} = \frac{N!}{d!(N-d)!} \hat{p}^d \cdot (1-\hat{p})^{(n_{sample}-d)}$

[Account of finite population size in estimating sampling uncertainties has to be taken for sample sizes which are more than 1/10th of the total entity population [47]]

The standard deviation of a discrete binomial distribution, $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, appears to be a natural choice for estimating a standard sampling uncertainty, u_{sample}, based on the corresponding confidence interval for $\hat{p}$ [48]. With $d = 2$ and $n = 50$, this uncertainty is about 2.8%.

In the present case, an optimised attribute sampling uncertainty based on consumer risk is given by the following version of eq. 5.1:

$$E_{USL,attr} = \frac{D_{np}}{u_{sample}^2} + \left(C_{USL} + C_{LSL}\right) \cdot \Phi_{binomial}(d, n, SL_{\hat{p}}), \qquad (5.2)$$

for a sample of size n and actual observed number of non-conforming entities, d; where the costs of sampling, D_{np}, were taken to be typically to be 147€ per sample.

The consumer risk, in percentage terms, i.e. the probability, P_{NC}, of passing non-conforming meters – can be expressed by the cumulative binomial distribution beyond a specification limit, SL_p, on the fraction non-conforming instruments

$$\Phi_{binomial}(d, n, SL_{\hat{p}}) = \sum_{p=SL_{\hat{p}}}^{100\%} \frac{n!}{d!(n-d)!} p^d \cdot (1-p)^{(n-d)}$$

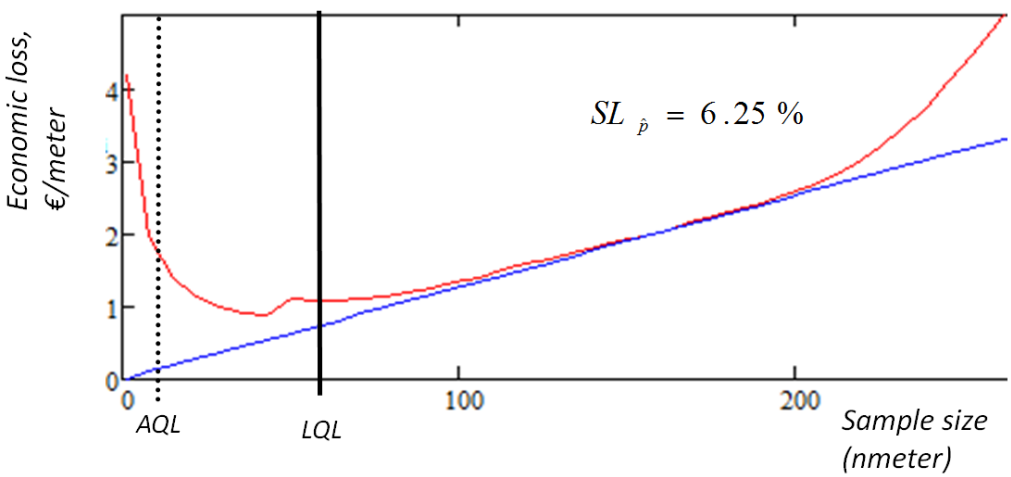
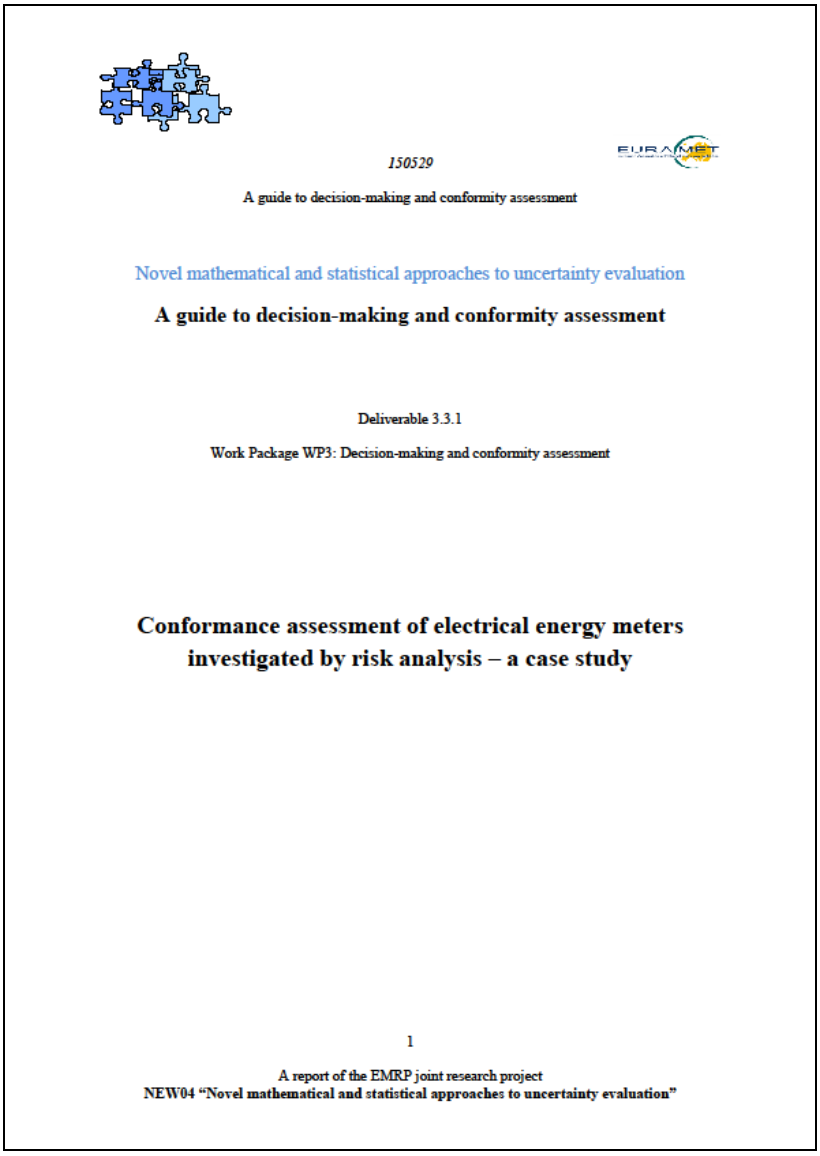


Figure 5.3 Costs as function of attribute sampling size for electricity meters (red line: eq. 5.2; blue line testing costs)

Figure 5.3 illustrates how overall costs vary with the number of instruments sampled (assuming an infinite population) and how an optimised sampling uncertainty and corresponding sample size [48] can be identified in relation to traditional sampling planning limits *AQL* and *LQL* [47] for the

example of electricity meters. Where the actual optimum sampling uncertainty (and sample size) will lie will of course be determined both by actual sampling costs as well as the choice of specification limit (even that economically motivated) on the maximum permissible fraction non-conforming product.

This new optimised sampling uncertainty methodology, extends traditional attribute sampling plans to include economic assessments of the costs of measuring, testing and sampling together with the costs of incorrect decision-making.


150529

A guide to decision-making and conformity assessment

Novel mathematical and statistical approaches to uncertainty evaluation

A guide to decision-making and conformity assessment

Deliverable 3.3.1

Work Package WP3: Decision-making and conformity assessment

Conformance assessment of electrical energy meters investigated by risk analysis – a case study

1

A report of the EMRP joint research project
NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"

Figure 5.4 Link to Conformance assessment of electrical energy meters investigated by risk analysis – a case study

6 Concluding remarks

The subject of this guide and the research of this work package of the NEW04 project - Decisions based on measurement – lies close to various stakeholder groups where the impact of limited measurement uncertainty is most tangible.

It is therefore natural that reference can be made to impact related activities in the NEW04 project, as reported by WP4. During the course of this project, many contacts have been made – for instance the IMEKO TC21, ISO TC 69/SC6 'Measurement methods and results' technical standardisation committees – where guides such as the present document are a primary source of information.

Acknowledgments

Thanks are due to other members of the NEW04 project for their constructive comments about the manuscript and for contributions throughout the project. The European Metrology Research Programme (EMRP, FP7 Art. 185) is jointly funded by the EMRP participating countries within EURAMET (www.euramet.org) and the European Union.

References

1. M Abu Zahid and A Sultana, *Assessment and comparison of multivariate process capability indices in ceramic industry.* Journal of Mechanical Engineering, 2008. **ME39**: p. 18-25.
2. Lisa Skedung, H.W., Izabela Buraczewska Norin (ACO), Camilla Öhgren, Leslie Pendrill, Lovisa Ringstad, *Evaluation of sensory properties of skin creams using Rasch transformation*, in *11th Pangborn Sensory Science Symposium*. 2015: Göteborg.
3. Pendrill, L.R., *Optimised Measurement Uncertainty and Decision-Making in Conformity Assessment.* NCSLI Measure J. Meas. Sci., 2007. **2**(2): p. 76 – 86.
4. C Elster et al, *A Guide to Bayesian Inference for Regression Problems, deliverable of EMRP project NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation".* 2015.
5. L Wright et al, *A Guide to Uncertainty Evaluation for Computationally Expensive Models, deliverable of EMRP project NEW04 "Novel mathematical and statistical approaches to uncertainty evaluation"*. 2015.
6. EURAMET. Available from: www.euramet.org.
7. AFNOR. *Use of uncertainty in measurement: presentation of some examples and common practices*. French standardisation 2004; FD x07-022].
8. ISO, *Conformity assessment – General vocabulary*, in *ISO/IEC 17000:2004*. 2004, International Organization for Standardization: Geneva.
9. ISO, *Statistical methods Guidelines for the evaluation of conformity with specified requirements*, in *Part 1: General principles,*. 2003, International Organization for Standardization: Geneva.
10. EURACHEM/CITAC, *Use of uncertainty information in compliance assessment*, in *Guide*. 2007.
11. JCGM, *Evaluation of measurement data – The role of measurement uncertainty in Conformity Assessment*, in *Joint Committee on Guides in Metrology (JCGM)*. 2012.
12. Kacker, R., N. Zhang, and C Hagwood, *"Real-time control of a measurement process,"* Metrologia, 1996. **33**: p. 433-45.
13. Rossi, G.B. and F Crenna, *A probabilistic approach to measurement-based decisions.* Measurement, 2006. **39**: p. 101-19.
14. JCGM, *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, in *Joint Committee on Guides in Metrology (JCGM)*. 2012.
15. Pendrill, L.R., *Using measurement uncertainty in decision-making & conformity assessment.* Metrologia, 2014. **51**: p. S206.
16. Rukhin, A.L., *Assessing compatibility of two laboratories: formulations as a statistical hypothesis testing problem.* Metrologia, 2013. **50**: p. 49-59.
17. AIAG, *Measurement Systems Analysis Reference Manual*, in *Chrysler, Ford, General Motors Supplier Quality Requirements Task Force*. 2002, Automotive Industry Action Group.
18. Linacre, J.M., *Optimizing Rating Scale Category Effectiveness.* J. Appl. Measurem., 2002. **3**: p. 85 - 106.

19. Bashkansky, E., et al., *Effectiveness of a Product Quality Classifier.* Quality Engineering, 2007. **19**(3): p. 235-244.
20. Iverson, G. and R Luce, *The representational measurement approach to psychophysical and judgmental problems*, in *Measurement, Judgment, and Decision Making*. 1998, Academic Press.
21. Pendrill, L.R. and W P Fisher Jr, *Counting and Quantification: Comparing Psychometric and Metrological Perspectives on Visual Perceptions of Number.* Measurement, 2015. **71**: p. 46–55.
22. Klir, G. and T Folger, *Fuzzy sets, uncertainty and information*. 1988, New Jersey, USA: Prentice Hall.
23. Linacre, J., *Bernoulli Trials, Fisher Information, Shannon Information and Rasch.* Rasch Measurement Transactions, 2006. **20**(3): p. 1062-1063.
24. Pendrill, L.R., *El ser humano como instrument de medida.* e-medida, 2014.
25. Pendrill, L.R., *Man as a Measurement Instrument.* NCSLI Measure J. Meas. Sci., 2014. **9**: p. 24 – 35.
26. Ellison, S. and T. Fearn, *Characterising the performance of qualitative analytical methods: Statistics and terminology.* TRAC-Trend Anal. Chem., 2005. **24**(6): p. 468–476.
27. Kennedy, M.C. and A O'Hagan, *Predicting the output from a complex computer code when fast approximations are available.* Biometrika, 1998. **87**: p. 1-13.
28. Gratiet, L.L., *MuFiCokriging, Multi-Fidelity Cokriging models*, in *R package*. 2012.
29. Rasmussen, C.E. and C K I Williams, *Gaussian Processes for Machine Learning*. 2006: MIT Press.
30. Oakley, J. and A O'Hagan, *Bayesian inference for the uncertainty distribution of computer model outputs.* Biometrika, 2002. **89**: p. 769–784.
31. Xiong, S., P. Qian, and J Wu, *Sequential design and analysis of high-accuracy and low-accuracy computer codes.* Technometrics, 2013. **55**: p. 37–46.
32. Bect, J., et al., *Sequential design of computer experiments for the estimation of a probability of failure.* Statistics and Computing, 2012. **22**: p. 773–793.
33. Roustant, O., D. Ginsbourger, and Y Deville, *DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization.* Journal of Statistical Software, 2012. **51**(1): p. 1-55.
34. JCGM, *Guide to the Expression of Uncertainty in Measurement — Extension to any number of output quantities*, in *Supplement 2*. 2011.
35. Subaric-Leitis, A., *Risikoanalyse multivariater Konformitätsprüfungen.* tm - Technisches Messen, 2010. **77**: p. 662-670.
36. CRAN-R, *Multivariate Normal and t Distributions: Computes multivariate Normal and t probabilities, quantiles, random deviates and densities*, in *Mvtnorm*. 2013.
37. Cox, M. and P Harris, *Uncertainty evaluation*, in *SSfM Best Practice Guide*. 2010: National Physical Laboratory, Teddington, UK.
38. Prall, J.K., *Instrumental evaluation of the effects of cosmetic products on skin surfaces with particular reference to smoothness.* J. Soc. Cosmet. Chem., 1973. **24**: p. 693-707.

39. Fall, A., *Towards surface softness by topographic modeling of haptic properties*, in *Master of Science Thesis*. 2008: Biomaterials, Department of Biological Physics, Chalmers university of technology, Gothenburg.
40. Wang, F. and J Chen, *Capability index using principal component analysis.* Quality Engineering, 1998. **11**: p. 21 - 7.
41. Thompson, M. and T Fearn, *What exactly is fitness for purpose in analytical measurement?* Analyst, 1996. **121**: p. 275 – 278.
42. Källgren, H. and L.R. Pendrill, *Exhaust gas analysers and optimised sampling, uncertainties and costs.* Accreditation and Quality Assurance – Journal for Quality, Reliability and Comparability in Chemical Measurement, 2006. **11**: p. 496- 505.
43. Pendrill, L.R. and H Källgren, *Optimised measurement uncertainty and decision-making in the metering of energy, fuel and exhaust gases,.* Izmerite'lnaya Technika (Measurement Techniques), 2008. **51**: p. 370 - 7.
44. Pendrill, L.R., *Operating 'cost' characteristics in sampling by variable and attribute.* Accred. Qual. Assur., 2008. **13**: p. 619 - 631.
45. Williams, E. and C Hawkins, *The economics of guardband placement*, in *24th IEEE International test Conference*. 1993: Baltimore.
46. SSB. Available from: http://www.ssb.no/en/elektrisitetaar.
47. Montgomery, D.C., *Introduction to Statistical Quality Control*. 1996, Hoboken, NJ: J. Wiley & Sons.
48. Pendrill, L.R., *Optimised measurement uncertainty and decision-making when sampling by variables or by attribute.* Measurement, 2006. **39**: p. 829-840.