UNIVERSITY OF OSLO
Department of informatics

Automatic evaluation of measurement data

Master thesis
60 credits

Brit Furnes-Wilkens

1. May 2007

Preface

This thesis is written as part of my Master's degree programme in Informatics (M.Sc.) at the Faculty of Mathematics and Natural Sciences, Department of Informatics, at the University of Oslo (UiO). The work was carried out at the National Laboratory (NL) at Justervesenet (JV) in cooperation with University Graduate Centre (UniK) in the period August 2006 to May 2007.

First of all, I would like to acknowledge UiO, UniK, and Justervesenet for making this thesis possible.

Secondly, I would like to express my sincere gratitude to my supervisors, Åsmund Sand and Tor Fjeldly, for tutoring me throughout my thesis. A special thank to Å. Sand for constant encouragements, guidance and supervision during this period.

Thirdly, I would like to thank Justervesenet for providing me with an office at NL and the employees for their cooperation and support. This has been vital for the completion of my thesis.

Finally, I express my greatest appreciation to my family. Especially my dear husband, Vidar Furnes-Wilkens, and my lovely children, Markus and Martine, who helped me keep my perspective on what is most important in life.

Brit Furnes-Wilkens
Kjeller, May 1st 2007

Abstract

Due to recent advances in computer technology and network infrastructure, databases and data programs are becoming increasingly important in the metrology area. This thesis introduces a new methodology for automatic evaluation of measurement data concerning calibration of electronic instruments. The methodology for developing a software-based analysis procedure is described.

The requirements for such an analysis procedure are discussed and necessary statistical methods are implemented in order to evaluate the measured data against the historical data. Hierarchical Bayesian [1] method is chosen since it meets the analysis requirements.

Implementation of the first version of the analysis procedure and tests has been performed. The range of use is specifically designed for electronic instruments. On the other hand there is nothing to suggest that the same approach cannot be used for other instruments.

The reader of this thesis would benefit from having some basic familiarity with both statistical methods and information technology (IT). However, some basic concepts are provided in the text.

Acronyms and Abbreviations

Here the most common symbols are given.

Abbreviations	*Description*
AC	Alternating Current
DC	Direct Current
ACV	AC Voltage
DCV	DC Voltage
ACI	AC Current
DCI	DC Current
ISO	International Standards Organization
OOT	Out of Tolerance
OOC	Out of Confidence
NL	National Laboratory
JV	Justervesenet
IT	Information technology
DUT	Device-under-test
PC	Personal computer
IDE	Integrated Development Environment
DLL	Dynamic Link Library
SOAP	Simple Object Access Protocol
HTTP	Hypertext Transfer Protocol
EU	European Union
ISG	Integrated Sciences Group
DLL	Dynamic-link library

Definitions

The following are definitions valid for the context of understanding the work of this thesis:

Measurand refer to the quantity intended measured.

Instrument specifications are the specific set of requirements agreed to by the manufacturer/producer of an instrument.

Calibration refers to the process of determining the relationship between to measurement instruments, of which one of them serves as a reference.

Metrology (from Greek 'metron' (measure), and -logy) is the science of measurement. Metrology includes all theoretical and practical aspects of measurement.

Metrologist is a person working on a NL that develops and evaluate calibrations systems that performs measurements on objects, in order to give the object a certificate.

Drift is a gradual and unintentional change in the reference value with respect to which measurements are made.

Error is the variance between read value and the true value.

Validity of a measurement is a statement of how well an instrument actually measures what it purports to measure.

Accuracy of a measurement refers to the freedom error, or the degree of conformance between the measurand and the standard.

Precision refers to the exactness of successive measurements.

Uncertainty of measurement is the quantified doubt of the result of a measurement.

Traceability refers to the absolute information about every step in a process chain. Concerning instruments, traceability can be used to certify their accuracy relative to a known standard, usually a national or international standard.

Contents

List of Tables

List of Figures

1 Introduction

This thesis theme is to find procedures, which will make it possible to conduct automatic evaluation of calibration. We will focus on building a methodology and develop an automatic evaluation application

In this first section an introduction and the background defines the problem addressed. Next we outline the motivation and the objectives. The last section describes the structure of this thesis, serving as a roadmap for the reader.

1.1 Introduction

Units of measurements have been used for all of antiquity. People have always used some type of set standard for trade. The Romans used the first known measurement device 2,000 years ago. They devised an equal beam scale that was shaped like the letter T with both arms measuring 7.4 in (18.8 cm) wide. Attached to each arm were metal pans that were typically 1.5 in (4 cm) in diameter.

> *"Weights and Measures may be ranked among the necessaries of life to every individual of human society. They enter into the economical arrangements and daily concerns of every family. They are necessary to every occupation of human industry; to the distribution and security of every species of property; to every transaction of trade and commerce; to the labours of husbandman; to the ingenuity of the artificer; to the studies of the philosopher; to the researches of the antiquarian; to the navigation of the mariner and the marches of the soldier; to all the exchanges of peace, and all the operations of war. The knowledge of them as in established use, is among the first elements of education and is often learned by those who learn nothing else, not even to read and write. This knowledge is riveted in the memory by the habitual application of it to the employments of men throughout life."*
>
> - John Quincy Adams

Weights and measures are mandated by state law to protect the interests of the buyer and seller to ensure honesty and integrity of everyday business transactions. This protection is accomplished through continuous and systematic inspection of all equipment that weighs or measures a commodity that is sold. Every transaction involving the exchange of goods, property, and service is affected in a very vital way by some form of weight and measures.

JV, a directorate of measuring technique placed under the Department of Trade and Industry, maintains metrology facilities and offers calibration services with high accuracy. Results of measurement values can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties. JVs references are traceable back to the national standards (see chapter 2.2). The calibration of electronic measuring instruments ensures that they are operating correctly and accurately.

Today, JV employees evaluate the calibration data differently and there is no centralized storage of historical data. This makes the process person-dependant where a person's past experience with the instruments plays a vital role.

To enable automatic evaluation of measurement data a centralized database and an analysis methodology must be established. Calibration history is stored in the database, which thereby maintains complete and traceable records. The analysis procedure includes statistical methods to predict the next measurement value based on historical data.

This chapter will introduce the concepts and main challenges related to the implementation of the analysis procedure. The objectives of the thesis then follow from these problems. Also, the motivation that started this work is mentioned.

1.2 Background

Today, when performing a calibration in a laboratory at JV, the employees manually evaluate the measurement data. This evaluation process includes comparing the measured values with the specification and the historical data of the instrument. Also the employee's past experience with the instrument is of vital importance for the evaluation of the measured values. The historical data is currently stored in different formats in spreadsheets by each employee.

It would be interesting to develop a new automated evaluation procedure and explore implementation and practical issues related to the procedure development. The idea is that such an approach gives easier access to the historical data, and the automatic evaluation procedure will be more effective and less person-dependant than the manual process today. Instead of using time on evaluating measurement values, the employees can focus their attention on the calibration process.

My supervisor's doctoral thesis deals with applying the Internet to instrumentation and metrology. My automatic evaluation method will be a part of this work and play an important role considering fast and accurate evaluations of the measurement data. The fact that the operator cannot see the instrument being calibrated using remote calibration via Internet, gives rise to challenges concerned with authentication of instruments. In chapter 10.1.3 we will discuss whether an improved version of the automatic evaluation procedure can be used for authentication purposes.

1.3 Motivation

The information day at UniK is an annual seminar for master students at UiO, where companies in the Kjeller area hold presentations of master thesis of current interest. On the seminar, 17 of august 2006, I met Hans Arne Frøystein, leader of the NL at JV. I told him about my background in IT and information security. I stated that I was interested in these areas. Frøystein mentioned that he had colleagues working with IT in his department, and that they could be interested in offering me a master thesis on this theme. Since I already had taken enough master subjects at Høgskolen i Agder (HiA), approved by UiO, I could start on a long master thesis, lasting two semesters.

By recommendation from Frøystein I contacted his colleague Åsmund Sand, candidate for the doctorate degree at the NL. I participated in a meeting with Sand, his supervisor Harald Slinde and Jeanne Espedalen. The last-mentioned takes a master degree in information security at Gjøvik. On this meeting we agreed that the theme for my master thesis should be "Authentication of electrical instruments by automatic consideration of measurement data". Sand was appointed external supervisor, while Tor Fjeldly, professor II at UiO, was to function as my internal supervisor. There were some initial concerns about the possibility of successfully authenticating instruments purely based on historical measurement data, and the theme of the thesis soon changed to "Automatic evaluation of measurement data." The possible uses of the automatic evaluation approach for authentication issues are described in more detail in the discussion part (see chapter 10.1.3).

1.4 Objectives

The first objective of this thesis is to define a methodology for automatic evaluation of measurement data during a calibration process. This will be based on a survey of the requirements of Justervesenet, the customers and the users (employees).

The second objective will be to start implementing the automatic evaluation method in JVs environment and handle the practical challenges that arise.

The third objective is to discuss the possibilities of using the method to authenticate instruments in an Internet-enabled calibration scenario. Will improvements of the automatic evaluation method solve the challenges involved?

1.5 Structure of the Thesis

This thesis is presented in twelve chapters:

Chapter 1: An introduction outlining the general context of the work. This includes the background of the study, the motivation and its objectives.

Chapter 2: This chapter presents electrical instruments, and describes the digital multimeters and measurement standards.

Chapter 3: An introduction to measurement techniques. Focus on the uncertainty related to a measurement value, the calibration process and evaluation of measurement data.

Chapter 4: Basic statistical theory. Overview of formulas of relevance for the research, including estimating predicted values and confidence intervals of a set of measurement data.

Chapter 5: An insight of the existing manual calibration process at JV. The employees' current evaluation procedures are examined.

Chapter 6: Two scenarios show the different ways to calibrate electrical instruments.

Chapter 7: The system requirements.

Chapter 8: The design of the database tables and their interactions. Relevant issues concerning the implementation of the automatic evaluation method are outlined in this chapter.

Chapter 9: The methodology for developing the automatic evaluation method.

Chapter 10: Discussion and evaluation of the work done. The issue of using the automatic evaluation method for authentication of electrical instruments is outlined.

Chapter 11: Conclusions and recommendations, reflecting on the objectives of this research.

Chapter 12: Possibilities for additional work on already implemented functionality as well as suggestions for new functionality.

In addition an appendix including complete equations, tables and explanation of theories used in this thesis is appended at the end.

2 Electrical instruments

In this chapter we give some basic knowledge of electrical instruments and their requisite for calibrations. Then we present the instruments we use in our testing (see chapter 8.5). Finally, the measurement standards used in order to perform a calibration are explained.

Electrical instruments include a large number and a wide variety of components. The characteristics of an instrument are determined by the circuit configuration and the values of the components. These values vary with time and the instrument therefore needs to be periodically calibrated (see chapter 3.4.3) to assure continued compliance with the specifications. Before the invention of the microprocessor a complex process was needed to calibrate electrical instruments. This included physical adjustment of components within the instrument. Today, internal software corrections have eliminated the need for physical adjustment. This applies to instruments, which support artifact calibration (see chapter 3.4.4).

2.1 Digital multimeters (DMMs)

A multimeter is an electrical measuring instrument, which combines three different meters in one, respectively an ammeter, a voltmeter, and an ohmmeter. The ammeter measures the electrical current running through a device (in amperes), the voltmeter measures the electrical potential difference across a device (in volts) and the ohmmeter measures the electrical resistance of a device (in ohms). Analog multimeters are sometimes referred to as "volt-ohm-meters", abbreviated VOMs. Digital multimeters are usually referred to as "digital-multi-meters", abbreviated DMMs.

DMMs are generally bench top or handheld. Handheld devices are useful for basic faultfinding and field service work. Bench top instruments on the other hand can measure to a very high degree of accuracy and therefore will commonly be found in calibration laboratories. Here they can be used to characterise resistance and voltage standards, or adjust and verify the performance of multi-function calibrators.

Measurements made by both VOMs and DMMs include DC voltage, AC voltage, DC current, AC current and frequency. In addition VOMs include measurements on decibel measurement and DMMs include measurements on resistance, capacitance, range, time period and several special measurements. Both VOMs and DMMs include battery power. Other common features for VOMs are overloaded protection, temperature compensation, mirrored scale, range switch, diode test and battery test. For DMMs other included features are analog bar graph, dB readings, auto-ranging, adjustable sampling rate, programmable, data acquisition, data storage and logging, removable data storage and trigging. DMMs also often come with the following output interfaces: RS232, BCD (Binary-coded decimal) and D/A (Digital to analog) and GPIB (General Purpose Interface Bus)/IEEE 488. GPIB/IEE-488 is used to connect the DMM to a computer so that data and control information can pass between them, like e.g. RS232.

Modern multimeters are exclusively digital making the VOMs destined to become obsolete. In DMMs, the signal under test is converted to a digital voltage. An amplifier with an electronically controlled gain preconditions the signal.

Similarly, better circuitry and electronics have improved meter accuracy. While older analog meters might have basic accuracies of five to ten percent, modern portable DMMs can have accuracies as good as ±0.025%, and high-end bench-top instruments can have accuracies in the hundredths of parts per million.

2.2 Measurement standards

A measurement standard is a material measure, measuring instrument, reference material or measuring system intended to conserve or reproduce a unit or one or more values of a quantity to serve as a reference. The most accurate version of a certain measurement is called a primary standard.

JVs voltage standards are groups of Zener-standards. These voltage standards are calibrated regularly against their Josephson voltage standard. JVs working standards for resistance in ohm, is realized by the quantised Hall standard.

3 Measurement technique

This chapter includes basic theory of measurement technique. Metrology and what a calibration involves are explained. Next, we describe the term measurement uncertainty. Dealing with measurements a vital part is to evaluate the uncertainties. At the end of this chapter we explain the evaluation process.

3.1 Theory

> *“One cannot really claim to know much about a thing until one can measure it”.*
>
> - William Thompson, Lord Kelvin (1824-1907), a British physicist and mathematician.

Much of the history of science and engineering in general, and of electronics in particular, has been involved in measuring things. Whether a measurement is made in order to troubleshoot an existing circuit, to characterize and define a new circuit, or to find the value of some non-electronic physical variable (e.g. temperature), the common thread is the need for using some electronic device to make a measurement.

A measurement is the process of estimating an object's magnitude relative to some unit of measurement. The measurement is expressed as a number of units of the standard (a real number times a unit), such as distance being indicated by a number of miles or kilometres. Measurements are always made using an instrument of some kind. Examples of measuring instruments include rulers, thermometers, speedometers, weighing scales and voltmeters. In this thesis I am concentrating on electronic measuring instruments, and my testing is performed using digital multimeters.

In order to measure accurately, measuring instruments must be carefully constructed and calibrated. However, all measurements have some degree of uncertainty associated with them, which is usually expressed as a standard error of measurement. This means that while a measurement is usually given as a number followed by a unit, every measurement has three components; *the estimate*, *an error bound*, and *a probability* that the actual magnitude lies within the error bound of the estimate.

For example, we might say that the length of a certain plank might result in a measurement of *9* meters plus or minus *0.01* meters, with a probability of *0.95.* This result could be written:

9 m ± 0.01 m, at a level of confidence of 95%.

3.2 SI system

The International System of Units (SI from French Le Système international d`unitès) is the most widely used system of units, both in everyday commerce and in science. SI is a living set of standards where units are created and definitions are modified with international agreement as measurement technology progresses. The system is a metric measurement system that builds on seven basis units viewed in the table below.

SI base units		
Name	**Symbol**	**Quantity**
Metre	**M**	Length
Kilogram	**Kg**	Mass
Second	**S**	Time
Ampere	**A**	Electrical current
Kelvin	**K**	Thermodynamic temperature
Mole	**Mol**	Amount of substance
Candela	**Cd**	Luminous intensity

Table 1: SI base units that are nominally dimensionally independent.

From these seven base units several other units are derived. When dealing with electrical instruments three units are fundamental quantities. The first one is electrical current in ampere, which is the SI base unit second, the electromotive force in volt and last the resistance in ohm. The last two are derived from the electrical current and the derivation is done in table 2.

Volt	V	Electrical potential difference, Electromotive force	W/A = J/sA	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-1}$
Ohm	Ω	Electric resistance, Impedance, Reactance	V/A	$m^2 \cdot kg \cdot s^{-3} \cdot A^{-2}$

Table 2: SI derived units that are derived from the base unit ampere.

3.3 An example of a measurement system

Figure 1 shows a measurement system from process to observer. A person, e.g. an employee at a National laboratory, who needs information about measurement variables from a process, is defined as an observer. The purpose of a measurement system is to link an observer to the measurement process. The input to the measurement system is the true value of the variable and the system output is the measured value of the variable. The accuracy of the system can be defined as the closeness of the measured value to the true value. The accuracy of a real system is quantified using *measurement system error*, E, where

$$E = Read\ value - True\ value \Rightarrow E = System\ output - System\ input$$

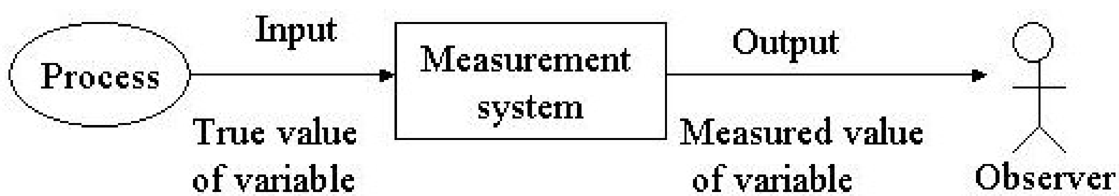

Figure 1: An overview of the measurement system from process to observer.

3.4 Metrology

Metrology is defined by the *International Bureau of Weights and Measures* (BIPM) as "...the science of measurement, embracing both experiment and theoretical determinations at any level of uncertainty in any field of Science and Technology."

Theoretically, metrology, as the science of measurement, attempts to validate the data obtained from test equipment. Though metrology is the science of measurement, in practical applications it is the enforcement and validation of predefined standards for precision, accuracy, traceability, and reliability. These standards can vary widely, but are often mandated by governments, agencies, and treaties such as the International Organization for Standardization, the Metre Convention, or the FDA (Food and Drug Administration). These agencies announce policies and regulations that standardize industries, countries, and streamline international trade, products, and measurements.

Metrology is, at its core, an analysis of the uncertainty of individual measurements, and attempts to validate each measurement made with a given instrument, and the data obtained from it. Dedicated calibration laboratories, often perform the dissemination of traceability to consumers in society. These laboratories possess recognized quality systems in compliance with standards. National laboratory accreditation schemes have been established to offer third-party assessment of such quality systems. A central requirement of these accreditations is documented traceability to national or international standards.

At the base of metrology are the definition, realisation and dissemination of units of measurement. The basic 'lineage' of measurement standards is:

1. The definition of a unit, based on some physical constant. Examples are absolute zero, the freezing point of water, or an agreed-upon arbitrary standard.

2. The realisation of the unit by experimental methods and the scaling into multiples and submultiples, by establishment of primary standards. In some cases an approximation is used, when the realisation of the units is less precise than other methods of generating a scale of the quantity in question. This is presently the situation for the electrical units in the SI, where voltage and resistance are defined in terms of the ampere, but are used in practice from realisations based on the Josephson effect and the quantised Hall effect. JV is realizing the unit for electrical resistance (Ohm), by the quantised Hall effect and the unit for electrical direct voltage, volt, by using the quant phenomenon Josephson effect.

3. The transfer of traceability from the primary standards to secondary and working standards. This is achieved by calibration.

3.4.1 Categories of metrology

Metrology is a very broad field and, in the European Union (EU), it is divided into three subfields with different levels of complexity and accuracy:

- **Scientific metrology** concerns the development and organisation of measurement standards (highest level). This includes establishment of measurement units, unit systems, development of new measurement methods and the transfer of traceability from the measurement standards to users in society.

- **Applied or industrial metrology** has to ensure adequate functioning of measurement instruments used in industry as well as in production and testing processes. This concerns the measurement instruments suitability, their calibration and quality control of measurements.

- **Legal metrology**, concerns regulatory requirements of measurements and measuring instruments for the protection of health, public safety and of economic transactions.

Fundamental metrology has no international definition. It may be described as scientific metrology since fundamental metrology signifies the highest level of accuracy within a given field, supplemented by those parts of legal and industrial metrology that require scientific competence.

3.4.2 Traceability

A core concept in metrology is traceability, defined as the ability to relate the results of individual measurements to national or international accepted standards, through an unbroken c hain of comparisons. Traceability can also be defined as the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties. The level of traceability establishes the level of comparability of the measurement, whether the result of the measurement can be compared to the previous one, a measurement result a year ago or to the result of a measurement performed anywhere else in the whole world.

In Europe, Industry ensures traceability to the highest international level by using accredited European Laboratories. In Norway, JV is the accredited laboratory with the highest national traceability level.

3.4.3 Calibration of electrical instruments

To ensure the traceability of a measurement, a basic tool is calibration of measuring instruments. Calibration involves the determination of the metrological characteristics of an instrument. It is achieved through direct comparisons with high-accuracy standards. A relationship between the indication of a measuring instrument and the value of a measurement standard is established. A calibration certificate is issued and, in most cases, a sticker is attached to the calibrated instrument. Calibrations performed at JV satisfy ISO 9000-requirements to traceability calibration for certified companies.

The main reasons why a customer is interested in having their electrical intruments calibrated are to ensure that readings from an instrument are consistent with other measurements, to determine the accuracy of the instrument readings and to establish the reliability of the instrument (i.e. that it can be trusted).

Calibration of DMMs will involve one or more of the following measurement areas: *AC-voltage, AC-current, DC-resistance, DC-voltage, DC-current, and Frequency.*

3.4.4 Artifact calibration

Some multifunction instruments have integrated components and software, which enable them to perform “self-calibration” by connecting a few external standards. Null detectors, AC/DC transfer standards and voltage dividers are examples of components that may be implemented internally in the instrument. A null detector is used to indicate a “balance” at zero volts. A measured value is balanced with an adjustable voltage source. The more sensitive the null detector is, the more precisely the adjustable source may be adjusted to equal the voltage under test. AC/DC transfer standards are among the basic electrical standards, which respond to both ac and dc in a known way. These standards are used to relate ac current and ac voltages to their counterparts. Voltage dividers are resistors that make available a voltage that is less than the potential difference between the two ends of the divider. They are used to create a voltage (Vout) that is proportional to another voltage (Vin). The “self-calibration” software is functionally identical to the manual calibration facility. The built-in software then performs the same functions as manual metrology functions.

In a traditional calibration the metrologist is performing comparisons using a total set of external devices. Artifact calibration on the other hand gives reduced process insight having parts of the calibration process placed inside the instrument. But artifact calibration also gives positive results like time-savings and equipment costs. Nevertheless a traditional calibration gives more precise results and a better overview of instrument drift.

A majority of multifunction calibrators and multimeters produced today support artifact calibration.

3.5 Uncertainty in measurement values

Uncertainty estimates play a critical role in quality assurance and control processes. Statistical analysis forms the basis of many of the decisions made in these areas. In fact, the ISO 9000 industry standards for quality assurance require that test measurements include an estimate of their uncertainty as specified in the International Standards Organization (ISO) Guide to the Expression of Uncertainty in Measurement (GUM). An example of a derived work is the National Institute for Standards and Technology (NIST) publication NIST Technical Note 1297 "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results".

The concept of uncertainty as a quantifiable attribute is relatively new in the history of measurement, although error and error analysis have since long been part of the practice of metrology. Uncertainty gives an indication of the quality of the measurement result. It gives a picture of the correctness of the stated result and the doubt about how well the result of the measurement represents the value of the quantity being measured. Expressing the uncertainty is important when comparing measurement results with other results, references, specifications or standards.

All measurements are subject to error, in that the result of a measurement differs from the true value of the measurand. Through calibration, given enough time and resources, most sources of measurement error can be identified and measurement errors can be quantified and corrected for.

The uncertainty of the result of a measurement generally consists of several components. The components are regarded as random variables, and may be grouped into two categories according to the method used to estimate their numerical values described below.

3.5.1 Type A

For type A uncertainty, evaluation of the standard deviation of the uncertainty means to evaluate the uncertainty using a statistical analysis of a series of observations. In this case the standard uncertainty is the experimental standard deviation of the mean that follows from an averaging procedure or an appropriate regression analysis.

3.5.2 Type B

For type B uncertainty, evaluation of the standard uncertainty involves evaluating the uncertainty by using other methods than statistical analysis of a series of observations. In this case the standard uncertainty is evaluated by scientific judgement based on all of the available information of possible variations of the measurement value, such as earlier measurement values, measurement uncertainty given in supplier specifications, measurement uncertainty given in a calibration report or certificate, and other reports or certificates, and experience with or general knowledge of the characteristics of relevant materials and instruments.

Type B evaluation of standard uncertainty of a measurement requires insight based on experience and general knowledge that can be learned with practice.

3.5.3 Evaluation of measurement data

Evaluation of an instrument's measured values deals with evaluating the uncertainty of the measurements, analysing measurement requirements and testing in order to decide if the measured values meet the specifications and customer criteria.

The employees at the NL must follow their department's measurement guidelines when calibrating a specific instrument. They also need to evaluate the measurement values according to the "Guide to the expression of uncertainty in measurement" (GUM) by the International Organization for Standardization, the specifications issued by the manufacturer of the instrument, and the criteria from the customer who owns the instrument. GUM gives information of the general rules for evaluating and expressing uncertainty in measurements (see chapter 3.5). The specifications issued by the manufacturer set the values, within which the measurand must lie. Customers, which own the instruments, also have different requirements for how accurate their instrument should measure. Therefore NL employees must evaluate the measurement values with these criteria in mind.

4 Basic Statistics on sets of numbers

In this chapter basic statistical methods are introduced. These formulas constitute an important role in the development of the automatic evaluation method. Two different analyses models are proposed. Next we introduce two different ways to do statistics. Finally, hierarchical Bayesian modelling and Bayesian metrology are described.

4.1 The mean and standard deviation

When performing measurements, mistakes can arise. Repeating measurements several times and carrying out some statistical calculations can prevent these mistakes. At the same time the amount of information from the measurements increases. By taking many readings, a mean or average can be estimated. This gives an estimate of the "true" value in events that gives variation in the repeated readings. Summing a series of values and dividing by the total number of values equals the mean or average.

The definition of the mean, $\overline{x}$, for N values x_1, ….., x_N:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \cdots + x_N}{N} \qquad (4.1)$$

The standard deviation of the measurements x_1, ….., x_N is an estimate of the average uncertainty of the measurements x_1, ….., x_N.

The definition of standard deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \qquad (4.2)$$

4.2 Confidence and Prediction intervals

A confidence interval (CI) is based upon sample standard deviations and is used in general statistical analysis, to tell us the range the mean of population will fall in. A CI gives an estimated range of values, which is likely to include an unknown population parameter. The estimated range is calculated from a given set of sample data. The mean ($\overline{X}$) of this random sample will, most likely, not be the true mean of the whole population (μ) but rather an estimate. If independent samples are taken repeatedly from the same population, and a CI is calculated for each sample, a certain percentage of the intervals will include the unknown population parameter. For the unknown parameter we can produce this percentage to be e.g. 90%, 95%, 99% or 99.9%. The lower and upper boundaries of a CI make the confidence limits. The width of the CI gives an idea about how uncertain the unknown parameter is. For a set of sample data the width increases as the confidence level increases. It is more likely that the true mean is included with a greater width.

When determining the CI for a sample of data the following values are needed:

1. M = sample mean

2. N = sample size

3. $s = \sqrt{\frac{\sum (X - M)^2}{(N-1)}}$,
   Where s = the sample standard deviation and X = each data value

4. $s_m = s / \sqrt{(N)}$ where s_m denotes the standard deviation of M

5. df = N-1 (df = degrees of freedom)

6. t= t-value obtained from a t-table according to confidence interval

The method to find the CI is given:

1. Take a random sample.

2. Find the mean of this sample.

3. Calculate s and s_m using the equations above.

4. Find the t-value using the t-table, which can be found in appendix A. In this table, to find a 90% confidence interval, use the .05 column, to find a 95% confidence interval, use the 0.25 column, due to the one-sided nature of the t-table.

Confidence intervals can be expressed as one-sided or two-sided. A one-sided CI (see Figure 17 in appendix A) includes either an upper or a lower boundary. The upper one-sided border defines a point that a certain percentage of the population is less than. Conversely, a lower one-sided border defines a point that a specified percentage of the population is greater than. A two-sided CI (see Figure 18 in appendix A) includes both an upper and a lower boundary.

While a CI estimates present population characteristics, a prediction interval (PI) estimates future values based on past background samples taken. A CI estimates the true population mean or other quantity of interest that cannot be observed whereas a PI predicts the distribution of individual points. PI is useful in determination of future values based upon present or past data.

4.3 T-Distribution

William S. Gosset discovered the t-distributions in 1908. He was a statistician employed at a Guinness brewery in Dublin and was not permitted to publish under his own name. Therefore, his paper was written under the pseudonym ‘Student’.

Student's t-distribution arises when (as in nearly all practical statistical work) the population standard deviation is unknown and has to be estimated from the data. The formula for student t-distribution is:

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \qquad (4.3)$$

Where,

n - sample size,

μ - the known expectation value of X,

$\overline{x}$ - sample mean,

s - sample analog to the standard deviation, σ.

When operating with a specific t-distribution, the degrees of freedom (df) must be specified. The quantity T in equation 4.3, has a t-distribution with n-1 df. This df is connected with the estimation of the sample standard deviation, s. The df for an estimate equals the sample size minus the number of additional parameters estimated for that calculation. The t-distribution is in conformity with the normal distribution. The larger the df, the closer the *t*-density is to the normal density. This reflects the fact that the standard deviation *s* approaches σ for large sample sizes.

The graph drawn in figure 2 can be used for both 1-sided and 2-sided (lower and upper) tests due to the symmetry of the *t*-distribution. Appropriate values of the significance level α must be chosen. The most common is α= 0.05.

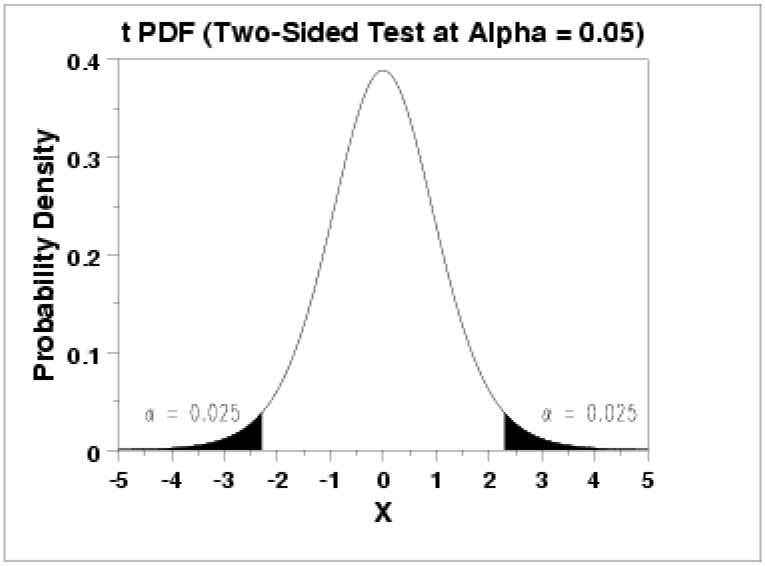

Figure 2: An overview of a graph that demonstrates the significance level, α. It plots a t-distribution with 10 df. The graph shows a two-sided test at significance level α= 0.05. Then the percent point function is estimated at α/2 (0.025).

4.4 Least-squares fitting

In chapter 4.1 repeated measurements are done on a single quantity. Investigation of the mathematical relationship between two variables, where the expected relation is linear, is probably the most important experiments of this type. The formula for a straight line is:

$$y = ax + b \qquad (4.4)$$

Given two measurement variables x and y that are linearly related and subject to no uncertainties, a graph of y against x, should be a straight line that has slope a and intersects the y-axis at y = b. Real measurement values involve uncertainties. A value point is presented as (x_i, y_i), where i runs from 1 to n (the total number of points). Such a value point can only be expected to lie close to the line, and a process that quantitatively estimates the trend of the points is necessary. Desired is a line that gives the minimal deviation from the value points, commonly known as the best-fitting curve. Considering a linear relation this curve is obtained by the method of least-squares line. In other cases a polynomial best fits the set of data points. A linear, second degree and third degree regression are derived in appendix A.

The general definition of the least-squares m^{th} degree least-squares:

$$f(x) = a_0 + a_1 + a_2 + ... + a_m x^m \qquad (4.5)$$

Then the least square error of the best fitting curve f(x):

$$\Pi = \sum_{i=1}^{n}[y_i - f(x_i)]^2 = \sum_{i=1}^{n}\left[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_m x_i^m)\right]^2 = \min.$$

Take the derivative with respect to the unknown coefficients a_0, a_1, a_2, …, and a_m, and set them equal zero:

$$
\begin{cases}
\dfrac{\partial \Pi}{\partial a_0} = 2\sum_{i=1}^{n} y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_m x^m)\Big] = 0 \\
\dfrac{\partial \Pi}{\partial a_1} = 2\sum_{i=1}^{n} x_i \Big[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_m x^m)\Big] = 0 \\
\dfrac{\partial \Pi}{\partial a_2} = 2\sum_{i=1}^{n} x_i^2 \Big[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_m x^m)\Big] = 0 \\
\vdots \\
\vdots \\
\dfrac{\partial \Pi}{\partial a_m} = 2\sum_{i=1}^{n} x_i^m \Big[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \ldots + a_m x^m)\Big] = 0
\end{cases}
$$

Expanding the above equations:

$$
\begin{cases}
\sum_{i=1}^{n} y_i = a_0 \sum_{i=1}^{n} 1 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 + \ldots + a_m \sum_{i=1}^{n} x_i^m \\
\sum_{i=1}^{n} x_i y_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 + \ldots + a_m \sum_{i=1}^{n} x_i^{m+1} \\
\sum_{i=1}^{n} x_i^2 y_i = a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 + \ldots + a_m \sum_{i=1}^{n} x_i^{m+2} \\
\vdots \\
\vdots \\
\sum_{i=1}^{n} x_i^m y_i = a_0 \sum_{i=1}^{n} x_i^m + a_1 \sum_{i=1}^{n} x_i^{m+1} + a_2 \sum_{i=1}^{n} x_i^{m+2} + \ldots + a_m \sum_{i=1}^{n} x_i^{2m}
\end{cases}
$$

The unknown coefficients a_0, a_1, a_2, …, and a_m can be obtained by solving the above linear equations [4].

The least square regression approach gives a curve that represents the general trend of the data. This approach is useful in cases where the data have significant error or noise, or when there are more data points than the number of unknown coefficients. In the first case there is possibility of error in any data point and the regression curve is made to follow the pattern of the data points taken as a group. A least square regression can also describe the relationship between a dependent variable and multiple independent variables, called multiple regression.

4.5 Considering analysis models

An analysis model describes the structure of a system or application. Such a model only includes the logical implementation of the functional requirements. How the technically implementation will be, is not described. An analysis model consists of class diagrams and sequence diagrams that describe the logic structure e.g. identified in a use case model.

I have considered two different analysis models and chosen the model that best meets the functional requirements of the automated evaluation method.

4.5.1 Procedural analysis

A procedural analysis (see figure 3) breaks down the physical steps that a learner must go through, so that a task can be successfully achieved. The steps that make up a task are arranged linearly and sequentially, illustrating where the learner begins and ends. The steps throughout the task, from start to finish, as well as any decisions that the learner must make are usually arranged in a flowchart. Examples of procedural analysis in nature are changing a tire or formatting a disk.

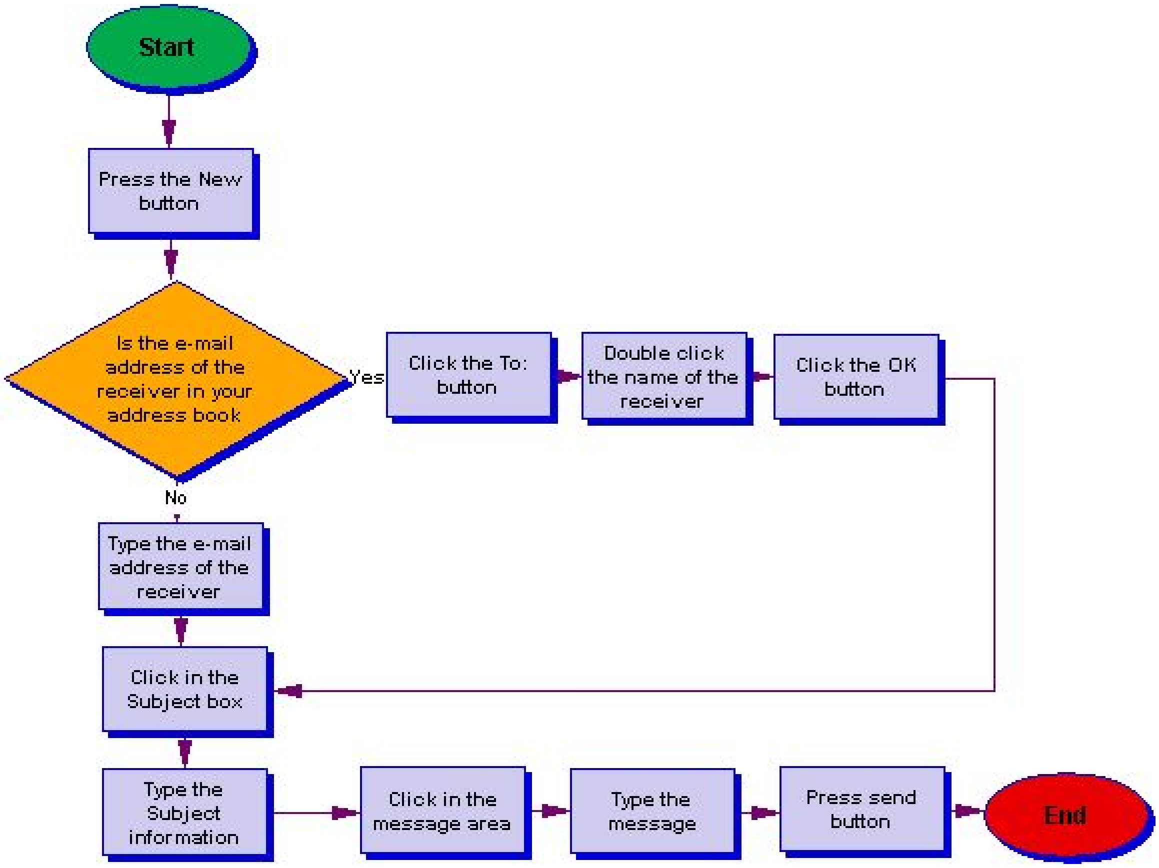

Figure 3: A flowchart showing a procedural task analysis from start to end.

4.5.2 Hierarchical analysis

"A hierarchy is an organization of elements that, according to prerequisite relationships, describes the path of experiences a learner must take to achieve any single behaviour that appears higher in the hierarchy"", [18]. In a hierarchical analysis, breaking down the tasks from top to bottom shows a hierarchical relationship amongst them. Thereafter instruction is sequenced bottom up.

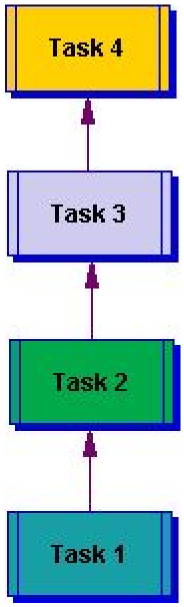

Figure 4: Example of four tasks having hierarchical relationships.

Figure 4 shows a hierarchical relationship. Task four has been broken down into its enabling tasks. The learner cannot perform the third task until he/she has performed the first and second tasks.

4.6 Two different ways to do statistics

"Bayesian" refers to the Reverend Thomas Bayes (b. 1702, London – d. 1761). Bayes was both a clergyman and an amateur scientist/mathematician.

In the early 18th century the development of probability theory arose to answer questions in gambling, and to back up the new and related ideas of insurance. A problem known as the question of inverse probability emerged: "the mathematicians of the time knew how to find the probability that, say, 4 people aged 50 die in a given year out of a sample of 60 if the probability of any one of them dying was known. But they did not know how to find the probability of one 50-year old dying based on the observation that 4 had died out of 60" [5]. Thomas Bayes had an answer to this problem, known as Bayes theorem, which was published in 1763. This theorem is a simple mathematical formula used for calculating conditional probabilities. Conditional probability can be explained as a measure of the probability that one event occurs given another.

In the first half of the nineteenth century R.A. Fisher, E. Pearson and J. Neyman developed the theory of statistics and probability using frequency probability. Frequency probability is any probability that can be found by sampling a large amount of data. For example if you want to find out how many red cars that drives past your house, you sample the next 100 cars, and estimates how many of them was red. Supporters of frequency probability are called Frequentists. Broadly speaking, the 20th century statistics was Frequentists while 19th century was Bayesians.

Bayesians and Frequentists have different views on the interpretation of the term "probable", and a controversy between them came to be. The background for the disagreement was that both groups wanted their own approach to reflect the commonly meaning of the term. Frequentists accuse Bayesians of being subjective, because their analyses are not determined from the data. On the other hand a negative thing with frequency probability is that it cannot assign probabilities to things outside their scope. Their scope is the sample that is taken from a large amount of data.

Importantly the two groups have agreed that Bayesian analysis and frequency analysis answer genuinely different questions. In a Bayesian view the information about that case is quantified as completely and consistent as possible. But the Bayesian approach does not support a process whereby a probability can be attached to the future occurrence of special case. However it allows such a probability to be specified. Bayesian methods are based on the data actually observed, and are therefore able to assign posterior probabilities to any number of hypotheses directly. Frequentists aim for universally acceptable conclusions, ones that will stand up to opponent investigation. A common frequentists tactic is to split problems up, and focus on the objectivity on a subset of the data, that can be analyzed optimally.

In many applications the Bayesian methods are more general and seem to give better results than frequency probability. The Bayesian spam filter is a recent example of an application that uses Bayesian techniques. The filter involves a set of e-mails that defines what originally is believed to be spam. The application uses these definitions to decide whether a new e-mail is spam or not. The new e-mails are regarded as new information. If a user discovers that an e-mail is being mistaken as legitimate or being spam he can update the defined set of e-mails with the new information. Hopefully future applications then will involve a more accurate set of definitions.

4.7 Bayesian hierarchical modelling

A framework for Bayesian hierarchical modelling consists of three Bayesian principles.

These principles are as followed:

- Combine information.
- Uncertainty quantification and management (inputs and outputs are probability distributions).
- Decisions making.

To carry out Bayesian modelling in practice, hierarchical thinking and seeking effective data-model compromises are requested. The goal of using a Bayesian framework will in our work be to find information about significant changes in an instruments measurement values. In our application we use a Bayesian line of though for the uncertainty calculation but frequency probability is applied to historical data in order to calculate the predicted next value.

4.8 Bayesian Metrology

Bayesian methods are likely to have impact in metrology. Many problems in metrology are suitable for Bayesian approaches. Often substantive prior knowledge is available in the form of prior calibration history. Each time the instrument is used to make a measurement we learn more about the non-ideal behaviour of the measuring device. By assigning prior distributions to the standard deviation we can predict the probability for a posterior distribution. Then we can compare the true measurement with the posterior value and see how good the true value fits the prior distributions.

Difficulties arise in analyzing data where it is expected that there is a drift or sudden changes in the travelling standard. In these problems, Bayesian methods offer significant advantages over classical approaches. The methods can be used to examine more comprehensively the quality of the fit of the model to the data and to select from a range of potential models on a probabilistic basis.

Recognizing the potential of Bayesian methods, statisticians from the National Institute of Standards and Technology (NIST) have begun exploring the use of these methods in several metrological applications. After some initial research, a five-year competence initiative on Bayesian metrology was started in FY99. Four specific areas were targeted: traceability, inter-laboratory comparisons, calibration, and part inspection. These areas have potential benefit from Bayesian methods.

So far, members of NIST have completed work on a Bayesian model in several areas namely in inter-laboratory comparisons, exploring the relationship between the ISO uncertainty procedure and Bayesian statistics, presenting a review of Bayesian statistics to NIST staff, examining the use of Bayesian statistics in the certification of reference materials and applying a Bayesian decision rule to the part inspection problem. Figure 5 displays the reduced cost of using a Bayesian decision rule in the part inspection problem, and the sensitivity of the result to various misspecifications. The left diagram in figure 5 shows the resulting cost of the decision rule not using prior information. The three points correspond to three estimates of the measurement uncertainty, where the value of 100% is correct. The right diagram in figure 5 contains the same information for the decision rule that uses prior information. The three lines correspond to three estimates of the prior uncertainty, again where the value of 100% is correct. This experiment shows that even using incorrect values of the prior or measurement uncertainty, the Bayesian approach might result in lower cost.

In the future, research on Bayesian statistical methods for metrological applications and demonstrating their value to NIST problems will continue.

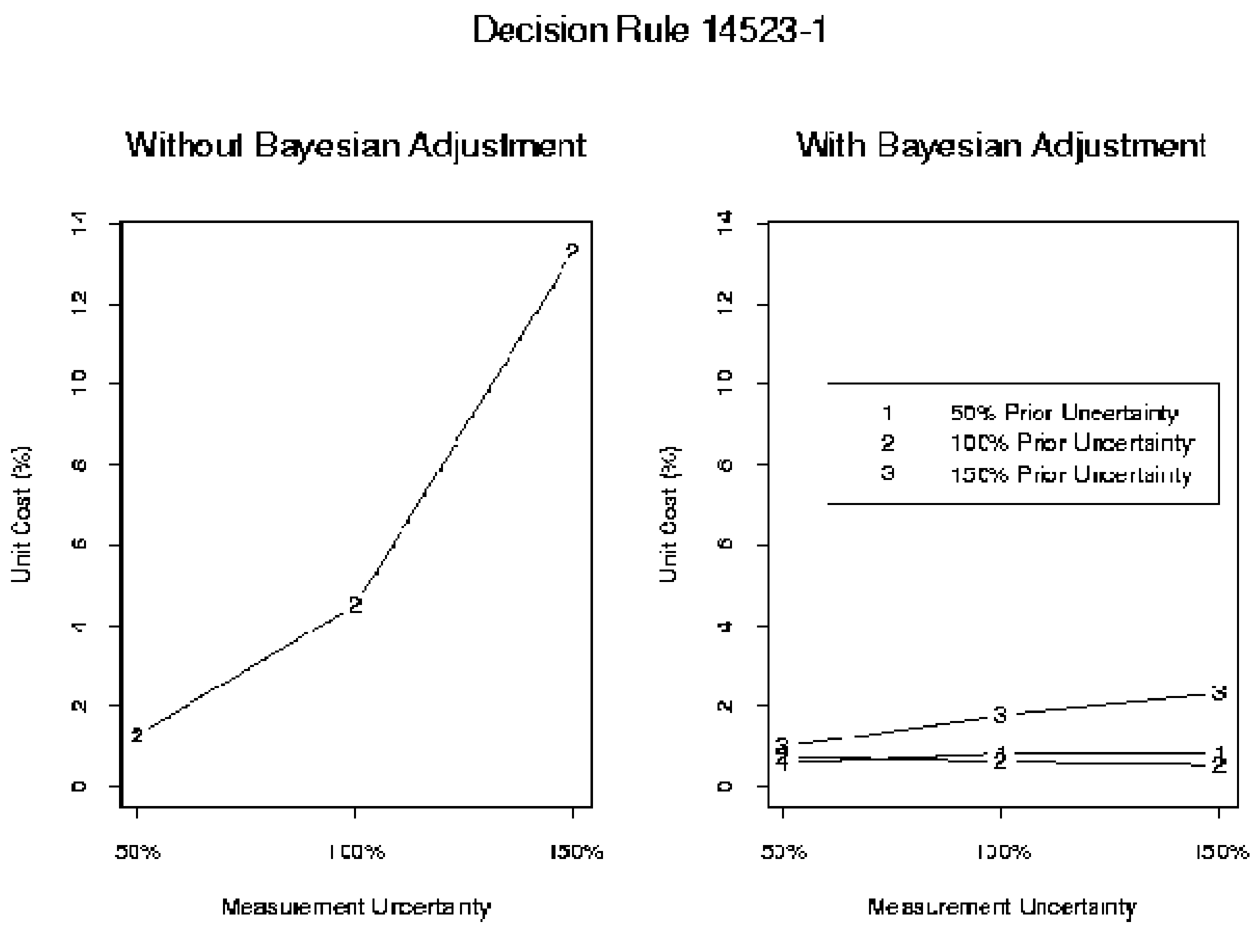

Figure 5: An example from NIST showing reduced costs when applying a Bayesian decision rule.

5 Existing solution at Justervesenet

For electrical instruments JV offers calibration services in the areas of dc current (DCI), dc voltage (DCV), ac current (ACI), ac voltage (ACV) and resistance. Today, the employees at the NL manually perform evaluation of the measurement data obtained from the calibration process. This chapter describes how this manual evaluation is performed. This includes the collection of the measurement values, the uncertainty budget and the procedures used.

5.1 Current calibration methodology

Calibration of an electrical instrument involves calibration of one or more of the following areas: DCI, DCV, ACI, ACV and resistance. Depending on the calibration area and the type of instrument, different standards are used in the calibration process. If the device-under-test (DUT) supports artifact calibration such a calibration is done first. Thereafter a procedure for resistance measurements is performed. Unless no suspicious values appear, measurements are performed in the following order: DCI, DCV, ACI and ACV. If nonconforming values appear, error-finding procedures are performed.

A DMM calibration includes all of the above areas as well as artifact calibration. A Multifunction Calibrator is (at JV) used as the standard.

5.1.1 The data collection

LabVIEW (Laboratory Virtual Instrumentation Engineering Workbench) is a platform and development environment for a visual programming language from National Instruments. In the laboratory, during a calibration, a standard and the DUT are connected to a PC running LabVIEW [23]. All the raw data from a measurement is collected in this program. The data is stored in a text file and then imported to a spreadsheet.

5.1.2 Algorithm

The metrologists follow certain procedures when calibrating multimeters using a calibrator as the standard. The procedures are different according to the measurement area used. The main similarity between these procedures is to perform several readings, usually between 5 and 20. The mean value of the readings is then estimated and corrected, by evaluating offset and thermal effects. The variance between corrected mean and the ’true’ calibrator value is calculated. Comparing the unknown resistance-standard against JV`s working-standards does calibration on resistance.

5.1.3 Usage of the data

The spreadsheet contains all information needed to evaluate the calibrated measurement values and to make a certificate for the DUT. It contains calculations of standard deviation in parts per million (PPM), specifications, uncertainty and the corrected value (the measured value minus the error on the standard) among others. Graphs based on the data in the spreadsheet are used to evaluate the measurement values. The axes of the graphs depend on which calibration area is evaluated. When performing resistance calibration, the measured values in PPM are plotted on the y-axis against the time on the x-axis. Calibration of DMMs involves performing measurements on many different nominal values. DC currents and DC voltages are therefore plotted with the nominal value on the x-axis and the variance of the measured value and the nominal value in PPM on the y-axis. The data is plotted against the frequency on the x-axis and the measured value in PPM on the y-axis when doing AC currents and AC voltages. A linear regression line is estimated and plotted in the graph. Each line represents a calibration performed in one year. This way all historical information can be plotted in the same graph where each line represents a different calibration year.

5.1.4 Measurement uncertainty

Metrologists calculate the total uncertainty in the calibration and they set up uncertainty budgets according to JV`s procedures. Such a budget is an analysis of the uncertainty for a measurement. The analysis should include all sources to uncertainty with the belonging standard uncertainties for the measurement, and the methods used for evaluating them. In the tables of a budget all quantities should be referred to with a physical symbol x_i, the belonging standard measurement uncertainty $u(x_i)$, the sensitivity factor c_i and the different contributions $u_i(y)$ should be specified. The units of every quantity should also be stated with the numerical values in the table.

Below is an example of a measurement function used when calibrating a multimeter in the DCI area:

$$I_x = I_{nom} + \lfloor (I_{meas} - I_{Corr}) - {*}I_{ref} \rfloor + I_{drift} + I_{res} \qquad (5.1)$$

Where,

I_x = The current the multimeter displays with forced nominal current.
I_{nom} = The nominal current of the Calibrator.
I_{meas} = Current read from the multimeter display.
I_{corr} = Correction of offset/thermal currents.
I_{ref} = The current value of the Calibrator.
I_{drift} = Drift/ stability based on history from the Calibrator.
I_{res} = The resolution of the multimeter.

An example on a uncertainty budget:
I_{nom} =10 mA

Quantity X_i	Estimate x_i	Standard uncertainty $u(x_i)$	Distribution function	Sensitivity factor c_i	Uncertainty contribution $u_i(y)$	Degrees of freedom ν_i
I_{nom}	10 mA	0 nA	-	1	0 nA	
I_{meas}	9,999 941 mA	1 nA	normal	1	1 nA	9
I_{corr}	0 mA	0,003 nA	normal	-1	0,003 nA	9
I_{ref}	9,999 952 mA	40 nA	normal	-1	40 nA	∞
I_{drift}	0 mA	40 nA	normal	1	65 nA	∞
I_{res}	0 mA	0,6 nA	square	1	0,6 nA	∞
I_x	9,999 99 mA			$u_c(y)$=	80 nA	∞

Table 3: The uncertainty budget for DCI calibration of a multimeter. Here there is no correlation between the input-parameters. If such correlations occur there will be an extra table.

5.2 Manual evaluation method

Evaluation of the measurement data is performed in different ways depending on the amount of historical data available, the area being calibrated and the person carrying out the calibration process. If the DUT has been calibrated before, the following procedures are performed resulting in verified measurement values:

Steps used for previously calibrated instruments:

1. Perform several, usually five, measurement series on each measurement point.

2. Estimate the mean value of the series for each measurement point and store the result in a spreadsheet file.

3. Compare the measured values with the results from the standard. Subtract the offset of the standard in order to subdue the instruments error level.

4. Compare error values with the instrument's specifications and results from historical measurement data.

5. Rely on your past experience with the instrument, and check if the results are according to the instrument's expected behaviour.

6. If the measurement data does not lie outside the specifications and if there is no drastic increase in the standard deviation, the data is verified.

If the DUT is calibrated for the first time, no stored historical data is available. Then the evaluation process is based on the specifications and the experience that metrologists have with similar instruments, if any. The following procedure is carried out resulting in verified measurement values:

Steps used for first time calibration:

1. Follow steps 1) and 2) in the list above.

2. Evaluate the measurement data by looking at the instrument's specifications and by comparing the results with the results from any similar instruments. If the measured data lies inside the specifications and they do not differ much from the behaviour of similar instrument the data is verified.

The calibration of a DMM goes through different areas in an orderly fashion. The first area must be completed and verified before progressing to the second area and so on. The five areas associated with the calibration of a DMM include DCV, DCI, ACV, ACI and Resistance. Figure 6 shows that the more components the area includes in order to make the intended measurements, the higher the margin for faults.

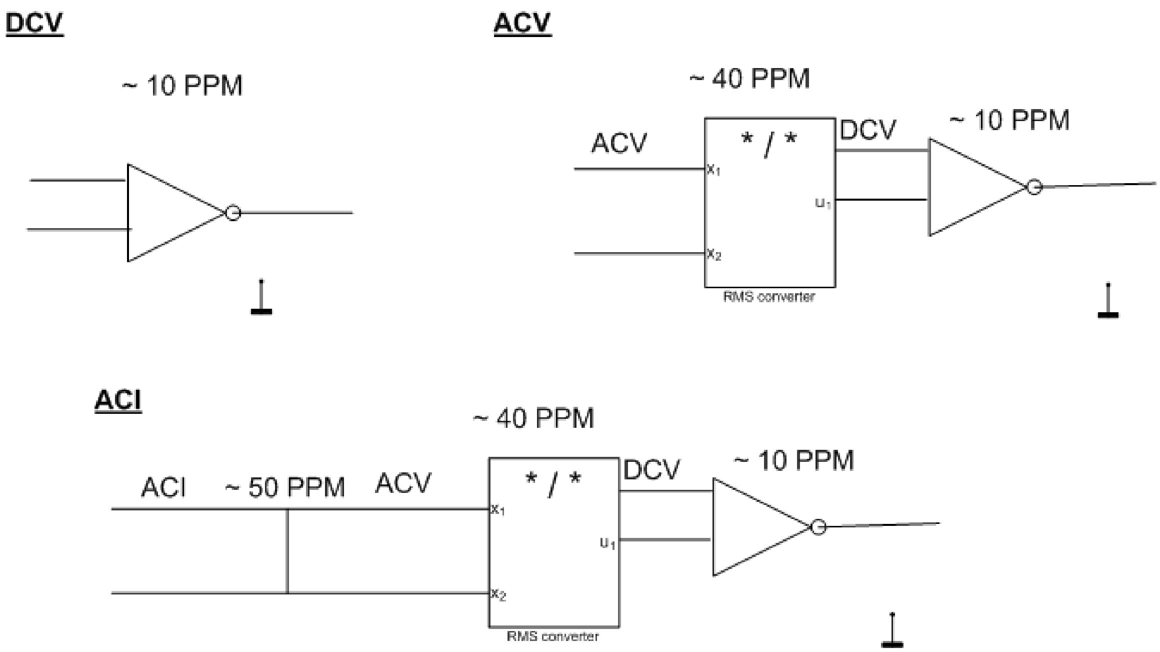

Figure 6: The more components included in order to perform the desired measurement area the more faults arise.

In most cases, several metrologists, each responsible for one or more calibration areas, are involved in the calibration of an instrument. Each metrologist possesses different experiences with different instruments and they evaluate the measurement data in their own way. The layouts of the spreadsheets are different according to which metrologist has developed it, and the fields included are different according to the calibration area. For example the following fields are stored along with the measured result during an AC current calibration: Nominal value, Range, Frequency, Standard deviation, Corrected "calibrator value" (the measurement standard used), Corrected result, Specifications and Uncertainty. For a resistance calibration the same fields as for AC is stored except from the Frequency field, since frequency is not used in resistance calibrations.

For cases where abnormal measurement values are obtained and the metrologist is suspicious that errors have occurred, the following steps are checked:

Procedure if observing abnormal values:

1. Do several measurement sets and see if the measurements still look abnormal.

2. Look at the setup. Find out if something is adjusted or/and connected in a wrong way. If error in the setup, fix it.

3. Use another standard.

4. Check if measurement values look plausible with regards to experience from similar instruments.

If the metrologist has followed the procedure above and still finds abnormal measurement values, he/she will suspect that something is wrong with the instrument. The metrologist then reports to the customer, who is the owner of the specific DUT. The owner then decides the further actions, e.g. adjusting the instrument, confronting the producer of the instrument or accepting the nonconforming values.

The triggering factors for the metrologist to suspect abnormal measurement values are that the values lie outside the specifications, that the standard deviation is out of range or that they deviate from the historical trend. The specifications are obtained from the producers of the instrument, and normal measurement values should lie inside the specification limits with good margins. The formula for the total measurement uncertainty can be written as:

$$U_{tot} = \sqrt{U_a{}^2 + U_b{}^2} \qquad (5.2)$$

Where,

U_a is the evaluation of the standard uncertainty based on statistical analysis described in chapter 3.5.1,
U_b is the evaluation of standard uncertainty based on other factors than statistical analysis described in chapter 3.5.2 and
U_{tot} is the total standard uncertainty.

The total uncertainty (U_{tot}) is the uncertainty that is given in the calibration certificate of the DUT. Usually the U_a is a lot smaller than U_b, resulting in no essential increase of the U_{tot}. Ua consists of estimation of the standard deviation. In cases where the standard deviation has not affected the U_{tot} the metrologist concludes that it has not increased enough to give abnormal values. On the other hand if the U_a has affect on the U_{tot} a new total uncertainty must be estimated and used in the certificate. In cases where the U_a is bigger than 20% of U_{tot} the metrologist will react and evaluate the measurement results to be abnormal. The metrologist can then decide to do a new calibration of the DUT to see if the new estimated standard deviation lies within the acceptable level.

Even if the standard deviation has not increased the measurement value can be abnormal and errors can occur. By looking at historical data from the DUT, the metrologist can see if there are considerable deviation between the measured values and the trend of the historical data. If such a deviation exists something can be wrong with the calibration setup or the instrument. The metrologist will then follow the procedure described above. If the measurement values still differ considerably from the historical data, something may have happened to the instrument. For example it can have been dropped to the floor or a component inside the instrument is not working properly.

There are different ways of using historical data when evaluating data from a calibration process. Some metrologists only use the measurement values stored from the previous year. They may use values from several earlier years if the variance between the measured values and the values from the previous year is too high (see yellow alarm in chapter 8.4.1). Other metrologists consistently use all historical calibration values from the DUT. The evaluation process today depends on the metrologist`s earlier experience with the DUT or similar instruments. If the metrologist knows the behaviour of the instrument or other similar instruments, this knowledge can be used in the evaluation process. If a measurement value looks nonconforming, the metrologist will compare it with the information he/she holds of the behaviour of the DUT or of other similar instruments. Also the information the metrologist has about the customers plays a vital role. For example if he/she knows that the customer of the DUT tolerates some uncertainty in the measurements they await reporting until the values lies outside the customer's tolerance limits.

6 Scenarios

In this chapter we demonstrate the two different ways to perform calibration. The method the majority use today, is manual evaluation of the measurement values performed in an NL. The other method is Internet-enabled calibration, currently established by Sand. These methods are described in two scenarios below.

6.1 Scenario 1: An Internet-enabled calibration

A remote calibration of electrical instruments via Internet is illustrated in figure 7. The DUT is connected to a computer with Internet access and remains in its operating environment at the customer site during calibration. The operator sits at an NL computer and performs the calibration via the Internet. He/she needs a method to verify the measurement data sent from the instrument. An automatic evaluation method must be able to obtain the stored historical data from the database and the measurement values from the electrical instrument.

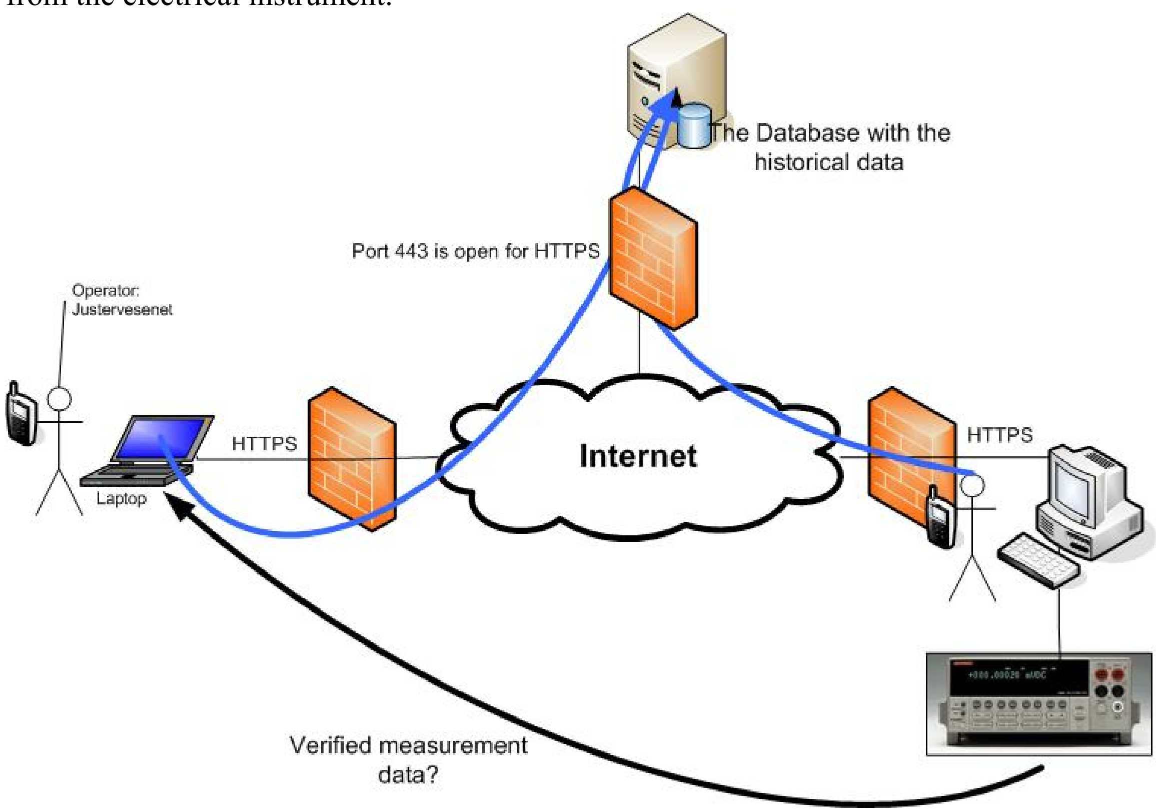

Figure 7: Schematic overview of the system used in remote calibration. The communication between the operator and the instrument runs through a web server on Internet. The operator accesses the historical instrument data from the database, from a web interface. Here the web server is linked to a local database, but it can also be linked to a database on a separate server.

6.2 Scenario 2: A calibration in a laboratory

When a calibration is performed locally the electrical instrument is transported from the customer to an NL. In this scenario a multimeter owned by a customer is sent to JV. In JV`s laboratories the temperature and the humidity lie within the following boundaries:

Temperature: 23°C ± 0,5°C
Humidity: 45%RH ± 5%RH

Where,

°C - Degrees Celsius and
RH - The Relative Humidity expressed as a percentage. RH is the ratio of the actual water vapour pressure to the saturation water vapour pressure at the prevailing temperature.

The setup for a calibration in a laboratory is shown in figure 8. The DUT and the standard are connected to a PC, which collects the measurement values in LabVIEW. Today, metrologists transfer the data stored in text files in LabVIEW, to spreadsheets and evaluate them manually. An automatic evaluation method depends on obtaining the historical data and the measurement data from a database.

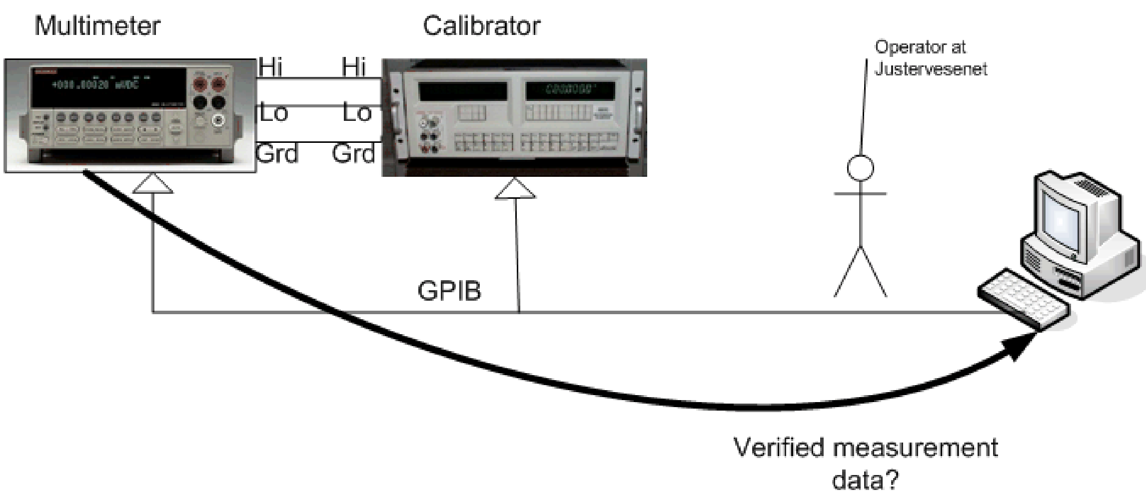

Figure 8: Calibration of a multimeter in a laboratory at JV using a calibrator as the standard. The calibrator and the multimeter are connected to a PC. LabVIEW is installed at the PC and all the measurement data is collected using this program.

7 System requirements

The system is intended to evaluate the calibration values automatically in real-time. To achieve this, a software program is developed. The program must implement the same statistical formulas and take up the same information, as the manual method possesses. This chapter will present the requirements for this automatic evaluation method implementation. The requirements presented here are a foundation for the design and implementation in chapter 8 and the methodology in chapter 9.

The following requirements should be met:

1) A framework and a methodology must be developed before the existing historical data, currently stored in spreadsheets, is stored in the database. This is a time-consuming task.

2) The methodology should be practical and at the same time ensure a satisfactory degree of quality. In order to employ the new evaluation method it must be practical and at least give the same quality as today's method.

3) It is desired that the analysis procedure be executed in parallel with both scenario 1 (see 6.1) and scenario 2 (see 6.2) calibration processes. Then a continuous analysis of the measurement data will be feasible.

4) Measurement data must be stored at a server in a database. The calculations can either be done on the client's computer programs or at the server. The measurements are stored in the DB first and thereafter the verification is stored.

5) A user interface that stores the measured values directly in the database must be developed. Should this be done via a web page? Or should it be done using an application that runs at the operator? The current system for Internet-enabled calibration stores data in the DB automatically, while for a calibration in a laboratory no solution for direct storing exists yet.

6) Statistical methods that can take part in an evaluation of the measurement data must be implemented. This includes estimations of CI and best-fit curve. A 95% CI should be used because this is the most common regarding calibration evaluation. Possibly, if the uncertainty exceeds the given borders, restricted intervals can be calculated. The analysis should give a best-fit regression of the measurement data. Both linear as well as non-linear effects, which include higher order approximations, must be considered.

7) The specifications of the electrical instruments should be implemented in order to analyze if the measurement values meets the specifications. A vital part of the evaluation of the measurement values is to compare them to the specifications.

8) The measurement values, the specifications and the results from the statistical methods and the verification result should be displayed graphically to the operator.

Metrologists today mainly base their evaluation of the measurement data using data graphs plotting the measured values against the specifications and the historical values. The automatic evaluation program should generate such graphs automatically. The program must also handle all other relevant information regarding measurement verification. The final evaluation result of the measurement data should be displayed in a graph.

The realization of the analysis method depends on getting all the measurement values needed in the calculations from the database. Therefore it is crucial that all measurement values, both historical and current calibration values are stored in the database.

The automatic evaluation method is intended employed both in laboratory calibrations and in Internet-enabled calibrations. Sand's Internet-enabled calibration system [6] deploys a graphical user interface (GUI) application. Via this GUI the metrologist can store the measurement data directly into a DB. For scenario 2 in 6.1 no solution for storing the measurement data into a DB exists. It is complicated to find a solution in this scenario because there is no joint way to store the data from LabVIEW. The metrologists today have different ways to store the data from text files in LabVIEW to spreadsheets.

It is important that the program has access to the historical data. They can be fetched into memory as the calibration starts. The analysis can then run continuously in parallel with the calibration. A picture of a suitable GUI application is shown in chapter 8.3. This application is feasible to directly implement in Sand's IMET system, but it is not so easy for scenario 2. Sand's system is designed and developed in the same programming environment as the automatic evaluation method. Laboratory-based calibrations are usually performed using LabVIEW.

8 Design and implementation

In this chapter we will discuss the design and implementation of the automatic evaluation procedure. Sand is currently setting up a database environment in order to realize Internet-enabled calibration. My analysis method will represent an extra function to this system. The database development must fulfil the calibration requirements and the database tables must include all necessary variables in order to perform a calibration. Since my automatic evaluation procedure is going to be a part of Sand's calibration system, it is convenient to use the same database and develop the database design in cooperation with him.

This chapter is structured as follows. Chapter 8.1 gives an overview of the DB layout. In chapter 8.2 we describe how the functionalities of our program are implemented. Chapter 8.3 shows the GUI application viewing the evaluation results. When employing the application for laboratory-based calibrations some challenges arise which need considerations. We will discuss different solutions to meet these difficulties. Chapter 8.4 describes how we have implemented the statistical functions in our program. Finally, in chapter 8.5 we go through a test of the program.

8.1 Database design

A database is designed to offer an organized manner of operation for storing, managing and retrieving information. It uses tables and table relationships to realize this. The database tables consist of rows and columns that define an entity. Each row represents a set of related data, and every row in the table has the same structure. Each column contains a specific attribute type. For example one column might contain text representing a device's name while another might contain an integer representing number of devices. Table 3 shows how two columns and three rows are represented in a database table.

	Column 1	Column2
Row 1	Device name 1	2
Row 2	Device name 2	4
Row 3	Device name 3	3

Table 4: An example on how a database table is divided into columns and rows.

Some columns are primary keys and they uniquely identify the record in the table. A primary key (PK) can be made up of one or more columns. A two column PK might consist of e.g. a "FirstName" and a "LastName" column. A foreign key (FK) is a column or combinations of columns that is used to establish a link between the data in two tables, because there is a logical relationship between them.

When planning the database design, we first thought of the fundamental purpose of our database: "To store all information obtained from a calibration process". Since the metrologists will be the users of the database we interviewed them to learn what information a calibration process involves. We considered this information and identified the broad categories they fell into. Accurately identifying these categories was critical to get a scalable and consistent database. After defining all the categories we considered the relationship between them. Fields in a table may be referred to from other tables. Using unique keys is a good way to reduce queries, because they identify single-rows.

We have developed a database diagram with fifteen tables, their fields and relationships, which is shown in figure 9. The automatic evaluation procedure makes use of all these tables. The tables are presented with the columns needed to properly store the information received from a calibration.

List of the main database tables:

- The *Measurement* table stores among others the nominal value (the value we wish to measure e.g. 10V), the real value (the real value the standard generates e.g. 10.002V) and the uncertainty in each measurement point. It also contains information about the measured value, the metrologist, the manufacturer, the type and the serial number of the reference instrument. For this table our evaluation method obtains all necessary information about the measurement that is vital in order to do measurement verifications.

- The *Reading* table stores the readings viewed in the DUT`s display. Our program gets these measurement values and uses them in the verification process. For example they are evaluated against the historical values and the specifications.

- The *Operator* table holds information about the metrologist that is responsible for the calibration. This table is used for logon in Sand's program. The metrologist enters a valid username and password (in addition to a valid server signed X.509 certificate [30]). Then he/she is authenticated by the system and is permitted to use it. By implementing my program in Sand's the measurement values will be evaluated and the metrologist will be warned if one of the alarms is set (refer to chapter 8.4.1).

- The *Quantity* table includes the measurement units; Ohm, V and A, and their names; Resistance, Current and Voltage. The operator chooses a unit and the nominal value he/she wants to evaluate on through the GUI. The analysis method then gets the measurement values stored on the chosen unit and value.

- The *Calibration* table keeps track of the number of calibrations. It gives information of manufacturer, type and serial number of the DUT. Our method fetches the calibration number based on the choices the operator makes in the GUI. The number is used to obtain the wanted measurement values from the DB.

- The *Company* table stores the name and the address of the company owners of the DUT`s. The method does not directly use this table at the moment, but future Out-Of-Tolerance evaluation results could be sent to the customer using this information.

- The *Device* table holds information about all devices used in a calibration. The verification process fetches this information to know which instrument the measurement values belong to.

- The *AccuracySpecification* table stores the instrument specifications given by the manufacturer. In an evaluation process a vital part is to compare the specifications against the measurement values. To verify the values they must not exceed the specifications. Our method gets the specifications and uses them to evaluate if the measurement values is OOT or not.

- The *Model* table stores what type of model the instruments in the system are. It also holds information of the manufacturer and category of the instruments. The verification process fetches this information from the DB based on which model chosen in the GUI.

- The *ModelCategory* involves the different categories an instrument can belong to. The verification process gets this information via the *Model* table. The *Model* table and the *Manufacturer* table are related. This way the program can know what category the instrument belongs to. One instrument can be of the model category digital multimeter and another instrument can belong to the category calibrator.

- The *Manufacturer* table holds information of the manufacturers of the different electrical instruments being calibrated. Our program fetches this information from the DB to fill the dropdown boxes in the GUI. This table is also related to the *Model* table.

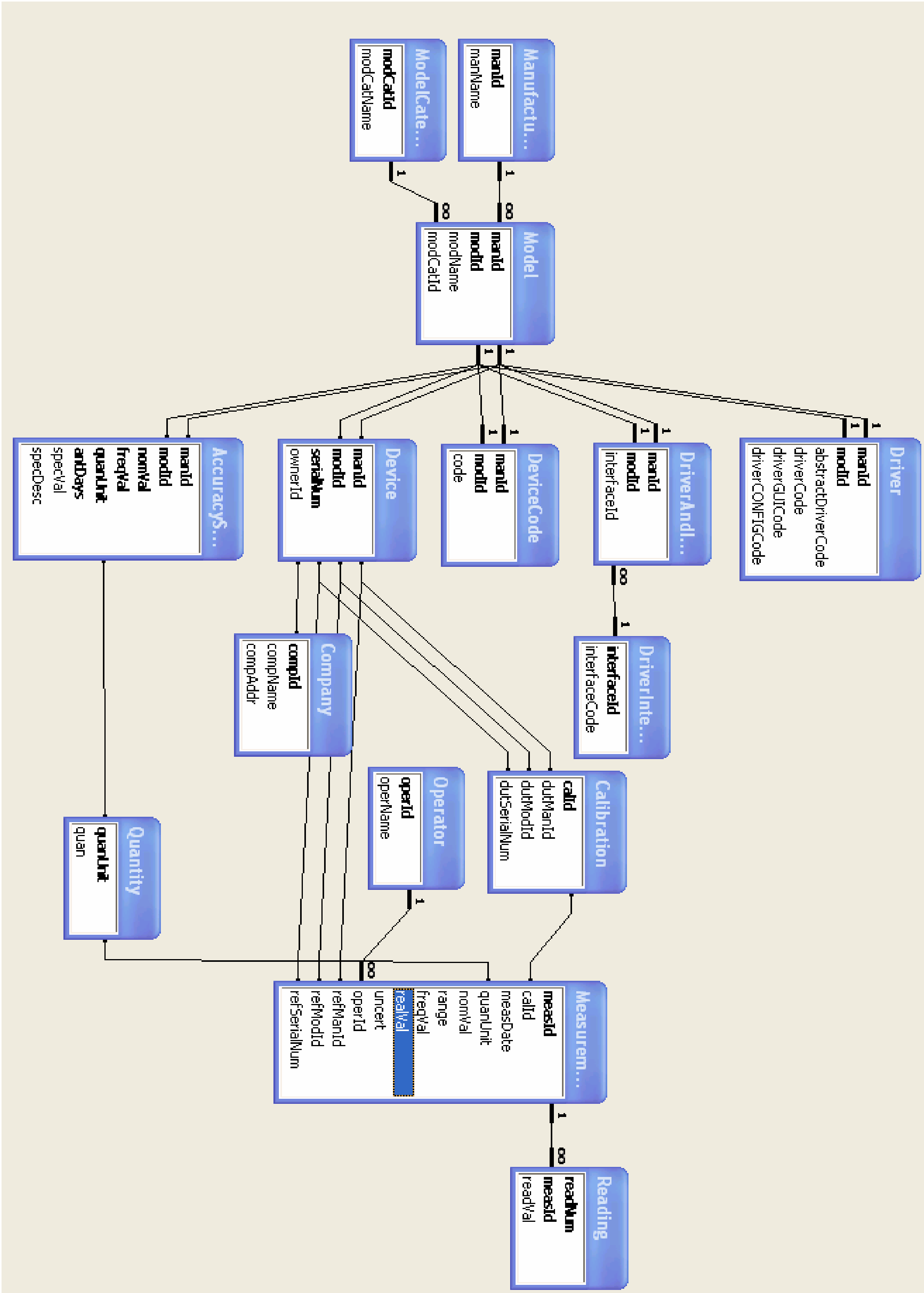

Figure 9: The database and its connections. The bold fields represent the PK of the tables. The line from a table to another represents the relationship between the tables.

8.2 Implementation

Since the database must be consistent with Sand's system we decided to use the same DB. The DB will be used by Sand to store all calibration values obtained using his Internet-enabled calibration system, and by my automatic evaluation procedure to obtain all data needed for the data evaluation.

To meet the first requirement given in chapter 7, which says that the method shall be running in parallel with Sand's method, the analysis program should be integrated into his system. A real-time evaluation of the measurement values is then possible. This is feasible in Scenario 1, by including my program in Sands work. To provide the same functionality in Scenario 2 is more difficult. A program to insert the raw measurement data from LabVIEW to the DB sequentially does not exist today.

At this point, we have only stored historical data for three test instruments. Much time and work would be required to store all existing historical data from spreadsheets to the DB. A script may be developed that stores existing spreadsheet data for an operator into the DB.

The mathematical formulas given in chapter 4 are implemented in software. We have not looked at the physics behind the historical trend of the instruments, and our evaluations purely consist of statistical considerations based on the measurement points. If the trend curves for an instrument can be explained using physical theory, this should be considered alongside the evaluation of historical data. For now, the evaluation is limited to a statistical approach, but this could be an extension to the system in the future.

We have used the programming language C# for the implementation of the mathematical functions [19]. This is a relatively new object-oriented language for software development on both Windows-based machines and UNIX/LINUX machines[1], and it is integrated with the .NET framework [24]. It has had a big influence in the programming community. C# is heavily based on C++ and includes several aspects of other programming languages like Java, Visual C++ and Delphi. It is in common use in the development of Web services. The definition of a Web service is: "A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically Web Service Definition Language (WSDL)). Other systems interact with the Web service in a manner prescribed by its description using Simple Object Access Protocol (SOAP) messages, typically conveyed using HTTP (Hypertext Transfer Protocol) with an XML serialization in conjunction with other Web-related standards" [20]. A web service allows a site to expose programmatic functionality via the Internet. Web services can accept messages and optionally return replies to those messages. In our work we use web services to communicate with the DB in order to fetch the historical data. There is no particular reason to use C# as the Web language over other programming languages, and Visual Basic and Java could just as easily be used. C# was used because Sand's system was developed for .NET.

[1] Using a ported version of the .NET Framework

8.3 User interface

My work has been to make a software program that automatically evaluates the measurement values. The calibration process consists of two main parts. One is to collect the measurement data and the other is to evaluate them. Sand's system [6] includes an approach for using a graphical user interface (GUI) application to insert the calibration values into a DB. In laboratory calibrations the measurement values are stored in spreadsheets and currently there is no procedure for storing these values in a DB. We have in this thesis focused on the development of an evaluation procedure that uses data already stored in a DB.

The concept of a graphical user interface (GUI), invented by researchers at the Stanford Research Institute, led by Doug Engelbart [17], is an important supplement to application programming. Its goal is to increase the usability of the underlying logical design of an application, e.g. our *DataAnalysis* program. We have developed a first edition of the GUI application, viewed in figure 10, which can be added directly to Sand's Internet-enabled system.

In scenario 2 (see chapter 6.2) it is more difficult developing a GUI application that stores the measurement values. In the existing evaluation method the measurement values are stored in regular text files using LabVIEW. The employees fetch these raw values and import them into spreadsheets in different ways. A practical way to fetch these values from LabVIEW and store them in the DB must be found. The evaluation procedure must run during the calibration process and evaluate the measured values continuously. The following are three possible solutions:

1. A parallel-running, stand-alone GUI- application that reads the text file generated by LabVIEW.

2. A DLL-component, containing the evaluation methods, that is integrated in all the LabVIEW programs.

3. A Web service that is called from all the LabVIEW programs.

The first solution consists of development of a GUI application that reads the generated text file. The application uses these data in the evaluation process. Since the text files are different for each operator, we also have to analyze the shape of the files for every operator. One benefit with GUIs is that the developer does not have to worry about the interface design. GUI design standards, in most environments, introduce conventions that make the applications look much the same on the same platform. The user gets access to the system functionality via menu bars, buttons and keyboards shortcuts. Also additional objects such as radio buttons, scrolling lists, check boxes and other graphics may be displayed or directly manipulated.

The second solution deals with an integrated DLL-component. A DLL (Dynamic Link Library) is a library of functions and procedures that can be called from an application or another DLL. The major advantage using a library in this way is that it permits the sharing of code. Many applications and DLLs can use the same DLL and multiple processes can share the code in a single DLL. Also component-based and modular development is allowed, resulting in simplified development and upgrade process. A separate DLL file is dynamically loaded, either when the application loads or when its member functions are needed. A DLL file can be called from many development environments and languages that ensure interoperability. In our case the DDL file will contain all the evaluation functions necessary when performing calibrations. These functions can for example be called from a custom made LabVIEW program (see chapter 5.1.1).

The last solution is to develop a web service (refer to chapter 8.2). The purpose of a web service is to give some functionality on behalf of its owner. This owner namely the *provider entity* is a person or organization. It provides an appropriate agent to implement a particular service. Then a *requester entity*, also being a person or organization, wishes to make use of the *provider entity's* web service. The user interface could thus utilize a web service to run all the evaluation procedures remotely. The local user interface could for example use the web service to check historical data to generate graphs. We will be interested in a web service that is the same for all the three above alternatives. In alternative 1) the GUI application will call it, in alternative 2) the DLL library will call it while in alternative 3) the LabVIEW program will call it directly.

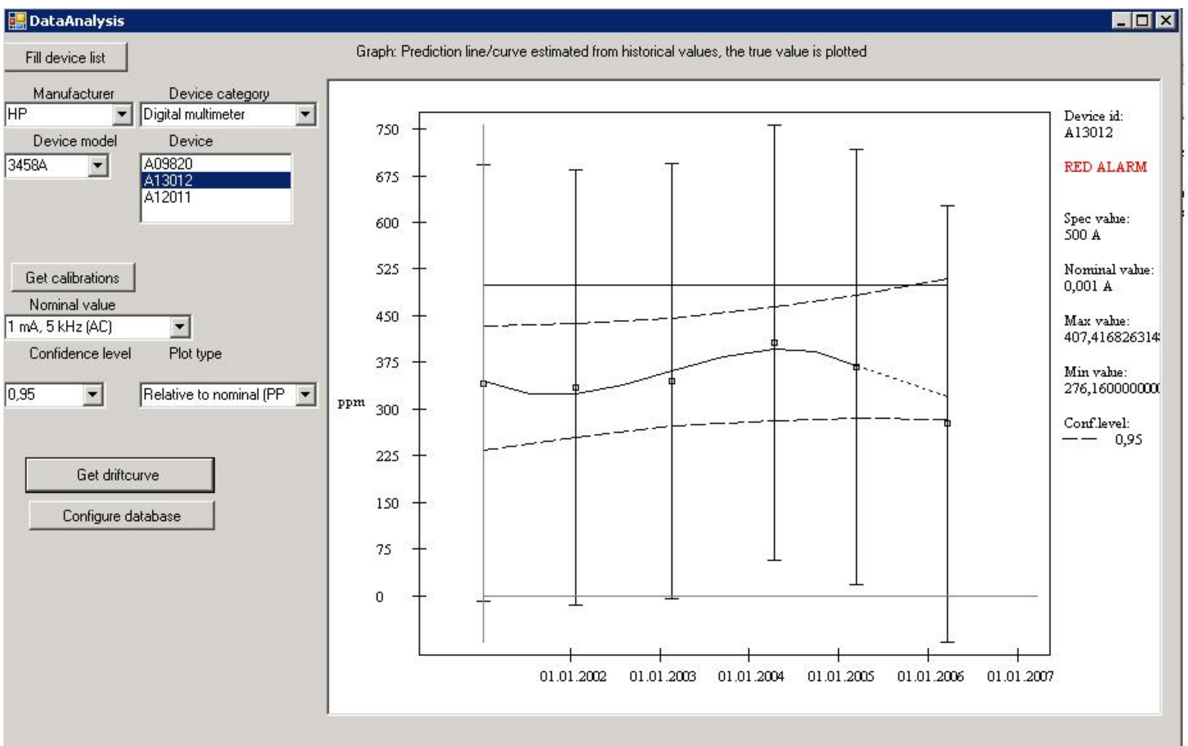

Figure 10: A picture of the graphical user interface (GUI) application for the automatic evaluation method. Here the six measurement values and their uncertainty are plotted. The predicted value is seen where the dotted regression line ends. Also the two dotted lines giving the confidence interval, the straight line indicating the specifications and a red alarm is plotted. The red alarm is here viewed because the uncertainty in the points exceeds the specifications. This is a temporary interface that probably will change.

8.4 Evaluation engine (statistics)

The goal of the automatic evaluation method is that it shall give equal or better evaluation functionality compared to the existing manual method. This means that the method must support the same statistical functions as the manual evaluation method. We have implemented these statistics into our *DataAnalysis* program. When evaluating measurement values it is vital to estimate the uncertainty for each value. Each measurement value is associated with an uncertainty due to the limits of instruments and the people using them (see chapter 3.5). Each measurement point with the corresponding type A uncertainty and type B uncertainty is stored into the DB by the metrologist during a calibration. The analysis program accesses these uncertainties and uses the method *RegressionPlot* to plot them along with the belonging measurement points in a graph, given in figure 10.

To evaluate a measurement value against historical measurement values we need to estimate the prediction interval (explained in chapter 4.2). The following formula is implemented from Koeman`s article [28], in our *CalcPredictionBandValue* method:

$$\text{Prediction Bands}:$$

$$v_{predict} = T * s * \sqrt{\frac{1 + \frac{1}{n} + \{(t - t_{avg})^2\}}{\sum_{i=1}^{n}\{(t_i - t_{avg})^2\}}} \qquad (8.1)$$

Where,

$v_{predict}$ is the predicted uncertainty

T is the Student-t number for a selected level of confidence with n-2 degrees of freedom;

$$s = \sqrt{\frac{\{\sum_{i=1}^{n}(v_i - v_0 - d * t_i)^2\}}{n-2}},$$

n is the total number of data points,

t is the time where the predicted uncertainty is to be computed,

t_{avg} is the average time of data points: $t_{avg} = \frac{1}{n} * \sum_{i-1}^{n} t_i$.

The *CalcPredictionBandValue* method calls the method *GetTValue.* This method includes input parameters for the degree of freedom (df) and the confidence level (CI). Since we have implemented the table for t-distribution in this method (shown in table 4 in appendix A), the method finds the t-value consulting this table. All the t-values in the table are one-sided. To find the one-sided confidence level α_1 that corresponds to the two-sided confidence level α_2, the following formula must be used (see figure 2 in chapter 4.3 and figure 13 in appendix A):

$$\alpha_1 = (1 + \alpha_2)/2 \qquad (8.2)$$

Inserting $\alpha_2 = 95\%$ results in a 97,5% confidence level in equation 8.2. E.g. when the degrees of freedom equals 3 the t-value will be 3.182 (see table 4 in appendix A).

The manufacturer of the DUT provides its specifications. There are different ways of estimating the specification depending on which area is being calibrated. The *AccuracySpecification* table in the DB stores the specification values and the number of days after a calibration they apply. Our analysis method retrieves these values from the DB and the *RegressionPlot* method plots them in a graph. In figure 9 the specification is plotted as a straight line. This happens when the graph is plotted with the nominal value against time. When plotting against the frequency the specification appears as a nonlinear curve.

The mathematical function for a linear regression line is implemented by making the method *FitLinearCurve*. This method holds arguments for the number of days, the measurement values, the number of measurement data at hand, the slope of the linear line, *a*, and the intercept, *b*, after the formula: y = ax + b (see equation 4.4).

Employing these arguments the defined Sx, Sx^2, Sy and Sxy (see appendix A) is calculated using the following code:

```
// tempX and tempY are of the same length.
double Sx = 0;
double Sx2 = 0;
double Sy = 0;
double Sxy = 0;
for ( int i = 0; i < tempX.Length; i++ )
{
      Sx += tempX[i];
      Sx2 += Math.Pow( tempX[i], 2 );
      Sy += tempY[i];
      Sxy += tempX[i]*tempY[i];
}
```

In the code above tempX is the parameter of the number of days while tempY is the parameter of the measurement value.

The constants, a and b, are then estimated using the following code:

$$a = \frac{(n * S_{xy}) - (S_y * S_x)}{(n * S_x{}^2) - Math.Pow(S_x, 2)}$$

$$b = \frac{(S_y * S_x{}^2) - (S_{xy} * S_x)}{(n * S_x{}^2) - Math.Pow(S_x, 2)}$$

Where,

Math.Pow raises the value Sx to the power of 2.

To implement a second order curve we use the method *FitSecondOrderCurve* and specifies the same parameters as in the *FitLinearCurve* adding a parameter c in order to adopt the formula $f(x) = ax^2 + bx + c$ given in appendix A. We create four matrixes in order to calculate the three parameters a, b and c. The example below shows the C# code that calculates the parameter a given the matrix m:

```
Matrix m = new Matrix( new double[,]{
                        {Sx4, Sx3, Sx2},
                        {Sx3, Sx2, Sx },
                        {Sx2, Sx , n  },
} );

Matrix ma = new Matrix( new double[,]{
                        {Sx2y, Sx3, Sx2},
                        {Sxy, Sx2, Sx },
                        {Sy, Sx , n  },
} );
double a = ma.Det() / m.Det();
```

Where Det() method returns the determinant of the matrix. A determinant is a square array of elements with following prescribed evaluation rule [25]:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} \cdots & a_{3n} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{n1} & a_{n2} & a_{n3} \cdots & a_{nn} \end{bmatrix} = \det(A) = \sum_{\sigma \in S_n} sng(\sigma) \prod_{i=1}^{n} A_{i,\sigma(i)}$$

The implementation of the third order curve is contained in the method *FitThirdOrderCurve*. It takes the same parameters as the *FitSecondOrderCurve*. A fourth parameter d is also added according to the formula $f(x) = ax^3 + bx^2 + cx + d$. The C# code is the same as for the *FitSecondOrderCurve* only that the code for the additional parameter is added, including one more matrix.

The *RegressionPlot* method includes code that estimates the error which each of the above regression functions gives. The one that gives the least error is the one that lies closest to the points, and is the one that is plotted in the graph. This line or curve is estimated based on the historical values stored in the DB. After a Bayesian approach we predict the next measurement point. This predicted value is estimated using the formula of the best-fit regression line or curve. We set the x-value to the number of days from the first calibration to the date the next calibration will take place. The output from the formula will be the predicted measurement value.

8.4.1 Alarm/Warning

The essential purpose of our analysis method is to give a report whether the measurement data is acceptable or not. There are discussions around which kind of drift model to use and which alarm tests that should be implemented. When shall an alarm be activated? How shall the uncertainties be used in the evaluation process? Should for instance the measurement uncertainty be combined with the prediction uncertainty? How much must the measurement value differ from the historical values before the alarm is activated, etc?

Our research has resulted in three different types of alarms with different levels. Each alarm has its own colour according to the error's significance level. The alarms are listed with a descending level of importance.

1) **Red alarm: The measured value is outside the specifications.**
As long as the measurement value included the measurement uncertainty, lies within the specification everything is fine and no alarm is set. If the measurement uncertainty overlap the specification, we can estimate a significance value for how certain we are that the value lies within the specification borders. For instance: "We are 65% certain that the value lies within the specifications".

2) **Orange alarm: The standard deviation has increased too much.**
Orange alarm is activated according to evaluations of the influence Ua constitutes in the uncertainty formula: $U_{tot} = \sqrt{U_a{}^2 + U_b{}^2}$ (see equation 5.2). Usually Ua is so little that is has no influence on Utot. If the standard deviation is increasing so that Ua has an effect on Utot the uncertainty budget must be changed.

3) **Yellow alarm: The measured value corresponds badly with the historical data.**
There are discussions when alarms and warnings shall be activated in the system according to verified data or not. Helge Karlson, a statistician at the JV, advised us to use a test for normalized error, E_n , to compare measurement results. Calibration laboratories use this E_n-value to compare results. In our method we wanted to activate an alarm if our predicted value differed too much compared to the measured value, at the date of measurement. This is our formula for E_n:

$$E_n = \frac{(X_{measured} - X_{pred})}{\sqrt{(U_{measured}{}^2 + U_{pred}{}^2)}} \quad , \qquad (8.3)$$

Where,

$X_{measured}$ – measured value (measurand),
X_{pred} – predicted value (the value on the regression line),
$U_{measured}$ – the measurement uncertainty (type A and B) in the measured value,
U_{pred} – predicted uncertainty (i.e. the range between the regression line and the prediction line).

If E_n is less than one the predicted value is consistent with the measured value and no alert is given. But if E_n is bigger than or equal to one there is a discrepancy between the results and the prediction, and an alert signal must warn the operator. Then the measurement value including the measurement uncertainty lies outside the prediction interval.

We have developed a flow chart in figure 11 that shows the alarm implementation. We have implemented the software code that gives a red alarm when the measurement value exceeds the specifications. This method is a framework that can be developed further. In the future one may want to change the terms for “red alarm”. The remaining alarms are not yet implemented.

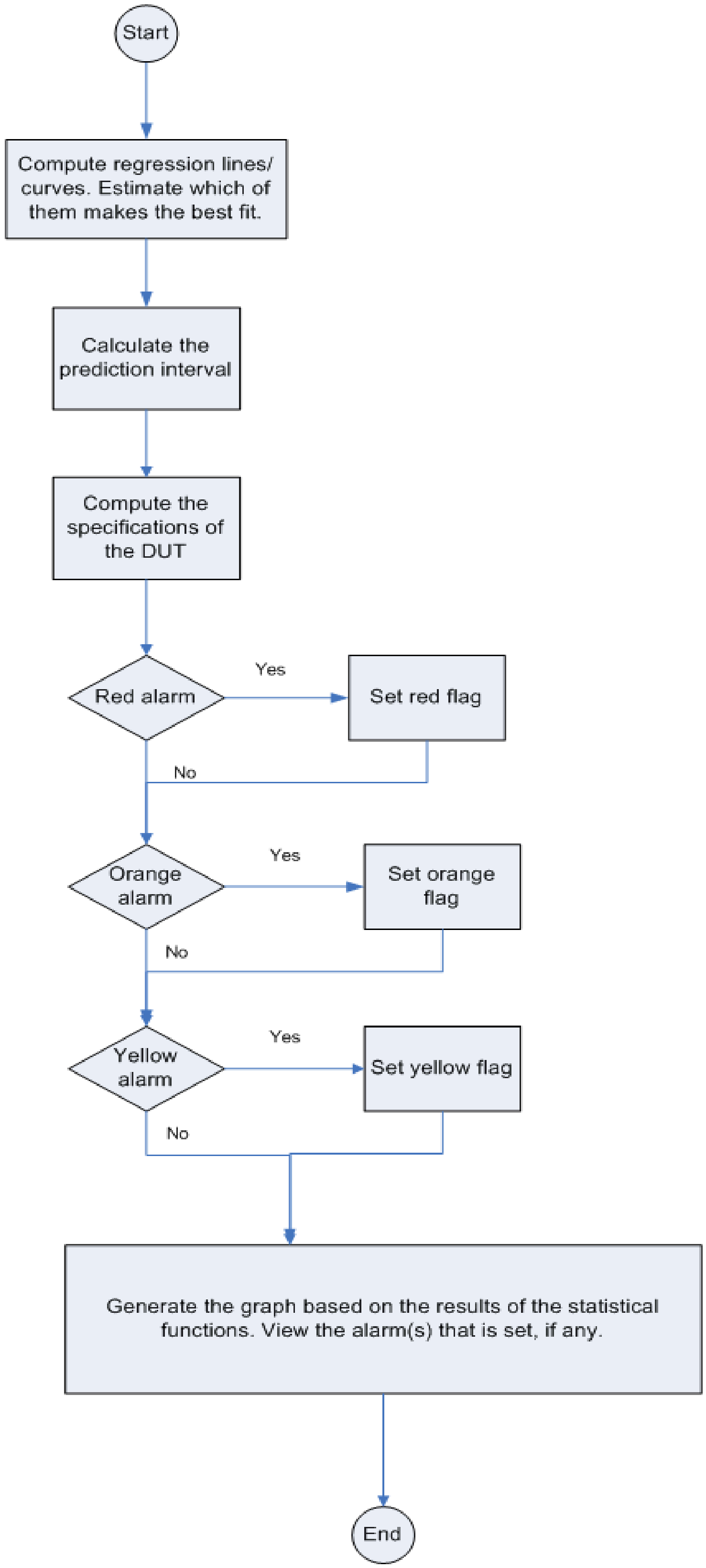

Figure 11: A flow chart for the three different alarm types.

8.5 Testing

We have performed tests of our evaluation method. Three digital multimeters of the same type, viewed in figure 12, were used as test instruments. For these multimeters we have manually inserted measurement values from the last six years into the DB. To simulate a real world calibration, we used the five first measurement values as the historical data. Then we used the newest value as the measured value gotten from the current calibration.

Figure 12: High-precision digital multimeter, the device-under-test.

The instruments' serial numbers are unique and consist of a series of numbers. For our three DUT`s the first five numbers refers to the instrument type. The remaining numbers constitute the unique numbers that gives the identity of each DUT.

We used the GUI in figure 9 and chose one or more of the three test instruments and different calibration selections from the dropdown boxes. The results of our choices were plotted and viewed in our plot window. The following is an example of one of our tests and the following steps using the GUI:

1. First we pressed the *Fill device list* button. The system then listed all manufacturers and devices stored in the DB.

2. Secondly we chose desired manufacturer from the *Manufacturer* dropdown box. We chose HP, and the device category, device model and device list boxes were then filled with the stored data of HP devices.

3. Thirdly we chose device category. Here we had only stored one category, namely digital multimeters, so we didn't have to make a choice.

4. We then chose device model. We had only stored one type of electrical instrument so we didn't have to make a choice here either.

5. Thereafter we chose serial number of the DUT. We chose one of the three serial numbers we had stored. (It is also possible to choose several serial numbers. Then the drift curves of the different instruments will appear in different colours.)

6. We pushed the *Get calibrations* button. Then all nominal values were listed. We wanted to evaluate a calibration in the ac current area and chose the desired nominal value to be: 1mA, 5kHz.

7. We chose confidence level to be 0,95 representing a two-sided 95% confidence level.

8. We then had to choose between two different plot types, either normal or relative to nominal in parts per million (PPM). We chose the plot type to be relative to the nominal value in PPM.

9. We then pushed the *Get driftcurve* button.

We then got a picture of the results of our selections shown in the plot window in figure 10. The five first points refers to the five historical values we stored for the DUT, its specifications on the area, the measured value (here we used the measurement value for the last year as the currently measured value), the best-fit curve estimated based on the five historical values, the uncertainties of the measurement points and red alarm. The red alarm is flagged because the uncertainties of the points (measurement values) lie outside the specifications (see chapter 9).

9 Methodology

"Methodology refers to the principles, procedures and practices that govern research, whereas research design refers to the plan used to examine the question of interest." [21] [22]. This chapter is intended to encompass the entire process of carrying out the automatic evaluation procedure and give guidelines for the phases this involves. Chapter 9.1 shows the procedure that consists of the different steps that should be done. In chapter 9.2 the practical use of this automatic evaluation method is given.

9.1 Setup

In order to set up an automatic evaluation method the goals and purposes of such a procedure must be established. The practical aspects compared with a manual method must be declared. The objectives for our procedure are given in chapter 1.4. In our case we are going to make the existing, manual evaluation method more efficient. It is vital to know the basis of this method. The following shows a step-by-step procedure for making an automatic evaluation method based on the existing manual method:

1. Collect information about the possibilities for developing an automated software program based on the manual process.

2. Interview the metrologists currently performing the manual evaluation. Information on how they perform manual evaluation should then be obtained. The result of our interviews of the metrologists on JV is given in chapter 5.2. We then got ideas and impressions of how the automatic evaluation method best could be designed in order to be practical and acceptable to the users.

3. Define the requirements of the users, customers and the company applying the method, refer to chapter 7. This is important in order to get the analysis program as applicable as possible, considering user acceptance and implementation.

4. Choose a modelling technique that focuses on the major steps of the process. We have chosen a use case diagram showing a hierarchical structure. A use case diagram will describe the proposed functionality of our system. The hierarchical relationship is modelled using the *Include* and the *precede* dependency. Our use case diagram is drawn in figure 13.

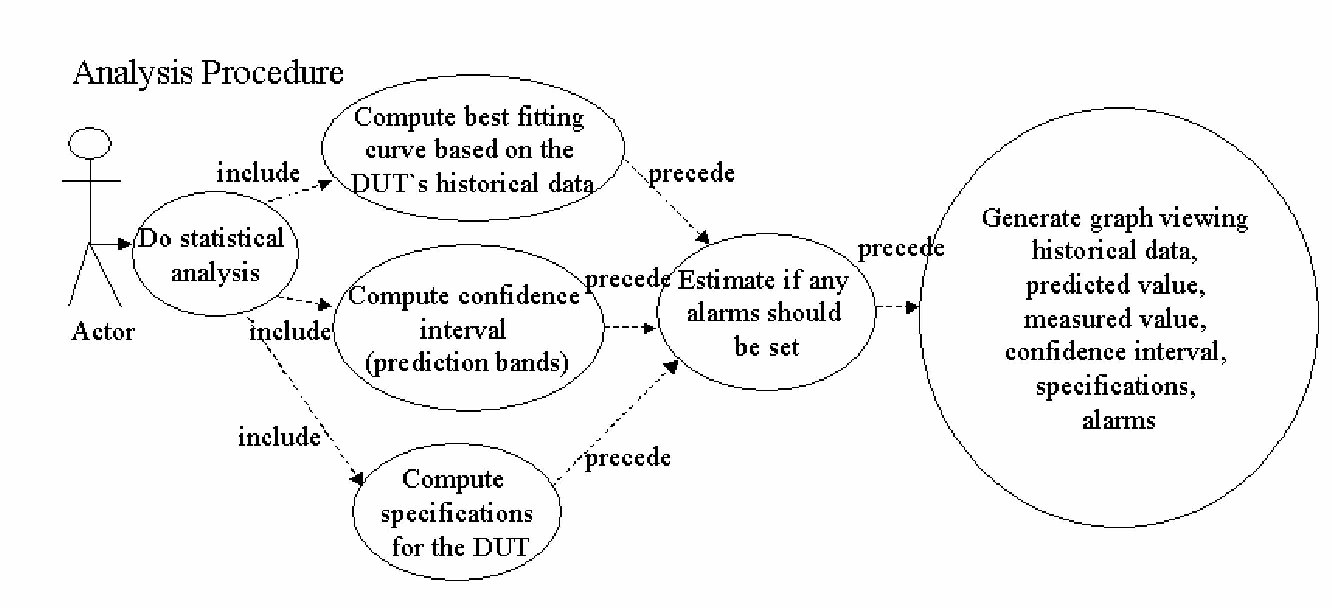

Figure 13: Use case diagram of the automatic evaluation method.

The actor is the user of the system including both human users and other computer systems. In our use case the actor is the analysis program. The *include* arrows between Use Cases indicate that one Use Case holds the functionality of another as part of its normal processing. One or more Use Cases may include a Use Case, so duplication of functionality is reduced. This means that one Use Case can extend the behaviour of another, expressed in our diagram using the *extend* dependency. In order to make an evaluation method the analysis program first must perform statistical analysis, Use Case number 1. This includes the next three Use Cases refer to chapter 4. Now the program has enough information to make an evaluation decision estimating if some of the three alarms should be set, refer to 8.4. Finally the program can generate the graph viewing the evaluation results performing the last Use Case in figure 11.

5. Evaluate different methods and articles of relevance for evaluation decisions. This includes evaluation of prediction bands, best-fit curve, the DUT`s specifications, measurement uncertainties and prediction of next measurement point considering Bayesian approaches.

6. Find out when the result of an evaluation should result in an alarm. Decide if different types of alarms exist. If so, these must be implemented after the Flow chart given in figure 11 (see chapter 8.4.1).

7. Draw a flowchart diagram of the evaluation process (see chapter 8.4.1).

8. Develop a database that consists of all the necessary tables that a calibration requires (see 8.1). Since the analysis program constitutes a part of the iMET system this development must be done in collaboration with the developer of this system. Refer to chapter 8.1.

9. Develop the analysis program and a GUI according to the Use Case diagram and the Flowchart diagram. Here there are many different programming languages to choose from. We wrote the program in C# mainly because the iMET system is written using this programming language. Therefore a Sharp Develop programming environment already was set up [26]. SharpDevelop is a free integrated development environment (IDE) for writing applications in C# or Visual Basic on the .NET platform. C# is a modern object-oriented programming language that are suitable for writing applications for both hosted and embedded systems. Refer to chapter 8.3 and 8.4.

10. Insert all calibration data for the test instruments and test the analysis program. We manually inserted the measurement values for our test instruments. An automation script should be developed to save time.

11. Identify possible improvements and implement these in the program. Performing tests may give ideas for better or new functionalities.

12. Integrate the analysis program into the calibration process. For the Internet-enabled method (IMET) this will be an easy task by performing a method call of our *DataAnalysis* method during the IMET process. Employing the program in a laboratory calibration is more complicated. Figure 14 illustrates how this can be done. Under a calibration the DUT is connected to the LabVIEW program. This program stores the calibration values in a text file. Our evaluation program will get access to the measurement data via a C# GUI application. This application knows the path to the text file and the operator responsible for the calibration. The application can only fetch the right measurement values if they are stored in a specific format.

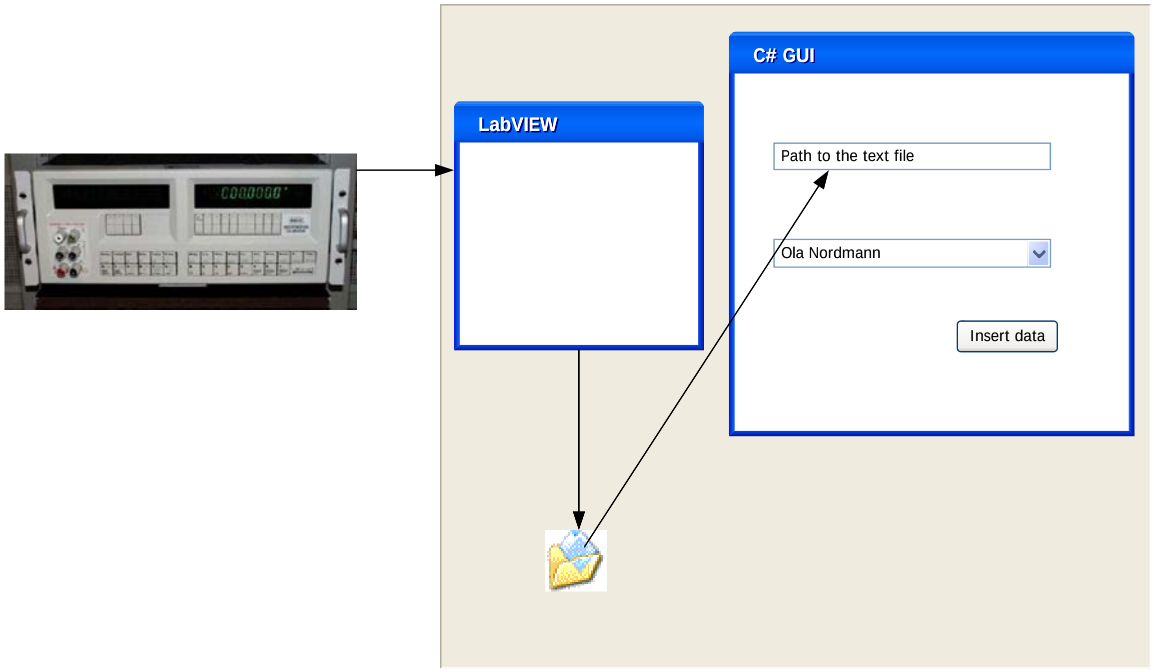

Figure 14: The shown solution will give our evaluation method access to the measurement data. LabVIEW stores the calibration values in a text file. Our program gets access to this file from a C# GUI application which also knows the operator that is performing the calibration. This setup is dependent on that the measurement values are stored in the text file in a specific format.

Other approaches, which give the same functionality, should also be considered. The operators` LabVIEW programs should be set to generate result files in a standardized format, e.g. XML-based. This would result in easier data exchange internally at JV or among different NMIs using the Internet. This would also make it more efficient to store the values into the database (see chapter 8.3).

13. Develop methods for inserting the historical calibration data or perform a manual insertion.

14. Develop methods for continuous storing of measurement values into the DB during a calibration (see chapter 8.3).

The manual evaluation decisions are based upon several statistical methods as well as the metrologists` knowledge about the DUT, other similar instruments and their customers. The statistical methods uses the measurement values obtained from a calibration process. In order to realize an automatic evaluation program it is vital that all measurement values are stored in a DB.

9.2 Performing automatic evaluation

When the automatic evaluation method is established and integrated in the calibration process (see chapter 8.3), the metrologists can use the GUI application shown in figure 10. A step-by-step procedure for the functionality of this application is given in chapter 8.5 which describes our tests of the program. Here we manually used the GUI application to test if it gives the desired evaluation results. During a future automatic evaluation it is desired that this will be an automatic event (see point 12 in the chapter above).

10 Discussion and evaluation

In this chapter we will discusses and evaluate the work presented in this thesis. It starts with a discussion of the results, before we evaluate the requirements and the device authentication question. Thereafter, we evaluate how well the objectives in chapter 1.4 have been answered.

10.1 Discussion

This discussion focuses on chapter 8 and chapter 9. Several issues are discussed, including how the requirements (see chapter 7) are met, and the possibility of using our program to authenticate instruments.

10.1.1 Discussion of results

Chapter 8 describes how we have designed the automatic evaluation procedure, the database development and the implementation of the necessary statistical functions. This has been our main work. We have through interviews of the metrologists and research developed a program that includes statistical functionalities and procedures, useful for automatic evaluation of measurement data. Our tests of the program showed that the evaluation gave reasonable results. The final graph outputs were the best-fit curve for the stored historical data, the specifications, the predicted next value and the measured value. Also the red alarm was flagged as intended (refer to the GUI in chapter 8.5). The program also fetched all historical data for the chosen DUT from the DB in a desired fashion.

The implemented statistical functions worked correctly and it seems that our program can be used for automatic evaluation of calibration values. As mentioned in chapter 4.7, we use a Bayesian line of though for the uncertainty calculation but frequency probability is applied to historical data in order to calculate the predicted next value. The two other alarms should also be implemented, in order to give a total evaluation of the data.

Although the statistical methods implemented in our automatic evaluation program give the desired results, there are considerations that must be made before a complete automatic evaluation will work in practice. Our tests (see chapter 8.5) are only performed on the functionalities of the evaluation program. There are challenges concerning integration of the method into the calibration process and also the framework needs to be considered. The framework must include the integration of the method in both a laboratory and an Internet-enabled calibration, in order to get a continuous evaluation of the measurements.

The focus of the performed tests has been on the application's evaluation functionalities, meaning its ability to evaluate all calibration data for a DUT. We have discussed issues concerning implementing the evaluation procedure into the calibration process, extended functionalities and improvements. The actual work concerning these issues is left for further work.

The results of chapter 8 and chapter 9 are promising. The methodology gives a useful step-by-step framework and our program includes the main evaluation functions necessary for an automatic evaluation application.

10.1.2 Coverage of the general requirements

Table 15 describes how the requirements presented in chapter 7 are covered by the framework and the methodology.

Requirement(s):	Coverage met:	Notes:
1) Methodology and framework.	Covered	We have built a methodology and a framework for development of the method. Add-ons can be considered in each case.
2) The program should be practical and produce the same or better quality than the present method.	Partly	We have tested that all the statistical functions give the same results as the manual method. It is more practical but the quality issue must be investigated further.
3) Integration in the calibration process.	Discussed	We have discussed how to integrate the evaluation procedure in chapter 8.2.
4) Storing data in a DB.	Partly	We have developed a DB with the tables needed to store all measurement values from a calibration. Here we have stored calibration data for three instruments. The verification results should also be stored in this DB.
5) GUI that stores the measurement data into the DB.	Partly and discussed	For Internet-enabled calibrations this will be done automatically. We have discussed and proposed a solution for calibrations done in a laboratory (see chapter 8.4).
6) Statistical functions used for evaluation of measurement data.	Covered, can be extended.	We have implemented all statistical methods that are needed to make evaluation decisions from the measurement data. More statistical methods can be added if desired e.g. higher order curves.
7) The specifications of the DUTs	Covered	We have implemented code in our program and in the DB, that will calculate the DUT`s specifications.
8) Graphical presentation of the results of the statistical methods and the evaluation results.	Covered	A GUI application that shows the outputs of the statistical methods as well as the final evaluation result. This includes the alarms that are set (see figure 11).

Table 5: The table gives an indication to what extent the requirements are met.

The methodology and framework given in this thesis is meant to be a helpful recipe for the development and integration of an automatic evaluation application. We have focused on JV's calibration environment and this is therefore the most suitable range of use. Improvements and supplements should be considered, particularly when employing the methodology and framework in other environments.

10.1.3 Can further development authenticate instruments that are calibrated via Internet?

We will here discuss whether an extended version of our program can be used to authenticate instruments. The question is whether unique properties of some instruments can be extracted by analyzing several measurement points over time? Do unique curves exist for calculated drift that can be used to authenticate the instrument/equipment?

10.1.3.1 Identification problem

With the introduction of remote calibration via Internet comes the challenge of trustworthy authentication of electrical instruments (see figure 15). The operator may know which type of instrument he/she is operating but he cannot be certain of its identity. Today, an identification string obtained by using the DUT`s command set [27], is used for instrument identification. Some instruments don't have the necessary functions for identification and don't possess an identification string. This string usually contains manufacturer, model number, serial number and firmware revision number. Even though an instrument is equipped with such functionality, it is not necessarily trustworthy. The reason is that the serial number in many cases easily can be overwritten. This functionality has obviously been designed for practical reasons and no thoughts to security have been made during implementation. The need for authenticating equipment has not been an issue up until now.

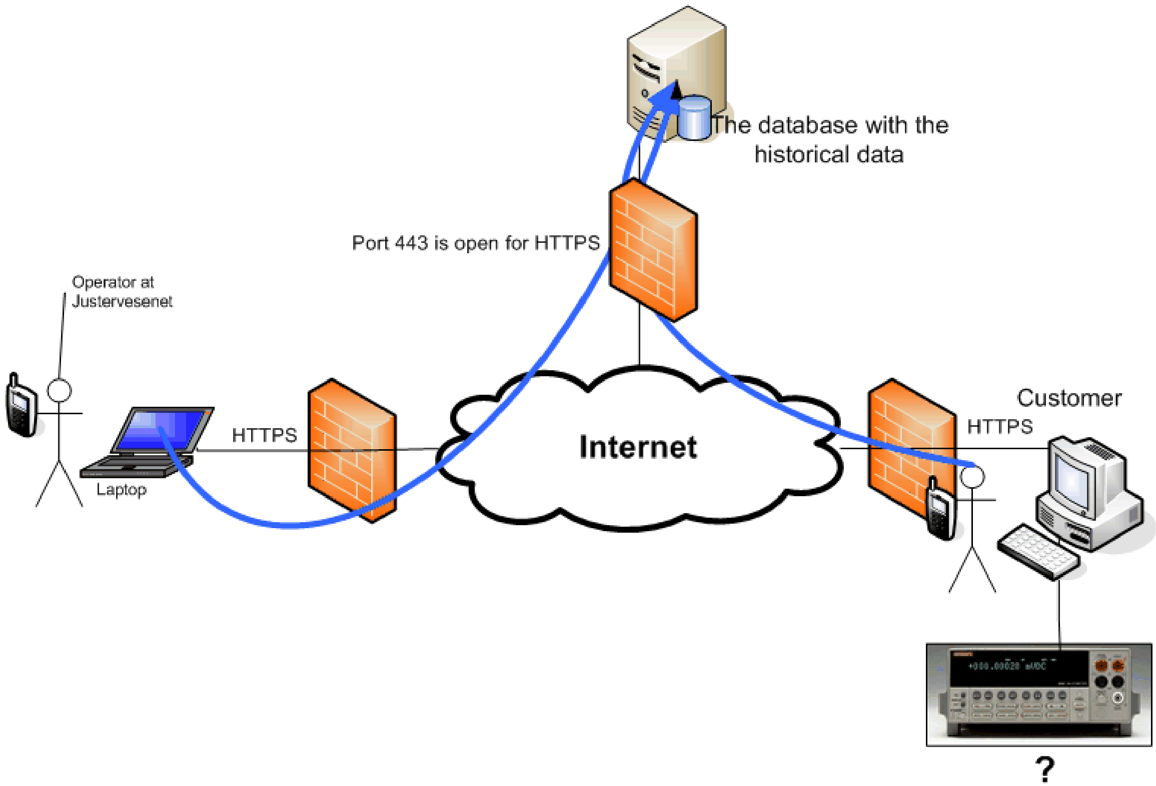

Figure 15: Schematic overview of the approach to the problem with the execution of Internet-enabled calibration. The operator cannot be certain of the identification of the instrument he/she is calibrating on.

10.1.3.2 The authentication issue

Our present program bases the evaluation decisions on the measurement data obtained from calibrations. Can the issue concerning authentication of instruments be solved only based on measurement values obtained from calibrations? If the answer is yes an extended version of our program may be useful.

Our program already contains useful functionalities that can be used in the authentication process. The graph plots the data against time (the time on the x-axis) representing the instruments drift curve. The y-axis can be plotted using different measurement areas within different measurement quantities. Considering the authentication issue it will be desirable to find unique drift curves for one specific instrument, which will yield as this instrument's "fingerprint". A "fingerprint" can consist of several drift curves. Also the implemented prediction of the next measurement point using a Bayesian approach [28] is useful. The present version compares this predicted value with the true value to see how the true value corresponds with the historical data (refer to yellow alarm in chapter 8.4.1). To make instrument authentication possible, our program could use this comparison test to decide if this is the right instrument or not. If the true value corresponds badly with the historical data (yellow alarm is flagged) the program will determine that there is a possibility that the metrologist is operating on another instrument. But how certain can the program be that this is another instrument? There will always be a probability that the true value corresponds badly because something has happened to the instrument, e.g.

dropped on the floor. It is possible that the method could only be used for multimeters and other instruments that involve measurements using several different measurement areas over many different measurement quantities.

10.1.3.3 Required Extensions

Our automatic evaluation program should be extended with the following proposed functionalities in order to be able to autenticate instruments:

1. A test functionality must be developed. All measurment points to a group of instruments of the same type must be compared. A graph that show all the drift curves for these instruments in the same window should be shown. We could then go through every measurment point and see it there were points that had unique instrument drift curves. If many unique drift curves were found for one instrument they together could indicate the instrument`s “fingerprint”. This is partly implemented in our present version since the next measurement point is calculated using a Bayesian approach [28].

2. The instruments “fingerprint” should be stored in the DB.

3. During a calibration of a DUT the authentication program should get the DUT`s “fingerprint” from the DB. In an Internet-enabled approach the customer will choose from a GUI application the device name of the DUT. The program should then find this DUT`s “fingerprint”. A distrusted customer can have overwritten the DUT`s true serial number with a serial number of another instrument, to con the operator.

4. If the program suspects that the DUT appears with false identity it should go through the stored fingerprints list and see if the measurment values (of the DUT) could belong to another instrument.

5. Finally the output of the authentication program should be that the instrument is authenticated with a certain degree. For instance that the instrument is authenticated with a probability of 78%. In cases where it is not authenticated and the program has found another fingerprint that the measured values corresponds better to, the program`s output statement could be: “The instrument cannot be authenticated as e.g. A09820 (the name the customer has key-entered). It is more likely that the instrument is A1312 (the identity of the other instrument where the predicted value of the “fingerprint” driftcurve, and the measured value corresponded with each other).

It is not certain that it is possible to establish uniqe driftcurves for instruments of the same type. It could be a good idea to start building the identification routines using equipment which is impossible to adjust, like resistors. Resistors will drift freely and have a unique driftcurve.

10.1.3.4 Required work

A vital issue that has to be met before the deployment of the authentication functionality proposed above is to find "fingerprints" for instruments used in Internet-enabled calibrations.

10.2 Evaluation

The success of this work depends on whether the objectives of the thesis are reached. This can be determined by looking at how well the research questions defined in chapter 1.4 have been answered.

Starting with the main research question, this thesis tried to develop a methodology for how to develop an automatic evaluation method. This question has been answered with a set of requirements and a step-by-step framework for the development and implementation of such a method. The requirements impose guidelines for the development of the automatic evaluation method. The framework presented in chapter 9 describes how an automatic method based on the manual method may be set up.

The second goal is the actual implementation of our program. We have only discussed the possibilities for the implementation of our method into the existing calibration process and in future Internet-enabled calibrations. We can conclude that it is a manageable task to call our program from the IMET method. In laboratory calibrations it will be more difficult, but we have proposed one solution, which could work.

The last goal was to discuss whether it is possible to identify instruments based on our program. We tried to see if historical calibration data and new measurement data could uniquely identify an instrument. Many metrologists are sceptical to such an approach. Our discussion concludes that there is a possibility for instrument authentication based on measurement data. This will however require analyzing large amount of drift curves, and might be a time-consuming task. The program would have to look through all possible drift curves for every measurement quantity for all measurement areas. It would then have to compare each of the curves with drift curves for all other instruments of the same type. For one instrument all drift curves that separates from all the other instruments must be stored as a part of the fingerprint for that instrument.

11 Conclusions

This chapter summarizes and concludes the achievements of the work presented in this thesis. It will cover the most interesting contributions in this thesis and the prospects for the future.

11.1 The problem statement

The purpose of the work of this thesis was to investigate the feasibility to automate the manual evaluation of measurement values. We have answered this problem statement by investigating the three objectives given in chapter 1.4. We have seen in this thesis that it is possible to develop an automatic evaluation program based on our established methodology. A more comprehensive conclusion of these three sub-goals is given in the next chapter.

11.2 The main themes

The presented methodology gives a step-by-step method to develop an automatic evaluation program based on the present manual method. It provides the solution of the central challenge to make the manual evaluation method more efficient as well as give better objectivity and quality assurance.

We have developed an automatic evaluation program that evaluates the measurement data from a calibration. This involves analysis methods that partly fulfil the desired requirements given in chapter 7. The automatic evaluation program is aimed at replacing the existing manual evaluation procedures and making the evaluation process more independent of the operators.

We have seen, after performing a couple of tests of the program (see chapter 8.5), that the results are promising. The program was able to evaluate the measurement data correctly and gave the desired evaluation results.

The analysis method could be a big improvement compared to the manual approach due to efficiency issues and error reduction. A vital factor for future success will be to complete the collection of all historical and future measurement values in one DB. This will centralize the storing of these values compared to today’s separated spreadsheets. The verification process will be more practical and less user-dependent compared to the manual method. The system should work as a good decision support system.

11.3 The future

Today, the metrologists spend much time and make unnecessary faults when evaluating the calibration data manually. This thesis shows that advances in information technology provide possibilities for quick and accurate evaluations of data, using databases and software. An automatic evaluation method will be favourable to the operator as well as to the customers. The operator can spend more time on other profitable work and the customers will be more pleased with the evaluation results that are performed faster and include less error.

12 Further work

This section will point out some directions for further work. Although the functionality of our evaluation program has been comprehensive, there are topics and interesting areas for further research. This involves issues concerning possible improvements and extensions of our work. The two most important issues vital to deploying the program in "the real world" is the storing of all calibration data into a DB, and the integration of the application in the calibration process.

12.1 Integrating the automatic evaluation method

Our automatic evaluation program depends on storing all historical and future calibration data in a DB. In JV's case this means that all the historical data currently stored in spreadsheets must be stored in a DB. Methods that fulfil this task may either include a script or a manual insertion by the metrologists. The manual approach will be time-consuming.

Another significant shortcoming is the integration of the application into the calibration process. This should not be a problem for the Internet-enabled method as described in 9.1. For a laboratory calibration this is more complicated. Three possible approaches are described in chapter 8.3.

12.2 Possible extensions

During the work with this thesis we have met some ideas of improvements and extensions of our program that are left for further work. Hopefully, they will give the metrologist more choices and more helpful evaluation functionalities. The topics for these ideas are described in this section.

12.2.1 Other graph plots

Our present program only gives a graph that views the time on the x-axis. Today, when the metrologists perform manual evaluations they draw graphs that also have the frequency and the nominal values on the x-axis, depending on the measurement area (see figure 16). When operating with AC voltage and AC current the graph is plotted using the frequencies on the x-axis. For DC voltage and DC current the graph is plotted with the nominal values on the x-axis.

To meet these user requirements the program should be extended with the option to choose between three different settings on the x-axis, respectively the time, the frequency and the nominal values.

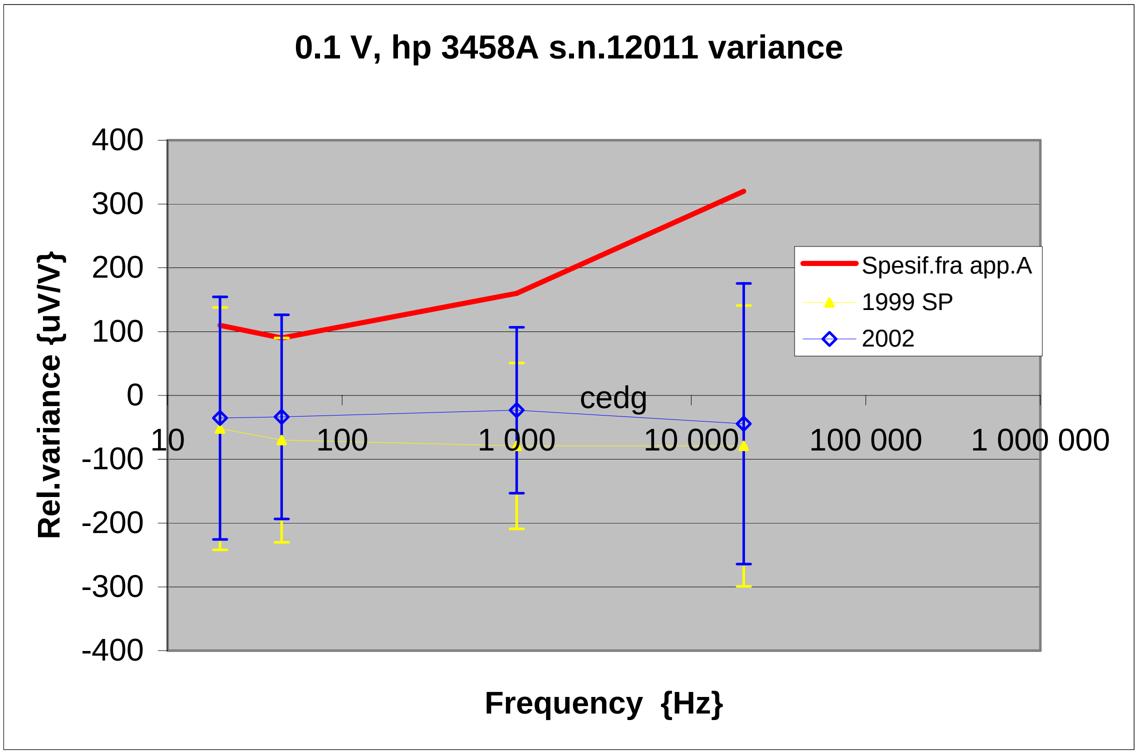

Figure 16: A picture of the graph the metrologists use for evaluation of the calibration data today. The frequency is used on the x-axis.

12.2.2 Calibration interval

The Integrated Sciences Group (ISG) has described a test for setting a calibration interval, also known as "method A3"[2]. An interval fails the test if its observed reliability differs significantly from the reliability target. The reliability target is the interval of time between calibrations that results in holding the percentage of items in use within a minimum acceptance level based on the operators' tolerance boundaries [29]. Respectively, this method involves testing an assigned calibration interval statistically, to determine if the reliability target has been achieved.

This method could be implemented into the program to suggest a date for the next calibration. The statistical calculations could however be difficult to conduct accurately because JV today calibrates too few instruments of the same type. In the future this situation could change and the functionality should be implemented.

[2] Establishment and Adjustment of Calibration Intervals, Recommended Practice RP-1, National Conference of Standards Laboratories, January 1996.

12.2.3 Instrument authentication

This issue has allready been discussed in chapter 10.1.3. Research on this area must include comparing drift curves for every measurement quantity for all measurement areas for instruments of the same type. A vital task should be to find several drift curves that make a unique fingerprint for one instrument.

12.2.4 Reporting

Our automtic evaluation program should be extended with a functionality that makes custom reports automatically. The program should be able to extract results from calculations and put them in a report template that is sent to the customer. It should also put in the graphs with belonging description automatically. The metrologsits should have the possibility to take out reports from historical data, e.g. "Give me all the measurment points on AC current for the instrument HP3500 from year 1997 to 2000".

Bibliography

[1] P. Congdon. *Applied Bayesian Modelling*. Wiley 2003.

[2] Digital Multimeter Measurement Fundamentals:
http://zone.ni.com/devzone/cda/tut/p/id/3295#toc0

[3] J. R. Taylor. *An introduction to Error Analysis – The study of uncertainties in physical measurements.* 2nd edition, University Science Books 1982.

[4] The Least-Squares *m*th Degree Polynomials:
http://www.efunda.com/math/leastsquares/lstsqrmdcurve.cfm

[5] Bayesian analysis: http://www.bayesian.org/

[6] Åsmund Sand. *Applying the Internet to Instrumentation and Metrology*, PhD thesis, NTNU, Dec. 2006.

[7] *Hierarchical Bayesian Statistical Analysis for a Calibration Experiment*, Reid D. Landes, Peter G. Loutzenhiser, and Stephen B. Vardeman, IEEE Transactions on Instrumentation and Measurement, Vol. 55, No. 6, Dec. 2006.

[8] *Expression of the Uncertainty of Measurement in Calibration.* Dec. 1999. Drafted by EAL Task Force of WECC on behalf od the EAL Committee 2 (Calibration and Testing Activities).

[9] Database models: http://unixspace.com/context/databases.html

[10] Fluke. *Artifact Calibration, Theory and Application*, Fluke Corporation, 1996.

[11] Basic Statistics:
http://www.statsoft.com/textbook/stbasic.html#Descriptive%20statistics

[12] Erling Hammer. *Instrumenter og måleteknikk.* Universitetsforlaget, 2utg. 1977.

[13] N. Kularatna. *Modern electronic test and measuring instruments.* 1996: The Institution of Electrical Engineers.

[14] Joseph J. Carr. *Elements of Electronic Instrumentation and Measurement*. Third Edition, 1996 by Prentice-Hall Inc.

[15] Peter Congdon. *Applied Bayesian Modelling.* 2003 John Wiley & Sons Ltd.

[16] James S. Clark & Alab E. Gelfand. *Hierarchical modelling for the environmental sciences.* 2006 Oxford University Press.

[17] Jeremy Reimer. *A history of the GUI.* May 05, 2005.

[18] Seels, B., & Glasgow, Z. *Exercises in instructional design*. 1990, Columbus, OH: Merrill.

[19] Web Services Architecture: http://www.w3.org/TR/ws-arch/

[20] John Hunt. *Guide to C# and object orientation.* 2002, Springer-Verlag London.

[21] Kazdin, A.E. (1992). *Research design in clinical psychology (2nd ed.)*. Boston: Allyn & Bacon.

[22] Kazdin, A.E. (2003a). *Methodology: What it is and why it is so important*. In A.E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (3rd ed., pp. 5-22). Whashington, DC: American Psychological Association.

[23] LabVIEW: http://www.ni.com/labview/

[24] .NET framework: http://msdn2.microsoft.com/en-us/netframework/default.aspx

[25] Determinants http://hyperphysics.phy-astr.gsu.edu/hbase/deter.html

[26] Metrology:
http://www.euromet.org/docs/pubs/docs/Metrology_in_short_2nd_edition_may_2004.pdf

[27] 5700A/5720A Series 11 Multi-Function Calibrator. *Remote Programming Reference Guide.* 1996 Fluke Corporation, Inc.

[28] H. Koeman. *Multifunction, Multirange Electronic Instrument Performance Prediction.* IEEE Transactions on Instrumentation and Measurement, vol. 41, no. 6, Dec. 1992.

[29] *Method A3 – Interval Test Method Description of the Methodology.* 2003-2004, Integrated Sciences Group.

[30] Å. Sand and H. Slinde. *IMet - A Secure and Flexible Approach to Internet-Enabled Calibration at Justervesenet.* In Series on Advances in Mathematics for Applied Science - Advanced Mathematical and Computational Tools in Metrology VII. World Scientific, 2006, vol. 72, pp. 229-236.

[31] Measurement history: http://en.wikipedia.org/wiki/History_of_measurement

[32] Gunnar G. Løvås. *Statistikk – for universiteter og høgskoler.* Universitetsforlaget 2001.

[33] Bradley L. Jones. *Teach Yourself C# på 21 dager.* Aug. 2002.

Appendix A

Least-squares fit for a line

This formula for a straight line: $f(x) = ax + b$, where a and b are constants.
The data points: $Y = [Y_1, Y_2, Y_3, ..., Y_n]$
Error i: ε_i
SSE = Sum of Square Errors

$$\varepsilon_i = f(x_i) - Y_i$$
$$\varepsilon_i{}^2 = (f(x_i) - Y_i)^2$$

$$SSE = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (ax_i + b - Y_i)^2$$

Derivate SSE regarding a and b, and set equal to zero:

$$\frac{\partial SSE}{\partial a} = 2\sum_{i=1}^{n} ax_i{}^2 + 2\sum_{i=1}^{n} bx_i - 2\sum_{i=1}^{n} Y_i x_i$$
$$= 2a\sum_{i=1}^{n} x_i{}^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i Y_i$$
$$= 0$$

$$\frac{\partial SSE}{\partial b} = 2\sum_{i=1}^{n} ax_i + 2bn - 2\sum_{i=1}^{n} Y_i$$
$$= 2a\sum_{i=1}^{n} x_i + 2bn - 2\sum_{i=1}^{n} Y_i$$
$$= 0$$

Define four sums:

$$Sx = \sum_{i=1}^{n} x_i \qquad Sx^2 = \sum_{i=1}^{n} x_i{}^2$$
$$Sy = \sum_{i=1}^{n} y_i \qquad Sxy = \sum_{i=1}^{n} x_i y_i$$

Simplify the expressions above:

$$aSx^2 + bSx - Sxy = 0 \qquad (4)$$

$$aSx + bn - Sy = 0 \qquad (5)$$

Isolate b from equation 5:

$$b = \frac{Sy - aSx}{n}$$

Use this in equation 4:

$$aSx^2 + \frac{Sy - aSx}{n} Sx - Sxy = 0$$

$$aSx^2 + \frac{SySx}{n} - \frac{a(Sx)^2}{n} - Sxy = 0$$

$$a(Sx^2 - \frac{(Sx)^2}{n}) = Sxy - \frac{SySx}{n}$$

$$a = \frac{Sxy - \frac{SySx}{n}}{Sx^2 - \frac{(Sx)^2}{n}}$$

$$a = \frac{nSxy - SySx}{nSx^2 - (Sx)^2}$$

Isolate a from equation 4:

$$a = \frac{Sy - bn}{Sx}$$

Use this in equation 5:

$$\frac{Sy - bn}{Sx} Sx^2 + bSx - Sxy = 0$$

$$\frac{SySx^2}{Sx} + \frac{bnSx^2}{Sx} + bSx - Sxy = 0$$

$$b(\frac{Sx^2 n}{Sx} - Sx) = \frac{SySx^2}{Sx} - Sxy$$

$$b(\frac{Sx^2 n - (Sx)^2}{Sx}) = \frac{SySx^2 - SxySx}{Sx}$$

$$b = \frac{SySx^2 - SxySx}{Sx^2 n - (Sx)^2}$$

The least-squares 2th degree polynomials

The method of least-squares 2th degree polynomials based on the general formula for mth degree polynomial in chapter 4.4:

This method uses the formula:

$$y = F(x) = ax^2 + bx + c$$

$$\varepsilon_i = F(x_i) - y_i$$

$$\varepsilon_i^{\,2} = (F(x_i) - y_i)^2$$

$$SSE = \sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} (ax_i^{\,2} + bx_i + c - y_i)^2$$

$$\frac{\partial SSE}{\partial a} = \sum_{i=1}^{n} 2(ax_i^{\,2} + bx_i + c - y_i) * x_i^{\,2}$$
$$= 2a\sum_{i=1}^{n} x_i^{\,4} + 2b\sum_{i=1}^{n} x_i^{\,3} + 2c\sum_{i=1}^{n} x_i^{\,2} - 2\sum_{i=1}^{n} x_i^{\,2} y_i$$
$$= 0$$

$$\frac{\partial SSE}{\partial b} = \sum_{i=1}^{n} 2(ax_i^{\,2} + bx_i + c - y_i) * x_i$$
$$= 2a\sum_{i=1}^{n} x_i^{\,3} + 2b\sum_{i=1}^{n} x_i^{\,2} + 2c\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i$$
$$= 0$$

$$\frac{\partial SSE}{\partial c} = \sum_{i=1}^{n} 2(ax_i^{\,2} + bx_i + c - y_i)$$
$$= 2a\sum_{i=1}^{n} x_i^{\,2} + 2b\sum_{i=1}^{n} x_i + 2nc - 2\sum_{i=1}^{n} y_i$$
$$= 0$$

Defines :

$$Sx^4 = \sum_{i=1}^{n} x_i^{\,4} \qquad Sx^3 = \sum_{i=1}^{n} x_i^{\,3} \qquad Sx^2 = \sum_{i=1}^{n} x_i^{\,2} \qquad Sx = \sum_{i=1}^{n} x_i$$

$$Sx^2y = \sum_{i=1}^{n} x_i^{\,2} y_i \qquad Sxy = \sum_{i=1}^{n} x_i y_i \qquad Sy = \sum_{i=1}^{n} y_i$$

$$\begin{matrix} Sx^4 & Sx^3 & Sx^2 & & a & & Sx^4y \\ Sx^3 & Sx^2 & Sx & * & b & = & Sxy \\ Sx^2 & Sx & n & & c & & Sy \end{matrix}$$

$$D = \begin{vmatrix} Sx^4 & Sx^3 & Sx^2 \\ Sx^3 & Sx^2 & Sx \\ Sx^2 & Sx & n \end{vmatrix}$$

$$= Sx^4(nSx^2 - (Sx)^2) - Sx^3(nSx^3 - Sx^2Sx) + Sx^2(Sx^3Sx - (Sx^2)^2)$$

$$Da = \begin{vmatrix} Sx^2y & Sx^3 & Sx^2 \\ Sxy & Sx^2 & Sx \\ Sy & Sx & n \end{vmatrix}$$

$$= Sx^2y(nSx^2 - (Sx)^2) - Sxy(nSx^3 - Sx^2Sx) + Sy(Sx^3Sx - (Sx^2)^2)$$

$$Db = \begin{vmatrix} Sx^4 & Sx^2y & Sx^2 \\ Sx^3 & Sxy & Sx \\ Sx^2 & Sy & n \end{vmatrix}$$

$$= Sx^4(nSxy - SySx) - Sx^3(nSx^2y - SySx^2) + Sx^2(Sx^2ySx - SxySx^2)$$

$$Dc = \begin{vmatrix} Sx^4 & Sx^3 & Sx^2y \\ Sx^3 & Sx^2 & Sxy \\ Sx^2 & Sx & Sy \end{vmatrix}$$

$$= Sx^4(SySx^2 - SxySx) - Sx^3(SySx^3 - Sx^2ySx) + Sx^2(SxySx^3 - Sx^2ySx^2)$$

$$a = \frac{Da}{D}, \qquad b = \frac{Db}{D}, \qquad c = \frac{Dc}{D}$$

The least-squares 3^{th} *degree polynomials*

The method of least-squares 3^{th} degree polynomials based on the general formula for m^{th} degree polynomial in chapter 4.4:

This method uses the formula:

$$y = F(x) = ax^3 + bx^2 + cx + d$$

$$\varepsilon_i = F(x_i) - y_i$$

$$\varepsilon_i^2 = (F(x_i) - y_i)^2$$

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (ax_i^3 + bx_i^2 + cx_i + d - y_i)^2$$

Defines :

$$Sx^6 = \sum_{i=1}^{n} x_i^{\ 6} \qquad Sx^5 = \sum_{i=1}^{n} x_i^{\ 5} \qquad Sx^4 = \sum_{i=1}^{n} x_i^{\ 4} \qquad Sx^3 = \sum_{i=1}^{n} x_i^{\ 3} \qquad Sx^2 = \sum_{i=1}^{n} x_i^{\ 2} \qquad Sx = \sum_{i=1}^{n} x_i$$

$$Sx^3y = \sum_{i=1}^{n} x_i^{\ 3} y_i \qquad Sx^2y = \sum_{i=1}^{n} x_i^{\ 2} y_i \qquad Sxy = \sum_{i=1}^{n} x_i y_i \qquad Sy = \sum_{i=1}^{n} y_i$$

From this formula we make four matrixes:

$$D = \begin{matrix} Sx^6 & Sx^5 & Sx^4 & Sx^3 \\ Sx^5 & Sx^4 & Sx^3 & Sx^2 \\ Sx^4 & Sx^3 & Sx^2 & Sx \\ Sx^3 & Sx^2 & Sx & n \end{matrix} \qquad Da = \begin{matrix} Sx^3y & Sx^5 & Sx^4 & Sx^3 \\ Sx^2y & Sx^4 & Sx^3 & Sx^2 \\ Sxy & Sx^3 & Sx^2 & Sx \\ Sy & Sx^2 & Sx & n \end{matrix}$$

$$Db = \begin{matrix} Sx^6 & Sx^3y & Sx^4 & Sx^3 \\ Sx^5 & Sx^2y & Sx^3 & Sx^2 \\ Sx^4 & Sxy & Sx^2 & Sx \\ Sx^3 & Sy & Sx & n \end{matrix} \qquad Dc = \begin{matrix} Sx^6 & Sx^5 & Sx^3y & Sx^3 \\ Sx^5 & Sx^4 & Sx^2y & Sx^2 \\ Sx^4 & Sx^3 & Sxy & Sx \\ Sx^3 & Sx^2 & Sy & n \end{matrix}$$

$$Dd = \begin{matrix} Sx^6 & Sx^5 & Sx^4 & Sx^3y \\ Sx^5 & Sx^4 & Sx^3 & Sx^2y \\ Sx^4 & Sx^3 & Sx^2 & Sxy \\ Sx^3 & Sx^2 & Sx & Sy \end{matrix}$$

We then get the values a, b, c and d:

$$a = \frac{Da}{D}, \qquad b = \frac{Db}{D}, \qquad c = \frac{Dc}{D}, \qquad d\,\frac{Dd}{D}$$

Table for t-distributions

The table lists a few selected values for t-distributions with *v* degrees of freedom for the 90%, 95%, 97.5%, and 99.5% *one-sided* confidence intervals (see chapter 13.4.1).

v	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**	**99.5%**	**99.75%**	**99.9%**	**99.95%**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140

15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646

Table 6: Extraction of the t-distribution table regarding the degree of freedom and confidence level.

One-sided confidence interval

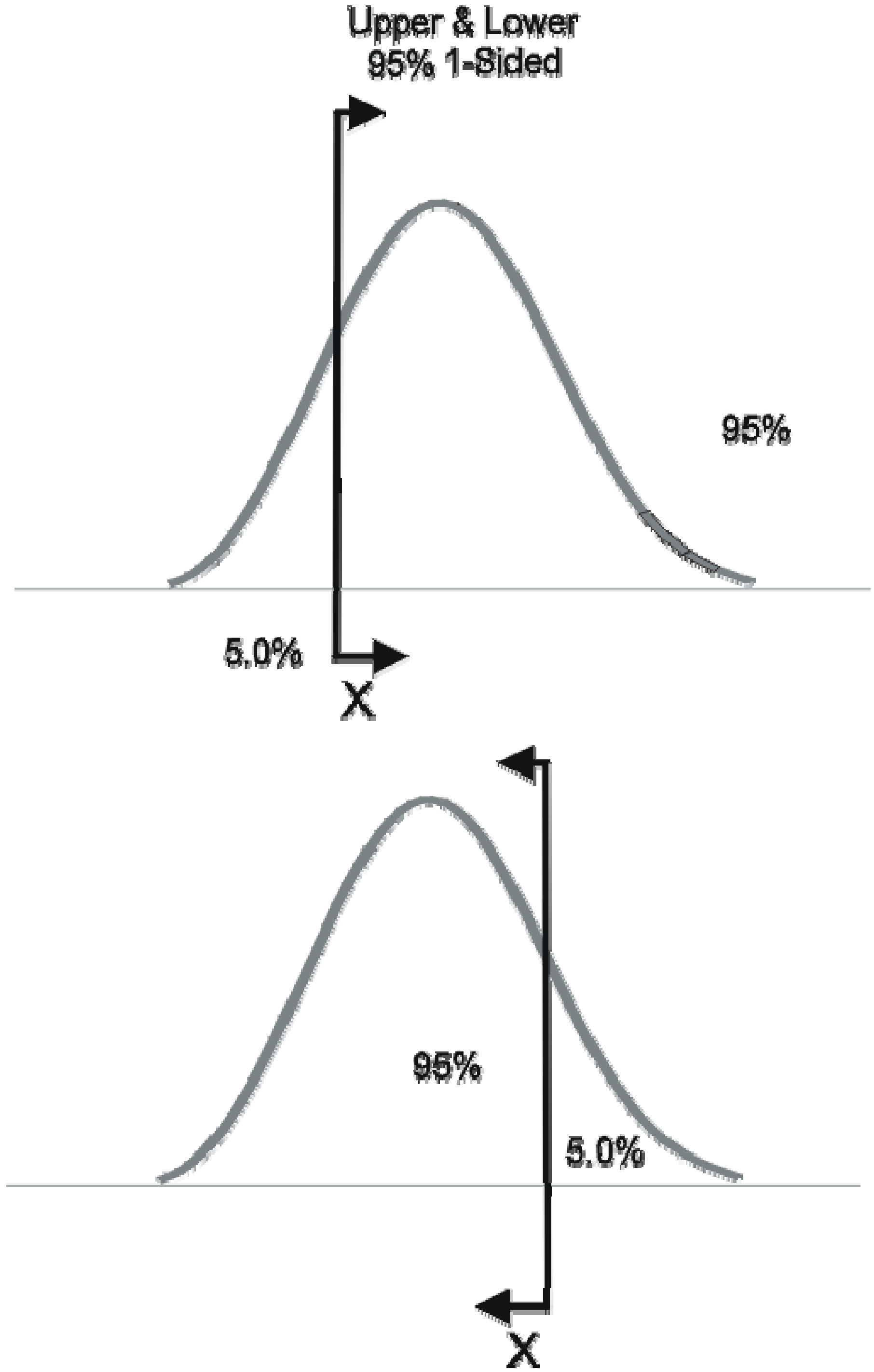

Figure 17: These figures give an example on a one-sided confidence interval. The first graph implies that 95% of the population is greater than X, where X is a 95% lower confidence border. If 95% is less than X *the* 95% is an upper confidence border. This would indicate that 95% of the population is less than X.

Two-sided confidence interval

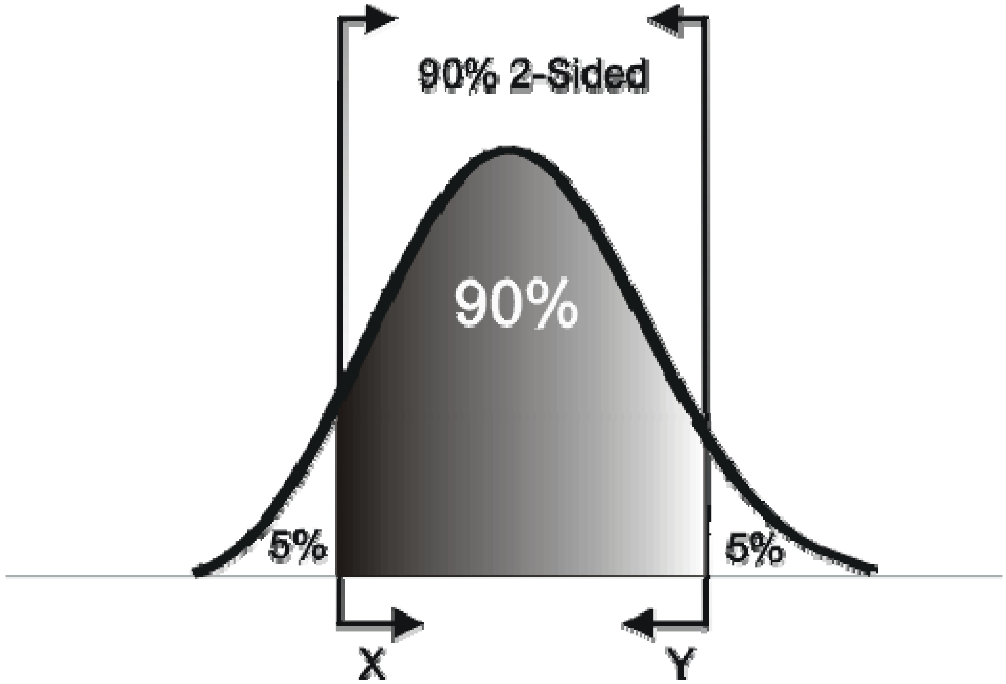

Figure 18: A 90% two-sided confidence interval where 90% lies between X and Y, with 5% less than X and 5% greater than Y.