NATIONAL TECHNICAL UNIVERSITY OF ATHENS
School of Chemical Engineering
Department of Synthesis and
Development of Industrial Processes

Measurement data analysis in quality management systems. Application to fuel test methods.

PhD Thesis
of
Dimitrios G. Theodorou

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering of the National Technical University of Athens

September 2015

The approval of this PhD dissertation by the School of Chemical Engineering of the National Technical University of Athens does not imply the acceptance of the author's opinions (Law 5343/32 Article 202 § 2)

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Χημικών Μηχανικών
Τομέας Σύνθεσης και
Ανάπτυξης Βιομηχανικών Διαδικασιών

Τεχνικές Ανάλυσης Δεδομένων στα Συστήματα Διαχείρισης της Ποιότητας. Εφαρμογή στις Αναλύσεις Καυσίμων.

Διδακτορική Διατριβή

του

Δημητρίου Γ. Θεοδώρου

Η παρούσα Διδακτορική Διατριβή υποβλήθηκε προς απόκτηση του διπλώματος του Διδάκτορος Μηχανικού του Εθνικού Μετσόβιου Πολυτεχνείου

Σεπτέμβριος 2015

Η έγκριση της διδακτορικής διατριβής από την Ανώτατη Σχολή Χημικών Μηχανικών του Εθνικού Μετσοβίου Πολυτεχνείου δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα (Ν. 5343/32, Άρθρο 202 § 2)

Advisory Committee

F. Zannikos –Professor – School of Chemical Engineering, NTUA (Supervisor)

K. Tzia - Professor – School of Chemical Engineering, NTUA

D. Karonis – Associate Professor - School of Chemical Engineering, NTUA

Examination Committee

F. Zannikos – Professor – School of Chemical Engineering, NTUA

K. Tzia - Professor – School of Chemical Engineering, NTUA

D. Karonis – Associate Professor - School of Chemical Engineering, NTUA

E. Lois - Professor – School of Chemical Engineering, NTUA

M. Koupparis – Professor – Department of Chemistry, National and Kapodistrian University of Athens

E. Bakeas – Assistant Professor - Department of Chemistry, National and Kapodistrian University of Athens

J. Thybaut – Full Professor - Faculty of Engineering and Architecture - Ghent University – Belgium

Note to the reader

The present PhD thesis has been originally written in English and is accompanied by an extended summary in Greek (Annex A). The Greek summary may be considered as a stand-alone text and presents all research achievements. The references are only presented in English at the end of the core text of the thesis.

Σημείωμα προς τον αναγνώστη

Η Γενική Συνέλευση της Σχολής Χημικών Μηχανικών (συνεδρίαση 23/04/2015), όρισε, σύμφωνα με τις διατάξεις του άρθρ. 9 παρ. 4.γ του Ν.3685/08, την αγγλική ως γλώσσα συγγραφής της παρούσας Διδακτορικής Διατριβής. Η διατριβή συνοδεύεται από εκτεταμένη περίληψη στα ελληνικά (Annex A), η οποία αποτελεί αυτόνομο κείμενο και περιέχει όλα τα ερευνητικά αποτελέσματα της διατριβής. Η βιβλιογραφία παρατίθεται στο τέλος του αγγλικού κειμένου μόνο.

Abstract

Test results should be reliable and accepted by all interested parties so that they reduce risks, shorten time to market and demonstrate the quality and safety of fuel products. The quality of a result and its fitness for purpose is directly related to the estimation of **measurement uncertainty**. The estimation of measurement uncertainty following recognized and valid methodologies is also a key requirement for laboratories or other organizations accredited or seeking accreditation according to international standards such as ISO/IEC 17025, ISO 15189, ISO/IEC 17043 or ISO Guide 34. The **measurement uncertainty** provides a quantitative indication of the quality of a measurement result and has implications for the interpretation of analytical results in the context of regulatory compliance or **conformity assessment**. There are two broad types of approaches for the estimation of uncertainty, **modeling** and **empirical**. Several of these approaches, which employ the use of statistical and numerical methods are considered. The work presented in this thesis concerns the estimation and use of measurement uncertainty in all parts of the measurement cycle: pre-analytical (sampling), analytical (including measurement processes using a calibration curve) and post-analytical (conformity assessment of products using test results).

Sampling is an important part of any measurement process and is therefore recognized as an important contributor to the measurement uncertainty. A reliable estimation of the uncertainty arising from sampling of fuels leads to a better control of risks associated with decisions concerning whether product specifications are met or not. Three empirical statistical methodologies (classical ANOVA, robust ANOVA and range statistics) using data from a balanced experimental design, which includes duplicate samples analyzed in duplicate from 104 sampling targets (petroleum retail stations), are described and compared. These methodologies are used for the estimation of the uncertainty arising from the manual sampling of fuel (automotive diesel) and the subsequent sulfur mass content determination. The results of the three methodologies statistically differ, with the expanded uncertainty of sampling being in the range of 0.34 – 0.40 mg kg^{-1}, while the relative expanded uncertainty lying in the range of 4.8 - 5.1%, depending on the methodology used. The estimation of robust ANOVA (sampling expanded uncertainty of 0.34 mg kg^{-1} or 4.8% in relative terms) is considered more reliable, because of the presence of outliers within the 104 datasets used for the calculations.

The Guide to the Expression of Uncertainty in Measurement (GUM) approach and the adaptive Monte Carlo method (MCM) provide two alternative approaches for the propagation stage of the uncertainty estimation of a **measurement procedure**. These two approaches are implemented and compared concerning the 95% coverage interval estimation of the measurement of Gross Heat of Combustion (GHC) of an automotive diesel fuel by bomb calorimetry. The GUM approach, which assumes either a Gaussian or a *t*-distribution for the output quantity (GHC) gives half width intervals of 0.28 MJ kg^{-1} or 66 cal g^{-1}(Gaussian distribution) and 0.29 MJ kg^{-1} or 70 cal g^{-1} (*t*- distribution). On the other hand, MCM, which provides a reliable probability density function of GHC through numerical approximation, gives a half width interval of 0.32 MJ kg^{-1} or 75 cal g^{-1}. Thus, the GUM approach underestimates the calculated uncertainties and coverage intervals by up to 7 –

12%. The main reasons of these differences are the approximations and the assumptions introduced by the GUM approach i.e. assumption for the GHC probability distribution and overestimation of effective degrees of freedom by the Welch-Satterwaite formula. Only if the GUM approach is combined with a Bayesian treatment of Type A uncertainties, the results are comparable with the MCM results. Moreover, the estimation and the use of sensitivity coefficients and uncertainty budget within GUM and MCM approaches are examined. Finally, it is shown that an initial estimate of measurement uncertainty may be obtained using the proficiency testing data.

The construction of a **calibration curve** using least square linear regression is common in many analytical measurements and comprises an important uncertainty component of the whole analytical procedure uncertainty. Various methodologies are applied concerning the estimation of the standard uncertainty of a calibration curve used for the determination of sulfur mass concentration in fuels. The methodologies applied include the GUM approach, the Kragten numerical method, the MCM as well as the approximate equation calculating the standard error of prediction. The standard uncertainty results obtained by all methodologies agree well (0.172 - 0.175 ng μL^{-1}). Aspects of inappropriate use of the approximate equation of the standard error of prediction, which leads to overestimation or underestimation of the calculated uncertainty, are discussed. Moreover, the importance of the correlation between the calibration curve parameters (slope and intercept) within GUM, MCM and Kragten approaches is examined.

In order to use a test result to decide whether it indicates compliance or non-compliance, it is necessary to take into account the dispersion of the values that can be attributed to the measurand. When dealing with **conformity assessment** of automotive fuel samples against European Union specification limits, this dispersion may be represented by uncertainty estimates based on either standard method precision data (ISO 4259 approach) or within laboratory precision data (intermediate precision approach). Possible decision rules based on these approaches are directly related to the required or acceptable level of probability of making a wrong decision. Acceptance limits for 95% and 99% confidence levels calculated for all the properties of automotive fuels, are presented. Moreover, the effect of different approaches for defining guard bands, different levels of confidence or different number of replicate measurements is investigated using the results of the analyses of 769 diesel fuel samples for the determination of sulfur mass concentration.

Σύντομη Περίληψη (Abstract in Greek)

Τα αποτελέσματα των αναλύσεων των καυσίμων θα πρέπει να είναι αξιόπιστα και αποδεκτά από όλα τα ενδιαφερόμενα μέρη, έτσι ώστε να μειωθούν οι κίνδυνοι προϊόντων εκτός προδιαγραφών, να συντομεύσει ο χρόνος διάθεσης των καυσίμων στην αγορά και να αποδεικνύεται η ποιότητα και η ασφάλεια αυτών. Η αξιολόγηση της «καταλληλόλητας για χρήση» μιας αναλυτικής μεθόδου είναι άρρηκτα συνδεδεμένη με την **εκτίμηση της αβεβαιότητας** της μέτρησης η οποία ουσιαστικά χαρακτηρίζει την ποιότητα ενός αποτελέσματος συνυπολογίζοντας τόσο συστηματικά όσο και τυχαία σφάλματα. Επιπλέον η εκτίμηση της αβεβαιότητας των μετρήσεων με τη χρήση μιας επιστημονικά τεκμηριωμένης και έγκυρης μεθοδολογίας είναι μια βασική απαίτηση συγκεκριμένων διεθνών προτύπων ποιότητας βάσει των οποίων διαπιστεύονται εργαστήρια και φορείς (ISO/IEC 17025, ISO 15189, ISO/IEC 17043 ή ISO Guide 34). Στα κεφάλαια της παρούσας Διδακτορικής Διατριβής παρουσιάζεται η ανάπτυξη και εφαρμογή στατιστικών και αριθμητικών μεθόδων για την εκτίμηση και χρήση της αβεβαιότητας μετρήσεων σε συγκεκριμένα στάδια του κύκλου της μέτρησης των καυσίμων: προ-αναλυτικά (δειγματοληψία), αναλυτικά (κυρίως μετρητική διαδικασία) και μετα-αναλυτικά (αξιολόγηση της συμμόρφωσης βάσει εργαστηριακών αποτελεσμάτων).

Η **δειγματοληψία** αποτελεί ένα βασικό στάδιο των μετρητικών διαδικασιών και συνεισφέρει σημαντικά στην αβεβαιότητα των εργαστηριακών μετρήσεων. Μια αξιόπιστη εκτίμηση της αβεβαιότητας λόγω δειγματοληψίας μπορεί να οδηγήσει σε καλύτερο έλεγχο των κινδύνων που συνδέονται με αποφάσεις περί συμμόρφωσης ή μη, ενός καυσίμου με προδιαγραφές που επιβάλει η νομοθεσία. Στα πλαίσια της παρούσας Διδακτορικής Διατριβής περιγράφονται και συγκρίνονται ως προς τα αποτελέσματά τους, τρεις εμπειρικές στατιστικές μεθοδολογίες («κλασική» ANOVA, ανθεκτική ANOVA και στατιστική εύρους τιμών) χρησιμοποιώντας δεδομένα ενός ισορροπημένου πειραματικού σχεδίου (balanced experimental design). Οι τρεις μεθοδολογίες χρησιμοποιούνται για την εκτίμηση της αβεβαιότητας λόγω δειγματοληψίας καυσίμου (ντήζελ κίνησης) και λόγω της αναλυτικής διαδικασίας προσδιορισμού περιεκτικότητας σε θείο. Η διευρυμένη αβεβαιότητα της δειγματοληψίας κυμαίνεται από 0,34 εώς 0,40 mg kg^{-1}, ενώ η σχετική διευρυμένη αβεβαιότητα από 4,8 εώς 5,1%, ανάλογα με τη στατιστική μεθοδολογία που χρησιμοποιήθηκε. Τα αποτελέσματα της ανθεκτικής ANOVA (διευρυμένη αβεβαιότητα της δειγματοληψίας 0,34 mg kg^{-1}), η οποία δεν επηρεάζεται από την παρουσία μικρού αριθμού ακραίων (έκτοπων) τιμών στα δεδομένα, μπορούν να θεωρηθούν ως περισσότερο αξιόπιστα.

Κοινά αποδεκτή μεθοδολογία για την εκτίμηση της αβεβαιότητας μιας **μετρητικής διαδικασίας** είναι αυτή που περιγράφεται στην Οδηγία ISO GUM "Guide to the Expression of Uncertainty in Measurement". Ωστόσο, η προσέγγιση GUM παρουσιάζει περιορισμούς στην εφαρμογή της, που μπορούν να ξεπεραστούν με την εφαρμογή της αριθμητικής μεθόδου Monte Carlo (MCM). Η εκτίμηση της αβεβαιότητας με τη χρήση της μεθοδολογίας Monte Carlo βασίζεται στην τεχνική διάδοσης κατανομών πιθανότητας και όχι αβεβαιοτήτων όπως ισχύει στην κλασσική προσέγγιση κατά GUM. Οι δύο ανωτέρω προσεγγίσεις χρησιμοποιούνται για την παράλληλη εκτίμηση της αβεβαιότητας μέτρησης της θερμογόνου δύναμης πετρελαίου κίνησης με τη χρήση θερμιδομέτρου όλμου. Η διευρυμένη αβεβαιότητα (για πιθανότητα κάλυψης 95%) εκτιμήθηκε μέσω της μεθοδολογίας GUM (υποθέτοντας κανονική κατανομή) στα 0,28 MJ kg^{-1} ή 66,3 cal g^{-1}. Η τιμή αυτή είναι 12% μικρότερη από την τιμή που έδωσε η εφαρμογή της μεθοδολογίας MCM (0,32 MJ kg^{-1} ή 75,3 cal g^{-1}). Χρησιμοποιώντας τη μεθοδολογία GUM σε συνδυασμό με την εξίσωση Welch-Satterthwaite για τον υπολογισμό βαθμών ελευθερίας και στη συνέχεια, αποδίδοντας μια κατανομή *t* - student στο μετρούμενο μέγεθος οδηγούμαστε σε

αυξημένη (σε σχέση με την υπόθεση της κανονικής κατανομής) διευρυμένη αβεβαιότητα (0,29 MJ kg^{-1} ή 70,4 cal g^{-1}), αλλά και πάλι κατά 7% χαμηλότερη από αυτή που δίνει η προσέγγιση MCM. Αυτές οι διαφορές μπορούν να αποδοθούν στη μικρή μη γραμμικότητα του μοντέλου μέτρησης και στη προσεγγιστική φύση της εξίσωσης Welch-Satterthwaite, η χρήση της οποίας στην προκειμένη περίπτωση υπερεκτιμά τους βαθμούς ελευθερίας. Μόνο στην περίπτωση χρήσης της Μπεϋζιανής προσέγγισης για την εκτίμηση τών αβεβαιοτήτων Τύπου Α και στη συνέχεια χρήσης τους στο ισοζύγιο αβεβαιοτήτων της μεθοδολογίας GUM τα αποτελέσματα συμφωνούν με τα αποτελέσματα της προσέγγισης MCM. Επιπλέον, η αβεβαιότητα της μέτρησης της θερμογόνου δύναμης εκτιμήθηκε και με τη χρήση δεδομένων από διεργαστηριακές συγκριτικές δοκιμές ικανότητας (εμπειρική προσέγγιση).

Η διαδικασία της **βαθμονόμησης** (μέσω της κατασκευής καμπύλης βαθμονόμησης) είναι ένα απαραίτητο στάδιο σε πολλές χημικές αναλύσεις που αφορούν στον προσδιορισμό της συγκέντρωσης μιας ουσίας με βάση την απόκριση (σήμα) ενός οργάνου. Στα πλαίσια της παρούσας Διδακτορικής Διατριβής, εφαρμόζονται και συγκρίνονται 4 μεθοδολογίες εκτίμησης της αβεβαιότητας (GUM, MCM, Kragten, εξίσωση τυπικού σφάλματος) λόγω καμπύλης βαθμονόμησης που χρησιμοποιείται για τον προσδιορισμό της περιεκτικότητας καυσίμων σε θείο σύμφωνα με τη μέθοδο υπεριώδους φθορισμού (ISO 20846, ASTM D5453). Τα αποτελέσματα όλων των μεθοδολογιών συμφωνούν μεταξύ τους (τυπική αβεβαιότητα 0,172 – 0,175 ng μL^{-1}). Πραγματοποιήθηκαν επίσης και υπολογισμοί αγνοώντας τη συνδιακύμανση μεταξύ κλίσης και τεταγμένης (με τις μεθοδολογίες GUM, MCM και Kragten) όπου φαίνεται ότι γίνεται κατά 62% υπερεκτίμηση της αβεβαιότητας. Αν στο αποτέλεσμα της προσεγγιστικής εξίσωσης του τυπικού σφάλματος δεν προστεθεί και η τυπική αβεβαιότητα της απόκρισης, οδηγούμαστε τότε σε υποεκτίμηση της αβεβαιότητας κατά 22%. Επιπλέον, δεδομένου ότι η εκτίμηση των 2 παραμέτρων μιας καμπύλης βαθμονόμησης βασίζεται σε ένα μοντέλο μέτρησης με πολλαπλά εξερχόμενα μετρούμενα μεγέθη (κλίση, τεταγμένη), εφαρμόζονται και οι βασικές αρχές της συμπληρωματικής οδηγίας του GUM (Supplement 2 - Extension to any number of output quantities).

Τέλος, η **αξιολόγηση της συμμόρφωσης** ενός προϊόντος όταν βασίζεται σε εργαστηριακές μετρήσεις, θα πρέπει να λαμβάνει υπόψη ότι καμία μέτρηση δεν είναι 100% ακριβής, καθώς η πραγματική τιμή κάθε μετρούμενου μεγέθους και τυχόν σφάλματα που σχετίζονται με τη μέτρηση δεν μπορούν να είναι γνωστά. Ιδιαίτερα, όταν το αποτέλεσμα της μέτρησης είναι κοντά στο όριο κάποιας προδιαγραφής, μόνο με τη χρήση της θεωρίας πιθανοτήτων και καταλλήλων κανόνων απόφασης μπορεί κανείς να έχει τον έλεγχο επί της πιθανότητας να λάβει μια λανθασμένη απόφαση (ρίσκο). Στα πλαίσια της παρούσας Διδακτορικής Διατριβής αναλύονται δύο διαθέσιμες προσεγγίσεις (με χρήση δεδομένων πιστότητας μεθόδων και χρήση εκτίμησης αβεβαιότητας) που μπορούν να χρησιμοποιηθούν για να υποστηρίξουν αξιόπιστες αποφάσεις σχετικά με την αξιολόγηση της συμμόρφωσης των καυσίμων. Τα αποτελέσματα των αναλύσεων 769 δειγμάτων ντήζελ κίνησης από αντίστοιχο αριθμό πρατηρίων για τον προσδιορισμό της περιεκτικότητας σε θείο χρησιμοποιήθηκαν για να γίνει η σύγκριση των διαφορετικών προσεγγίσεων ορισμού κανόνων απόφασης για την αξιολόγηση της συμμόρφωσης τους σε σχέση με το νομοθετικό όριο των 10 mg kg^{-1}. Χρησιμοποιώντας κανόνες λήψης αποφάσεων που βασίζονται στη χρήση ζωνών προστασίας πάνω ή κάτω από το όριο της προδιαγραφής οδηγούμαστε σε αριθμό μη συμμορφούμενων δειγμάτων είτε μικρότερο («ελαστική» αποδοχή) είτε μεγαλύτερο («αυστηρή» αποδοχή) σε σχέση με τη μη χρήση ζωνών. Ο ακριβής αριθμός των μη συμμορφούμενων αποτελεσμάτων εξαρτάται από το επιλεγμένο επίπεδο εμπιστοσύνης και τον αριθμό επαναληπτικών εργαστηριακών μετρήσεων. Επιπλέον υπολογίστηκαν όρια απόφασης για όλες τις παραμέτρους - προδιαγραφές που περιγράφονται στα πρότυπα EN 228 (αμόλυβδη βενζίνη) και EN 590 (ντήζελ κίνησης).

Acknowledgments

I would like to thank my supervisor (Professor Fanourios Zannikos) initially for giving me the opportunity to do this PhD and then for effectively supporting and advising me until the end of my research work. He has given me the freedom to pursue various research areas without objection always providing insightful discussions about my work. I would also like to thank the members of my advisory committee (Professor Konstantina Tzia and Associate Professor Dimitris Karonis) for their support, their comments as well as the sincere interest they showed concerning the progress of my research work. I also thank all the members of the Laboratory of Fuels and Lubricants Technology for always being so helpful and friendly.

I especially thank my wife Ada for her love, support and encouragement throughout the course of this research. It would also have been impossible for me to contemplate this road, without the love and support of my parents. Finally I thank my beloved daughter Sophia, who was born during the 1st year of my PhD, and was always a source of inspiration and mental relaxation (even providing me with some "innovative" interpretation for bell shaped distributions (mountains) and joint PDF simulations (sun) produced by MATLAB).

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Φανούριο Ζαννίκο καταρχήν για το γεγονός ότι μου έδωσε την ευκαιρία να εκπονήσω αυτή τη Διδακτορική Διατριβή καθώς για την μετέπειτα αποτελεσματική υποστήριξη και καθοδήγηση έως το τέλος της ερευνητικής μου εργασίας. Μου έδωσε την ελευθερία να εμβαθύνω σε συγκεκριμένους ερευνητικούς τομείς χωρίς αντιρρήσεις, ενώ οι συζητήσεις μας για την εργασία μου ήταν πάντα χρήσιμες και εποικοδομητικές. Θα ήθελα επίσης να ευχαριστήσω τα δύο άλλα μέλη της τριμελούς συμβουλευτικής μου επιτροπής, την Καθηγήτρια Κωνσταντίνα Τζιά και τον Αναπληρωτή Καθηγητή Δημήτριο Καρώνη, για τα χρήσιμα σχόλια τους και το ειλικρινές ενδιαφέρον που έδειξαν για την πρόοδο της ερευνητικής μου δουλειάς. Ένα μεγάλο ευχαριστώ επίσης σε όλα τα μέλη του Εργαστηρίου Τεχνολογίας Καυσίμων και Λιπαντικών που διακατέχονταν πάντα από φιλική διάθεση και προθυμία να βοηθήσουν.

Ιδιαιτέρως θα ήθελα να ευχαριστήσω τη σύζυγό μου Άντα για την αγάπη, την υποστήριξη και την ενθάρρυνσή της καθ 'όλη τη διάρκεια της εκπόνησης της Διδακτορικής Διατριβής. Επίσης θα ήταν αδύνατο να ακολουθήσω αυτό το δρόμο χωρίς τα πολύτιμα εφόδια που μου έδωσε η στήριξη και η αγάπη των γονιών μου. Τέλος θέλω να ευχαριστήσω την αγαπημένη κόρη μου Σοφία η οποία γεννήθηκε τον πρώτο χρόνο της εκπόνησης της Διδακτορικής Διατριβής και έκτοτε αποτελούσε πάντα μια πηγή έμπνευσης αλλά και ψυχικής χαλάρωσης (δίνοντας μου ακόμη και «καινοτόμες» ερμηνείες των διαγραμμάτων του MATLAB με τις κατανομές Gauss να είναι τα βουνά και η προσομοίωση σημείων διμμεταβλητής κατανομή ο ήλιος….)

Dedication / Αφιέρωση

To my parents and my wife, Ada for their constant support and unconditional love.

To my daughter Sophia for inspiring me and having a positive impact on my work and everyday life.

Στους γονείς μου και την σύζυγό μου Άντα για την συνεχή υποστήριξη και την ανιδιοτελή αγάπη τους.

Στην κόρη μου Σοφία για την έμπνευση και το θετικό αντίκτυπο που δίνει στη δουλειά και την καθημερινή μου ζωή.

Measurement uncertainty under a philosophical perspective…

The Allegory of the Cave is presented by the Greek philosopher Plato in his work "The Republic" (380 BC). Plato has Socrates describe a gathering of people who have lived chained to the wall of a cave all of their lives, facing a blank wall. People watch shadows projected on the wall by things passing in front of a fire behind them. Socrates suggests that the shadows constitute reality for the prisoners because they have never seen anything else; they do not realize that what they see are shadows of objects in front of a fire, much less that these objects are inspired by real living things outside the cave. They can't see the real things unless they leave the cave.

Relocating the allegory from the world of philosophy into the world of measurements, we could liken the shadows on the wall to the results of measurements, the fire to the measurement model, and the real objects to the true value of measured quantities. The estimation of uncertainty is, in a sense, a conquest of liberty because allows us to have a clearer picture of the range where the measurement's true values lie.

Αβεβαιότητα της μέτρησης και φιλοσοφία…

Η Πολιτεία του Πλάτωνα (380 π.Χ) περιλαμβάνει την Αλληγορία του σπηλαίου, με την οποία ο Πλάτων εξηγεί μέσω του Σωκράτη τη Θεωρία των Ιδεών του. Σε ένα σπήλαιο, κάτω από τη γη, βρίσκονται μερικοί άνθρωποι αλυσοδεμένοι με τέτοιο τρόπο, ώστε να μπορούν να δουν μόνο τον απέναντί τους τοίχο. Πίσω τους ωστόσο είναι αναμμένη μια φωτιά. Έτσι οτιδήποτε εκδηλώνεται πίσω από την πλάτη τους αναπαριστάνεται ως σκιά στον απέναντι τους τοίχο. Επειδή οι άνθρωποι αυτοί σε ολόκληρη τη ζωή τους τα μόνα πράγματα που έχουν δει είναι οι σκιές των πραγμάτων, έχουν την εντύπωση ότι οι σκιές που βλέπουν πάνω στον τοίχο είναι τα ίδια τα πράγματα. Εάν όμως κάποιος από τους αλυσοδεμένους ανθρώπους του σπηλαίου κατορθώσει να ελευθερωθεί, να βγει από τη σπηλιά και να ανέβει πάνω στη γη και, κάτω από το φως του ήλιου πλέον, δει τα πράγματα, θα καταλάβει την πλάνη στην οποία ζούσε όσο ήταν μέσα στη σπηλιά.

Μεταφέροντας την αλληγορία από τον κόσμο της φιλοσοφίας στον κόσμο των μετρήσεων θα μπορούσε κάνεις να παρομοιάσει τις σκιές στον τοίχο με τα αποτελέσματα των μετρήσεων, τη φωτιά με το μηχανισμό μέτρησης (μετρητικό μοντέλο), τα πραγματικά αντικείμενα με την αληθή τιμή των μετρούμενων μεγεθών και την κατάκτηση της ελευθερίας με την εκτίμηση της αβεβαιότητας που μας επιτρέπει να έχουμε μια πιο ξεκάθαρη εικόνα για το εύρος στο οποίο βρίσκεται η πραγματική τιμή αυτού που μετράμε.

Contents

List of figures

Chapter 1 – Introduction

Chapter 2 - Theoretical background

Chapter 3 - Measurement uncertainty arising from sampling

Chapter 4 - Measurement uncertainty estimation of an analytical procedure

Chapter 5 - Estimation of the standard uncertainty of a calibration curve

Chapter 6 - The use of measurement uncertainty and precision data in conformity assessment

Chapter 7 - Conclusions

List of Tables

Chapter 1 – Introduction

Chapter 2 - Theoretical background

Chapter 3 - Measurement uncertainty arising from sampling

Chapter 4 - Measurement uncertainty estimation of an analytical procedure

Chapter 5 - Estimation of the standard uncertainty of a calibration curve

Chapter 6 - The use of measurement uncertainty and precision data in conformity assessment

1. Introduction

Test results should be reliable and accepted by all interested parties so that they reduce risks, shorten time to market and demonstrate the quality and safety of fuel products. The quality of a result and its fitness for purpose is directly related to the estimation of **measurement uncertainty**. The estimation of measurement uncertainty following recognized and valid methodologies is also a key requirement for laboratories or other organizations accredited or seeking accreditation according to international standards such as ISO/IEC 17025, ISO 15189, ISO/IEC 17043 or ISO Guide 34. Measurement uncertainty is defined as the "*non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used"*. A consistent and transferable evaluation of measurement uncertainty should follow the basic principles described in the document "**Guide to the Expression of Uncertainty in Measurement**" (GUM) produced by the Joint Committee for Guides in Metrology (JCGM).

1.1 Motivation and objectives

The market of fuels and related products becomes more diverse and sophisticated, turning quality control of fuels into an essential risk management activity for producers, traders and distributors. The fuels produced and placed on market should comply with strict requirements introduced by relevant legislation. In European Union (EU), for example, several directives [1,2] set technical specifications for automotive fuels used with positive ignition engines (petrol) or with compression ignition engines (diesel). Several test methods are used for the evaluation and assessment of the physical, mechanical, rheological, thermal, and chemical properties of crude oils, lubricating grease/oils, automobile and aviation gasoline, hydrocarbons, and other naturally occurring energy resources used for various industrial applications. These products are tested for their composition, purity, density, toxicity, thermal stability and miscibility / compatibility with other fluids and materials, among others. Test results should be reliable and accepted by all interested parties so that they reduce risks, shorten time to market and demonstrate the quality and safety of fuel products.

In general, laboratories produce results that are passed on to someone else (e.g. the customer) who will use them to solve a problem or answer questions. The social and economic impact of the laboratory getting a wrong result and the customer consequently reaching a false conclusion can be enormous. Thus, the laboratory should provide a high quality service to its customers. Quality in this context is not necessarily getting the most accurate results. Quality is providing results that are "fit for purpose", i.e. match the service with the requirements of the customer. This is achieved by providing results that:

- meet the specific needs of the customer,
- attract the confidence of the customer and all others who make use of the results, and
- represent value for money.

The judgement of "fitness for purpose" of a test method is inseparably related to the estimation of the **measurement uncertainty** which actually characterizes the quality of a result by accounting for both systematic and random errors. **Pre-analysis**, **analysis** and **post-analysis** are essential parts of the measurement cycle from client issue to decision on measurement result. In the various **pre-analytical**, **analytical** and **post-analytical** steps, the measurement uncertainty has to be either estimated or taken into account. Figure 1.1 shows the various steps of measurement cycle,

indicating the steps that measurement uncertainty has be estimated or taken into account, as well as the chapters of the thesis that deal with measurement uncertainty estimation and use.

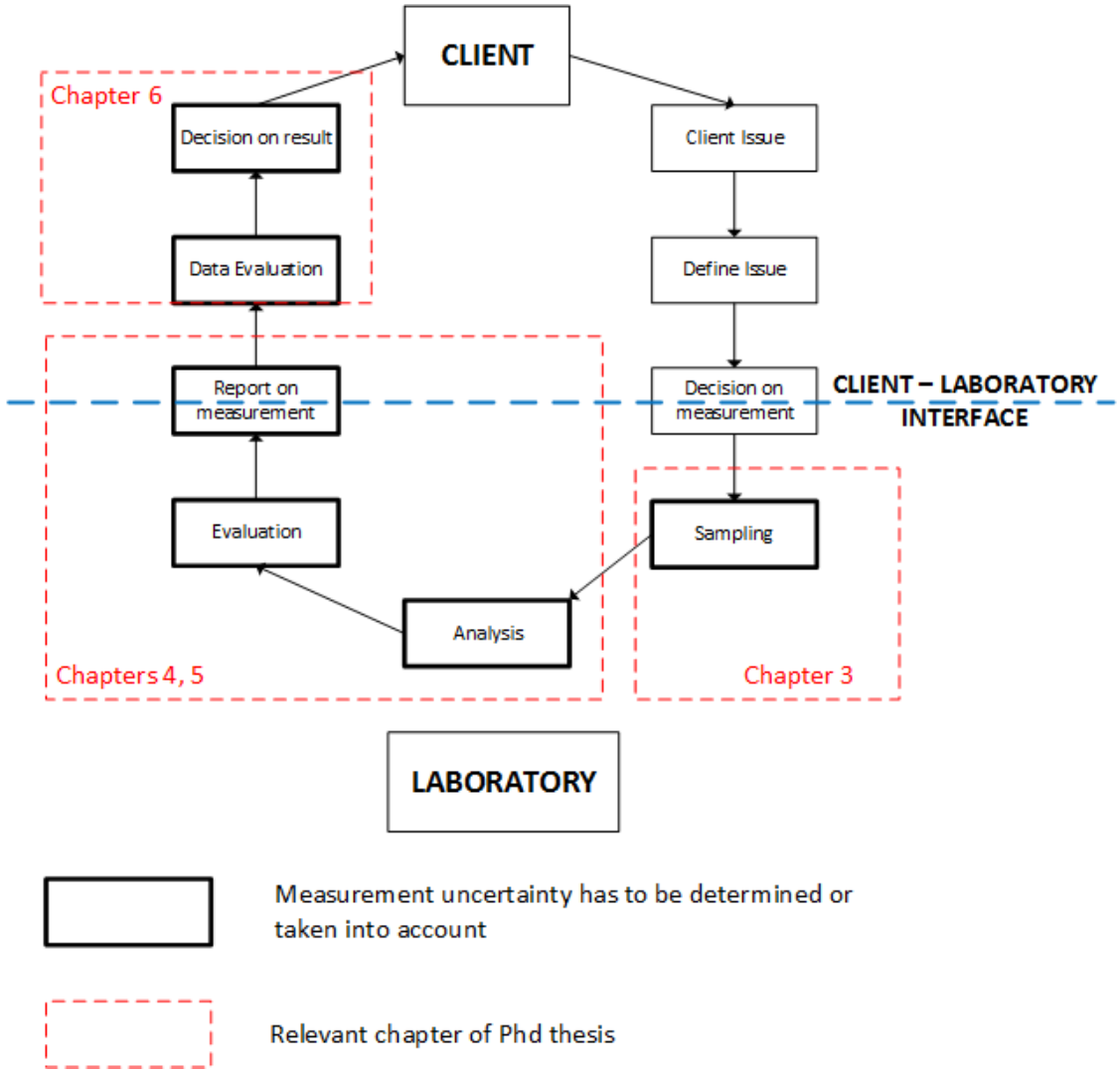


Figure 1.1 Measurement cycle

1.2 Quality management systems

Laboratories performing tests and calibrations as well as other organizations that support laboratories in their activities (e.g. proficiency testing providers, reference material producers) should implement quality management systems specifically designed to prove their technical competence. This competence is often confirmed by an independent authorative third party (accreditation body). The main standards against which, accreditation bodies accredit organizations are:

- ISO/IEC 17025 [3], for testing and calibration laboratories
- ISO 15189 [4], for medical laboratories,
- ISO/IEC 17043 [5], for proficiency testing providers,
- ISO Guide 34 [6], for reference material producers.

These standards define what is required by an organization in order for it to demonstrate both technical competence of personnel and the availability of all the technical resources needed to deliver reliable and fit for purpose services and/or products. In addition to the technical requirements, there is a second major component of the standards which requires that the management systems of the organizations meet the principles of ISO 9001 [7].

A key requirement for laboratories or other organizations accredited or seeking accreditation according to standards mentioned above is the estimation of measurement uncertainty following recognized and valid methodologies, which in most of the cases involve measurement data analysis. Table 1.1 presents the requirements of the accreditation standards concerning uncertainty estimation as stated in certain clauses.

Table 1.1 Requirements of accreditation standards making reference to uncertainty estimation

Standard	Clause	Requirement
ISO/IEC 17025	5.4.1	The laboratory shall use appropriate methods and procedures for all tests and/or calibrations within its scope. These include sampling, handling, transport, storage and preparation of items to be tested and/or calibrated, and, where appropriate, an estimation of the **measurement uncertainty** as well as statistical techniques for analysis of test and/or calibration data.
	5.4.6.1	A calibration laboratory, or a testing laboratory performing its own calibrations, shall have and shall apply a procedure to estimate the **uncertainty of measurement** for all calibrations and types of calibrations.
	5.4.6.2	Testing laboratories shall have and shall apply procedures for estimating **uncertainty of measurement**. In certain cases the nature of the test method may preclude rigorous, metrologically and statistically valid, calculation of **uncertainty of measurement**. In these cases the laboratory shall at least attempt to identify all the components of uncertainty and make a reasonable estimation, and shall ensure that the form of reporting of the result does not give a wrong impression of the uncertainty. Reasonable estimation shall be based on knowledge of the performance of the method and on the measurement scope and shall make use of, for example, previous experience and validation data.

Standard	Clause	Requirement
	5.4.6.3	When estimating the **uncertainty of measurement**, all uncertainty components which are of importance in the given situation shall be taken into account using appropriate methods of analysis.
	5.10.3.1	…test reports shall, where necessary for the interpretation of the test results, include the following: … c) where applicable, a statement on the estimated **uncertainty of measurement**; information on uncertainty is needed in test reports when it is relevant to the validity or application of the test results, when a customer's instruction so requires, or when the **uncertainty** affects compliance to a specification limit; …
	5.10.4.1	…calibration certificates shall include the following, where necessary for the interpretation of calibration results: … b) the **uncertainty of measurement** and/or a statement of compliance with an identified metrological specification or clauses thereof; …
	5.10.4.2	When statements of compliance are made, the **uncertainty of measurement** shall be taken into account.
ISO 15189	5.5.1.4	The laboratory shall determine **measurement uncertainty** for each measurement procedure in the examination phase used to report measured quantity values on patients' samples. The laboratory shall define the performance requirements for the **measurement uncertainty** of each measurement procedure and regularly review estimates of **measurement uncertainty**. The laboratory shall consider **measurement uncertainty** when interpreting measured quantity values. Upon request, the laboratory shall make its estimates of **measurement uncertainty** available to laboratory users. Where examinations include a measurement step but do not report a measured quantity value, the laboratory should calculate the **uncertainty of the measurement** step where it has utility in assessing the reliability of the examination procedure or has influence on the reported result.
ISO/IEC 17043	4.4.5.1	The proficiency testing provider shall document the procedure for determining the assigned values for the measurands or characteristics in a particular proficiency testing scheme. This procedure shall take into account the metrological traceability and **measurement uncertainty** required to demonstrate that the proficiency testing scheme is fit for its purpose.
	4.4.5.2	Proficiency testing schemes in the area of calibration shall have assigned values with metrological traceability, including **measurement uncertainty**.
	4.4.5.3	For proficiency testing schemes in areas other than calibration, the relevance, needs and feasibility for metrological traceability and associated **measurement uncertainty** of the assigned value shall be determined by taking into account specified requirements of participants or other interested parties, or by the design of the proficiency testing scheme.
	4.4.5.4	When a consensus value is used as the assigned value, the proficiency testing provider shall document the reason for that selection and shall estimate the **uncertainty** of the assigned value as described in the plan for the proficiency testing scheme.
	4.6.1.2	The proficiency testing provider shall give detailed documented instructions to all participants. Instructions to participants shall

Standard	Clause	Requirement
		include: … f) specific and detailed instructions on the manner of recording and reporting test or measurement results and associated **uncertainties**. If the instructions include reporting of the **uncertainty** of the reported result or measurement, this shall include the coverage factor and, whenever practicable, the coverage probability; …
	4.8.2	Reports shall include the following, unless it is not applicable or the proficiency testing provider has valid reasons for not doing so: … m) details of the metrological traceability and **measurement uncertainty** of any assigned value; …
ISO Guide 34	5.4.3	In planning the production processes, the reference material producer shall have procedures and service facilities, for … m) establishing uncertainty budgets and estimating **uncertainties** of the assigned property values, if applicable; n) defining acceptance criteria for verifying that **uncertainty** estimates are applicable for replacement batches of reference materials produced…
	5.9.1	The reference material producer shall meet the requirements of ISO/IEC 17025 with respect to tests, calibrations and measurements under their responsibility (including preparation of items, sampling, handling, preservation, storage, packaging, transport to subcontractors, estimation of **measurement uncertainty** and analysis of measurement data).
	5.14.2	The stability of the reference material shall be assessed… In case of detectable degradation, both the degradation and its **uncertainty** shall be included in the assessment…
	5.16.2	An important aspect of establishing the property values of the reference material being produced is an assessment of their **uncertainties**. The reference material producer shall carry out an assessment of the **measurement uncertainties** to be included in the assignment of the property values in accordance with the requirements of the GUM (ISO/IEC Guide 98-3). In the process of estimating **uncertainties** of the property values of interest, any uncertainties resulting from between-unit variations and/or from possible doubts on stability (both during storage and during transportation) shall be assessed in accordance with ISO Guide 35 and shall be included in the assigned uncertainty. A statement of the **measurement uncertainty** is mandatory for certified values. In case values are assigned to non-certified reference materials (e.g. “indicative values” or “information values”), a statement of **uncertainties** is highly recommended to improve the use of the material.
	5.18.4	The reference material producer shall employ best efforts to notify customers of any change to the assigned value or **uncertainty** for any products not expired.

1.3 What is uncertainty of measurement?

The purpose of a measurement is to provide information about a quantity of interest (a measurand). Unfortunately, no measurement is exact and many factors (visible or invisible) can undermine that measurement. When a quantity is measured, the outcome depends on the item being measured, the measuring system, the measurement procedure, the skill of the operator, the environment, sampling issues or other effects. These effects give rise to uncertainty in measurement due to measurement errors that can be either systematic or random. A systematic error (an estimate of which is known as a measurement bias) is associated with the fact that a measured quantity value contains an offset. A random error is associated with the fact that when a measurement is repeated it will generally provide a measured quantity value that is different.

Prior to the introduction of the ISO Guide to the Expression of Uncertainty in Measurement (GUM) [8], it was common to express systematic and random error values relating to the measurement, along with a best estimate of the measurand, in order to express what is learned about the measurand. The GUM provided a different way of thinking about measurement, in particular about how to express the perceived quality of the result of a measurement. Rather than expressing the result of a measurement by providing a best estimate of the measurand, along with information about systematic and random error values, the GUM approach expresses the result of a measurement as a best estimate of the measurand, along with an associated measurement uncertainty. Using the GUM approach, it is possible to characterize the quality of a measurement by accounting for both systematic and random errors on a comparable footing, and a method is provided for doing that. This method refines the information previously provided in an “error analysis”, and puts it on a probabilistic basis through the concept of measurement uncertainty.

The definition the term uncertainty of measurement is provided by the International Vocabulary of Metrology (VIM) [9]:

“non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.”

Notes to the VIM definition additionally indicate that measurement uncertainty includes components arising from systematic effects, such as components associated with corrections and the assigned quantity values of measurement standards, as well

as the definitional uncertainty. Sometimes estimated systematic effects are not corrected for, but instead, associated measurement uncertainty components are incorporated. The parameter characterizing the dispersion of the quantity values may be, for example, a standard deviation called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability.

Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by Type A evaluation of measurement uncertainty from the statistical distribution of the quantity values from a series of measurements and can be characterized by standard deviations. The other components, which may be evaluated by Type B evaluation of measurement uncertainty, can also be characterized by standard deviations, evaluated from probability density functions based on experience or other information. In general, for a given set of information, the measurement uncertainty is associated with a stated quantity value attributed to the measurand. A modification of this value results in a modification of the associated uncertainty. Figure 1.2 presents a concept diagram intended to provide a visual presentation of the relations between the various measurement uncertainty concepts defined and termed in VIM.

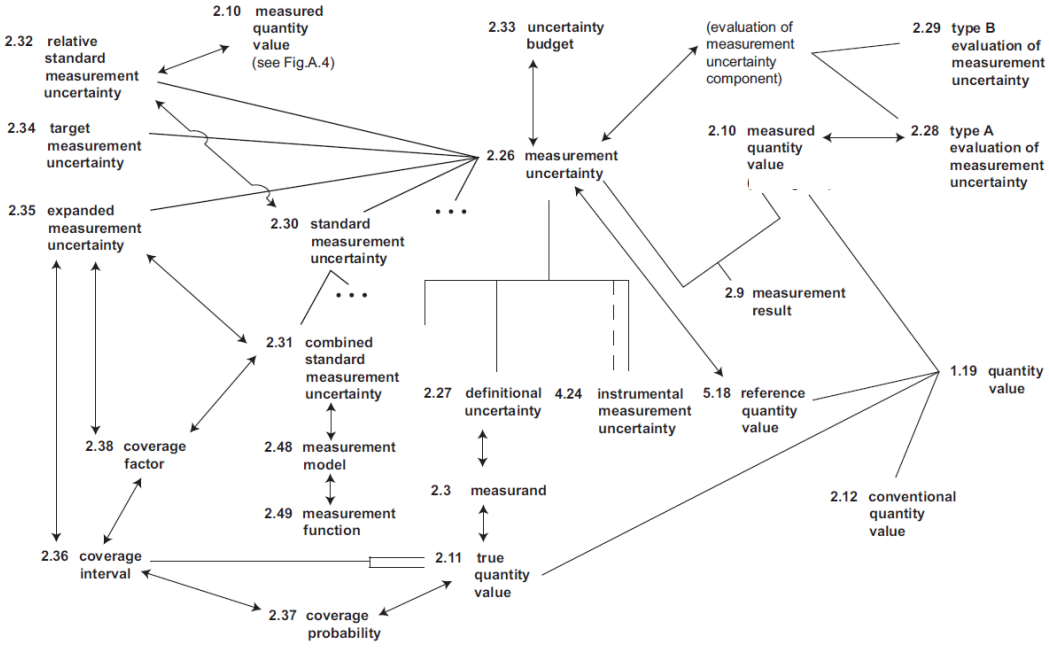


Figure 1.2 Concept diagram for measurement uncertainty [9]

A statement of measurement uncertainty is indispensable in judging the fitness for purpose of a measured quantity value and plays a central role in quality assessment and quality standards. Paul De Bievre [10] has given a definition of the quality of measurement result established by metrological criteria:

"Quality of a measurement result is the fitness-for intended- use of that result, expressed by its metrological traceability including a comparison of the ensuing measurement uncertainty to a pre-set target measurement uncertainty"

Measurement uncertainty enables users of a measured quantity value to make comparisons, in the context of conformity assessment, to obtain the probability of making an incorrect decision based on the measurement, and to manage the consequential risks.

1.4 Historical background

In 1977, recognizing the lack of international consensus on the expression of uncertainty in measurement, the world's highest authority in metrology, the Comité International des Poids et Mesures (CIPM), requested the Bureau International des Poids et Mesures (BIPM) to address the problem in conjunction with the national standards laboratories and to make a recommendation. The recognized need to arrive at an internationally accepted procedure for expressing measurement uncertainty and for combining individual uncertainty components into a single total uncertainty resulted in the Recommendation INC–1 (1980), "*Expression of experimental uncertainties*" [11], of the Working Group on the Statement of Uncertainties, convened in 1980 by the Bureau International des Poids et Mesures (BIPM). The CIPM approved the Recommendation in 1981 [12], and reaffirmed it in 1986 [13].

The responsibility for developing a detailed guide based on the Working Group Recommendation was given to the Technical Advisory Group on Metrology (TAG4) of the International Organization for Standardization (ISO), in which six other international organizations were represented, namely, the BIPM, the International Electrotechnical Commission (IEC), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Union of Pure and Applied Chemistry (IUPAC), the International Union of Pure and Applied Physics (IUPAP) and

the International Organization of Legal Metrology (OIML). The resulting document was published in 1993 and reprinted with minor corrections in 1995 [8].

In 1997 a Joint Committee for Guides in Metrology (JCGM), chaired by the Director of the BIPM, was created by the seven international organizations that had originally in 1993 prepared the "*Guide to the expression of uncertainty in measurement*' (GUM) and the `*International vocabulary of metrology - basic and general concepts and associated terms*" (VIM). The JCGM assumed responsibility for these two documents from the ISO Technical Advisory Group 4 (TAG4). In 1998 one more organization joined these seven international organizations, namely, the International Laboratory Accreditation Cooperation (ILAC). The JCGM has two Working Groups. Working Group 1, "Expression of uncertainty in measurement", has the task of promoting the use of the GUM and preparing supplements for its broad application. Working Group 2, "Working Group on International Vocabulary of Basic and General Terms in Metrology (VIM)", has the task of revising and promoting the use of the VIM.

Following wide international dissemination of the GUM over several years, it has been decided to supplement the Guide with a number of documents. These documents comprise an introductory document, a document concerned with concepts and basic principles, four supplements to the GUM, and two documents concerned with the use of measurement uncertainty in the context of conformance to specified requirements and the application of the method of least squares. The titles of the documents, under the banner "Evaluation of measurement data", are:

(i) JCGM 104:2009 - Evaluation of measurement data – An introduction to the "Guide to the expression of uncertainty in measurement" and related documents [14]

(ii) JCGM 101:2008 - Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method [15]

(iii) JCGM 102:2011 - Evaluation of measurement data – Supplement 2 to the "Guide to the expression of uncertainty in measurement" – Extension to any number of output quantities [16]

(iv) JCGM 106:2012 - Evaluation of measurement data – The role of measurement uncertainty in conformity assessment [17]

(v) JCGM 105 - Evaluation of measurement data – Concepts, principles and methods for the evaluation of measurement uncertainty

(vi) JCGM 103 - Evaluation of measurement data – Supplement 3 to the "Guide to the expression of uncertainty in measurement" – Developing and using measurement models

(vii) JCGM 107 - Evaluation of measurement data – Applications of the least-squares method

(viii) JCGM 108 - Evaluation of measurement data — Supplement 4 to the “Guide to the expression of uncertainty in measurement” – Bayesian methods

The first four of these documents have been approved and are available for the scientific community, while the others are in the stage of preparation.

Moreover, there are other general guides available for uncertainty estimation, such as those produced by National Metrology Institutes (e.g NPL [18,19], NIST [20]) and accreditation bodies/organizations (e.g. UKAS [21], ILAC [22], EA [23,24]). There are also sector specific guides such as those produced by EURACHEM [25] or Nordtest [26]. These guides, even if sometimes employ the application of empirical approaches, remain compliant with the basic principles of GUM, resulting in consistent and transferable evaluation of measurement uncertainty.

1.5 Thesis outline

The chapters of the thesis deal with statistical and numerical methods concerning the estimation and use the measurement uncertainty in certain parts of the measurement cycle (Figure 1.1). In particular:

- **Chapter 2** provides the background theory concerning the approaches for the estimation and use of uncertainty.
- **Chapter 3** presents the development and application of methodologies for the estimation of measurement uncertainty arising from sampling.
- **Chapter 4** focuses on the estimation of measurement uncertainty of an analytical procedure using ISO GUM and Monte Carlo Method.
- **Chapter 5** considers uncertainty estimation in analytical methods employing the construction of a calibration function using linear regression.
- **Chapter 6** investigates the use of measurement uncertainty and precision data in conformity assessment of automotive fuel products.
- **Chapter 7** summarizes the achieved results and proposes some future work.

Each of Chapters 3, 4, 5 and 6 can be read independently having their own summary, introduction and conclusions. Nevertheless all together give the integrated picture

about the statistical and numerical techniques that have been developed and applied in all the parts of the measurement cycle for the estimation or use of measurement uncertainty.

1.6 PhD publication list

(i) D. Theodorou, Y. Zannikou, G. Anastopoulos, F. Zannikos. Coverage interval estimation of the measurement of Gross Heat of Combustion of fuel by bomb calorimetry: Comparison of ISO GUM and adaptive Monte Carlo method. **Thermochimica Acta** (2011) 526: 122– 129

(ii) D. Theodorou, Y. Zannikou, F. Zannikos. Estimation of the standard uncertainty of a calibration curve: application to sulfur mass concentration determination in fuels. **Accreditation and Quality Assurance** (2012) 17: 275–281

(iii) D. Theodorou, N. Liapis, F. Zannikos. Estimation of measurement uncertainty arising from manual sampling of fuels. **Talanta** (2013) 105: 360-365

(iv) D. Theodorou, F. Zannikos. The use of measurement uncertainty and precision data in conformity assessment of automotive fuel products. **Measurement** (2014) 50: 141-151

(v) D. Theodorou, Y. Zannikou, F. Zannikos. Components of measurement uncertainty from a measurement model with two stages involving two output quantities. **Chemometrics and Intelligent Laboratory Systems** (2015) 146: 305–312

2. Theoretical background

The **measurement uncertainty** provides a quantitative indication of the quality of a measurement result and has implications for the interpretation of analytical results in the context of regulatory compliance or **conformity assessment**. There are two broad approaches for the estimation of uncertainty, modeling and empirical.

The "**modeling approach**" quantifies all sources of uncertainty individually, and then combines (propagates) them through a mathematical model. The most widely understood modelling approach is described in the "**Guide to the Expression of Uncertainty in Measurement**", known as the **GUM**. An alternative approach to uncertainty evaluation is **Monte Carlo** simulation, where the propagation of uncertainties is undertaken numerically rather than analytically. Such techniques are useful for validating the results returned by the application of the GUM, as well as in circumstances where the assumptions made by the GUM do not apply. The principles for the uncertainty of the modelling approaches can be extended to the evaluation of uncertainties associated with measurement models with **two stages** and/or involving **multiple output quantities**.

On the other hand, there are "**empirical approaches**", which are based on whole-method performance investigations designed and conducted so as to comprise the effects from as many relevant uncertainty sources as possible. Empirical approaches are particularly appropriate where major contributions to uncertainty cannot readily be modelled in terms of measurable input quantities, and where many laboratories use essentially identical test methods and equipment.

When prior knowledge (before a measurement is made) is available, **Bayesian statistics** can be used. Type A uncertainties or uncertainties near zero may be evaluated through a Bayesian approach.

Different approaches exist for the assessment of laboratory and measurement procedure **bias** and its treatment related to the evaluation of uncertainty. Uncorrected bias and its uncertainty should be incorporated into measurement uncertainty.

2.1 Introduction

The main purpose of measurement is to enable decisions to be made. The reliability of these decisions depends on knowing the uncertainty of the measurement results. If the uncertainty of measurements is underestimated, for example because the sampling is not taken into account, then erroneous decisions may be made that can have large financial consequences. The fitness for purpose of measurement results can only be judged by having reliable estimates of their uncertainty. For this reason it is essential that effective procedures are available for estimating the uncertainties arising from all parts of the measurement process. These must include uncertainties arising from any relevant sampling and physical preparation, and within laboratory measurement procedures. Measurement uncertainty interacts closely with decision making. Therefore, in order to utilize a result to decide whether it indicates compliance or non-compliance with a specification, it is necessary to take into account the measurement uncertainty. Fig. 2.1 shows the correlation between the measurement process stages and the work presented in the various thesis chapters.

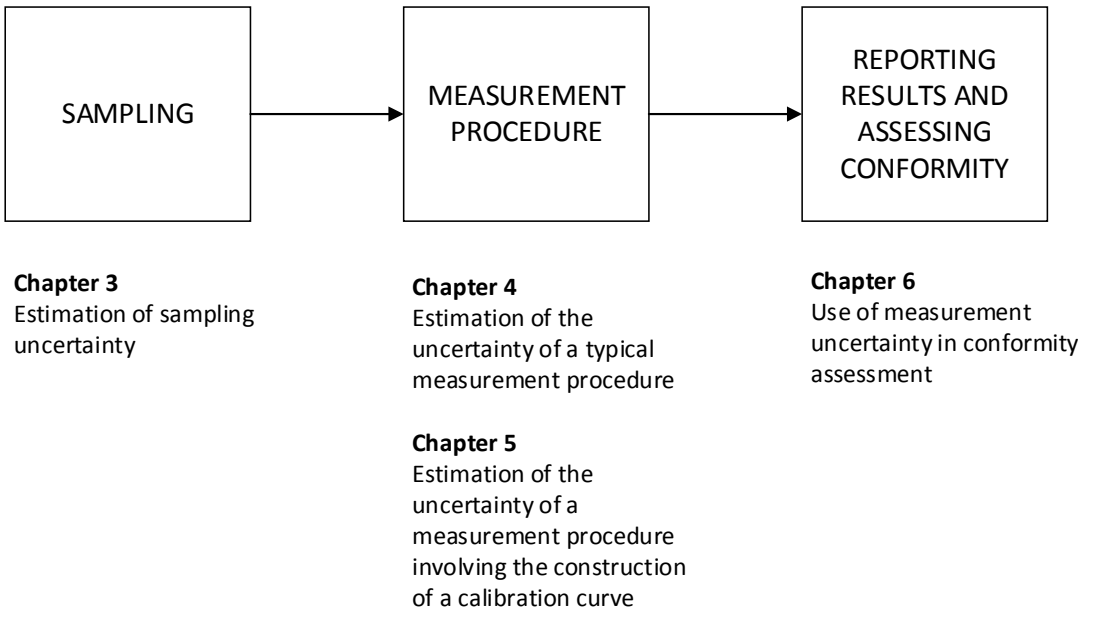


Figure 2.1 Measurement process stages and correlation with thesis chapters

The present Chapter sets the theoretical basis and outlines the statistical and numerical methods used in the work presented in Chapters 3, 4, 5 and 6.

2.2 Uncertainty – An overview

The "Guide to the Expression of Uncertainty in Measurement", known as the GUM[8], is acknowledged as the master document on measurement uncertainty throughout the scientific community. Through this document, certain principles have been established that require that:

- uncertainty evaluation is comprehensive, accounting for all relevant sources of measurement error
- uncertainties arising from random and systematic effects are treated alike, i.e. are expressed and combined as variances of associated probability distributions
- statistical evaluation of measurements (Type A) and alternative techniques, based on other data / information (Type B), are recognized and utilized as equally valid tools
- uncertainties of final results are expressed as standard deviations (standard uncertainty) or by multiples of standard deviations (expanded uncertainty) with a specified numerical factor (coverage factor).

There are two broad methodologies for the estimation of uncertainty, the modeling approach and the empirical approach. The "modeling approach" which is consistent with GUM and is described as a "bottom up" approach, quantifies all sources of uncertainty individually, and then combines (propagates) them through a mathematical model. On the other hand, there are "empirical approaches", which are based on whole-method performance investigations designed and conducted so as to comprise the effects from as many relevant uncertainty sources as possible. The data utilized in these approaches are typically precision and bias data obtained from within-laboratory validation studies, quality control, interlaboratory method validation studies, or proficiency tests (PTs) [27].

Figure 2.2 shows a classification of uncertainty approaches based on a distinction between uncertainty estimation using data produced by the laboratory itself (called intralaboratory approach) and uncertainty estimation based on data produced by collaborative studies (called interlaboratory approach).

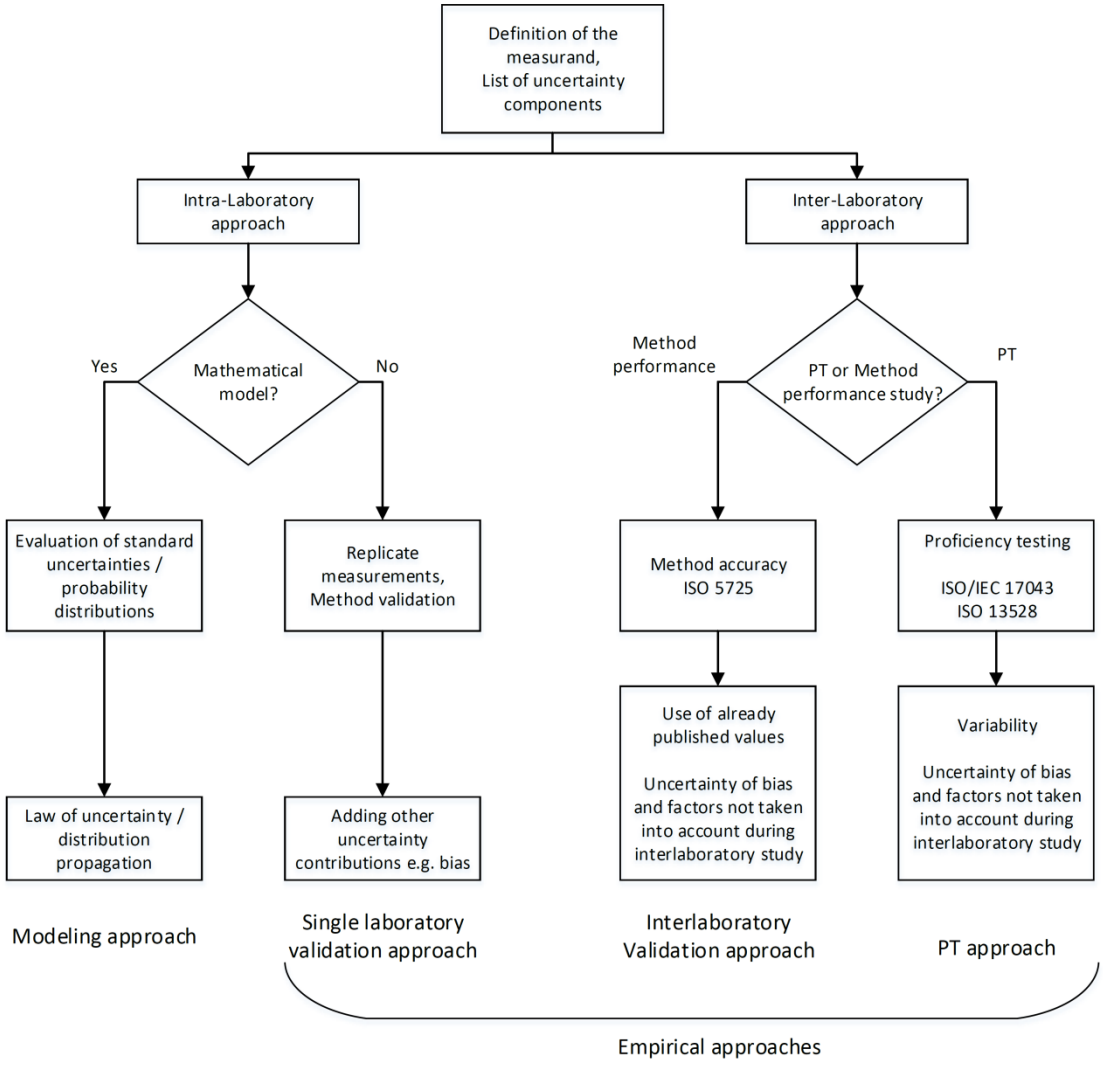


Figure 2.2 Uncertainty estimation approaches [27]

2.3 Basic concepts and terms

2.3.1 Measurement

According to the International Vocabulary of Metrology (VIM) [9] **measurement** is the process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity. The **measurement result** is the outcome of any measurement activity and is what is reported to the end-user, be it a regulatory body, an accreditation body or a commercial customer.

A measurement result is generally expressed as a single **measured quantity value** accompanied by a **measurement uncertainty**. This can be interpreted as a "set of quantity values", meaning that any value, within the interval defined by the

measurement uncertainty is a possible value for the measurand. This information provides the end-user with sufficient information on the reliability of the measurement result, to be taken into account when for example it is to be compared with a stated limit. The measurement uncertainty and the level of confidence associated with it are part of a measurement result.

2.3.2 Accuracy, trueness and precision

The action of measuring introduces changes in the system subjected to measurement. This leads to a **measurement error**, affecting each individual measurement. In principle, the measurement error is represented by the difference between the measured quantity value and a reference quantity value. In practice, for an individual measurement on a test sample/ item, the measurement error is unknowable. This is because, in this case, the reference quantity value is the unknown true quantity value for the measurand. The measurement error consists of two components, systematic and random, which represent respectively, the constant or predictable variation and the unpredictable variation in a series of replicate measurements.

The concepts of measurement error, random measurement error and systematic measurement error are illustrated in Figure 2.3, where y_i indicates an individual measured value of the measurand. A histogram is also shown in the figure for a set of such individual measured values, where the area of each rectangle of the histogram represents the number of times that an individual measured value was obtained within the interval defined by the width of the rectangle. And then a fit to the histogram data is shown by the black curve, and $\overline{y}$ is the location of the maximum of the curve. The unknown true value, y_{true} that is desired to be obtained from the measurement is plotted to the left in the figure.

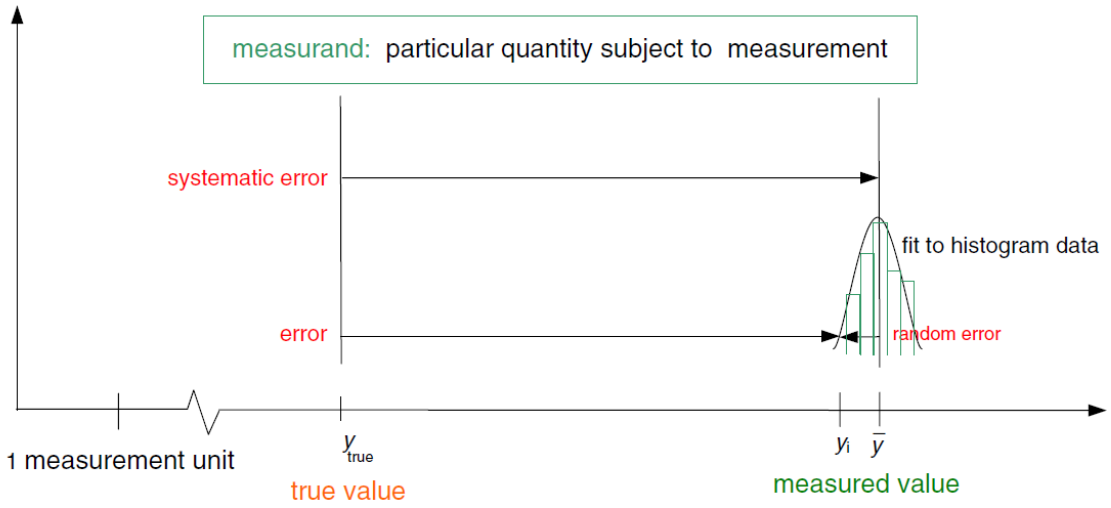


Figure 2.3 Illustration of the concept of measurement error and its components, random measurement error and systematic measurement error [28].

Systematic and random errors are terms closely related to measurement trueness and measurement precision, respectively. **Measurement trueness** is defined as the closeness of agreement between average of an infinite number of replicate measured quantity values and a reference quantity value. Measurement trueness expresses the hypothetical ability of a measurement procedure to yield results close to expected reference quantity values, such as the value of a certified reference material (CRM). Trueness is not a quantity and therefore cannot be expressed numerically. However, trueness is inversely related to **systematic measurement error** (component of measurement error that in replicate measurements remains constant or varies in a predictable manner). The systematic measurement error may be estimated as **measurement bias**.

Measurement precision is defined as the closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions. Measurement precision is related to **random measurement error** and is a measure of how close results are to one another. Measurement results cannot be corrected to remove the effect of random error but the size of the random error can be reduced by making replicate measurements and calculating the mean value. Measurement precision is expressed numerically using measures of **imprecision** such as the standard deviation calculated from results obtained by carrying out replicate measurements on a suitable material under specified conditions. VIM defines three measurement conditions: repeatability condition, intermediate precision condition and reproducibility condition. Estimates

of measurement repeatability and intermediate measurement precision are obtained in a single laboratory. **Repeatability condition of measurement** refers to measurements being made on portions of the same material by a single analyst, using the same procedure, under the same operating conditions over a short time period. Measurement repeatability is often used to provide an estimate of within-batch variability in results. Under **intermediate measurement conditions**, measurements are made on portions of the same material using the same procedure, but over an extended time period and possibly by different analysts who may be using different pieces of equipment. Intermediate measurement precision is often used to provide an estimate of between batch variability. Intermediate measurement conditions are user-defined and the conditions used should always be recorded. Estimates of measurement reproducibility are obtained from measurement results produced at different laboratories. **Reproducibility condition of measurement** refers to measurements being made on portions of the same material by different analysts working in different locations.

The effect of both precision and trueness are included in accuracy. **Measurement accuracy** describes how close a single measurement result is to the true quantity value. Accuracy cannot be given a numerical value but measurement results are said to be "more accurate" when the measurement errors and the measurement uncertainty, are reduced. Figure 2.4 illustrates trueness, precision and accuracy, through the use of the "target model". The "shots" on the target represent individual measurement results; the reference quantity value (true value) is the centre of the target. The best accuracy is achieved in case b) where the individual results are all close to the reference value. In cases a) and b) there is no significant bias as the results are all clustered in the centre of the target. However, the precision is poorer in case a) as the results are more widely scattered. The precision in case d) is similar to that in case b). However, there is a significant bias in case d) as all the results are off-set from the centre in the same area of the target. The accuracy is poorest in case c) as the results are widely scattered and are off-set to the right of the target. Measurement accuracy cannot be used to give a quantitative indication of the reliability of measurement results. An estimate of measurement uncertainty is required for this reason.

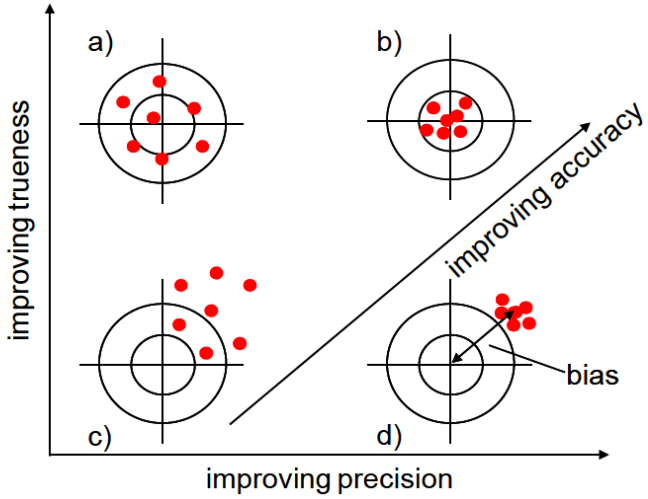


Figure 2.4 Target model to illustrate trueness, precision and accuracy.

2.3.3 Measurement uncertainty

The **measurement uncertainty** provides a quantitative indication of the quality of a measurement result. It is defined in VIM as a "*non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used*". This definition expresses the fact that parameters used to describe the dispersion of distributions, e.g. standard deviations, are usually positive. The statement, "*based on the information used*", explains why it is necessary to declare what is included in the estimate of measurement uncertainty. This does not mean that one can choose what to include and what to leave out. There are certain approaches to evaluating measurement uncertainty and these are described in the following sections. The result of a measurement consists of two quantitative parts: i) the measured quantity value, often an average or median of individual measurements, and ii) the measurement uncertainty. If, when the result is reported the uncertainty is included, it can be presented in the format (value ± uncertainty) and units. The uncertainty is interpreted as providing an interval within which the value of the measurand is believed to lie.

Estimates of measurement uncertainty can be expressed in a number of different ways, e.g. as a standard deviation or a confidence interval. However, in order to be able to combine uncertainty estimates they must be expressed in the same form, so some conversion may be necessary. Following GUM guidelines, uncertainty

estimates should be expressed as standard measurement uncertainties before they are combined. There are different forms of uncertainty:

- $u(x_i)$ – the **standard measurement uncertainty** for quantity x_i is an uncertainty expressed as a standard deviation.
- $u(y)$ – the **combined standard measurement uncertainty** for the measurand Y, is a mathematical combination of several individual standard measurement uncertainties.
- U – the **expanded measurement uncertainty** is normally what the laboratory reports to the customer. The expanded uncertainty provides an interval within which the value of the measurand is believed to lie with a higher level of confidence. The value of U is obtained by multiplying the combined standard measurement uncertainty $u(y)$ by a coverage factor k, i.e. $U = k \cdot u(y)$. The choice of the factor k is based on the level of confidence desired.

2.4 Modelling approach

2.4.1 GUM uncertainty framework

The most widely understood modelling approach to evaluation of uncertainty is described in GUM. This procedure is based on a mathematical-measurement model formulated to account for the interrelation of all the influence quantities that significantly affect the measurand. Corrections are assumed to be included in the model to account for all recognized, significant systematic effects. The application of the law of propagation of uncertainty enables evaluation of the combined uncertainty on the result. The steps to be followed for evaluating and expressing the uncertainty of the result of a measurement as presented in GUM are summarized in Table 2.1 and Figure 2.5.

Table 2.1 Procedure for evaluating and expressing the uncertainty according to GUM uncertainty framework [8, 21].

No	Step	Comments
1	Determine the mathematical relationship between the values of the input quantities, X_1, $X_{2,..,}X_N$ and that of the measurand, Y: $Y = f(X_1, X_2, ..., X_N)$	The set of input quantities X_1, X_2,…,X_N may be categorized as: • quantities whose values and uncertainties are directly determined in the current measurement • quantities whose values and uncertainties are brought into the measurement from external sources
2	Identify all corrections that have to be applied to the results of measurements of quantity (measurand) for the stated conditions of measurement.	The function f should contain every quantity, including all corrections and correction factors that can contribute a significant component of uncertainty to the result of the measurement.
3	Determine x_i, the estimated value of input quantity X_i, either on the basis of the statistical analysis of series of observations or by other means.	
4	Calculate the standard uncertainty $u(x_i)$ of each input estimate x_i. For an input estimate obtained from the statistical analysis of series of observations, the standard uncertainty is evaluated using Type A evaluation of standard uncertainty. For an input estimate obtained by other means, the standard uncertainty $u(x_i)$ is evaluated using Type B evaluation of standard uncertainty.	In a Type A, evaluation the standard uncertainty $u(x_i)$ is calculated as the standard deviation $s(x_i)$ of the mean of m measurements. The degrees of freedom associated with Type A standard uncertainties based on m measurements are $v_i = m - 1$. In a Type B, evaluation the standard uncertainty $u(x_i)$ is evaluated by scientific judgment based on information such as previous measurement data, experience with or general knowledge of materials and instruments involved, manufacturer's specifications, calibration data etc. When Type B uncertainties are used, it may be necessary to convert an interval into a standard uncertainty using information about the distribution of the value and the degrees of freedom (see Table 2.2).

No	Step	Comments
5	Evaluate the covariances associated with any input estimates that are correlated.	The degree of correlation between x_i and x_j is characterized by the estimated correlation coefficient: $r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)}$ where $u(x_i, x_j) = u(x_j, x_i)$ is the estimated covariance associated with x_i and x_j.
6	Calculate the result of the measurement, that is, the estimate y of the measurand Y, from the functional relationship f using for the input quantities X_i the estimates x_i obtained in step 3.	In some cases, the estimate y is taken as the arithmetic mean or average of a series of independent determinations, each determination having the same uncertainty and each being based on a complete set of observed values of the input quantities X_i obtained at the same time.
7	Determine the combined standard uncertainty $u(y)$ of the measurement result y from the standard uncertainties and covariances associated with the input estimates, using the law of propagation of uncertainty: $u^2(y) = \sum_i c_i^2 u^2(x_i) + 2\sum_{i<j} c_i c_j u(x_i) u(x_j) r(x_i, x_j)$ where c_i is the partial derivative: $\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial X_i}\Big\|_{x_1, x_2, \ldots, x_N}$ (also known as sensitivity coefficient).	If the estimates x_i and x_j are independent then $r(x_i, x_j)$=0 and the combined standard uncertainty is given by: $u^2(y) = \sum_i c_i^2 u^2(x_i)$ For the very special case where all of the input estimates are correlated with correlation coefficients $r(x_i, x_j)$=1 the combined standard uncertainty is given by: $u^2(y) = \left[\sum_i c_i u(x_i)\right]^2$ i.e. the combined standard uncertainty $u(y)$ is simply a linear sum of terms representing the variation of the output estimate y generated by the standard uncertainty of each input estimate x_i.

No	Step	Comments
8	Obtain the expanded uncertainty U by multiplying the combined standard uncertainty $u(y)$ by a coverage factor k depending on the level of confidence required: for normal distribution, a value k=2 corresponds to an approximate confidence level (coverage probability) of 95%, and k=3 of 99.7 %.	In some cases it may not be practical to base the Type A evaluation on a large number of readings, which could result in the coverage probability being significantly less than 95% if a coverage factor of k= 2 is used. In such cases a t-distribution is assumed and the degrees of freedom, ν_{eff}, associated with $u(y)$ are calculated using the Welch-Satterthwaite formula: $$\nu_{\text{eff}} = \frac{u^4(y)}{\sum_i \frac{u_i^4(y)}{\nu_i}}$$ where ν_i corresponds to the degrees of freedom of $u(x_i)$.

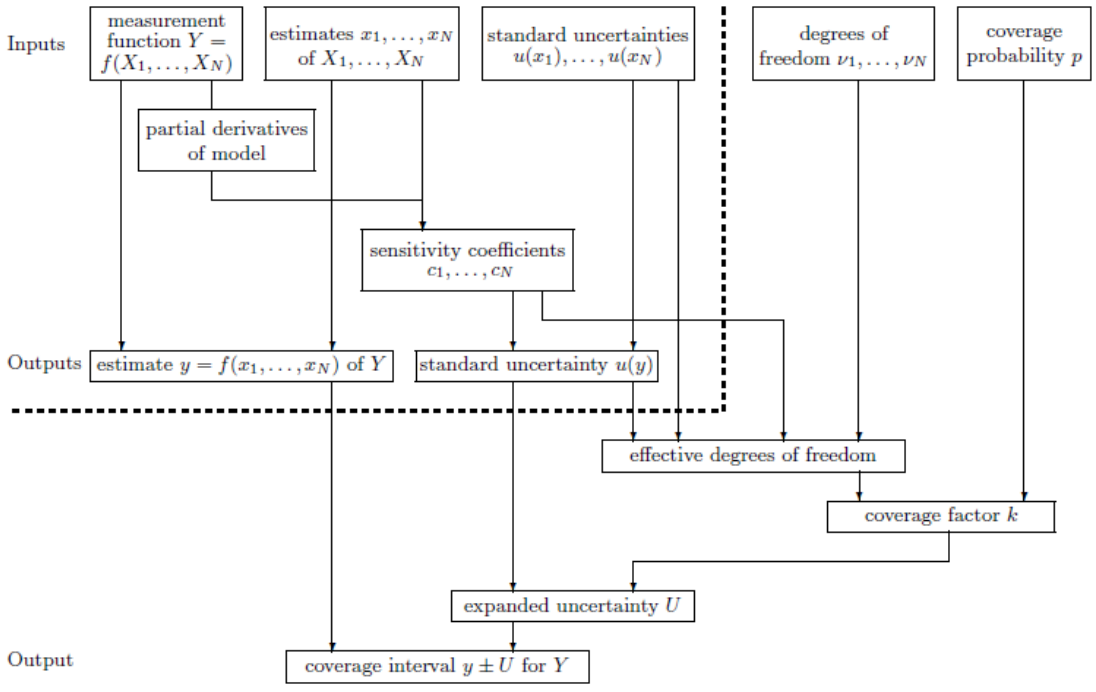
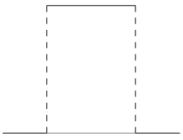
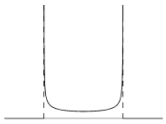
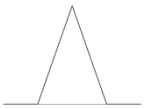
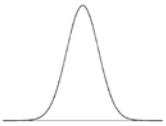
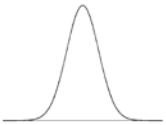
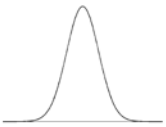


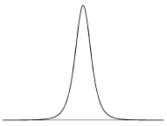
Figure 2.5 Measurement uncertainty evaluation using the GUM uncertainty framework. The top-left part of the figure, bounded by broken lines relates to obtaining an estimate y of the output quantity Y and the associated standard uncertainty u(y), and the remainder relates to the determination of a coverage interval for Y [14].

Table 2.2 Typical assumed probability distributions for input quantities and evaluation of their standard uncertainties [15,21]

Assumed probability distribution (and illustration)	Expression used to obtain the standard uncertainty	Comments
Rectangular 	$u(x_i) = \frac{b-\alpha}{2\sqrt{3}} = \frac{c}{\sqrt{3}}$	If the only available information regarding a quantity X_i is a lower limit α and an upper limit b (range b-α = 2c) with $\alpha < b$, then, according to the principle of maximum entropy, a rectangular distribution R(α, b) over the interval [α, b] would be assigned to X_i.
U-shaped 	$u(x_i) = \frac{b-a}{2\sqrt{2}} = \frac{c}{\sqrt{2}}$	If a quantity X_i is known to cycle sinusoidally, with unknown phase Φ, between specified limits α and b (range b- α = 2c), with $\alpha < b$, then, according to the principle of maximum entropy, a rectangular distribution R(0, 2π) would be assigned to X_i. The distribution assigned to X_i is the arc sine (U-shaped) distribution U(α, b).
Triangular 	$u(x_i) = \frac{b-a}{2\sqrt{6}} = \frac{c}{\sqrt{6}}$	If a quantity X_i is known to lie between specified limits α and b (range b- α = 2c), with $\alpha < b$, but there is reason to expect that extreme values are unlikely, it is normally appropriate to assume a triangular distribution T(α, b) over the interval [α, b].

Assumed probability distribution (and illustration)	Expression used to obtain the standard uncertainty	Comments
Gaussian (Normal) (from repeated measurements) 	Single measurement: $u(x_i) = s(x_i)$ Results subjected to averaging: $u(x_i) = \frac{s(x_i)}{\sqrt{m}}$	Where the uncertainty component was evaluated experimentally from the dispersion of repeated measurements, it can readily be expressed as a standard deviation. For the contribution to uncertainty in single measurements, the standard uncertainty is simply the observed standard deviation; for results subjected to averaging (*m* replicates), the standard deviation of the mean.
Gaussian (Normal) (from a calibration certificate) 	$u(x_i) = \frac{U}{k}$	A calibration certificate normally quotes an expanded uncertainty *U* at a specified, high coverage probability. A coverage factor, *k*, will have been used to obtain this expanded uncertainty from the combination of standard uncertainties. It is therefore necessary to divide the expanded uncertainty by the same coverage factor to obtain the standard uncertainty.

Assumed probability distribution (and illustration)	Expression used to obtain the standard uncertainty	Comments
Gaussian (Normal) (from a manufacturer's specification) 	$u(x_i) = \frac{tolerance}{k}$	Some manufacturers' specifications are quoted at a given coverage probability (sometimes referred to as "confidence level"), e.g. 95% or 99%. In such cases, a normal distribution can be assumed and the tolerance limit is divided by the coverage factor *k* for the stated coverage probability. For a coverage probability of 95%, *k* = 2 and for a coverage probability of 99%, *k* = 2.58. If a coverage probability is not stated then a rectangular distribution should be assumed.
t-distribution (from repeated measurements) 	$u(x_i) = \frac{\sqrt{n-1}}{\sqrt{n-3}} \frac{s(x_i)}{\sqrt{m}}$	When the uncertainty component was evaluated experimentally from the dispersion of a small number of repeated measurements and a Gaussian distribution cannot be assumed then a scaled and shifted t-distribution $t_v(\bar{x}, s^2/m)$ with $v=m-1$ degrees of freedom is assigned for X_i, where $\bar{x}$ and s^2 are the average and the variance of the measurement results, respectively.

Assumed probability distribution (and illustration)	Expression used to obtain the standard uncertainty	Comments
t-distribution (from a calibration certificate) 	$u(x_i) = \frac{U_p}{k_p}$	If the source of information about a quantity X_i is a calibration certificate in which a best estimate x_i, the expanded uncertainty U_p, the coverage factor k_p and the effective degrees of freedom ν_{eff} are stated, then a scaled and shifted *t*-distribution $t_\nu(x_i,(U_p/k_p)^2)$ with $\nu = \nu_{eff}$ degrees of freedom is assigned to X_i.

It has to be noted that the law of propagation of uncertainty in the GUM uncertainty framework applies exactly for a linear measurement model, since it is obtained from the formula for the variance of a linear combination of random variables given the variances and covariances of those variables. However, determination of a coverage interval requires knowledge of the probability density function for the measurand *Y*. In the GUM uncertainty framework it is assumed that this probability density function is Gaussian, or Student's *t* for a specific degrees-of-freedom parameter. The probability density function has Gaussian form when the input quantities are independent and Gaussian, or when they are multivariate Gaussian, or when the central limit theorem practically holds. For a non-linear model there is a further condition for the GUM uncertainty framework to apply: the model can be well approximated locally by a linear model within several standard deviations of the best estimates of the input quantities [29].

The situations where the GUM uncertainty framework might not be satisfactory, are summarized as follows:

(i) the measurement function is non-linear,
(ii) the probability distributions for the input quantities are asymmetric,
(iii) the uncertainty contributions are not of approximately the same magnitude,
(iv) the probability distribution for the output quantity is either asymmetric, or not a Gaussian or a t-distribution.

2.4.2 Monte Carlo method

Sampling techniques, such as Monte Carlo simulation, provide an alternative approach to uncertainty evaluation in which the propagation of uncertainties is undertaken numerically rather than analytically. Such techniques are useful for validating the results returned by the application of the GUM uncertainty framework, as well as in circumstances where the assumptions made by the GUM uncertainty framework do not apply. In fact, these techniques are able to provide much richer information, by propagating the distributions (rather than just the uncertainties) of the inputs, X_i through the measurement model f to provide the distribution of the output Y [15, 21].

The Monte Carlo method (MCM) for uncertainty evaluation is straightforward to apply in the sense that the only interaction with the measurement model is in terms of evaluating the model and sensitivity coefficients are not required. A further feature of the method is that it is a general and broadly applicable approach to the propagation of distributions because it makes no assumption about the measurement model (e.g., that it is only mildly non-linear), no assumption about the probability distributions for the input quantities (e.g., that no one distribution dominates), and no assumption about the distribution for the output quantity (e.g., that it takes a particular form). Although the method always produces an approximate solution to the propagation of distributions, the quality of the approximation can be controlled by the number of Monte Carlo trials [30].

In a Monte Carlo simulation the mathematical model of the underlying measurement system is run over and over again, each time using a different set of random numbers representing the input variables. Each of these sets of random numbers combines via the model to represent a different outcome. Each run of the model is called a simulation, or trial, and at the end of each trial the outcome of the process is recorded. Each different outcome arises, through the measurement model, corresponding to a particular set of random numbers being applied to it. If the model is a good representation of the real -world system, then, by running a large enough number of trials (each with a different set of random numbers), the whole range of possible outputs can be produced; these form the distribution of the output. The Monte Carlo numerical simulation tends to require up to 10^6 trials for calculating a 95% coverage interval which is correct to one or two significant decimals digits. In fact, it is often more reliable to implement an adaptive Monte Carlo procedure which involves carrying out an increasing number of trials until the

results have stabilized in a statistical sense [15]. Coverage intervals (either shortest or probabilistically equal) and other statistical information can then straightforwardly be produced from the PDF of the output quantity. The estimate *y* of the output quantity *Y* is estimated by the average of the *M* MCM trials which produce *M* measurement model values (y_r, $r = 1, \ldots, M$):

$$y = \frac{1}{M}\sum_{r=1}^{M} y_r \tag{2.1}$$

From the output distribution confidence intervals can be produced, as can other statistical information. The standard uncertainty $u(y)$ associated with *y* is estimated as the standard deviation of the *M* model values:

$$u(y) = \sqrt{\frac{1}{M-1}\sum_{r=1}^{M}(y_r - y)} \tag{2.2}$$

The 95% coverage interval for the output can be obtained by the 2.5- and 97.5-percentiles of the distribution of MCM results (probabilistically equal coverage interval). The GUM Supplement 1 [15] provides guidance for an alternative coverage, the "shortest coverage interval." This interval is the shortest among all possible intervals for a distribution of MCM results where each interval has the same coverage probability (95%). Figures 2.6 and 2.7 illustrate the way the MCM works for a measurement model *f* consisting of input quantities (X_1,…,X_N) which are used for estimating the output quantity *Y*.

Chapter 4 presents the estimation of measurement uncertainty of an analytical procedure using GUM and MCM.

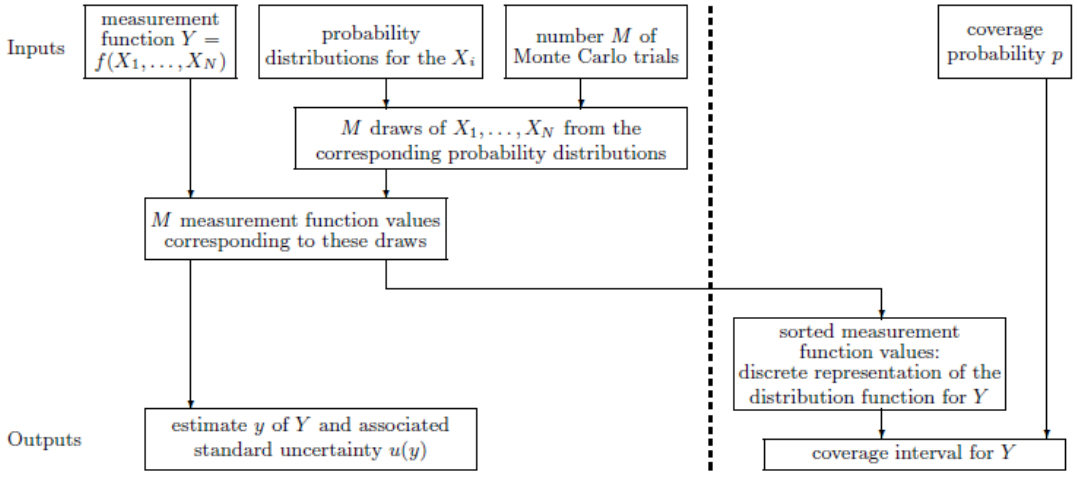


Figure 2.6 Measurement uncertainty evaluation using a Monte Carlo method, where the left part of the figure (left of the broken line) relates to obtaining an estimate *y* of the output quantity *Y* and the associated standard uncertainty *u*(*y*), and the remainder relates to the determination of a coverage interval for *Y* [14].

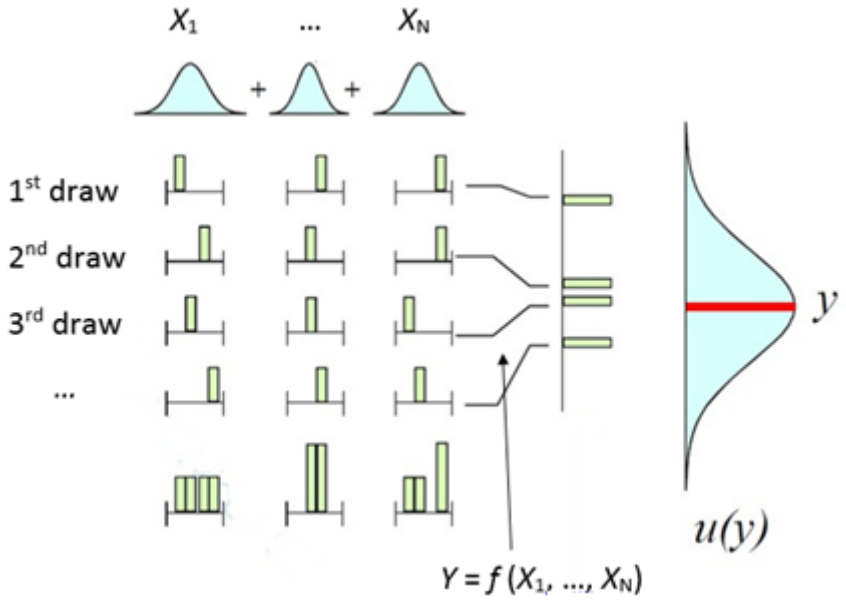


Figure 2.7 The basic principle of the Monte Carlo method. Samples are drawn from the distributions of the input quantities (X_1,…,X_N) and through the measurement model *f* a distribution for the measurand *Y* is generated.

2.4.3 Kragten approximation

The Kragten approximation is recommended in the Eurachem/CITAC Guide [25] and provides an approximation of the first derivative required by the GUM's first-order

formula (Table 2.1 – Step No7). This approach uses a numerical method of differentiation and relies on the approximation:

$$\frac{\partial f}{\partial x_i} \approx \frac{f(x_i + u(x_i)) - f(x_i)}{u(x_i)} \tag{2.3}$$

where x_i and $u(x_i)$ are the value of an influence quantity and the corresponding standard uncertainty and y is the value of the measurand inferred from a measurement model $Y=f(X_i)$. Since the GUM's first-order formula gives the corresponding standard uncertainty $u_i(y)$ in y as the partial derivative multiplied by $u(x_i)$ this approximation leads to the very simple form:

$$u_i(y) = c_i u(x_i) = \frac{\partial f}{\partial x_i} u(x_i) \approx f(x_i + u(x_i)) - f(x_i) \tag{2.4}$$

The approximation is quite good enough for a fit for purpose estimate required of testing laboratories where uncertainties are small and $f(X_i)$ close to linear (the conditions assumed by the first-order GUM formula). Moreover, as discussed by Ellison [31], in certain cases, while the Kragten estimate may be a poor approximation to the first derivative at the estimate, the result can, paradoxically, provide a better indication of the correct standard uncertainty than a good first-order calculation. Reference [32] discusses the point of Kragten approach assumptions more fully and suggests methods of checking the validity of the assumptions.

In **Chapter 5** the Kragten approximate numerical method is applied for the estimation of the uncertainty associated with the construction and use of a calibration curve used for the determination of sulfur mass concentration in fuels. The results of the estimations are compared with the results of the GUM uncertainty framework and the MCM approach.

2.5 Empirical approaches

Empirical approaches, which include interlaboratory comparisons and method validation studies, are particularly appropriate where major contributions to uncertainty cannot readily be modelled in terms of measurable input quantities, and where many laboratories use essentially identical test methods and equipment. The

ISO/IEC 17025 standard [3] accordingly references ISO 5725 series of standards "Accuracy (trueness and precision) of measurement methods and results" [33-38], as well as the GUM, among its uncertainty evaluation requirements applicable to testing laboratories. When applying empirical approaches, it is however, important to retain the consistency provided by adherence to the GUM concepts and recommendations. Careful application of the different approaches can ensure that all the different approaches remain compliant with the basic principles of the GUM [27].While GUM (modelling approach) uses mathematical measurement models, empirical approaches use statistical models as the basis for data analysis.

As shown in Figure 2.1 empirical approaches may be classified either as interlaboratory or intralaboratory depending on the source and the type of the data used. The intralaboratory empirical approaches use data from single laboratory method validation studies. The interlaboratory empirical approaches use data from collaborative method performance data or from interlaboratory comparisons (proficiency testing). Table 2.3 presents the basic principles of the 3 main categories of empirical approaches.

Table 2.3 Basic principles of the main categories of empirical approaches.

Approach	Basic principles
Single laboratory validation approach	The major sources of variability are assessed by method validation study. Estimates of bias, repeatability, and within laboratory reproducibility can be obtained by organizing experimental work inside the laboratory. Information can also be obtained from quality control data (control charts). Combined with experimental investigation of important individual effects, this approach provides essentially all of the data required for uncertainty estimation.
Interlaboratory validation approach	The major sources of variability are assessed by interlaboratory studies which provide estimates of repeatability (repeatability standard deviation), reproducibility (reproducibility standard deviation and (sometimes) trueness of the method (measured as a bias with respect to a known reference value). ISO 21748 [39] provides guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.

Proficiency testing approach	If the same method is used by all the participants to the Proficiency Testing scheme, the standard deviation is equivalent to an estimate of interlaboratory reproducibility and can, in principle, be used in the same way as the reproducibility standard deviation obtained from collaborative study (interlaboratory validation). Further, over several rounds, the deviations of laboratory results from the assigned value can provide a preliminary evaluation of the bias for that laboratory.

In analytical methods, certain empirical approaches are widely used for the estimation of uncertainty arising from sampling, when it is difficult or costly to establish mathematical models reflecting the influence of the various uncertainty contributions. The empirical approaches use repeated sampling and analysis, under various conditions, to quantify the effects caused by factors such as the heterogeneity of the analyte in the sampling target and variations in the application of one or more sampling protocols. Data obtained from repeated sampling and analysis protocols are then analyzed using suitable statistical methods.

In **Chapter 3** three statistical methods used for the estimation of the uncertainty caused by manual sampling of fuels from petroleum retail stations are developed, applied and compared. Moreover, in **Chapter 4** data from interlaboratory comparisons (proficiency testing) are used for the estimation of uncertainty and the result is compared with the results of other approaches.

2.6 Bayesian uncertainty analysis

Another method of uncertainty analysis is by the use of Bayesian statistics. This is based on the concept of degree of belief. Bayes' theorem gives the "posterior" odds on the correctness of a belief (given the new evidence that has just been observed), making use of the "prior" odds that the belief is correct, i.e. an estimate of the plausibility of the belief before one has the new data. A concept of likelihood ratio allows the "prior" odds to be adjusted according to the each piece of new evidence [21]. In Bayesian statistics, the measurement data are constants and the value of the measurand is a random variable. The probability distribution for the value of the measurand is a "state of knowledge distribution" that describes the degrees of belief about all possible values that could be attributed to the value of the measurand. The degrees of belief are based on all available information including scientific

judgement, current measurements, and ancillary knowledge. Similar state of knowledge probability distributions apply to the other unknown quantities involved in measurement.

A Bayesian uncertainty analysis starts with prior distributions, which represent the states of knowledge before measurements are made, for the values of unknown quantities. Negligible prior knowledge is expressed by using non-informative probability distributions. The measurements are then used to update the prior distributions using Bayes' theorem to obtain posterior distributions. The posterior distribution for the value of the measurand is a probability distribution that could reasonably be attributed to the value of the measurand after measurements are made. A measure of centrality (such as the expected value) and a measure of dispersion (such as the standard deviation) of the Bayesian posterior distribution quantify, respectively, the result of measurement and its associated standard uncertainty [40].

The GUM contains elements from both classical and Bayesian statistics. For example, in a Type B uncertainty evaluation probability distributions are used to express one's state of knowledge about the quantities concerned. Proceeding in such a way is related to a Bayesian point of view. On the other hand, the Type A evaluation of standard uncertainty associated with the mean of repeated observations can be understood as an estimate of the standard deviation of the corresponding sampling distribution and hence relates to classical statistics. Moreover, although GUM-Supplement 1 [15] and GUM-Supplement 2 [16] do not explicitly apply Bayes' theorem, these supplements appear to adopt a Bayesian point of view. For example, the introduction of GUM-Supplement 1 says: "*The use of PDFs as described in this Supplement is generally consistent with the concepts underlying the GUM. The PDF for a quantity expresses the state of knowledge about the quantity, i.e. it quantifies the degree of belief about the values that can be assigned to the quantity based on the available information. The information usually consists of raw statistical data, results of measurement, or other relevant scientific statements, as well as professional judgement*" This intention is clearly associated with a Bayesian view, and the distributions obtained by the application of GUM-Supplement 1 may equal those derived in a Bayesian uncertainty analysis [41-43].

Type A uncertainties can be evaluated through a Bayesian approach. Under classical approach the standard uncertainty $u(x_i)$ is calculated as the standard deviation $s(x_i)$ of the mean of m measurements. When the number m of measurements is small,

the standard deviation $s(x_i)$ is uncertain leading to unreliable estimations. Even, the application of Welch-Satterthwaite formula described in GUM, Section G4 may lead to incorrect results. Under the Bayesian approach the standard uncertainty $u(x_i)$ is calculated as:

$$u_{Bayes}(x_i) = \sqrt{\frac{m-1}{m-3}} s(x_i) \tag{2.5}$$

As the number m of mutually independent and normally distributed measurements increases, the t-distribution tends to normal distribution. The uncertainty $u(x_i) = s(x_i)$ from classical statistics may be interpreted as an approximation to the Bayesian uncertainty. The approximation is poor when m is small but improves as m increases. The factor $\sqrt{(m-1)/(m-3)}$ built into the Bayesian uncertainty accounts for the statistical uncertainty that arises from a small number of measurements. It turns out that the estimates from a classical statistical analysis are either equal or approximately equal to the corresponding estimates from a Bayesian analysis with non-informative prior probability distributions [44]. The results of the application of the Bayesian approach in Type A uncertainties and the comparison with the results of other approaches are presented in **Chapter 4**.

Bayesian statistics may also be used to provide expanded uncertainty estimates near zero [25]. If the expanded uncertainty has been calculated using classical statistics, the interval – including any part lying below zero – will, by definition, have 95 % coverage. However, since the (true) value of the measurand cannot lie outside the possible range, it is possible to simply truncate this interval at the edge of the possible range and yet retain the required 95 % coverage. This truncated classical confidence interval therefore maintains exact 95 % coverage (Figure 2.8).

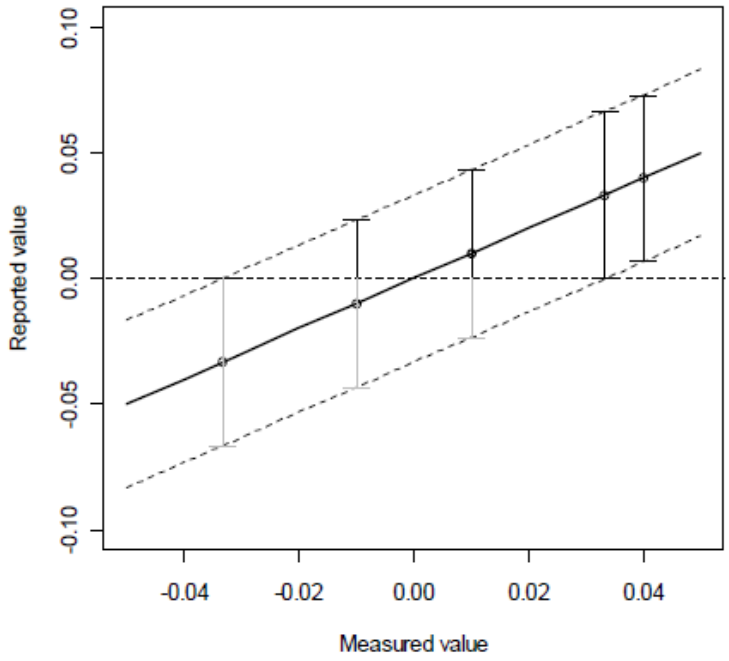


Figure 2.8 Example of truncating classical confidence intervals close to zero. The solid, partial bars show the reported uncertainty interval after truncation [25]

In the case of a quantity known to be limited to lie above zero and a measurement which provides information in the form of a *t*-distribution, it can be shown [45] that the resulting distribution of possible values is approximately a truncated *t*-distribution. To obtain a minimally biased result and an expanded uncertainty interval with appropriate coverage, it is recommended that:

i. The mode of the posterior distribution be reported. For a truncated *t*-distribution, this is either the observed mean value or zero if the observed mean value is below zero.
ii. The expanded uncertainty interval is calculated as the maximum density interval containing the required fraction of the posterior distribution. The maximum density interval is also the shortest interval that contains the required fraction of the distribution.

The Bayesian interval provides the same minimal bias as the classical approach, with the useful property that as the observed mean value falls further below zero, the reported uncertainty increases (Figure 2.9). This makes it particularly appropriate for reporting results which are expected to fall consistently very close to a limit such as zero or 100 %, such as in the estimation of purity for highly pure materials [25].

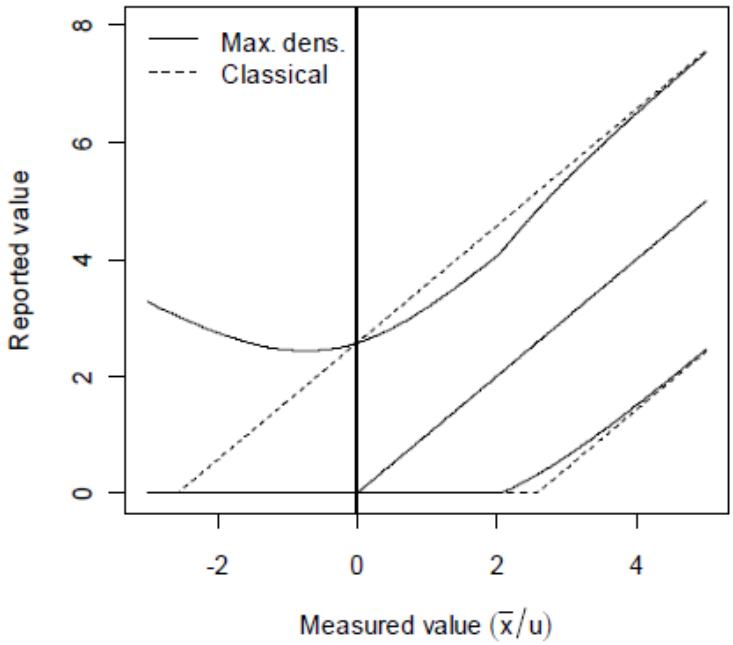


Figure 2.9 Example of bayesian maximum density interval (solid lines) for 5 degrees of freedom as a function of *x*. The dashed line shows the corresponding classical interval [25].

2.7 Treatment of bias in estimating uncertainty

Whereas intermediate precision and repeatability are estimated as standard deviations and are included as contributions to the uncertainty budget, different approaches exist for the assessment of laboratory and measurement procedure bias and its treatment related to the evaluation of uncertainty [46, 47]. Nevertheless, consistent treatment of measurement bias, including the question of whether or not to correct measurement results for bias, is essential for the comparability of measurement results [48].

A prerequisite for the application of the GUM is that "*the result of a measurement has been corrected for all recognized significant systematic effects*" (GUM 3.2.4). In practice, however, correction for potential bias is not always possible or justifiable. The bias may not be statistically significant. The bias magnitude may, or may not, be significant compared with the required uncertainty. Even where a bias is detectable on a related material (such as a reference material), the cause of the bias may not be known, and without a known mechanism, it may be unsafe to apply a correction

based on studies of another material. It may, as GUM paragraph F2.4.5 recognizes, not be economical, practical or even possible to investigate this bias in detail.

If a single test item with a reference value x_{ref} is used to assess the bias and the mean value $\bar{x}_r$ is obtained with p replicates x_i using the measurement procedure under any of the specified conditions above, the bias can be calculated as follows:

$$bias = \bar{x}_r - x_{\text{ref}} = \frac{\sum x_i}{p} - x_{\text{ref}} \tag{2.6}$$

An example of the estimation of bias as the difference between the mean value of several measurement results and a reference quantity value is shown in Figure 2.10.

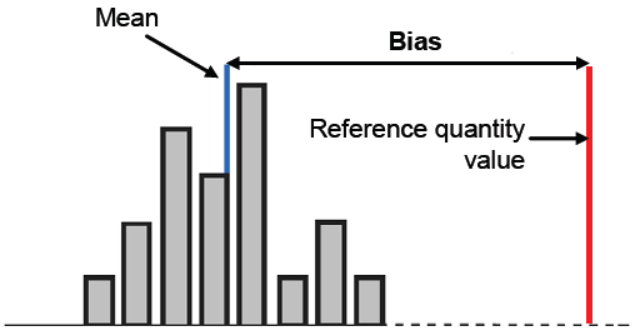


Figure 2.10 Schematic illustration of the estimation of measurement bias. The mean of several measurement results is compared with a reference quantity value.

If a CRM is used in the bias estimation, then the bias uncertainty, $u(bias)$, can be calculated:

$$u(bias) = \sqrt{\frac{s_r^2}{p} + u^2(x_{ref})} \tag{2.7}$$

where $u(x_{\text{ref}})$ is the standard uncertainty in the certified value of the CRM and s_r is the sample standard deviation of the p replicate results of the CRM analysis. The estimated bias is tested for its significance. The bias is not significant when:

$$|bias| \leq ku(bias) \tag{2.8}$$

where *k* is usually a value of the coverage factor if the number of effective degrees of freedom is high enough. Sometimes it is more precisely required to replace *k* by the two-tailed value from Student's *t*-distribution for the effective degrees of freedom associated to *u*(*bias*) at a given level of confidence.

When the systematic error is assumed to be relative, recovery, *R*, is calculated

$$R = \frac{\overline{x}_r}{x_{\text{ref}}} \tag{2.9}$$

and its difference from 1 is tested on significance:

$$\left|R - 1\right| \leq ku(R) \tag{2.10}$$

where *u*(*R*) is the uncertainty of *R*, which can be estimated by analogy with Equation (2.7). When the systematic error is statistically significant all observed results, y_{obs}, measured on routine samples should be corrected by using the determined value of bias or *R*:

$$y_{\text{corr}} = y_{\text{obs}} - bias \tag{2.11}$$

or

$$y_{\text{corr}} = \frac{y_{\text{obs}}}{R} \tag{2.12}$$

where y_{corr} denotes the measurement results after the correction. The standard uncertainty of bias or recovery is then included in the evaluation of the combined standard uncertainty of the corrected result.

When the systematic error is not statistically significant, or it is statistically significant but the analyst decides not to apply a correction, then the uncorrected bias should be incorporated into the expanded uncertainty. Table 2.4 presents the main approaches to increase uncertainty in the presence of bias information [48].

Table 2.4 Approaches to increase uncertainty in the presence of bias information

Uncertainty increase strategy	Abbreviation	Remark	References
Correction applied			
$U = k\sqrt{u(y_{\text{obs}})^2 + u(bias)^2}$	GUM	Required for significant bias. Covariance assumed equal to zero	GUM [8]
$U = k\sqrt{u(y_{\text{obs}})^2 + u(bias)^2}$	EURACHEM	$u(bias)$ included irrespective of significance of bias	EURACHEM [5]
Correction not applied			
$U = k\sqrt{u(y_{\text{obs}})^2 + bias^2 + u(bias)^2}$	RSSu		Lira & Wögler [47] Nordtest [26], NIST [49]
$U = k\sqrt{u(y_{\text{obs}})^2 + \left(\frac{bias}{k}\right)^2 + u(bias)^2}$	RSSU		IUPAC [50], Barwick et al. [51]
$U_+ = \max\left(0, k\sqrt{u(y_{\text{obs}})^2 + u(bias)^2} + bias\right)$ $U_- = \max\left(0, k\sqrt{u(y_{\text{obs}})^2 + u(bias)^2} - bias\right)$	SUMU	Asymmetrical interval. Equation applies when $ku(y_{obs}) \le bias$	NIST [52]
$U = k\sqrt{u(y_{\text{obs}})^2 + u(bias)^2} + \lvert bias\rvert$	$SUMU_{MAX}$	Also described in GUM F 2.4.5	IUPAC [50], Maroto et al. [53]
$U = k\sqrt{u(y_{\text{obs}})^2 + u(bias)^2} + E\lvert bias\rvert$ where E is dependent on the bias and is in the range 0–2	U_e	For $\frac{\lvert bias\rvert}{\sqrt{u(y_{\text{obs}})^2 + u(bias)^2}} > 0.8$ *U* may also be calculated using: $U = 1.65\sqrt{u(y_{\text{obs}})^2 + u(bias)^2} + \lvert bias\rvert$	Synek [54]

2.8 Using measurement uncertainty in decision-making and conformity assessment

Regulatory compliance (or conformity assessment) often requires that a measurand, is shown to be within particular limits. Measurement uncertainty clearly has implications for the interpretation of analytical results in this context. Detailed guidance on how to take uncertainty into account when assessing compliance is given in the EURACHEM Guide "Use of uncertainty information in compliance assessment" [55] and JCGM 106 "Evaluation of measurement data – The role of measurement uncertainty in conformity assessment" [17].

According to Pendrill [56] the essential steps in conformity assessment are:

i. Define entity and its quality characteristics to be assessed for conformity with specified requirements.
ii. 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
iii. Produce test results by performing measurements of the quality characteristics together with expressions of measurement uncertainty.
iv. Decide if test results indicate that the entity and the measurements themselves are within specified requirements or not.
v. Assess risks of incorrect decisions of conformity.
vi. Assess the conformity of entity to specified requirements in terms of impact.

More simply, the basic requirements for deciding whether or not to accept the test item are:

- A specification giving upper and/or lower permitted limits of the characteristics (measurands) being controlled.
- A decision rule that describes how the measurement uncertainty will be taken into account with regard to accepting or rejecting a product according to its specification and the result of a measurement.
- The limit(s) of the acceptance or rejection zone (i.e. the range of results), derived from the decision rule, which leads to acceptance or rejection when the measurement result is within the appropriate zone.

For example, a decision rule that is currently widely used is that a result implies non compliance with an upper limit if the measured value exceeds the limit by the expanded uncertainty. With this decision rule, then only case (i) in Figure 2.11 would imply non compliance. Similarly, for a decision rule that a result implies compliance only if it is below the limit by the expanded uncertainty, only case (iv) would imply compliance.

The use of uncertainty in conformity assessment is further discussed in **Chapter 6** of the thesis, where measurement uncertainty and precision data are used in conformity assessment of automotive fuel products.

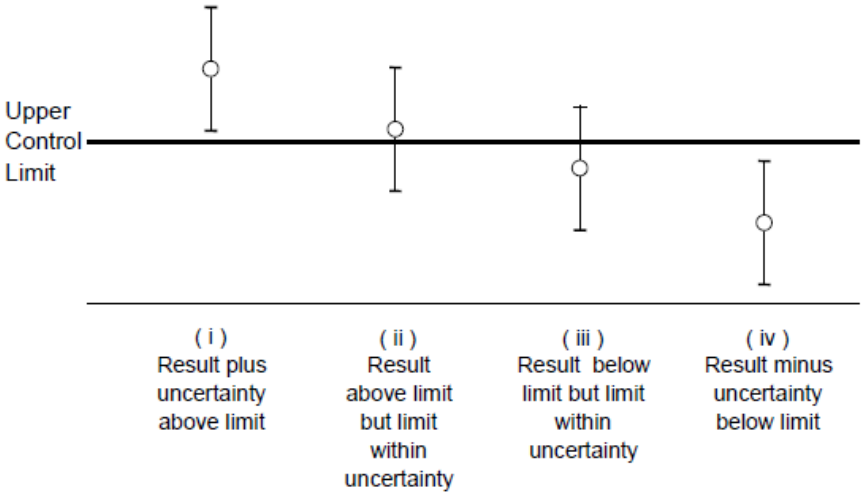


Figure 2.11 Uncertainty and compliance limits

2.9 Uncertainty of measurement models with two stages involving multiple output quantities – Uncertainty of a calibration curve

Until recently there has been a focus on the estimation of uncertainty of simple measurement models consisting of one stage and having a single output quantity (univariate scalar models). There are measurement models, however, where: (i) the output quantities from previous stages become input quantities for subsequent stages (multistage measurement models) and, (ii) have more than one output quantity, depending on a common set of input quantities. The principles for the uncertainty estimation described in the GUM and its Supplement describing the application of Monte Carlo method can be extended to the evaluation of uncertainties associated with vector estimates of multivariate output quantities. In 2011, JCGM has published GUM Supplement 2 [16] which provides guidance concerning the application of a generalized of GUM uncertainty framework and MCM in measurement models with any number of output quantities.

A typical multistage measurement model, that also includes a measurement stage with more than one output quantity is the calculation of the coefficients of a calibration curve, constructed by means of least squares regression and the subsequent use this calibration curve for the estimation of a quantity value. In analytical chemistry, linear regression is very often used in the construction of calibration functions required for techniques such as liquid or gas chromatography and atomic absorption spectrometry. Calibration establishes a relationship between the values of a standard (reference values) and the output quantities (response of

the instrument). Once this relationship (often assumed to be represented by a straight line) is established, the calibration model is used in reverse, that is, to predict a value from an instrument response [57-59]. Figure 2.12 illustrates the two-stage measurement model involving construction and use of a calibration function. As with most statistics, calibration curve coefficients (slope and intercept for linear calibration model) are only estimates based on a finite number of measurements, and therefore, their values are associated with uncertainties. This leads to an uncertainty of the predicted value as well.

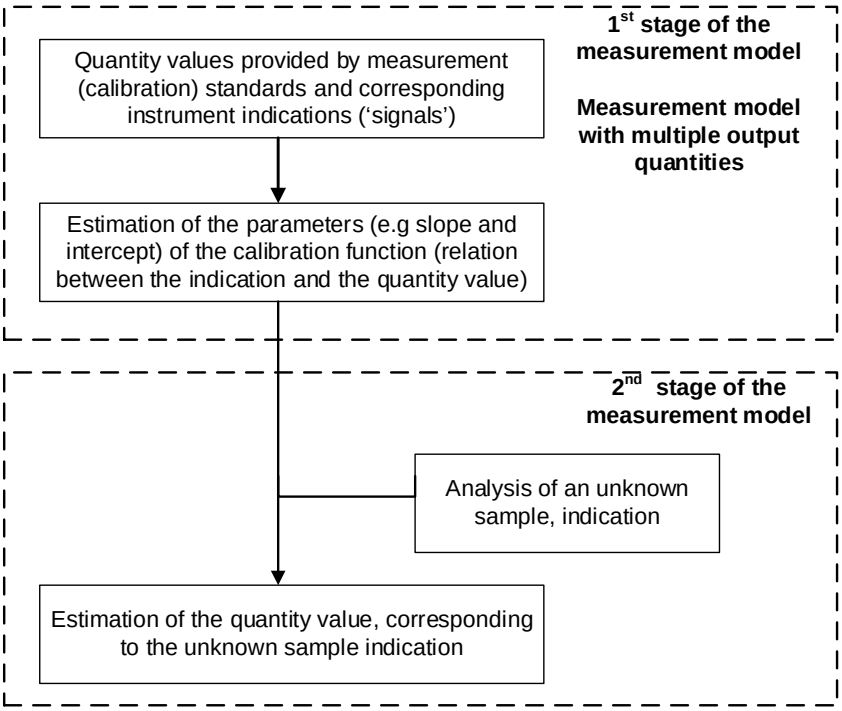


Figure 2.12 Two-stage measurement model involving the estimation of the parameters of a calibration function (1st stage of the measurement model having multiple output quantities) and the use of a calibration function for the estimation of a quantity value (2nd stage of the measurement model)

In **Chapter 5** the uncertainty of a calibration curve is estimated. The slope and the intercept of the calibration curve are treated as a vector output quantity characterized by a joint probability distribution and two types of coverage regions are estimated.

.

3. Measurement uncertainty arising from sampling

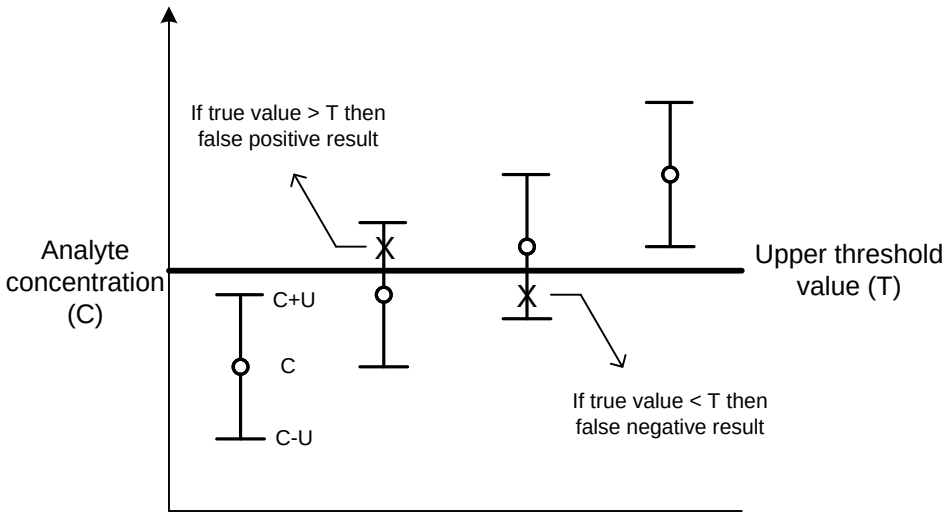
Sampling is an important part of any measurement process and is therefore recognized as an important contributor to the measurement uncertainty. A reliable estimation of the uncertainty arising from sampling of fuels leads to a better control of risks associated with decisions concerning whether product specifications are met or not. The present chapter of the thesis describes and compares the results of three empirical statistical methodologies (classical ANOVA, robust ANOVA and range statistics) using data from a balanced experimental design, which includes duplicate samples analyzed in duplicate from 104 sampling targets (petroleum retail stations). These methodologies are used for the estimation of the uncertainty arising from the manual sampling of fuel (automotive diesel) and the subsequent sulfur mass content determination. The results of the three methodologies statistically differ, with the expanded uncertainty of sampling being in the range of 0.34 – 0.40 mg kg^{-1}, while the relative expanded uncertainty lying in the range of 4.8 - 5.1%, depending on the methodology used. The estimation of robust ANOVA (sampling expanded uncertainty of 0.34 mg kg^{-1} or 4.8% in relative terms) is considered more reliable, because of the presence of outliers within the 104 datasets used for the calculations. Robust ANOVA, in contrast to classical ANOVA and range statistics, accommodates outlying values, lessening their effects on the produced estimates. The results of the work presented in this chapter also show that, in the case of manual sampling of fuels, the main contributor to the whole measurement uncertainty is the analytical measurement uncertainty, with the sampling uncertainty accounting only for the 29% of the total measurement uncertainty.

3.1 Introduction

The aim of sampling is to obtain a small portion of material (sample) from a selected system (sampling target) within a container which is representative of the material in that system [60,61]. The sampling process should ensure that the sample is an unbiased reflection of the composition of the sampling target [60]. Representative samples of petroleum and petroleum products are required for the determination of their chemical and physical properties, which are often used to establish compliance with commercial and regulatory specifications [61].

When a measurement result is compared with specified limits in order to make a decision relating to conformance or compliance, it is very likely that measurement uncertainty will have implications for the interpretation of the result. Not accounting for the uncertainty (deterministic approach) may lead to incorrect decisions i.e. false positive or false negative classifications that may have financial, health, environmental or other consequences [62,63]. Figure 3.1 shows the four situations apparent for a case of compliance with an upper limit and the conclusions drawn under the probabilistic and deterministic approach (assuming that an upper limit is set with no allowance for uncertainty). EURACHEM/ CITAC Guide "Use of uncertainty information in compliance assessment" [55] covers the above matters extensively.

Sampling becomes extremely important when considering the uncertainty of measurement. Until recently a "metrological gap" existed between analysts and end-users concerning the interpretation of measurement results and their associated uncertainties. Analysts concentrated on the analytical measurement process and estimated the uncertainty of the measurand of the sample received at the laboratory while the end user naturally interpreted the measurement result together with its uncertainty in order to characterize the sampling target as a whole [65,66]. Therefore, the end user needs to know a precise estimate of an uncertainty that includes the uncertainty caused by sampling i.e. the combined uncertainty from sampling and analysis [65,67,68]. Reliable estimations of the uncertainties of fuel sampling and analysis are important as they are associated with the application of legal requirements and the identification of events of cross contamination of incompatible fuels and fuel adulteration.



Compliant	Possibly Compliant	Probably Compliant	Non Compliant	*Probabilistic Classification*
Compliant	Compliant	Non Compliant	Non Compliant	*Deterministic Classification*
< 0.025	0.025-0.500	0.500-0.975	>0.975	*Probability of non compliance (C>T)*

Figure 3.1 Deterministic and probabilistic classification for compliance assessment against an upper limit. Adapted from [64].

Sampling uncertainty is defined as the part of the total measurement uncertainty attributable to sampling [62,66]. Principles and procedures for estimating the uncertainty of measurement arising from sampling are described in the Guide published by Eurachem and CITAC [69] as well as in the Nordtest handbook [66] which is intended for practical applications.

There are two broad approaches for the estimation of uncertainty, the modeling method and the empirical method [68,70,71]. The modeling approach which is consistent with ISO GUM [8] and is described as a "bottom up" approach [72], quantifies all sources of uncertainty individually, and then combines (propagates) them through a mathematical model. The implementation of the modeling approach reveals difficulties in establishing reliable estimates for the input variables of the model [70]. On the other hand the empirical approach, which is described as "top down" approach [72], uses replicated measurements in order to obtain a reliable

estimate of the uncertainty, without necessarily knowing any of the sources individually [69-71]. One of the most commonly used empirical methods is the duplicate method with a balanced experimental design. This method involves the formation of duplicate samples from the sampling targets by applying the same sampling protocol and duplicate analysis of samples under repeatability conditions. Appropriate statistical analysis applied to the resulting data leads to the estimation of the sampling uncertainty.

The aim of the work of this chapter is to present and compare three statistical approaches used for the estimation of the uncertainty caused by manual sampling of fuels from petroleum retail stations, utilizing the duplicate sampling method. Duplicate samples of automotive diesel from 104 petroleum retail stations (10.9% of the petroleum retail stations monitored for fuel quality purposes) were analyzed in duplicate for the determination of sulfur content according to ASTM D 5453 [73]. The sulfur mass content is one of the most critical parameters associated with automotive diesel specifications. The results of the measurements of the samples were analyzed using three statistical approaches, classical ANOVA, robust ANOVA and range statistics [66,74] and the sampling uncertainty under each approach was calculated. Sampling (and analytical) bias has been assumed to be zero in this study.

3.2 Sampling protocol and experimental design

A balanced nested experimental design was used. Duplicate samples were taken from 104 petroleum retail stations, which were selected at random and comprised the 10.9% of the 950 petroleum retail stations monitored by the laboratory. The scheme of sampling is shown in Figure 3.2. The duplicated samples were taken by repeating the same sampling protocol. The sampling protocol used was consistent with the standard method ASTM D 4057 [61] concerning the manual sampling of petroleum and petroleum products. Instructions were given to the samplers to introduce variations to the sampling process provided that they do not violate any requirement of the sampling protocol. These variations actually represent variations which may arise due to the random nature of the sampling process. All automotive diesel samples were maintained in special closed containers. During transport and storage samples were protected to prevent weathering or degradation from light, heat or other potential detrimental condition.

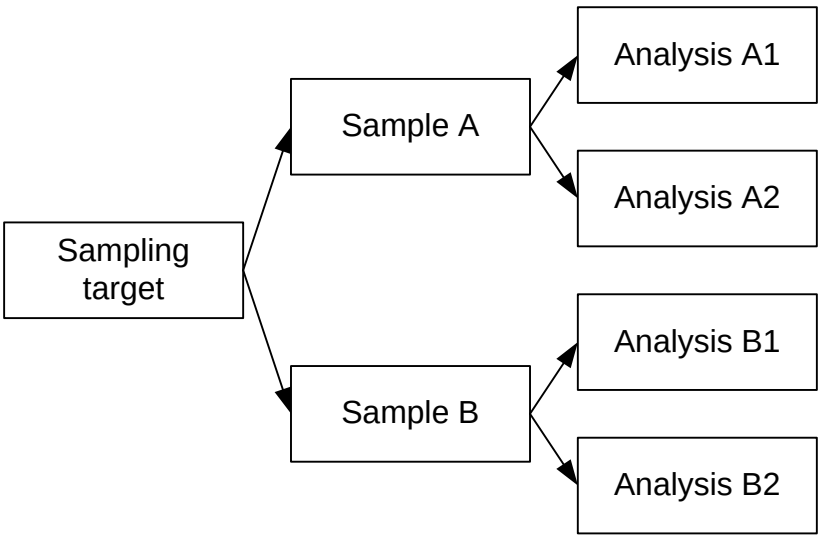


Figure 3.2 Balanced experimental design employed for the estimation of sampling uncertainty

3.3 Experimental work

The duplicated samples were analyzed in duplicate under repeatability conditions for sulful mass content determination. An ANTEK 9000S sulfur analyzer equipped with an automatic sampler was employed in the work presented in this chapter. This analyzer fully complies with ASTM D 5453 [73] and ISO 20846 [75]. Table 3.1 presents the operating conditions of the instrument. The sample sulfur content in ng μL^{-1} was calculated using a calibration curve. Six calibration standards (VHG Labs, Petrochemical Test Standard, Range set 1) with concentrations of 0, 1.00, 2.50, 5.01, 7.50 and 10.0 ng μL^{-1} were analyzed in triplicate. The instrument responses were recorded and a calibration curve with 18 points was constructed. The samples and the calibration standards were injected using a 10 µL syringe. Using density measurements the results were converted into mg kg^{-1} (ppm m/m). The majority of the automotive diesel sample measurement results were found to be under or close to 10 mg kg^{-1} which is the EU regulatory limit for sulfur mass content.

Table 3.1 Instrument parameters used for total sulfur determination in petroleum products

Parameter	Value
Volume injected (μL)	10
Syringe drive rate (μL s^{-1})	1
Furnace temperature (^{o}C)	1080
Furnace oxygen flowmeter setting (mL min^{-1})	470
Inlet oxygen flowmeter setting (mL min^{-1})	15
Inlet carrier (Argon) flowmeter setting (mL min^{-1})	150

3.4 Data analysis methods

3.4.1 Statistical model of the empirical approach and estimation of uncertainties

The statistical model describing the relationship between the measured and the true values of analyte concentration based on a single measurement, *x*, has the following form [69]:

$$x = X_{\text{true}} + \varepsilon_{\text{sampling}} + \varepsilon_{\text{analysis}} \quad (3.1)$$

where X_{true}, is the true average concentration of the analyte at the sampling target, $\varepsilon_{\text{sampling}}$, is the total error due to sampling, associated with a variance of $\sigma^2_{\text{sampling}}$ and $\varepsilon_{\text{analysis}}$, is the total analytical error, associated with a variance of $\sigma^2_{\text{analysis}}$. If the sources of variation are independent [76], then the total variance of the measurement, $\sigma^2_{\text{measurement}}$, for a single sampling target is given by:

$$\sigma^2_{\text{measurement}} = \sigma^2_{\text{sampling}} + \sigma^2_{\text{analysis}} \quad (3.2)$$

If statistical estimates, s^2, are used to approximate variances, σ^2, then Equation (3.2) becomes:

$$s^2_{\text{measurement}} = s^2_{\text{sampling}} + s^2_{\text{analysis}} \quad (3.3)$$

The components of measurement variation, the sampling and the analytical variance, which represent the corresponding uncertainties, can be separated and

estimated using appropriate statistical methods, such as the statistical approaches presented in this chapter.

In order to obtain a coverage interval (expanded uncertainty) corresponding to an approximately 95% coverage probability, standard deviations, *s*, are multiplied by a coverage factor of two. The expanded uncertainties for the measurement, sampling and analysis, $U_{\text{measurement}}$, U_{sampling} and U_{analysis}, respectively, are calculated using the following equations.

$$U_{\text{measurement}} = 2s_{\text{measurement}} \tag{3.4}$$

$$U_{\text{sampling}} = 2s_{\text{sampling}} \tag{3.5}$$

$$U_{\text{analysis}} = 2s_{\text{analysis}} \tag{3.6}$$

The expanded uncertainties can also be expressed relative to the reported value *x* (as a percentage), as relative expanded uncertainties *U*(%):

$$U_{measurement}(\%) = 100\frac{2s_{\text{measurement}}}{x}\% \tag{3.7}$$

$$U_{sampling}(\%) = 100\frac{2s_{\text{sampling}}}{x}\% \tag{3.8}$$

$$U_{analysis}(\%) = 100\frac{2s_{\text{analysis}}}{x}\% \tag{3.9}$$

3.4.2 Estimation of uncertainty using classical Analysis of Variance

Classical ANOVA (Analysis of Variance) is a statistical technique by which variations associated with different sources can be isolated and estimated [77]. The simplest type of classical ANOVA is the one-way classical ANOVA, which deals with one independent variable and one dependent variable. Classical ANOVA may be applied to the data produced by the implementation of the balanced experimental design in order to estimate the sampling uncertainty. Classical ANOVA estimations are based on the differences from the mean values, not on the range as in the approach of the

range statistics. Table 3.2 presents the one way classical ANOVA calculations used for the estimation of sampling and analytical uncertainty.

Table 3.2 Calculation of sampling and analysis uncertainty components by one-way classical ANOVA using data from measurements from a balanced experimental design with n targets (*i*=1,2,…,n), 2 samples (*j*=A, B) from each target and 2 analyses (*k*=1, 2) of each sample [66,70].

	Source of variation	
	Sampling	**Analytical**
Differences from the mean values (*D*)[a,b,c]	$D_{i(\overline{x})} = \left\lvert \overline{X}_i - \overline{x}_{iA} \right\rvert = \left\lvert \overline{X}_i - \overline{x}_{iB} \right\rvert$	$D_{iA(\overline{x})} = \left\lvert x_{iA1} - \overline{x}_{iA} \right\rvert = \left\lvert x_{iA2} - \overline{x}_{iA} \right\rvert$ $D_{iB(\overline{x})} = \left\lvert x_{iB1} - \overline{x}_{iB} \right\rvert = \left\lvert x_{iB2} - \overline{x}_{iB} \right\rvert$
Sum of Squares (*SS*)	$SS_{\text{sampling}} = 4\sum_i D^2_{i(\overline{x})}$	$SS_{analysis} = 2\sum_i \left(D^2_{iA(\overline{x})} - D^2_{iB(\overline{x})} \right)$
Degrees of freedom (*df*)	*2 n – n=n*	*2·2 n - 2n =2n*
Mean square (*MS*)	$MS_{\text{sampling}} = \frac{SS_{\text{sampling}}}{df_{\text{sampling}}}$	$MS_{\text{analysis}} = \frac{SS_{\text{analysis}}}{df_{\text{analysis}}}$
Variances (*V*)	$V_{\text{sampling}} = \frac{MS_{\text{sampling}} - MS_{\text{analysis}}}{2}$	$V_{\text{analysis}} = MS_{\text{analysis}}$
Uncertainty parameter (*s*)	$s_{\text{sampling}} = \sqrt{V_{\text{sampling}}}$	$s_{\text{analysis}} = \sqrt{V_{\text{analysis}}}$

[a] $\overline{X}_i$: mean value of target *i* (two samples – four analyses)

[b] $\overline{x}_{ij}$: mean value of the 2 analyses of sample *j* (A or B) of target *i*

[c] x_{ijk} : measured value from target *i*, sample *j* and split (analysis) *k* (1 or 2)

3.4.3 Estimation of uncertainty using robust Analysis of Variance

The term “robustness” in the statistics is used to describe methods designed to be insensitive to distributional assumptions (such as normality) and tolerate a certain amount of unusual observations (outliers). Robust statistics are characterized by the accommodation rather than the rejection of outlying values within a certain dataset, avoiding the risk – often associated with classical statistics - of skewing the statistics [71,78]. The valid application of classical ANOVA is based on three assumptions

summarized by Ramsey *et. al* [79] as: (i) the variances should be independent, (ii) each level of variance should be homogeneous, not varying systematically within one level and (iii) the distribution of errors within each level of variance should be approximately Gaussian. The application of robust ANOVA which utilizes robust statistics has been shown to be particularly appropriate for providing estimated of variances, in cases where the validity of classical ANOVA is doubtful [69]. Robust ANOVA uses robust estimates of the mean and standard deviation which are calculated by an iterative process [78,79]. In this process, also known as Huber's method, extreme values that exceed a certain distance (product of a constant *c* and the standard deviation) from the sample mean are downweighted or brought in. Actually these data are assigned a new value equal to that distance. A value of c=1.5 is widely accepted as optimal for datasets containing a small proportion of outliers (up to 10% of outliers). Initial values for sample mean and standard deviation estimates can be obtained by classical or robust statistics (e.g. median, mean absolute deviation). After the population has been modified, new sample mean and standard deviation are estimated. This process is repeated until the estimated values converge to the so called robust estimates. In the present work, robust ANOVA was implemented using a specifically written computer program called, Roban.exe, developed from the Analytical Methods Committee (AMC) in Great Britain [80].

3.4.4 Estimation of uncertainty using range statistics

Range statistics may also be used for calculating standard deviations by treating data produced by a balanced experimental design [64,81]. Range statistics, like classical ANOVA assume normal distribution and the calculations are done from the differences between duplicate measurements. The relation between standard deviation and differences requires the application of a certain factor depending on the replication chosen, e.g. 1.128 for duplicate, 1.693 for triplicate etc. [82]. Actually the variance of sampling is calculated indirectly as the difference of the variances of measurement and analysis. Table 3.3 presents the range statistics calculations used for the estimation of sampling and analytical uncertainty.

Table 3.3 Calculation of sampling and analysis uncertainty components by range statistics using data from measurements from measurements from a balanced experimental design with n targets (*i*=1,2,…,n), 2 samples (*j*=A, B) from each target and 2 analyses (*k*=1, 2) of each sample [64].

Parameter	Equation of calculation
Differences of duplicates of sample A ($D_{i\mathrm{A}}$)[a]	$D_{i\mathrm{A}} = \left\lvert x_{i\mathrm{A}1} - x_{i\mathrm{A}2} \right\rvert$
Differences of duplicates of sample B ($D_{i\mathrm{B}}$)[a]	$D_{i\mathrm{B}} = \left\lvert x_{i\mathrm{B}1} - x_{i\mathrm{B}2} \right\rvert$
Differences of the means of the two measurements (D_i)[b]	$D_i = \left\lvert \overline{x}_{i\mathrm{A}} - \overline{x}_{i\mathrm{B}} \right\rvert$
Mean range of measurement ($D_{\mathrm{measurement}}$)	$D_{\mathrm{measurement}} = \frac{\sum_i D_i}{\mathrm{n}}$
Mean range of analysis (D_{analysis})	$D_{\mathrm{analysis}} = \frac{1}{2}\left(\frac{\sum_i D_{i\mathrm{A}}}{\mathrm{n}} + \frac{\sum_i D_{i\mathrm{B}}}{\mathrm{n}} \right)$
Standard deviation of analysis (s_{analysis})	$s_{\mathrm{analysis}} = \frac{D_{\mathrm{analysis}}}{1.128}$
Standard deviation of measurement ($s_{\mathrm{measurement}}$)	$s_{\mathrm{measurement}} = \frac{D_{\mathrm{measurement}}}{1.128}$
Standard deviation of sampling (s_{sampling})	$s_{\mathrm{sampling}} = \sqrt{s^2_{\mathrm{measurement}} - \left(\frac{s_{\mathrm{analysis}}}{\sqrt{2}} \right)^2}$ The s_{analysis} is divided by a square root of 2 because the result of the analysis of each sample is the mean of two measurements – standard error of the mean.

[a] x_{ijk} : measured value from target *i*, sample *j* and split (analysis) *k* (1 or 2)

[b] $\overline{x}_{ij}$: mean value of the 2 analyses of sample *j* (A or B) of target *i*

3.5 Results and discussion

The main objective of the work presented in this chapter was to estimate the uncertainty components resulting from sampling of fuels from petroleum retail stations. The results of the evaluation of the measurement results (104 duplicate sample analyzed twice) using three statistical methodologies are presented in Table

3.4 and Figure 3.3. The expanded uncertainty of sampling is in the range of 0.34 – 0.40 mg kg^{-1}, while relative expanded uncertainty lies in the range of 4.8 - 5.1%, depending on the statistical methodology used.

Table 3.4 Results calculated using range statistics, classical ANOVA and robust ANOVA. All uncertainties *U* are estimated using a coverage of 2 which corresponds to an approximately 95% coverage probability.

	Range statistics	**Classical ANOVA**	**Robust ANOVA**
Mean (mg kg^{-1})	7.988	7.988	7.079
$s_{analysis}$ (mg kg^{-1})	0.205	0.378	0.265
$s_{sampling}$ (mg kg^{-1})	0.202	0.200	0.169
$s_{measurement}$ (mg kg^{-1})	0.288	0.427	0.314
$U_{analysis}$ (mg kg^{-1})	0.411	0.755	0.529
$U_{analysis}$ (%)	5.1	9.5	7.5
$U_{sampling}$ (mg kg^{-1})	0.404	0.401	0.337
$U_{sampling}$ (%)	5.1	5.0	4.8
$U_{measurement}$ (mg kg^{-1})	0.576	0.855	0.628
$U_{measurement}$ (%)	7.2	10.7	8.9
Analysis uncertainty contribution (%)	51	78	71
Sampling uncertainty contribution (%)	49	22	29

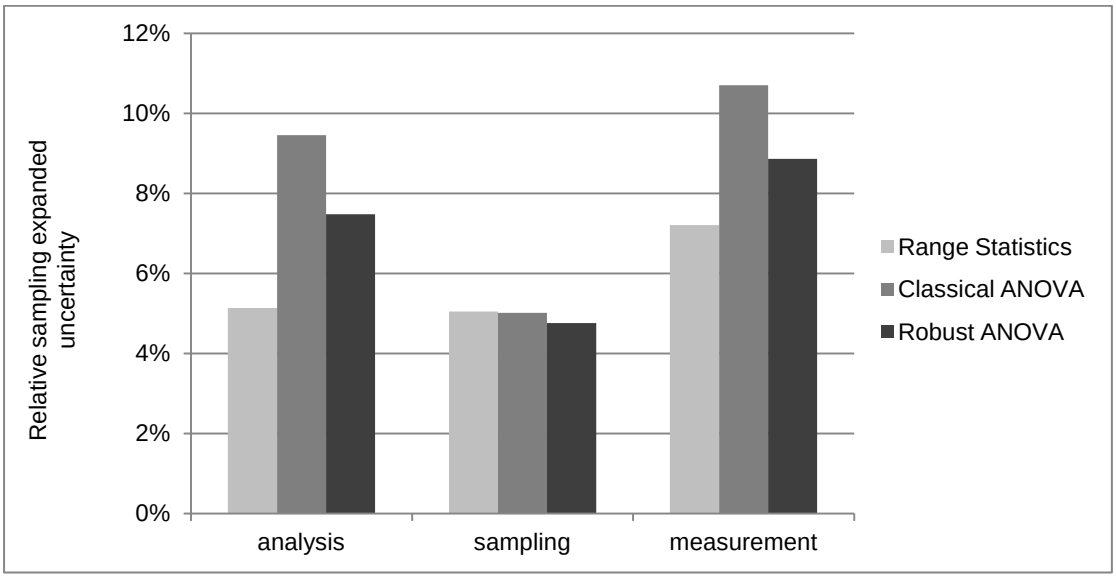


Figure 3.3 Graphical comparison of the relative uncertainty estimates obtained by range statistics, classical ANOVA and robust ANOVA. The relative uncertainties are estimated using a coverage of 2 which correspond to an approximately 95% coverage probability

The differences between the results of the 3 methodologies concerning sampling and analysis variances were evaluated using *F*-test (with 208 and 104 degrees of freedom for the variances of analysis and sampling, respectively). Table 3.5 presents the results of the *F*-tests performed. *F*-test results should be treated with caution as the test is used "out-of-its-scope" and possibly some of its assumptions (e.g. normality, independence) are violated. Nevertheless, it may provide a quick and gross evaluation of the significance of the differences of the estimated variances. The differences are statistically significant for all cases compared, with the exception of the sampling variances estimated by range statistics and classical ANOVA. Classical ANOVA and range statistics are typical tools of classical statistics that are strongly affected by the presence of outlying values. Indeed, as shown in the Boxplot diagram (produced by PASW 18 [83]) of Figure 3.4 created using the averages of the 104 datasets used for the calculations, there exist seven outliers and extreme values (between target outliers). These seven values, were also confirmed as outliers by applying the Grubbs test. Moreover, three datasets were identified as analytical or sampling outlier by applying the Cochran test. In total 9 out of 104 datasets (8.7%) were flagged by the outlier tests (one dataset was flagged by both Cochran and Grubbs test). Therefore, the results of robust ANOVA, which is insensitive to small number of outliers (less than 10%) can be considered as more reliable than the results of range statistics and classical ANOVA.

Table 3.5 Results of the *F*-test used for the comparison of the results of the three statistical methodologies

Methodologies compared	*F*-statistic for analysis variances (Probability)	*F*-statistic for sampling variances (Probability)
Range Statistics – Classical ANOVA	3.38* (<0.001)	1.01 (0.470)
Robust ANOVA Classical ANOVA	2.04* (0.001)	1.41* (0.039)
Range Statistics – Robust ANOVA	1.66* (0.001)	1.43* (0.034)

**Indicating significant difference (probability<0.05)*

It is obvious from the results of robust ANOVA that the measurement uncertainty is dominated by the analytical variance. In fact the analysis uncertainty accounts for the 71 % of the measurement uncertainty. This leaves "room" for an effective reduction of the measurement uncertainty. Uncertainty reduction may be accomplished by making more measurements and calculating their average, instead

of making a single measurement. Then the standard deviation of the mean gets smaller as the number of data increases leading to smaller random error uncertainty contributions. It has to be noted that multiple samplings and analyses required for collecting data used in uncertainty calculations are often associated with unreasonably high costs. These costs cost may be reduced, if the classical ANOVA and robust ANOVA are implemented using results from an unbalanced design [69,84], which is more economical as it requires 33% less analyses. Moreover, it has been shown that a minimum of eight duplicate samples leads to uncertainty estimates that are often fit for purpose [70].

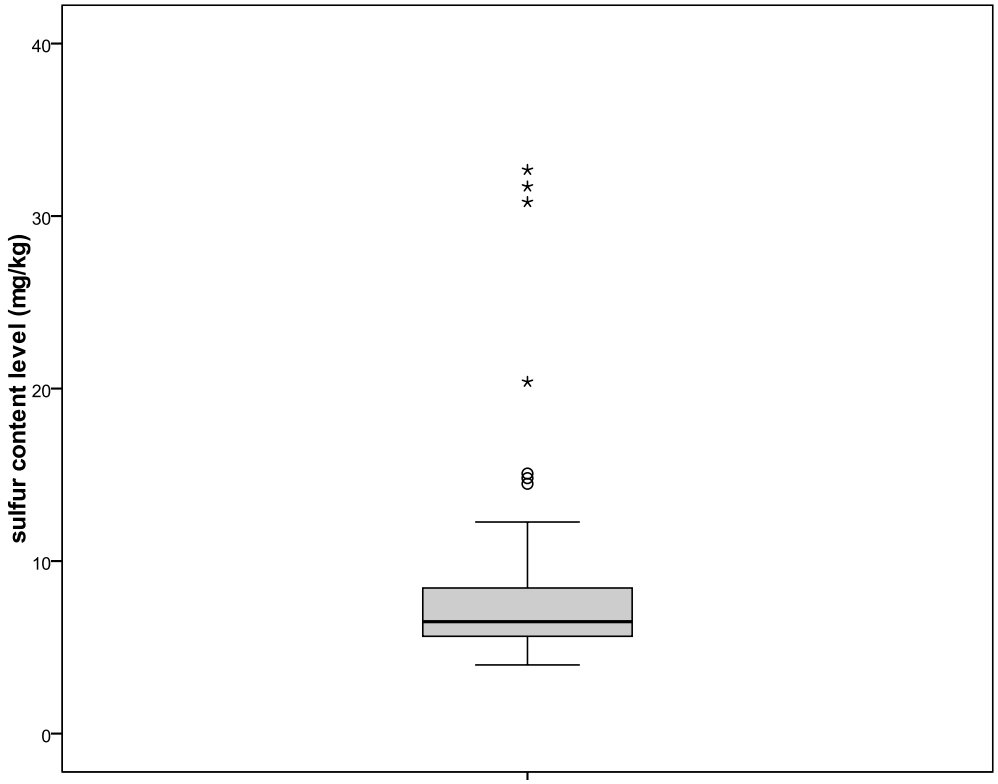


Figure 3.4 Box Plot of the average values of the 104 datasets (produced by PASW 18 [83]). The box length represents the interquantile range. Values which are more than three box lengths from either end of the box are extreme values are denoted by an asterisk. Values which are between one and a half and three box lengths from either end of the box are extreme values and are denoted by a circle.

For the purposes of the work presented, sampling bias has not been considered. It is widely accepted that it is difficult to establish the presence of bias in a sampling protocol because it is difficult to define a reference sampling method or a reference sampling target. This often leads to the characterization of the sampling methods as

"empirical" (by analogy to empirical analytical method). This means that the results are dependent on the method (and are usually reported with reference to the method), which by definition is associated with a zero bias [65]. It has to be noted though that there are methods available that may be used for the estimation and the inclusion of analytical or sampling bias in the sampling uncertainty estimates [69,85].

3.6 Conclusions

Three alternative statistical approaches for data analysis concerning the estimation of measurement uncertainty of manual sampling of fuel were described and compared in the work presented in this chapter. A balanced experimental design was used, which included duplicate samples of automotive diesel from 104 petroleum retail stations (10.9% of the petroleum retail stations monitored for fuel quality purposes) and duplicate analyses of these samples for sulfur mass content determination. The sulfur mass content is a critical fuel quality parameter associated with automotive diesel specifications and is often used for identifying fuel cross contamination or fuel adulteration incidents. The results were treated using classical ANOVA, robust ANOVA and range statistics. The three methodologies gave statistically different estimates with the expanded uncertainty of sampling being in the range of 0.34 – 0.40 mg kg^{-1}, while the relative expanded uncertainty lying in the range of 4.8 - 5.1%. The fact that the robust ANOVA leads to different results compared to the other two methodologies is an indication that the assumptions of classical ANOVA and range statistics are not justified. This fact is also confirmed by the presence of a small but unignorable number of outliers (8.7%) within the data used for the calculations. Therefore, robust ANOVA, which is not influenced by less than 10% outliers, is considered as the method providing the most reliable estimates for the sampling expanded uncertainty (0.34 mg kg^{-1} or 4.8% in relative terms). The results of robust ANOVA show that the analytical measurement uncertainty accounts for the 71%, while the sampling measurement uncertainty accounts only for the 29% of the total measurement uncertainty. Thus, minimizing or reducing repeatability errors may lead to substantial reduction of the total measurement uncertainty. Finally, it has to be noted that besides providing realistic estimates of uncertainty which allow end users make more informed decisions, the estimates of sampling and analytical variance may be also used for establishing an internal quality control method that will monitor the whole measurement procedure, including sampling, ensuring that it remains in statistical control.

4. Measurement uncertainty estimation of an analytical procedure

The Guide to Expression of Uncertainty in Measurement (GUM) approach and the adaptive Monte Carlo method (MCM) provide two alternative approaches for the propagation stage of the uncertainty estimation. These two approaches are implemented and compared concerning the 95% coverage interval estimation of the measurement of Gross Heat of Combustion (GHC) of an automotive diesel fuel by bomb calorimetry. The GUM approach, which assumes either a Gaussian or a *t*- distribution for the output quantity (GHC) gives half width intervals of 0.28 MJ kg^{-1} or 66 cal g^{-1}(Gaussian distribution) and 0.29 MJ kg^{-1} or 70 cal g^{-1} (*t*- distribution). On the other hand, MCM, which provides a reliable probability density function of GHC through numerical approximation, gives a half width interval of 0.32 MJ kg^{-1} or 75 cal g^{-1}. Thus, the GUM approach underestimates the calculated uncertainties and coverage intervals by up to 7 – 12%. The main reasons of these differences are the approximations and the assumptions introduced by GUM approach i.e. assumption for the GHC probability distribution and overestimation of effective degrees of freedom by the Welch-Satterwaite formula. Only if the GUM approach is combined with a Bayesian treatment of Type A uncertainties, the results are comparable with the MCM results. Moreover, the estimation and the use of sensitivity coefficients and uncertainty budget within GUM and MCM approaches are examined. Finally, it is shown that an initial estimate of measurement uncertainty may be obtained using the proficiency testing data.

4.1 Introduction

Bomb calorimetry is widely accepted as one of the most reliable and accurate methods for the determination of the heat of combustion of materials. Many procedures applied for the determination of the heat of combustion are based on standards issued by standardization organizations like ISO, CEN, DIN or ASTM. Bomb calorimetry is a test method which is directly related to many areas of science and engineering such as the production and utilization of solid and liquid fuels, the incineration of waste and refuse materials, the production of explosives, the formulation of energy balances or the implementation of thermodynamic studies [86].

Correct utilization of the result of any measurement requires some quantitative indication of its quality and reliability [87]. The result of heat of combustion determination is no exception. Fuels in particular, are characterized by their heat of combustion for technological, environmental and financial purposes. Heat of combustion of fuels represents the energy available and comprises an input parameter for planning and control of generators. It is also used for efficiency calculations and plays an important role in the estimation of greenhouse gas emission factors under the European Union Emissions Trading Scheme (EU ETS) [88].

The uncertainty of a measurement result is an index of its quality and its estimation is a key requirement of the international standard ISO/IEC 17025 [3], used for the accreditation of laboratories. As no measurement is exact, the true value of any measured quantity or any errors associated with the measurement cannot be known exactly [89]. The concept of uncertainty, which is relatively new in measurement history, reflects this lack of knowledge. The Guide to the Expression of Uncertainty in Measurement (GUM) [8], first published in 1993 describes an internationally agreed approach to the estimation and expression of measurement uncertainty, applicable to a wide range of measurements [89,90]. The ultimate goal is the estimation of a coverage interval for the measurand which according to International Vocabulary of Metrology (VIM) [9] is defined as "*an interval containing the set of true quantity values of a measurand with a stated probability, based on the information available*".

Working Group 1 (WG1) of the Joint Committee for Guides in Metrology (JCGM) which has the responsibility for maintaining the GUM, has decided to supplement the guide with a series of documents. Two of these supplementary that documents have been already approved are: the introduction to the "Guide to the expression of

uncertainty in measurement" and related documents [14] and the Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method (JCGM 101) [15]. The latter document describes a general numerical approach using Monte Carlo method (MCM) for carrying out the calculations required for the process of uncertainty estimation as a practical alternative to GUM.

In the present chapter of the thesis, the two methodologies for the estimation of uncertainty described in GUM and Supplement 1 to GUM (MCM) are compared. The two methodologies are used for the estimation of the uncertainty of the Gross Heat of Combustion (GHC) (or Higher Calorific Value) determination of a diesel fuel using a bomb calorimeter and following the standard method ASTM D240 [91]. ASTM D240 describes a widely accepted standard procedure for the determination of GHC in liquid fuel samples. It has to be noted though that ASTM D240 is used for routine technical measurements and not scientific ones that may have high precision requirements. MCM algorithm was implemented in the mathematical program MATLAB® [92]. As MCM performs random sampling from probability distributions, the quality of the results depends on the number of Monte Carlo trials made. Therefore, the implementation of an adaptive Monte Carlo procedure is recommended, that involves carrying out an increasing number of Monte Carlo trials until the various results of interest reach the desired degree of numerical accuracy [15]. This part of the thesis presents the results of MCM employing both fixed number of trials and number of trials selected adaptively. Furthermore, the concepts of uncertainty budget and sensitivity coefficients within the application of GUM and MCM are examined. The use of Bayesian statistics within GUM approach and the estimation of uncertainty using proficiency testing data are also discussed.

4.2 Main stages of uncertainty evaluation

Formulation, propagation and summarizing (statement of the complete measurement result) comprise the main stages of uncertainty evaluation (Figure 4.1) [15].

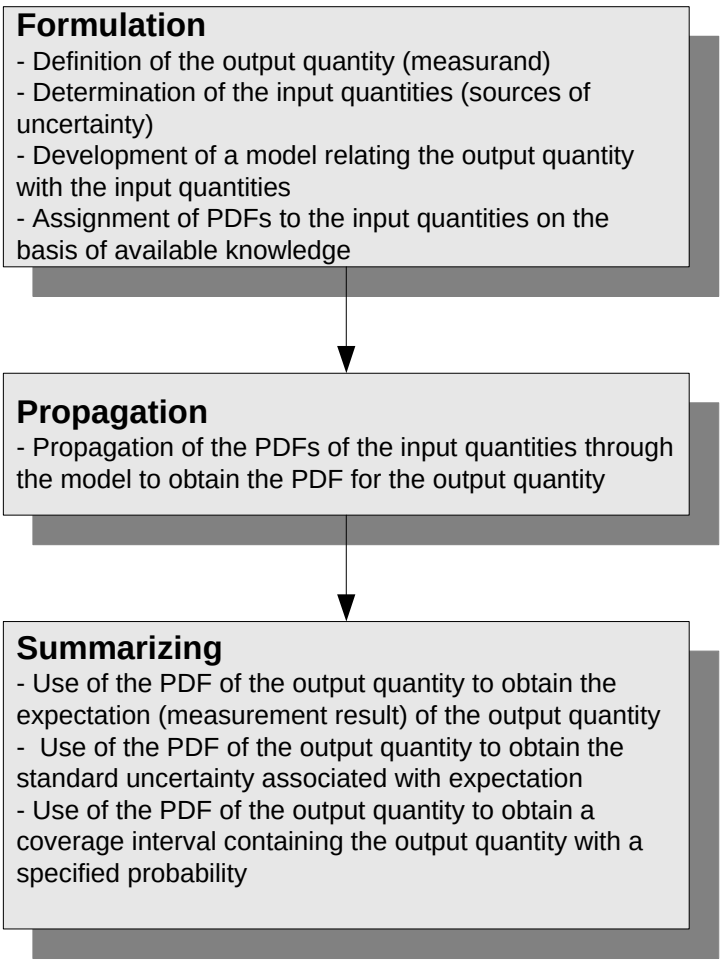


Figure 4.1 Flowchart depicting the main stages of measurement uncertainty evaluation [15].

The propagation stage of uncertainty evaluation can be implemented using one of the following approaches [15]:

- analytic methods, in which mathematical analysis is used to derive an algebraic form for the probability distribution for the output quantity. This approach is exact (introduces no approximations) but is applicable to relatively simple cases only.
- the GUM uncertainty framework, which involves the application of the law of propagation of uncertainty (based on a first-order Taylor series approximation) and the characterization of the output quantity by a Gaussian or a *t*-distribution,
- a Monte Carlo method (MCM), in which an approximation to the distribution function for the output quantity is established numerically by making random draws from the probability distributions of the input quantities, and evaluating the model at the resulting values. This approach provides a solution with a numerical accuracy that can be controlled.

4.3 Measurement procedure and identification of uncertainty sources

4.3.1 Experimental work

The GHC of 20 diesel fuel samples were determined by a Parr® 6200 calorimeter equipped with a Parr® 1108 Oxygen Bomb and a Parr® 6510 water handling system. The Parr® 6200 is the evolution of the two new automatic isoperibol calorimeters (1271 and 1281) which were introduced by Parr® in 1992-1994 [86]. The determinations were carried out under reproducibility conditions (measurements of 2 analysts over a period of approximately one month).

The calorimeter was calibrated by the combustion of one gram pellets of standardized benzoic acid purchased from Parr®, having combustion energy of 6318 ± 2.1 cal g^{-1} (26452.2 ± 9.0 J g^{-1}) and being traceable to SRM 39j Benzoic Acid, National Institute of Standards and Technology. According to Parr®, the value of the benzoic acid is referenced to the following standard bomb conditions: combustion reaction referred to 25 K, sample burned in a bomb of constant volume in pure oxygen at an initial absolute pressure of 3.0 MPa measured at 25 K, the number of grams of the sample burned is equal to three times the volume of the bomb in liters and the number of grams of water placed in bomb is equal to three times the volume of the bomb in liters. Seven pre-weighted benzoic acid samples of size around one gram were tested according to standard procedures described in ASTM D240 and Parr® Instruction Manual [93] and the energy equivalent of the calorimeter, W, calculated was 10093.57 J K^{-1} (2410.81 cal K^{-1}). The fuel samples were placed directly in a stainless steel capsule (Parr® 43AS, 2.5 cm of diameter and 1 cm of deep) and burned with oxygen at a pressure of 3.0 MPa. Table 4.1 presents the experimental data of the energy equivalent determination. Relevant quantities are explained in Table 4.3.

Concerning the 20 diesel fuel samples, approximately 0.6 g of sample were placed inside the sample cup in a pressure vessel (oxygen bomb) which was in contact with an ignition wire (Parr® 45C10 nickel - chromium fuse wire) connected to two electrodes. The bomb was then sealed, purged, and pressurized up to 3.0 MPa with pure oxygen. The sealed bomb was placed inside a 2 lt water bath which is inside a controlled temperature jacket. Once the jacket temperature stabilized and came within 0.5 K of 30 K, the sample was ignited and completely combusted.

Table 4.1 Experimental data of the energy equivalent determination

Benzoic acid sample	**Benzoic acid heat of combustion**	**Temperature rise**	**Fixed Corrections**		**Energy Equivalent**
***g'* (g)**	***Q* (cal/g)**	***t'* (K)**	**e_1' (cal)**	**e_2' (cal)**	***W* (cal/g)**
0.9589	6318	2.5625	10	50	2387.6411
0.9845	6318	2.6134	10	50	2403.0271
0.9971	6318	2.6361	10	50	2412.5328
0.9818	6318	2.5961	10	50	2412.4696
0.9957	6318	2.6315	10	50	2413.3888
0.9328	6318	2.4474	10	50	2432.5531
0.9845	6318	2.6009	10	50	2414.5761
				Average	2410.8077
				St. deviation	13.5194

Note: Raw data are presented with the units provided by the instrument (1 cal = 4.1868 J)

The GHC of the sample at room temperature was calculated from the temperature rise of the water bath, which is typically a few degrees Kelvin, the energy equivalent of the calorimeter and the mass of the sample. Three correction factors were also taken into account, to compensate for the heat of combustion of the ignition wire and the heat of formation of acids (nitric and sulphuric acid). For benzoic acid and fuel measurements, fixed corrections for fuse wire (e_2', e_3) and nitric acid (e_1', e_1) of 50 cal (209.34 J) and 10 cal (41.868 J), respectively, were used in accordance with Parr® Operating Instruction Manual. Table 4.2 presents the experimental data of the gross heat of combustion determinations.

Table 4.2 Experimental data of the gross heat of combustion determinations

Reproducibility Experiment

Mass sample	Temperature rise	Corrections			Energy Equivalent	GHC
g **(g)**	***t*** **(K)**	e_1 **(cal)**	e_2 **(cal)**	e_3 **(cal)**	***W*** **(cal/K)**	Q_g **(cal)**
0.6105	2.7725	10	0.0285	50	2410.8077	10850.00
0.6096	2.7550	10	0.0285	50	2410.8077	10797.00
0.5757	2.6085	10	0.0285	50	2410.8077	10819.25
0.5797	2.6164	10	0.0285	50	2410.8077	10777.35
0.5774	2.6113	10	0.0285	50	2410.8077	10798.80
0.5781	2.6142	10	0.0285	50	2410.8077	10797.76
0.5739	2.5921	10	0.0285	50	2410.8077	10783.99
0.5739	2.5899	10	0.0285	50	2410.8077	10774.89
0.5740	2.5951	10	0.0285	50	2410.8077	10794.97
0.5746	2.5993	10	0.0285	50	2410.8077	10801.31
0.5697	2.5835	10	0.0285	50	2410.8077	10827.39
0.5768	2.6160	10	0.0285	50	2410.8077	10830.02
0.5904	2.6686	10	0.0285	50	2410.8077	10795.28
0.5833	2.6282	10	0.0285	50	2410.8077	10759.57
0.5946	2.6844	10	0.0285	50	2410.8077	10783.04
0.5989	2.7002	10	0.0285	50	2410.8077	10769.00
0.5931	2.6794	10	0.0285	50	2410.8077	10790.02
0.6087	2.7515	10	0.0285	50	2410.8077	10799.04
0.5922	2.6795	10	0.0285	50	2410.8077	10806.56
0.6079	2.7391	10	0.0285	50	2410.8077	10763.83
					St. deviation	**22.93**

Sample Measurement

Mass sample	Temperature rise	Corrections			Energy Equivalent	GHC
g **(g)**	***t*** **(K)**	e_1 **(cal)**	e_2 **(cal)**	e_3 **(cal)**	***W*** **(cal/K)**	Q_g **(cal)**
0.5872	2.6541	10	0.0285	50	2410.8077	10794.44

Note: Raw data are presented with the units provided by the instrument (1 cal = 4.1868 J)

4.3.2 Measurement system modeling

The formulation of a model relating measured or influence quantities (the input quantities) and measurand (the output quantity) is the first critical stage of uncertainty evaluation [15].

GHC of a liquid non-volatile fuel (measurand) using bomb calorimetry is calculated using Equations (4.1)-(4.3). Various quantities of the equations are defined in Table 4.3.

$$Q_g = \frac{t \cdot W - e_1 - e_2 - e_3}{g} + \delta_{rep} \tag{4.1}$$

$$W = \frac{Q \cdot g' + e'_1 + e'_2}{t'} + \delta'_{rep} \tag{4.2}$$

$$e_2 = 58 \cdot S \cdot g \tag{4.3}$$

The automotive diesel fuel sample was tested under reproducibility conditions (twenty measurements, one measurement per day) in order to estimate uncertainty contributions arising from random variations of the influence parameters. The measurement result i.e. the estimate of the measurand was calculated 45194.17 J g^{-1} (10794.44 cal g^{-1}) (one measurement).

Table 4.3 List of quantities of Equations (4.1)-(4.3) (measurement model of GHC using bomb calorimetry).

Quantity	Units	Definition
Q_g	J g^{-1}	Gross heat of combustion at constant volume
t	K	Temperature rise
W	J K^{-1}	Energy equivalent of calorimeter
e_1	J	Correction for heat of formation of nitric acid
e_2	J	Correction for heat of formation of sulphuric acid
e_3	J	Correction for heat of combustion of fire wire
g	g	Weight of sample
δ_{rep}	J g^{-1}	Zero factor used to incorporate the reproducibility component of uncertainty into the uncertainty calculation of Q_g
t'	K	Temperature rise during benzoic acid combustion
Q	J g^{-1}	Heat of combustion of standard benzoic acid
e'_1	J	Correction for heat of formation of nitric acid
e'_2	J	Correction for heat of combustion of fire wire
g'	g	Weight of benzoic acid sample
δ'_{rep}	J K^{-1}	Zero factor used to incorporate the repeatability component of uncertainty into the estimation of W
S	%	Percentage of sulphur in sample

4.4 GUM uncertainty framework

4.4.1 Basic concepts of GUM uncertainty framework

The Guide to the Expression of Uncertainty in Measurement (GUM) introduces a method to unify the evaluation and the statement of measurement uncertainties [8,89]. The GUM uncertainty framework estimates the overall uncertainty by identifying, quantifying and combining all the sources of uncertainty associated with the measurement (bottom up approach) [27]. The method used for the evaluation of measurement uncertainty requires the formulation of one or more algebraic relationships between the measurand *Y* (output quantity) and the parameters or individual factors X_i that have an influence on the measurand (input quantities).

$$Y = f(X_1, X_2, ..., X_n) \tag{4.4}$$

Each input quantity is described by a probability density function (PDF). The expectation value of that PDF is a best estimate for the value of quantity, and the standard deviation of the PDF is the standard uncertainty associated with the estimate.

Therefore, an estimate, *y*, of the measurand *Y* is determined by substituting the best estimates $x_1, x_2, ..., x_n$ for $X_1, X_2, ..., X_n$ in Equation (4.4). Thus:

$$y = f(x_1, x_2, ..., x_n) \tag{4.5}$$

The standard uncertainty of *y*, $u(y)$, often called combined uncertainty, is obtained by appropriately combining the standard uncertainties of the input estimates $x_1, x_2, ..., x_n$, denoted by $u(x_1), u(x_2), ..., u(x_n)$. This is performed using the so called "law of propagation of uncertainty" which is based on a first-order Taylor series approximation of the Equation (4.4). Thus:

$$Y = f(X_1, X_2, ..., X_n) \approx y(x_1, x_2, ..., x_n) + \sum_{i=1}^{n} c_i X_i \tag{4.6}$$

where c_i is the so called sensitivity coefficient of x_i:

$$c_i = \frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial X_i}\bigg|_{x1,x2,\ldots,xn} \tag{4.7}$$

A change caused by the standard uncertainty $u(x_i)$ leads to a variation $u_i(y)$ of the estimate y (Equation 4.8), which is called the contribution to uncertainty $u(y)$ from $u(x_i)$:

$$u_i(y) = c_i u(x_i) \tag{4.8}$$

The law of propagation of uncertainty leads to the expression:

$$u^2(y) = \sum_i c_i^2 u^2(x_i) + 2\sum_{i<j} c_i c_j u(x_i) u(x_j) r(x_i, x_j) \tag{4.9}$$

where $r(x_i, x_j)$ is the coefficient of correlation (also called the correlation coefficient) between X_i and X_j for $i, j = 1,2,\ldots, n$. Standard uncertainties $u(x_i)$ may be evaluated either by observation of repeated experiments (Type A evaluation) or by other means (Type B evaluation). In a Type A evaluation the standard uncertainty $u(x_i)$ is calculated as the standard deviation $s(x_i)$ of the mean of m measurements. The degrees of freedom associated with Type A standard uncertainties based on m measurements are $v_i = m - 1$. In a Type B evaluation the standard uncertainty $u(x_i)$ is evaluated by scientific judgment based on information such as previous measurement data, experience with or general knowledge of materials and instruments involved, manufacturer's specifications, calibration data etc. When Type B uncertainties are used, it may be necessary to convert an interval into a standard uncertainty using information about the distribution of the value and the degrees of freedom. If no information is available, the distribution can be assumed to be rectangular. The degrees of freedom associated with Type B standard uncertainties may be taken to be infinite.

The GUM uncertainty framework makes the assumption that the probability distribution of the output quantity is a Gaussian distribution or a *t*-distribution. The expanded uncertainty U is obtained by multiplying the combined standard uncertainty $u(y)$ by a coverage factor k depending on the level of confidence required: for normal distribution, a value k=2 corresponds to an approximate confidence level (coverage probability) of 95%, and k=3 of 99.7 %.

In some cases the evaluation of a Type A standard uncertainty may not be based on a large number of readings or Type A uncertainty may be the dominant source of uncertainty. This could result in the coverage probability being significantly less than 95% if a coverage factor of $k = 2$ is used. In such cases a *t*-distribution is assumed and the degrees of freedom, v_{eff}, associated with $u(y)$ are calculated using the Welch-Satterthwaite formula [8,21]:

$$v_{eff} = \frac{u^4(y)}{\sum_i \frac{u_i^4(y)}{v_i}} \tag{4.10}$$

where v_i corresponds to the degrees of freedom of $u(x_i)$. The value of k, often denoted as k_p, where p is the coverage probability, will now give an expanded uncertainty, U_p, that maintains the coverage probability at approximately the required level p.

4.4.2 Application of GUM uncertainty framework

According to GUM approach [8], the definition of the measurand (in our case Q_g) is followed by the identification and quantification of all uncertainty sources. All the parameters involved in Equations (4.1) – (4.3), are regarded as uncertainty sources. It has to be noted that these uncertainty sources are associated with the application of ASTM D240 standard method which is intended for routine technical measurements. Other highly precision methods which may be used for scientific measurements are subject to different limitations and parameter determinations. In this case different or additional uncertainty components have to be taken into account. Zero factors δ_{rep} and δ'_{rep} have uncertainties that require Type A evaluation with 19 and 6 degrees of freedom, respectively. All other parameters have uncertainties based on Type B evaluations with infinite degrees of freedom. Table 4.4 presents the method of estimation as well as the mathematical formulas used for the estimation of the standard uncertainties of the various factors. Best estimates, half width intervals of the estimates and attributed probability distributions of each input quantity are presented in Table 4.5. For Type B evaluations, a certain divisor is selected in order to estimate standard uncertainty $u_i(x)$ from half width interval, according to the distribution of each input quantity. The uncertainty contribution $u_i(y)$ is then estimated as the product of standard uncertainty $u_i(x)$ multiplied by sensitivity coefficient, which is calculated as a partial derivative (Equations (4.7)-(4.8)). Results of these calculations are also presented in Table 4.5 (uncertainty budget for the measurand Q_g).

Table 4.4 Quantification of uncertainty sources and calculation of standard uncertainties.

i	Uncertainty source	Symbol	Method of estimation	Mathematical formula
1	**Reproducibility component of uncertainty estimation of Q_g**	δ_{rep}	The uncertainty due to random variations is estimated as a Type A uncertainty from twenty independent determinations of gross heat of combustion of a stable automotive diesel sample. The determinations were carried out under reproducibility conditions (measurements of 2 analysts over a period of approximately one month).	The standard uncertainty is calculated as the sample standard deviation, $s(Q_g)$, of the 20 measurements. $u(\delta_{rep}) = s(Q_g)$ Sensitivity coefficient: $c_1 = \frac{\partial Q_g}{\partial \delta_{rep}} = 1$
2	**Temperature rise (sample measurement)**	t	Temperature rise is calculated as the difference of two temperature measurements, each of which has an expanded uncertainty of 0.00005 K (caused by rounding)	Having in mind that combination of two identical rectangular distributions leads to a triangular distribution, the temperature rise standard uncertainty can be calculated from an expanded uncertainty of U_t= 0.0001 K assuming triangular distribution i.e. dividing the expanded uncertainty by square root of 6. $u(t) = \frac{U_t}{\sqrt{6}}$ Sensitivity coefficient: $c_2 = \frac{\partial Q_g}{t} = \frac{W}{g}$

i	Uncertainty source	Symbol	Method of estimation	Mathematical formula
3	**Correction for heat of formation of nitric acid (sample measurement)**	e_1	The fixed value of 41.868 J (10 cal) used for the correction for the heat of formation of nitric acid, has an expanded uncertainty U_{e1} of 12.560 J (3 cal) (based on manufacturer information).	Assuming rectangular distribution, the standard uncertainty of the correction for heat of formation of nitric acid is calculated by dividing the expanded uncertainty by square root of 3. $u(e_1) = \frac{U_{e_1}}{\sqrt{3}}$ Sensitivity coefficient: $c_3 = \frac{\partial Q_g}{\partial e_1} = \frac{1}{g}$
4	**Correction for heat of combustion of fire wire (sample measurement)**	e_3	The fixed value of 209.34 J (50 cal) used for the correction for the heat of combustion of fire wire has an expanded uncertainty U_{e3} of 20.934 J (5 cal) (based on manufacturer information).	Assuming rectangular distribution, the standard uncertainty of the correction for heat of combustion of fire wire is calculated by dividing the expanded uncertainty by square root of 3. $u(e_3) = \frac{U_{e_3}}{\sqrt{3}}$ Sensitivity coefficient: $c_4 = \frac{\partial Q_g}{\partial e_3} = \frac{1}{g}$

i	Uncertainty source	Symbol	Method of estimation	Mathematical formula
5	**Weight of sample**	*g*	The mass measurement uncertainty is obtained from data of the calibration certificate of the balance used. According to the certificate, the expanded uncertainty of a weighing result is given by the following equation: $U_G = 0.00002 + 8.8714 \cdot 10^{-6}\ G$ where U_G is the expanded uncertainty of the weighing result calculated for a confidence level of approximately 95% (k=2) and G is the mass of the sample weighted.	Substitution of the value of the sample mass, g , to the balance uncertainty equation gives the expanded uncertainty of the weight of the sample U_g, which divided by a coverage factor of 2 gives the corresponding standard uncertainty. $u(g) = \frac{U_g}{2}$ Sensitivity coefficient: $c_5 = \frac{\partial Q_g}{\partial g} = \frac{t \cdot W - e_1 - e_2}{g^2}$
6	**Percentage of sulphur in sample**	*S*	Sulphur content determination was carried out using an energy dispersive X-ray fluorescence (EDXRF) test method, which has an estimated maximum expanded uncertainty U_S (95%, k=2) of 6 ppm (0.0006%) at the level of interest (35 ppm).	The expanded uncertainty of the sulphur content determination divided by a coverage factor of 2, gives the corresponding standard uncertainty. $u(S) = \frac{U_S}{2}$ Sensitivity coefficient: $c_6 = \frac{\partial Q_g}{\partial S} = \frac{\partial Q_g}{\partial e_2}\frac{\partial e_2}{\partial S} = 58$

i	Uncertainty source	Symbol	Method of estimation	Mathematical formula
7	**Weight of benzoic acid sample**	*g'*	The mass measurement uncertainty is obtained from data of the calibration certificate of the balance used. According to the certificate, the expanded uncertainty of a weighing result is given by the following equation: $U_G = 0.00002 + 8.8714 \cdot 10^{-6}\ G$ where U_G is the expanded uncertainty of the weighing result calculated for a confidence level of approximately 95% (k=2) and G is the mass of the sample weighted.	Substitution of the value of benzoic acid sample mass, *g'*, gives the expanded uncertainty of the weight of the sample $U_{g'}$, which divided by a coverage factor of 2 gives the corresponding standard uncertainty. $u(g') = \frac{U_{g'}}{2}$ Sensitivity coefficient: $c_7 = \frac{\partial Q_g}{\partial g'} = \frac{\partial Q_g}{\partial W}\frac{\partial W}{\partial g'} = \frac{Q \cdot t}{t' \cdot g}$
8	**Heat of combustion of standard benzoic acid**	*Q*	The expanded uncertainty of the certified heat of combustion of benzoic acid, U_Q, is provided by the manufacturer's certificate 8.99 J g^{-1} (2.1 cal g^{-1})	The standard uncertainty of the certified heat of combustion of benzoic acid is calculated using uncertainty information from the manufacturer's certificate, assuming rectangular distribution, i.e. dividing the expanded uncertainty U_Q by square root of 3 $u(Q) = \frac{U_Q}{\sqrt{3}}$ Sensitivity coefficient: $c_8 = \frac{\partial Q_g}{\partial Q} = \frac{\partial Q_g}{\partial W}\frac{\partial W}{\partial Q} = \frac{g' \cdot t}{t' \cdot g}$

i	Uncertainty source	Symbol	Method of estimation	Mathematical formula
9	**Correction for heat of formation of nitric acid (benzoic acid measurement)**	e'_1	The fixed value of 41.868 J (10 cal) used for the correction for the heat of formation of nitric acid, has an expanded uncertainty $U_{e'1}$ of 12.560 J (3 cal) (based on manufacturer information).	Assuming rectangular distribution, the standard uncertainty of the correction for heat of formation of nitric acid is calculated by dividing the expanded uncertainty by square root of 3. $u(e'_1) = \frac{U_{e'1}}{\sqrt{3}}$ Sensitivity coefficient: $c_9 = \frac{\partial Q_g}{\partial e'_1} = \frac{\partial Q_g}{\partial W}\frac{\partial W}{\partial e'_1} = \frac{t}{t' \cdot g}$
10	**Correction for heat of combustion of fire wire (benzoic acid measurement)**	e'_2	The fixed value of 209.34 J (50 cal) used for the correction for the heat of combustion of fire wire has an expanded uncertainty $U_{e'2}$ of 20.934 J (5 cal) (based on manufacturer information).	Assuming rectangular distribution, the standard uncertainty of the correction for heat of combustion of fire wire is calculated by dividing the expanded uncertainty by square root of 3. $u(e'_2) = \frac{U_{e'2}}{\sqrt{3}}$ Sensitivity coefficient: $c_{10} = \frac{\partial Q_g}{\partial e'_2} = \frac{\partial Q_g}{\partial W}\frac{\partial W}{\partial e'_2} = \frac{t}{t' \cdot g}$

i	Uncertainty source	Symbol	Method of estimation	Mathematical formula
11	**Temperature rise (benzoic acid measurement)**	*t'*	Temperature rise is calculated as the difference of two temperature measurements, each of which has an expanded uncertainty of 0.00005 K (caused by rounding)	Having in mind that combination of two identical rectangular distributions leads to a triangular distribution, the temperature rise standard uncertainty can be calculated from an expanded uncertainty of $U_{t'}$= 0.0001 K assuming triangular distribution i.e. dividing the expanded uncertainty by square root of 6. $u(t') = \frac{U_{t'}}{\sqrt{6}}$ Sensitivity coefficient: $c_{11} = \frac{\partial Q_g}{\partial t'} = \frac{\partial Q_g}{\partial W}\frac{\partial W}{\partial t'} = \frac{W \cdot t}{t' \cdot g}$
12	**Repeatability component of uncertainty estimation of *W***	*δ'rep*	The uncertainty due to random variations is estimated as a Type A uncertainty from seven independent determinations of energy equivalent values using benzoic acid samples. The determinations were carried out under repeatability conditions.	The standard uncertainty is calculated as the standard deviation of the mean, *s*(*W*), of the 7 measurements. $u(\delta_{rep}') = \frac{s(W)}{\sqrt{7}}$ Sensitivity coefficient: $c_{12} = \frac{\partial Q_g}{\partial \delta'_{rep}} = \frac{\partial Q_g}{\partial W}\frac{\partial W}{\partial \delta'_{rep}} = \frac{t}{g}$

Table 4.5 Uncertainty budget for the measurand Q_g

i	Quantity	Estimate, x_i	Half width interval of the estimate	Probability distribution (divisor)	Standard uncertainty, $u_i(x)$	Sensitivity coefficient, c_i	Uncertainty contribution, $u_i(y)$	Degrees of freedom, v_i
1	δ_{rep}	0 J g^{-1} (0 cal g^{-1})	-	t- distribution	96.0 J g^{-1} (22.9 cal g^{-1})	1.000	96.0 J g^{-1} (22.9 cal g^{-1})	19
2	t	2.6541 K	0.00010 K	Triangular $(\sqrt{6})$	0.00004 K	17189.32 J g^{-1} K^{-1} (4105.6 cal g^{-1} K^{-1})	0.7 J g^{-1} (0.17 cal g^{-1})	∞
3	e_1	41.9 J (10 cal)	12.6 J (3 cal)	Rectangular $(\sqrt{3})$	7.3 J (1.7 cal)	1.703 g^{-1}	12.3 J g^{-1} (2.95 cal g^{-1})	∞
4	e_3	209.3 J (50 cal)	20.9 J (5 cal)	Rectangular $(\sqrt{3})$	12.1 J (2.9 cal)	1.703 g^{-1}	20.6 J g^{-1} (4.92 cal g^{-1})	∞
5	g	0.5872 g	0.00021 g	Normal (2)	0.0001 g	76966 J g^{-2} (18383 cal g^{-2})	7.9 J g^{-1} (1.88 cal g^{-1})	∞
6	S	0.0035 %	0.0006 %	Normal (2)	0.0003 %	58.00 J (% g)$^{-1}$ (13.85 % J (% g)$^{-1}$)	0.02 J g^{-1} (0.004 cal g^{-1})	∞
7	g'	0.9765 g	0.00021 g	Normal (2)	0.0001 g	46270.11 J g^{-2} (11051.43 cal g^{-2})	4.8 J g^{-1} (1.15 cal g^{-1})	∞
8	Q	26452.2 J g^{-1} (6318 cal g^{-1})	9.0 J g^{-1} (2.1 cal g^{-1})	Rectangular $(\sqrt{3})$	5.2 J g^{-1} (1.2 cal g^{-1})	1.71	8.9 J g^{-1} (2.12 cal g^{-1})	∞
9	e'_1	41.9 J (10 cal)	12.6 J (3 cal)	Rectangular $(\sqrt{3})$	7.3 J (1.7 cal)	1.75 g^{-1}	12.7 J g^{-1} (3.03 cal g^{-1})	∞
10	e'_2	209.3 J (50 cal)	20.9 J (5 cal)	Rectangular $(\sqrt{3})$	12.1 J (2.9 cal)	1.75 g^{-1}	21.1 J g^{-1} (5.05 cal g^{-1})	∞
11	t'	2.5840 K	0.00010 K	Triangular $(\sqrt{6})$	0.00004 K	17655.64 J g^{-1} K^{-1} (4216.98 cal g^{-1} K^{-1})	0.7 J g^{-1} (0.17 cal g^{-1})	∞
12	δ'_{rep}	0 J K^{-1}	-	*t*- distribution	21.4 J K^{-1} (5.1 cal K^{-1})	4.52 K g^{-1}	96.7 J g^{-1} (23.1cal g^{-1})	6
				Covariance term			152.5 J^2 g^{-2} (8.70 cal^2 g^{-2})	
				Combined standard uncertainty, $u(Q_g)$			141.7 J g^{-1} (33.84 cal g^{-1})	

Having quantified all the main uncertainty contributions, the standard uncertainty of the measurand $u(Q_g)$ (combined uncertainty) can be estimated using the law of propagation of uncertainties (Equation (4.9)). The mass of the sample, g, and the mass of the benzoic acid, g', are determined using the same analytical balance, therefore a significant correlation between these two input quantities exists. Assuming a correlation coefficient $r(g,g')=1$, a covariance term (second term of Equation (4.9)) of 152.49 $J^2\ g^{-2}$ (8.70 $cal^2\ g^{-2}$) is calculated. The correlation between the temperature rise of the sample combustion, t, and the temperature rise of the benzoic acid combustion, t', has negligible effect to the standard uncertainty estimation of Q_g (covariance term equal to 1 $J^2 g^{-2}$). The calculations give a standard uncertainty $u(Q_g)$ equal to 141.7 $J\ g^{-1}$ (33.84 $cal\ g^{-1}$). Figure 4.2 presents the combined uncertainty of the measurand together with the uncertainty contributions (last column of uncertainty budget presented in Table 4.5).

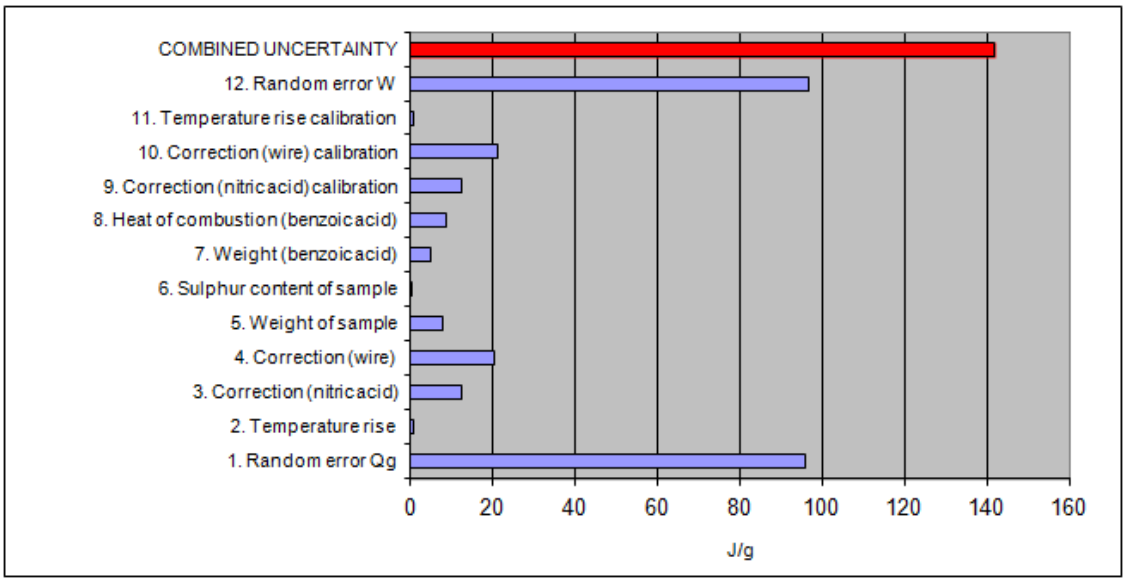


Figure 4.2 Combined uncertainty of the measurand and uncertainty contributions.

If the probability distribution of Q_g is taken as a Gaussian distribution the expanded uncertainty $U(Q_g)$ may be obtained by multiplying the standard uncertainty $u(Q_g)$ by a coverage factor k=1.96, which corresponds to a 95% level of confidence. The resulting expanded uncertainty is 277.7 $J\ g^{-1}$ (66.3 $cal\ g^{-1}$).

However, a closer look at the uncertainty budget (Table 4.5 and Figure 4.2) shows that the combined uncertainty is dominated by two components (δ_{rep} and δ'_{rep}) which are based on Type A uncertainty evaluation with a limited number of measurements. The reproducibility component of uncertainty, δ_{rep}, is calculated

from the standard deviation of twenty fuel sample measurements. The reproducibility component of uncertainty, δ_{rep}, is calculated from the standard deviation of twenty fuel sample measurements under reproducibility conditions, while the repeatability uncertainty component of the energy equivalent determination, δ'_{rep}, is calculated from the standard deviation of the mean of seven energy equivalent determinations using benzoic acid under reproducibility conditions. Therefore, according to GUM, it is reasonable to assume that the value of the coverage factor, *k*, should be based on a *t*-distribution rather than a Gaussian distribution. This value will give an expanded uncertainty that maintains the coverage probability at approximately the required level (95%).

Using as a coverage factor the two sided *t* tabulated value for the 95 % level of confidence and the effective degrees of freedom, v_{eff}, calculated using the Welch-Satterthwaite formula (Equation (4.10)) increases the resulting expanded uncertainty. The Welch-Satterthwaite formula gives 21.16 effective degrees of freedom which, for a coverage probability *p*=95%, correspond to a coverage factor k_p equal to 2.08. The resulting expanded uncertainty using k_p is 294.6 J g^{-1} (70.4 cal g^{-1}).

4.5 Monte Carlo method

4.5.1 Basic concepts of Monte Carlo method

Monte Carlo method (MCM) is actually any technique that uses pseudo-random numbers to solve a problem. It can be applied to problems which are characterized by the establishment of a formal equivalence between the desired result and the expected behaviour of a stochastic system. Monte Carlo methods have been successfully introduced to many scientific and engineering applications [94].

MCM comprises also an alternative method for calculating uncertainty. This method is described in the first supplement of GUM and involves no restrictions for valid application concerning the linearity of the measurement model and the applicability of Central Limit Theorem [87]. MCM carries out propagation of the probability density functions (PDF's) (not just the uncertainties like GUM) of the input quantities X_i through the measurement model *f* to provide the PDF of the measurand *Y.* In case of pairs of input quantities that are not independent (non zero covariance), multivariate distributions are used.

Figures 4.3 and 4.4 illustrate examples of the propagation of distributions (MCM) and the propagation of uncertainties (GUM), respectively, for a measurement model with three independent parameters. The major difference of the two methodologies is clear. Within GUM approach the standard uncertainty of the output is estimated and then distribution is assumed, while within MCM approach a distribution of the output is directly estimated. The GUM approach introduces approximations when estimating the coverage interval for the measurand. Information needed to determine the standard deviation of the estimate of the output quantity and a coverage interval (corresponding to a specified level of confidence) for the measurand is quite different. The mean and standard deviation can be determined knowing the distribution of the output, but the converse is not true.

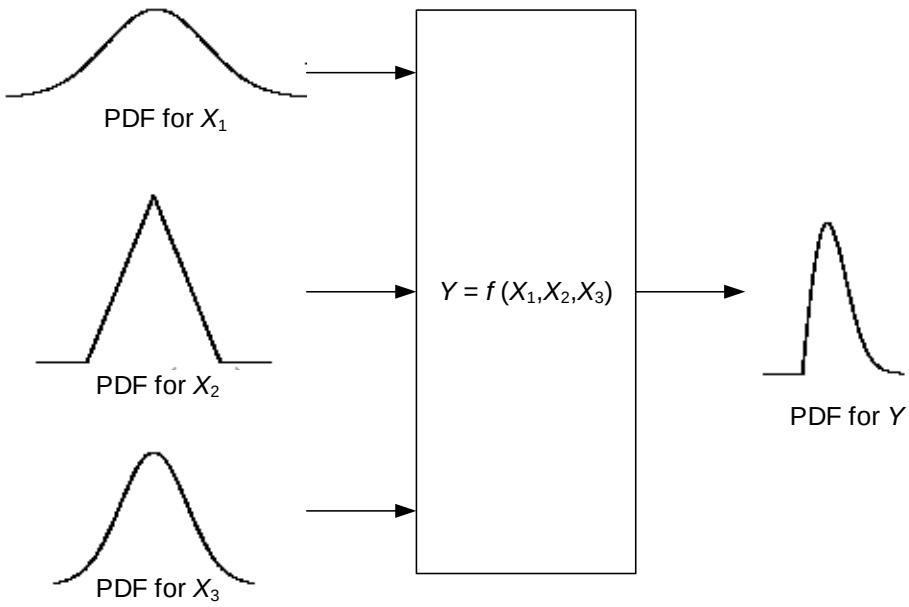


Figure 4.3 Propagation of distributions for three independent input quantities (MCM approach)

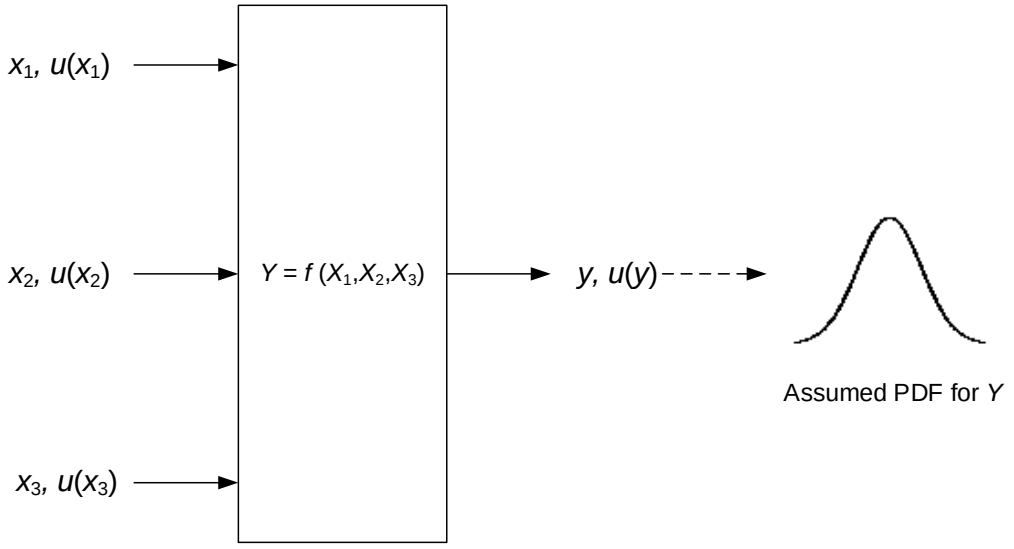


Figure 4.4 Propagation of uncertainties for three independent input quantities (GUM approach)

The Monte Carlo numerical simulation tends to require up to 10^6 trials for calculating a 95% coverage interval which is correct to one or two significant decimals digits. In fact, it is often more reliable to implement an adaptive Monte Carlo procedure which involves carrying out an increasing number of trials until the results have stabilized in a statistical sense [15]. Coverage intervals (either shortest or probabilistically equal) and other statistical information can then straightforwardly be produced from the PDF of the output quantity.

The estimate y of the output quantity Y is estimated by the average of the M MCM trials which produce M measurement model values (y_r , r = 1,…,M):

$$y = \frac{1}{M}\sum_{r=1}^{M} y_r \tag{4.11}$$

while the standard uncertainty $u(y)$ associated with y is estimated as the standard deviation of the M model values

$$u(y) = \sqrt{\frac{1}{M-1}\sum_{r=1}^{M}(y_r - y)^2} \tag{4.12}$$

The 95% coverage interval for the output can be obtained by the 2.5- and 97.5-percentiles of the distribution of MCM results (probabilistically equal coverage interval). The GUM Supplement 1 [15] provides guidance for an alternative coverage, the "shortest coverage interval". This interval is the shortest among all possible intervals for a distribution of MCM results where each interval has the same coverage probability (95%). When the distribution is symmetric, the two types of coverage intervals are identical [95].

One of the uses of MCM may also be the validation of the results of the GUM uncertainty framework. Since there are fewer restrictions for the validity of MCM compared to GUM uncertainty framework, it is recommended that both approaches are applied and the results compared. When adaptive MCM is used for validation purposes, a numerical tolerance $\delta' = \delta / 5$ is selected (δ denoting the numerical tolerance associated with standard uncertainty $u(y)$). In order to validate GUM results the absolute difference of the endpoints of GUM and MCM coverage intervals (lower and upper) are calculated. Then, if both differences (for lower and upper endpoints) are no larger than the numerical tolerance δ, then the GUM approach is considered validated [15].

4.5.2 Calculation algorithm of the adaptive Monte Carlo method

The calculation algorithm of the adaptive MCM has the following steps [15, 96]:

(i) Choose the desired coverage probability p for the interval to be obtained.

(ii) Choose the number of significant decimal digits n_{dig} for the uncertainty $u(y)$ (normally 1 or 2).

(iii) Choose the number M of trials to perform in each sequence of the application of the process.

$$M = \max\left(\frac{100}{1-p}, 10^4\right) \tag{4.13}$$

(iv) The variable h counts the number of MCM simulations. In order to carry out the first one, $h = 1$ is established.

(v) For each h sequence, M trials or evaluations of the model are carried out, which gives the values y_r ($r = 1,\ldots, M$) and the following estimated parameters:

- Average is taken as an estimate y of Y:

$$y^{(h)} = \frac{1}{M}\sum_{r=1}^{M} y_r \tag{4.14}$$

- Standard deviation is taken as the standard uncertainty $u(y)$ associated with y:

$$u(y^{(h)}) = \sqrt{\frac{1}{M-1}\sum_{r=1}^{M}(y_r - y)^2} \tag{4.15}$$

- Let q the integer part of $pM+1/2$. Sorting the values y_r (r = 1,…, M) in a non-decreasing order $y_{(r)}$ (r = 1,…, M), the probabilistically symmetric coverage interval for Y will be [$y^{(h)}_{low}$, $y^{(h)}_{high}$]. The interval extremes are $y^{(h)}_{low}$= $y_{(r)}$ and $y^{(h)}_{high}$= $y_{(r+q)}$, where r is the integer part of (M - q)/2 + 1/2. If the desired result is the shortest coverage interval, the value r* should be determined such that $y_{(r^*-q)} - y_{(r^*)} \leq y_{(r-q)} - y_{(r)}$ for each of the values r = 1,…, (M - q).

(vi) In order to analyze the variability of the parameters, more than one sequence is required, thus if h = 1, it should be increased by one unit and then return to step (v).

(vii) After every sequence, the average and standard deviation of these parameters must be calculated:

- For the estimate:

$$\hat{y} = y = \frac{1}{h}\sum_{i=1}^{h} y^{(i)} \tag{4.16}$$

$$s_{\hat{y}} = \sqrt{\frac{1}{h(h-1)}\sum_{i=1}^{h}(y^{(i)} - \hat{y})^2} \tag{4.17}$$

- For the standard uncertainty:

$$\hat{u}(y) = \frac{1}{h}\sum_{i=1}^{h} u(y^{(i)}) \tag{4.18}$$

$$s_{\hat{u}(y)} = \sqrt{\frac{1}{h(h-1)}\sum_{i=1}^{h}(u(y^{(i)}) - \hat{u}(y))^2} \tag{4.19}$$

- For the lowest extreme of the coverage interval:

$$\hat{y}_{\text{low}} = \frac{1}{h}\sum_{i=1}^{h} y^{(i)}_{\text{low}} \tag{4.20}$$

$$s_{y_{\text{low}}} = \sqrt{\frac{1}{h(h-1)}\sum_{i=1}^{h}(y_{\text{low}}^{(i)} - \hat{y}_{\text{low}})^2} \tag{4.21}$$

- For the highest extreme of the coverage interval:

$$\hat{y}_{\text{high}} = \frac{1}{h}\sum_{i=1}^{h} y_{\text{high}}^{(i)} \tag{4.22}$$

$$s_{\hat{y}_{\text{high}}} = \sqrt{\frac{1}{h(h-1)}\sum_{i=1}^{h}(y_{\text{high}}^{(i)} - \hat{y}_{\text{high}})^2} \tag{4.23}$$

(viii) In order to apply the stabilization criterion to the results, the numerical tolerance δ related to the standard uncertainty $u(y)$ must be calculated. The uncertainty $u(y)$ is calculated as in step (v), but using all the values $h \cdot M$ of the model. The tolerance is simply half of the last significant digit of the uncertainty. For its computer calculation, the uncertainty must be expressed in the form $u(y) = c \cdot 10^d$, where c is an n_{dig} decimal digit integer and d an integer. Then, the related numerical tolerance is:

$$\delta = \frac{1}{2}10^d \tag{4.24}$$

(ix) The stabilization criterion for the results establishes that if any of the values of $2s_{\hat{y}}$, $2s_{\hat{u}(y)}$, $2s_{\hat{y}_{\text{low}}}$ or $2s_{\hat{y}_{\text{high}}}$ are greater than δ, h should be increased by one unit and step (v) should be repeated.

(x) Once the stabilization criterion has been verified, all the values $h \cdot M$ of the model should be used to calculate y_{low} and y_{high} in the same way that it was done in step (v) for each sequence. The values of y and $u(y)$ have already been calculated in steps (vii) and (viii), respectively.

4.5.3 Programming in MATLAB®

MATLAB® [96] is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB®, one can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions may be reach a solution faster than with spreadsheets or traditional programming languages. A program in MATLAB®, as in any programming language, consists of a series of instructions that run sequentially, with the possibility

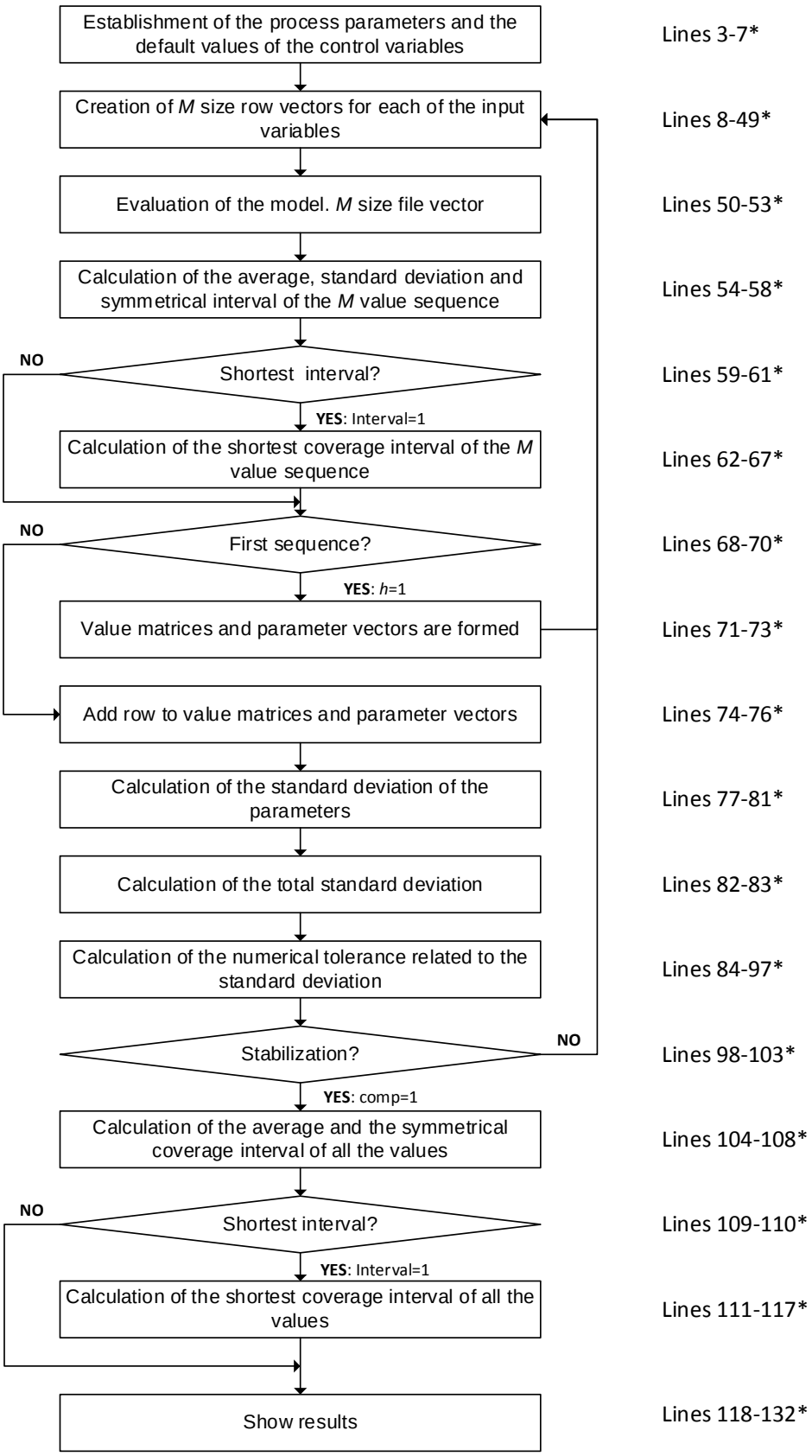
of containing certain instructions that control those which will run in each circumstance.

Implementation of the Monte Carlo method in MATLAB®, according to the algorithm described in 4.5.2 and summarized in the flow diagram shown in Figure 4.5, requires the user to enter a series of information: the functional relation of the measurement process, the probability distributions of the input quantities, their parameters, the required coverage probability, whether the coverage interval being sought is probabilistically symmetrical or the shortest and the number of significant decimal digits for the standard uncertainty.

4.5.4 Application of Monte Carlo method

The adaptive MCM performs the Monte Carlo simulation divided into h sequences of M' trials until twice the standard deviation associated with certain parameters (estimate, standard uncertainty, lowest extreme of the coverage interval, highest extreme of the coverage interval) is less than the chosen numerical tolerance, δ or δ', of the standard uncertainty, $u(y)$. In our case, M' was set to 10^4 for a coverage probability of 95%. The numerical tolerance δ was set to 0.5. The standard uncertainty $u(y)$ was calculated using all the values, M, of the model ($M=h\,M'$).

The algorithm of the adaptive MCM was implemented in MATLAB®. Table 4.6 details the program developed for the calculation of uncertainty. The MATLAB® code used, simulated samples from the distributions of the input quantities t, e_1, e_3, g, δ_{rep}, t', Q, e'_1, e'_2, g', δ'_{rep} and S of the measurement model. Univariate distributions were used for the input parameters, except for the cases which involved correlation (pairs $g - g'$ and $t - t'$) where multivariate distributions were used. These draws were then used to obtain the distributions of the interim quantities W and e_2, and the distribution of the measurand, Q_g. The whole process stabilized after 3.5 10^5 evaluations and gave the estimate Q_g = 45194.4 J g^{-1} (10794.5 cal g^{-1}) with associated standard uncertainty $u(Q_g)$ = 160.0 J g^{-1} (38.2 cal g^{-1}) and a 95% coverage interval [44877.8, 45510.2] J g^{-1} ([10718.9, 10869.9] cal g^{-1}). The PDF obtained by MCM is bell-shaped and symmetrical, which leads to almost identical shortest and probabilistically equal coverage intervals.



**MATLAB code presented in Table 4.6*

Figure 4.5 Flow diagram of the Adaptive Monte Carlo algorithm implemented in MATLAB®

Table 4.6 Implementation program of Adaptive Monte Carlo Method in MATLAB®

Line	MATLAB Code
1	`function y=MC_adapt_Qg`
2	`Tic`
3	`% ESTABLISHMENT OF PROCESS PARAMETERS`
4	`p=0.95; n_digit=2;`
5	`tol_divisor=1; % (set 1 for δ=0.5, set 5 for δ=0.1)`
6	`interval=0; %  (1 'shortest', other value 'symmetric')`
7	`M=max((100/1-p),10000); comp=0; h=0;`
8	`while comp==0  % Stabilization index (comp=1 when stabilized)`
9	`h=h+1; % hth application of MCM`
10	
11	`% MODEL INPUTS`
12	`S=normrnd(0.0035,0.0003,1,M);  % simulate sample of S`
13	
14	`% simulate sample of g and g' (correlated)`
15	`%covariance matrix (SIGMA)`
16	`L=[0.0001026^2 0.0001026*0.0001043; 0.0001026*0.0001043 0.0001043^2];`
17	`%mean matrix (MU)`
18	`K=[0.5872 0.9765];`
19	`% simulate g,g' from joint pdf`
20	`J=mvnrnd(K,L,M); % Mx2 matrix`
21	`g=J(:,1); % 1st column of Mx2 matrix`
22	`g=transpose(g);`
23	`g_i=J(:,2); % 2nd column of Mx2 matrix`
24	`g_i=transpose(g_i);`
25	
26	`e2=(58/4.1868)*S.*g;  % Obtain e2 draws`
27	
28	`drep_i=5.1099*random('t',6,1,M);  % simulate sample of drep'`
29	`[a,b]=triang_param(2.5840,0.00004082);  % simulate sample of t'`
30	`c=(a+b)/2;`
31	`t_i=trirnd(a,c,b,M);`
32	`[a,b]=unif_param(50,2.8867);  % simulate sample of e2'`
33	`e2_i=unifrnd(a,b,1,M);`
34	`[a,b]=unif_param(10,1.7320);  % simulate sample of e1'`
35	`e1_i=unifrnd(a,b,1,M);`
36	
37	`[a,b]=unif_param(6318,1.2401);  % simulate sample of Q`
38	`Q=unifrnd(a,b,1,M);`
39	
40	`W=((Q.*g_i+e1_i+e2_i)./t_i)+drep_i; % obtain W draws`
41	
42	`[a,b]=unif_param(50,2.8867);  % simulate sample of e3`
43	`e3=unifrnd(a,b,1,M);`
44	`[a,b]=unif_param(10,1.7320);  % simulate sample of e1`
45	`e1=unifrnd(a,b,1,M);`
46	`[a,b]=triang_param(2.6541,0.00004082);  % simulate sample of t`
47	`c=(a+b)/2;`
48	`t=trirnd(a,c,b,M);`
49	`drep=22.9347*random('t',19,1,M);  % simulate sample of drep`
50	`% MODEL EVALUATION`
51	`Qg=((W.*t-e1-e2-e3)./g)+drep;  % obtain Qg draws`
52	`y=Qg;`
53	`y=sort(y);`
54	`% CALCULATION OF AVERAGE, STANDARD DEVIATION AND PERCENTILES`
55	`y_mean=mean(y);`
56	`y_std=std(y);`
57	`q=round(p*M); r=round((M-q)/2);`
58	`y_low=y(r); y_high=y(r+q);`
59	`% CALCULATION OF THE SHORTEST INTERVAL`
60	`if interval==1`
61	`r=1;`
62	`    while r<=(M-q)`
63	`        if y(r+q)-y(r)<=y_high-y_low;`

Line	MATLAB Code
64	y_low=y(r); y_high=y(r+q);
65	End
66	r=r+1;
67	End
68	End
69	
70	if h==1 % when h=1 value matrices and parameter vectors are formed
71	Y=y;
72	Y_mean=y_mean; Y_std=y_std; Y_low=y_low; Y_high=y_high;
73	else % when h>1 a new row is added to matrices and vectors
74	Y=[Y;y];
75	Y_mean=[Y_mean;y_mean]; Y_std=[Y_std;y_std];
76	Y_low=[Y_low;y_low]; Y_high=[Y_high;y_high];
77	% CALCULATES STANDARD DEVIATION OF PARAMETERS
78	y_mean_std=std(Y_mean)/(h^0.5);
79	y_std_std=std(Y_std)/(h^0.5);
80	y_low_std=std(Y_low)/(h^0.5);
81	y_high_std=std(Y_high)/(h^0.5);
82	y_values=reshape(Y,1,h*M); % Transforms Y into 1xhM matrix
83	y_standard=std(y_values); % Calculates the TOTAL standard deviation
84	% CALCULATION OF NUMERICAL TOLERANCE REL ATED TO TOTAL STANDARD DEVIATION
85	a=y_standard; b=0;
86	if a>=10^n_digit
87	while a>=10^n_digit
88	a=a/10; b=b+1;
89	End
90	tol=(0.5/tol_divisor)*10^b;
91	elseif a<10^(n_digit-1)
92	while a<10^(n_digit-1)
93	a=a*10; b=b+1;
94	End
95	tol=(0.5/tol_divisor)*10^-b;
96	else tol=(0.5/tol_divisor);
97	End
98	% STABILIZATION CRITERION
99	if (2*y_mean_std<tol & 2*y_std_std<tol & 2*y_low_std<tol & 2*y_high_std<tol)
100	comp=1;
101	End
102	End
103	End
104	% CALCULATION OF THE AVERAGE AND THE SYMMETRICAL COVERAGE INTERVAL
105	y_values=sort(y_values);
106	y_mean=mean(y_values);
107	q=round(p*h*M); r=round((h*M-q)/2);
108	y_limit_low=y_values(r); y_limit_high=y_values(r+q);
109	if interval==1
110	r=1;
111	while r<=(h*M-q)
112	if y_values(r+q)-y_values(r)<=y_limit_high-y_limit_low;
113	y_limit_low=y_values(r); y_limit_high=y_values(r+q);
114	End
115	r=r+1;
116	End
117	End
118	% Show Results
119	Toc
120	trials=h*M;
121	y_mean, y_standard, y_limit_low, y_limit_high, trials
122	lower=min(y_values)-y_mean; upper=max(y_values)-y_mean;
123	xc=lower:(upper-lower)/499:upper;
124	y_values=y_values-y_mean;
125	y_mean=0;

Line	MATLAB Code
126	n=hist(y_values,xc);
127	bar(xc,n./(((upper-lower)/499)*h*M),1);
128	hold on;
129	Z=normpdf(y_values,y_mean,33.8395); % GUM Gaussian pdf
130	T=tpdf((y_values-y_mean)/ 33.8395,21.85)/ 33.8395; % GUM t-distribution pdf
131	plot(y_values,Z,'-.r',y_values,T,'-g')
132	end
133	function [a,b]=unif_param(m,v)
134	% A function to obtain the parameters of the uniform (rectangular)
135	% distribution given its mean and standard deviation
136	b=m+v*sqrt(12)/2;
137	a=2*m-b;
138	end
139	function [a,b]=triang_param(m,v)
140	% A function to obtain the parameters of the triangular
141	% distribution given its mean and standard deviation
142	b=m+v*sqrt(6);
143	a=2*m-b;
144	end
145	function X=trirnd(a,c,b,N)
146	% This function generates a vector of triangular distributed
147	% continuous random variable. By specifying minimum value(a),
148	% maximum value(b), mode(c), and number of variables to be generated(n),
149	% the function gives a vector of random variables as output (X).
150	X=zeros(1,N);
151	for i=1:N
152	%Assume a<X<c
153	z=rand;
154	if sqrt(z*(b-a)*(c-a))+a<c
155	X(i)=sqrt(z*(b-a)*(c-a))+a;
156	else
157	X(i)=b-sqrt((1-z)*(b-a)*(b-c));
158	end
159	end %for
160	end %function

The algorithm of adaptive MCM was also implemented with a numerical tolerance δ'=0.1 (one fifth of δ), in order to validate the results of the GUM approach. $5.65 \cdot 10^6$ MCM trials were required, while the results produced an estimate Q_g = 45194.1 J g^{-1} (10794.4 cal g^{-1}) with associated standard uncertainty $u(Q_g)$ =160.0 J g^{-1} (38.2 cal g^{-1}) and a 95% coverage interval [44878.5, 45509.8] J g^{-1} ([10719.0, 10869.8] cal g^{-1}). The PDF obtained by MCM is again bell-shaped and symmetrical, which leads to identical shortest coverage interval and the probabilistically equal shortest interval.

Furthermore, an algorithm of MCM was implemented using a fixed number of trials (M=10^6). The MATLAB code for the implementation of MCM using fixed number of trial is presented in Table 4.7. The results compared to adaptive show no significant differences. Final results in MJ kg^{-1} and cal g^{-1} are summarized in the Table 4.11 of Section 4.8 (Discussion).

Table 4.7 Implementation program of Monte Carlo Method in MATLAB using fixed number of trials

Line	MATLAB Code
1	`function [Qg]=MC_Qg_2_correlation(N)`
2	`tic`
3	`S=normrnd(0.0035,0.0003,1,N);  % simulate sample of S`
4	
5	`% simulate sample of g and g' (correlated)`
6	`%covariance matrix (SIGMA)`
7	`L=[0.0001026^2 0.0001026*0.0001043; 0.0001026*0.0001043 0.0001043^2];`
8	`%mean matrix (MU)`
9	`K=[0.5872 0.9765];`
10	`% simulate g,g' from joint pdf`
11	`J=mvnrnd(K,L,N); % Nx2 matrix`
12	`g=J(:,1); % 1st column of Nx2 matrix`
13	`g=transpose(g);`
14	`g_i=J(:,2); % 2nd column of Nx2 matrix`
15	`g_i=transpose(g_i);`
16	
17	`e2=(58/4.1868)*S.*g;  % Obtain e2 draws`
18	
19	`drep_i=5.1099*random('t',6,1,N);  % simulate sample of drep'`
20	`[a,b]=triang_param(2.5840,0.00004082);  % simulate sample of t'`
21	`c=(a+b)/2;`
22	`t_i=trirnd(a,c,b,N);`
23	`[a,b]=unif_param(50,2.8867);  % simulate sample of e2'`
24	`e2_i=unifrnd(a,b,1,N);`
25	`[a,b]=unif_param(10,1.7320);  % simulate sample of e1'`
26	`e1_i=unifrnd(a,b,1,N);`
27	
28	`[a,b]=unif_param(6318,1.2401);  % simulate sample of Q`
29	`Q=unifrnd(a,b,1,N);`
30	
31	`W=((Q.*g_i+e1_i+e2_i)./t_i)+drep_i; % obtain W draws`
32	
33	`[a,b]=unif_param(50,2.8867);  % simulate sample of e3`
34	`e3=unifrnd(a,b,1,N);`
35	`[a,b]=unif_param(10,1.7320);  % simulate sample of e1`
36	`e1=unifrnd(a,b,1,N);`
37	`[a,b]=triang_param(2.6541,0.00004082);  % simulate sample of t`
38	`c=(a+b)/2;`
39	`t=trirnd(a,c,b,N);`
40	`drep=22.9347*random('t',19,1,N);  % simulate sample of drep`
41	
42	`Qg=((W.*t-e1-e2-e3)./g)+drep;  % obtain Qg draws`
43	`Qg=sort(Qg);`
44	`toc`
45	`Gross_Heat=mean(Qg)`
46	`Standard_Uncertainty=std(Qg)`
47	`q=round(0.95*N); r=round((N-q)/2);`
48	`Qg_0_025=Qg(r) % Symmetric Interval (low end point)`
49	`Qg_0_975=Qg(r+q) % Symmetric Interval (upper end point)`
50	`% Shortest Interval`
51	`r=1; q=0.95*N; Qg_low=Qg_0_025; Qg_high=Qg_0_975;`
52	`while r<=(N-q)`
53	`    if Qg(r+q)-Qg(r)<=Qg_high-Qg_low;`
54	`        Qg_low=Qg(r); Qg_high=Qg(r+q);`
55	`    end`
56	`    r=r+1;`
57	`end`
58	`disp(['Shortest Interval: ', num2str(Qg_low), ' - ', num2str(Qg_high)]);`
59	
60	`% Qg=(Qg-Gross_Heat);`

Line	MATLAB Code
61	hist(Qg,100)
62	
63	end
64	
65	function [a,b]=unif_param(m,v)
66	% A function to obtain the parameters of the uniform (rectangular)
67	% distribution given its mean and standard deviation
68	b=m+v*sqrt(12)/2;
69	a=2*m-b;
70	end
71	function [a,b]=triang_param(m,v)
72	% A function to obtain the parameters of the triangular
73	% distribution given its mean and standard deviation
74	b=m+v*sqrt(6);
75	a=2*m-b;
76	end
77	function X=trirnd(a,c,b,N)
78	% This function generates a vector of triangular distributed
79	% continuous random variable. By specifying minimum value(a),
80	% maximum value(b), mode(c), and number of variables to be generated(n),
81	% the function gives a vector of random variables as output (X).
82	X=zeros(1,N);
83	for i=1:N
84	%Assume a<X<c
85	z=rand;
86	if sqrt(z*(b-a)*(c-a))+a<c
87	X(i)=sqrt(z*(b-a)*(c-a))+a;
88	else
89	X(i)=b-sqrt((1-z)*(b-a)*(b-c));
90	end
91	end %for
92	%hist(X,50); Remove this comment % to look at histogram of X
93	end %function

4.5.5 Sensitivity coefficients / uncertainty budget

The calculation of the uncertainty contributions $u_i(y)$ of the input quantities as the product of the standard uncertainty multiplied by the sensitivity coefficient (Equation (4.8)) is a necessary step when evaluating uncertainty using the GUM uncertainty framework. These uncertainty contributions listed in the uncertainty budget (Table 4.5) are also a valuable tool, revealing dominant uncertainty sources, i.e. the degree of the contribution of the uncertainty of each input quantity to the estimate of the uncertainty of the output quantity [97]. On the other hand, propagation of distributions and its implementation using MCM do not involve the calculation of sensitivity coefficient and consequently of uncertainty contributions, at any point. However, this "limitation" of MCM can be overcome, as, by simply holding all input quantities but one fixed at their best estimates, MCM provides the PDF for the output quantity for the model having just that input quantity as a variable. Then the sensitivity coefficient can be estimated as the ratio of the

standard deviation of the resulting model values, s_{Yi} and the standard uncertainty associated with the best estimate of the relevant input quantity, $u(x_i)$ [15].

$$c_i = \frac{s_{Yi}}{u(x_i)} \tag{4.25}$$

It has to be noted, though, that for a nonlinear model, sensitivity coefficients are in general approximate, the quality of the approximations being less reliable with increased standard deviations for the input PDF's. [97]. Table 4.8 presents the results of the application of MCM (10^6 trials) for the estimation of sensitivity coefficient of all input quantities.

Table 4.8 Results of the application of MCM for the estimation of sensitivity coefficients

Input quantity	Standard deviation of Q_g, s_{Yi} [1]	Standard uncertainty of input quantity $u(x_i)$	Sensitivity coefficient, c_i
δrep	101.5681302 J g^{-1}	101.5681302 J g^{-1}	1.00
t	0.7013953 J g^{-1}	0.0000408 K	17189.32 J g^{-1} K^{-1}
e_1	12.3497478 J g^{-1}	7.2517719 J	1.70 g^{-1}
e_3	20.5860067 J g^{-1}	12.0881031 J	1.70 g^{-1}
g	7.8898975 J g^{-1}	0.0001025 g	76965.89 J g^{-2}
S	0.0173783 J g^{-1}	0.0002996 %	58.00 % J g^{-1}
g'	4.8283460 J g^{-1}	0.0001044 g	46270.11
Q	8.8754566 J g^{-1}	5.1961267 J g^{-1}	1.71
e'_1	12.6819466 J g^{-1}	7.2501534 J	1.75 g^{-1}
e'_2	21.1396086 J g^{-1}	12.0853218 J	1.75 g^{-1}
t'	0.7208433 J g^{-1}	0.0000408 K	17655.64 J g^{-1} K^{-1}
$\delta rep'$	118.3137066 J g^{-1}	26.1760328 J K^{-1}	4.52 K g^{-1}
$\sqrt{\sum_i s_{Yi}^2}$	160.20 J g^{-1}		

[1] standard deviation of the values of a measurement model having just input quantity X_i as variable and all other input quantities fixed at their best estimates.

4.6 Uncertainty evaluated from Bayesian statistics

Bayesian methods allow the combination of information from measurements with prior information about the possible (or likely) distribution of values of the measurand. The approach combines a "prior" distribution with a likelihood (the distribution inferred from the measurement results alone) to obtain a "posterior distribution" which describes the distribution of values reasonably attributable to the measurand.

Under the classical statistics approach the Type A standard uncertainty associated with x_i is $u(x_i)= s(x_i)$. When the number m of measurements is small, the classical uncertainty is uncertain. The uncertainty in $s(x_i)$ arising from small m is a statistical uncertainty, which is accounted for by the degrees of freedom (GUM, section E.4.3).

Under the Bayesian approach a Type A standard uncertainty $u(x_i)$ is calculated as:

$$u_{Bayes}(x_i) = \sqrt{\frac{m-1}{m-3}} s(x_i) \tag{4.26}$$

where m is the number of independent measurements and $s(x_i)$ is the experimental standard deviation. The factor $\sqrt{(m-1)/(m-3)}$ built into the Bayesian uncertainty accounts for the statistical uncertainty that arises from a small number of measurements. It turns out that the estimates from a classical statistical analysis are either equal or approximately equal to the corresponding estimates from a Bayesian analysis with non-informative prior probability distributions [44].

The measurement model studied in this chapter includes, two components (δ_{rep} and δ'_{rep}) which are based on Type A uncertainty evaluation with a limited number of measurements. Table 4.9 presents the calculations of Type A standard uncertainties associated with these two components. The inclusion of the Bayesian statistics estimates in the ISO GUM uncertainty budget (Table 4.5) gives a combined standard uncertainty $u(Q_g)$ of 160.7 J·g^{-1} (38 cal g^{-1}). This leads to a 95% coverage interval [44.88 – 45.51] MJ kg^{-1} or [10719 – 10870] cal g^{-1}. These results are consistent with the results of the Monte Carlo method, presented in Table 4.11.

Table 4.9 Type A standard uncertainties estimations under classical and Bayesian statistics

Uncertainty Source	Symbol	Classical statistics approach	Bayesian statistics approach
Reproducibility component of uncertainty estimation of Q_g	δ_{rep}	The standard uncertainty is calculated as the sample standard deviation, $s(Q_g)$, of the 20 measurements. $u(\delta rep) = s(Q_g) =$ 96.0 J g^{-1} (22.9 cal g^{-1})	$u_{Bayes}(\delta rep) = \sqrt{\frac{20-1}{20-3}} s(Q_g) =$ 101.5 J g^{-1} (24.2 cal g^{-1})
Repeatability component of uncertainty estimation of W	δ'_{rep}	The standard uncertainty is calculated as the standard deviation of the mean, $s(W)$, of the 7 measurements. $u(\delta' rep) = \frac{s(W)}{\sqrt{7}} =$ 21.4 J K^{-1} (5.1 cal K^{-1})	$u_{Bayes}(\delta' rep) = \sqrt{\frac{7-1}{7-3}} \frac{s(W)}{\sqrt{7}} =$ 26.2 J K^{-1} (6.3 cal K^{-1})

4.7 Uncertainty evaluated from proficiency testing data

An empirical approach utilizing data from a proficiency testing scheme (PTS) was also used to estimate the measurement uncertainty of the test method. The reproducibility standard deviation between laboratories, s_R is normally given directly in reports from the PTS provider. These data may well be used by a laboratory (having performed satisfactorily in the comparisons) as a first estimation of the standard uncertainty of the analysed parameter, provided that the comparison covers all relevant uncertainty components and steps.

In order to estimate the uncertainty of the measurement of Gross Heat of Combustion (GHC) by bomb calorimetry, data from several rounds of a PTS were used (Table 4.10). The PTS provider is accredited according to ISO/IEC 17043 [5] and most of the participants used the standard method ASTM D240 [91] for the measurement. The precision data of this method are also used by the PTS provider for the evaluation of the performance of the participants.

Table 4.10 Proficiency testing data concerning the measurement of Gross Heat of Combustion (GHC) by bomb calorimetry.

Round, *i*	Year	Number of participating laboratories, l_i	Reproducibility standard deviation, $s_R^{(i)}$
1	2014	99	0.132
2	2013	85	0.157
3	2012	56	0.193
4	2011	42	0.115
5	2010	32	0.148
6	2009	34	0.182
7	2008	32	0.188
8	2006b	14	0.075
9	2006a	60	0.118

The standard uncertainty was estimated as the pooled reproducibility standard deviation between laboratories:

$$s_R^{pooled} = \sqrt{\frac{\sum_{i=1}^{z}\left[(l_i - 1)\left(s_R^{(i)}\right)^2\right]}{\sum_{i=1}^{z} l_i - z}} \tag{4.27}$$

Where $s_R^{(i)}$ is the reproducibility standard deviation of round *i*, l_i is the number of participating laboratories in round *i*, and *z* is the number of rounds. The application of the Equation 4.27 gave an s_R^{pooled} =0.15 MJ kg^{-1} (36 cal g^{-1}), which multiplied by a coverage factor *k*=1.96 leads to an 95% expanded uncertainty of 0.30 MJ kg^{-1} (71 cal g^{-1}). Using the experimental results of Section 4.3.1, the estimated 95% coverage interval is [44.90 – 45.49] MJ kg^{-1} or [10724 – 10865] cal g^{-1}.

4.8 Discussion

MCM and GUM results for all cases examined are summarized in Table 4.11. All of the GUM and MCM approaches gave almost identical results for the estimate of Q_g. On the other hand, appreciable differences exist regarding the estimated coverage intervals between MCM and GUM results. Figure 4.6 shows the assumed PDF's of

the GUM approaches (broken line curve for Gaussian distribution, dotted line curve for *t*-distribution), the scaled frequency distribution obtained by MCM in the form of an histogram as well as the endpoints of the probabilistically symmetric 95% coverage interval provided by these approaches. The inner pair of broken vertical lines indicates the probabilistically symmetric 95% coverage intervals provided by GUM when assuming a Gaussian distribution. The outer pair of continuous vertical lines indicates the probabilistically symmetric 95% coverage interval provided by MCM. The pair of dotted vertical lines indicates the probabilistically symmetric 95% coverage interval provided by GUM when assuming a *t*- distribution.

Table 4.11 Results of the application of GUM and MCM to the estimation of the uncertainty of Gross Heat of Combustion determination

Method	MCM trials, M	Time required [1]	Measurand estimate, Q_g	Standard uncertainty, $u(Q_g)$	95% Coverage interval	Number of significant digits	Numerical tolerance, δ
GUM - Gaussian distribution	-	-	45.19 MJ kg^{-1} (10794 cal g^{-1})	0.14 MJ kg^{-1} (34 cal g^{-1})	[44.92 – 45.47] MJ kg^{-1} [10728 – 10861] cal g^{-1}	-	-
GUM - *t*-Student distribution	-	-	45.19 MJ kg^{-1} (10794 cal g^{-1})	0.14 MJ kg^{-1} (34 cal g^{-1})	[44.90 – 45.49] MJ kg^{-1} [10724 – 10865] cal g^{-1}	-	-
MCM	10^6	1.04 sec	45.19 MJ kg^{-1} (10794 cal g^{-1})	0.16 MJ kg^{-1} (38 cal g^{-1})	[44.88 – 45.51] MJ kg^{-1} [10719 – 10870] cal g^{-1}	-	-
Adaptive MCM	$3.5 \cdot 10^5$	0.60 sec	45.19 MJ kg^{-1} (10794 cal g^{-1})	0.16 MJ kg^{-1} (38 cal g^{-1})	[44.88 – 45.51] MJ kg^{-1} [10719 – 10870] cal g^{-1}	2	0.005 MJ kg^{-1} (0.5 cal g^{-1})
Adaptive MCM	$5.7 \cdot 10^6$	39.9 sec	45.19 MJ kg^{-1} (10794 cal g^{-1})	0.16 MJ kg^{-1} (38 cal g^{-1})	[44.88 – 45.51] MJ kg^{-1} [10719 – 10870] cal g^{-1}	2	0.001 MJ kg^{-1} (0.1 cal g^{-1})

[1] *using a PC equipped with Intel® Core™ i3 M330, 2.13GHz, 4GB RAM*

The probabilistically symmetric 95% coverage interval provided by GUM when assuming a Gaussian distribution is shorter than that obtained by MCM. Using the Welch-Satterthwaite formula to calculate the effective degrees of freedom v_{eff}, and then assigning a *t*-distribution to the measurand leads to an increased width of coverage interval but still more optimistic than that provided by MCM.

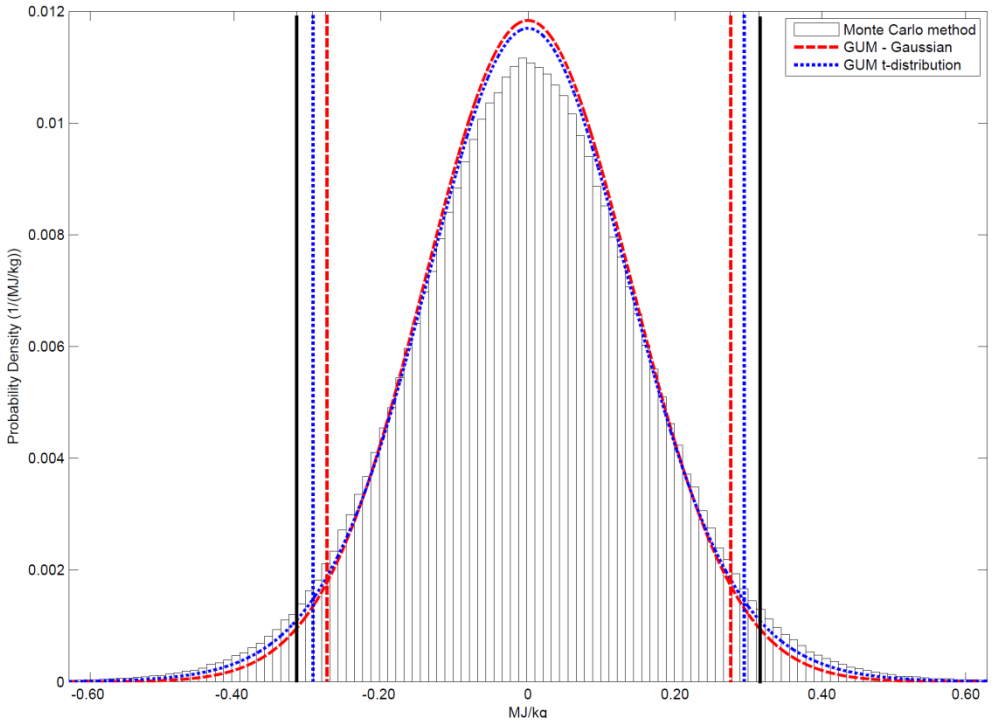


Figure 4.6 Probability Density Function of the measurand provided by GUM – Gaussian distribution, GUM – *t* - distribution and MCM

Specifically, the semi-width of the coverage interval (expanded uncertainty for 95% coverage probability) provided by GUM assuming Gaussian distribution (0.28 MJ kg^{-1} or 66.3 cal g^{-1}) is 12% smaller than the value (0.32 MJ kg^{-1} or 75.3 cal g^{-1}) returned by MCM. When a *t* - distribution is assumed, the semi-width of the coverage interval provided by GUM (0.29 MJ kg^{-1} or 70.4 cal g^{-1}) is underestimated by 7%, compared to the value returned by MCM. These differences may be attributed to the slight non linearity of the measurement model and to the approximations introduced by the use of the Welch-Satterthwaite formula. The Welch-Satterthwaite formula, as it has been pointed out by other work [98,99], does not provide an adequate approximation, as it overestimates the effective degrees of freedom, when there are dominant uncertainty contribution terms with relatively few degrees of freedom. This unavoidably leads to significantly underestimated coverage intervals. There is good agreement between estimates and coverage intervals determined by adaptive MCM and that provided by MCM with *a priori* chosen number of trials (M=10^6).

Furthermore, both GUM approaches are not validated using MCM, since the magnitudes of the coverage intervals endpoints differences d_{low}, d_{high} are larger than the numerical tolerance, δ, associated with the uncertainty of the measurand (Table 4.12).

Table 4.12 Results of the validation of GUM approaches

	GUM – Gaussian distribution	**GUM – *t*-distribution**	**GUM validated (δ=0.5)**
d_{low} = |**low endpoint (GUM) – low endpoint (MCM)**|	38.0 J g^{-1} (9.1 cal g^{-1})	21.0 J g^{-1} (5.0 cal g^{-1})	No
d_{high} = |**high endpoint (GUM) – high endpoint (MCM)**|	37.9 J g^{-1} (9.1 cal g^{-1})	21.0 J g^{-1} (5.0 cal g^{-1})	No

If the GUM approach is combined with a Bayesian treatment of Type A uncertainties, the results are comparable with the MCM results (see Section 4.5.4). Actually, as stated by Kacker and Jones [44] GUM contains an inconsistency, as it recommends classical (frequentist) statistics for evaluating the Type A components of uncertainty, but it interprets the combined uncertainty from a Bayesian viewpoint. In order to overcome this inconsistency, it is suggested that all Type A uncertainties should be evaluated through a Bayesian approach.

Sensitivity coefficients calculated by differentiating the function with respect to each input quantity (GUM approach) are identical to the sensitivity coefficients indirectly estimated by implementing MCM with one input quantity as a variable each time and all other input quantities fixed at their best estimates (Table 4.5, Table 4.8). The "non linear" sensitivity coefficients produced by MCM may be used in an uncertainty budget for the measurement. As the standard uncertainty of the output quantity is almost equal to the square root of the sum of the individual uncertainty contributions (Table 4.8), it is concluded that there are not contributions to $u(Q_g)$ that cannot be assigned to any individual input quantity and therefore uncertainty budget can be reliably used to detect dominant terms [97].

The uncertainty of GHC is dominated by the reproducibility component of uncertainty (calculated as the standard deviation of twenty fuel sample measurements under reproducibility conditions) and the repeatability uncertainty component of the energy equivalent determination (calculated as the standard deviation of the mean of seven energy equivalent determinations using benzoic acid). The magnitude of the estimated uncertainty is fit for purpose, especially considering the fact that there is no specification limit for the GHC of fuels. Therefore, due to practical reasons, it is acceptable for a single measurement to be

made, even though it is known that the method has imperfect repeatability and reproducibility. Nevertheless, a smaller uncertainty can be obtained by decreasing the uncertainty contributions of the dominant factors. By making more GHC measurements and calculating their average, instead of making a single measurement, a better estimate of the true value is obtained with smaller standard deviation of the mean (standard uncertainty of the reproducibility uncertainty component, δ_{rep}). The standard deviation of the mean gets smaller as the number of data increases. For example, if the result of the GHC is the average of four measurements instead of one, the standard uncertainty is halved (the standard deviation divided by square root of four) and the resulting expanded uncertainty is about 15 % smaller.

The measurement uncertainty derived from proficiency data (0.30 MJ kg^{-1} or 71 cal g^{-1}) is smaller than that obtained by the MCM (Section 4.5.5). Nevertheless, if the demand on uncertainty is low, then the reproducibility standard deviation between laboratories from a proficiency testing scheme can provide an uncertainty estimate that is fit for purpose. Generally, it has to be noted that such an empirical approach should be used with caution as it may give overestimated results depending on the quality of the participating laboratories – worst case scenario or underestimated results due to sample inhomogeneity or matrix variations [27].

4.9 Conclusions

Gross Heat of Combustion (GHC) is an important characteristic of fuels. It is determined using standard methods, like ASTM D 240 which describes a GHC determination procedure using a bomb calorimeter. The results of the determination are used for a wide range of technological, environmental and financial reasons, including the calculation of green house gas emissions under the European Union Emissions Trading Scheme (EU ETS). These calculations require accurate measurements accompanied with reliable uncertainty estimates.

Two alternative uncertainty estimation approaches were presented and compared, GUM and MCM (Supplement 1 to GUM). GUM approach propagates the uncertainties of the input quantities through a linearized model, while MCM provides an alternative approach in which the PDF's of the input quantities are propagated through the model. A coverage interval is obtained from the PDF of the output quantity without making a Gaussian or any other assumption concerning the

form of this distribution. Variations of both approaches were applied in the estimation of the uncertainty of the Gross Heat of Combustion (GHC) determination of a diesel fuel by bomb calorimetry. Appreciable differences were observed between the results of GUM and MCM approaches. These differences are inextricably related to the particular uncertainty sources used in this work and are aligned to the measurement parameters described to ASTM D240. If another method is used (e.g. more precise) the uncertainty sources may be different. The half width interval or expanded uncertainty results (at 95% level of confidence) obtained by GUM assuming Gaussian distribution, GUM assuming *t*-distribution and MCM were 0.28 MJ kg^{-1} (66 cal g^{-1}), 0.29 MJ kg^{-1} (70 cal g^{-1}) and 0.32 MJ kg^{-1} (75 cal g^{-1}), respectively. This means that GUM approaches are optimistic concerning the magnitude of the coverage interval of the GHC determination. This may be attributed to the slight non linearity of the measurement model and to the limitations involved in the use of the Welch-Satterthwaite formula (calculation of effective degrees of freedom of the assigned *t*-distribution) when dominant terms with relatively few degrees of freedom exist. Although, MCM does not need to calculate uncertainty contributions and sensitivity coefficients, these parameters which are necessary for the construction of an uncertainty budget and the identification of dominant uncertainty sources can be easily obtained by MCM. In our case the uncertainty is dominated by the reproducibility component of uncertainty and the repeatability uncertainty component of the energy equivalent determination. Only if the GUM approach is combined with a Bayesian treatment of Type A uncertainties the results are consistent with the results of the Monte Carlo method. Overall, MCM proves to be a more reliable tool for the estimation of the uncertainty of the determination of the GHC of diesel fuel, as it is not based on approximations or assumptions and it does not have the limitations of the GUM approach. The differences of uncertainty estimates using different methodologies might be significant for some uses of GHC results, especially when these results are equivalent to amounts of money, for example when selling and buying emission allowances under the EU ETS.

If the demand on uncertainty is low, it can be possible to directly use the reproducibility standard deviation between laboratories from a proficiency testing scheme as an approximation of the standard uncertainty. The expanded uncertainty (half width interval) obtained may be often fit for the intended use.

5. Estimation of the standard uncertainty of a calibration curve

The construction of a calibration curve using least square linear regression is common in many analytical measurements and it comprises an important uncertainty component of the whole analytical procedure uncertainty. In the present chapter of the thesis various methodologies are applied concerning the estimation of the standard uncertainty of a calibration curve used for the determination of sulfur mass concentration in fuels. The methodologies applied include the GUM uncertainty framework, the Kragten numerical method, the Monte Carlo method (MCM) as well as the approximate equation calculating the standard error of prediction. The standard uncertainty results obtained by all methodologies agree well (0.172 - 0.175 ng μL^{-1}). Aspects of inappropriate use of the approximate equation of the standard error of prediction, which leads to overestimation or underestimation of calculated uncertainty, are discussed. Moreover, the importance of the correlation between calibration curve parameters (slope and intercept) within GUM, MCM and Kragten approaches is examined.

5.1 Introduction

The calibration process is an essential stage of many chemical analyses which involve the prediction of an analyte concentration from a single instrumental response. Calibration establishes a relationship between the value of a standard (reference value) and the output quantity (response of the instrument). Once this relationship (often assumed to be represented by a straight line) is established, the calibration model is used in reverse, i.e. to predict a value from an instrument response [58-59].

As with most statistics, the slope and the intercept of a linear calibration model are only estimates based on a finite number of measurements and therefore their values are associated with uncertainties. This leads to an uncertainty of the predicted value as well. This uncertainty, also known as prediction interval, can be estimated using various methods and approximations. As calibration often comprises an important uncertainty component of the uncertainty of the whole analytical procedure, a reliable estimation of its uncertainty is crucial.

The sulfur content in petroleum products is determined using various spectrometric techniques, which almost always involve the construction and use of a calibration curve. The total sulfur content in petroleum products is a significant variable, as sulfur compounds are associated with problems of storage, processing, transportation and quality of fuel products. Sulfur also causes atmospheric pollution as the oxidation of sulfur compounds releases large quantities of SO_x into the atmosphere. For this reason European Union and many countries worldwide have legislation which specifies maximum sulfur content for automotive or other fuels. In order to use a result to decide whether it indicates compliance or non-compliance with a regulatory limit, it is necessary to take into account the measurement uncertainty [55]. Therefore, correct decisions have as prerequisite a reliable uncertainty estimation of the result.

In the present part of the thesis several methodologies are applied in order to estimate the uncertainty associated with the calibration curve used for the determination of sulfur mass concentration in fuels. In particular, the methodologies described in the Guide to the expression of uncertainty in measurement (GUM) [8] and the Supplement 1 to GUM (Monte Carlo method - MCM) [15] as well as the approximate numerical method Kragten [25,32] and the approximate equation giving the standard error of the estimate are applied and their results are compared. Moreover, as the estimation of the slope and the intercept of a calibration curve is

based on a measurement model of multiple outputs, the basic principles of Supplement 2 to GUM (Extension to any number of output quantities) [16] are applied, as well. All sulfur mass concentration measurements follow the standard test methods ASTM D5453 [100] and ISO 20846 [75] which describe the determination of the total sulfur in fuels and oil by ultraviolet fluorescence.

5.2 Experimental work

An ANTEK 9000S sulfur analyzer equipped with an automatic sampler was employed in this work. This analyzer fully complies with ASTM D 5453. Table 5.1 presents the operating conditions of the instrument. Six calibration standards (VHG Labs, Petrochemical Test Standard, Range set 1) with concentrations of 0, 1.00, 2.51, 5.00, 7.50 and 10.0 ng μL^{-1} were analyzed in triplicate. The instrument responses were recorded and a calibration curve with 18 points was constructed. Then a diesel sample with sulfur content near the EU regulatory limit (10 mg kg^{-1}) and a density of 0.830 g ml^{-1} was measured and its sulfur mass concentration in ng μL^{-1} was predicted using the calibration curve constructed. The sample and the calibration standards were injected using a 10 μL syringe. Data produced are presented in Table 5.2.

Table 5.1 Instrument parameters used for total sulfur determination in petroleum products

Parameter	Value
Volume injected (μL)	10
Syringe drive rate (μL s^{-1})	1
Furnace temperature (^{o}C)	1080
Furnace oxygen flowmeter setting (mL min^{-1})	470
Inlet oxygen flowmeter setting (mL min^{-1})	15
Inlet carrier (Argon) flowmeter setting (mL min^{-1})	150

Table 5.2 Calibration and sample measerement raw data

	Response (*y*)		
Concentration (*x*)	**1**	**2**	**3**
0.00	1464.9	1558.4	1594.8
1.00	2882.7	2851.6	2844.3
2.51	4889.8	4925.3	4917.3
5.00	6852.5	6934.5	7150.1
7.50	9344.5	9466.2	9265.8
10.00	13253.4	13567	13277.7
Unknown sample			
Response y_0	10603.0		

5.3 Linear calibration by least square regression

The establishment of a calibration function is a common procedure in many quantitative analyses. The main purpose is to obtain a function that allows one to calculate the value of a measurand - in chemistry often the concentration of the analyte - as a function of an instrumental signal. A set of standards are prepared containing a known amount of the analyte of interest and the instrument response of each standard is measured. Then a relationship is established between the instrument response and the analyte concentration. The most frequent type of calibration in chemical analyses is the indirect calibration, where the equipment gives a value (signal or instrument response), which has a different quantity from that of the standard (reference value) [59].

Calibration data (pairs of analyte concentrations, x_i, and instrument responses, y_i) are used as an input to the least square regression analysis which fits a predetermined measurement model to these data. The simplest measurement model is the one described by the linear function:

$$Y = b_0 + b_1 x \tag{5.1}$$

where Y is the instrument response (dependent variable), x is the analyte concentration (independent variable) and b_0 and b_1 are the coefficients of the model known as the intercept and slope, respectively.

The slope and the intercept are calculated from the calibration data (x_i, y_i) by the following equations:

$$b_1 = \frac{\sum_{i=1}^{n}\left[\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)\right]}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{5.2}$$

$$b_0 = \bar{y} - b_1\bar{x} \tag{5.3}$$

where n is the number of calibration data (pairs of x_i, y_i), $\bar{x}$ is the average concentration of all standards used and $\bar{y}$ is the average of all measured responses.

It has to be noted that least square regression process can validly be applied under the following conditions [57,58,101]:

a) The linear model holds for the data (i.e. the response does indeed vary linearly with concentration).
b) The errors in the independent variable are insignificant compared with those of the dependent variable.
c) These errors are normally distributed and independent of the values of the dependent variable (i.e. the data are homoscedastic).

Only when these conditions hold, the application of linear least square regression is expected to realize the best fit to the calibration data. Table 5.3 in Section 5.5.1 presents the information that a typical regression analysis produces.

When analyzing unknown test samples the relationship of Equation (5.1) is used in reverse in order to predict a value from an instrument response:

$$x_{\text{pred}} = \frac{y_0 - b_0}{b_1} \tag{5.4}$$

where y_0 is the instrument response of the unknown test sample and x_{pred} is the prediction (or estimation) of the measurement model concerning the true value of the measurand (e.g. analyte concentration)

5.4 Uncertainty estimation methods

The result obtained from a calibration curve is actually an estimate of the true value of the measurand. Therefore, it is associated with an uncertainty due to lack of fit of the linear model to the data, which leads to uncertainties (standard errors) of the parameters b_0 and b_1. This is actually a typical example of measurement model of multiple outputs, which involves correlated data. Some approaches used to estimate the uncertainty of the predicted value x_{pred} are presented in the following Sections.

5.4.1 Propagation of uncertainties – GUM

The Guide to the Expression of the Uncertainty in Measurement (GUM) provides an uncertainty estimation framework which uses as information the best estimates and

the standard uncertainties of the input quantities (b_0, b_1, y_0). This information is propagated, using the law of propagation of uncertainty through a first-order Taylor series approximation to the measurement model (Equation (5.4)) to provide the estimate of the output quantity x_{pred} and the standard uncertainty $u(x_{pred})$ associated with x_{pred} [14]. Thus, the standard uncertainty $u(x_{pred})$ of the value x_{pred} due to variations of b_0, b_1 and y_0 is given by the expression:

$$u(x_{\text{pred}}) = \sqrt{\left[c_0 u(b_0)\right]^2 + \left[c_1 u(b_1)\right]^2 + \left[c_2 u(y_0)\right]^2 + 2c_0 c_1 u(b_0) u(b_1) r(b_0, b_1)} \tag{5.5}$$

where c_0, c_1 and c_2 are the sensitivity coefficients of b_0, b_1 and y_0, $u(b_0)$, $u(b_1)$ and $u(y_0)$ are the standard uncertainties of b_0, b_1 and y_0, and $r(b_0,b_1)$ is the correlation coefficient of b_0 and b_1.

The sensitivity coefficients describe how the output estimate varies with changes in the values of input estimates and are calculated as partial derivatives:

$$c_0 = \frac{\partial x_{\text{pred}}}{\partial b_0} = -\frac{1}{b_1} \tag{5.6}$$

$$c_1 = \frac{\partial x_{\text{pred}}}{\partial b_1} = -\frac{y_0 - b_0}{{b_1}^2} \tag{5.7}$$

$$c_2 = \frac{\partial x_{\text{pred}}}{\partial y_0} = \frac{1}{b_1} \tag{5.8}$$

The standard uncertainties $u(b_0)$ and $u(b_1)$ are provided by the regression statistics as standard error of intercept, $SE(b_0)$, and slope, $SE(b_1)$, respectively. The standard uncertainty of the instrument response can be obtained from manufacturer specification (± 2% tolerance).

The correlation coefficient $r(b_0, b_1)$ is calculated by the following equation:

$$r(b_0, b_1) = -\frac{\sum_{i=1}^{n} x_i}{\sqrt{n \sum_{i=1}^{n} x_i^2}} \tag{5.9}$$

5.4.2 Kragten approach

A slightly modified approach for propagating uncertainties is provided by Kragten [25,32]. This approach uses an approximate numerical method of differentiation and is valid when the measurement model is linear or the standard uncertainties of the input quantities are small compared to the value of the respective input quantity. According to Kragten approach the sensitivity coefficients (partial derivatives) in Equation (5.5) can be approximated by:

$$c_0 = \frac{\partial x_{\text{pred}}}{\partial b_0} \approx \frac{x_{\text{pred}}(b_0 + u(b_0)) - x_{\text{pred}}(b_0)}{u(b_0)} \tag{5.10}$$

$$c_1 = \frac{\partial x_{\text{pred}}}{\partial b_1} \approx \frac{x_{\text{pred}}(b_1 + u(b_1)) - x_{\text{pred}}(b_1)}{u(b_1)} \tag{5.11}$$

$$c_2 = \frac{\partial x_{\text{pred}}}{\partial y_0} \approx \frac{x_{\text{pred}}(y_0 + u(y_0)) - x_{\text{pred}}(y_0)}{u(y_0)} \tag{5.12}$$

Multiplying c_0, c_1 and c_2 by $u(b_0)$, $u(b_1)$ and $u(y_0)$, respectively, in order to obtain the uncertainty contributions $u(x_{\text{pred}}, b_0)$, $u(x_{\text{pred}}, b_1)$ and $u(x_{\text{pred}}, y_0)$ gives:

$$u(x_{\text{pred}}, b_0) = c_0 u(b_0) = x_{\text{pred}}((b_0 + u(b_0)), b_1, y_0) - x_{\text{pred}}(b_0, b_1, y_o) \tag{5.13}$$

$$u(x_{\text{pred}}, b_1) = c_1 u(b_1) = x_{\text{pred}}(b_0, (b_1 + u(b_1)), y_0) - x_{\text{pred}}(b_0, b_1, y_0) \tag{5.14}$$

$$u(x_{\text{pred}}, y_0) = c_0(y_0) = x_{\text{pred}}(b_0, b_1, (y_0 + u(y_0))) - x_{\text{pred}}(b_0, b_1, y_0) \tag{5.15}$$

In this way each uncertainty contribution is calculated by a difference between two values of x_{pred}. The square root of the sum of the squares of the uncertainty contributions, provided that an extra term is added to account for the correlation between b_0 and b_1, gives an estimate of the standard uncertainty $u(x_{\text{pred}})$.

$$u(x_{\text{pred}}) = \sqrt{u(x_{\text{pred}}, b_0)^2 + u(x_{\text{pred}}, b_1)^2 + u(x_{\text{pred}}, y_0)^2 + 2u(x_{\text{pred}}, b_0)u(x_{\text{pred}}, b_1)r(b_0, b_1)} \tag{5.16}$$

Kragten approach can be easily implemented using a spreadsheet software. EURACHEM/ CITAC Guide "Quantifying Uncertainty in Analytical Measurement" [25] provides a detailed description of the actions required for the set up of a spreadsheet.

5.4.3 Propagation of distributions – Monte Carlo method

Monte Carlo method (MCM) comprises an alternative method for calculating uncertainty. This method is described in the first supplement of GUM [15] and involves no restrictions for valid application concerning the linearity of the measurement model and the applicability of Central Limit Theorem [87]. The MCM actually combines and propagates distributions rather than propagating uncertainties as in the GUM uncertainty framework. Uncertainty estimation is based on a probabilistic approach that combines the whole distribution of the input parameters and is not just based on their best estimates and standard uncertainties. It has to be noted though that MCM requires prior knowledge concerning the type of the probability distribution assigned to each input parameter.

The MCM consists of simulating draws from the distribution of the output quantity (x_{pred}), based on simulated draws from the distributions of the input quantities (b_0, b_1, y_0). In the present case two input quantities (b_0, b_1) are not independent (non zero covariance) and therefore multivariate distribution is used for the draws. The Monte Carlo numerical simulation tends to require up to 10^6 trials for calculating a standard uncertainty or a coverage interval which is correct to one or two significant decimals digits.

The estimate of the output quantity x_{pred} is estimated by the average of the M MCM trials which produce M measurement model values ($x_{\text{pred}}^{(k)}$, k = 1,…,M):

$$x_{\text{pred}} = \frac{1}{M}\sum_{k=1}^{M} x_{\text{pred}}^{(k)} \tag{5.17}$$

while the standard uncertainty $u(x_{pred})$ associated with x_{pred} is estimated as the standard deviation of the M model values:

$$u(x_{\text{pred}}) = \sqrt{\frac{1}{M-1}\sum_{k=1}^{M}(x_{pred}^{(k)} - x_{pred})^2} \tag{5.18}$$

5.4.4 Standard error of prediction

The estimation of the uncertainty of a result from a linear calibration is very often approximated through the so called error of prediction, $s(x_{\text{pred}})$ which is given by the following equation:

$$s(x_{\text{pred}}) = \frac{SE_{\text{regression}}}{b_1}\sqrt{\frac{1}{n} + \frac{1}{N} + \frac{\left(y_0 - \bar{y}\right)^2}{b_1^2\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \tag{5.19}$$

where N is the number of repeated measurements made on the unknown test sample.

The derivation of the Equation (5.19) is described in detail by Hibbert [57]. Ellison has also presented some modified equations for the calculation of the standard error of prediction as a function of the correlation coefficient [102]. It has to be noted that Equation (5.19) actually calculates the combined uncertainty of the regression line and the uncertainty compound due to the repeatability of the response. The latter has a variance which is estimated as:

$$Var(y_0) = \frac{SE_{\text{regression}}^2}{N} \tag{5.20}$$

If we want to calculate the uncertainty of the regression line only, then the term $1/N$ in Equation (5.19) should be omitted. This leads to the expression:

$$s'(x_{\text{pred}}) = \frac{SE_{\text{regression}}}{b_1}\sqrt{\frac{1}{n} + \frac{\left(y_0 - \bar{y}\right)^2}{b_1^2\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \tag{5.21}$$

The above expression does not take into account the uncertainty $u(y_0)$ of the response y_0 which is an input parameter of the measurement model (Equation (5.4)). This may lead to results underestimated and not comparable with the results of the

other methodologies. The standard uncertainty of the predicted value including the uncertainty of the response is given by the equation:

$$u(x_{\text{pred}}) = \sqrt{s'(x_{\text{pred}})^2 + \left[c_2 u(y_0)\right]^2} \tag{5.22}$$

5.5 Results and discussion

5.5.1 Calibration results

A least square linear regression was carried out for the calibration data using the LINEST function of Microsoft® Excel. The results of the regression analysis are presented in Table 5.3. The coefficient of determination for the straight line, r^2, is 0.989, and the regression ANOVA gave a Fisher – Snedecor value *F* equal to 1398 with a significance (*p* – value) of 5.3 10^{-17}. This implies a strong linear relationship between concentration and instrument response values. Figure 5.1 shows the plot produced by the calibration data and the fitted line.

Table 5.3 Results of least square linear regression

Symbol	Description	Value
b_1, b_0	The estimates of the slope and the intercept.	1118.89 counts $ng^{-1}\mu L$, 1651.87 counts
$SE(b_0)$	The standard error of the value of the intercept.	167.75 counts
$SE(b_1)$	The standard error of the value of the slope.	29.92 counts $ng^{-1}\mu L$
r^2	The coefficient of determination. It is a measure of the significance of the degree of correlation between *Y* and *x* values.	0.989
$SE_{\text{regression}}$	The standard error of regression (or residual standard deviation). It is a statistical measure of the deviation of the data from the fitted regression line.	451.25 counts
F	The *F* statistic, or the Fisher – Snedecor *F*-value. It is used to determine whether the observed relationship between the dependent and independent variables occurs by chance.	1398
df	The degrees of freedom.	16
$SS_{\text{regression}}$	The regression sum of squares. It represents the variability in the data that can be accounted for by the fitted regression line.	284674634 $counts^2$
SS_{residual}	The residual sum of squares. It represents the variability that can be accounted for by the observed residuals.	3258057$counts^2$

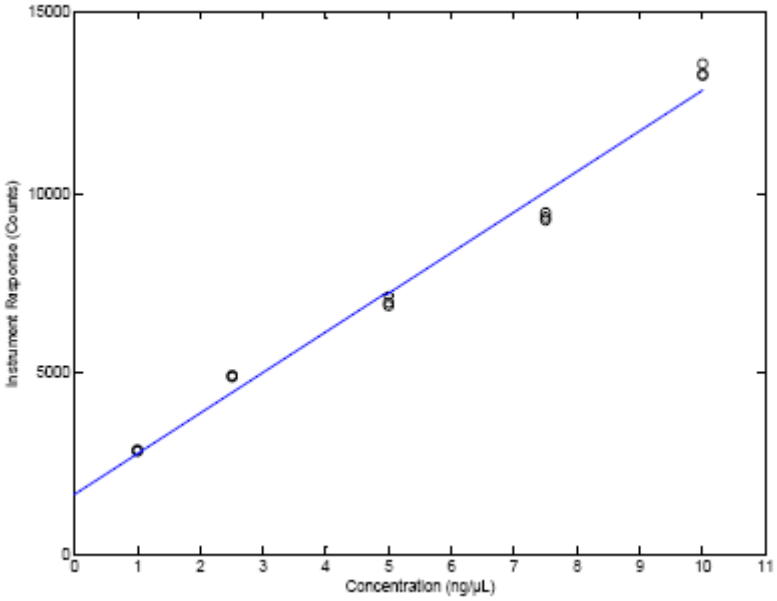


Figure 5.1 Calibration graph

5.5.2 Uncertainty estimation results

The application of the GUM uncertainty framework (described in Section 5.4.1) gave the results presented in Table 5.4 (uncertainty budget). The calculations gave a standard uncertainty $u(x_{pred})$ equal to 0.175 ng μL^{-1}. In order to show the importance of the existing correlation between b_0 and b_1, which is often neglected when applying GUM, the standard uncertainty was also calculated without taking into account the covariance term. The calculations gave a standard uncertainty $u(x_{pred})$ equal to 0.283 ng μL^{-1}. Thus, when applying the GUM uncertainty framework, ignoring the correlation leads to a 62 % overestimation of the standard uncertainty. It has to be pointed out though that the uncertainty overestimation when neglecting correlation between input parameters is not always the case. There are cases where the inclusion of the correlation results in higher uncertainties.

Table 5.4 Uncertainty budget for x_{pred}

Quantity	Estimate	Probability distribution	Standard uncertainty	Sensitivity coefficient	Uncertainty contribution
Instrument response, y_0	10603.01	Rectangular	122.43 counts	0.0008937 counts^{-1} ng µL^{-1}	0.1094 ng µL^{-1}
Intercept, b_0	1651.87 counts	Normal	167.75 counts	0.0008937 counts^{-1} ng µL^{-1}	0.1499 ng µL^{-1}
Slope, b_1	1118.89 counts ng^{-1}µL	Normal	29.92 counts ng^{-1}µL	0.0071499 counts^{-1} ng^2 µL^{-2}	0.2140 ng µL^{-1}
		Combined standard uncertainty (no correlation included)			**0.283** ng µL^{-1}
		Covariance term			-0.0496 ng^2 µL^{-2}
		Combined standard uncertainty			**0.175** ng µL^{-1}

The implementation of Kragten spreadsheet approach (Figure 5.2) according to the procedure described in Section 5.4.2 gave results very close to the results of the GUM uncertainty framework. The standard uncertainty $u(x_{pred})$ was calculated 0.172 ng µL^{-1} when the correlation term was included and 0.279 ng µL^{-1} without the inclusion of the correlation term. This closeness of the results was rather expected as the assumptions for linearity and small standard uncertainties compared to the value of the respective input quantities are valid in our case.

		$u(y_0)$	$u(b_0)$	$u(b_1)$
		122,43	167,75	29,92
y_0	10603,0	10725,43	10603,00	10603,00
b_0	1651,87	1651,87	1819,62	1651,87
b_1	1118,89	1118,89	1118,89	1148,81
i		y_0	b_0	b_1
x_{pred}	**8,000**	8,109	7,850	7,792
$u(x_{pred}, i)$		-0,109	0,150	0,208
$u(x_{pred}, i)^2$		0,012	0,022	0,043
$\Sigma u(x_{pred}, i)^2$	0,078			
Results				
Standard uncertainty without accounting for correlation				
$u(x_{pred})$	**0,279**			
Standard uncertainty accounting for correlation				
$r(b_0, b_1)$	-0,773			
correlation term	-0,048			
$u(x_{pred})$	**0,172**			

Figure 5.2 Kragten approach spreadsheet implementation

The algorithm of Monte Carlo method (described in Section 5.4.3) was implemented in MATLAB® [92] using a fixed number of trials (M=10^6). MATLAB® code is presented in Table 5.5. Samples for b_0 and b_1 were drawn from a bivariate (or joint) Gaussian distribution $N(\boldsymbol{E},\boldsymbol{V})$ characterized by the expectation and the covariance (or uncertainty) matrices, $\boldsymbol{E}$ and $\boldsymbol{V}$, respectively:

$$E = \begin{bmatrix} b_o \\ b_1 \end{bmatrix} \tag{5.23}$$

$$V = \begin{bmatrix} u^2(b_0) & u(b_0,b_1) \\ u(b_0,b_1) & u^2(b_1) \end{bmatrix} = \begin{bmatrix} SE^2(b_0) & r(b_0,b_1)SE(b_0)SE(b_1) \\ r(b_0,b_1)SE(b_0)SE(b_1) & SE^2(b_1) \end{bmatrix} \tag{5.24}$$

Sample for y_0 was drawn from rectangular (uniform) distribution $R(y_0-\sqrt{3}u(y_0), y_0+\sqrt{3}u(y_0))$.

Table 5.5 Implementation program of Monte Carlo method in MATLAB®

Line	MATLAB Code
1	`function [x_pred]=regression_cov(N)`
2	
3	`covar=2; % (1 'covariance no', other value 'covariance yes')`
4	
5	`% Calibration Data`
6	`x1=0; x2=1; x3=2.51; x4=5; x5=7.5; x6=10; x7=0; x8=1; x9=2.51; x10=5; x11=7.5; x12=10; x13=0; x14=1; x15=2.51; x16=5; x17=7.5; x18=10;`
7	
8	`y1=1464.9; y2=2882.7; y3=4889.8; y4=6852.5; y5=9344.5; y6=13253.4; y7=1558.4; y8=2851.6; y9=4925.3; y10=6934.5; y11=9466.2; y12=13567; y13=1594.8; y14=2844.3; y15=4917.3; y16=7150.1; y17=9265.8; y18=13277.7;`
9	
10	`%mean of all concentrations`
11	`x_av=(x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12+x13+x14+x15+x16+x17+x18)/18;`
12	
13	`% mean of all responses`
14	`y_av=(y1+y2+y3+y4+y5+y6+y7+y8+y9+y10+y11+y12+y13+y14+y15+y16+y17+y18)/18;`
15	
16	`% Obtain estimates for the slope and the intercept`
17	`Xi=[x1; x2; x3; x4; x5; x6; x7; x8; x9; x10; x11; x12; x13; x14; x15; x16; x17; x18];`
18	
19	`Yi=[y1; y2; y3; y4; y5; y6; y7; y8; y9; y10; y11; y12; y13; y14; y15; y16; y17; y18];`
20	
21	`b1=sum((Xi-x_av).*(Yi-y_av))./sum((Xi-x_av).^2)`
22	`b0=y_av-b1.*x_av`
23	
24	`%standard errors of slope and intercept`
25	`u_b1=29.92494948;`
26	`u_b0=167.7533433;`
27	`if covar==1 % ignore covariance`

Line	MATLAB Code
28	% simulate b1
29	B1=normrnd(b1,u_b1,1,N);
30	B1=transpose(B1);
31	
32	% simulate b0
33	B0=normrnd(b0,u_b0,1,N);
34	B0=transpose(B0);
35	
36	else % do not ignore covariance
37	r_b1b0=-0.773305935; % correlation coefficient
38	u_b1b0=u_b1*u_b0*r_b1b0;
39	
40	%covariance matrix
41	V=[u_b1^2 u_b1b0; u_b1b0 u_b0^2];
42	
43	%mean matrix
44	
45	E=[b1; b0];
46	
47	% simulate b1,b0 from joint pdf
48	
49	X=mvnrnd(E,V,N); % Nx2 matrix
50	
51	B1=X(:,1); % 1st column of Nx2 matrix
52	
53	B0=X(:,2); % 2nd column of Nx2 matrix
54	end
55	
56	% simulate sample of y0
57	[a,b]=unif_param(10603.0087,122.4329981);
58	y0=unifrnd(a,b,N,1);
59	% obtain x_pred
60	x_pred=(y0-B0)./B1;
61	% Results
62	x_pred=sort(x_pred);
63	Mean_Value=mean(x_pred)
64	Standard_Uncertainty=std(x_pred)
65	
66	figure;
67	lower=min(x_pred);
68	upper=max(x_pred);
69	xc=lower:(upper-lower)/499:upper;
70	n=hist(x_pred,xc);
71	bar(xc,n./(((upper-lower)/499)*N),1);
72	hold on;
73	Z=normpdf(x_pred,8.000,0.175);
74	plot(x_pred,Z,'-.r')
75	%ESTIMATE AND PLOT COVERAGE REGIONS FOR b1, b0
76	data=[B0 B1];
77	
78	figure; % 95% rectangular and elliptical coverage region
79	
80	%Plot the original data
81	plot(data(:,1), data(:,2), 'k.');
82	%mindata = min(min(data));
83	%maxdata = max(max(data));
84	xlim([min(B0)-50, max(B0)+50]);
85	ylim([min(B1)-50, max(B1)+50]);
86	hold on;
87	
88	% Set the axis labels
89	hXLabel = xlabel('b0');
90	hYLabel = ylabel('b1');
91	% Plot 95% rectangular coverage region
92	
93	rectangle('Position',[b0-2.24*u_b0, b1-2.24*u_b1, 2*2.24*u_b0,

Line	MATLAB Code
	2*2.24*u_b1], 'Linestyle', '--')
94	Plot 95% rectangular and elliptical coverage region
95	
96	
97	% Calculate the eigenvectors and eigenvalues
98	covariance = cov(data);
99	[eigenvec, eigenval] = eig(covariance);
100	
101	% Get the index of the largest eigenvector
102	[largest_eigenvec_ind_c, r] = find(eigenval == max(max(eigenval)));
103	largest_eigenvec = eigenvec(:, largest_eigenvec_ind_c);
104	
105	% Get the largest eigenvalue
106	largest_eigenval = max(max(eigenval));
107	
108	% Get the smallest eigenvector and eigenvalue
109	if(largest_eigenvec_ind_c == 1)
110	smallest_eigenval = max(eigenval(:,2))
111	smallest_eigenvec = eigenvec(:,2);
112	else
113	smallest_eigenval = max(eigenval(:,1))
114	smallest_eigenvec = eigenvec(1,:);
115	end
116	
117	% Calculate the angle between the x-axis and the largest eigenvector
118	angle = atan2(largest_eigenvec(2), largest_eigenvec(1));
119	
120	% This angle is between -pi and pi.
121	% Shift it such that the angle is between 0 and 2pi
122	if(angle < 0)
123	angle = angle + 2*pi;
124	end
125	% Get the coordinates of the data mean
126	avg = mean(data);
127	
128	% Get the 95% confidence interval error ellipse
129	chisquare_val = 2.4477;
130	theta_grid = linspace(0,2*pi);
131	phi = angle;
132	X0=avg(1);
133	Y0=avg(2);
134	a=chisquare_val*sqrt(largest_eigenval);
135	b=chisquare_val*sqrt(smallest_eigenval);
136	
137	% the ellipse in x and y coordinates
138	ellipse_x_r = a*cos(theta_grid);
139	ellipse_y_r = b*sin(theta_grid);
140	
141	%Define a rotation matrix
142	R = [cos(phi) sin(phi); -sin(phi) cos(phi)];
143	
144	%Rotate the ellipse to some angle phi
145	r_ellipse = [ellipse_x_r;ellipse_y_r]' * R;
146	
147	% Draw the error ellipse
148	plot(r_ellipse(:,1) + X0,r_ellipse(:,2) + Y0,'k-')
149	figure % marginal probabilities of b0, b1
150	
151	subplot(1,2,1);
152	B1=sort(B1);
153	Mean_value=mean(B1);
154	lower=min(B1)
155	upper=max(B1)
156	xc=lower:(upper-lower)/199:upper;
157	n=hist(B1,xc);

Line	MATLAB Code
158	`bar(xc,n./(((upper-lower)/199)*N),1);`
159	`hold on;`
160	`Z=normpdf(B1,b1,u_b1);`
161	`plot(B1,Z,'-.r')`
162	`title('b1')`
163	
164	`subplot(1,2,2);`
165	`B0=sort(B0);`
166	`Mean_value=mean(B0);`
167	`lower=min(B0)`
168	`upper=max(B0)`
169	`xc=lower:(upper-lower)/199:upper;`
170	`n=hist(B0,xc);`
171	`bar(xc,n./(((upper-lower)/199)*N),1);`
172	`hold on;`
173	`Z=normpdf(B0,b0,u_b0);`
174	`plot(B0,Z,'-.r')`
175	`title('b0')`
176	
177	`end`
178	`function [a,b]=unif_param(m,v)`
179	`% A function to obtain the parameters of the uniform (rectangular)`
180	`% distribution given its mean and standard deviation`
181	`b=m+v*sqrt(12)/2;`
182	`a=2*m-b;`
183	`end`

The results produced by the implementation of MCM were an estimate x_{pred} =8.003 ng µL^{-1} with an associated standard uncertainty $u(x_{pred})$ =0.175 ng µL^{-1}. Figure 5.3 presents the probability density function of x_{pred} provided by GUM – Gaussian distribution and MCM. MCM was also implemented assuming zero covariance between b_0 and b_1, drawing samples from two independent Gaussian distributions, $N(b_0, SE(b_0))$ and $N(b_1, SE(b_1))$. The MCM algorithm gave an estimate x_{pred} =8.005 ng µL^{-1} with an associated standard uncertainty $u(x_{pred})$ =0.284 ng µL^{-1}.

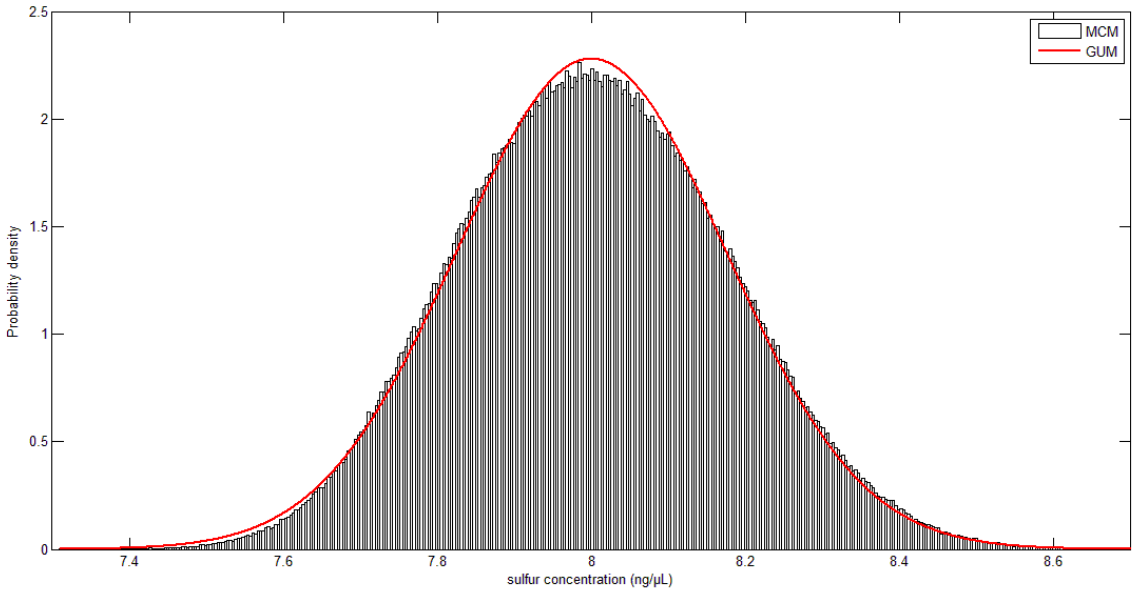


Figure 5.3 Probability density function of x_{pred} provided by GUM – Gaussian distribution and MCM

Concerning the use of the approximate equation (Section 5.4.4), the application of the Equations (5.21) and (5.22) to our calibration data provides an estimate of the standard uncertainty $u(x_{pred})$ equal to 0.175 ng μL^{-1} which agrees well with the results of the other methodologies. The application of Equation (5.19) with N=1 would lead to a falsely overestimated result of 0.426 ng μL^{-1}. This overestimation is due to the fact that Equation (5.19) includes the estimation of the variance of the observed instrument response (repeatability component), $Var(y_0)$ as the variance of the regression (squared standard error of regression). On the other hand, the response y_0 has a standard uncertainty $u(y_0)$, which if omitted, setting $u(x_{pred})$ equal to $s'(x_{pred})$, leads to a 22% underestimated result (0.137 ng μL^{-1}). Results of all approaches are summarized in Table 5.6.

Table 5.6 Results of the application of various approaches to the estimation of uncertainty of the calibration curve used for the determination of sulfur mass concentration in fuels.

	Mean value (ng μL^{-1})	Standard uncertainty (ng μL^{-1})
GUM (correlation included)	8.000	0.175
Kragten method (correlation included)	8.000	0.172
MCM (correlation included)	8.003	0.175
Standard error of prediction equation (including response uncertainty)	8.000	0.175
Standard error of prediction equation (no response uncertainty included)	8.000	0.137
GUM (no correlation included)	8.000	0.283
Kragten method (no correlation included)	8.000	0.279
MCM (no correlation included)	8.005	0.284

5.5.3 Treating calibration curve as a bivariate measurement model

The estimation of the slope and the intercept of a calibration curve is based on a measurement model of multiple outputs, which involves correlated data. Given an estimate of the output quantity $\boldsymbol{E}$ (Equation 5.23), its associated covariance matrix $\boldsymbol{V}$ (Equation 5.24) and a coverage probability p, and once the probability density function (PDF) is established (bivariate Gaussian in our case) a coverage region can be determined. A coverage region specifies a region in 2-dimensional space that contains $\boldsymbol{E}$ with probability p. Two types of coverage region can be considered [16]:

- rectangle centered coverage region (separately determined coverage intervals for b_1 and b_0).
- ellipse centered coverage region

The rectangle centered coverage region (sides parallel to the axes) is determined by the marginal PDFs of b_1 and b_0 and the coverage intervals are: $b_1 \pm k_q\, u(b_1)$ and $b_0 \pm k_q\, u(b_0)$. For coverage probability p=0.95, a coverage factor k_q = 2.24, is used.

The ellipse centered coverage region is determined as:

$$(\boldsymbol{\eta} - \boldsymbol{E})^T\, \boldsymbol{V}^{-1}\, (\boldsymbol{\eta} - \boldsymbol{E}) = k_p^{\,2} \qquad (5.25)$$

or

$$\begin{bmatrix} n_0 - b_0 & n_1 - b_1 \end{bmatrix} \begin{bmatrix} u^2(b_0) & u(b_0, b_1) \\ u(b_0, b_1) & u^2(b_1) \end{bmatrix}^{-1} \begin{bmatrix} n_0 - b_0 \\ n_1 - b_1 \end{bmatrix} = k_p^2 \tag{5.26}$$

where $\boldsymbol{\eta} = (\eta_0, \eta_1)^T$ is a vector variable describing possible values of the output quantity $\boldsymbol{E}$. For coverage probability p=0.95, a coverage factor k_p = 2.45, is used.

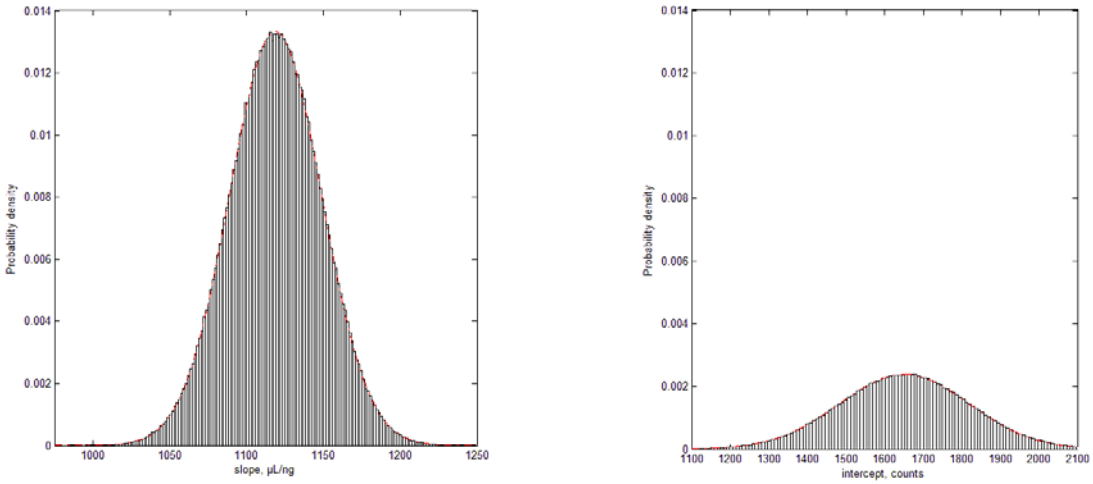


Figure 5.4 Marginal distributions of the slope (b_1) and the intercept (b_0)

Figure 5.4 shows the marginal distributions for the slope (b_1) and the intercept (b_0) produced by the application of MCM. Figure 5.5 shows elliptical and rectangular coverage regions (coverage probability p=0.95) for the bivariate quantity $\boldsymbol{E}$ characterized by a Gaussian probability distribution for which the component quantities b_1 and b_0 are correlated. 1000 draws were carried out (instead of 10^6) for illustration purposes.

As in our case the two outputs of the measurement model (b_1 and b_0) are mutually correlated, the ellipse centered coverage region is more appropriate than the rectangular one. Generally the ellipse is the smallest coverage region for the stipulated probability, while the rectangle does not reflect at all the correlation between component quantities. Therefore, the rectangular coverage region might be considered inappropriate as a coverage region (the coverage probability for the rectangular region exceeds 0.95). A rectangle with sides parallel to the axes of the ellipse would have smaller area and might be considered more appropriate, but

might be inconvenient since it would be expressed in terms of variables that would be artificial in terms of the application.

Figure 5.6 shows elliptical and rectangular coverage regions (coverage probability p=0.95) for the bivariate quantity $\boldsymbol{E}$ characterized by a Gaussian probability distribution not taking into account the correlation between b_1 and b_0. Finally, Figure 5.7 depicts, in the form of contour lines, elliptical coverage regions (for various coverage probabilities p). The parameters k_p that were used (Equation 5.25) were calculated from the chi-squared distribution so that:

$$p = \Pr(\chi_2^2 \le k_p^2) \tag{5.27}$$

where χ_2^2 has a chi-squared distribution with two degrees of freedom.

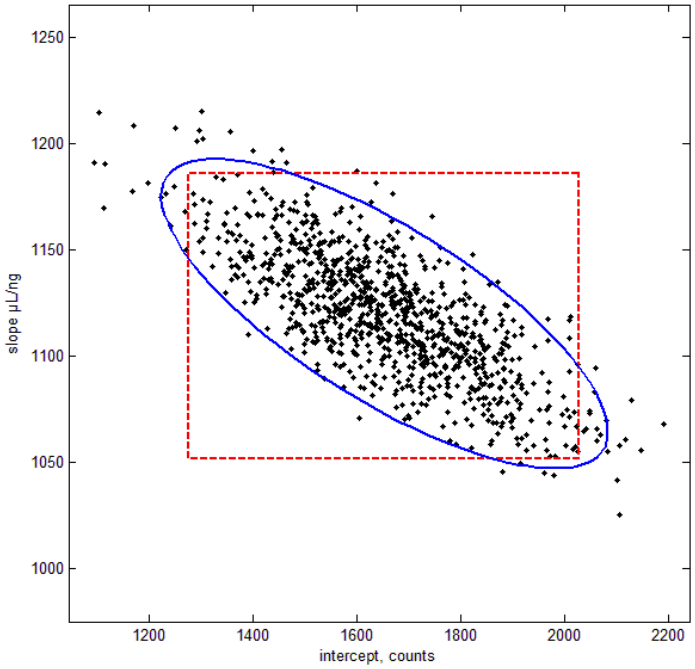


Figure 5.5 Elliptical and rectangular coverage regions (coverage probability p=0.95) of the joint probability density function of the slope, b_1 and the intercept,b_0.

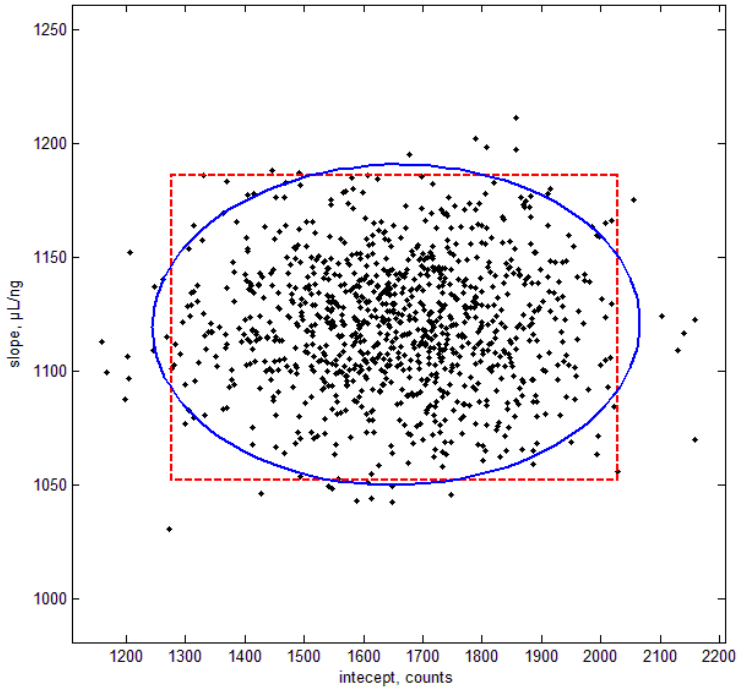


Figure 5.6 Elliptical and rectangular coverage regions (coverage probability p=0.95) of the joint probability density function of the slope, b_1 and the intercept,b_0, ignoring the correlation between b_1 and b_0 .

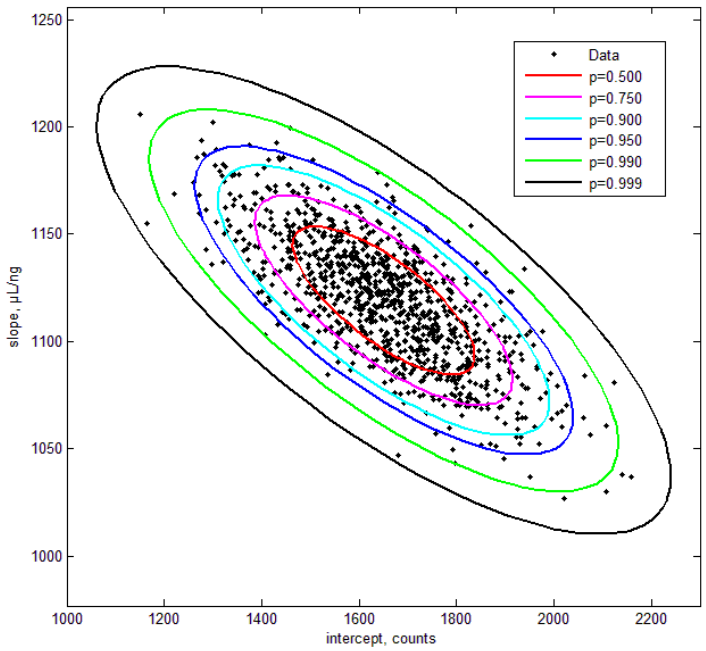


Figure 5.7 Elliptical coverage regions (for various coverage probabilities p) of the joint probability density function of the slope, b_1 and the intercept,b_0.

5.6 Conclusions

Many analytical procedures involve the construction and the use of a calibration curve in order to determine the value of the measurand (e.g. analyte concentration). As the linear calibration function is defined by the slope and the intercept, we actually have to deal with a measurement model of multiple (two) outputs, which also involves correlated data. The work presented in this part of the thesis compared different methodologies concerning the estimation of the uncertainty of a calibration curve used for the determination of sulfur mass concentration in fuels. The methodologies applied included: the GUM uncertainty framework, the Kragten numerical method, the Monte Carlo Method (MCM) as well as the approximate equation calculating the standard error of prediction.

The standard uncertainty results obtained by the four methodologies (0.172 – 0.175 ng μL^{-1}) agree well, as there is no appreciable non linearity in the measurement model and there are no dominant parameters whose distributions are far from normal. The use of the approximate equation calculating the standard error of prediction leads to correct results only if it used appropriately i.e. omitting the term $1/N$ accounting for the repeatability of the response and including its standard uncertainty estimated from manufacturer information.

When applying GUM, MCM or Kragten approach, the correlation between calibration curve parameters (slope and intercept) is a significant component of uncertainty and cannot be ignored. The data for the calculation of correlation terms and/or covariance matrices are readily available or easily calculated. As correlation between slope and intercept reduce the uncertainty, if it is ignored, it leads to 62 % overestimated standard uncertainties of the predicted value, x_{pred}.

The calculations shown in this chapter are applicable to the most of the cases of calibration curve construction using least square linear regression. However, more general regression models that take account of significant uncertainties of the dependent parameters or correlations between the dependent and independent variables require special treatment, which can be found in ISO TS 28037, Determination and use of straight-line calibration functions [103].

6. The use of measurement uncertainty and precision data in conformity assessment

In order to use a test result to decide whether it indicates compliance or non-compliance, it is necessary to take into account the dispersion of the values that can be attributed to the measurand. When dealing with conformity assessment of automotive fuel samples against European Union specification limits, this dispersion may be represented by uncertainty estimates based on either standard method precision data (ISO 4259 approach) or within laboratory precision data (intermediate precision approach). The present part of the thesis presents possible decision rules based on these approaches and directly related to the required or acceptable level of probability of making a wrong decision. Acceptance limits for 95% and 99% confidence levels are calculated for all the properties of automotive fuels. Moreover, the effect of different approaches for defining guard bands, different levels of confidence or different number of replicate measurements is investigated using the results of the analyses of 769 diesel fuel samples for the determination of sulfur mass concentration.

6.1 Introduction

The automotive fuels placed on market should comply with strict requirements introduced by relevant legislation. In European Union (EU), several directives [1,2] set technical specifications for fuels used with positive ignition engines (petrol) or with compression ignition engines (diesel). These directives aim at the reduction of direct and indirect health and environmental risks and are supported by documents prepared by CEN (European Committee for Standardization) such as EN 228:2008 [104] and EN 590:2009 [105] that specify requirements as well as test methods for marketed and delivered unleaded petrol and automotive diesel.

Evaluation of conformity with specified requirements should provide adequate confidence that the product under test fulfills (or not) these requirements [106], minimizing the risk of incorrect decisions, which often have financial consequences [107]. As no measurement is exact, the true value of any measured quantity or any errors associated with the measurement cannot be known exactly and the measurement result is actually only an estimate. This estimate should be accompanied by an uncertainty statement or a coverage interval, which summarizes the knowledge of the possible values of the measured quantity [8]. Therefore, the assessment of conformity with specified requirements, especially when the measurement result is close to a specification limit, is closely related to the probability density function of the measurement data and should be approached using the probability theory [17]. In these cases, appropriate decision rules may permit a control over the probability of taking the wrong decision [108].

In the present part of the thesis an insight is given to the available approaches that can be used to support reliable decisions - expressed by a certain confidence level - in conformity assessment of fuels. These approaches are applied and compared for the assessment of conformity of automotive diesel fuel samples against the EU sulfur mass concentration specification. The results of the analyses of diesel fuel samples from 769 petroleum retail stations, monitored for fuel quality purposes, are used for the calculations.

6.2 Evaluation of conformity with specified requirements

The evaluation of conformity (or conformity assessment) has the objective to determine whether specified requirements relating to a product, process, system,

person or body are fulfilled or not [17,106,109]. Often, a conformity test is involved in the activity of the conformity assessment, which actually has three distinct stages: measurement of the property of interest, comparison of the measurement result with the specified requirement (or tolerance limit) and finally, decision on the action that will follow. The measurement result has to be obtained using a validated procedure, which should guarantee its metrological traceability [10]. The subsequent comparison of the result with the specified requirements should be based on predefined decision rules, which are of key importance when the result is close to the tolerance limit. The decision rules take into account the measurement process variability (expressed as standard deviation or uncertainty) in order to determine acceptance and rejection zones or intervals [55,110]. Figures 6.1 and 6.2 show acceptance intervals and their relation to the tolerance intervals defined by upper and lower specification limits, T_U and T_L, respectively. Figure 6.1 shows a case which involves an acceptance interval constructed by reducing the tolerance interval on either side by a guard band of width, w (guarded or stringent acceptance). On the other hand, Figure 6.2 shows a case which involves an acceptance interval constructed by increasing the tolerance interval on either side by a guard band of width, w (relaxed acceptance or guarded rejection) [111,112]. The guard bands are defined as the magnitude of the offset from a specification limit to the acceptance interval boundary [112]. The selected decision rules should minimize the consequences of an incorrect decision and are thus indispensably related to the determination of a minimum acceptable level of probability that the measurand lies within or outside specification limits [113].

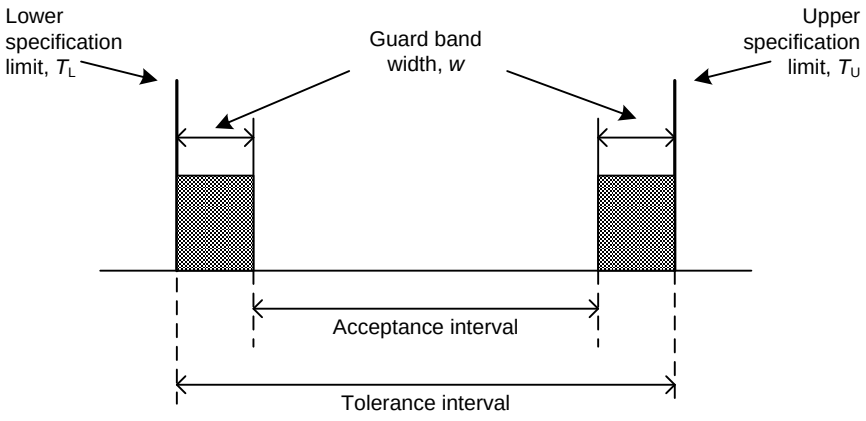


Figure 6.1 Symmetric two-sided acceptance interval created by reducing tolerance interval, defined by the lower specification limit T_L and the upper specification limit T_U, on either side by a guard band of width, w (guarded or stringent acceptance)

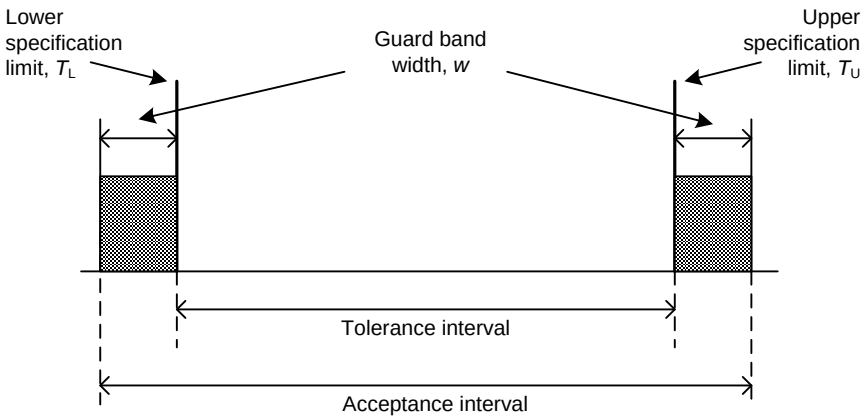


Figure 6.2 Symmetric two-sided acceptance interval created by expanding the tolerance interval, defined by the lower specification limit T_L and the upper specification limit T_U, on either side by a guard band of width, w (guarded rejection or relaxed acceptance)

There are two types of possible errors in the conformity assessment procedure, Type I and Type II. In Type I errors, conforming products are incorrectly rejected. Minimizing Type I error of a conformity assessment test means minimizing the probability of the measurand lying within specification when the test result is outside the specification limit. On the other hand, in Type II errors, non conforming products are incorrectly accepted. Minimizing Type II error of a conformity assessment test means minimizing the probability of the measurand lying outside the specification when the test result is inside the specification limit [110,114].

Guidance regarding the design and use of decision rules is provided by several documents [55,111,115-117]. Although many of them are sector specific, the principles they describe, may be applied in any kind of conformity assessment. Decision rules may be based either on the simple acceptance/ rejection or on guard bands. Applying simple acceptance/ rejection decision rules means that Figures 6.1 and 6.2 would present a situation with guard bands of zero magnitude and acceptance and tolerance intervals that coincide with each other. This decision rule is insufficient as it can lead to high (up to 50%) probabilities of Type I and Type II errors when a measured value is close to the specification limit. These probabilities can be controlled or reduced by using acceptance intervals that differ by tolerance intervals. The acceptance interval can be inside the tolerance interval (Figure 6.1) leading to reduced probability of false acceptance (Type II error). Alternatively, the acceptance interval may be wider than the tolerance interval (Figure 6.2) leading to reduced of probability of false rejection (Type I error). The reduction of these

probabilities is proportional to the width of the guard band, *w*. It has to be noted though, that the probability of Type I error is reduced at the cost of increasing the probability of Type II error and vice versa. Therefore the risks associated with making the wrong decision have to be taken into account when formulating decision rules [17,112].

6.3 Probability of conformity

6.3.1 General

Probability density function (PDF), $g_Y(\eta)$, may be employed to describe the dispersion of probable values η of a measurand *Y* about the best estimate *y*, given a measured value η_m. In many cases, this PDF is or can be approximated by a normal distribution, described by the Gaussian function:

$$g_Y(\eta|\eta_m) = \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\eta - y}{s}\right)^2\right] \tag{6.1}$$

where $y=y(\eta_m)$ and *s* is the standard deviation associated to the best estimate. In many cases there is little or no knowledge of Y before the measurement, that it can be assumed $y \approx \eta_m$ [17]. The probability that *Y* lies in the interval [*a*,*b*] is given by the equation:

$$\Pr(a \le Y \le b) = \int_a^b g_Y(\eta|\eta_m) d\eta = \Phi\left(\frac{b-y}{s}\right) - \Phi\left(\frac{a-y}{s}\right) \tag{6.2}$$

where $\Phi(z)$ is the cumulative density function (CDF) of the standard normal distribution which gives the probability of the random variable *z* falling in the interval $[-\infty, t]$:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2) dt \tag{6.3}$$

When the dispersion of the values of the measurand can be described by a normal distribution, Equation (6.2) may be used to calculate the maximum probabilities of Type I or Type II errors in the case of guarded rejection or guarded acceptance

decision rules. These maximum probabilities occur when the measurement result (best estimate, *y*) coincides with the boundary of an acceptance interval. Figures 6.3 and 6.4 show guarded acceptance decision rules for upper and lower specification limits and display the maximum probability of false acceptance (Type II error) when a guard band of width *w* is used. These maximum probabilities are calculated by the following equations:

Upper specification limit: $$\Pr(Y \geq T_{\mathrm{U}}) = 1 - \Phi\left(\frac{T_{\mathrm{U}} - (T_{\mathrm{U}} - w)}{s}\right) = 1 - \Phi\left(\frac{w}{s}\right) \quad (6.4)$$

Lower specification limit: $$\Pr(Y \leq T_{\mathrm{L}}) = \Phi\left(\frac{T_{\mathrm{L}} - (T_{\mathrm{L}} + w)}{s}\right) = \Phi\left(-\frac{w}{s}\right) \quad (6.5)$$

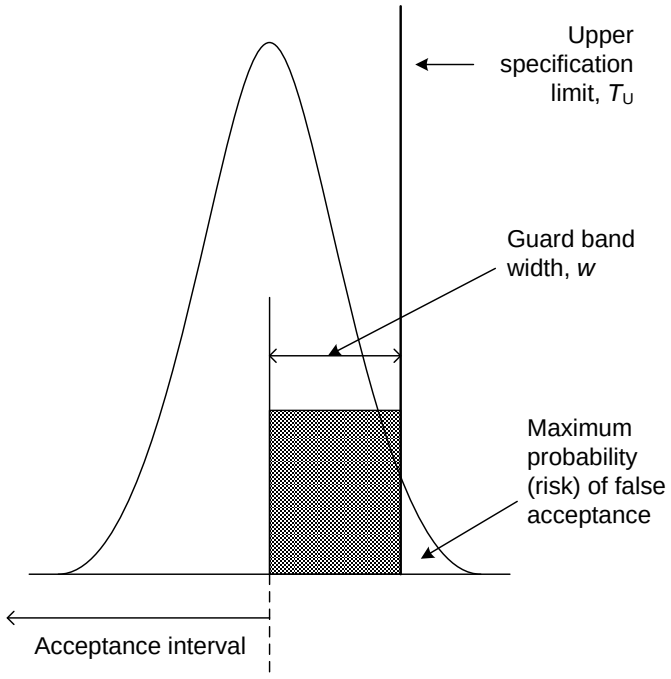


Figure 6.3 Guarded acceptance decision rule for an upper specification limit T_U and maximum probability of false acceptance (Type II error) when a guard band of width *w* is used.

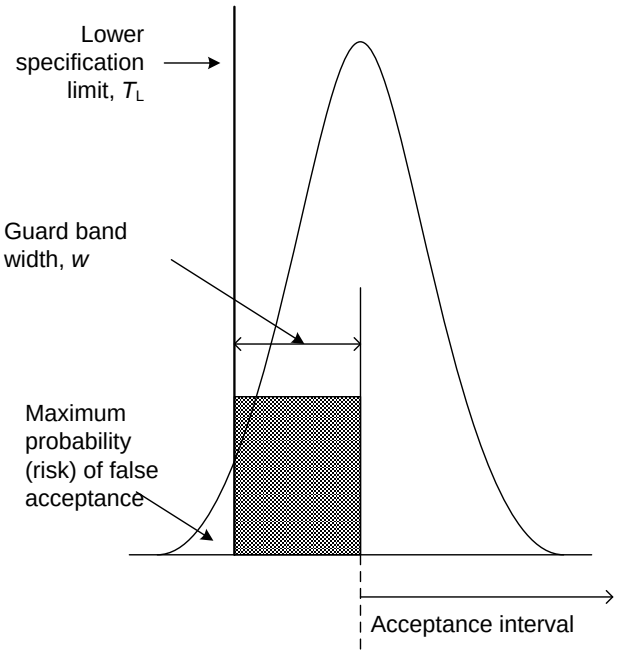


Figure 6.4 Guarded acceptance decision rule for a lower specification limit T_L and maximum probability of false acceptance (Type II error) when a guard band of width w is used.

Figures 6.5 and 6.6 show guarded rejection decision rules for upper and lower specification limits and display the maximum probability of false rejection (Type I) when a guard band of width w is used. These maximum probabilities are calculated by the following equations:

Upper specification limit: $$\Pr(Y \le T_U) = \Phi\left(\frac{T_U - (T_U + w)}{s}\right) = \Phi\left(-\frac{w}{s}\right) \quad (6.6)$$

Lower specification limit: $$\Pr(Y \ge T_L) = 1 - \Phi\left(\frac{T_L - (T_L - w)}{s}\right) = 1 - \Phi\left(\frac{w}{s}\right) \quad (6.7)$$

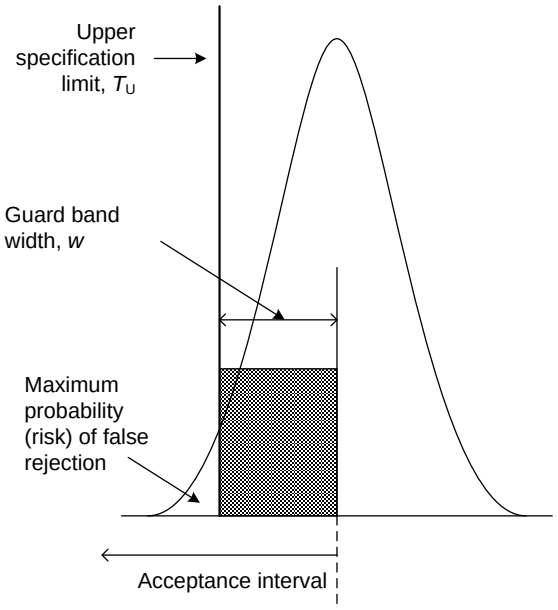


Figure 6.5 Guarded rejection decision rule for an upper specification limit T_U and maximum probability of false rejection (Type I error) when a guard band of width w is used.

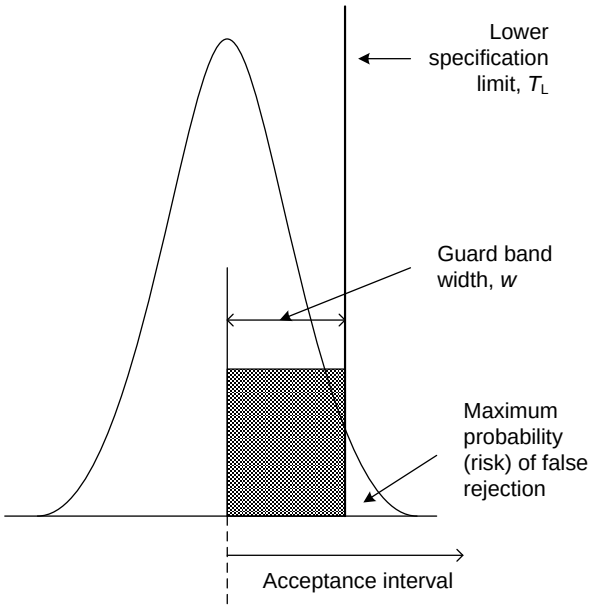


Figure 6.6 Guarded rejection decision rule for a lower limit T_L and maximum probability of false rejection (Type I error) when a guard band of width w is used.

In the present part of the thesis we are concerned with "specific" conformity assessment, which is related to the decision for conformity based on testing single fuels samples from petroleum retails stations and not from a production process. In the case of samples from a production process, "global" conformity assessment would be more appropriate as it would take into account the variability of both the measuring system and the process [17,114].

The final decision of the conformity assessment and the way it will be stated, depends on the type of test performed and whether it is a test of conformity or a test for non-conformity. This distinction is discussed in detail by Desimoni and Brunetti [112,118].

6.3.2 The between laboratory precision data approach (ISO 4259 approach)

Fuels are generally bought and sold in accordance with certain specifications, and sales terms refer to the test methods that should be used in case of a quality dispute. For example, European Union legislation for automotive fuels, that adopts the requirements referred in EN 590:2009 and EN 228:2008, makes reference to ISO 4259 [119] for dispute resolution. ISO 4259 describes the application of test data and precision data (repeatability and reproducibility) in order decide whether a fuel product meets the specifications. Suppliers and recipients of fuel products may use information provided in ISO 4259 to judge their quality based on either single or multiple results.

Based on a single test result n_m for a certain parameter, when officially monitoring fuel quality (guarded rejection), a fuel product shall be considered out of the specification, with 95% confidence, only if the test result is such that, in the case of an upper specification limit (T_U):

$$n_m > T_U + 0.59R \tag{6.8}$$

and in the case of a lower specification limit (T_L):

$$n_m < T_L - 0.59R \tag{6.9}$$

where R is the reproducibility limit of the test method used to measure the parameter compared with the specification limit.

On the other hand, based again on a single test result n_m, when receiving fuel products (guarded acceptance) and in order to ensure compliance with absolute limits and avoid disputes and extra test costs, the fuel has to be accepted from the supplier as compliant with 95% confidence only if the test result is such that, in the case of an upper specification limit (T_U):

$$n_m \leq T_U - 0.59R \quad (6.10)$$

and in the case of a lower specification limit (T_L):

$$n_m \geq T_L + 0.59R \quad (6.11)$$

When the decision considering acceptance or rejection is based on multiple test results with the n_m being the average of k number of acceptable results, the reproducibility limit R in the expressions (6.8), (6.9), (6.10) and (6.11) has to be replaced by the variable R_1, which is given by the following equation:

$$R_1 = \sqrt{R^2 - r^2\left(1 - \frac{1}{k}\right)} \quad (6.12)$$

where r is the repeatability limit of the test method used to measure the parameter compared with the specification limit.

Moreover, if a level of confidence of 99% is required for the compliance assessment decisions, then a coefficient of 0.83 instead of 0.59, has to be used in expressions (6.8), (6.9), (6.10) and (6.11). The expressions described above are summarized in Table 6.1.

Table 6.1 Expressions for calculating acceptance limits according to ISO 4259 approach for 95% and 99% confidence levels.

	Acceptance limits when receiving fuel (guarded acceptance)		**Acceptance limits when officially assessing fuel conformity (guarded rejection)**	
	95% confidence level	**99% confidence level**	**95% confidence level**	**99% confidence level**
Lower specification limit	$T_L + 0.59R$	$T_L + 0.83R$	$T_L - 0.59R$	$T_L - 0.83R$
Upper specification limit	$T_U - 0.59R$	$T_U - 0.83R$	$T_U + 0.59R$	$T_U + 0.83R$

Note: The expressions of this table are applicable when a single test result is available. In case of multiple test results the reproducibility limit R has to be replaced by R_1 (Eq. (6.12))

The expressions of Table 6.1 have been applied for all the parameters related to automotive fuel quality referred in EN 590:2009 and EN 228:2008 and the acceptance limits were calculated using the precision data of the relevant test methods. The results are presented in Tables 6.2 and 6.3.

Table 6.2 Acceptance limits for automotive diesel according to ISO 4259 approach based on the requirements specified in EN 590

Property	Units	Limits (EN 590)		Test method	Method reproducibility limit at:		Acceptance limits when receiving fuel (guarded acceptance)				Acceptance limits when officially assessing fuel conformity (guarded rejection)			
							95% confidence level		99% confidence level		95% confidence level		99% confidence level	
		min	max		min	max	min	max	min	max	min	max	min	max
Cetane number [a]		51	-	ISO 5165:1998	4.2	-	53	-	54	-	49		48	
				EN 15195:2007	3.2		53		54		49		48	
Cetane index [b]		46	-	ISO 4264:2007	not applicable									
Density at 15 °C [a]	kg/m^3	820	845	ISO 3675:1998	1.2	1.2	820.7	844.3	821.0	844.0	819.3	845.7	819.0	846.0
				ISO 12185:1996	0.5	0.5	820.3	844.7	820.4	844.6	819.7	845.3	819.6	845.4
Polycyclic aromatic hydrocarbons [a,c]	%(m/m)	-	11	EN 12916:2006	-	0.80	-	10.5	-	10.3	-	11.5	-	11.7
Sulfur content [a]	mg/kg		10	ISO 20846:2011	-	2.24	-	8.7	-	8.1	-	11.3	-	11.9
Flash point	°C	55		ISO 2719:2002	3.91	-	57	-	58	-	53	-	52	-
Carbon residue (on 10% distillation residue)	%(m/m)	-	0.30	ISO 10370:1993	-	0.110	-	0.24	-	0.21	-	0.36	-	0.39
Ash content	%(m/m)	-	0.01	ISO 6245:2001	-	0.005	-	0.007	-	0.006	-	0.013	-	0.014
Water content	mg/kg	-	200	ISO 12937:2000	-	0.97	-	199	-	199	-	201	-	201
Total contamination	mg/kg	-	24	EN 12662:2008	-	7.2	-	20	-	18	-	28	-	30
Copper strip corrosion (3h at 50°C)	rating	class 1		ISO 2160:1998	not applicable									
Fatty acid methyl ester (FAME) content [a]	%(V/V)	-	7	EN 14078:2009	-	0.5	-	6.7	-	6.6	-	7.3	-	7.4
Oxidation stability	g/m^3	-	25	ISO 12205:1995	-	13.3	-	17	-	14	-	33	-	36
	h	20	-	EN 15751:2009	4.2	-	22	-	23	-	18		17	-
Lubricity, corrected wear scar diameter at 60°C	µm	-	460	ISO 12156-1:2006	-	102	-	400	-	375	-	520	-	545

Table 6.2 (*continued*)

Property	Units	Limits (EN 590)		Test method	Method reproducibility limit at:		Acceptance limits when receiving fuel (guarded acceptance)				Acceptance limits when officially assessing fuel conformity (guarded rejection)			
							95% confidence level		99% confidence level		95% confidence level		99% confidence level	
		min	max		min	max	min	max	min	max	min	max	min	max
Viscocity at 40°C	mm^2/s	2	4.5	ISO 3104:1994	0.01	0.03	2.0	4.5	2.0	4.5	2.0	4.5	2.0	4.5
Distillation[d]	%(V/V)													
- % recovered at 250 °C		-	65		-	2.66	-	63	-	63	-	67	-	67
- % recovered at 550 °C	%(V/V)	85	-		2.66		87	-	87		83		83	-
- 95% recovered at [a]	°C	-	360	ISO 3405:2011	-	9.31	-	355	-	352	-	365	-	368
	°C			EN 116:1997										
			5		-	2.1	-	4	-	3	-	6	-	7
CFPP [e]			0		-	2.6	-	-2	-	-2	-	2	-	2
Grade A			-5		-	3.1	-	-7	-	-8	-	-3	-	-2
Grade B			-10		-	3.6	-	-12	-	-13	-	-8	-	-7
Grade C			-15		-	4.1	-	-17	-	-18	-	-13	-	-12
Grade D			-20		-	4.6	-	-23	-	-24	-	-17	-	-16
Grade E														
Grade F														

[a] Requirement referred to the European Fuels Directive 98/70/EC, including Amendment 2003/17/EC.
[b] The precision of the cetane index equation is dependent on the precision of the original density and distillation recovery temperature determinations which enter into the calculation
[c] precision data for Total Aromatic Hydrocarbons
[d] Precision limits taken from the equivalent method ASTM D86 (precision not dependent on the rate of change of temperature)
[e] Climate related requirement - Limits presented are applicable to temperate climates

Table 6.3 Acceptance limits for unleaded petrol (gasoline) according to ISO 4259 approach based on the requirements specified in EN 228

Property	Units	Limits (EN 228)		Test method	Method reproducibility limit at:		Acceptance limits when receiving fuel (guarded acceptance)				Acceptance limits when officially assessing fuel conformity (guarded rejection)			
							95% confidence level		99% confidence level		95% confidence level		99% confidence level	
		min	max		min	max	min	max	min	max	min	max	min	max
Research Octane Number, RON [a]	-	95	-	ISO 5164:2005	0.7	-	95.4	-	95.6	-	94.6	-	94.4	-
Motor Octane Number, MON [a]	-	85	-	ISO 5163:2005	0.9	-	85.5	-	85.7	-	84.5	-	84.3	-
Lead content [a]	mg/l	-	5	EN 237:2004	-	0.62	-	4.6	-	4.5	-	5.4	-	5.5
Density (at 15°C)	kg/m^3	720	775	ISO 3675:1998		1.2	-	774.3	-	774.0	-	775.7	-	776.0
				ISO 12185:1996	-	0.5	-	774.7	-	774.6	-	775.3	-	775.4
Sulfur content [a]	mg/kg	-	10	ISO 20846:2011	-	2.71	-	8.4	-	7.8	-	11.6	-	12.2
Oxidation stability	minutes	360	-	ISO 7536:1994	36	-	381.2	-	389.9	-	338.8	-	330.1	-
Existent gum content (solvent washed)	mg/100ml	-	5	ISO 6246:1995	-	3.84	-	2.7	-	1.8	-	7.3	-	8.2
Copper strip corrosion (3h at 50°C)	rating	class 1		EN 2160:1998	not applicable									
Appearance	-	clear and bright		visual inspection	not applicable									
Hydrocarbon type content [a] - olefins	% (V/V)	-	18	EN 15553:2007	-	4.6	-	15.3	-	14.2	-	20.7	-	21.8
- aromatics		-	35		-	3.7	-	32.8	-	31.9	-	37.2	-	38.1
Benzene content [a]	% (V/V)	-	1	EN 238:1996	-	0.17/ 0.3	-	0.9	-	0.9	-	1.2	-	1.2
				EN 12177:1998	-	0.1	-	0.94	-	0.92	-	1.06	-	1.08
Oxygen content [a]	%(m/m)		2.7	EN 1601:1997										
				EN 13132:2000	-	0.3	-	2.52	-	2.45	-	2.88	-	2.95

Table 6.3 (*continued*)

Property	Units	Limits (EN 228)		Test method	Method reproducibility limit at:		Acceptance limits when receiving fuel (guarded acceptance)				Acceptance limits when officially assessing fuel conformity (guarded rejection)			
							95% confidence level		99% confidence level		95% confidence level		99% confidence level	
		min	max		min	max	min	max	min	max	min	max	min	max
Oxygenates content: [a]	%(V/V)			EN 1601:1997										
- methanol		-	3.0	EN 13132:2000	-	0.3/0.4	-	2.8	-	2.8	-	3.2	-	3.3
- ethanol		-	5.0		-	0.4/0.5	-	4.8	-	4.7	-	5.3	-	5.4
- iso-propyl alcohol		-	10.0		-	0.8	-	9.5	-	9.3	-	10.5	-	10.7
- iso-butyl alcohol		-	10.0		-	0.8	-	9.5	-	9.3	-	10.5	-	10.7
- tert-butyl alcohol		-	7.0		-	0.5/0.6	-	6.7	-	6.6	-	7.4	-	7.5
- ethers (5 or more C atoms)		-	15.0		-	1	-	14.4	-	14.2	-	15.6	-	15.8
- other oxygenates		-	10.0		-	0.8	-	9.5	-	9.3	-	10.5	-	10.7
Vapour pressure (DVPE) [a,b]	kPa			EN 13016-1:2007										
- class A		45.0	60.0		2.75	2.75	46.6	58.4	47.3	57.7	43.4	61.6	42.7	62.3
- class B		45.0	70.0		2.75	2.75	46.6	68.4	47.3	67.7	43.4	71.6	42.7	72.3
- class C/C1		50.0	80.0		2.75	2.75	51.6	78.4	52.3	77.7	48.4	81.6	47.7	82.3
- class D/D1		60.0	90.0		2.75	2.75	61.6	88.4	62.3	87.7	58.4	91.6	57.7	92.3
- class E/E1		65.0	95.0		2.75	2.75	66.6	93.4	67.3	92.7	63.4	96.6	62.7	97.3
- class F/F1		70.0	100.0		2.75	2.75	71.6	98.4	72.3	97.7	68.4	101.6	67.7	102.3
Distillation [b,c]				ISO 3405:2011										
- % evaporated at 70 °C	% (V/V)													
- classes A/B		20	48		2.6	2.04	21.5	46.8	22.2	46.3	18.5	49.2	17.8	49.7
- classes C/C1/D/D1/E/E1/F/F1		22	50		2.56	2	23.5	48.8	24.1	48.3	20.5	51.2	19.9	51.7
- % evaporated at 100 °C [a]	% (V/V)	46.0	71.0		2.08	1.58	47.2	70.1	47.7	69.7	44.8	71.9	44.3	72.3

Table 6.3 (*continued*)

Property	Units	Limits (EN 228)		Test method	Method reproducibility limit at:		Acceptance limits when receiving fuel (guarded acceptance)				Acceptance limits when officially assessing fuel conformity (guarded rejection)			
							95% confidence level		99% confidence level		95% confidence level		99% confidence level	
		min	max		min	max	min	max	min	max	min	max	min	max
- % evaporated at 150 ^{o}C [a]	% (V/V)	75.0	-		1.5	-	75.9	-	76.2	-	74.1	-	73.8	-
- Final Boiling Point (FBP)	^{o}C	-	210		-	6.78	-	206.0	-	204.4	-	214.0	-	215.6
Vapour Lock Index (VLI) [b,d]					not applicable									
- classes C/D/E/F		-	-											
- class C1		-	1050											
- class D1		-	1150											
- class E1		-	1200	calculation										
- class F1		-	1250	method										

[a] Requirement referred to the European Fuels Directive 98/70/EC, including Amendment 2003/17/EC
[b] Climate related requirement
[c] Precision limits taken from the equivalent method ASTM D86 (precision not dependent on the rate of change of temperature)
[d] The precision of the vapour lock index equation is dependent on the precision of the original vapour pressure and distillation evaporated volume determinations which enter into the calculation

6.3.3 The intermediate precision approach

The needs of the laboratories for guidance on reliable compliance assessment decisions led to the development of the Guide "Use of uncertainty information in compliance assessment" [55] by the Working Group on Measurement Uncertainty and Traceability of EURACHEM/CITAC. The key concept of the EURACHEM/CITAC Guide is the use of decision rules for the determination of acceptance and rejection zones that take into account the measurement uncertainty [113]. Therefore, whereas the ISO 4259 defines the guard bands using solely the test method precision data, the EURACHEM/CITAC guide is more generic and defines the guard bands using measurement uncertainty estimates. The laboratory may use any accepted methodology for estimating its measurement uncertainty. In the present part of the thesis, intermediate precision data (standard deviation) calculated using robust ANOVA were used. The size of the guard band, *w*, is actually a multiple of the standard uncertainty *u*. This uncertainty should ideally include sampling uncertainty, as well. If the distribution of the likely values of the measurand is approximately normal, then a value of 1.64*u* may enable a confidence level of 95% concerning correct rejection or correct acceptance (depending on the decision rule selected). If a level of confidence of 99% is required for the compliance assessment decision, then a guard band of width 2.33*u*, has to be used. The expressions that may be used in fuel compliance assessment using measurement uncertainty information are summarized in Table 6.4.

Table 6.4 Acceptance limits according to the intermediate precision approach for 95% and 99% confidence levels.

	Acceptance limits when receiving fuel (guarded acceptance)		**Acceptance limits when officially assessing fuel conformity (guarded rejection)**	
	95% confidence level	**99% confidence level**	**95% confidence level**	**99% confidence level**
Lower specification limit	$T_L + 1.64u$	$T_L + 2.33u$	$T_L - 1.64u$	$T_L - 2.33u$
Upper specification limit	$T_U - 1.64u$	$T_U - 2.33u$	$T_U + 1.64u$	$T_U + 2.33u$

An interesting aspect of this approach is that the larger the value of the standard uncertainty *u*, the larger the number of samples that will be assessed incorrectly. This highlights the need for a target measurement uncertainty based on the acceptable probability of making an incorrect decision of compliance [110]. On the other hand, smaller uncertainty values are associated with high cost of analysis. Therefore, a balance between expenditure in higher accuracy measurements and potential costs associated with incorrect decisions, is needed [114,120-122].

6.4 Experimental work

Automotive diesel fuel samples were taken from 769 petroleum retail stations and their sulfur mass content was determined in order to assess their compliance with the EU regulatory limit of 10 mg kg^{-1}. The samples were analyzed in triplicate under repeatability conditions for sulfur mass concentration determination. An ANTEK 9000S sulfur analyzer equipped with an automatic sampler was employed in this work. This analyzer fully complies with ISO20846 [75] and ASTM 5453 [73]. The uncertainty of the measurement method has been previously estimated using replicated measurements of an extended nested experimental design (top down approach) and robust analysis of variance. The results of this estimation, which are described in detail in Chapter 3, gave a sampling standard uncertainty, $u_{sampling}$, of 0.169 mg kg^{-1} and an analysis standard uncertainty, $u_{analysis}$, of 0.265 mg kg^{-1}. The measurement standard uncertainty, *u*, which is the combination of both analytical and sampling uncertainties, is 0.314 mg kg^{-1}. The analysis standard uncertainty counts for the 71% of the measurement uncertainty. This means that the measurement uncertainty can be substantially reduced if the analysis uncertainty is decreased i.e. by making more measurements and calculating their average. In this case the analysis standard uncertainty is divided by the square root of the number, *k*, of repeated measurements leading to reduced measurement uncertainty, *u*:

$$u = \sqrt{u_{\text{sampling}}^2 + \left(\frac{u_{\text{analysis}}}{\sqrt{k}}\right)^2} \tag{6.13}$$

For *k*=2 the sulfur mass concentration measurement uncertainty was estimated at 0.252 mg kg^{-1}, while for *k*=3 at 0.228 mg kg^{-1}.

6.5 Results and discussion

An overview of the range and the dispersion of the sulfur mass content determination results is shown in the Boxplot diagram (produced by PASW 18 [83] of Figure 6.7 using the average values of the triplicate measurements of the fuel samples.

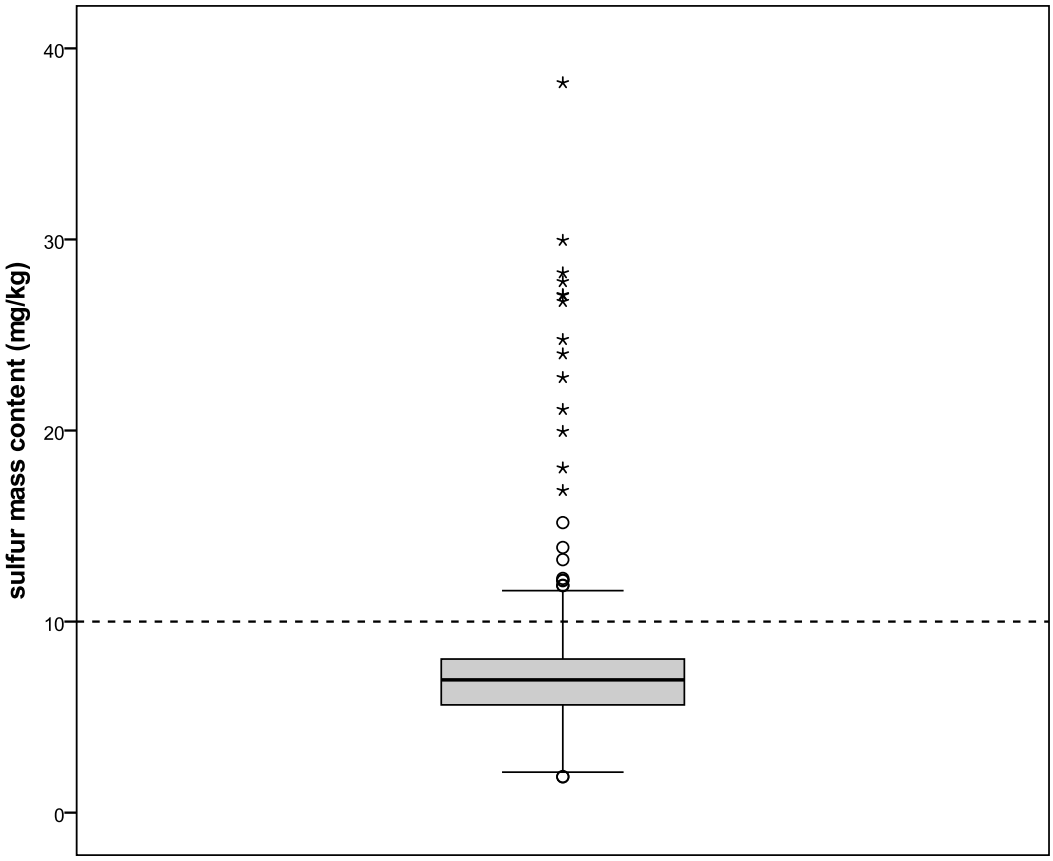


Figure 6.7 Box Plot of the average values of the 769 datasets (produced by PASW 18 [83]). The box length represents the interquantile range. Values which are more than three box lengths from either end of the box are denoted by an asterisk. Values which are between one and a half and three box lengths from either end of the box are denoted by a circle.

The majority of the samples comply with the specification of 10 mg kg^{-1}, while there are some results, that beyond any reasonable doubt can be considered non-compliant. There are some measurements (close to the specification limit) though, that require the application of certain decision rules in order to be classified as compliant or non-compliant.

The type of the decision rule and whether it will be based on guarded acceptance or guarded rejection, depends on the risks, related to making a wrong decision, that the interested parties are willing take. The approaches described in Sections 6.3.2 and 6.3.3 have been applied in the sulfur mass concentration determination results for both cases i.e. when receiving fuel (guarded acceptance) and when officially assessing fuel conformity (guarded rejection). Moreover some additional calculations have been made using Equation (6.12) (for ISO 4259 approach) and Equation (6.13) (for intermediate precision approach), in order to study the effect of multiple measurements. For these calculations, the values used for the assessment were treated as if they were obtained from either duplicate measurement (k=2) or triplicate measurements (k=3).

Tables 6.5 and 6.6 present the guard band widths, the estimated decision limits as well as the number and the percentage of non-compliant results for both 95% and 99 % confidence levels. If no guard band (or a guard band of zero magnitude) is used, then 47 (6.1%) results can be considered non-compliant. Using decision rules based on guard bands leads to either smaller (guarded rejection) or higher (guarded acceptance) number of non-compliant results. The exact number of non compliant results depends on the level of confidence used and the number of replicate measurements on each sample.

Table 6.5 Guard band widths, estimated decision limits and number of non-compliant results for both 95% and 99 % confidence level according to ISO 4259 approach.

Level of Confidence	Number of measurements, *k*	Guard band width, *w* (mg kg^{-1})	Acceptance limits when officially assessing fuel conformity (guarded rejection)		Acceptance limits when receiving fuel (guarded acceptance)	
			Decision limit (mg kg^{-1})	Number of non - compliant results	Decision limit (mg kg^{-1})	Number of non - compliant results
50%	*any number*	0.00	10.00	47 (6.1%)	10.00	47 (6.1%)
95%	1	1.32	11.32	25 (3.3%)	8.68	124 (16.1%)
	2	1.24	11.24	25 (3.3%)	8.76	114 (14.8%)
	3	1.21	11.21	25 (3.3%)	8.79	113 (14.7%)
99%	1	1.86	11.86	22 (2.9%)	8.14	173 (22.5%)
	2	1.74	11.74	22 (2.9%)	8.26	159 (20.7%)
	3	1.70	11.70	22 (2.9%)	8.30	156 (20.3%)

Table 6.6 Guard band widths, estimated decision limits and number of non-compliant results for both 95% and 99 % confidence level according to intermediate precision approach.

Level of Confidence	Number of measurements, *k*	Guard band width, *w* (mg kg^{-1})	Acceptance limits when officially assessing fuel conformity (guarded rejection)		Acceptance limits when receiving fuel (guarded acceptance)	
			Decision limit (mg kg^{-1})	Number of non - compliant results	Decision limit (mg kg^{-1})	Number of non - compliant results
50%	*any number*	0.00	10.00	47 (6.1%)	10.00	47 (6.1%)
95%	1	0.51	10.51	39 (5.1%)	9.49	63 (8.2%)
	2	0.41	10.41	41 (5.3%)	9.59	58 (7.5%)
	3	0.37	10.37	42 (5.5%)	9.63	57 (7.4%)
99%	1	0.73	10.73	35 (4.6%)	9.27	72 (9.4%)
	2	0.59	10.59	36 (4.7%)	9.41	66 (8.6%)
	3	0.53	10.53	39 (5.1%)	9.47	64 (8.3%)

Using the ISO 4259 approach (Table 6.5), guard band widths ranging from 1.21 mg kg^{-1} (95 % confidence level, 3 measurements) to 1.86 mg kg^{-1} (99 % confidence level, 1 measurement) are calculated. Using these guard bands to define decision limits for guarded rejection (official assessment of fuel conformity) leads to 25 (or 3.3%) and 22 (or 2.9%) non-compliant results, respectively. On the other hand, if these guard bands are used for defining guarded acceptance decision rules (reduce the probability of receiving a non-compliant fuel) then the number of non-compliant results is 113 (14.7%) and 173 (22.5%), respectively.

The use of the intermediate precision approach (Table 6.6), leads to considerably narrower guard bands with widths ranging from 0.37 mg kg^{-1} (95 % confidence level, 3 measurements) to 0.73 mg kg^{-1} (99 % confidence level, 1 measurement). Using these two extremes of guard bands (0.37 and 0.73 mg kg^{-1}) to define decision limits for guarded rejection (official assessment of fuel conformity) leads to 42 (or 5.5 %) and 35 (or 4.6 %) non-compliant results, respectively. On the other hand, if these guard bands are used for defining guarded acceptance decision rules (reduce the probability of receiving a non-compliant fuel) then the number of non-compliant results is 57 (7.4%) and 72 (9.4%), respectively. Compared to the ISO 4259 approach, the intermediate precision approach leads to smaller number of non-compliant results for guarded acceptance and higher number of non-compliant results for guarded rejection.

In all cases (with the exception of guarded rejection with the ISO 4259 approach), as the number of replicate measurements increases, the guard band width becomes smaller leading to changes in the number of test results classified as non-compliant. This effect is more intense in the case of guarded acceptance, because the decision limit is below the tolerance limit in a region where many test results are clustered.

It has to be noted that, the larger the width of the guard band used, the larger the proportion of samples that will be judged incorrectly. For example, in the case of guarded rejection for a confidence level of 95% with 3 measurements, there is a difference of 17 (42 minus 25) test results considered non-compliant between the intermediate precision and the ISO 4259 approach. Similarly, in the case of guarded acceptance again for a confidence level of 95% with 3 measurements, there is a difference of 56 (113 minus 57) test results considered non-compliant between the two approaches. The intermediate precision approach uses uncertainty estimates for defining the width of the guard bands. These estimates represent more precisely the dispersion of the values of the measurand. Therefore, the differences in the

calculations of the two approaches reflect a possible number of samples judged incorrectly when using the ISO 4259 approach (17 possibly non-conforming samples accepted as conforming in the guarded rejection case or 56 possibly conforming samples rejected as non-conforming, in the case of guarded acceptance). Minimizing the guard band width by reducing the measurement uncertainty (more replicates, more accurate measurement method) leads to fewer cases of false acceptance or false rejection decisions, reducing as well the costs associated with these decisions. As at the same time the cost of analysis becomes higher, there is a need that these two costs are balanced against each other in order to find an optimum measurement uncertainty [114,120-122]. Of course, fuel product quality failures are associated with high costs [123], which justify the investment in preventive actions like the use of more expensive - but at the same time more accurate – compliance assessment test methods or procedures.

6.6 Conclusions

Most analytical measurements are carried out in order to decide whether the result indicates compliance or non-compliance with a regulatory limit or specification. When the result is close to the limit, the decision cannot be that straightforward and certain decision rules have to be used. These rules, which should be accepted by all interested parties, are based on the acceptable level of the probability of making a wrong decision. The conformity assessment of automotive fuel samples against specifications set by EU directives and listed in EN 590 (for automotive diesel) and EN 228 (for unleaded petrol), is presented in this chapter. Two approaches for defining decision rules, utilizing the concept of guard bands, are presented, based on uncertainty estimates using either standard method precision data (ISO 4259 approach) or within laboratory precision data (intermediate precision approach). Under the ISO 4259 approach, acceptance limits for guarded acceptance and guarded rejection for 95% and 99% confidence levels were calculated. Moreover, the results of the analyses of 769 diesel fuel samples for the determination of sulfur mass concentration were used to highlight the differences in the resulting number of non-conforming results when using different approaches for defining guard bands, different levels of confidence or different number of replicate measurements. As the ISO 4259 approach employs “wider” guard bands compared to the intermediate precision approach, it leads to considerable more non-compliant results in the case of guarded acceptance and less non-compliant results in the case guarded rejection. The guard bands defined under the intermediate precision approach offer safer

decision rules because of the fact that they use uncertainty estimates which represent the “true” dispersion of the values of the measurand (sulfur mass concentration). On the other hand the ISO 4259 approach may result in large number of samples classified incorrectly as compliant or non-compliant. Nevertheless, this approach may be useful for end users of fuel sample analysis results when uncertainty estimates are not available.

7. Conclusions

The work presented in this thesis dealt with statistical and numerical methods concerning the estimation and use the measurement uncertainty in all parts of the measurement cycle. The development and application of methodologies for the estimation of measurement uncertainty arising from sampling was presented in Chapter 3. Chapter 4 focused on the estimation of measurement uncertainty of an analytical procedure using ISO GUM and Monte Carlo method. The uncertainty estimation in analytical methods employing the construction of a calibration function, using linear regression was considered in Chapter 5. Chapter 6 investigated the use of measurement uncertainty and precision data in conformity assessment of automotive fuel products.

The program codes developed in MATLAB in order to apply the Monte Carlo method (adaptive and fixed trials) and the empirical approaches for the estimation of sampling uncertainty may be used in any type of measurement. Moreover, based on the work presented in Chapter 6 the requirements and the key features of a software for the evaluation of conformity of fuel products were defined. Concerning future work, a focus should be put on the development of statistical tools that will establish a coherent uncertainty estimation framework based on Bayesian analysis. Moreover, more research is required for conformity assessment of qualitative and subjective properties and conformity assessment that takes into account the variability of both the measuring system and the production system as well as economic impacts.

7.1 Summary of thesis achievements

7.1.1 Outline

The judgement of "fitness for purpose" of a test method is inseparably related to the estimation of the measurement uncertainty which actually characterizes the quality of a result by accounting for both systematic and random errors. In the various pre-analysis, analysis and post - analysis steps of the measurement cycle, the measurement uncertainty has to be either estimated or taken into account. The measurement uncertainty has implications for the interpretation of analytical results in the context of regulatory compliance or conformity assessment. The estimation of measurement uncertainty following recognized and valid methodologies is also a key requirement for laboratories or other organizations accredited or seeking accreditation according to standards like ISO/IEC 17025, ISO 15189, ISO/IEC 17043 and ISO Guide 34. This, in most of the cases, involves measurement data analysis. A consistent and transferable evaluation of measurement uncertainty should follow the basic principles described in the document "Guide to the expression of uncertainty in measurement" (GUM) produced by the Joint Committee for Guides in Metrology (JCGM). The work presented in this thesis dealt with statistical and numerical methods concerning the estimation and use the measurement uncertainty in all parts of the measurement cycle.

7.1.2 Estimation of sampling uncertainty

The development and application of methodologies for the estimation of measurement uncertainty arising from sampling was presented in **Chapter 3**. Three alternative statistical approaches for data analysis concerning the estimation of measurement uncertainty of manual sampling of fuel were described and compared. A balanced experimental design was used, which included duplicate samples of automotive diesel from 104 petroleum retail stations and duplicate analyses of these samples for sulfur mass content determination. The results were treated using classical ANOVA, robust ANOVA and range statistics. The three methodologies gave statistically different estimates with the expanded uncertainty of sampling being in the range of 0.34 – 0.40 mg kg^{-1}. The fact that the robust ANOVA leads to different results compared to the other two methodologies is an indication that the assumptions of classical ANOVA and range statistics are not justified. The estimation

of robust ANOVA is considered more reliable, because of the presence of outliers within the 104 datasets used for the calculations.

7.1.3 Estimation of the uncertainty of a typical measurement procedure

Chapter 4 focused on the estimation of measurement uncertainty of an analytical procedure using ISO GUM and Monte Carlo method (MCM). The GUM approach propagates the uncertainties of the input quantities through a linearized model, while MCM provides an alternative approach in which the probability density functions (PDF's) of the input quantities are propagated through the model. Variations of both approaches were applied for the estimation of the uncertainty of the Gross Heat of Combustion (GHC) determination of a diesel fuel by bomb calorimetry. Appreciable differences were observed between the results of GUM and MCM approaches. The half width interval or expanded uncertainty results (at 95% level of confidence) obtained by GUM assuming Gaussian distribution, GUM assuming *t*-distribution and MCM were 0.28 MJ kg^{-1} (66 cal g^{-1}), 0.29 MJ kg^{-1} (70 cal g^{-1}) and 0.32 MJ kg^{-1} (75 cal g^{-1}), respectively. This means that GUM approaches are optimistic concerning the magnitude of the coverage interval of the GHC determination. This may be attributed to the slight non linearity of the measurement model and to the limitations involved in the use of the Welch-Satterthwaite formula (calculation of effective degrees of freedom of the assigned *t*-distribution) when dominant terms with relatively few degrees of freedom exist. When the GUM approach was combined with a Bayesian treatment of Type A uncertainties, the results were consistent with the results of the MCM. Overall, MCM proved to be a more reliable tool for the estimation of the uncertainty of the determination of the GHC of diesel fuel, as it is not based on approximations or assumptions and it does not have the limitations of the GUM approach.

7.1.4 Estimation of the uncertainty of a measurement procedure involving the construction of a calibration curve

The uncertainty estimation of the construction of a calibration function, used for the determination of sulfur mass concentration in fuels, using linear regression was considered in **Chapter 5.** As the linear calibration function is defined by the slope and the intercept, we actually had to deal with a measurement model of multiple (two) outputs, which also involves correlated data. The slope and the intercept were

described by a bivariate (or joint) Gaussian distribution characterized by the expectation and the covariance. Two types of coverage regions were considered: a rectangle centered coverage region and an ellipse centered coverage region. It was shown that the ellipse centered coverage region was the most appropriate one. Different methodologies were compared concerning the estimation of the uncertainty of a calibration curve used for the determination of sulfur mass concentration in fuels. The methodologies applied included: the GUM uncertainty framework, the Kragten numerical method, the Monte Carlo method (MCM) as well as the approximate equation calculating the standard error of prediction. The standard uncertainty results obtained by the four methodologies (0.172 – 0.175 ng μL^{-1}) agreed well, as there is no appreciable non linearity in the measurement model and there are no dominant parameters whose distributions are far from normal. The use of the approximate equation calculating the standard error of prediction leads to correct results only if it used appropriately i.e. omitting the term accounting for the repeatability of the response and including its standard uncertainty estimated from manufacturer information. When applying GUM, MCM or Kragten approach, the correlation between calibration curve parameters (slope and intercept) was a significant component of uncertainty and if it is ignored, it leads to 62% overestimated standard uncertainties of the predicted value.

7.1.5 Use of measurement uncertainty in conformity assessment

Chapter 6 investigated the use of measurement uncertainty and precision data in conformity assessment of automotive fuel products. The conformity assessment of automotive fuel samples against specifications set by EU directives and listed in EN 590 (for automotive diesel) and EN 228 (for unleaded petrol), was presented in this chapter. Two approaches for defining decision rules, utilizing the concept of guard bands, were presented, based on uncertainty estimates using either standard method precision data (ISO 4259 approach) or within laboratory precision data (intermediate precision approach). Under the ISO 4259 approach, acceptance limits for guarded acceptance and guarded rejection for 95% and 99% confidence levels were calculated. Moreover, the results of the analyses of 769 diesel fuel samples for the determination of sulfur mass concentration were used to highlight the differences in the resulting number of non-conforming results when using different approaches for defining guard bands, different levels of confidence or different number of replicate measurements. As the ISO 4259 approach employed “wider” guard bands compared to the intermediate precision approach, it led to considerable

more non-compliant results in the case of guarded acceptance and less non-compliant results in the case guarded rejection. The guard bands defined under the intermediate precision approach offered safer decision rules because of the fact that they use uncertainty estimates which represent the “true” dispersion of the values of the measurand (sulfur mass concentration). On the other hand the ISO 4259 approach may result in large number of samples classified incorrectly as compliant or non-compliant. Nevertheless, that approach may be useful for end users of fuel sample analysis results when uncertainty estimates are not available.

7.2 Applications

The program codes developed in MATLAB in order to apply the Monte Carlo method (adaptive and fixed trials) may be used in any type of measurement, if the mathematical model of the measurement is known and is inserted in the code (Chapters 4 and 5). Moreover the empirical approaches for the estimation of measurement uncertainty arising from sampling are also applicable to all measurement methods involving sampling (Chapter 3).

Based on the work presented in Chapter 6 the requirements and the key features of a software named “Decision Support System for the Evaluation of Conformity of Fuel Products” were defined. This software aims to support those who deliver, receive or for any reason, check fuel products and based on laboratory measurements should determine whether they comply with legislative requirements. The potential users of the software may belong in any stage of the supply chain of fuel (refineries, distribution companies, petroleum retail stations, large consumers, industries, airlines, shipping companies, etc.) and may have a buyer role (receiving fuel product) or seller role (delivering fuel product) or both. In addition, the software can be used by authorities carrying out official assessment of fuel conformity.

The key features of the software are as follows:

- The software will be available online (website) and available for installation on a PC or a Tablet / Smartphone (application)
- The user will enter the results of the laboratory measurements and will receive as information the probability of the fuel product compliance (or non compliance) with the relevant specifications (Figure 7.1).

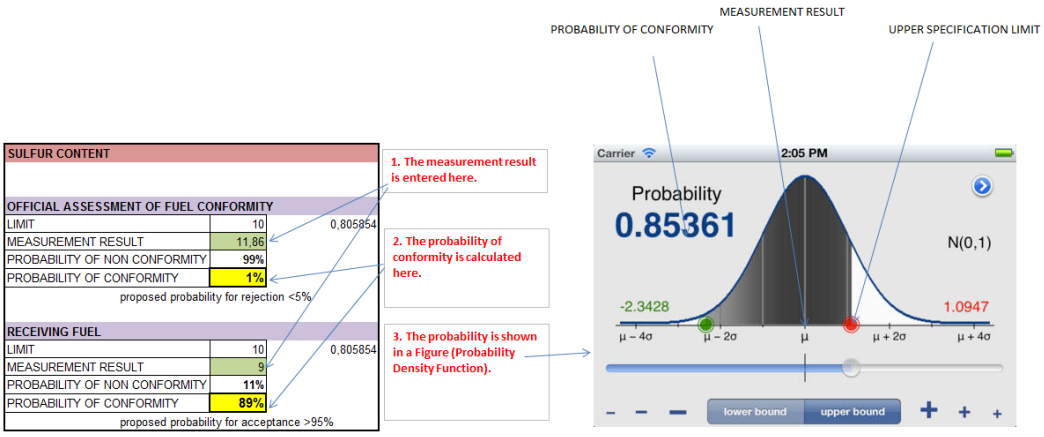


Figure 7.1 Example calculations of the probability of compliance or non-compliance.

- Depending on the type of user (buyer, seller, authority) different decision rules may be defined.
- The software will have already imported all the specifications of different types of fuels derived from the relevant legislation and standards (eg EN 228, EN 590) as well as the precision data of the standard methods (ISO, EN, ASTM) used for the determination of the fuel properties. The specifications and data will be updated every time there is a change either in the legislation or in a standard method.
- The software will keep a log of previous analysis results and export statistics.
- The software will enable the user to create validated official reports, with information on the quality of the tested fuel.

7.3 Future work

7.3.1 Bayesian uncertainty analysis

The GUM mixes elements from both Bayesian and classical statistics. Furthermore, the supplements 1 and 2 to the GUM made a shift towards the Bayesian point of view (they produce a probability distribution that shall encode one's state of knowledge about the measurand). It was shown in Chapter 4 that Bayesian treatment of Type A uncertainties “corrects” the underestimation of measurement uncertainty of the GUM approaches compared to the Monte Carlo method. Nevertheless, the “Bayesian scope” of GUM and its supplements is limited, and key features of a Bayesian uncertainty analysis, such as the use of informative prior

distributions, are missing. In order to utilize those features explicit use of Bayes' theorem is required and, consequently, different and more involved numerical techniques would have to be developed and employed.

7.3.2 Conformity assessment

In Chapter 6 we were concerned with “specific” conformity assessment, which was related to the decision for conformity based on testing single fuels samples from petroleum retails stations and not from a production process. In the case of samples from a production process, a “global” conformity assessment can be implemented that would take into account the variability of both the measuring system and the process (production system). Moreover, adding measures of economic impact, such as the costs of measurement and the costs incorrect decisions, can give more objective and more readily appreciated bases for decisions for all parties concerned. Such costs are associated with a variety of consequences, such as unnecessary re-manufacturing by the supplier as well as various consequences for the customer, arising from incorrect measures of quantity, poor product performance etc. Finally, it is recognized that common tool of statistics work readily for traditional quantitative measurements. Nevertheless, there is a growing need for development of statistical methodologies concerning the evaluation of measurement uncertainty and assessment of conformity, in the measurement of qualitative and subjective properties.

7.3.3 Sampling bias

In Chapter 3 we were concerned with sampling uncertainty without accounting for sampling bias. Although it is difficult to establish the presence of bias in a sampling protocol as it requires the definition of a reference sampling method or a reference sampling target, it would be useful to develop and apply methods that may be used for the estimation and the inclusion of analytical or sampling bias in the sampling uncertainty estimates.

Bibliography

[1] Directive 98/70/EC of the European Parliament and of the Council of 13 October 1998 relating to the quality of petrol and diesel fuels and amending Council Directive 93/12/EEC, Official Journal of the European Union, Brussels, Belgium.

[2] Directive 2003/17/EC of the European Parliament and of the Council of 3 March 2003 amending Directive 98/70/EC relating to the quality of petrol and diesel fuels.

[3] ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories, International Organization for Standardization (ISO), Geneva, Switzerland, 2005.

[4] ISO 15189, Medical laboratories - Requirements for quality and competence, International Organization for Standardization (ISO), Geneva, Switzerland, 2012.

[5] ISO/IEC 17043, Conformity assessment - General requirements for proficiency testing, International Organization for Standardization (ISO), Geneva, Switzerland, 2010.

[6] ISO Guide 34, General requirements for the competence of reference material producers, International Organization for Standardization (ISO), Geneva, Switzerland, 2009.

[7] ISO 9001 Quality management systems – Requirements, International Organization for Standardization (ISO), Geneva, Switzerland, 2008.

[8] JCGM 100 (GUM 1995 with minor corrections), Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, 2008 (available from www.bipm.org).

[9] JCGM 200, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition), Joint Committee for Guides in Metrology, 2012 (available from www.bipm.org).

[10] P. De Bièvre, “Quality of a measurement result” is established by means of metrological criteria, Accreditation and Quality Assurance 17 (2012) 479–480.

[11] R. Kaarls, BIPM Proceedings-Verb. Comité International des Poids et Mesures 49 (1981) A1-A12 (in French) / P. Giacomo, Metrologia 17 (1981) 73 -74 (in English).

[12] CIPM, BIPM Proceedings-Verb. Comité International des Poids et Mesures 49 (1981) 8-9, 26 (in French) / P. Giacomo, Metrologia 18 (1982) 43-44 (in English).

[13] CIPM, BIPM Proceedings-Verb. Comité International des Poids et Mesures 54(1986) 14, 35 (in French) / P. Giacomo, Metrologia 24 (1987) 49-50 (in English).

[14] JCGM 104, Evaluation of measurement data – An introduction to the "Guide to the expression of uncertainty in measurement" and related documents, Joint Committee for Guides in Metrology, 2009 (available from www.bipm.org).

[15] JCGM 101, Evaluation of measurement data - Supplement 1 to the “Guide to the expression of uncertainty in measurement" - Propagation of distributions using a Monte Carlo method, Joint Committee for Guides in Metrology, 2008 (available from www.bipm.org).

[16] JCGM 102, Evaluation of measurement data – Supplement 2 to the "Guide to the expression of uncertainty in measurement" – Extension to any number of output quantities, Joint Committee for Guides in Metrology, 2011 (available from www.bipm.org).

[17] JCGM 106, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment, Joint Committee for Guides in Metrology, 2012 (available from www.bipm.org).

[18] S.A. Bell, Measurement Good Practice Guide No. 11, A beginner's guide to uncertainty of measurement (Issue 2), National Physical Laboratory, Teddington, Middlesex, United Kingdom, 2001 (available from www.npl.co.uk).

[19] K. Birch, Measurement Good Practice Guide No. 36, Estimating uncertainties in testing, National Physical Laboratory, Teddington, Middlesex, United Kingdom, 2001 (available from www.npl.co.uk).

[20] B.N. Taylor and C.E. Kuyatt, NIST Technical Note 1297 - Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, National Institute of Standards and Technology (NIST), Gaithersburg, USA, 1994 (available from www.nist.gov).

[21] M3003, The Expression of Uncertainty and Confidence in Measurement (Edition 3), United Kingdom Accreditation Service (UKAS), Middlesex, UK, 2012 (available from www.ukas.com).

[22] ILAC-G17, Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025, International Laboratory Accreditation Cooperation (ILAC), Silverwater, Australia, 2002 (available from www.ilac.org).

[23] EA-4/02, Evaluation of the Uncertainty of Measurement in Calibration, European Co-operation for Accreditation (EA), 2013 (available from www.european-accreditation.org).

[24] EA-4/16, EA guidelines on the expression of uncertainty in quantitative testing, European Co-operation for Accreditation (EA), 2003 (available from www.european-accreditation.org).

[25] S.L.R. Ellison and A. Williams (Eds), EURACHEM /CITAC guide: Quantifying Uncertainty in Analytical Measurement (Edition 3) EURACHEM/CITAC, 2012 (available from www.eurachem.org).

[26] B. Magnusson, T. Näykki, H. Hovind and M. Krysell, NORDTEST report TR 537 “Handbook for calculation of measurement uncertainty in environmental laboratories” (Edition 3.1), Nordic Innovation Centre, Oslo, Norway, 2012 (available from www.nordtest.info).

[27] EUROLAB Technical Report, Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation, EUROLAB, Paris, France, 2007 (available from www.eurolab.org).

[28] C. Ehrlich, Terminological aspects of the Guide to the *Expression of Uncertainty in Measurement* (GUM), Metrologia 51 (2014) S145-S154.

[29] M.G. Cox and P.M. Harris, Validating the applicability of the GUM procedure, Metrologia 51 (2014) S167-S175.

[30] P.M. Harris and M.G. Cox, On a Monte Carlo method for measurement uncertainty evaluation and its implementation, Metrologia 51 (2014) S176-S182.

[31] S.L.R. Ellison, Implementing measurement uncertainty for analytical chemistry: the *Eurachem Guide* for measurement uncertainty, Metrologia 51 (2014) S199-S205.

[32] J. Kragten, Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique, Analyst 119 (1994) 2161-2166.

[33] ISO 5725-1, Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions, International Organization for Standardization (ISO), Geneva, Switzerland, 1994.

[34] ISO 5725-2, Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method, International Organization for Standardization (ISO), Geneva, Switzerland, 1994.

[35] ISO 5725-3, Accuracy (trueness and precision) of measurement methods and results - Part 3: Intermediate measures of the precision of a standard measurement method, International Organization for Standardization (ISO), Geneva, Switzerland, 1994.

[36] ISO 5725-4, Accuracy (trueness and precision) of measurement methods and results - Part 4: Basic methods for the determination of the trueness of a standard measurement method, International Organization for Standardization (ISO), Geneva, Switzerland, 1994.

[37] ISO 5725-5, Accuracy (trueness and precision) of measurement methods and results - Part 5: Alternative methods for the determination of the precision of a standard measurement method, International Organization for Standardization (ISO), Geneva, Switzerland,1998.

[38] ISO 5725-6, Accuracy (trueness and precision) of measurement methods and results - Part 6: Use in practice of accuracy values, International Organization for Standardization (ISO), Geneva, Switzerland, 1994.

[39] ISO 21748, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, International Organization for Standardization (ISO), Geneva, Switzerland, 2010.

[40] C. Elster, Bayesian uncertainty analysis compared with the application of the GUM and its supplements, Metrologia 51 (2014) S159-S166.

[41] C. Elster, W. Wöger and M.G. Cox, Draft GUM Supplement 1 and Bayesian analysis, Metrologia 44 (2007) L31-L32.

[42] C. Elster and B. Toman, Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using the Supplement 1 to the *Guide*: a comparison, Metrologia 46 (2009) 261-266.

[43] I. Lira and D. Grientschnig, Equivalence of alternative Bayesian procedures for evaluating measurement uncertainty, Metrologia 47 (2010) 334-336.

[44] R. Kacker and A. Jones, On use of Bayesian statistics to make the *Guide to the Expression of Uncertainty in Measurement* consistent, Metrologia 40 (2003) 235-248.

[45] Analytical Methods Committee, Measurement uncertainty evaluation for a non-negative measurand: an alternative to limit of detection, Accreditation and Quality Assurance 13 (2008) 29-32.

[46] W. Hasselbarth, Accounting for bias in measurement uncertainty estimation, Accreditation and Quality Assurance 9 (2004) 509-514.

[47] I. Lira and W. Wöger, Evaluation of the uncertainty associated with a measurement result not corrected for systematic effects, Measurement Science and Technology 9 (1998) 1010-1011.

[48] B. Magnusson and S.L.R. Ellison, Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements, Analytical and Bioanalytical Chemistry 390 (2008) 201-213.

[49] A. Kallner, L. Khorovskaya and T. Pettersson, A method to estimate the uncertainty of measurements in a conglomerate of instruments/laboratories, Scandinavian Journal of Clinical & Laboratory Investigation 65 (2005) 551–558.

[50] M. Thompson, S.L.R. Ellison, A. Fajgelj, P. Willetts and R. Wood, Harmonised guidelines for the use of recovery information in analytical measurement (Technical Report), Pure and Applied Chemistry 71 (1999) 337-348.

[51] V.J. Barwick and S.L.R. Ellison, Measurement uncertainty: Approaches to the evaluation of uncertainties associated with recovery, Analyst 124 (1999) 981-990.

[52] S.D. Phillips, K.R. Eberhardt and B. Parry, Guidelines for Expressing the Uncertainty of Measurement Results Containing Uncorrected Bias, Journal of Research of NIST 102 (1997) 577-585.

[53] A. Maroto, R. Boqué, J. Riu and F.X. Rius, Should non-significant bias be included in the uncertainty budget?, Accreditation and Quality Assurance 7 (2002) 90-94.

[54] V. Synek, Attempts to include uncorrected bias in the measurement uncertainty, Talanta 65 (2005) 829-837

[55] S.L.R. Ellison and A. Williams (Eds), EURACHEM/CITAC guide: Use of uncertainty information in compliance assessment (Edition 1), EURACHEM/CITAC, 2007 (available from www.eurachem.org).

[56] L.R. Pendrill, Using measurement uncertainty in decision-making and conformity assessment, Metrologia 51 (2014) S206-S218.

[57] D.B. Hibbert, The uncertainty of a result from a linear calibration, Analyst 131 (2006) 1273–1278.

[58] V.J. Barwick, Preparation of calibration curves—a guide to best practice, VAM, Teddington UK, 2003.

[59] L. Cuadros-Rodríguez, L. Gámiz-Gracia, E. Almansa-López, J. Laso-Sánchez Calibration in chemical measurement processes: I. A metrological approach, TrAC - Trends in Analytical Chemistry 20 (2001) 195–206.

[60] K. Heydorn and K. Esbensen, Sampling and metrology, Accreditation and Quality Assurance 9 (2004) 391-396.

[61] ASTM D 4057-06, Standard Practice for Manual Sampling of Petroleum and Petroleum Products, ASTM International, West Conshohocken, PA, 2011

[62] J.B. Hansen, C. Grøn, U. Lund, B. Magnusson, A. Nordbotten, A. Oberender and M.H. Ramsey, Uncertainty from sampling: workshop to launch a Nordtest handbook on sampling uncertainty estimation and control, Accreditation and Quality Assurance 12 (2007) 377-381.

[63] K.A. Boon, P. Rostron and M.H. Ramsey, An Exploration of the Interplay between the Measurement Uncertainty and the Number of Samples in Contaminated Land Investigations, Geostandards and Geoanalytical Research 35 (2011) 353-367.

[64] M.H. Ramsey and A. Argyraki, Estimation of measurement uncertainty from field sampling: implications for the classification of contaminated land, Science of the Total Environment 198 (1997) 243-257.

[65] M. Thompson, Uncertainty of sampling in chemical analysis, Accreditation and Quality Assurance 3 (1998) 117-121.

[66] C. Grøn, J.B. Hansen, B. Magnusson, A. Nordbotten, M. Krysell, K.J. Andersen and U. Lund, NORDTEST report TR 604 “Uncertainty from sampling. A Nordtest handbook for sampling planners on sampling quality assurance and uncertainty estimation”, Nordic Innovation Centre, Oslo, Norway, 2007 (available from www.nordtest.info).

[67] M.H. Ramsey and M. Thompson, Uncertainty from sampling, in the context of fitness for purpose, Accreditation and Quality Assurance 12 (2007) 503-513.

[68] M.H. Ramsey, When is sampling part of the measurement process, Accreditation and Quality Assurance 9 (2004) 727-728.

[69] M.H. Ramsey and S.L.R. Ellison (eds), EURACHEM/ EUROLAB/ CITAC/ NORDTEST/ AMC Guide "Measurement Uncertainty arising from sampling: a guide to methods and approaches" (Edition 1), EURACHEM/CITAC, 2007 (available from www.eurachem.org).

[70] J.A. Lyn, M.H. Ramsey, A.P. Damant and R. Wood, Empirical *versus* modelling approaches to the estimation of measurement uncertainty caused by primary sampling, Analyst 132 (2007) 1231-1237.

[71] M. H. Ramsey, Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources, Journal of Analytical Atomic Spectrometry 13 (1998) 97-104.

[72] J. Galbán and C. Ubide, Uncertainty due to the quantification step in analytical methods, Talanta 71 (2007) 1339-1344.

[73] ASTM D 5453-09, Standard Test Method for Determination of Total Sulfur in Light Hydrocarbons, Spark Ignition Engine Fuel, Diesel Engine Fuel, and Engine Oil by Ultraviolet Fluorescence, ASTM International, West Conshohocken, PA, 2009

[74] E.V. Reiter, M.F. Dutton, A. Agus, E. Nordkvist, M.F. Mwanza, P.B. Njobeh, D. Prawano, P. Häggblom, E. Razzazi-Fazeli, J. Zentek and M.G. Andersson, Uncertainty from sampling in measurements of aflatoxins in animal feedingstuffs: application of the Eurachem/CITAC guidelines, Analyst 136 (2011) 4059-4069.

[75] ISO 20846, Petroleum products - Determination of sulfur content of automotive fuels - Ultraviolet fluorescence method, International Organization for Standardization (ISO), Geneva, Switzerland, 2011.

[76] C.E. Efstathiou, On the sampling variance of ultra-dilute solutions, Talanta 52 (2000) 711-715.

[77] S.L.R. Ellison, V.J. Barwick and T.J.D. Farrant, Practical Statistics for the Analytical Scientist: A Bench Guide (Edition 2), The Royal Society of Chemistry, Cambridge, UK, 2009.

[78] Royal Society of Chemistry – Analytical Methods Committee, Robust statistics–how not to reject outliers. Part 1. Basic concepts, Analyst 114 (1989), 1693-1697.

[79] M.H. Ramsey, M. Thompson and M. Hale, Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance, Journal of Geochemical Exploration 44 (1992) 23-36.

[80] ROBAN, Royal Society of Chemistry – Analytical Methods Committee, http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/ROBAN.asp (Accessed March 11, 2015).

[81] ISO 3085, Iron ores - Experimental methods for checking the precision of sampling, sample preparation and measurement, International Organization for Standardization (ISO), Geneva, Switzerland, 2002.

[82] H. Hovind, B. Magnusson, M. Krysell, U. Lund and I. Mäkinen, NORDTEST report TR 569 "Internal Quality Control – Handbook for Chemical Laboratories" (Edition 4), Nordisk Innovation Centre, Oslo, Norway, 2011 (available from www.nordtest.info).

[83] PASW Statistics 18, release Version 18.0.0, SPSS, Inc., 2009, Chicago, IL, www.spss.com

[84] P.D. Rostron and M.H. Ramsey, Cost effective, robust estimation of measurement uncertainty from sampling using unbalanced ANOVA, Accreditation and Quality Assurance 17 (2012) 7-14

[85] M.H. Ramsey, B. Geelhoed,R.Woodb and A. P. Damant, Improved evaluation of measurement uncertainty from sampling by inclusion of between-sampler bias using sampling proficiency testing, Analyst 136 (2011) 1313-1321.

[86] H.J. Albert, New concepts in bomb calorimeter design and operation, Thermochimica Acta 310 (1998) 243-251.

[87] M. Cox, P. Harris and B.R.-L. Siebert, Evaluation of Measurement Uncertainty based on the propagation of distributions using Monte Carlo Simulation, Measurement Techniques 46 (2003) 824-833.

[88] 2004/156/EC, Commission Decision of 29 January 2004 establishing guidelines for the monitoring and reporting of greenhouse gas emissions pursuant to Directive 2003/87/EC of the European Parliament and of the Council, Official Journal of the European Union, Brussels, Belgium.

[89] W. Kessel, Measurement uncertainty according to ISO/BIPM-GUM, Thermochimica Acta 382 (2002) 1-16.

[90] W. Bich, M.G. Cox, P.M. Harris, Evolution of the 'Guide to the Expression of Uncertainty in Measurement', Metrologia 43 (2006) S161-S166.

[91] ASTM D240-14, Standard test method for heat of combustion of liquid hydrocarbon fuels by bomb calorimeter, 2014, ASTM International, West Conshohocken, PA, USA.

[92] MATLAB The Mathworks, http://www.mathworks.com (accessed March 11, 2014)

[93] Parr Operating Instruction Manual No. 442M, 6200 Oxygen Bomb Calorimeter, Parr Instruments Company, Illinois, USA, 2010.

[94] K.M. Decker, The Monte Carlo method in science and engineering: Theory and application, Computer Methods in Applied Mechanics and Engineering 89 (1991) 463-483.

[95] H.W. Coleman and W.G. Steel, Experimentation, Validation, and Uncertainty Analysis for Engineers (Edition 3), John Wiley & Sons, Inc., Hoboken, New Jersey, 2009.

[96] M. Solaguren-Beascoa Fernández, J.M. Alegre Calderón and P.M. Bravo Díez, Implementation in MATLAB of the adaptive Monte Carlo method for the evaluation of measurement uncertainties, Accreditation and Quality Assurance 14 (2009) 95-106.

[97] M. Solaguren-Beascoa Fernández, On the use of the 'uncertainty budget' to detect dominant terms in the evaluation of measurement uncertainty, Accreditation and Quality Assurance 14 (2011) 83-88.

[98] B.D. Hall and R. Willink, Does "Welch-Satterthwaite" make a good uncertainty estimate?, Metrologia 38 (2001) 9-15.

[99] M. Ballico, Limitations of the Welch-Satterthwaite approximation for measurement uncertainty calculations, Metrologia 37 (2000) 61-64.

[100] ASTM D5453-12, Standard Test Method for Determination of Total Sulfur in Light Hydrocarbons, Spark Ignition Engine Fuel, Diesel Engine Fuel, and Engine Oil by Ultraviolet Fluorescence, 2012, ASTM International, West Conshohocken, PA, USA.

[101] D.B. Hibbert and J.J. Gooding, Data Analysis for Chemistry. An Introductory Guide for Students and Laboratory Scientists, Oxford University Press, New York, 2006.

[102] S.L.R. Ellison, In defence of the correlation coefficient, Accreditation and Quality Assurance 11 (2006) 146-152.

[103] ISO/TS 28037, Determination and use of straight-line calibration functions, International Organization for Standardization (ISO), Geneva, Switzerland, 2010.

[104] EN 228, Automotive fuels – Unleaded petrol – Requirements and test methods, European Committee for Standardization (CEN), Brussels, Belgium, 2008.

[105] EN 590, Automotive fuels – Diesel – Requirements and test methods, European Committee for Standardization (CEN), Brussels, Belgium, 2009.

[106] ISO 10576-1, Statistical methods - Guidelines for the evaluation of conformity with specified requirements - Part 1: General principles, International Organization for Standardization (ISO), Geneva, Switzerland, 2003.

[107] P. Carbone, D. Macii and D. Petri, Measurement uncertainty and metrological confirmation in quality-oriented organizations, Measurement 34 (2003) 263–271.

[108] G.B. Rossi and F. Crenna, A probabilistic approach to measurement-based decisions, Measurement 39 (2006) 101–119.

[109] ISO/IEC 17000, Conformity assessment -Vocabulary and general principles, International Organization for Standardization (ISO), Geneva, Switzerland, 2004.

[110] M.L.J. Weitzel and W.M. Johnson, Using target measurement uncertainty to determine fitness for purpose, Accreditation and Quality Assurance 17 (2012) 491-495.

[111] ASME B89.7.3.1, Guidelines for Decision Rules: Considering Measurement Uncertainty in Determining Conformance with Specifications, American Society for Mechanical Engineers (ASME), Fairfield, NJ, USA, 2001.

[112] E. Desimoni and B. Brunetti, Uncertainty of measurement and conformity assessment: a review, Analytical and Bioanalytical Chemistry 400 (2011) 1729–1741.

[113] A. Williams, Principles of the EURACHEM/CITAC guide "Use of uncertainty information in compliance assessment", Accreditation and Quality Assurance 13 (2008) 633–638.

[114] L.R. Pendrill, Operating 'cost' characteristics in sampling by variable and attribute, Accreditation and Quality Assurance 13 (2008) 619–631.

[115] ISO 14253-1, Geometrical Product Specifications (GPS) - Inspection by measurement of workpieces and measuring equipment - Part 1: Decision rules for proving conformance or non-conformance with specifications, International Organization for Standardization (ISO), Geneva, Switzerland, 1998.

[116] ILAC-G8:03, Guidelines on the Reporting of Compliance with Specification, International Laboratory Accreditation Cooperation (ILAC), Silverwater, Australia, 2009 (available from www.ilac.org).

[117] WADA Technical Document – TD2010DL, Decision limits for the confirmatory quantification of threshold substance**s,** World Anti-Doping Agency (WADA)**,** Montreal (Quebec), Canada, 2010.

[118] E. Desimoni and B. Brunetti, About acceptance and rejection zones as defined in the EURACHEM/CITAC Guide (2007) "Use of uncertainty information in compliance assessment", Accreditation and Quality Assurance 15 (2010) 45–47.

[119] ISO 4259, Petroleum products - Determination and application of precision data in relation to methods of test, International Organization for Standardization (ISO), Geneva, Switzerland, 2006.

[120] G. Beges, J. Drnovsek and L.R. Pendrill, Optimising calibration and measurement capabilities in terms of economics in conformity assessment, Accreditation and Quality Assurance 15 (2010) 147–154.

[121] L.R. Pendrill, Optimised measurement uncertainty and decision-making when sampling by variables or by attribute, Measurement 39 (2006) 829–840.

[122] L.R. Pendrill and H. Källgren, Exhaust gas analysers and optimised sampling, uncertainties and costs, Accreditation and Quality Assurance 11 (2006) 496–505.

[123] N. Liapis, D. Theodorou and F. Zannikos, Absence of TQM across the fuel supply chain: quality failure-associated costs, Total Quality Management & Business Excellence 24 (2013) 452-461.

Annex A: Extended Abstract in Greek

Τεχνικές Ανάλυσης Δεδομένων στα Συστήματα Διαχείρισης της Ποιότητας. Εφαρμογή στις Αναλύσεις Καυσίμων.

Περιεχόμενα Εκτεταμένης Περίληψης

Α. ΕΙΣΑΓΩΓΗ

Α.1 Κίνητρα και στόχοι

Καθώς η αγορά των καυσίμων γίνεται όλο και περισσότερο διαφοροποιημένη, ο έλεγχος της ποιότητας των καυσίμων έχει μετατραπεί σε μια βασική και υψηλής προστιθέμενης αξίας δραστηριότητα καθώς συμβάλει στη διαχείριση των κινδύνων για παραγωγούς, εμπόρους, διανομείς και καταναλωτές – τελικούς χρήστες. Τα καύσιμα που παράγονται και διατίθενται στην αγορά πρέπει να συμμορφώνονται με αυστηρές απαιτήσεις που προδιαγράφονται από τη σχετική νομοθεσία. Στην Ευρωπαϊκή Ένωση για παράδειγμα, διάφορες οδηγίες θέτουν τεχνικές προδιαγραφές για τα καύσιμα των αυτοκινήτων που χρησιμοποιούνται σε κινητήρες επιβαλλόμενης ανάφλεξης (βενζίνη) ή σε κινητήρες ανάφλεξης με συμπίεση (ντήζελ).

Μέσω της εφαρμογής ευρείας γκάμας μεθόδων ανάλυσης γίνεται αξιολόγηση και εκτίμηση των φυσικών, μηχανικών, ρεολογικών, θερμικών και χημικών ιδιοτήτων του αργού πετρελαίου των προϊόντων επεξεργασίας του (λιπαντικά, βενζίνες, ντήζελ, αεροπορικά καύσιμα, μαζούτ), καθώς και άλλων καυσίμων (π.χ. στερεών, αερίων) και ενεργειακών πόρων που χρησιμοποιούνται ως καύσιμα (π.χ. βιομάζα). Τα αποτελέσματα των αναλύσεων των καυσίμων θα πρέπει να είναι αξιόπιστα και αποδεκτά από όλα τα ενδιαφερόμενα μέρη, έτσι ώστε να μειωθούν οι κίνδυνοι προϊόντων εκτός προδιαγραφών, να συντομεύσει ο χρόνος διάθεσης των καυσίμων στην αγορά και να αποδεικνύεται η ποιότητα και η ασφάλεια αυτών.

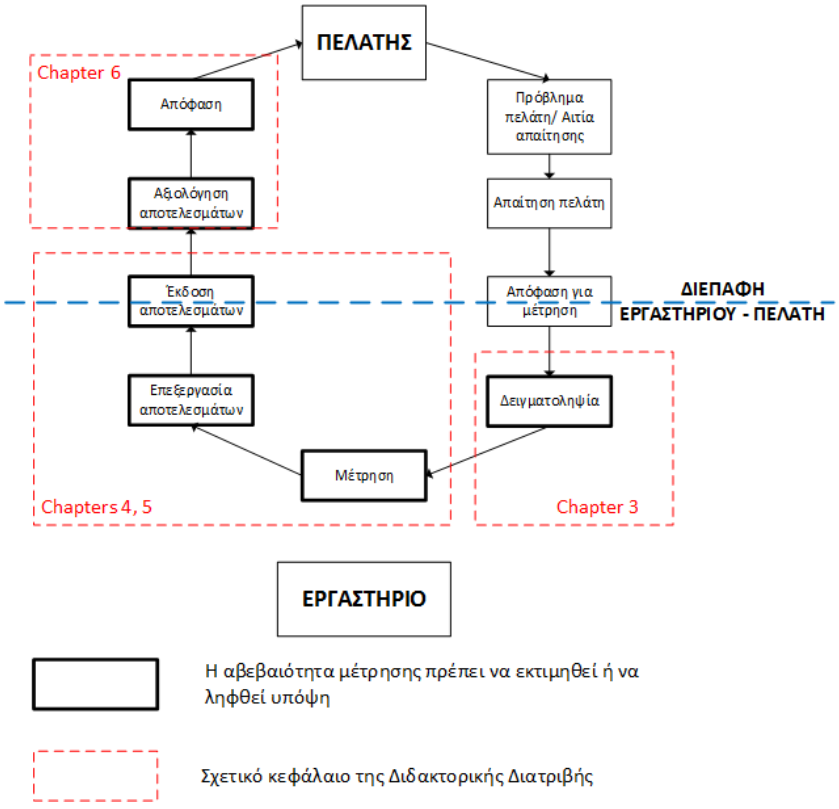
Γενικά, τα εργαστήρια διενεργούν αναλύσεις, δοκιμές και ελέγχους τα αποτελέσματα των οποίων παρέχονται σε κάποιον τρίτο (π.χ. τον πελάτη), ο οποίος θα τα χρησιμοποιήσει για να επιλύσει κάποιο πρόβλημα ή για να καταλήξει σε κάποιο συμπέρασμα. Ένα λάθος αποτέλεσμα μπορεί να έχει τεράστιες οικονομικές ή κοινωνικές επιπτώσεις καθώς είναι πιθανόν να έχει ως συνέπεια τη λήψης μιας εσφαλμένης απόφασης. Κατά συνέπεια, ένα εργαστήριο θα πρέπει να παρέχει υψηλής ποιότητας υπηρεσίες στους πελάτες του. Η έννοια «ποιότητα» σε ένα εργαστήριο δεν είναι κατ' ανάγκη η παροχή αποτελεσμάτων που χαρακτηρίζονται από τη μέγιστη δυνατή ακρίβεια. Ποιότητα σημαίνει η παροχή εργαστηριακών αποτελεσμάτων που είναι κατάλληλα για τη σκοπούμενη χρήση (fit for purpose). Αυτό επιτυγχάνεται μέσω:

- της κάλυψης των ειδικών απαιτήσεων των πελατών.

- της εμφύσησης εμπιστοσύνης στο πελάτη και σε όλα τα ενδιαφερόμενα μέρη που θα κάνουν χρήση των εργαστηριακών αποτελεσμάτων.
- τη βελτιστοποίηση της σχέσης τιμής – ποιότητας.

Η αξιολόγηση της «καταλληλόλητας για χρήση» μιας μεθόδου δοκιμής είναι άρρηκτα συνδεδεμένη με την εκτίμηση της αβεβαιότητας της μέτρησης η οποία ουσιαστικά χαρακτηρίζει την ποιότητα ενός αποτελέσματος συνυπολογίζοντας τόσο συστηματικά όσο και τυχαία σφάλματα.

Ο κύκλος μιας μέτρησης ξεκινά από μια απαίτηση που διατυπώνεται για συγκεκριμένο λόγο από τον πελάτη και κλείνει με την λήψη μιας απόφασης βάσει των εργαστηριακών αποτελεσμάτων. Σε αυτό τον κύκλο υπάρχουν σημαντικά στάδια (προ-αναλυτικά, αναλυτικά και μετα-αναλύτικά) όπου η αβεβαιότητα της μέτρησης εμπλέκεται σε μεγάλο βαθμό. Το Σχήμα Α.1 δείχνει τα διάφορα στάδια του κύκλου μέτρησης, υποδεικνύοντας τα στάδια όπου η αβεβαιότητα της μέτρησης πρέπει είτε να εκτιμηθεί είτε να ληφθεί υπόψη, καθώς και τα αντίστοιχα κεφάλαια της παρούσας Διδακτορικής Διατριβής που περιλαμβάνουν θέματα σχετικά με την εκτίμηση και τη χρήση της αβεβαιότητας.



Σχήμα Α.1 Κύκλος μέτρησης. Στάδια όπου η αβεβαιότητα της μέτρησης πρέπει είτε να εκτιμηθεί είτε να ληφθεί υπόψη – Σύνδεση με τα αντίστοιχα κεφάλαια της παρούσας Διδακτορικής Διατριβής.

A.2 Συστήματα διαχείρισης ποιότητας

Τα εργαστήρια που διενεργούν δοκιμές και διακριβώσεις καθώς και άλλοι οργανισμοί που υποστηρίζουν τα εργαστήρια στις δραστηριότητές τους (π.χ. διοργανωτές διεργαστηριακών δοκιμών ικανότητας, παραγωγοί υλικών αναφοράς) πρέπει να εφαρμόζουν συστήματα διαχείρισης ποιότητας που ως στόχο έχουν την απόδειξη της τεχνικής τους επάρκειας. Η ορθή εφαρμογή ενός τέτοιου συστήματος ποιότητας επιβεβαιώνεται από έναν ανεξάρτητο οργανισμό (φορέα διαπίστευσης).

Τα βασικά πρότυπα με τα οποία οι φορείς διαπίστευσης διεθνώς, διαπιστεύουν εργαστήρια και οργανισμούς είναι:

- το ISO/IEC 17025, για τα εργαστήρια δοκιμών και διακριβώσεων
- το ISO 15189, για τα ιατρικά εργαστήρια
- το ISO/IEC 17043, για τους διοργανωτές διεργαστηριακών δοκιμών ικανότητας
- το ISO Guide 34, για τους παραγωγούς υλικών αναφοράς

Μια βασική απαίτηση των παραπάνω προτύπων διαπίστευσης είναι η εκτίμηση της αβεβαιότητας των μετρήσεων με τη χρήση επιστημονικά τεκμηριωμένης και έγκυρης μεθοδολογίας. Η εκτίμηση αυτή στη πλειονότητα των περιπτώσεων απαιτεί μεταξύ άλλων επεξεργασία εργαστηριακών μετρήσεων.

A.3 Δομή διδακτορικής διατριβής

Στα κεφάλαια της παρούσης Διδακτορικής Διατριβής παρουσιάζεται η ανάπτυξη και εφαρμογή στατιστικών και αριθμητικών μεθόδων για την εκτίμηση και χρήση της αβεβαιότητας μετρήσεων σε συγκεκριμένα στάδια του κύκλου της μέτρησης (Σχήμα Α.1). Συγκεκριμένα:

- Το **Κεφάλαιο 2** περιγράφει το θεωρητικό υπόβαθρο όσον αφορά τις προσεγγίσεις-μεθοδολογίες για την εκτίμηση και τη χρήση της αβεβαιότητας.
- Το **Κεφάλαιο 3** παρουσιάζει την ανάπτυξη και εφαρμογή μεθοδολογιών για την εκτίμηση της αβεβαιότητας που προκύπτει λόγω δειγματοληψίας.
- Το **Κεφάλαιο 4** επικεντρώνεται στην εκτίμηση της αβεβαιότητας μιας μετρητικής διαδικασίας χρησιμοποιώντας τη μέθοδο GUM και τη μέθοδο προσομοίωσης Monte Carlo.

- Το **Κεφάλαιο 5** εξετάζει την εκτίμηση της αβεβαιότητας σε αναλυτικές μεθόδους που χρησιμοποιούν καμπύλη βαθμονόμησης η οποία έχει κατασκευαστεί μέσω γραμμικής παλινδρόμησης.
- Το **Κεφάλαιο 6** διερευνά τη χρήση της αβεβαιότητας των μετρήσεων και των δεδομένων επίδοσης των μεθόδων στην αξιολόγηση της συμμόρφωσης των προϊόντων.
- Το **Κεφάλαιο 7** συνοψίζει τα αποτελέσματα που έχουν επιτευχθεί και προτείνει περιοχές μελλοντικής ερευνητικής δραστηριότητας.

A.4 Λίστα δημοσιεύσεων

Στα πλαίσια της παρούσας διδακτορικής διατριβής πραγματοποιήθηκαν οι παρακάτω δημοσιεύσεις σε διεθνή περιοδικά με κριτές:

(i) D. Theodorou, Y. Zannikou, G. Anastopoulos, F. Zannikos. Coverage interval estimation of the measurement of Gross Heat of Combustion of fuel by bomb calorimetry: Comparison of ISO GUM and adaptive Monte Carlo method. (2011) **Thermochimica Acta** 526: 122– 129

(ii) D. Theodorou, Y. Zannikou, F. Zannikos. Estimation of the standard uncertainty of a calibration curve: application to sulfur mass concentration determination in fuels. **Accreditation and Quality Assurance** (2012) 17: 275–281

(iii) D. Theodorou, N. Liapis, F. Zannikos. Estimation of measurement uncertainty arising from manual sampling of fuels. **Talanta** (2013) 105: 360-365

(iv) D. Theodorou, F. Zannikos. The use of measurement uncertainty and precision data in conformity assessment of automotive fuel products. **Measurement** (2014) 50: 141-151

(v) D. Theodorou, Y. Zannikou, F. Zannikos. Components of measurement uncertainty from a measurement model with two stages involving two output quantities. **Chemometrics and Intelligent Laboratory Systems** (2015) 146: 305–312

Β. ΕΚΤΙΜΗΣΗ ΑΒΕΒΑΙΟΤΗΤΑΣ – ΓΕΝΙΚΑ

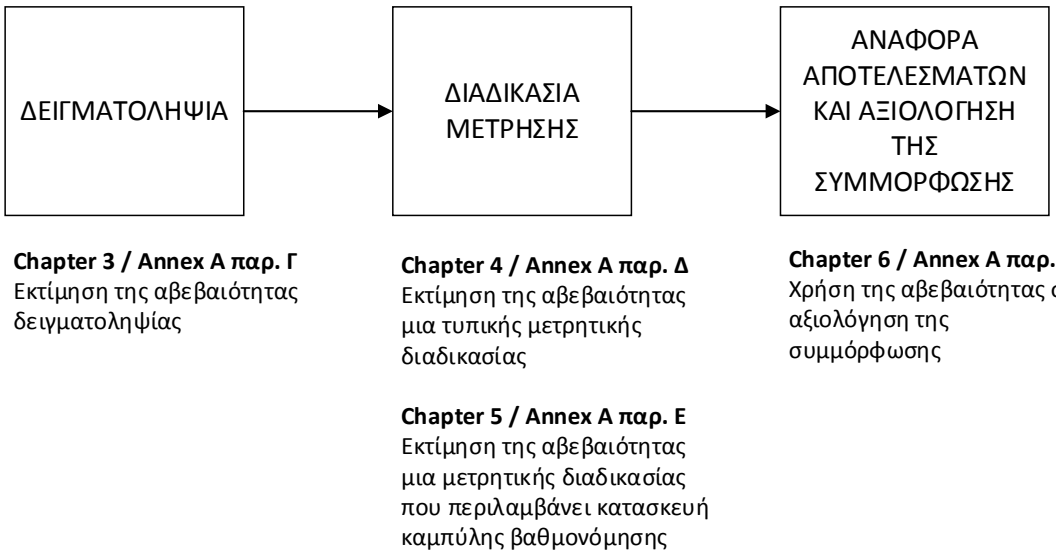
Β.1 Αβεβαιότητα και μέτρηση

Η αξιολόγηση της ορθότητας του αποτελέσματος μιας μέτρησης απασχολεί διαχρονικά όλους τους κλάδους των επιστημών. Στο παρελθόν, πηγές συστηματικών και τυχαίων μετρητικών σφαλμάτων θεωρούνταν η αιτία του λιγότερο ή περισσότερο ακριβούς προσδιορισμού μιας ποσότητας. Βάσει αυτής της άποψης τα μετρητικά σφάλματα περιγράφουν την απόκλιση μιας μέτρησης από την «πραγματική τιμή» μιας ποσότητας. Δεδομένου όμως ότι ο προσδιορισμός της «πραγματικής τιμής» μιας ποσότητας είναι και ο σκοπός πραγματοποίησης της μέτρησης, ο ορισμός των μετρητικών σφαλμάτων σε σχέση με αυτή την άγνωστη «πραγματική τιμή» προκαλεί εννοιολογικά προβλήματα στην κατανόηση της ορθότητας του αποτελέσματος. Αναγνωρίζοντας αυτό το πρόβλημα καθώς επίσης και το γεγονός ότι δεν υπάρχει ένας και μοναδικός τρόπος ποσοτικής έκφρασης της αμφιβολίας για την ορθότητα ενός μετρητικού αποτελέσματος, η Διεθνής Συνέλευση Μέτρων και Σταθμών (CIPM), που είναι η ανώτατη επιτροπή για θέματα μετρολογίας, το 1978 ζήτησε από το εκτελεστικό της όργανο, το Διεθνές Γραφείο Μέτρων και Σταθμών (BIPM), να καθορίσει συμφωνημένες θεμελιώδεις αρχές που θα αφορούσαν στο πρόβλημα έκφρασης και υπολογισμού της διασποράς των τιμών ενός μετρούμενου μεγέθους (measurand) σε σχέση με την καλύτερη εκτίμησή του. Ταυτόχρονα, υιοθετήθηκε ο όρος «αβεβαιότητα μέτρησης» για την περιγραφή αυτής της διασποράς των τιμών. Αποτέλεσμα αυτής της προσπάθειας ήταν η έκδοση της σύστασης INC-1 το 1980. Οι γενικές αρχές που περιέχονταν στην παραπάνω σύσταση διαμορφώθηκαν στη συνέχεια σε μία πρακτική τεχνική οδηγία από την Τεχνική Συμβουλευτική Επιτροπή στην Μετρολογία (TAG 4) του διεθνούς οργανισμού ISO σε συνεργασία με το BIPM, το IEC και το OIML και το 1993 εκδόθηκε το “Guide to the Expression of Uncertainty in Measurement” ή “GUM”.

Η αβεβαιότητα σύμφωνα και με το Λεξικό Όρων Μετρολογίας (ISO VIM) ορίζεται ως «μια παράμετρος (μη αρνητική) που χαρακτηρίζει την διασπορά των τιμών που θα μπορούσαν να αποδοθούν στην μετρούμενη ποσότητα βάσει της διαθέσιμης πληροφορίας για αυτή». Σήμερα είναι κοινά αποδεκτό ότι μια ποσοτική δήλωση σχετική με οποιοδήποτε μετρούμενο μέγεθος δεν μπορεί να είναι πλήρης εάν δεν περιλαμβάνει, εκτός από το αποτέλεσμα της μέτρησης, και μια αναφορά στην αβεβαιότητα που συνοδεύει το αποτέλεσμα αυτό. Το να συμπεριλαμβάνεται η αβεβαιότητα στις πληροφορίες ποσοτικού προσδιορισμού ενός μεγέθους εξυπηρετεί ένα διττό σκοπό: αφενός επισημαίνεται στο χρήστη της μέτρησης η

πιθανότητα ύπαρξης σφαλμάτων, εφιστώντας την προσοχή του στον πεπερασμένο χαρακτήρα της γνώσης μας για μια συγκεκριμένη ποσότητα, αφετέρου δίνεται μια ποσοτική εκτίμηση του διαστήματος μέσα στο οποίο περιέχεται η αληθής τιμή του μετρούμενου μεγέθους, καθώς και της πιθανότητας να βρίσκεται η αληθής αυτή τιμή σε μια συγκεκριμένη περιοχή του διαστήματος αυτού. Επιπλέον, η αβεβαιότητα επηρεάζει την ερμηνεία των αποτελεσμάτων όταν για παράδειγμα, με δεδομένο το αποτέλεσμα μιας εργαστηριακής μέτρησης αξιολογείται η συμμόρφωση ή μη, ενός προϊόντος σε σχέση με κάποια προδιαγραφή.

Η αβεβαιότητα της μέτρησης θα πρέπει να εκτιμάται ή να χρησιμοποιείται σε όλα τα στάδια μιας μετρητικής διαδικασίας από την δειγματοληψία έως και την αναφορά – αξιολόγηση των αποτελεσμάτων. Στο Σχήμα Β.1 απεικονίζονται τα βασικά στάδια της μετρητικής διαδικασίας και η συσχέτιση τους με αντίστοιχα κεφάλαια της Διδακτορικής Διατριβής.



Σχήμα Β.1 Βασικά στάδια της μετρητικής διαδικασίας και συσχέτιση με αντίστοιχα κεφάλαια της Διδακτορικής Διατριβής.

Β.2 Προσεγγίσεις εκτίμησης αβεβαιότητας

Υπάρχουν δύο γενικές προσεγγίσεις-μέθοδοι για την εκτίμηση της αβεβαιότητας, η μέθοδος της μοντελοποίησης και η εμπειρική μέθοδος.

Η «μέθοδος μοντελοποίησης» προσδιορίζει όλες τις πηγές αβεβαιότητας ξεχωριστά, οι οποίες στη συνέχεια συνδυάζονται (διαδίδονται) μέσω ενός μαθηματικού

μοντέλου. Η ευρύτερα κατανοητή μέθοδος μοντελοποίησης βασίζεται στην Οδηγία GUM. Σύμφωνα με τη Οδηγία GUM, η αβεβαιότητα στο αποτέλεσμα μιας μέτρησης αποτελείται γενικά από πολλές συνιστώσες, οι οποίες μπορούν να κατηγοριοποιηθούν σε δύο είδη ανάλογα με τον τρόπο υπολογισμού τους: στις αβεβαιότητες Τύπου Α, που υπολογίζονται με στατιστικές μεθόδους, και στις αβεβαιότητες Τύπου Β, που υπολογίζονται με άλλα μέσα. Η κατηγοριοποίηση αυτή δεν αντιστοιχεί στη διάκριση μεταξύ «τυχαίων» και «συστηματικών» αβεβαιοτήτων. Οι συνιστώσες Τύπου Α προκύπτουν από τη διακύμανση (variance) ή την τυπική απόκλιση (standard deviation) και τους βαθμούς ελευθερίας του αποτελέσματος, ενώ είναι συχνά απαραίτητη και η γνώση της συνδιακύμανσης (covariance). Οι συνιστώσες Τύπου Β, παρά το ότι δεν προκύπτουν απευθείας από κάποια στατιστική επεξεργασία, οφείλουν να παρουσιάζονται με όρους τυπικής αβεβαιότητας. Η τυπική αυτή αβεβαιότητα μπορεί να θεωρηθεί ως προσέγγιση της αντίστοιχης διακύμανσης, η ύπαρξη της οποίας υφίσταται ως υπόθεση. Η τελική συνδυασμένη αβεβαιότητα (combined uncertainty) προκύπτει από το συνδυασμό όλων των επιμέρους συνιστωσών, εκφραζόμενων με τη μορφή τυπικών αποκλίσεων.

Παρόλο που η φιλοσοφία της Οδηγίας GUM στηρίζεται στην χρήση της συνδυασμένης τυπικής αβεβαιότητας ως μία γενική παράμετρο χαρακτηρισμού της ποιότητας της μέτρησης, στην περίπτωση της τεκμηρίωσης της συμμόρφωσης η παράμετρος αυτή δεν είναι η πιο κατάλληλη δεδομένου ότι επιπλέον είναι απαραίτητος ο καθορισμός ενός εύρους στο οποίο θα εμπεριέχεται ένα μεγάλο ποσοστό (π.χ. 95%) των εξαγόμενων τιμών y, συμβατών με τις συνθήκες μέτρησης του μετρούμενου μεγέθους Y. Το εύρος αυτό είναι απαραίτητο ιδιαίτερα στην περίπτωση λήψης αποφάσεων που αφορούν την ασφάλεια, την υγεία και τις εμπορικές συναλλαγές. Για τους παραπάνω λόγους στην μεθοδολογία GUM καθιερώθηκε η χρήση της αποκαλούμενης διευρυμένης αβεβαιότητας, U, η οποία δίνεται από το γινόμενο της συνδυασμένης τυπικής αβεβαιότητας $u(y)$ με έναν ένα συντελεστή κάλυψης k, $U=k\cdot u(y)$, έτσι ώστε να ορίζεται ένα διάστημα πιθανοτήτων $[y-U, y+U]$ εκατέρωθεν της εξαγόμενης εκτίμησης y στο οποίο εμπεριέχεται το απαιτούμενο μεγάλο ποσοστό τιμών του μετρούμενου μεγέθους. Το διάστημα αυτό αντιστοιχεί σε μία πιθανότητα κάλυψης ή επίπεδο εμπιστοσύνης p. Για κάποιο επιζητούμενο επίπεδο εμπιστοσύνης p, ο συντελεστής κάλυψης k προκύπτει από την κατανομή t-student για συγκεκριμένο αριθμό βαθμών ελευθερίας v, δηλ. $k=t_p(v)$.

Ένας εναλλακτικός τρόπος για την εκτίμηση της αβεβαιότητας είναι με τη χρήση προσομοίωσης Monte Carlo. Στα πλαίσια εφαρμογής της προσομοίωσης Monte Carlo (πιθανοκρατική μέθοδος) πραγματοποιείται συνδυασμός κατανομών πιθανοτήτων μέσω αριθμητικής προσομοίωσης σε αντίθεση με τη κλασσική μεθοδολογία GUM (ντετερμινιστική μέθοδος) όπου πραγματοποιείται συνδυασμός αβεβαιοτήτων. Η μέθοδος Monte Carlo είναι χρήσιμη για την επικύρωση των αποτελεσμάτων που δίνει η μεθοδολογία GUM, καθώς και για τις περιπτώσεις όπου δεν ισχύουν οι παραδοχές που γίνονται κατά την εφαρμογή της μεθοδολογίας GUM. Οι βασικές αρχές για την εκτίμηση της αβεβαιότητας μέσω της μεθόδου της μοντελοποίησης (εφαρμόζοντας είτε μέθοδο GUM είτε προσομοίωση Monte Carlo) μπορούν να επεκταθούν και στην εκτίμηση της αβεβαιότητας μετρητικών μοντέλων που περιλαμβάνουν πολλαπλά στάδια μέτρησης ή / και εμπεριέχουν περισσότερα του ενός μετρούμενα μεγέθη.

Από την άλλη πλευρά, υπάρχουν «εμπειρικές προσεγγίσεις», οι οποίες βασίζονται σε δεδομένα επιδόσεων των μεθόδων, τα οποία για να δώσουν αξιόπιστες εκτιμήσεις αβεβαιότητας θα πρέπει να έχουν ενσωματώσει επιδράσεις από όσο το δυνατόν περισσότερες πηγές αβεβαιότητας. Οι εμπειρικές προσεγγίσεις, οι οποίες δίνουν αποτελέσματα συμβατά με την Οδηγία GUM, είναι ιδιαίτερα κατάλληλες για τις περιπτώσεις όπου σημαντικές συνεισφορές στην αβεβαιότητα δεν είναι εύκολο να ενσωματωθούν στο μαθηματικό μετρητικό μοντέλο ή όπου πολλά εργαστήρια χρησιμοποιούν ουσιαστικά ταυτόσημες μεθόδους και εξοπλισμό.

Στις περιπτώσεις όπου εκ των προτέρων (prior) γνώσεις και πεποιθήσεις σχετικά με μια με μέτρηση είναι διαθέσιμες, αυτές μπορούν να αξιοποιηθούν με τη χρήση Μπεϋζιανής (Bayesian) στατιστικής μεθοδολογίας, η οποία σε συνδυασμό με το αποτέλεσμα και την αβεβαιότητα της μέτρησης θα μας δώσει μια αξιόπιστη εκ των υστέρων (posterior) κατανομή για το μετρούμενο μέγεθος. Οι αβεβαιότητες Τύπου Α ή οι αβεβαιότητες κοντά στο μηδέν μπορούν να εκτιμηθούν μέσω μιας Μπεϋζιανής προσέγγισης.

Υπάρχουν διάφορες προσεγγίσεις για την αξιολόγηση των εργαστηριακών και μετρητικών συστηματικών σφαλμάτων (bias) καθώς και για την μεταχείρισή τους σε σχέση με την εκτίμηση της αβεβαιότητας. Το συστηματικό σφάλμα και η αβεβαιότητα υπολογισμού του, όταν δεν πραγματοποιείται διόρθωση του αποτελέσματος, πρέπει να ενσωματώνεται στην αβεβαιότητα της μέτρησης.

Στο **Κεφάλαιο 2** της παρούσας Διδακτορικής Διατριβής περιγράφεται πλήρως το θεωρητικό υπόβαθρο όσον αφορά τις προσεγγίσεις-μεθοδολογίες για την εκτίμηση και τη χρήση της αβεβαιότητας.

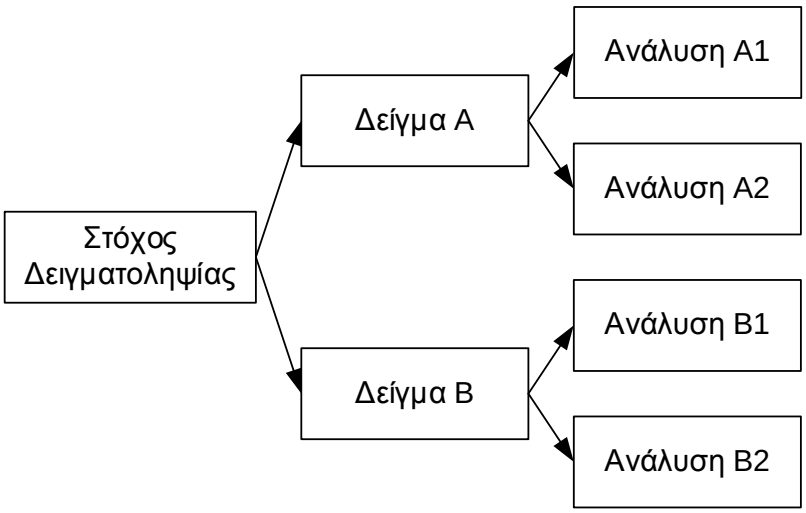
Γ. ΕΚΤΙΜΗΣΗ ΑΒΕΒΑΙΟΤΗΤΑΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΥΣΙΜΩΝ

Γ.1 Εισαγωγή

Η δειγματοληψία αποτελεί ένα βασικό στάδιο των μετρητικών διαδικασιών και συνεισφέρει σημαντικά στην αβεβαιότητα των εργαστηριακών μετρήσεων. Μια αξιόπιστη εκτίμηση της αβεβαιότητας λόγω δειγματοληψίας μπορεί να οδηγήσει σε καλύτερο έλεγχο των κινδύνων που συνδέονται με αποφάσεις περί συμμόρφωσης ή μη, ενός καυσίμου με προδιαγραφές που επιβάλει η νομοθεσία. Στο **Κεφάλαιο 3** της παρούσας Διδακτορικής Διατριβής περιγράφονται και συγκρίνονται ως προς τα αποτελέσματά τους, τρεις εμπειρικές στατιστικές μεθοδολογίες («κλασική» ANOVA, ανθεκτική ANOVA και στατιστική εύρους τιμών) χρησιμοποιώντας δεδομένα ενός ισορροπημένου πειραματικού σχεδίου (balanced experimental design). Οι τρεις μεθοδολογίες χρησιμοποιούνται για την εκτίμηση της αβεβαιότητας λόγω δειγματοληψίας καυσίμου (ντήζελ κίνησης) και λόγω της αναλυτικής διαδικασίας προσδιορισμού περιεκτικότητας σε θείο.

Γ.2 Πρωτόκολλο δειγματοληψίας και σχεδιασμός μετρήσεων

Υλοποιήθηκε ένα ισορροπημένο πειραματικό σχέδιο. Διπλά δείγματα ελήφθησαν από 104 πρατήρια υγρών καυσίμων, τα οποία επιλέχθηκαν τυχαία και αποτελούσαν το 10,9% των 950 πρατηρίων συγκεκριμένης εταιρείας τα οποία παρακολουθούνται από το εργαστήριο σε συστηματική βάση. Το σύστημα δειγματοληψίας – ανάλυσης που χρησιμοποιήθηκε παρουσιάζεται στο Σχήμα Γ.1. Τα διπλά δείγματα ελήφθησαν επαναλαμβάνοντας το ίδιο πρωτόκολλο δειγματοληψίας. Το πρωτόκολλο δειγματοληψίας που χρησιμοποιήθηκε ήταν βασισμένο στην πρότυπη μέθοδο ASTM D 4057 (2012a) που περιγράφει μεθοδολογίες μη αυτόματης δειγματοληψίας πετρελαίου και προϊόντων πετρελαίου.



Σχήμα Γ.1 Ισορροπημένο πειραματικό σχέδιο για την εκτίμηση της αβεβαιότητας της δειγματοληψίας.

Τα διπλά δείγματα αναλύθηκαν εις διπλούν κάτω από συνθήκες επαναληψιμότητας. Για τον προσδιορισμό της περιεκτικότητας σε θείο χρησιμοποιήθηκε ο αναλυτής θείου ANTEK 9000S, εξοπλισμένος με αυτόματο δειγματολήπτη. Ο εξοπλισμός αυτός συμμορφώνεται πλήρως με τα πρότυπα ASTM D 5453 (2012b) και ISO 20846 (2011) για τον προσδιορισμό θείου σε καύσιμα αυτοκινήτων. Η πλειοψηφία των αποτελεσμάτων των μετρήσεων των δειγμάτων βρέθηκε να περιέχει θείο λιγότερο από ή πολύ κοντά στα 10 mg kg^{-1}, το οποίο είναι το νομοθετικό (άνω) όριο της ΕΕ για την περιεκτικότητα του ντήζελ κίνησης σε θείο.

Γ.3 Μεθοδολογίες ανάλυσης δεδομένων

Το στατιστικό μοντέλο που περιγράφει τη σχέση μεταξύ μετρούμενης και πραγματικής τιμής της συγκέντρωσης μιας ουσίας, βάσει μια μέτρησης, *x*, έχει την ακόλουθη μορφή:

$$x = X_{\text{true}} + \varepsilon_{\text{sampling}} + \varepsilon_{\text{analysis}} \quad (\Gamma.1)$$

όπου X_{true}, είναι η πραγματική μέση συγκέντρωση της ουσίας στο σημείο δειγματοληψίας, $\varepsilon_{\text{sampling}}$, είναι το συνολικό σφάλμα λόγω δειγματοληψίας, που εκφράζεται από μια διακύμανση $\sigma^2_{\text{sampling}}$ και $\varepsilon_{\text{analysis}}$, είναι το συνολικό αναλυτικό σφάλμα, που εκφράζεται από μια διακύμανση, $\sigma^2_{\text{analysis}}$. Αν οι πηγές της διακύμανσης είναι ανεξάρτητες, τότε η συνολική διακύμανση της μέτρησης, $\sigma^2_{\text{measurement}}$, για ένα σημείο δειγματοληψίας δίνεται από τη σχέση:

$$\sigma^2_{\text{measurement}} = \sigma^2_{\text{sampling}} + \sigma^2_{\text{analysis}} \quad (\Gamma.2)$$

Αν χρησιμοποιηθούν οι στατιστικές εκτιμήτριες, s^2, για την προσέγγιση των διακυμάνσεων, σ^2, τότε η εξίσωση (Γ.2) γίνεται:

$$s^2_{\text{measurement}} = s^2_{\text{sampling}} + s^2_{\text{analysis}} \quad (\Gamma.3)$$

Τα συστατικά της διακύμανσης της μέτρησης, οι διακυμάνσεις της δειγματοληψίας και της ανάλυσης, που αντιπροσωπεύουν τις αντίστοιχες αβεβαιότητες, μπορούν να διαχωριστούν και να εκτιμηθούν χρησιμοποιώντας κατάλληλες στατιστικές μεθόδους:

- ***«Κλασική» ANOVA.*** Η ANOVA (ανάλυση διακύμανσης) είναι μία στατιστική τεχνική με την οποία οι διακυμάνσεις που προέρχονται από διαφορετικές πηγές μπορούν να απομονωθούν και να εκτιμηθούν. Ο πιο απλός τύπος ANOVA είναι η ANOVA κατά ένα παράγοντα (one way), η οποία εξετάζει μία ανεξάρτητη μεταβλητή και μία εξαρτημένη μεταβλητή. Η ANOVA μπορεί να εφαρμοστεί στα δεδομένα που παράγονται από την εφαρμογή ενός ισορροπημένου πειραματικού σχεδίου, προκειμένου να εκτιμηθεί η αβεβαιότητα δειγματοληψίας. Οι εκτιμήσεις της ANOVA βασίζονται στις διαφορές από τις μέσες τιμές.
- ***Ανθεκτική ANOVA.*** Η εφαρμογή της ανθεκτικής ANOVA που χρησιμοποιεί ανθεκτικά στατιστικά μεγέθη έχει αποδειχθεί ότι είναι ιδιαίτερα κατάλληλη για την εκτίμηση διακυμάνσεων, σε περιπτώσεις όπου η εγκυρότητα της «κλασικής» ANOVA είναι αμφισβητήσιμη (π.χ λόγω ύπαρξης ακραίων τιμών). Η ανθεκτική ANOVA χρησιμοποιεί ανθεκτικές εκτιμήσεις του μέσου όρου και της τυπικής απόκλισης που υπολογίζονται από μια επαναληπτική διαδικασία.
- ***Στατιστική εύρους τιμών.*** Η στατιστική εύρους τιμών, όπως και η «κλασική» ANOVA κάνει τη παραδοχή ύπαρξης κανονικής κατανομής ενώ οι υπολογισμοί γίνονται χρησιμοποιώντας διαφορές μεταξύ διπλών μετρήσεων. Στην πραγματικότητα η διακύμανση της δειγματοληψίας υπολογίζεται έμμεσα ως η διαφορά των διακυμάνσεων της μέτρησης και της ανάλυσης.

Για να υπολογιστεί ένα διάστημα κάλυψης (διευρυμένη αβεβαιότητα) που αντιστοιχεί σε πιθανότητα κάλυψης περίπου 95%, οι τυπικές αποκλίσεις, *s*, πολλαπλασιάζονται με ένα συντελεστή κάλυψης που ισούται με δύο (2). Η

διευρυμένη αβεβαιότητα μπορεί επίσης να εκφραστεί σε σχέση με την αναφερόμενη τιμή *x* (ως ποσοστό), ως σχετική διευρυμένη αβεβαιότητα *U* (%).

Γ.4 Αποτελέσματα

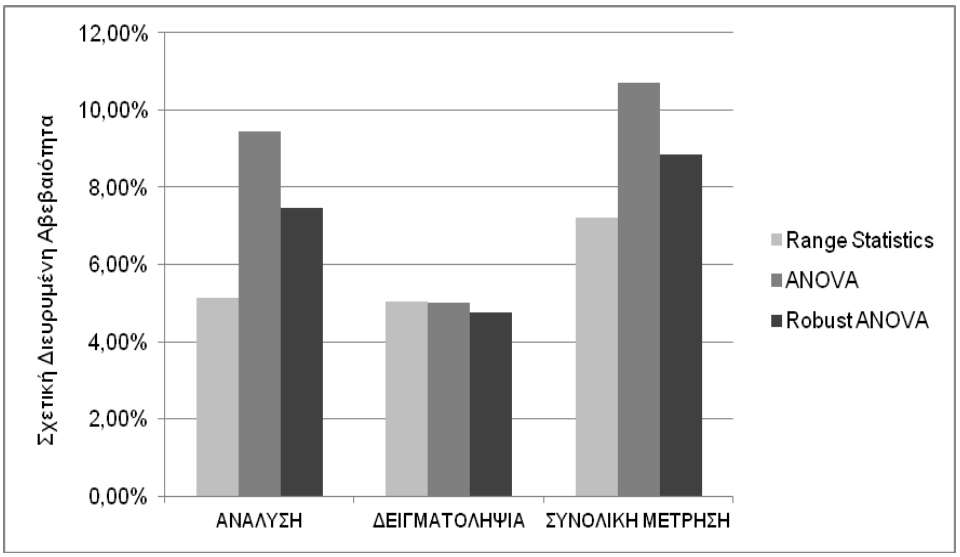
Τα αποτελέσματα της επεξεργασίας των μετρήσεων (εις διπλούν αναλύσεις 104 διπλών δειγμάτων) με τη χρήση των τριών στατιστικών μεθοδολογιών παρουσιάζονται στον Πίνακα Γ.1 και στο Σχήμα Γ.2. Η διευρυμένη αβεβαιότητα της δειγματοληψίας κυμαίνεται από 0,34 εώς 0,40 mg kg^{-1}, ενώ η σχετική διευρυμένη αβεβαιότητα από 4,8 εώς 5,1%, ανάλογα με τη στατιστική μεθοδολογία που χρησιμοποιήθηκε. Όλες οι αβεβαιότητες *U* έχουν εκτιμηθεί με χρήση συντελεστή κάλυψης ίσο με 2 που αντιστοιχεί σε επίπεδο εμπιστοσύνης 95% περίπου.

Μεταξύ των δεδομένων υπάρχουν επτά ακραίες (έκτοπες) τιμές μεταξύ σημείων δειγματοληψίας, οι οποίες επιβεβαιώθηκαν ως τέτοιες εφαρμόζοντας το τεστ Grubbs. Επιπλέον, τρεις ομάδες δεδομένων αναγνωρίστηκαν ως ακραίες τιμές λόγω ανάλυσης ή δειγματοληψίας εφαρμόζοντας το τεστ Cochran. Συνολικά, 9 από τις 104 ομάδες δεδομένων (8,7%) αναγνωρίστηκαν στατιστικά ως ακραίες τιμές. Ως εκ τούτου, τα αποτελέσματα της ανθεκτικής ANOVA, η οποία δεν επηρεάζεται από μικρό αριθμό ακραίων τιμών (λιγότερο από 10%), μπορεί να θεωρηθούν ως περισσότερο αξιόπιστα σε σύγκριση με τα αποτελέσματα της Στατιστικής εύρους τιμών και της «κλασικής» ANOVA.

Είναι φανερό από τα αποτελέσματα της ανθεκτικής ANOVA ότι η αβεβαιότητα της μέτρησης προέρχεται κυρίως από την διακύμανση της ανάλυσης. Συγκεκριμένα, η αβεβαιότητα της ανάλυσης αντιπροσωπεύει το 71% της συνολικής αβεβαιότητας της μέτρησης. Αυτό αφήνει περιθώρια για την αποτελεσματική μείωση της συνολικής αβεβαιότητας, η οποία μπορεί να επιτευχθεί κάνοντας περισσότερες μετρήσεις και υπολογίζοντας τη μέση τιμή τους, αντί για τη διενέργεια μιας μονής μέτρησης. Σε αυτή την περίπτωση, η τυπική απόκλιση της μέσης τιμής ελαττώνεται οδηγώντας σε μικρότερες συνεισφορές στην αβεβαιότητα λόγω τυχαίου σφάλματος.

Πίνακας Γ.1: Αποτελέσματα υπολογισμών με χρήση στατιστικής εύρους τιμών, «κλασικής» ANOVA και ανθεκτικής ANOVA.

	Στατιστική Εύρους Τιμών	**«Κλασική» ANOVA**	**Ανθεκτική ANOVA**
Μέση τιμή (mg kg^{-1})	7,988	7,988	7,079
$s_{analysis}$ (mg kg^{-1})	0,205	0,378	0,265
$s_{sampling}$ (mg kg^{-1})	0,202	0,200	0,169
$s_{measurement}$ (mg kg^{-1})	0,288	0,427	0,314
$U_{analysis}$ (mg kg^{-1})	0,411	0,755	0,529
$U_{analysis}$ (%)	5,1	9,5	7,5
$U_{sampling}$ (mg kg^{-1})	0,404	0,401	0,337
$U_{sampling}$ (%)	5,1	5,0	4,8
$U_{measurement}$ (mg kg^{-1})	0,576	0,855	0,628
$U_{measurement}$ (%)	7,2	10,7	8,9
Συνεισφορά στην αβεβαιότητα λόγω ανάλυσης (%)	51	78	71
Συνεισφορά στην αβεβαιότητα λόγω δειγματοληψίας (%)	49	22	29



Σχήμα Γ.2 Σύγκριση των σχετικών διευρυμένων αβεβαιοτήτων όπως αυτές εκτιμήθηκαν με χρήση στατιστικής εύρους τιμών, «κλασικής» ANOVA και ανθεκτικής ANOVA.

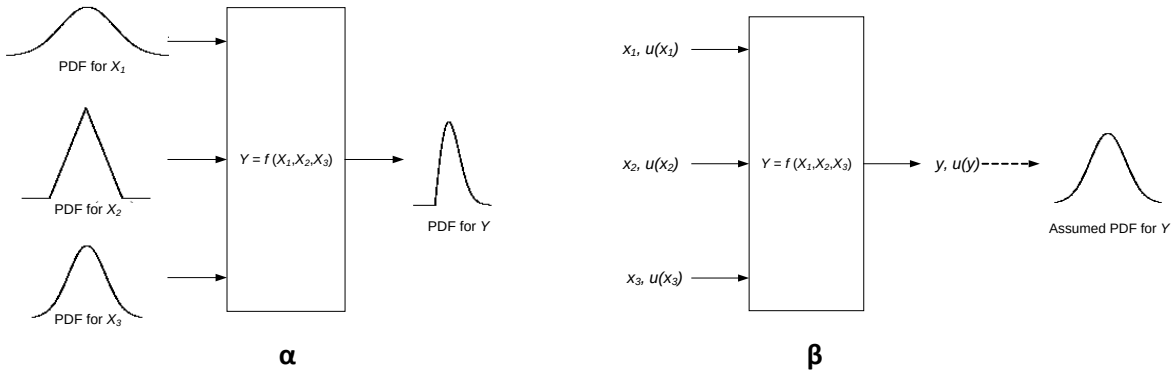
Δ. ΕΚΤΙΜΗΣΗ ΑΒΕΒΑΙΟΤΗΤΑΣ ΜΕΤΡΗΤΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ

Δ.1 Εισαγωγή

Η εκτίμηση της αβεβαιότητας και η καθιέρωση μετρολογικής ιχνηλασιμότητας είναι συνυφασμένες με την αξιοπιστία των αναλυτικών μετρήσεων. Κοινά αποδεκτή μεθοδολογία για την εκτίμηση της αβεβαιότητας είναι αυτή που περιγράφεται στην Οδηγία ISO "Guide to the Expression of Uncertainty in Measurement - GUM". Ωστόσο, η προσέγγιση GUM παρουσιάζει περιορισμούς στην εφαρμογή της, που σχετίζονται με την γραμμικότητα του μοντέλου μέτρησης, την ισχύ του θεωρήματος κεντρικού ορίου και τον υπολογισμό βαθμών ελευθερίας. Προκειμένου να ξεπεραστούν οι ανωτέρω περιορισμοί εκδόθηκε από την επιτροπή Joint Committee for Guides in Metrology (JCGM) μία συμπληρωματική οδηγία του GUM όπου περιγράφεται η εφαρμογή της αριθμητικής μεθόδου Monte Carlo (MCM) ως εναλλακτικής διαδικασίας για την εκτίμηση της αβεβαιότητας. Η εκτίμηση της αβεβαιότητας με τη χρήση της μεθοδολογίας Monte Carlo βασίζεται στην τεχνική διάδοσης κατανομών πιθανότητας και όχι αβεβαιοτήτων (κάτι που ισχύει στην κλασσική προσέγγιση κατά GUM). Στο **Κεφάλαιο 4** της παρούσας Διδακτορικής Διατριβής οι δύο ανωτέρω προσεγγίσεις χρησιμοποιούνται για την παράλληλη εκτίμηση της αβεβαιότητας μέτρησης της θερμογόνου δύναμης πετρελαίου κίνησης με τη χρήση θερμιδομέτρου όλμου.

Δ.2 Εκτίμηση της αβεβαιότητας με χρήση μεθοδολογίας GUM και με προσομοίωση Monte Carlo (MCM)

Στα πλαίσια της προσέγγισης GUM εκτιμάται η τυπική αβεβαιότητα, $u(y)$, του εξερχόμενου μεγέθους Y και στη συνέχεια γίνεται η πάντα υπόθεση ότι το εξερχόμενο μέγεθος ακολουθεί κανονική κατανομή ή κατανομή *t*-Student. Στην προσέγγιση Monte Carlo, αντιθέτως, προσδιορίζεται απευθείας η κατανομή του εξερχόμενου μεγέθους χωρίς να γίνεται κάποια υπόθεση ή παραδοχή για το είδος της. Στο Σχήμα Δ.1 δίνεται ένα παράδειγμα εφαρμογής των 2 μεθοδολογιών για ένα μοντέλο μέτρησης με 3 ανεξάρτητες εισερχόμενες μεταβλητές.



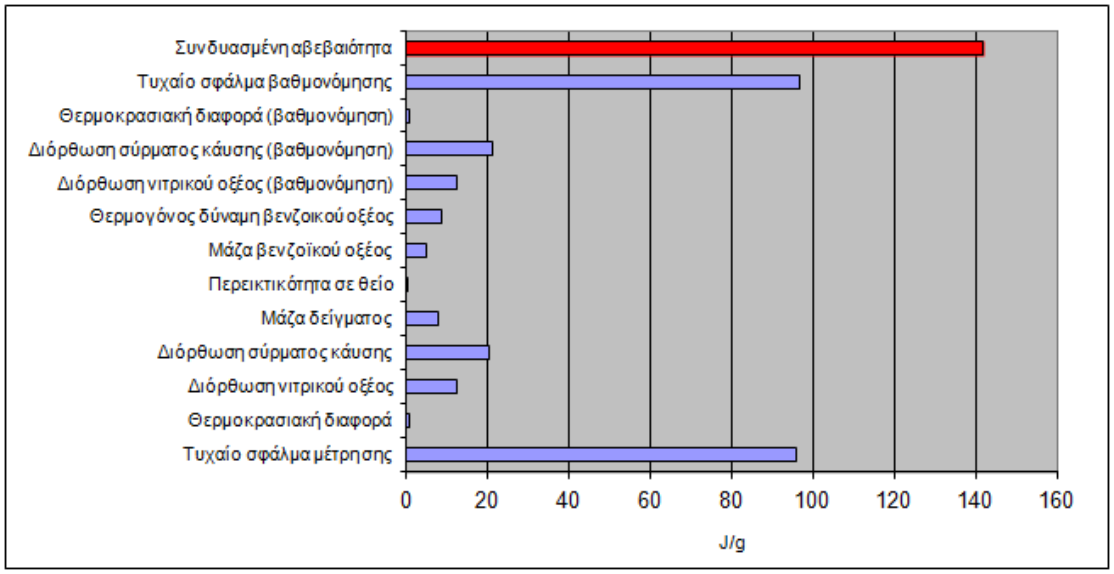
Σχήμα Δ.1 α. Διάδοση κατανομών πιθανότητας 3 ανεξάρτητων μεταβλητών (Προσέγγιση Monte Carlo) **β**. Διάδοση αβεβαιοτήτων 3 ανεξάρτητων μεταβλητών (Προσέγγιση GUM)

Η προσέγγιση Monte Carlo γενικά απαιτεί 10^6 επαναληπτικές δειγματοληψίες στις κατανομές των εισερχόμενων μεγεθών για να προσδιορίσει με αξιοπιστία ένα διάστημα κάλυψης με πιθανότητα κάλυψης 95% και με ακρίβεια ενός ή δύο σημαντικών ψηφίων. Στην πράξη είναι προτιμητέο να εφαρμοστεί μια διαδικασία προσαρμοστικής (adaptive) Monte Carlo κατά την οποία πραγματοποιείται ένας συνεχώς αυξανόμενος αριθμός επαναληπτικών δειγματοληψιών μέχρι τα αποτελέσματα (αναμενόμενη τιμή, τυπική αβεβαιότητα, όρια διαστήματος κάλυψης) να «σταθεροποιηθούν» στατιστικά. Με βάση την συνάρτηση κατανομής πιθανότητας του εξερχόμενου μεγέθους μπορούν να προσδιοριστούν στατιστικές παράμετροι όπως το διάστημα κάλυψης.

Δ.3 Εφαρμογή μεθοδολογιών GUM και MCM - Αποτελέσματα

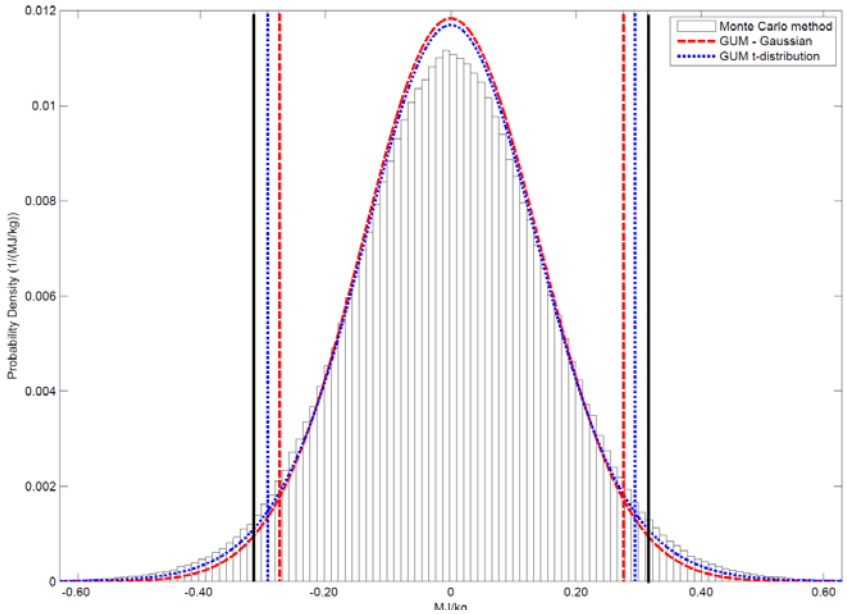
Στα πλαίσια της **προσέγγισης GUM** αναγνωρίστηκαν μέσω του μαθηματικού μοντέλου της μέτρησης όλα τα εισερχόμενα μεγέθη που συνεισφέρουν στην αβεβαιότητα. Στο μετρητικό μοντέλο ενσωματώθηκαν και 2 μηδενικοί όροι, δ_{rep} και το δ_{rep}', για να εισάγουν την αβεβαιότητα λόγω τυχαίων σφαλμάτων κατά τη μέτρηση και κατά τη βαθμονόμηση. Οι τυπικές αβεβαιότητες των εισερχομένων μεγεθών εκτιμήθηκαν είτε βάσει μετρήσεων και εφαρμογής στατιστικών μεθόδων (Τύπου Α) είτε με άλλα μέσα, όπως διαθέσιμες πληροφορίες από πιστοποιητικά διακρίβωσης, χαρακτηριστικά του οργάνου, προηγούμενα δεδομένα κλπ. (Τύπου Β). Στη συνέχεια μέσω υπολογισμού των συντελεστών ευαισθησίας για κάθε

εισερχόμενο μέγεθος, υπολογίστηκε η συνεισφορά στην αβεβαιότητα. Τέλος, εφαρμόζοντας το νόμο διάδοσης των αβεβαιοτήτων, εκτιμήθηκε η τυπική (συνδυασμένη) αβεβαιότητα, $u(Q_g)$, του εξερχόμενου μεγέθους Q_g (θερμογόνος δύναμη), η διευρυμένη αβεβαιότητα (για πιθανότητα κάλυψης 95%) και το αντίστοιχο διάστημα κάλυψης. Η διευρυμένη αβεβαιότητα υπολογίστηκε θεωρώντας είτε κανονική κατανομή (συντελεστής κάλυψης k=1,96), είτε κατανομή t-Student με 21 βαθμούς ελευθερίας που υπολογίστηκαν με εφαρμογή της εξίσωσης Welch-Satterwaite (συντελεστής κάλυψης k_p=2,08). Στο Σχήμα Δ.2 φαίνονται οι εκτιμήσεις της συνδυασμένης αβεβαιότητας και οι συνεισφορές στην αβεβαιότητα όλων των εισερχομένων μεγεθών.



Σχήμα Δ.2 Συνδυασμένη αβεβαιότητα θερμογόνου δύναμης και συνεισφορές στην αβεβαιότητα.

Η **μέθοδος Monte Carlo** υλοποιήθηκε στο περιβάλλον του μαθηματικού λογισμικού MATLAB. Ο κώδικας που αναπτύχθηκε διενήργησε δειγματοληψίες από τις κατανομές των εισερχόμενων μεγεθών. Για όλα τα μεγέθη χρησιμοποιήθηκαν μονομεταβλητές κατανομές, με εξαίρεση ζεύγη μεγεθών που εμφανίζουν συσχέτιση λόγω κοινής χρήσης εξοπλισμού όπου οι δειγματοληψίες έγιναν από διμεταβλητές κατανομές (joint distributions). Η μεθοδολογία Monte Carlo εφαρμόστηκε τόσο με τη προσαρμοστική (adaptive) της μορφή για 2 επίπεδα αριθμητικών ανοχών όσο και με *a priori* επιλεγμένο αριθμό επαναλήψεων (10^6). Ο Πίνακας Δ.1 και το Σχήμα Δ.3 εμφανίζουν τα αποτελέσματα όλων των μεθοδολογιών.



Σχήμα Δ.3 Κατανομές πιθανότητας και διαστήματα κάλυψης 95% των προσεγγίσεων Monte Carlo, GUM (κανονική κατανομή) και GUM (κατανομή *t*-student).

Πίνακας Δ.1 Αποτελέσματα εφαρμογής μεθοδολογιών GUM και MCM για την εκτίμηση της αβεβαιότητας μέτρησης της θερμογόνου δύναμης πετρελαίου κίνησης.

Μέθοδος	**Αριθμός επαναληπτικών δειγματοληψιών, *M***	**Απαιτούμενος χρόνος υπολογισμών [1]**	**Εκτιμούμενη τιμή, Q_g**	**Τυπική Αβεβαιότητα, $u(Q_g)$**	**Διάστημα κάλυψης 95%**	**Αριθμός σημαντικών ψηφίων**	**Αριθμητική ανοχή, *δ***
GUM - Gaussian distribution	-	-	45,19 MJ kg^{-1} (10794 cal g^{-1})	0,14 MJ kg^{-1} (34 cal g^{-1})	[44,92 – 45,47] MJ kg^{-1} [10728 – 10861] cal g^{-1}	-	-
GUM - *t*-Student distribution	-	-	45,19 MJ kg^{-1} (10794 cal g^{-1})	0,14 MJ kg^{-1} (34 cal g^{-1})	[44,90 – 45,49] MJ kg^{-1} [10724 – 10865] cal g^{-1}	-	-
MCM	10^6	1,04 sec	45,19 MJ kg^{-1} (10794 cal g^{-1})	0,16 MJ kg^{-1} (38 cal g^{-1})	[44,88 – 45,51] MJ kg^{-1} [10719 – 10870] cal g^{-1}	-	-
Adaptive MCM	$3.5 \cdot 10^5$	0,60 sec	45,19 MJ kg^{-1} (10794 cal g^{-1})	0,16 MJ kg^{-1} (38 cal g^{-1})	[44,88 – 45,51] MJ kg^{-1} [10719 – 10870] cal g^{-1}	2	0,005 MJ kg^{-1} (0,5 cal g^{-1})
Adaptive MCM	$5.7 \cdot 10^6$	39,9 sec	45,19 MJ kg^{-1} (10794 cal g^{-1})	0,16 MJ kg^{-1} (38 cal g^{-1})	[44,88 – 45,51] MJ kg^{-1} [10719 – 10870] cal g^{-1}	2	0,001 MJ kg^{-1} (0,1 cal g^{-1})

[1] Χρησιμοποιήθηκε Η/Υ με τα εξής χαρακτηριστικά: Intel® Core™ i3 M330, 2.13GHz, 4GB RAM

Όλες οι μεθοδολογίες GUM και MCM έδωσαν ταυτόσημα αποτελέσματα σε σχέση με την εκτιμούμενη τιμή για τη θερμογόνο δύναμη, Q_g. Από την άλλη πλευρά, σημαντικές διαφορές παρατηρούνται σε σχέση με τις διευρυμένες αβεβαιότητες και τα εκτιμώμενα διαστήματα κάλυψης μεταξύ των αποτελεσμάτων MCM και GUM. Συγκεκριμένα, η διευρυμένη αβεβαιότητα (για πιθανότητα κάλυψης 95%) εκτιμήθηκε μέσω της μεθοδολογίας GUM (υποθέτοντας κανονική κατανομή) στα 0,28 MJ kg^{-1} ή 66,3 cal g^{-1}. Η τιμή αυτή είναι 12% μικρότερη από την τιμή που έδωσε η εφαρμογή της μεθοδολογίας MCM (0,32 MJ kg^{-1} ή 75,3 cal g^{-1}). Χρησιμοποιώντας τη μεθοδολογία GUM σε συνδυασμό με την εξίσωση Welch-Satterthwaite για τον υπολογισμό βαθμών ελευθερίας και στη συνέχεια, αποδίδοντας μια κατανομή *t* - student στο μετρούμενο μέγεθος οδηγούμαστε σε αυξημένη (σε σχέση με την υπόθεση της κανονικής κατανομής) διευρυμένη αβεβαιότητα (0,29 MJ kg^{-1} ή 70,4 cal g^{-1}), αλλά και πάλι κατά 7% χαμηλότερη από αυτή που δίνει η προσέγγιση MCM. Αυτές οι διαφορές μπορούν να αποδοθούν στη μικρή μη γραμμικότητα του μοντέλου μέτρησης και στη προσεγγιστική φύση της εξίσωσης Welch-Satterthwaite, που στην προκειμένη περίπτωση υπερεκτιμά τους βαθμούς ελευθερίας. Τέλος, από τα αποτελέσματα φαίνεται ότι υπάρχει συμφωνία μεταξύ των αποτελεσμάτων της προσέγγισης Monte Carlo τόσο στην προσαρμοστική (adaptive) της μορφή όσο στη μορφή της με *a priori* επιλεγμένο αριθμό επαναλήψεων (10^6).

Δ.4 Υπολογισμοί συντελεστών ευαισθησίας

Ο υπολογισμός συντελεστών ευαισθησίας και των συνεισφορών του κάθε εισερχόμενου μεγέθους στην τελική αβεβαιότητα δεν είναι απαραίτητος όταν εφαρμόζει κανείς την τεχνική Monte Carlo (σε αντίθεση με την προσέγγιση κατά GUM όπου απαιτούνται για την κατασκευή του ισοζυγίου αβεβαιοτήτων). Παρόλα αυτά ο υπολογισμός των συνεισφορών των αβεβαιοτήτων χρησιμεύει στο να προσδιοριστούν κυρίαρχοι παράγοντες αβεβαιότητας. Το «μειονέκτημα» αυτό της προσέγγισης Monte Carlo μπορεί εύκολα να ξεπεραστεί αφού με μια μικρή τροποποίηση του αλγόριθμου και την εφαρμογή του για κάθε εισερχόμενο μέγεθος μπορούν να υπολογιστούν οι τιμές των συντελεστών ευαισθησίας και οι συνεισφορές στην αβεβαιότητα. Από τα αποτελέσματα αυτών των υπολογισμών φαίνεται ότι οι δύο κυρίαρχοι παράγοντες είναι τα μεγέθη δ_{rep} και το δ_{rep}' τα οποία συνδέονται με τα τυχαία σφάλματα της μέτρησης και της βαθμονόμησης (βλ. Σχήμα Δ.2).

Δ.5 Εκτίμηση αβεβαιοτήτων Τύπου Α μέσω Μπεϋζιανής προσέγγισης

Με χρήση της «κλασσικής» στατιστικής, η τυπική αβεβαιότητα Τύπου Α ενός μεγέθους x_i εκτιμάται ως η τυπική απόκλιση από επαναλαμβανόμενες μετρήσεις δηλ. $u(x_i)= s(x_i)$. Όταν ο αριθμός των επαναληπτικών μετρήσεων m, είναι μικρός αυτή η εκτίμηση καθίσταται αρκετά αναξιόπιστη (στατιστικά αβέβαιη) και για αυτό το λόγο γίνεται χρήση των βαθμών ελευθερίας.

Μέσω της Μπεϋζιανής προσέγγισης η τυπική αβεβαιότητα Τύπου Α εκτιμάται ως:

$$u_{Bayes}(x_i) = \sqrt{\frac{m-1}{m-3}}s(x_i) \qquad (\Delta.1)$$

Ο συντελεστής $\sqrt{(m-1)/(m-3)}$ που είναι ενσωματωμένος στη Μπεϋζιανή προσέγγιση εξαλείφει ουσιαστικά τη στατιστική αβεβαιότητα που προκύπτει από ένα μικρό αριθμό μετρήσεων.

Στην περίπτωση της εκτίμησης της αβεβαιότητας της μέτρησης της θερμογόνου δύναμης, οι δύο κυρίαρχοι παράγοντες αβεβαιότητας είναι τα μεγέθη δ_{rep} και το δ_{rep}' τα οποία συνδέονται με τα τυχαία σφάλματα της βαθμονόμησης και της μέτρησης και εκτιμώνται ως αβεβαιότητες Τύπου Α. Η χρήση της Μπεϋζιανής προσέγγισης για την εκτίμηση αυτών αβεβαιοτήτων Τύπου Α και στη συνέχεια η χρήση τους στο ισοζύγιο αβεβαιοτήτων της μεθοδολογίας GUM έδωσε μια συνδυασμένη αβεβαιότητα $u(Q_g)$ 160,7 J g^{-1} ή 38 cal g^{-1} που αντιστοιχεί σε 95% διάστημα κάλυψης [44,88 – 45,51] MJ kg^{-1} ή [10719 – 10870] cal g^{-1}. Τα αποτελέσματα αυτά συμφωνούν με τα αποτελέσματα της προσέγγισης MCM.

Δ.6 Εκτίμηση της αβεβαιότητας με χρήση δεδομένων από διεργαστηριακές συγκριτικές δοκιμές ικανότητας

Η αβεβαιότητα της μέτρησης της θερμογόνου δύναμης εκτιμήθηκε και με τη χρήση δεδομένων από διεργαστηριακές συγκριτικές δοκιμές ικανότητας (εμπειρική προσέγγιση). Χρησιμοποιώντας δεδομένα από 9 κύκλους ενός συγκεκριμένου διεργαστηριακού προγράμματος, εκτιμήθηκε η συνολική (pooled) τυπική απόκλιση σε συνθήκες αναπαραγωγιμότητας, s_R^{pooled}, της μεθόδου μέτρησης της θερμογόνου δύναμης καυσίμων. Οι υπολογισμοί έδωσαν s_R^{pooled}=0,15 MJ kg^{-1} (36 cal g^{-1}), το

οποίο πολλαπλασιαζόμενο με ένα συντελεστή κάλυψης *k*=1,96 δίνει μια διευρυμένη αβεβαιότητα 0,30 MJ kg^{-1} ή 71 cal g^{-1}. Η εκτίμηση αυτή είναι μικρότερη από τα αποτελέσματα της προσέγγισης MCM. Παρόλα αυτά σε πολλές περιπτώσεις όπου οι απαιτήσεις ακρίβειας των χρηστών είναι χαμηλές, μια εκτίμηση αβεβαιότητας από δεδομένα διεργαστηριακής αναπαραγωγιμότητας μπορεί να αποδειχτεί κατάλληλη για τη σκοπούμενη χρήση (fit for purpose).

Ε. ΕΚΤΙΜΗΣΗ ΤΥΠΙΚΗΣ ΑΒΕΒΑΙΟΤΗΤΑΣ ΚΑΜΠΥΛΗΣ ΒΑΘΜΟΝΟΜΗΣΗΣ

Ε.1 Εισαγωγή

Η διαδικασία της βαθμονόμησης είναι ένα απαραίτητο στάδιο σε πολλές χημικές αναλύσεις που αφορούν στον προσδιορισμό της συγκέντρωσης μιας ουσίας με βάση την απόκριση (σήμα) ενός οργάνου. Η βαθμονόμηση δημιουργεί μια σχέση μεταξύ της τιμής ενός προτύπου (τιμή αναφοράς) και της μετρούμενης ποσότητας (απόκριση του οργάνου). Μόλις δημιουργηθεί αυτή η σχέση, η οποία συνήθως περιγράφεται από μια ευθεία γραμμή (γραμμικό μοντέλο), το μοντέλο βαθμονόμησης χρησιμοποιείται ανεστραμμένα, με στόχο να προβλέψει μια τιμή από μία απόκριση του οργάνου.

Για την κατασκευή της καμπύλης βαθμονόμησης συχνά χρησιμοποιείται η γραμμική μέθοδος προσαρμογής ελαχίστων τετραγώνων (linear least square regression) από την οποία προκύπτουν η κλίση και η τεταγμένη του γραμμικού μοντέλου. Όπως και με τα περισσότερα στατιστικά μεγέθη, οι αριθμητικές τιμές της κλίσης και της τεταγμένης αποτελούν μόνο εκτιμήσεις που βασίζονται σε έναν πεπερασμένο αριθμό μετρήσεων και ως εκ τούτου συνοδεύονται από αβεβαιότητα. Αυτό όμως οδηγεί και σε μια αβεβαιότητα στην μέτρηση μιας ποσότητας (π.χ. συγκέντρωσης) με τη χρήση καμπύλης βαθμονόμησης. Αυτή η αβεβαιότητα, μπορεί να εκτιμηθεί χρησιμοποιώντας διάφορες μεθόδους και προσεγγίσεις. Καθώς η βαθμονόμηση αποτελεί συχνά μια σημαντική συνιστώσα της συνολικής αβεβαιότητας της μέτρησης, μια αξιόπιστη εκτίμηση αυτής είναι κρίσιμης σημασίας.

Ο προσδιορισμός της περιεκτικότητας σε θείο των προϊόντων πετρελαίου προσδιορίζεται συνήθως με τη χρήση διαφόρων φασματομετρικών τεχνικών, οι οποίες περιλαμβάνουν σχεδόν πάντα την κατασκευή και τη χρήση μίας καμπύλης βαθμονόμησης. Στο **Κεφάλαιο 5** της παρούσας Διδακτορικής Διατριβής, εφαρμόζονται και συγκρίνονται 4 μεθοδολογίες εκτίμησης της αβεβαιότητας λόγω καμπύλης βαθμονόμησης που χρησιμοποιείται για τον προσδιορισμό της περιεκτικότητας σε θείο σε καύσιμα σύμφωνα με τη μέθοδο υπεριώδους φθορισμού (ISO 20846, ASTM D5453). Επιπλέον, δεδομένου ότι η εκτίμηση των 2 παραμέτρων μιας καμπύλης βαθμονόμησης βασίζεται σε ένα μοντέλο μέτρησης με πολλαπλά εξερχόμενα μετρούμενα μεγέθη (κλίση, τεταγμένη), εφαρμόζονται και οι βασικές αρχές της συμπληρωματικής οδηγίας του GUM (Supplement 2 - Extension to any number of output quantities).

E.2 Κατασκευή καμπύλης βαθμονόμησης

Για την κατασκευή της καμπύλης βαθμονόμησης χρησιμοποιήθηκαν πρότυπα δείγματα γνωστής (πιστοποιημένης) συγκέντρωσης τα οποία αναλύθηκαν εις τριπλούν. Με δεδομένα τις συγκεντρώσεις των προτύπων, $x_{i,}$ και τις αντίστοιχες αποκρίσεις του εξοπλισμού, y_i, κατασκευάστηκε καμπύλη βαθμονόμησης 18 σημείων η οποία στη συνέχεια χρησιμοποιήθηκε για να μετρηθεί ένα δείγμα καυσίμου ντήζελ. Η εξίσωση της καμπύλης βαθμονόμησης ήταν της μορφής:

$$Y = b_0 + b_1 x \tag{E.1}$$

όπου *Y* είναι η απόκριση του οργάνου (εξαρτημένη μεταβλητή), *x* είναι η συγκέντρωση του αναλύτη (ανεξάρτητη μεταβλητή) και b_0 και b_1 είναι οι συντελεστές του μοντέλου, γνωστοί ως τεταγμένη και κλίση, αντίστοιχα. Κάτα την ανάλυση του άγνωστου δείγματος η παραπάνω εξίσωση χρησιμοποιείται ανεστραμμένη, ώστε να υπολογιστεί η συγκέντρωση x_{pred}, με δεδομένη την απόκριση, y_0, του οργάνου:

$$x_{\text{pred}} = \frac{y_0 - b_0}{b_1} \tag{E.2}$$

E.3 Μεθοδολογίες εκτίμησης της αβεβαιότητας

E.3.1 Μεθοδολογία GUM

Η Οδηγία GUM (Guide to the Expression of Uncertainty in Measurement) περιγράφει ένα πλαίσιο εκτίμησης της αβεβαιότητας το οποίο χρησιμοποιεί ως πληροφορίες τις εκτιμήτριες και τις τυπικές αβεβαιότητες των εισερχομένων μεγεθών (b_0, b_1, y_0) του μετρητικού μοντέλου. Στη συνέχεια, χρησιμοποιώντας το νόμο της διάδοσης των αβεβαιοτήτων μέσω προσέγγισης του μοντέλου με ανάπτυγμα σειράς Taylor $1^{ης}$ τάξης εξάγεται η εκτίμηση του μετρούμενου μεγέθους x_{pred} και η τυπική αβεβαιότητα $u(x_{\text{pred}})$ που οφείλεται στις διακυμάνσεις των b_0, b_1 και y_0:

$$u(x_{\text{pred}}) = \sqrt{\left[c_0 u(b_0)\right]^2 + \left[c_1 u(b_1)\right]^2 + \left[c_2 u(y_0)\right]^2 + 2c_0 c_1 u(b_0) u(b_1) r(b_0, b_1)} \quad \text{(E.3)}$$

όπου c_0, c_1 και c_2 είναι οι συντελεστές ευαισθησίας των b_0, b_1 και y_0, $u(b_0)$, $u(b_1)$ και $u(y_0)$ είναι οι τυπικές αβεβαιότητες των b_0, b_1 και y_0, και $r(b_0, b_1)$ είναι ο συντελεστής συσχέτισης των b_0 και b_1.

E.3.2 Μεθοδολογία Kragten

Η μέθοδος Kragten, βασίζεται στην προσέγγιση των μερικών παραγώγων με πεπερασμένες διαφορές με στόχο τον υπολογισμό των συντελεστών ευαισθησίας στη σχέση υπολογισμού της αβεβαιότητας μέτρησης. Με βάση την προσέγγιση αυτή, η μέθοδος εισάγει έναν αριθμητικό τρόπο υπολογισμού της συνδυασμένης αβεβαιότητας μέτρησης. Η αριθμητική προσέγγιση αναιρεί την ανάγκη αναλυτικών υπολογισμών των μερικών παραγώγων.

Η αριθμητική προσέγγιση των μερικών παραγώγων της συνάρτησης που παρέχει την τιμή του μετρούμενου μεγέθους ως προς τις επιμέρους μεταβλητές, με αντίστοιχες πεπερασμένες διαφορές, υλοποιείται λαμβάνοντας τη διαφορά – διαταραχή δx, ίση με την αντίστοιχη τυπική αβεβαιότητα $u(x)$:

$$c_0 = \frac{\partial x_{\text{pred}}}{\partial b_0} \approx \frac{\delta x_{\text{pred}}}{\delta b_0} = \frac{x_{\text{pred}}(b_0 + u(b_0)) - x_{\text{pred}}(b_0)}{u(b_0)} \quad \text{(E.4)}$$

$$c_1 = \frac{\partial x_{\text{pred}}}{\partial b_1} \approx \frac{\delta x_{\text{pred}}}{\delta b_1} = \frac{x_{\text{pred}}(b_1 + u(b_1)) - x_{\text{pred}}(b_1)}{u(b_1)} \quad \text{(E.5)}$$

$$c_2 = \frac{\partial x_{\text{pred}}}{\partial y_0} \approx \frac{\delta x_{\text{pred}}}{\delta y_0} = \frac{x_{\text{pred}}(y_0 + u(y_0)) - x_{\text{pred}}(y_0)}{u(y_0)} \quad \text{(E.6)}$$

E.3.3 Μεθοδολογία Monte Carlo

Η μέθοδος Monte Carlo (MCM) υπολογίζει την κατανομή της μετρούμενης ποσότητας (x_{pred}), με βάση δειγματοληψίες που διενεργούνται στις κατανομές των εισερχομένων μεγεθών (b_0, b_1, y_0). Στην προκειμένη περίπτωση, δύο εισερχόμενες ποσότητες (b_0, b_1) δεν είναι ανεξάρτητες (μη μηδενική συνδιακύμανση) και ως εκ

τούτου χρησιμοποιείται για τις δειγματοληψίες μια από κοινού κατανομή πιθανότητας (joint probability distribution).

Η εκτίμηση της μετρούμενης ποσότητας (x_{pred}) υπολογίζεται από το μέσο όρο των M δειγματοληψιών που παράγουν M τιμές ($x_{\text{pred}}^{(k)}$, k = 1,…,M):

$$x_{\text{pred}} = \frac{1}{M}\sum_{k=1}^{M} x_{\text{pred}}^{(k)} \tag{E.7}$$

ενώ η τυπική αβεβαιότητα $u(x_{\text{pred}})$ που συνδέεται με το x_{pred} υπολογίζεται ως η τυπική απόκλιση των M τιμών:

$$u(x_{\text{pred}}) = \sqrt{\frac{1}{M-1}\sum_{k=1}^{M}(x_{\text{pred}}^{(k)} - x_{\text{pred}})^2} \tag{E.8}$$

E.3.4 Χρήση εξίσωσης τυπικού σφάλματος

Η αβεβαιότητα της συγκέντρωσης που υπολογίζεται μέσω μιας καμπύλης βαθμονόμησης μπορεί να εκτιμηθεί και από την θεωρητική εξίσωση του τυπικού σφάλματος:

$$s'(x_{\text{pred}}) = \frac{SE_{\text{regression}}}{b_1}\sqrt{\frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{E.9}$$

όπου $SE_{\text{regression}}$ είναι το τυπικό σφάλμα εκτίμησης, n ο αριθμός των πειραματικών δεδομένων (ζεύγη x_i, y_i), $\bar{x}$, η μέση τιμή των συγκεντρώσεων των προτύπων και $\bar{y}$, η μέση τιμή των αποκρίσεων.

Η παραπάνω έκφραση δεν λαμβάνει υπόψη την αβεβαιότητα $u(y_0)$ της απόκρισης, y_0, του άγνωστου δείγματος, η οποία είναι εισερχόμενο μέγεθος του μοντέλου μέτρησης. Αυτό μπορεί να οδηγήσει σε αποτελέσματα που έχουν υποεκτιμηθεί και δεν είναι συγκρίσιμα με τα αποτελέσματα των άλλων μεθοδολογιών. Η τυπική αβεβαιότητα της μετρούμενης τιμής x_{pred}, συμπεριλαμβανομένης της αβεβαιότητας της απόκρισης δίνεται από την εξίσωση:

$$u(x_{\text{pred}}) = \sqrt{s'(x_{\text{pred}})^2 + \left[c_2 u(y_0)\right]^2} \quad \text{(E.10)}$$

E.4 Αποτελέσματα

Τα αποτελέσματα των τεσσάρων μεθοδολογιών συνοψίζονται στον Πίνακα Ε.1. Τα αποτελέσματα όλων των μεθοδολογιών συμφωνούν μεταξύ τους. Πραγματοποιήθηκαν επίσης και υπολογισμοί μη λαμβάνοντας υπόψη τη συνδιακύμανση μεταξύ κλίσης και τεταγμένης (με τις με τις μεθοδολογίες GUM, MCM, Kragten) όπου φαίνεται ότι γίνεται κατά 62% υπερεκτίμηση της αβεβαιότητας. Επίσης αν στο αποτέλεσμα της προσεγγιστικής εξίσωσης του τυπικού σφάλματος δεν προστεθεί και τυπική αβεβαιότητα της απόκρισης οδηγούμαστε σε υποεκτίμηση της αβεβαιότητας κατά 22%.

Πίνακας Ε.1 Αποτελέσματα της εφαρμογής διαφόρων μεθοδολογιών για την εκτίμηση της αβεβαιότητας της καμπύλης βαθμονόμησης που χρησιμοποιείται για τον προσδιορισμό της συγκέντρωσης του θείου σε καύσιμα.

	Μέση τιμή, x_{pred} (ng μL^{-1})	**Τυπική αβεβαιότητα, $u(x_{\text{pred}})$ (ng μL^{-1})**
GUM (με συνδιακύμανση)	8,000	0,175
Μέθοδος Kragten (με συνδιακύμανση)	8,000	0,172
MCM (με συνδιακύμανση)	8,003	0,175
Εξίσωση τυπικού σφάλματος (με αβεβαιότητα απόκρισης)	8,000	0,175
Εξίσωση τυπικού σφάλματος (χωρίς αβεβαιότητα απόκρισης)	8,000	0,137
GUM (χωρίς συνδιακύμανση)	8,000	0,283
Μέθοδος Kragten method (χωρίς συνδιακύμανση)	8,000	0,279
MCM (χωρίς συνδιακύμανση)	8,005	0,284

E.5 Η καμπύλη βαθμονόμησης ως μετρητικό μοντέλο πολλαπλών εξόδων

Η δημοσίευση της συμπληρωματικής Οδηγίας GUM S2 (Supplement 2 - Extension to any number of output quantities), δίνει τη δυνατότητα μιας αποτελεσματικότερης και μετρολογικά ορθότερης αποτίμησης του αποτελέσματος μιας μέτρησης, όταν αυτό συνίσταται σε περισσότερα του ενός, συσχετιζόμενα μεταξύ τους δεδομένα, στη βάση των διαθέσιμων εκτιμήσεων των εισερχομένων μεγεθών, των

αβεβαιοτήτων που χαρακτηρίζουν τις εκτιμήσεις αυτές, καθώς επίσης και των αμοιβαίων αβεβαιοτήτων λόγω συνδιακύμανσης.

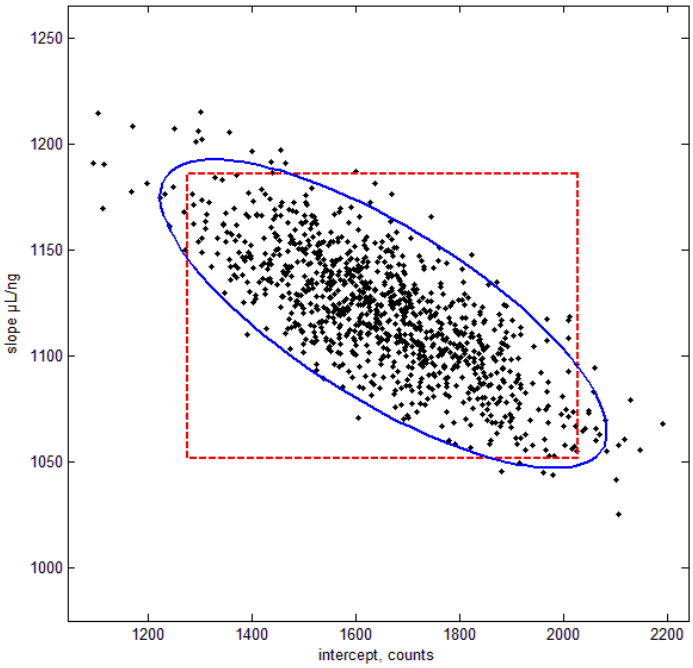
Η κατασκευή μιας γραμμικής καμπύλης βαθμονόμησης είναι ουσιαστικά ένα μετρητικό μοντέλο πολλαπλών εξόδων. Το αποτέλεσμα της «μέτρησης» είναι οι συντελεστές b_0 και b_1, της εξίσωσης της καμπύλης. Το «μοντέλο μέτρησης» είναι ο μηχανισμός υπολογισμού των b_0 και b_1 από τα πρωτογενή πειραματικά δεδομένα της βαθμονόμησης (συγκεντρώσεις και αποκρίσεις). Οι τιμές των b_0 και b_1 χαρακτηρίζονται από τις αντίστοιχες αβεβαιότητες $u(b_0)$ και $u(b_1)$, καθώς και από την μεταξύ τους συνδιακύμανση $u(b_0,b_1)$. Δεδομένου ότι πρόκειται για διμεταβλητό μοντέλο μέτρησης, το διάστημα κάλυψης έχει τη μορφή δυσδιάστατης περιοχής στο επίπεδο που καθορίζεται από τους άξονες τιμών των b_0 και b_1.

Υπάρχουν 2 τύποι περιοχών κάλυψης που μπορούν να κατασκευαστούν:

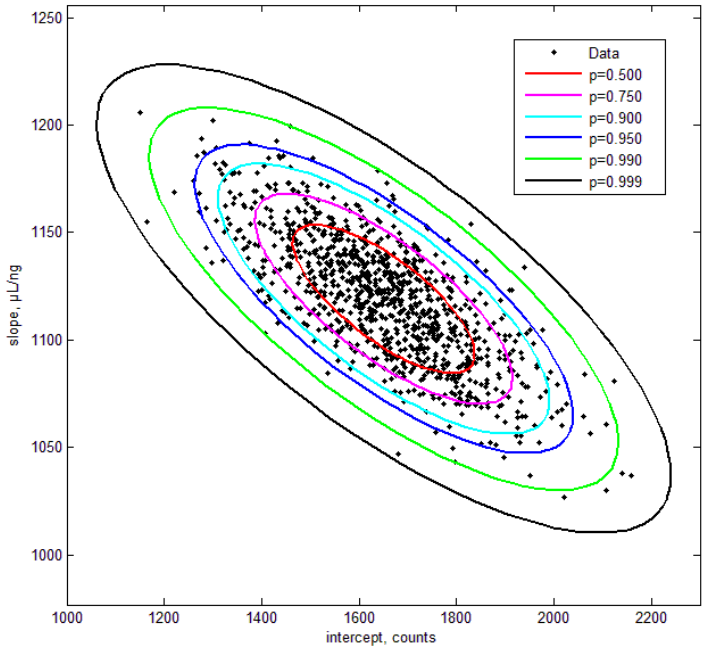
- Η ορθογωνική περιοχή κάλυψης με κέντρο το σημείο (b_0 , b_1)
- Η ελλειπτική περιοχή κάλυψης με κέντρο το σημείο (b_0 , b_1)

Καθώς στην περίπτωση του μετρητικού μοντέλου της καμπύλης βαθμονόμησης τα δυο εξερχόμενα μεγέθη (κλίση και τεταγμένη) χαρακτηρίζονται από συνδιακύμανση, η ελλειπτική περιοχή κάλυψης είναι περισσότερο κατάλληλη από την ορθογωνική. Γενικά, η ελλειπτική περιοχή κάλυψης είναι η μικρότερη δυνατή περιοχή κάλυψης για μια συγκεκριμένη πιθανότητα κάλυψης, ενώ η ορθογωνική περιοχή κάλυψης δεν αντικατοπτρίζει καθόλου την συσχέτιση μεταξύ των μετρούμενων μεγεθών καθώς ορίζεται θεωρώντας δύο διακριτά διαστήματα κάλυψης για το b_0 και το b_1.

Στο Σχήμα Ε.1 φαίνονται τα όρια της ελλειπτικής περιοχής κάλυψης (συνεχής γραμμή), μαζί με τα όρια της ορθογωνικής περιοχής κάλυψης (διακεκομμένη γραμμή) για πιθανότητα κάλυψης p=0,95. Στο ίδιο σχήμα έχουν τοποθετηθεί ενδεικτικά και σημεία τα οποία έχουν προκύψει με τη βοήθεια προσομοίωσης Monte Carlo, θεωρώντας διμεταβλητή κανονική κατανομή. Στο Σχήμα Ε.2 φαίνονται τα όρια ελλειπτικών περιοχών κάλυψης για διάφορες πιθανότητες κάλυψης. Αντίστοιχοι υπολογισμοί και απεικονίσεις έγιναν και αγνοώντας της συνδιακύμανση μεταξύ b_0 και b_1.



Σχήμα Ε.1 Ελλειπτική και ορθογωνική περιοχή κάλυψης (πιθανότητα κάλυψης *p*=0,95) της από κοινού συνάρτησης πυκνότητας πιθανότητας της κλίσης και της τεταγμένης.



Σχήμα Ε.2 Ελλειπτικές περιοχές κάλυψης (για διάφορες πιθανότητες κάλυψης *p*) της από κοινού συνάρτησης πυκνότητας πιθανότητας της κλίσης και της τεταγμένης.

ΣΤ. ΧΡΗΣΗ ΤΗΣ ΑΒΕΒΑΙΟΤΗΤΑΣ ΚΑΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΕΠΙΔΟΣΗΣ ΜΕΘΟΔΩΝ ΣΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΣΥΜΜΟΡΦΩΣΗΣ

ΣΤ.1 Εισαγωγή

Η αξιολόγηση της συμμόρφωσης ενός προϊόντος με συγκεκριμένες προδιαγραφές θα πρέπει να γίνεται με μεθοδολογίες που παρέχουν επαρκή εμπιστοσύνη ότι το υπό εξέταση προϊόν πληροί (ή όχι) αυτές τις προδιαγραφές, ελαχιστοποιώντας τον κίνδυνο των εσφαλμένων αποφάσεων, οι οποίες συχνά έχουν οικονομικές επιπτώσεις. Ειδικά, όταν η αξιολόγηση βασίζεται σε εργαστηριακές μετρήσεις, η διαδικασία της αξιολόγησης της συμμόρφωσης θα πρέπει να λαμβάνει υπόψη ότι καμία μέτρηση δεν είναι 100% ακριβής, καθώς η πραγματική τιμή κάθε μετρούμενου μεγέθους και τυχόν σφάλματα που σχετίζονται με τη μέτρηση δεν μπορούν να είναι γνωστά. Ιδιαίτερα, όταν το αποτέλεσμα της μέτρησης είναι κοντά στο όριο προδιαγραφής, μόνο με τη χρήση της θεωρίας πιθανοτήτων και καταλλήλων κανόνων απόφασης μπορεί κανείς να έχει τον έλεγχο επί της πιθανότητας να λάβει μια λανθασμένη απόφαση (ρίσκο).

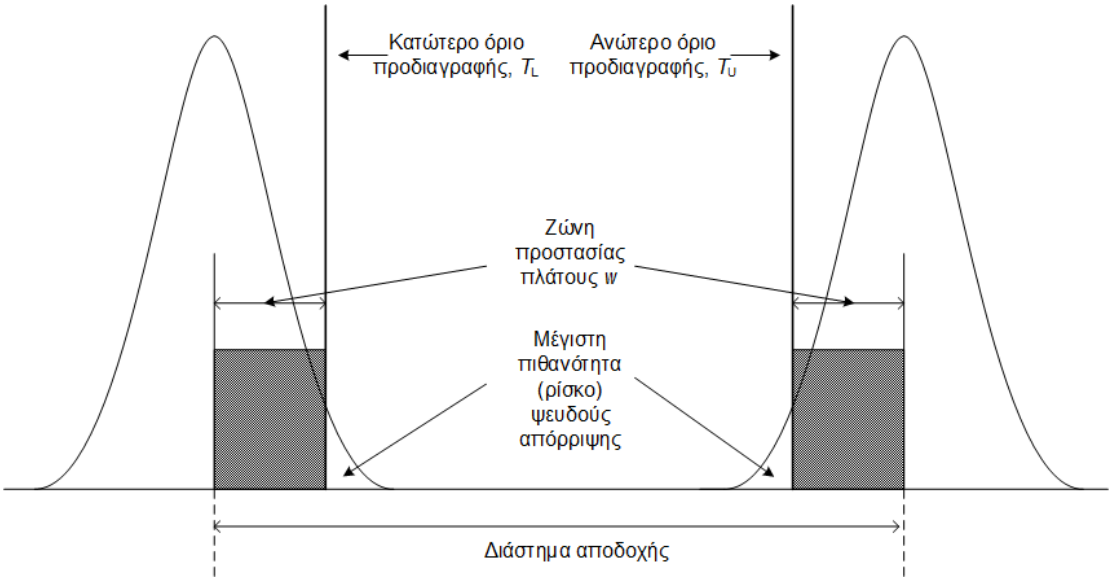
Στο **Κεφάλαιο 6** της παρούσας Διδακτορικής Διατριβής αναλύονται οι διαθέσιμες προσεγγίσεις που μπορούν να χρησιμοποιηθούν για να υποστηρίξουν αξιόπιστες αποφάσεις σχετικά με την αξιολόγηση της συμμόρφωσης των καυσίμων. Οι προσεγγίσεις αυτές εφαρμόζονται και συγκρίνονται για την αξιολόγηση της συμμόρφωσης δειγμάτων ντήζελ κίνησης τα οποία ελέγχονται με βάση την προδιαγραφή της Ευρωπαϊκής Ένωσης για την περιεκτικότητα τους σε θείο. Τα αποτελέσματα των αναλύσεων των δειγμάτων ντήζελ κίνησης από 769 πρατήρια καυσίμων χρησιμοποιήθηκαν για τους υπολογισμούς.

ΣΤ.2 Αξιολόγηση της συμμόρφωσης

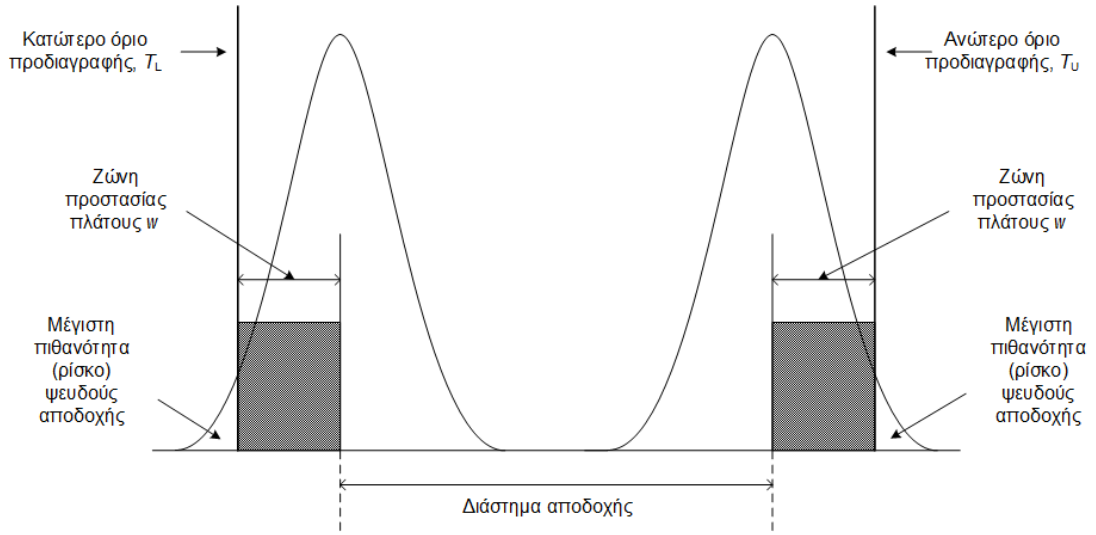
Ένα μετρούμενο μέγεθος Y μπορεί πάντα να περιγραφεί από μία κατανομή πιθανοτήτων, η οποία είναι αντιπροσωπευτική και συμβατή με τον βαθμό γνώσης του μεγέθους. Η καλύτερη εκτίμηση y για το μέγεθος προκύπτει από την αναμενόμενη τιμή (expectation) $E[Y]$ της κατανομής, ενώ η αβεβαιότητα της εκτιμάται από την τυπική απόκλιση της κατανομής ως $u^2(y)= \mathrm{Var}[Y]$, όπου $\mathrm{Var}[Y]$ είναι η διακύμανσή (variance) της. Συχνά, στις εργαστηριακές μετρήσεις οι αναμενόμενες τιμές ενός μετρούμενου μεγέθους περιγράφονται με την κανονική

κατανομή (Gaussian distribution). Έτσι, έχοντας ως δεδομένα το αποτέλεσμα της μέτρησης και την πιθανή διακύμανση των τιμών (αβεβαιότητα) μπορεί κανείς να υπολογίσει την πιθανότητα η πραγματική τιμή του μετρούμενου μεγέθους να είναι πάνω ή κάτω από ένα συγκεκριμένο όριο. Με τον τρόπο αυτό υπολογίζεται και η πιθανότητα ψευδούς αποδοχής ή ψευδούς απόρριψης ενός προϊόντος σε σχέση με κάποια προδιαγραφή.

Προϋπόθεση για την εφαρμογή των παραπάνω αποτελεί η χρήση κάποιου κανόνα απόφασης ο οποίος θα πρέπει να έχει συμφωνηθεί ή κοινοποιηθεί στους χρήστες των αποτελεσμάτων των μετρήσεων. Οι κανόνες απόφασης συνήθως ορίζουν μια ζώνη προστασίας πάνω ή κάτω από το όριο μιας προδιαγραφής μετατρέποντας το διάστημα αποδοχής ενός προϊόντος είτε σε περισσότερο αυστηρό είτε σε περισσότερο ελαστικό σε σχέση με το διάστημα που ορίζεται από τα όρια των προδιαγραφών. Στα Σχήματα ΣΤ.1 και ΣΤ.2 παρουσιάζονται περιπτώσεις ελαστικότερων ή αυστηρότερων διαστημάτων αποδοχής και απεικονίζονται οι μέγιστες πιθανότητες λήψης λανθασμένης απόφασης (ψευδούς απόρριψης ή ψευδούς αποδοχής).



Σχήμα ΣΤ.1 Απεικόνιση μέγιστης πιθανότητας ψευδούς απόρριψης ενός προϊόντος του οποίου η μέτρηση συγκρίνεται με το άνω ή το κάτω όριο συγκεκριμένης προδιαγραφής στην περίπτωση «ελαστικής» αποδοχής (relaxed acceptance).



Σχήμα ΣΤ.2 Απεικόνιση μέγιστης πιθανότητας ψευδούς αποδοχής ενός προϊόντος του οποίου η μέτρηση συγκρίνεται με το άνω ή το κάτω όριο συγκεκριμένης προδιαγραφής στην περίπτωση «αυστηρής» αποδοχής (stringent acceptance).

Οι πιθανότητες ψευδούς αποδοχής ή ψευδούς απόρριψης ενός προϊόντος σχετίζονται με τα σφάλματα Τύπου I και Τύπου II. Στα σφάλματα Τύπου I, συμμορφούμενα προϊόντα εσφαλμένα απορρίπτονται. Η ελαχιστοποίηση των σφάλματων Τύπου I κατά την αξιολόγηση της συμμόρφωσης σημαίνει ελαχιστοποίηση της πιθανότητας η πραγματική τιμή της μετρούμενης παραμέτρου να βρίσκεται εντός των προδιαγραφών όταν το αποτέλεσμα της δοκιμής είναι έξω από τα όρια προδιαγραφής. Από την άλλη πλευρά, στα σφάλματα Τύπου II, μη συμμορφούμενα προϊόντα εσφαλμένα γίνονται αποδεκτά. Η ελαχιστοποίηση των σφάλματων Τύπου II κατά την αξιολόγηση της συμμόρφωσης σημαίνει ελαχιστοποίηση της πιθανότητας η πραγματική τιμή της μετρούμενης παραμέτρου να βρίσκεται εκτός των προδιαγραφών, όταν το αποτέλεσμα της δοκιμής είναι μέσα στα όρια προδιαγραφής.

ΣΤ.3 Μεθοδολογίες αξιολόγησης της συμμόρφωσης

ΣΤ.3.1 Αξιολόγηση συμμόρφωσης με χρήση δεδομένων πιστότητας μεθόδων (ISO 4259)

Τα καύσιμα αγοράζονται και πωλούνται, σύμφωνα με συγκεκριμένες προδιαγραφές και όρους πώλησης, ενώ συχνά γίνεται και αναφορά στις μεθόδους δοκιμών που πρέπει να χρησιμοποιούνται σε περίπτωση επίλυσης διαφορών πωλητή – αγοραστή για θέματα ποιότητας. Για παράδειγμα, η νομοθεσία της Ευρωπαϊκής Ένωσης για τα καύσιμα αυτοκινήτων, που υιοθετεί τις απαιτήσεις που αναφέρονται στα πρότυπα

EN 590 και EN 228, κάνει αναφορά στο πρότυπο ISO 4259 για την επίλυση διαφορών. Το πρότυπο ISO 4259 περιγράφει μια μεθοδολογία σύμφωνα με την οποία το αποτέλεσμα μια δοκιμής μπορεί να συνδυαστεί με τα δεδομένα πιστότητας της μεθόδου δοκιμής (όρια επαναληψιμότητας *r*, όρια αναπαραγωγιμότητα *R*), προκειμένου να αποφασιστεί κατά πόσον ένα δείγμα καυσίμου πληροί συγκεκριμένες προδιαγραφές. Οι προμηθευτές και οι αποδέκτες των προϊόντων καυσίμων μπορούν να χρησιμοποιήσουν τις πληροφορίες που περιλαμβάνονται στο πρότυπο ISO 4259 για να αξιολογήσουν την ποιότητα τους βασιζόμενοι σε μία ή περισσότερες εργαστηριακές μετρήσεις. Στον Πίνακα ΣΤ.1 παρουσιάζονται οι μαθηματικές εκφράσεις για τον υπολογισμό ελαστικότερων ή αυστηρότερων ορίων αποδοχής σύμφωνα με όσα προβλέπονται στο ISO 4259.

Πίνακας ΣΤ.1 Εξισώσεις για τον υπολογισμό ορίων αποδοχής για 95% και 99% επίπεδα εμπιστοσύνης σύμφωνα με το ISO 4259.

	Όρια αποδοχής κατά την παραλαβή καυσίμων («αυστηρή» αποδοχή)		**Όρια αποδοχής κατά τον επίσημο έλεγχο καυσίμων («ελαστική» αποδοχή)**	
	95% επίπεδο εμπιστοσύνης	**99% επίπεδο εμπιστοσύνης**	**95% επίπεδο εμπιστοσύνης**	**99% επίπεδο εμπιστοσύνης**
Κατώτερο όριο	$T_L + 0{,}59R$	$T_L + 0{,}83R$	$T_L - 0{,}59R$	$T_L - 0{,}83R$
Ανώτερο όριο	$T_U - 0{,}59R$	$T_U - 0{,}83R$	$T_U + 0{,}59R$	$T_U + 0{,}83R$

Οι εκφράσεις του παραπάνω πίνακα εφαρμόζονται στην περίπτωση αποτελεσμάτων από μεμονωμένες μετρήσεις. Σε περίπτωση αποτελεσμάτων που προκύπτουν ως μέσος όρος πολλαπλών (*k*) μετρήσεων, το όριο αναπαραγωγιμότητας *R* πρέπει να αντικατασταθεί από το:

$$R_1 = \sqrt{R^2 - r^2\left(1 - \frac{1}{k}\right)} \qquad (\Sigma\Tau.1)$$

Με βάση τις εκφράσεις του Πίνακα ΣΤ.1 υπολογίστηκαν όρια απόφασης για όλες τις παραμέτρους που περιγράφονται στα πρότυπα EN 228 (αμόλυβδη βενζίνη) και EN 590 (ντήζελ κίνησης) (βλ Πίνακα ΣΤ.2). Στα πρότυπα αυτά καθορίζονται τόσο οι προδιαγραφές όσο και οι πρότυπες μέθοδοι δοκιμών με σύμφωνα με τις οποίες θα πρέπει τα καύσιμα να ελέγχονται. Από τα κείμενα των προτύπων μεθόδων

ελήφθησαν τα δεδομένα πιστότητας που χρησιμοποιήθηκαν για τον υπολογισμό των ορίων αποδοχής.

Πίνακας ΣΤ.2 Ιδιότητες καυσίμων αυτοκινήτων για τις οποίες έχουν καθοριστεί προδιαγραφές βάσει των EN 590 και EN 228.

Ιδιότητες ντήζελ κίνησης - EN 590	**Ιδιότητες αμόλυβδης βενζίνης - EN 228**
o Αριθμός κετανίου o Δείκτης κετανίου o Πυκνότητα στους 15 °C o Πολυκυκλικοί αρωματικοί υδρογονάνθρακες o Περιεκτικότητα σε θείο o Σημείο ανάφλεξης o Ανθρακούχο υπόλειμμα (επί 10% υπολείμματος αποστάξεως) o Περιεκτικότητα σε τέφρα o Περιεκτικότητα σε νερό o Αιωρούμενα σωματίδια o Διάβρωση χάλκινου ελάσματος (3 ώρες σε 50 °C) o Περιεκτικότητα σε FAME o Αντοχή στην οξείδωση o Λιπαντικότητα, διορθωμένη διάμετρος φθοράς σφαιριδίου στους 60°C o Ιξώδες στους 40°C o Θερμοκρασία αποφράξεως ψυχρού φίλτρου o Απόσταξη - % συμπύκνωμα στους 250 °C - % συμπύκνωμα στους 550 °C - συμπύκνωμα 95% (v/v) στους	o Ερευνητικός αριθμός οκτανίου, RON o Αριθμός οκτανίου κινητήρα, MON o Περιεκτικότητα σε μόλυβδο o Πυκνότητα (στους 15°C) o Περιεκτικότητα σε θείο o Αντοχή στην οξείδωση o Περιεχόμενα κομμιώδη o Διάβρωση χάλκινου ελάσματος (3 ώρες σε 50 °C) o Ανάλυση υδρογονανθράκων - ολεφίνες - αρωματικές ενώσεις o Περιεκτικότητα σε βενζόλιο o Περιεκτικότητα σε οξυγόνο o Τάση ατμών (DVPE) o Δείκτης ατμόφραξης (VLI) o Οξυγονούχες ενώσεις: - μεθανόλη - αιθανόλη - ισοπροπυλική αλκοόλη - ισοβουτυλική αλκοόλη - τριτοταγής βουτυλική αλκοόλη - αιθέρες (με 5 ή περισσότερα άτομα C) - άλλες οξυγονούχες ενώσεις o Απόσταξη: - % απόσταγμα στους 70 °C - % απόσταγμα στους 100 °C - % απόσταγμα στους 150 °C - τελικό σημείο βρασμού (FBP)

ΣΤ.3.2 Αξιολόγηση συμμόρφωσης με χρήση εκτίμησης αβεβαιότητας

Η Οδηγία EURACHEM/CITAC "Use of uncertainty information in compliance assessment" περιγράφει τη χρήση των κανόνων απόφασης για τον καθορισμό ορίων αποδοχής και απόρριψης λαμβάνοντας υπόψη και την αβεβαιότητα των μετρήσεων. Έτσι, ενώ το πρότυπο ISO 4259 καθορίζει ζώνες προστασίας πάνω ή κάτω από τα όρια προδιαγραφής χρησιμοποιώντας αποκλειστικά δεδομένα

πιστότητας των προτύπων μεθόδων δοκιμών, η Οδηγία της EURACHEM / CITAC είναι πιο γενική στην εφαρμογή της και καθορίζει τις ζώνες προστασίας χρησιμοποιώντας τις εκτιμήσεις της αβεβαιότητας της μέτρησης. Συγκεκριμένα το πλάτος της ζώνης προστασίας ορίζεται ως ένα πολλαπλάσιο της τυπικής (συνδυασμένης) αβεβαιότητας, *u*, ανάλογα με το επίπεδο εμπιστοσύνης που έχει επιλεχθεί και το οποίο στην πράξη εκφράζει τη μέγιστη πιθανότητα λήψης λανθασμένης απόφασης (ψευδούς αποδοχής ή ψευδούς απόρριψης).

Η τυπική αβεβαιότητα που θα χρησιμοποιηθεί, ιδανικά θα πρέπει να περιλαμβάνει τόσο την αβεβαιότητα της αναλυτικής διαδικασίας, u_{analysis}, όσο και την αβεβαιότητα λόγω δειγματοληψίας, u_{sampling}. Στην περίπτωση αποτελέσματος που δίνεται ως μέσος όρος πολλαπλών (*k*) μετρήσεων η συνολική τυπική αβεβαιότητα δίνεται από τη σχέση:

$$u = \sqrt{u_{\text{sampling}}^2 + \left(\frac{u_{\text{analysis}}}{\sqrt{k}}\right)^2} \qquad \text{(ΣΤ.2)}$$

Στον Πίνακα ΣΤ.3 παρουσιάζονται οι μαθηματικές εκφράσεις για τον υπολογισμό ελαστικότερων ή αυστηρότερων ορίων αποδοχής λαμβάνοντας υπόψη εκτιμήσεις αβεβαιοτήτων.

Πίνακας ΣΤ.3 Εξισώσεις για τον υπολογισμό ορίων αποδοχής για 95% και 99% επίπεδα εμπιστοσύνης, λαμβάνοντας υπόψη εκτιμήσεις αβεβαιοτήτων.

	Όρια αποδοχής κατά την παραλαβή καυσίμων («αυστηρή» αποδοχή)		**Όρια αποδοχής κατά τον επίσημο έλεγχο καυσίμων («ελαστική» αποδοχή)**	
	95% επίπεδο εμπιστοσύνης	**99% επίπεδο εμπιστοσύνης**	**95% επίπεδο εμπιστοσύνης**	**99% επίπεδο εμπιστοσύνης**
Κατώτερο όριο	$T_L + 1{,}64u$	$T_L + 2{,}33u$	$T_L - 1{,}64u$	$T_L - 2{,}33u$
Ανώτερο όριο	$T_U - 1{,}64u$	$T_U - 2{,}33u$	$T_U + 1{,}64u$	$T_U + 2{,}33u$

ΣΤ.4 Εφαρμογή και σύγκριση μεθοδολογιών αξιολόγησης της συμμόρφωσης

Τα αποτελέσματα των αναλύσεων 769 δειγμάτων ντήζελ κίνησης από αντίστοιχο αριθμό πρατηρίων για τον προσδιορισμό της περιεκτικότητας σε θείο

χρησιμοποιήθηκαν για να γίνει η σύγκριση των διαφορετικών προσεγγίσεων σχετικά με τον ορισμό κανόνων απόφασης για την αξιολόγηση της συμμόρφωσης τους σε σχέση με το νομοθετικό όριο των 10 mg kg^{-1}.

Αν δεν γίνει χρήση ειδικών κανόνων απόφασης και τα αποτελέσματα συγκριθούν απλά με το νομοθετικό όριο (μηδενική ζώνη προστασίας), τότε 47 δείγματα (6,1%) θεωρούνται μη συμμορφούμενα. Χρησιμοποιώντας κανόνες λήψης αποφάσεων που βασίζονται στη χρήση ζωνών προστασίας πάνω ή κάτω από το όριο της προδιαγραφής οδηγούμαστε σε αριθμό μη συμμορφούμενων δειγμάτων είτε μικρότερο («ελαστική» αποδοχή) είτε μεγαλύτερο («αυστηρή» αποδοχή). Ο ακριβής αριθμός των μη συμμορφούμενων αποτελεσμάτων εξαρτάται από το επιλεγμένο επίπεδο εμπιστοσύνης και τον αριθμό επαναληπτικών εργαστηριακών μετρήσεων.

Δεδομένου ότι η προσέγγιση του ISO 4259 χρησιμοποιεί «ευρύτερες» ζώνες προστασίας σε σχέση με την προσέγγιση που κάνει χρήση εκτιμήσεων αβεβαιότητας, καταλήγει σε αυξημένο αριθμό μη συμμορφούμενων αποτελεσμάτων στην περίπτωση της «αυστηρής» αποδοχής και σε μειωμένο αριθμό μη συμμορφούμενων αποτελεσμάτων στην περίπτωση της «ελαστικής» αποδοχής. Οι ζώνες προστασίας που ορίζονται λαμβάνοντας υπόψη τις εκτιμήσεις αβεβαιοτήτων προσφέρουν ασφαλέστερους κανόνες λήψης αποφάσεων, λόγω του γεγονότος ότι οι εκτιμήσεις αβεβαιότητας αντιπροσωπεύουν την «αληθινή» διασπορά των τιμών του μετρούμενου μεγέθους (περιεκτικότητα σε θείο). Οι διαφορές των δύο προσεγγίσεων ορισμού ζωνών προστασίας και ορίων αποδοχής αλλά και η επίδραση του αριθμού των επαναληπτικών μετρήσεων είναι ιδιαίτερα έντονη στην περίπτωση της «αυστηρής» αποδοχής όπου το όριο απόφασης βρίσκεται χαμηλότερα του ορίου της προδιαγραφής σε μια περιοχή όπου συσσωρεύονται αρκετά αποτελέσματα μετρήσεων. Στους Πίνακες ΣΤ.4 και ΣΤ.5 παρουσιάζονται τα αποτελέσματα από την εφαρμογή όλων των προσεγγίσεων.

Πίνακας ΣΤ.4 Πλάτη ζωνών προστασίας, όρια απδοχής και αριθμός μη συμμορφούμενων δειγμάτων που έχουν εκτιμηθεί με χρήση δεδομένων πιστότητας μεθόδων (ISO 4259).

Επίπεδο εμπιστοσύνης	Αριθμός επαναληπτικών μετρήσεων, *k*	Πλάτος ζώνης προστασίας, *w* (mg kg^{-1})	Όρια αποδοχής κατά τον επίσημο έλεγχο καυσίμων («ελαστική» αποδοχή)			Όρια αποδοχής κατά την παραλαβή καυσίμων («αυστηρή» αποδοχή)		
			Όριο αποδοχής (mg kg^{-1})	Αριθμός μη συμμορφούμενων δειγμάτων		Όριο απδοχής (mg kg^{-1})	Αριθμός μη συμμορφούμενων δειγμάτων	
50%	*οποιοσδήποτε*	0,00	10,00	47	(6,1%)	10,00	47	(6,1%)
95%	1	1,32	11,32	25	(3,3%)	8,68	124	(16,1%)
	2	1,24	11,24	25	(3,3%)	8,76	114	(14,8%)
	3	1,21	11,21	25	(3,3%)	8,79	113	(14,7%)
99%	1	1,86	11,86	22	(2,9%)	8,14	173	(22,5%)
	2	1,74	11,74	22	(2,9%)	8,26	159	(20,7%)
	3	1,70	11,70	22	(2,9%)	8,30	156	(20,3%)

Πίνακας ΣΤ.5 Πλάτη ζωνών προστασίας, όρια αποδοχής και αριθμός μη συμμορφούμενων δειγμάτων που έχουν εκτιμηθεί με χρήση εκτιμήσεων αβεβαιότητας.

Επίπεδο εμπιστοσύνης	Αριθμός επαναληπτικών μετρήσεων, *k*	Πλάτος ζώνης προστασίας, *w* (mg kg^{-1})	Όρια αποδοχής κατά τον επίσημο έλεγχο καυσίμων («ελαστική» αποδοχή)			Όρια αποδοχής κατά την παραλαβή καυσίμων («αυστηρή» αποδοχή)		
			Όριο αποδοχής (mg kg^{-1})	Αριθμός μη συμμορφούμενων δειγμάτων		Όριο αποδοχής (mg kg^{-1})	Αριθμός μη συμμορφούμενων δειγμάτων	
50%	*οποιοσδήποτε*	0,00	10,00	47	(6,1%)	10,00	47	(6,1%)
95%	1	0,51	10,51	39	(5,1%)	9,49	63	(8,2%)
	2	0,41	10,41	41	(5,3%)	9,59	58	(7,5%)
	3	0,37	10,37	42	(5,5%)	9,63	57	(7,4%)
99%	1	0,73	10,73	35	(4,6%)	9,27	72	(9,4%)
	2	0,59	10,59	36	(4,7%)	9,41	66	(8,6%)
	3	0,53	10,53	39	(5,1%)	9,47	64	(8,3%)

ΣΤ.5 Προδιαγραφές αυτοματοποιημένου λογισμικού για τον υπολογισμό της πιθανότητας συμμόρφωσης καύσιμου με προδιαγραφές

Το λογισμικό «Υποστήριξης Αποφάσεων περί Συμμόρφωσης Καυσίμων ως προς Προδιαγραφές» έχει ως στόχο να υποστηρίξει όσους παραδίδουν, παραλαμβάνουν ή για οποιοδήποτε λόγο ελέγχουν προϊόντα καυσίμων και πρέπει βασιζόμενοι σε εργαστηριακές μετρήσεις να αποφασίσουν κατά πόσο αυτά συμμορφώνονται με νομοθετικές απαιτήσεις. Οι χρήστες του λογισμικού μπορεί να προέρχονται από όλα τα στάδια της εφοδιαστικής αλυσίδας των καυσίμων (διυλιστήρια, εταιρείες

εμπορίας – διακίνησης, πρατήρια, μεγάλοι καταναλωτές, βιομηχανίες, αεροπορικές εταιρείες, ναυτιλιακές εταιρείες κλπ.) και μπορεί να έχουν ρόλο αγοραστή (παραλαμβάνοντος) ή πωλητή (παραδίδοντος) ή και τα δύο. Επιπλέον, το λογισμικό μπορεί να χρησιμοποιηθεί και από ελεγκτικές αρχές που διενεργούν τον επίσημο έλεγχο των καυσίμων. Τα βασικά χαρακτηριστικά του λογισμικού είναι τα εξής:

- Το λογισμικό θα είναι διαθέσιμο διαδικτυακά αλλά και διαθέσιμο για εγκατάσταση σε προσωπικό υπολογιστή ή σε Tablet / Smartphone (application)
- Ο χρήστης θα εισάγει αποτελέσματα εργαστηριακών μετρήσεων και θα λαμβάνει ως αποτέλεσμα την πιθανότητα το υπό εξέταση προϊόν να συμμορφώνεται ή όχι με τις αντίστοιχες προδιαγραφές.

Σχήμα ΣΤ.3: Παράδειγμα υπολογισμών πιθανότητας συμμόρφωσης ή μη συμμόρφωσης.

- Ανάλογα με το είδος του χρήστη (αγοραστής, πωλητής, ελεγκτική αρχή) θα μπορούν να ορίζονται και συγκεκριμένοι κανόνες απόφασης.
- Στο λογισμικό θα είναι εισηγμένες οι προδιαγραφές διαφόρων ειδών καυσίμων όπως προκύπτουν από τη σχετική νομοθεσία και πρότυπα (π.χ. EN 228, EN 590) αλλά και τα δεδομένα πιστότητας των προτύπων μεθόδων (ISO, EN, ASTM) βάσει των οποίων ελέγχονται τα καύσιμα. Οι προδιαγραφές και τα δεδομένα αυτά θα επικαιροποιούνται κάθε φορά που υπάρχει μεταβολή είτε στη νομοθεσία είτε σε κάποια πρότυπη μέθοδο.
- Το λογισμικό θα παρέχει τη δυνατότητα τήρησης ιστορικού αναλύσεων και εξαγωγής στατιστικών.
- Το λογισμικό θα δίνει τη δυνατότητα στο χρήστη να δημιουργεί επίσημες επικυρωμένες αναφορές, με πληροφορίες για την ποιότητα του υπό έλεγχο καυσίμου.

Ε. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η αξιολόγηση της «καταλληλόλητας για χρήση» μιας αναλυτικής μεθόδου είναι άρρηκτα συνδεδεμένη με την **εκτίμηση της αβεβαιότητας** της μέτρησης η οποία ουσιαστικά χαρακτηρίζει την ποιότητα ενός αποτελέσματος συνυπολογίζοντας τόσο συστηματικά όσο και τυχαία σφάλματα. Επιπλέον η εκτίμηση της αβεβαιότητας των μετρήσεων με τη χρήση μιας επιστημονικά τεκμηριωμένης και έγκυρης μεθοδολογίας είναι μια βασική απαίτηση συγκεκριμένων διεθνών προτύπων ποιότητας βάσει των οποίων διαπιστεύονται εργαστήρια και φορείς (ISO/IEC 17025, ISO 15189, ISO/IEC 17043 ή ISO Guide 34). Στην παρούσα Διδακτορική Διατριβή παρουσιάζεται η ανάπτυξη και εφαρμογή στατιστικών και αριθμητικών μεθόδων για την εκτίμηση και χρήση της αβεβαιότητας μετρήσεων σε συγκεκριμένα στάδια του κύκλου της μέτρησης των καυσίμων: προ-αναλυτικά (δειγματοληψία), αναλυτικά (κυρίως μετρητική διαδικασία) και μετα-αναλυτικά (αξιολόγηση της συμμόρφωσης βάσει εργαστηριακών αποτελεσμάτων) (βλ. Σχήμα Β.1).

Στο **Κεφάλαιο 3** (Παράγραφος Γ Εκτεταμένης Περίληψης), εξετάζεται η **δειγματοληψία**, η οποία συνεισφέρει σημαντικά στην αβεβαιότητα των εργαστηριακών μετρήσεων καθώς αποτελεί ένα βασικό στάδιο των μετρητικών διαδικασιών. Περιγράφονται και συγκρίνονται ως προς τα αποτελέσματά τους, τρεις εμπειρικές στατιστικές μεθοδολογίες («κλασική» ANOVA, ανθεκτική ANOVA και στατιστική εύρους τιμών) χρησιμοποιώντας δεδομένα ενός ισορροπημένου πειραματικού σχεδίου (balanced experimental design). Οι τρεις μεθοδολογίες χρησιμοποιούνται για την εκτίμηση της αβεβαιότητας λόγω δειγματοληψίας καυσίμου (ντήζελ κίνησης) και λόγω της αναλυτικής διαδικασίας προσδιορισμού περιεκτικότητας σε θείο. Η διευρυμένη αβεβαιότητα της δειγματοληψίας κυμαίνεται από 0,34 εώς 0,40 mg kg^{-1}, ενώ η σχετική διευρυμένη αβεβαιότητα από 4,8 εώς 5,1%, ανάλογα με τη στατιστική μεθοδολογία που χρησιμοποιήθηκε. Τα αποτελέσματα της ανθεκτικής ANOVA (διευρυμένη αβεβαιότητα της δειγματοληψίας 0,34 mg kg^{-1}), η οποία δεν επηρεάζεται από την παρουσία μικρού αριθμού ακραίων (έκτοπων) τιμών στα δεδομένα, μπορούν να θεωρηθούν ως περισσότερο αξιόπιστα.

Στο **Κεφάλαιο 4** (Παράγραφος Δ Εκτεταμένης Περίληψης) εξετάζεται η εκτίμηση της αβεβαιότητας μιας τυπικής μετρητικής διαδικασίας. Η μεθοδολογία για την εκτίμηση της αβεβαιότητας μιας **μετρητικής διαδικασίας** όπως περιγράφεται στην Οδηγία ISO GUM “Guide to the Expression of Uncertainty in Measurement” και η αριθμητική μέθοδος Monte Carlo (MCM) χρησιμοποιούνται για την παράλληλη

εκτίμηση της αβεβαιότητας μέτρησης της θερμογόνου δύναμης πετρελαίου κίνησης με τη χρήση θερμιδομέτρου όλμου. Η διευρυμένη αβεβαιότητα (για πιθανότητα κάλυψης 95%) εκτιμήθηκε μέσω της μεθοδολογίας GUM (υποθέτοντας κανονική κατανομή) στα 0,28 MJ kg^{-1} ή 66,3 cal g^{-1}. Η τιμή αυτή είναι 12% μικρότερη από την τιμή που έδωσε η εφαρμογή της μεθοδολογίας MCM (0,32 MJ kg^{-1} ή 75,3 cal g^{-1}). Χρησιμοποιώντας τη μεθοδολογία GUM σε συνδυασμό με την εξίσωση Welch-Satterthwaite για τον υπολογισμό βαθμών ελευθερίας και στη συνέχεια, αποδίδοντας μια κατανομή *t* - student στο μετρούμενο μέγεθος οδηγούμαστε σε αυξημένη (σε σχέση με την υπόθεση της κανονικής κατανομής) διευρυμένη αβεβαιότητα (0,29 MJ kg^{-1} ή 70,4 cal g^{-1}), αλλά και πάλι κατά 7% χαμηλότερη από αυτή που δίνει η προσέγγιση MCM. Αυτές οι διαφορές μπορούν να αποδοθούν στη μικρή μη γραμμικότητα του μοντέλου μέτρησης και στη προσεγγιστική φύση της εξίσωσης Welch-Satterthwaite, η χρήση της οποίας στην προκειμένη περίπτωση υπερεκτιμά τους βαθμούς ελευθερίας. Μόνο στην περίπτωση χρήσης της Μπεϋζιανής προσέγγισης για την εκτίμηση των αβεβαιοτήτων Τύπου Α και στη συνέχεια χρήσης τους στο ισοζύγιο αβεβαιοτήτων της μεθοδολογίας GUM τα αποτελέσματα συμφωνούν με τα αποτελέσματα της προσέγγισης MCM. Επιπλέον, η αβεβαιότητα της μέτρησης της θερμογόνου δύναμης εκτιμήθηκε και με τη χρήση δεδομένων από διεργαστηριακές συγκριτικές δοκιμές ικανότητας (εμπειρική προσέγγιση).

Στο **Κεφάλαιο 5** (Παράγραφος Ε Εκτεταμένης Περίληψης) εξετάζεται η εκτίμηση της αβεβαιότητα μιας ειδικής κατηγορίας μετρητικής διαδικασίας που περιλαμβάνει και τη κατασκευή καμπύλης **βαθμονόμησης**. Εφαρμόζονται και συγκρίνονται 4 μεθοδολογίες εκτίμησης της αβεβαιότητας (GUM, MCM, Kragten, εξίσωση τυπικού σφάλματος) λόγω καμπύλης βαθμονόμησης που χρησιμοποιείται για τον προσδιορισμό της περιεκτικότητας καυσίμων σε θείο σύμφωνα με τη μέθοδο υπεριώδους φθορισμού (ISO 20846, ASTM D5453). Τα αποτελέσματα όλων των μεθοδολογιών συμφωνούν μεταξύ τους (τυπική αβεβαιότητα 0,172 – 0,175 ng μL^{-1}). Πραγματοποιήθηκαν επίσης και υπολογισμοί αγνοώντας τη συνδιακύμανση μεταξύ κλίσης και τεταγμένης (με τις με τις μεθοδολογίες GUM, MCM και Kragten) όπου φαίνεται ότι γίνεται κατά 62% υπερεκτίμηση της αβεβαιότητας. Αν στο αποτέλεσμα της προσεγγιστικής εξίσωσης του τυπικού σφάλματος δεν προστεθεί και η τυπική αβεβαιότητα της απόκρισης, οδηγούμαστε τότε σε υποεκτίμηση της αβεβαιότητας κατά 22%. Επιπλέον, δεδομένου ότι η εκτίμηση των 2 παραμέτρων μιας καμπύλης βαθμονόμησης βασίζεται σε ένα μοντέλο μέτρησης με πολλαπλά εξερχόμενα μετρούμενα μεγέθη (κλίση, τεταγμένη), εφαρμόζονται και οι βασικές αρχές της

συμπληρωματικής οδηγίας του GUM (Supplement 2 - Extension to any number of output quantities).

Τέλος, στο **Κεφάλαιο 6** (Παράγραφος ΣΤ Εκτεταμένης Περίληψης) εξετάζεται η **αξιολόγηση της συμμόρφωσης** ενός προϊόντος (καυσίμου) που βασίζεται σε εργαστηριακές μετρήσεις. Εφαρμόζονται δύο διαθέσιμες προσεγγίσεις (με χρήση δεδομένων πιστότητας μεθόδων και χρήση εκτίμησης αβεβαιότητας) για να υποστηρίξουν αξιόπιστες αποφάσεις σχετικά με την αξιολόγηση της συμμόρφωσης των καυσίμων. Τα αποτελέσματα των αναλύσεων 769 δειγμάτων ντήζελ κίνησης από αντίστοιχο αριθμό πρατηρίων για τον προσδιορισμό της περιεκτικότητας σε θείο χρησιμοποιήθηκαν για να γίνει η σύγκριση των διαφορετικών προσεγγίσεων ορισμού κανόνων απόφασης για την αξιολόγηση της συμμόρφωσης τους σε σχέση με το νομοθετικό όριο των 10 mg kg^{-1}. Χρησιμοποιώντας κανόνες λήψης αποφάσεων που βασίζονται στη χρήση ζωνών προστασίας πάνω ή κάτω από το όριο της προδιαγραφής οδηγούμαστε σε αριθμό μη συμμορφούμενων δειγμάτων είτε μικρότερο («ελαστική» αποδοχή) είτε μεγαλύτερο («αυστηρή» αποδοχή) σε σχέση με τη μη χρήση ζωνών. Ο ακριβής αριθμός των μη συμμορφούμενων αποτελεσμάτων εξαρτάται από το επιλεγμένο επίπεδο εμπιστοσύνης και τον αριθμό επαναληπτικών εργαστηριακών μετρήσεων. Επιπλέον υπολογίστηκαν όρια απόφασης για όλες τις παραμέτρους που περιγράφονται στα πρότυπα EN 228 (αμόλυβδη βενζίνη) και EN 590 (ντήζελ κίνησης).