Tommi Ristinen

Expert Elicitation in Technology Readiness Assessment

Faculty of Electronics, Communications and Automation

Thesis submitted for examination for the degree of Master of Science in Technology.

22.10.2010

Thesis supervisor:

Prof. Jouko Lampinen

Thesis instructor:

D.Sc.(Tech.) Laura Kauhanen

Aalto University
School of Science
and Technology

AALTO-YLIOPISTO
TEKNILLINEN KORKEAKOULU

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Tommi Ristinen

Työn nimi: Asiantuntijamielipiteiden hyödyntäminen teknologian valmiustason määrityksessä

Päivämäärä: 22.10.2010 Kieli: Englanti Sivumäärä:10+73

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Professuuri: Laskennallinen tekniikka Koodi: S-114

Valvoja: Prof. Jouko Lampinen

Ohjaaja: TkT Laura Kauhanen

Tässä työssä käsitellään asiantuntijamielipiteiden hyödyntämistä teknologian valmiustason määrityksessä. Työssä suunnitellaan ja toteutetaan prosessi sekä käytännön työkalu asiantuntija-arvioiden hankkimiseen, analysointiin ja raportointiin. Asiantuntijamielipiteet kerätään sähköpostikyselyiden avulla ja syötetään analyysiä varten kehitettyyn ohjelmaan, joka tiivistää tiedon helposti ymmärrettäviksi kuviksi ja tunnusluvuiksi. Työssä vertaillaan eri analyysimenetelmiä ja niiden paremmuutta sekä arvioidaan yleisesti asiantuntijamielipiteiden hyödyntämisen luotettavuutta ja prosessiin liittyviä harhoja. Erot erilaisten analyysimenetelmien välillä todetaan varsin pieniksi ja tärkeimmät huomion kohteet asiantuntijamielipiteiden hyödyntämisessä liittyvät datan keräykseen ja epävarmuustekijöiden arviointiin ja niiden minimointiin. Lopputuloksena työssä toteutettu prosessi ja analyysiohjelma ovat käyttökelpoinen ja toimiva tapa teknologian valmiustason määritykseen.

Avainsanat: teknologian valmiustason arviointi, teknologia-arviointi, asiantuntijamielipiteiden hyödyntäminen, asiantuntijamielipide, päätösanalyysi, latenttimuuttujamalli, ordinaaliasteikko

AALTO UNIVERSITY
SCHOOL OF SCIENCE AND TECHNOLOGY

ABSTRACT OF THE
MASTER'S THESIS

Author: Tommi Ristinen

Title: Expert Elicitation in Technology Readiness Assessment

Date: 22.10.2010 Language: English Number of pages:10+73

Faculty of Electronics, Communications and Automation

Professorship: Computational Engineering Code: S-114

Supervisor: Prof. Jouko Lampinen

Instructor: D.Sc.(Tech.) Laura Kauhanen

This thesis covers expert opinion elicitation for Technology Readiness Assessment. A process and a tool is designed and implemented for obtaining, analysing and reporting expert opinion. The experts are contacted by web questionnaires and the data is inputted in the analysis program, which composes the data in easily interpretable diagrams and figures. This thesis compares different analysis methods in terms of their performance and discusses in general the usage of expert opinion, its reliability and related biases. The differences between the analysis methods are found to very small and most important indications of the final results are that most attention in expert elicitation should be paid to collection of data and estimating and minimising the uncertainties in the process. In conclusion, the process and analysis tool developed in this thesis are a practical and well working technique for Technology Readiness Assessment.

Keywords: Technology Readiness Level, Technology Readiness Assessment, expert elicitation, expert opinion, decision analysis, latent variable model, ordinal data

Preface

This work was done at Spinverse Oy as part of my studies in the Department of Biomedical Engineering and Computational Science of Aalto University. The work is closely related to European Commission FP7 funded project ObservatoryNANO. The thesis was instructed by Dr. Laura Kauhanen at Spinverse and supervised by Prof. Jouko Lampinen at the Aalto University.

I would like to express my gratitude to Laura Kauhanen for all guidance and dedication throughout the whole process. I also want to thank Jouko Lampinen for his supervision and valuable comments and Tom Crawley from Spinverse for his help in defining the work and making it possible altogether.

Finally, I want to thank Riikka and my parents for supporting me during the whole course of my studies.

Otaniemi, 22.10.2010

Tommi Ristinen

Contents

Abstract (in Finnish) ii

Abstract iii

Preface iv

Contents v

List of Figures viii

List of Tables ix

Abbreviations x

1 Introduction 1
1.1 General introduction . . . 1
1.2 Concept of technology readiness . . . 3
1.2.1 Technology in this context . . . 3
1.2.2 Technology Readiness Level and technology impact . . . 6
1.3 Technology assessment . . . 8
1.3.1 Introduction to technology assessment . . . 8
1.3.2 Key methodology . . . 10
1.3.3 Summary and integration of methods . . . 14
1.4 Use of expert opinion in decision analysis . . . 15
1.4.1 Decision analysis . . . 15
1.4.2 Expert use in decision analysis . . . 17
1.4.3 Assessment of uncertainty . . . 18
1.4.4 Sources of bias . . . 19
1.4.5 Summary and implications . . . 20
1.5 Expert opinion aggregation . . . 21
1.5.1 Measuring and modeling expert opinion . . . 21
1.5.2 Axiomatic and Bayesian models . . . 24
1.5.3 Measures for rater agreement . . . 26
1.5.4 Latent variable models for ordinal data . . . 26
1.5.5 Subjective logic . . . 28
1.5.6 Other methods . . . 30

2 Method 31
2.1 Introduction to methodology . . . 31
2.2 Requirements specification . . . 31
2.3 Data . . . 33
2.3.1 Collection and description . . . 33
2.3.2 Interpretation and analysis . . . 36
2.4 Modeling . . . 38
2.4.1 Multinomial and weighted multinomial models . . . 38

2.4.2 Judge-specific model . . . 39
2.4.3 Intraclass correlation based model - "ICC" . . . 40
2.4.4 Rater behaviour based model - "Ratings" . . . 41
2.4.5 Subjective logic . . . 41
2.5 Reliability . . . 44
2.5.1 Sensitivity analysis . . . 44
2.5.2 Verification and validation . . . 45
2.5.3 Estimating and reducing uncertainty and bias . . . 46
2.6 Implementation . . . 47

3 Results **48**
3.1 Comparison . . . 48
3.2 Reliability, robustness and sensitivity . . . 50
3.3 Final estimates . . . 54
3.4 Validation of results: case universal memory . . . 56
3.5 Other means for estimating technology readiness level . . . 58

4 Discussion **60**
4.1 Process . . . 60
4.2 Methodology . . . 61
4.3 Future research . . . 62

References **64**

Appendix A **68**

Appendix B **70**

Appendix C **71**

Appendix D **73**

List of Figures

1 ObservatoryNANO TRL scheme (left) corresponding to the defence standard nine-point scheme (NASA, right)(ObservatoryNANO, 2010d) 7

2 A Framework for Technology Futures Analysis shows relations between business or public policy, decision makers and decision studies. The outputs of TFA are used to e.g. increase awareness or inputs in policy making and the findings may be the starting point for a new TFA study. Adapted from Porter et al. (2004). 9

3 The evaluation and organisation of futures research methodology adapted from Aaltonen (2009). The figure shows relative position of relevant futures research methodology regarding this work in terms of nature of the system and level of ambiguity. 10

4 A decision-analysis process flowchart adapted from (Clemen, 1996). The process is followed in the structure of this study. 16

5 Normal latent trait for four category ordinal data. Numbered x axis denotes the true value of the measured variable and the letters between vertical lines (cutoff points) mark the ordinal categories. 27

6 Example of a frame of discernment (Jøsang and Knapskog, 1998). Each x_i presents an elementary state for which the belief, disbelief and uncertainty is defined by 9. Elementary states can also be overlapping and subjective logic provides operators for computing the opinion tuple for any combination (union, intersection, etc.) of the elementary states. 29

7 Triangle showing the opinion space in subjective logic. Adapted from Jøsang and Knapskog (1998). Variable ω denotes the opinion defined in equation 9. 29

8 Mapping of TRL data to subjective logic framework. The dots represent the effect of uncertainty in the values of belief and disbelief. When uncertainty approaches 1 in the top corner of the triangle, the numerical differences in belief values between different categories become smaller. Therefore opinions with high uncertainty possess less information than opinions with low uncertainty. 42

9 Median Absolute Percentage Error (MAPE) (a) and Median Absolute Deviation (MAD)(b) for estimates of methods compared. Clear groups with different performance levels are present in figure (a) whereas the results in (b) are more dispersed. Both measures, however, lead to same conclusion of methods' performances. 51

10 MAPE for TRL 1 probability of Nanoimprint lithography as function of number of respondents. Sampling with replacement (bootstrap), number of samples 100. 52

11 Distribution of the final estimate of TRL 1 for Nanoimprint lithography. (a) Sampling with replacement (bootstrap), (b) sampling without replacement. Number of samples 100. The most averaging subjective logic variant has least variation whereas the basic subjective logic has most variation. 53
12 Final estimates obtained by the tool. Analysis method used is subjective logic 54
13 Final estimates obtained by the tool for Universal Memory technologies. Original figure provided by the tool slightly modified manually in order to fit it in report template. Circle size denotes the number of respondents to the questionnaire, which is directly related to reliability of the results. Therefore the conclusion for RRAM, PCRAM and MRAM can be considered most reliable. Red color used shows what are the most potential technologies according to qualitative feedback from the experts. The number inside circles tells the median estimate of time to market entry or maturation. 57
A.1 Data on TRL related claims for printed electronic manufacturing technologies: Nanoimprint lithography (a) and ink-jet printing (b) . . 68
A.2 Data on TRL related claims for printed electronic manufacturing technologies: gravure printing (a) and flexography printing (b) 69

List of Tables

1 NASA definition of Technology Readiness Levels (DoD, 2009; Mankins, 1995) 6
2 Classification of scales of measurement (Stevens, 1946) 23
3 Response rates for questionnaires conducted 35
4 Basic statistics for printed electronics data. Barplots of category frequencies for each technology can be found in Appendix A. 37
5 Sample data for TRL 1 of Nanoimprint lithography in printed electronics 37
6 Suggested TRL interpretation scheme. The pattern describes ordinal category frequencies for: strongly agree-moderately agree-moderately disagree-strongly disagree. The sum of probabilities exceeds 100% due to rounding. 38
7 The effect of models in comparison to the basic multinomial model. The first row indicates the share of individual TRL scores being closer to neutral opinion. The second row contains the average effect of the model: e.g. -4% means that on average subjective logic scores were 4 percentage points closer to neutral than with the basic model. Notable deviations from expected value labeled by exclamation marks. 49
8 Results broken down on the level of individual technologies show that nanoimprint lithography is drawn most aggressively towards neutral values. Weighed multinomial model on the other hand provides more extreme results for each technology. ICC and Judge-specific models provide the most robust results in broad sense. Notable deviations from expected value labeled by exclamation marks. 49
9 Impact values for nanoscale printing technologies for the moment being and at estimated time of market entry. Resulting impacts are plotted in figure 12 55
10 TRL level estimates on different decision limits. The limit setting greatly affects the results, which need to be considered when analysing the results. 55
11 TRL level estimates on different decision limits. The limit setting greatly affects the results, which need to be considered when analysing the results. 56
12 Estimating TRL by year estimates. Experts gave their answer on scale {Now, 1-3 years (Short), 4-7 years (Medium), Later (Long), Never}. The scale is also transformed to numerical values and means are calculated for the values. TRL estimated from expert given year estimates gives slightly more optimistic results than the detailed claims. 58
B.1 A simple taxonomy of Futures Research Methods (Glenn and Gordon, 2009) 70
D.1 Full results of method comparisons for the printed electronics data . . 73

Abbreviations

AHP	Analytic Hierarchy Process
CBRAM	Conductive Bridging RAM
CEO	Chief Executive Officer
CNT RAM	Carbon Nanotube RAM
CTO	Chief Technology Officer
CTE	Critical Technology Element
DoD	United States Department of Defense
ESA	European Space Agency
FeRAM	Ferroelectric RAM
FLE	Flexography printing
GRA	Gravure printing
ICC	Intraclass Correlation Coefficient
ICT	Information and Communication Technology
IRL	Innovation Readiness Level
INK	Ink-jet printing
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
MCMC	Markov Chain Monte Carlo
MRAM	Magnetoresistive RAM
MRL	Manufacturing Readiness Level
MSA	Measure System Analysis
NASA	National Aeronautics and Space Administration
NIL	Nanoimprint Lithography
PCRAM	Phase-change RAM
QD RAM	Quantum dot RAM
R&D	Research and Development
RAM	Random Access Memory
RRAM	Resistive RAM
RDM	Robust Decisionmaking
SL	Subjective logic
SONOS	Silicon-Oxide-Nitride-Oxide-Silicon
SRL	System Readiness Level
TFA	Technology Futures Analysis
TRA	Technology Readiness Assessment
TSA	Technology Sequence Analysis
TRL	Technology Readiness Level
VAS	Visual Analog Scale

1 Introduction

1.1 General introduction

The knowledge of technological development stage is crucial in all decision making related to emerging technologies. Moreover, the assessment of technological development stage is a difficult decision making problem itself mainly due to data and information available for decision making. There is a huge amount of information available in the form of patents, scientific publications and news articles but it is very difficult for the decision maker to find the relevant information for the problem in hand.

Technology Readiness Assessment (TRA) is a concept developed by NASA in the 1970s for risk management purposes in research and technology development programmes (Mankins, 2009). The aim of TRA is to find out the Technology Readiness Level (TRL) for all technologies in a system (e.g. a space shuttle) and the information of individual TRLs is used for estimating whether the system is ready to be used in missions. The TRL is defined on a nine point scale ranging from the observation and reporting of basic scientific principles to the actual system proven to work through successful mission operations. The basic idea behind the scale is that a system consisting of components still under development is more likely to fail than a system consisting of components that are proven to work in previous missions.

Among many other organisations, the concept of TRLs has been adopted by the European Commission's ObservatoryNANO project[1], which aims to support European decision makers with information and analysis on developments in nanoscience and nanotechnology. The final goal of the project is to provide ongoing and independent support to European decision makers by analysing scientific and technological trends as well as economic realities and expectations. This information is used to help the EU make strategic judgements on funding priorities, prioritise research and technology development programmes and provide investors with early indicators of opportunities.

Within the ObservatoryNANO project, there is a need for assessing the TRL as well as the potential impact of certain nanotechnologies. TRL scale used by the project resembles the NASA scale but is simplified to five levels due to more market oriented and more future looking approach of the project. Most of the technologies assessed by the project are still in very early stage of development but it is necessary for decision makers in Europe to know what are the technologies that will have highest impact in five to twenty years of time and what will be their impact for applications and the community.

So far, most of the TRL assessment has been based on interviews and desk research conducted by analysts. The sources of information include patents and scientific

[1] http://www.observatorynano.eu/project/

publications along with statistical information from respective databases. Most of the analyst work and decision making is, however, based on subjective opinions. TRL assessment conducted by a single analyst both takes a lot of time and is subject to many sources of bias and uncertainty. This study aims to create a more systematic and automatic process for the TRL assessment and provide a simple tool for conducting the TRL analysis quickly, economically and reliably. The tool is intended to visualise the technological development stage by linking together the TRL and technology impact in a visually effective way. The manual analyst work, however, cannot be fully replaced by an automatic tool but the goal is rather to support the TRL assessment process by providing information in a compact form. This information is then supported by links to relevant publications and descriptions related to the technologies being assessed.

Technology assessment can be considered as a part of larger field of futures research where one aim is to find out how technological development will affect the society. Technology Futures Analysis (TFA) and futures research methodology include general decision support methods for various purposes. Along with NASA's TRA scheme, they represent the background of this study in the sense of collecting and analysing data for future looking decision making problems. Relevant methodology for TRL assessment is very diverse including quantitative, qualitative, normative and exploratory methods as well as different kinds of mixtures of these. Section 1.2 of this study gives a general introduction to the concept of technology readiness and TRL from both NASA and ObservatoryNANO project point of view. Methodology used for TFA and futures research is reviewed in section 1.3 in order to find out what kind of approaches have been successful for decision making problems similar to this study.

The use of expert opinion is important to decision making problems where little or no data is available or it is very difficult to use. This applies for this study, as there is no clear set of data describing the TRL of a certain nanotechnology, for instance quantum dot technology for optical interconnects in microprocessor chips. A straightforward approach is to use relevant experts such as researchers in universities, research organisations and companies to give their assessment on the technological development stage on a quantitative scale. The use of expert opinion has been widely studied in literature which is reviewed in section 1.4. This includes an introduction to decision analysis and expert use, modeling of uncertainties and bias as well as presenting relevant methodology that will be used in section 2.

The aim of this study is to develop a robust decision making process and tool for TRL assessment for the ObservatoryNANO project. This includes designing of a web questionnaire for expert data collection, developing a method for analysing and reporting the data and producing the analysis along with estimation of different uncertainties in the process. Moreover, this study aims to find out whether there is a need for advanced mathematical methods for combining expert opinion or if simple aggregation schemes such as calculating the average of expert opinion is enough.

Section 2 of this study specifies the requirements for the tool, introduces the data and its collection process, presents different methods for analysing the data and assessing the reliability of the results. The results, that is TRL and technology impact estimates, for three technology areas are presented in Section 3. Technologies assessed as part of this study are printed electronics manufacturing technologies, nanostructures for on-chip and chip-to-chip optical interconnects and universal memory technologies. Discussion and suggestions for future research are in section 4.

1.2 Concept of technology readiness

1.2.1 Technology in this context

In order to assess technological readiness, the term technology needs to be defined. The European Space Agency defines technology as our species' ability to make and use new tools: "Technology is the practical application of knowledge so that something entirely new can be done, or so that something can be done in a completely new way" (ESA, 2009). Another definition by Merriam-Webster Online Dictionary (2010b) defines technology as "a capability given by the practical application of knowledge" and "a manner of accomplishing a task especially using technical processes, methods, or knowledge". Different fields of technology include information and communication technology or ICT, construction technology and energy technology to name a few. On the other hand the word technology can be found in different functions such as manufacturing technology, measurement technology or materials technology.

Another concept closely related to technology is science. By the definition of Merriam-Webster Online Dictionary (2010a), science means "knowledge or a system of knowledge covering general truths or the operation of general laws especially as obtained and tested through scientific method" and "such knowledge or such a system of knowledge concerned with the physical world and its phenomena". Different fields of science include physics, chemistry or information sciences and deal with, for instance, theories about different physical properties and interactions. In a nutshell, science is about knowledge of the physical world and its phenomena and technology is the practical application of this knowledge for accomplishing tasks and solving problems.

Nanotechnology is a relatively new area of technology that originates from the 1950s. In 1959, Richard Feynman gave his famous speech where he presented a technological vision of extreme miniaturisation (Bhushan, 2006). However, the emergence of appropriate methods of fabrication of nanostructures in 1980s made a number of significant technological developments possible. In a nutshell, nanotechnology can be defined as dealing with various structures of matter having dimensions of the order of a billionth of a meter ($1 \times 10^{-9} m = 1 nm$) (Poole et al., 2003). Nanoscale structures as themselves are not a novel thing as they exist in nature but controlling the matter at nanoscale provides novel properties of materials that are not present in the larger scales. Nanoparticles are aggregates of atoms bonded together with a radius between 1 and 100nm (Bhushan, 2006). In summary, nanotechnology has been defined by National Nanotechnology Initiative as "Research and technology development at the atomic, molecular or macromolecular levels, in the length scale of approximately 1 - 100 nanometer range, to provide a fundamental understanding of phenomena and materials at the nanoscale and to create and use structures, devices and systems that have novel properties and functions because of their small and/or intermediate size" (Roco, 2001).

For assessing the technological development stage, there must be a distinction between nanoscience and nanotechnology. An application focused approach is used in this study because it offers a practical starting point for defining technology areas and furthermore the questions in expert elicitation. In this approach, a nanoscale feature or structure is not meaningful alone and it requires an application or a problem its solves. This can be, for instance, a novel feature that nanotechnology enables for given technology area or performance increment that cannot be achieved using non-nanotechnology. A few examples of technology definitions in this study are given below.

Technology areas assessed as part of this study include printed electronics, optical interconnects and universal memory technologies. Printed electronics is about printing electronic systems on a substrate by methods similar to conventional printing techniques (ObservatoryNANO, 2010c). Printed electronics as a whole is not nanotechnology but nanotechnology will likely have a positive effect on various printed electronics applications as it provides miniaturisation of printed devices and novel features as well as offers cost savings and performance increase. Optical interconnect technology is used for data transmission in very short range connections such as on and between microprocessor chips (ObservatoryNANO, 2010b). Due to nature of light, optical interconnects are naturally nanotechnology. Universal memory technologies are a group of emerging technologies that potentially offer very high speed operation, high density storage and non-volatility with low power consumption (ObservatoryNANO, 2010a).

There are several fields of nanoscience that are present in printed electronics, optical interconnects and universal memory. For all these, relevant fields of nanoscience include (but are not limited to) semiconductor physics, materials science and nanophotonics. All of these areas are, however, too broad for conducting Technology Readiness Assessment. For example, quantum dots are a three dimensional nanostructure (sometimes referred as a nanomaterial) that can be fabricated of various materials such as Silicon or III-V semiconducting materials. Quantum dots are a discovery originating from scientific research but they have no relevance as such and therefore it does not make sense to assign a TRL for them.

As technologies are closely related to application of science, a fruitful viewpoint for choosing technologies to be assessed is to start from the real applications. In this approach, a discovery typically from scientific research is applied to some real world challenge, which eventually defines a nanotechnology for which Technology

Readiness Assessment is feasible. Quantum dots can be applied in several areas of nanotechnology, for instance in optical interconnects or photovoltaics. Moreover, the material used for fabrication of quantum dots is also important and with these constraints a nanotechnology called "Silicon quantum dot laser" can be defined. Similarly "Quantum dot laser based on III-V materials" makes sense.

As a fundamental feature of nanotechnology is that controlling the matter in nanoscale changes the properties of matter, the role of materials in defining nanotechnologies is difficult. A good example are nanoparticle inks, which are used for printed electronics. As the properties of the ink depend on the material used in the ink and the diameter of particles, conducting Technology Readiness Assessment does not make sense for nanoparticle inks. If one were to assess and compare the technological readiness of nanoparticle inks, there would be in theory infinite amount of different technologies for assessment.

In conclusion, the following list of nanotechnologies are assessed and used as practical examples for developing the TRA framework:

- Manufacturing technologies for nanoscale printed electronics
  - Roll-to-roll nanoimprint lithography
  - Ink-jet printing
  - Gravure printing
  - Flexography printing
- Nanotechnologies/nanostructures for optical interconnects
  - Quantum wells of Silicon
  - Quantum dots of Silicon
  - Quantum wells of III-V materials
  - Quantum dots of III-V materials
  - High index-contrast structures (e.g. photonic crystals)
  - Surface plasmon polaritons
- Universal Memory technologies
  - Magnetoresistive RAM (MRAM)
  - Phase-change RAM (PCRAM)
  - Ferroelectric RAM (FeRAM)
  - Resistive RAM (RRAM)
  - SONOS
  - Conductive Bridging RAM (CBRAM)
  - Carbon nanotube based RAM (CNT RAM)
  - Quantum dot RAM (QD RAM)
  - Racetrack memory

1.2.2 Technology Readiness Level and technology impact

The success of advanced technology research and development (R&D) efforts is crucial for the development of new system capabilities. The main challenges any system development project face inevitably are related to performance, schedule and budget. The National Aeronautics and Space Administration (NASA) introduced the concept of Technology Readiness Levels (TRLs) in the mid 1970's to allow more effective assessment of, and communication regarding the maturity of new technologies (Mankins, 2009). Technology Readiness Assessment (TRA) is a formal, systematic, metrics-based process for assessing and reporting the maturity of technologies. The United States Department of Defense (DoD) and NASA use TRA for assessing the maturity of certain technologies called Critical Technology Elements (CTEs) to be used in systems. TRLs are the metric used assessing the maturity of CTEs and they are based on a scale from one through nine (see table 1)(DoD, 2009). In addition to measuring the technological development stage, TRLs provide a systematic measurement system for consistent comparison of maturity between different types of technology (Mankins, 1995).

Table 1: NASA definition of Technology Readiness Levels (DoD, 2009; Mankins, 1995)

TRL	Description
1	Basic principles observed and reported
2	Technology concept and/or application formulated
3	Analytical and experimental critical function and/or characteristic proof-of-concept
4	Component and/or breadboard validation in laboratory environment
5	Component and/or breadboard validation in relevant environment
6	System/subsystem model or prototype demonstration in a relevant environment
7	System prototype demonstration in an operational environment
8	Actual system completed and qualified through test and demonstration
9	Actual system proven through successful mission operations

As the concept of TRL has originated from military organisations and defence industry, it is still mostly used for such applications. Organisations using TRLs as a tool for managing their research and development efforts include national defence organisations such as UK Ministry of Defence (UK MOD, 2010), NATO, Australian Defence Organisation and recently the Turkish defence industry (Altunok and Cakmak, 2010). The main reason for using TRLs in military organisations is technology related risk management for multiple interdependent technologies. The users among these organisations include especially the technology management and systems engineering.

In addition to defence organisations, both the European Space Agency (ESA) and

NASA in United States are among the users of TRLs. Both use TRLs for technology risk management that is especially important for missions in space. In industry, TRLs have been applied by Institut national d'optique and Brio Conseils in Canada and by Network Rail in the United Kingdom. The former uses TRLs for technology transfer and managing innovation efficiently. Benefits can be found in project management, business development, intellectual property strategies and resources allocation. The latter employs TRLs as part of new product introduction process with business benefits, safety assurance, asset protection and supply diversity.

For some uses, the concept of TRL has been considered too restrictive and similar application specific measures are suggested to solve this issue. These include Manufacturing Readiness Levels (MRL), System Readiness Levels (SRL) and Innovation Readiness Level (IRL). These measures extend the original TRL framework and are not considered any further in this study.

Different organisations use TRLs for different purposes and a few parallel systems for measuring TRLs are used. NASA and DoD use a 9 level scale, which is mainly used for risk management purposes in technology development. The definition of TRLs used by NASA and DoD are listed in table 1. The ObservatoryNANO project uses a simpler 5 level scheme that is more suitable for market oriented technology assessment because it describes the technology readiness by intuitively named levels that are easy to understand. The downside in the 5 level scheme is that there a lot of different operations performed under each TRL and transitions between different levels require a large effort. The mapping between the schemes can be found in figure 1.

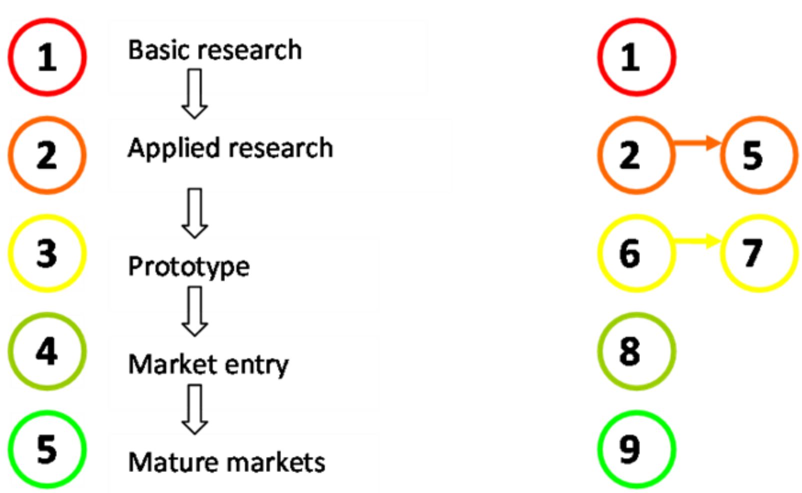


Figure 1: ObservatoryNANO TRL scheme (left) corresponding to the defence standard nine-point scheme (NASA, right)(ObservatoryNANO, 2010d)

The first level in ObservatoryNANO scale is fundamental research that corresponds to the first level in NASA scale. Fundamental research phase includes the observation and reporting of basic principles of the technology. Applied research phase (levels 2-5 in NASA scale) includes formulation of the technology concept and demonstrating the main components of the technology in both laboratory and relevant environments. The prototype level includes prototype demonstration in a relevant and in the final operating environment, which are levels 6 and 7 in the NASA scale. Despite very different naming schemes for the last two levels of technology readiness, the same idea exists behind both ObservatoryNANO and NASA scales. The second to last level, "Market entry" or "Flight qualification" means that the actual product or system is successfully demonstrated in its final operating environment. Similarly the last TRL, either "Mature markets" or "Flight proven" means that the technology is successfully operated in the final operating environment for several times. Mature markets phase also implies that there are several technology providers in the market.

NASA's TRL calculator[2] originated from the need to have a standard, consistent method for assessing and implementing the TRLs. The TRL Calculator attempts to address the issue of lacking guide on "How to use TRLs" by providing technology program managers with a tool that can be used to provide a snapshot of technology maturity at a given point of time (Nolte et al., 2003). According to a survey by Graettinger et al. (2002), TRLs account to 30% of the factors needed in making technology selections and in total up to 50 different factors need to be considered in technology transfer from university research to commercial use. The authors did not manage to find any commercially available tool suitable for DoD in their TRA. Statistical validity of the TRL Calculator has not been demonstrated and Nolte et al. (2003) suggests that a formal validation should be performed.

1.3 Technology assessment

1.3.1 Introduction to technology assessment

Technology Readiness Assessment can be considered a part of a broader area of Technology Futures Analysis. According to Porter et al. (2004) the analysis of emerging technologies and their implications are vital to today's economies, societies and companies. Decision makers in organisations need to be well-informed in order to prioritising research and development (R&D) efforts, understanding and managing risks, exploiting intellectual property and enhancing technological competitiveness.

Multiple, often overlapping methodologies are used for technology intelligence, forecasting, roadmapping, assessment and foresight. Porter et al. (2004) introduce the umbrella concept Technology Futures Analysis (TFA) to cover the field of technology-oriented forecasting methods and practices. They divide the field into "Technology foresight", "Technology forecasting" and "Technology assessment". TFA itself intersects with a wider concept of futures research. Most methodology used for futures research can be applied in technology futures analysis, too

[2]The most recent version of the calculator can be found in `https://acc.dau.mil/CommunityBrowser.aspx?id=25811&lang=en-US`

and technology futures analysis can be considered as a subset of futures research. A framework of TFA in Fig. 2 summarises the inputs, outputs, applications and relationships between different players of the TFA process.

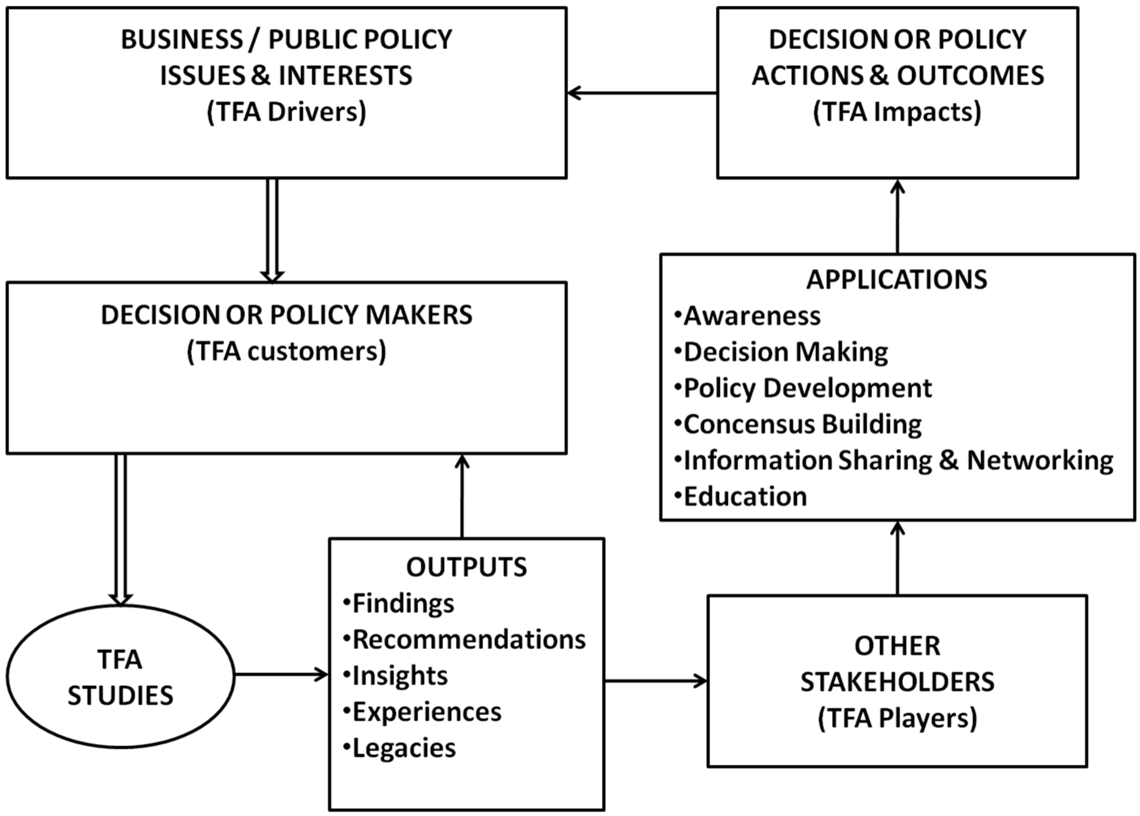


Figure 2: A Framework for Technology Futures Analysis shows relations between business or public policy, decision makers and decision studies. The outputs of TFA are used to e.g. increase awareness or inputs in policy making and the findings may be the starting point for a new TFA study. Adapted from Porter et al. (2004).

As their primary reference of Technology Futures Analysis, Porter et al. use the Futures Research Methodology V2.0 (version cited is V3.0)(Glenn and Gordon, 2009) by the United Nations' Millennium Project. The book includes a large collection of futures research methods for various different situations where forecasting the future is necessary. A simple taxonomy suggested by the book suggests classifying the methods in quantitative, qualitative, normative and exploratory categories. This classification and listing of methods included in the book can be found in Appendix B.

Another way for classifying and organising futures research methodology is suggested by Aaltonen (2009). Aaltonen classifies the methods in a two by two matrix (Fig. 3) which includes the means of controlling or directing the system on horizontal dimension and nature of possible understanding of system on vertical dimension. The measure used in the evaluation of controlling or directing the system is based on the level of ambiguity of the results. The nature of possible understanding takes into account the standpoint of the researcher trying to understand the system. With designed systems the expert can stand outside the system whereas with emergent systems the system cannot be understood or managed as a whole by the researcher. This is because the system emerges through the interaction of the agents involved: people, processes, technology, etc.

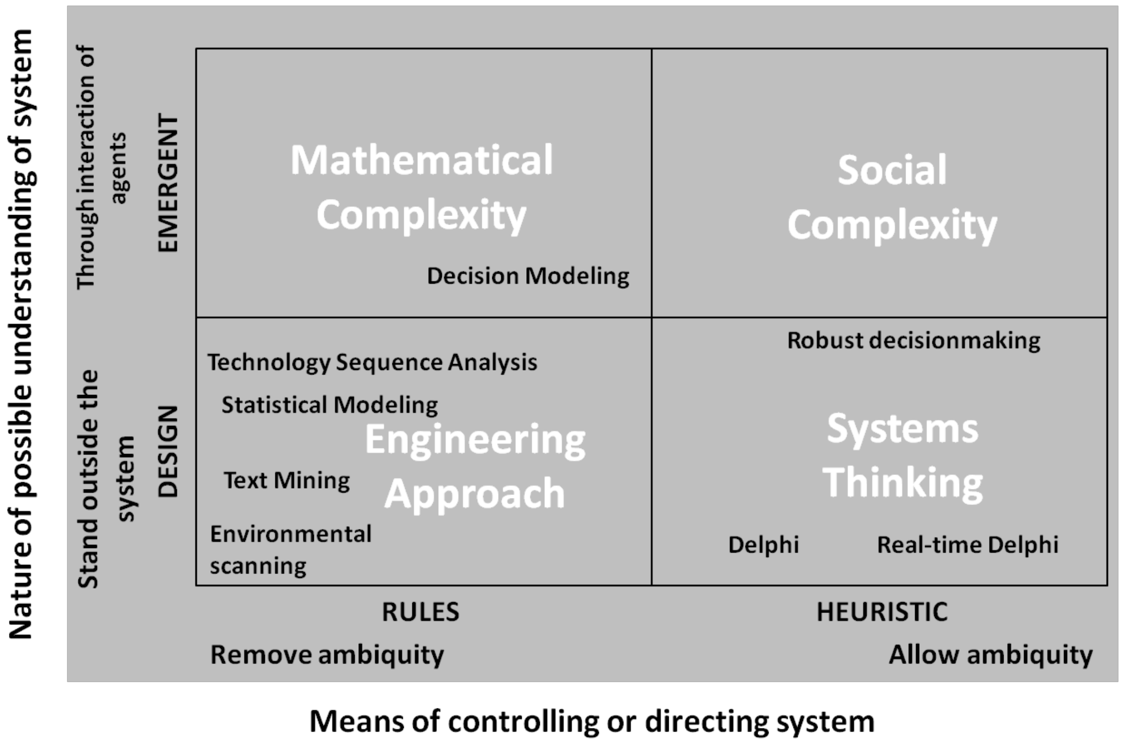


Figure 3: The evaluation and organisation of futures research methodology adapted from Aaltonen (2009). The figure shows relative position of relevant futures research methodology regarding this work in terms of nature of the system and level of ambiguity.

1.3.2 Key methodology

This section presents relevant key methodology for futures research. The methods introduced in this section will not be used as such later on in this thesis but they provide an important insight to the methodology used in similar forecasting or technology assessment problems. All methods have both advantages and disadvantages for any given decision making problem and therefore they provide a valuable input for the development of efficient methodology. Descriptions of methods presented are based on Glenn and Gordon (2009).

Delphi is a forecasting method using expert panels originally developed by RAND Corporation in the early 1960s. Delphi method begins with inviting a set of experts from different disciplines to participate in a study to establish a forecast on, for instance, the date by which a manned Mars landing would occur. Delphi technique is based on several rounds of sequential questionnaires used for refining relevant questions, providing new insights from the experts and finally reaching a consensus between the expert panel. To eliminate biases in the process, there is no direct expert to expert interaction in Delphi and the communication between experts is achieved through the researchers.

The questions are first refined by the researchers conducting the study and presented to each expert individually in a form of a questionnaire. The experts then provide their feedback to the questionnaire, e.g. their judgement on the date of the Mars landing, and researchers analyse the results of the first questionnaire and present the range of opinions to the experts in the form of a second round questionnaire. The experts having extreme opinions are then asked to reassess their opinion and provide reasons for their initial assessments. After the second round, researcher synthesises the group opinion and forms the third questionnaire based on it, which is again presented to the group. Several rounds of questionnaires can be performed but finally the group opinion is expected to reach a consensus.

The Delphi methodology by no means provides statistically significant results but its strength lies in the ideas it generates whether or not a consensus is reached. It is difficult to perform Delphi studies well and a great deal of time is required for executing multiple rounds of questionnaires. A lot of effort needs to be put in the selection of experts and designing the questionnaire.

Environmental scanning is a general monitoring procedure that is involved in all futures research in one way or another. Plans for future are typically based on forecasts and forecasts are based on assumptions. Environmental scanning systems provide information on future threats and opportunities as well as make it possible to detect weak signals or get early warnings about important changes.

Environmental scanning is based on defining a set of sources from which the information on future is acquired. This information is then analysed and synthesised by a scanning team that forwards the conclusions to decision makers. Sources used in environmental scanning may include expert panels (e.g. Delphi), database literature reviews, web based alert systems (e.g. Google Alerts or web crawlers), websites, electronic or hard-copy literature reviews, essays by experts or key person tracking. Requirements for the environmental scanning team are set by the decision makers and they also give feedback to the team for adjusting the scanning process and set of sources.

The work in ObservatoryNANO project can be considered environmental scanning as it combines information from multiple sources including literature, websites, experts as well as scientific publication and patents databases. On the other hand, reports produced by ObservatoryNANO members could be used as part of a larger environmental scanning scheme.

Robust decisionmaking (RDM) is essentially what this study is all about. In the context of futures research, RDM is a theoretic framework aiming to make systematic use of a large number of highly imperfect forecasts of the future. Instead of probabilistic predictions, RDM tries to produce a best possible representation of the future using the information that is available. Robustness means that the predictions given by RDM promise to do a reasonable job of achieving the decision makers' goals.

The methodology has been applied in a wide variety of decision problems including defence, climate change and science and technology planning. In summary, "RDM is an iterative, quantitative approach for identifying decision strategies whose good performance is relatively insensitive to key uncertainties facing decision makers and characterising the residual vulnerabilities of these strategies."

Technology Sequence Analysis (TSA) is a method for forecasting the time of technology dependent system becomes available. This system based approach somewhat resembles the TRL scheme used by NASA. The main difference is that TRL scheme used by NASA assigns a TRL to all components of a system in order to assess the technological risks related to the whole system. In TSA, the estimates of time required to achieve intermediate technology steps are combined statistically to produce an estimate when the full system could become available. The TSA is a trademark of The Futures Group and therefore the software is not available to the public.

In TSA a model network of subsystems, components and related technologies is built representing the structure of the system. The network uses boolean logic for relationships (paths) between subsystems and technologies, which means that a working subsystem may require multiple mature technologies (AND gate) or only one technology might be required (OR gate). In case of a harvesting robot, a fully working system needs guidance technology, position sensing, ripeness sensing, cleaning technology and packaging technology. If one of these subsystems is missing, the robot is not functional. On the other hand, there is only need for one ripeness sensing technology, which could be one of hormonal content sensor, color sensor or odour sensor.

A typical TSA network may consist of hundreds of nodes and hundreds of paths of either "AND" or "OR" relationship. The analysis is performed using a Monte Carlo method, which produces a probabilistic estimate of availability of intermediate technologies and components as well as the final system. TSA can be used for estimating the costs related to use of different technologies, identifying technological risks to be considered in R&D programmes and for estimating uncertainties in the development schedule. The TRL estimates produced by the methodology of this study could be used as inputs for TSA. The main disadvantage, however, is that acquiring large amount of data for building a network of hundreds of nodes and paths is very costly.

Decision modeling is a very general framework that attempts to model the human behaviour in decision making. It is based on identification of specific criteria for the decision task and assessing how well competing technologies meet those criteria. Each criteria is weighed by its relative importance and each technology is given a score per criteria. In case of technology assessment, decision modeling can be used for estimating the market potential by comparing new technologies to current market requirements. Based on the criteria weightings and item scores, a compound score can be calculated for each technology.

The main weaknesses of decision modeling are related to the modeling process itself. First of all, identifying the criteria and their weights is not an easy task and psychological factors distort the selection of what is important. A lot of information is also needed for establishing the criteria and their weights. Moreover, the perceptions of decision makers and the markets or customers change over time and therefore both the criteria and weights need to be evaluated again at a later point of time.

The main strength of decision modeling is that market research data can easily be used as an input. Both the decision criteria and their weights can be determined by customer feedback acquired through a questionnaire. This fact also promotes the use of questionnaires as a tool for this study. One practical example of decision modeling is the Analytic Hierarchy Process (AHP), which models the problem in hand as a hierarchy structuring decision into smaller parts. These parts can include social, political, technical and economic factors. The decision makers then perform the assessment by making simple pairwise comparisons on the lowest level of hierarchy, which result in the final high level assessment done by the software used. Using a light hierarchy for both TRL and technology impact is considered later on in this study.

Patent analysis, publication analysis and text mining Patent and publication analysis along with text mining share similar targets with methodology considered in this study. They all aim at providing information on technology readiness by well defined processes and report their findings in compact form, for instance in statistical tables. Patent and publication analysis offer a simple temporal perspective for technology readiness by measuring the number of patents and publications on a certain technology area during a period of time.

Publication count data is a purely quantitative measure of scientific production but its problem is that it does not capture scientific progress very well. Patent data can be better used for measuring innovation but its main flaw is that every technological innovation does not receive a patent. Martin and Daim (2007). Despite their flaws, they are a good addition to any TRA process because they can be included in the process by relatively simple means and as the outputs are numerical, they can be used as independent variables in regression models. Moreover, patent and publication data can be used to produce time series on how the R&D efforts on a certain technology have developed over time.

Text mining, or tech mining, is a bibliometric method that goes deeper than pure patent or publication analysis by taking into account the content of those documents as well. Text mining for TRA can also include data sources such as news articles, commercial documents released by companies or even blogs from experts dedicated to follow certain technology areas. The information sources for text mining can be divided in ones tracking developments in fundamental research (e.g. publication databases) and others in applied research (engineering oriented publications), invention (patent databases) and commercial application (product databases and marketing data).

Text mining along with patent and publication analysis can be potentially a very powerful tool for TRA. However, it faces significant challenges in terms of costs of obtaining data, assessment of its credibility and the process for automating and using the results obtained. The quality of results achieved by text mining is also naturally limited by the quality of data sources used. Subscription fees to the best available patent and publication databases are very high, which limit the usability of statistical analysis for low intensity use. A practical implementation of text mining is presented in Britt et al. (2008), where automatic document classification is used for Technology Readiness Level analysis achieving accuracy of 86%.

1.3.3 Summary and integration of methods

This section has presented a number of methods for assessing technology future and technology readiness. In conclusion, a complete TRA scheme would consist of a multitude of methods including environmental scanning, the use of expert opinion such as Delphi and quantitative methods in form of patent and publication analysis along with text mining. All these methods are situated in the bottom of figure 3, which states that they try to understand the technology readiness by standing outside the system.

Implementing statistical or decision making models moves the decision maker closer to the system and requires the decision making scheme to understand and model the interactions between different agents in the system. Technology sequence analysis, for instance, requires careful design when it comes to understanding the system structure and results in a model that defines interactions between components and subsystems, which correspond to the agents in the system. In summary, the use of decision modeling techniques ultimately means that the decision maker needs to understand the relations between data and the technology under assessment well. It is always easy to collect large amount of data using independent measures but the work becomes increasingly difficult when a conclusion needs to made.

Technology readiness and estimated impact of new technologies are arguably a very important part of predicting technology success. However, it is not only the technological side that will determine what are the technologies having largest impact in future. Galbraith et al. (2006) develop a model for estimating technology success including a number of factors such as: company size, age, R&D strategy, external funding, level of education of research team and formal partnerships. They achieve almost 50% goodness-of-fit with their model, where company age, technological development stage and amount of external funding (with a negative coefficient) are the most significant variables. Company age and amount of external funding can in most cases be determined quite easily. Therefore the assessment of technological development stage is crucial in predicting technology success.

1.4 Use of expert opinion in decision analysis

1.4.1 Decision analysis

The fundamental background of this thesis lies in the area of decision analysis. The object of decision analysis is to help a decision maker think hard about the specific problem at hand, including the overall structure of the problem as well as his or her preferences or beliefs. Decision analysis provides both an overall paradigm and a set of tools with which a decision maker can construct and analyse a model of a decision situation. The main idea of decision analysis is to be able to represent real-world problems using models that can be analysed to gain insight and understanding. The ultimate goal of decision analysis is then that the decisions can be improved using this insight and understanding gained (Clemen, 1996).

Different problems in decision making involve different and often case-specific difficulties. According to Clemen (1996), there are four basic sources of difficulty in decision analysis problems. First, the decision making problem can be hard because of its complexity, which means that there may be, for instance, multiple different sources of uncertainties, different possible courses of action and economic impacts to name a few. In complex problems, keeping all of the issues in mind at one time is nearly impossible and decision analysis aims to provide tools to structuring complex problems to make them possible to analyse. Second, decision making can be difficult due to uncertainties involved in the decision making. Uncertainties can be found in data as well as in the projected consequences of the decision. Therefore, the decision maker can never be sure about the basis of his decision nor the impact of his decision.

Third is the trade-off between multiple objectives found in some decision making problems. In these cases, the decision maker must trade off benefits in one area against costs in others. Typical trade-off situations include economic versus environmental effects in environmental decision making as well as expected return and riskiness in investment decisions. Finally, different approaches in decision making may often lead to different conclusions and small changes in the input data may lead to different choices and decisions. Different views of individuals are also a challenge in decision making as they may disagree on the uncertainty or value of the various outcomes. The use of decision making tools aims to resolve these differences whether the decision maker is an individual or a group of stakeholders with diverse opinions.

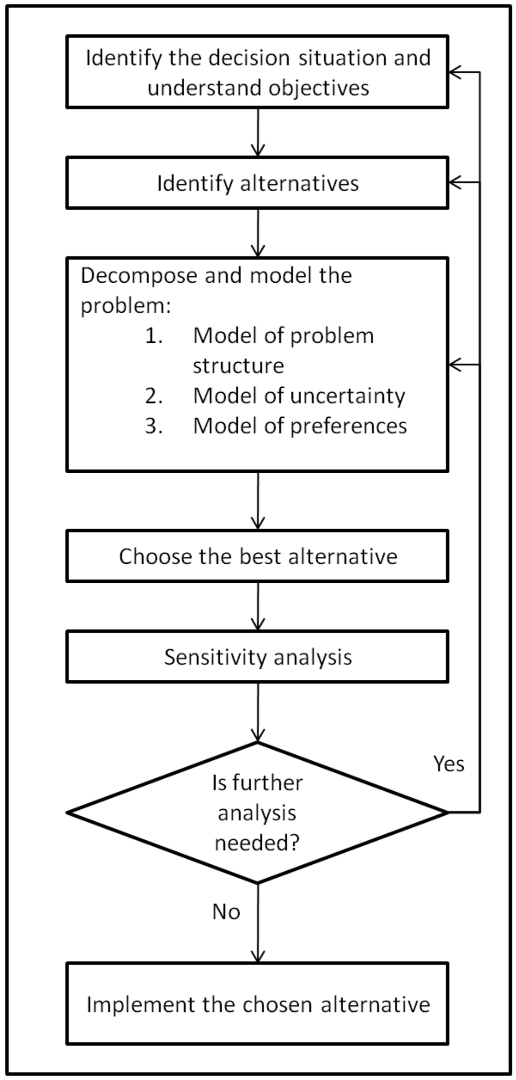


Figure 4: A decision-analysis process flowchart adapted from (Clemen, 1996). The process is followed in the structure of this study.

The advantage of decision analysis is to make complex problems easier to understand. Clemen (1996) presents a flowchart (Fig. 4) that defines the decision analysis process in simple steps that are: identifying the problem, identifying different alternatives for decision making, decomposing and modeling the problem, choosing the best decision alternative and inspecting its sensitivity and finally implementing the best alternative. The decision situation and alternatives were discussed earlier in the introduction section. Modeling the problem, uncertainties and preferences are presented in the following sections.

1.4.2 Expert use in decision analysis

Decision making problems often include a significant amount of uncertainty. As stated earlier, one elemental part of decision making is modeling of the problem, which requires the definition of data to be used. TRL is an artificial concept created to help decision making in various situations and therefore there is no fundamentally clear set of data that can be used for assessing it. There are, of course, publicly available databases such as patent or scientific publication databases but using them for assessing the Technology Readiness Level of a certain technology would require a significant amount of effort. Moreover, data retrieved from patent or publication databases can in general cases be included in technology readiness assessment as a qualitative source, which is taken into account by the decision maker.

The fact that quantitative data is in practice not available for technology readiness assessment gives rise to the use of expert opinion in decision making. Expert opinion is often used in real world decision making problems that include a significant amount of uncertainty (Morris, 1974). The uncertainty can be attempted to resolve by relying upon the judgement of one or more other persons referred as experts. An expert can be defined simply as a person who provides a judgement concerning uncertain matters (Morris, 1974).

Another view for using expert opinion in technology readiness assessment is that experts can be considered to reflect the state of the TRL of a certain technology because they have a large amount of knowledge on the contents of relevant information, such as patents and scientific publications. Similar grounds are also discussed in Daneshkhah (2004) who suggest that expert judgement is appropriate when data are sparse or difficult to obtain, data are too costly to obtain, data are open to different interpretations or there is need to perform an initial screening of problems.

In most situations requiring decision making under uncertainty, there is no absolutely correct answer to the question involved. Moreover, consulting multiple persons often produces contradictory opinions which make the decision making process difficult despite the increased amount of information received from the experts. Johnson and Albert (2001) compare combining of multiple expert opinions to having more than one wristwatch. "With one watch, you know the time – or at least you think you do. But with two watches, you are never sure."

There are several ways for obtaining expert opinion including mail/email surveys, expert interviews, expert meetings, interactive groups and the Delphi. Each of the situations have both advantages and disadvantages extensively discussed in Meyer and Booker (1991). In this study, only email questionnaires are used even though the process designed is not limited to using any particular form of communication and it is also suitable for a Delphi study. The mail surveys are good for eliciting simple data from a large sample (Meyer and Booker, 1991). The disadvantage is naturally the lack of communication with the expert and typically low response rates. Face-to-face communication and telephone on the other hand requires much more time and effort but results in larger amount of qualitative data and observations.

Obviously, no single best method exists and email questionnaire is a good starting point for TRA because of its low cost nature. Individual interviews should be used if interaction between the analyst and expert is required and meetings with a group of experts can be established if expert to expert interaction is considered helpful. These can also be achieved by interactive group sessions or workshops. The Delpi, as presented earlier, tries to combine advantages of questionnaires and group work and as such can be considered as a natural extension to questionnaire based TRA.

1.4.3 Assessment of uncertainty

The mathematical way for measuring, quantifying and modeling uncertainty is probability. Mathematical rules of probability are based on a simple set of axioms and they are well understood and noncontroversial. However, the interpretation of probability for modeling uncertainties is controversial and so are different attempts for trying to distinct between types of uncertainty. (Winkler, 1996) The types of uncertainty can be divided in aleatory and epistemic (i.e. reducible and irreducible or stochastic and subjective) uncertainties, and even though for instance Winkler does not consider this kind of classifications useful they provide an insight into the sources of uncertainty in probabilistic models.

Aleatory uncertainty is a type of uncertainty that cannot be expected to reduce because it comes from natural, unpredictable variation in the system under study (Daneshkhah, 2004). In modeling, there is no way one could obtain information that could be used for reducing aleatory uncertainty because of physical or even economic reasons. Winkler (1996) gives an example of tossing a coin where, at first, the uncertainty about how a fair coin lands can be thought as aleatory. However, if one knew all the conditions surrounding the toss one could use the laws of physics for predicting the outcome of the toss. Measuring all the initial conditions is in practice infeasible and therefore there is always some aleatory uncertainty included in the process of predicting the outcome of the toss of a fair coin.

Epistemic uncertainty originates from the lack of knowledge about the system under study and can be conceptually resolved (Daneshkhah, 2004). Epistemic uncertainty can be decreased or eliminated by collecting a sufficient amount of information from the system under study. In practice decreasing epistemic uncertainty could mean observing the system for a longer period of time or by collecting data from as many sources (e.g. experts) as possible. The fundamental reason Winkler (1996) does not consider the distinctions between types of uncertainty fruitful is that there is always uncertainty in the final probabilistic model and it is more important to improve the modeling and analysis part, including data collection, rather than focusing on the semantics.

Instead of using aleatory and epistemic uncertainties, Winkler (1996) argues that it can be useful to think about certain types of uncertainties that are related to the modeling process itself. First of all, he distinguishes between uncertainties about events or variables that are observable and uncertainties that are not. Outside temperature can be considered as an observable variable that can be measured to find out how good was the model used for forecasting tomorrow's temperature. There may also be some unobservable variables in the model trying to forecast the temperature and these parameters might not be easy to understand and they might not have an intuitively reasonable interpretation. Winkler (1996) furthermore points out that it is easier for experts to think about their subjective probabilities for observable quantities than about unobservable ones such as parameters in the probabilistic model.

This reasoning has an important role in this context. TRLs are an abstract concept and clearly unobservable variables as such. Therefore asking experts to try to assess the TRL straight away is a difficult task. Instead, a decomposition of the TRL assessment problem into smaller subproblems makes it easier for experts to express their opinion (Kynn, 2008). The discussion of question setting for TRL assessment is continued in later parts of this study.

1.4.4 Sources of bias

In addition to uncertainty, every decision analysis problem includes several sources of bias, which can be defined as skewing of results in an unfavourable direction. Based on this, one can say that a result in decision making that is not biased is reality or truth (Meyer and Booker, 1991). Bias is a common term in statistics, where estimates that are biased are generally considered harmful. Biases can occur in many forms and from many sources but in using expert opinion, biases originate from human decision making processes where the experts interpret reality through their previous experiences or mental models. It is important to take the effect of bias into account in modeling by identifying different sources of bias and controlling bias properly when modeling the problem and interpreting the results.

The sources of bias can be classified in two categories: motivational bias and cognitive bias (Meyer and Booker, 1991; Clemen, 1996). Motivational bias typically originates from other people or factors affecting experts thinking and results in experts not reporting their actual opinions. When other people, such as the interviewer or other experts, are involved when expert is expressing her opinion there tends to be social pressure for answering what people expect the expert to answer. Moreover, in group working situations, the group may steer towards a consensus which increases the bias of expert opinion due to an effect called group think. Sometimes, an experts wants to answer in a way that is favourable to them, which is called the wishful thinking effect. Wishful thinking might occur for instance when the expert receives money from assessing the results of her own work. Another good example of motivational bias is a salesperson doing the sales forecast for future. The person may be tempted to estimate the sales lower than the actual expectation to make him look good when results are assessed (Clemen, 1996).

Motivational bias is also included in the decision maker side in the form of misinterpretation. Human interpretation of incoming information is selective and decision makers tend to trust information that supports their original views (Meyer and Booker, 1991). In an interview situation, for instance, the analyst is already somewhat familiar with the expert's field and therefore interprets the expert opinion based on her current knowledge on the subject. Misinterpretation bias can also originate from problem modeling. The decision maker often encodes expert opinion and forms a mathematical model to combine collected data and is subject to introducing misinterpretation bias when making subjective assumptions in the modeling phase.

Cognitive biases originate from human mind not being able to process and remember all information available. People tend to make solving of complex problems easier by taking short cuts to reduce the cognitive burden (Meyer and Booker, 1991). Forming an opinion starts with the first impression which is then adjusted slightly based on new information received and therefore subsequent information is not used as much as the first impression. This effect is called anchoring bias referring to anchoring the opinion on the first impression. In case of the salesperson, it is likely that sales forecasts are based on past results rather than the person really trying to predict what factors affect future sales. The salesperson is, on other words, anchoring to past results (Clemen, 1996).

Other sources of cognitive bias include inconsistency bias that deals with humans forgetting their assumptions and therefore being inconsistent in the problem solving process. Some information is also easier to remember than other, which results in availability bias. This means that people tend to overestimate frequencies of events based on familiar, concrete or recent events. Finally, people are poor in estimating uncertainties and probabilities in general.

1.4.5 Summary and implications

The previous section makes many relevant points with important implications regarding the expert elicitation process. The decision analysis process flowchart by Clemen (1996) (Fig. 4) can be used in defining the TRA scheme in this study. Assessing TRL of competing technologies is a task where the real answer cannot be derived from any traditional information source because the TRL is a very abstract concept. Therefore, TRL assessment needs to be done by either qualitative research conducted by the analyst or by expert elicitation that can be either qualitative or quantitative.

There are several alternatives for data collection, analysis and modeling that will be discussed later on in this study. Biases can be effectively reduced by using multiple experts instead of one (Kuhnert et al., 2009) but choosing the experts is non trivial when many of them are needed for the assessment. Daneshkhah (2004) suggestion for general principles that should be applied in expert elicitation include noteworthy points in using expert elicitation, expert selection, modeling and dealing with uncertainties. As stated before, expert elicitation is appropriate when data are sparse, costly to obtain, subject to different interpretations or an initial screening of the problem is required. In risk assessment, experts are used to either structure the problem or provide estimates, which is the case in this thesis.

In the analysis process, only an expert's opinion is worth eliciting and therefore effort needs to be put in the selection of experts. The questions asked should consider observable variables to make experts more comfortable with their task. Significant attention must be paid on reducing or at least accounting for expert biases and uncertainties. In order to do this, the elicitation process should involve feedback and an uncertainty model to deal with e.g. overconfidence or anchoring (Daneshkhah, 2004).

Both Daneshkhah (2004) and Kuhnert et al. (2009) also point out that the use of multiple experts should be taken into account in models. The experts are and should be dependent at least to some extent. Modeling of dependencies has been dealt widely in literature but it is left out from the scope of this study because dependency considerations would significantly increase the complexity of the models. Calibration is a technique used in both single-expert and multi-expert models to ensure that the expert opinion is interpreted correctly. Similarly to dependency modeling, there is no clear definition of correct calibration of the expert and therefore it is not considered in this study either.

1.5 Expert opinion aggregation

1.5.1 Measuring and modeling expert opinion

The format for collecting expert opinion plays a big role in expert opinion aggregation. A number of data acquisition techniques are suggested by Meyer and Booker (1991). All techniques have their advantages and disadvantages, which are first dealt in this section and later relevant methodology related to each type of data is presented.

- Probability estimates and probability distributions
- Odds ratio
- Continuous scales
- Pairwise comparisons
- Ranks or ratings
- Estimate of physical quantity
- Bayesian updating

Probability estimates and probability distributions are very commonly used in decision analysis because they allow flexible use of statistical methods by nature. Their main disadvantage is that most people are not good at estimating probabilities and estimating probabilities or probability distributions is very fatiguing and time-consuming to the expert (Meyer and Booker, 1991). Mathematical aggregation of probability distributions reflecting expert opinion has been widely studied in literature especially for risk analysis purposes (Clemen and Winkler, 1999). The experts may also be asked to assess quantiles instead of full probability distributions, which makes the assessment simpler. This approach is presented, for instance, by Garthwaite and Dickey (1985).

Other common expert opinion acquisition methods are different sort of scales. Continuous scales have continuous number lines and the experts mark their answers in between the extreme values of the scale. Continuous scales might additionally be labeled with integers, text, probabilities or categories. Their main advantages include ease of use for the expert and simple mathematical representation. According to Meyer and Booker (1991) they may be, difficult to develop and care must be taken to guard against biased wording of the labels or their definitions.

Ranks, ratings and pairwise comparisons are among the easiest for the expert to perform (Meyer and Booker, 1991). Pairwise comparison is about the expert comparing two objects at a time. People are generally good at estimating pairwise comparisons but the disadvantage is that pairwise comparisons are time consuming when the amount of objects is large. Pairwise comparisons also provide only relative differences between objects compared and no absolute quantities, which might be an issue for some analysis. This problem is addressed with ranks and ratings, which involve assigning numbers or descriptions to the objects in question. The descriptions can be used to provide the baseline scale required for absolute assessment of the object. The main disadvantage of ranks and ratings is that the difference between different ratings is not defined and therefore a numerical representation of results is not mathematically valid.

The acquisition techniques are closely related to the classification of scales of measurement presented by Stevens (1946). The type of data collected has important implications on the mathematical and statistical methods allowed for manipulation of the data. The scales can be either nominal, ordinal, interval or ratio and they are presented in table 2.

The nominal scale is the most unrestricted in terms of assignment of numerals but is the most restrictive in terms of mathematical operations that can be performed on them. The numbers assigned on nominal data are only used as labels in identification and they have no mathematical meaning. The ordinal scale includes a rank-ordering and it is widely used in, for instance, psychology. Because of the rank ordering, means and standard deviations are strictly speaking prohibited statistics when dealing with ordinal data.

For instance, the difference between an expert agreeing strongly versus agreeing moderately cannot be explicitly defined to have the same distance as the difference between the expert agreeing moderately versus having a neutral opinion. The use

Table 2: Classification of scales of measurement (Stevens, 1946)

Scale	Basic Empirical Operations	Permissible Statistics
Nominal	Determination of equality	Number of cases Mode Contingency correlation
Ordinal	Determination of greater or less	Median Percentiles
Interval	Determination of equality of intervals or differences	Mean Standard deviation Rank-order correlation Product-moment correlation
Ratio	Determination of equality of ratios	Coefficient of variation

of means and standard deviations would imply knowledge of something more than the relative rank-order of data and therefore they do not typically have a statistical meaning for ordinal data. However, Stevens (1946) among many other researchers have argued that using this kind of illegal statistical operations leads to fruitful results in numerous instances.

As suggested in table 2, interval and ratio data are the most flexible from analysis and modeling point of view. They are what is normally understood by the word quantitative. The main difference between the two scales is that in case of interval data, the zero point is a result of agreement whereas for ratio data there exists a true zero point. Most scales in physics are ratio scales and constructing either ratio or interval scales is often tried in psychometrics but quantifying, for instance, an expert opinion leads to scales where the definition for the zero point is usually not clear. This thesis later focuses in obtaining expert opinion by ordinal scales, interval and ratio data are not considered need not be considered any further.

As discussed before, expert opinion aggregation methods have to deal with many different sources of uncertainty and bias. The most flexible model for TRLs would be modeling them as a probability distribution of reaching a certain level. In this kind of setting, the expert attempts to estimate the probability distribution of the TRL with lower bound in TRL not reached at all and upper bound in TRL fully reached. If the expert is very certain about his opinion, he places most of the probability mass on a very short interval and if he is very uncertain about the TRL he will give a distribution resembling the uniform distribution. This means that the variance (σ_i) of the expert opinion, regarding estimated item i, is low when the expert is certain about his opinion and high when expert is uncertain. The aggregation of expert probability distributions then produces one probability distribution for the TRL based on assessments of multiple experts. In this setting, it is comfortable for the aggregation method to report how much of probability mass is distributed close to "TRL fully reached".

Another aspect of uncertainty is related to the expertise level of experts. A university professor studying ink-jet printing is likely to be more aware of TRL of ink-jet printing than a junior researcher studying nanoimprint lithography. From the TRL assessment point of view, it is still important to get both to give their TRL estimate for both technologies given that they both are familiar with each others field of research. Otherwise, it would be very difficult to perform pairwise comparisons between TRL estimates of nanoimprint lithography and ink-jet printing. This aspect can be taken into account by including the level of expertise in the model. The level of expertise tries to assess how likely the expert is to correctly assess different TRL items. Mathematically this means that the variance (σ_j) of an expert j is lower for more experienced experts and higher for more junior ones. More examples of including expert uncertainty are given along with introduction of models.

The methodology for combining expert opinion can be classified in two categories: mathematical and behavioural approaches. Furthermore, mathematical approaches can be further classified in three groups: non-Bayesian axiomatic models, Bayesian models and psychological scaling models (Ouchi, 2004). Various methods for combining expert opinion in risk analysis are reviewed by Clemen and Winkler (1999) and Ouchi (2004). Combination of expert opinion is a popular subject in risk analysis because in many cases "hard" quantitative data is not available for decision making and the experts can provide valuable information in the decision making process. Application areas of using expert judgement include environmental risk assessment, military intelligence, nuclear engineering and various types of forecasting (economic, technological, etc.) among many others.

1.5.2 Axiomatic and Bayesian models

Axiomatic approaches are the simplest methods for combining probability distributions. In axiomatic approaches, certain properties and regularity conditions are established for combining probability distributions (Ouchi, 2004). The simplest axiomatic combination schemes are opinion pools that can be either linear or logarithmic. A linear opinion pool corresponds to the calculation of a simple weighted arithmetic mean whereas the logarithmic opinion pool corresponds to weighted geometric mean. The definitions of these opinion pools are in equations 1 and 2, respectively. The problem with opinion pools is that the weights need to be set subjectively case by case. Typical interpretation of the weights is "the better an expert, the heavier the weight"(Morris, 1977), which leads to a problem of measuring and balancing between expert goodness.

$$p(\theta) = \sum_i w_i p_i(\theta) \tag{1}$$

$$p(\theta) = k \prod_i p_i(\theta)^{w_i} \tag{2}$$

The basic Bayesian approach for expert opinion combination allows much more flexibility than the axiomatic models. A model developed by Morris (1977) is illustrated in equation 3. Opinions f_n from multiple experts about interesting quantity x are combined in the likelihood function and the decision maker can affect the process by choosing the model for the likelihood function $p(f_1, \ldots, f_n|x, d)$ as well as by including his own prior beliefs d in the prior density function $p(x|d)$.

$$p(x|f_1, \ldots, f_n, d) \propto p(f_1, \ldots, f_n|x, d)p(x|d) \tag{3}$$

Most of the early work, in 1970's and 1980's, on using Bayesian aggregation schemes has focused on using independent normal distributions for experts' assessments (Ouchi, 2004). The assumption of independence between experts' probability distributions makes it easy to form the likelihood function because independence implies $p(f_1, \ldots, f_n|x, d) = \prod_i p(f_i|x, d)$. Independence between experts is, however, a property difficult to satisfy because they are assessing the same object x and therefore using overlapping sets of information for making their conclusion on the probability distribution. In case of normal model for data, the dependencies can be modelled simply in a covariance matrix (Clemen, 1987). One can argue that dependencies between experts are by no means linear and their modelling would be a very difficult task for example in the setting of this study and therefore they are not considered further.

A more recent study by Lipscomb et al. (1998) develops a hierarchical Bayesian model for physician staffing. The hierarchical model means that a set of hyperparameters are used for estimating probability distributions for real parameters. Hierarchical modeling approach makes it possible to take into account dependencies among the experts as well as each expert's ability to accurately estimate the time required for performing certain medical procedures. This approach essentially decreases the sensitivity of the model to small variations of data, which is also a favourable requirements for this study, too. Similar problems of the one handled in Lipscomb et al. (1998) have been typically solved by Delphi processes and the authors discuss it could be possible to replace the interactive Delphi process by using the simple hierarchical approach that produces statistically interpretable results. The main advantage in Delphi process is that errors and misunderstandings can be easily detected and corrected during the interactive process.

Most of the traditional theory on Bayesian models does not apply for this study because of the type of data collected. The traditional Bayesian settings expect the data in the form of probability distributions, either full or quantile based. As the data used for this study is ordinal, there are some fundamental limitations on mathematical operations that can be applied (Table 2). Most of the models designed for ordinal data use logits of cumulative probabilities also called cumulative logits. They are, however, often intended for regression problems where the goal is to estimate the behaviour of a dependent variable by one or more independent variables. For an extensive review of ordinal regression models for different purposes, see Liu and Agresti (2005).

1.5.3 Measures for rater agreement

Measurement System Analysis (MSA) studies the quality of measurements and ways to improve their usefulness, accuracy, precision and meaningfulness. Precision of measurement system means a different thing for different users. In industry, most emphasis is on measurement spread whereas psychometrics focuses on reliability. When analysing ordinal data, the goal is often to draw one concluding figure of the data available and therefore measurements for the rate of agreement among data are needed. Mast and Wieringen (2004) present the concept of intraclass correlation coefficient (ICC) and Kappa that are commonly used for measuring reliability and agreement and used as a background for more advanced latent variable models (e.g. in (Johnson and Albert, 2001)) that are presented later in this section.

In ICC the observations X_{ij} are assumed to follow the model

$$X_{ij} = Z_i + \epsilon_{ij}, \tag{4}$$

where $Z_i \sim N(\mu_p, \sigma_p^2)$ denotes the true value of an item i and $\epsilon_{ij} \sim N(0, \sigma_e^2)$ the measurement error. Index j denotes the different measurements, e.g. experts, that have rated item i. ICC (Eq. 5) is then defined as the correlation between different measurements X_{ij} of an object i.

$$ICC = \frac{Cov(X_{ij}, X_{ik})}{\sqrt{Var(X_{ij})Var(X_{ik})}} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} \tag{5}$$

By equation 5, ICC is defined as the ratio between variance of interest over the total variance.

Kappa is a measure used for evaluating rater agreement typically for nominal scales. It is a measure of agreement corrected for agreement by chance. The definition of Kappa is

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \tag{6}$$

In the equation (6), P_o denotes the observed proportion of agreement and P_e the expected proportion of agreement, i.e. the agreement by change. The basic version of Kappa is not intended for multivariate data and the theory has been developed further for solving this issue. ICC and Kappa play an important role in development of latent variable models used in this study but the multivariate version is out of the scope of this study.

1.5.4 Latent variable models for ordinal data

One possible view to ordinal data is to understand the data as a representation of an underlying latent (i.e. unobserved) variable associated with each response. This is somewhat against the strict statistical requirements set in table 2 but offers greater flexibility in modeling. In the latent variable approach, the data is assumed to be drawn from a continuous distribution centred on a mean value that is different for each individual respondent. Figure 5 represents the latent variable approach in case of ordinal data with four categories A-D. Each expert expresses his opinion of the latent variable by one of the categories. The data is then combined to form a latent probability distribution that is assumed to be behind the expert opinion. Vertical lines in the figure denote the category cutoff points.

In latent variable models (e.g Johnson and Albert (2001)), the true value of the measured property is denoted by $Z \in \mathbb{R}$. The quantity $t = Z + \epsilon$ denotes the perception of the latent trait on the underlying trait scale where ϵ denotes the measurement error. The latent variable model is linked to the ordinal scale used according to equation 7. Measured variable is assigned the category c if it satisfies the condition in the equation. The category cutoffs are denoted by γ_j, where $\gamma_0 = -\infty$, $\gamma_K = \infty$ and $\gamma_j = \{\gamma_1, \ldots, \gamma_{K-1}\}$. The number of categories is denoted by K.

$$\gamma_{c-1} < t \leq \gamma_c \tag{7}$$

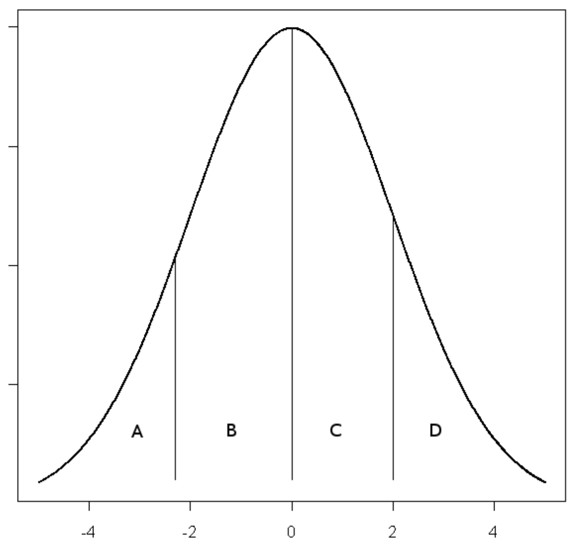


Figure 5: Normal latent trait for four category ordinal data. Numbered x axis denotes the true value of the measured variable and the letters between vertical lines (cutoff points) mark the ordinal categories.

Three latent variable models are considered in more detail in this study. Their features and applications are briefly presented here and more thoroughly in section 2. All models presented here have a different approach and applications but they are also closely linked with each other and the subject of this study regarding the fundamental problem of finding out the underlying variable from a dataset. More models for ordinal data, especially for regression purposes, are presented in Liu and Agresti (2005).

Mast and Wieringen (2004) base their model on the ICC, which is modified for ordinal data. They use a link function similar to one used in logistic regression for mapping from the ordinal scale to real numbers. Estimates for latent parameters are calculated by optimisation of a function that combines together multiple items rated by multiple experts. In addition to latent variables, there is one common variance parameter σ_e^2 in the model, which denotes the stochastic error in the data. Mast and Wieringen (2004) have applied their model on a dataset from visual inspection of printer assembly data. The inspection is done by 3 different raters and the authors find out that visual inspection is completely inadequate for testing of the printing quality.

Latent variable model by Johnson and Albert (2001) is an attempt to develop a consistent measure for essay grade data in case of multiple raters. Their model includes the use of judge specific category cutoffs as well as judge specific rater precision. This approach offers much greater flexibility compared to the model by Mast and Wieringen (2004), which only has population specific link function and variance. Due to the complexity from judge specific category cutoffs, precision variables and their prior distributions, Monte Carlo sampling is required which greatly increases the time required to run the model.

Yet another latent variable model for rating data is suggested by Ho and Quinn (2008). They derive their model from Johnson and Albert (2001) as a special case with population wide category cutoffs but introduce additional parameters for finding out how positive and discriminating the raters are. The model is intended to be used in online rating systems, e.g. ones used in Youtube or iTunes, where the rating system has no prior knowledge on how consistent or "good" the raters are.

1.5.5 Subjective logic

Subjective logic is a form of logic which operates on subjective beliefs about the state of the world. Subjective logic is closely related to belief theory originally developed by Dempster and Shafer (Jøsang, 2010). This work is called Dempster-Shafer theory of evidence and it provides means for representing and working with subjective beliefs. In standard logic used in computing, propositions can be either true or false whereas the belief theory takes uncertainty into account, too. The theory behind subjective logic and its relation to Dempster-Shafer theory of evidence is thoroughly presented in Jøsang (2010) and Jøsang (2001).

Subjective logic provides an interesting framework for combining subjective opinions from multiple experts even though the theory behind both Dempster-Shafer theory of evidence and subjective logic is slightly controversial. Pearl (1990), for instance, states that belief functions should only be used in analysis of deterministic systems, such as electronic circuits. There are many advantages in subjective logic for purposes of combining expert opinion and therefore it is reasonable to assess its performance. This section presents the fundamentals of subjective logic necessary to be utilised for the purpose of this study.

A key concept in Dempster-Shafer belief model is a set of possible situations which is called the frame of discernment. The frame consists a set of possible states and exactly one of these states is assumed to be true at any given time. Figure 6 illustrates a simple example case of frame of discernment Θ, which contains four elementary states $x_1, x_2, x_3, x_4 \in \Theta$. The powerset of Θ, denoted by 2^Θ, contains all atomic states and all their possible unions along with Θ itself. Any combination of these elementary states therefore belongs in the powerset: $x \in 2^\Theta$.

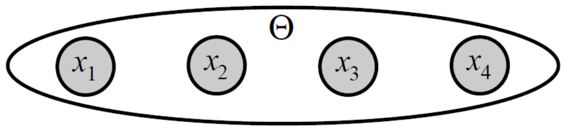


Figure 6: Example of a frame of discernment (Jøsang and Knapskog, 1998). Each x_i presents an elementary state for which the belief, disbelief and uncertainty is defined by 9. Elementary states can also be overlapping and subjective logic provides operators for computing the opinion tuple for any combination (union, intersection, etc.) of the elementary states.

Without going too much into detail of the belief function theory, the belief mass of the whole system is distributed in belief (b), disbelief (d) and uncertainty (u) functions according to equation 8. The subjective opinion can be then defined by a tuple ω_x defined in equation 9. Term (a) is called relative atomicity, which is used for including the decision makers prior belief in the model. The opinion space can be visualised by an equilateral triangle illustrated in figure 7.

$$b(x) + d(x) + u(x) = 1, x \in 2^\Theta, x \neq \emptyset \tag{8}$$

$$\omega_x \equiv (b(x), d(x), u(x), a(x)) \tag{9}$$

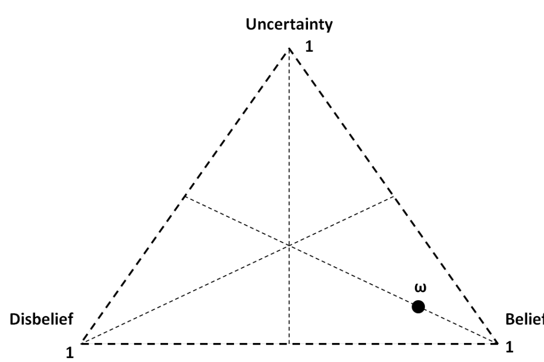


Figure 7: Triangle showing the opinion space in subjective logic. Adapted from Jøsang and Knapskog (1998). Variable ω denotes the opinion defined in equation 9.

Subjective logic provides a flexible set of operators and functions for mathematical processing of expert opinion. These include addition and subtraction, multiplication and division, deduction and abduction as well as fusion of opinions. Subjective logic, by definition, makes it possible to include expert uncertainties in the data collection phase which is very favourable for expert opinion aggregation. More detailed explanation of applying subjective logic is presented in section 2.

1.5.6 Other methods

Traditional statistics offer a multitude of methods more or less suitable for analysis of ordinal data. The most advanced methods in traditional statistics, such as variance analysis are not to be used for ordinal data and therefore one needs to rely on non-parametric methodology. Non-parametric statistical tests permissible for ordinal data include Mann-Whitney U test, Wilcoxon matched pairs tests, Kruskal-Wallis tests and Friedman's test. Contingency tables are also a popular method for presenting categorical and original data in a nonparametric way but they do not offer the possibility for combining expert opinion especially needed in this study.

Nonparametric tests, as well as other traditional statistics, are best suited for relative comparisons between groups. This kind of information is helpful but does not solve the problem for deriving a point estimate for decision making. Therefore, traditional statistics methods are left out of the scope of this study.

Elicitator (James et al., 2010a,b) is a recent tool for eliciting expert opinion from one or multiple experts. Elicitator is based on Bayesian framework using logistics regression and beta distributions for deriving the final expert opinion on the measured quantity. The expert is presented a number of covariates and he is to assess the probability of success for several cases based on these covariates.

A very practical approach for analysing ordinal questionnaire data has been taken by Hassall (1999), who uses so called triangular fuzzy numbers for combination of multiple ordinal values. Triangular numbers include a centre point in the original ordinal value and minimum and maximum artificially set by the analyst. In case of ordinal value of 3, the minimum could be 2 and maximum 4. The fuzzy approach offers the advantage of mapping the numerical results with linguistic values of the scale. However, the approach means in practice no more than computing a weighted average of the results and does not offer any added value for this work.

2 Method

2.1 Introduction to methodology

This section describes the methodology of the TRL assessment scheme that is developed as part of this thesis. The section is divided in 5 subsections covering the requirements, data collection and analysis, modeling, estimation of reliability and practical implementation. The subsections are described in different levels of detail reflecting their importance of the thesis.

The requirements specification subsection explicitly specifies the requirements set broadly in the general introduction of this thesis. It combines the theoretical background from literature review with practical requirements from the Observatory-NANO project. The main high level requirements for the method are economic approach, robustness and reliability of the model as well as practical implementation. The analysis tool produced as part of this study is finally evaluated against the requirements specified.

Data subsection introduces the collection, interpretation and analysis of TRL data used in this study. The approach to data collection is very practical due to the fact that there was no prior information on how one should collect expert opinion data for TRL assessment purposes from a group of experts all around the world. All the methods chosen are suitable for analysing ordinal data. Multinomial models were chosen because of their simplicity and they are used as a reference for more complex models. Subjective logic was chosen because of its simple structure and easy practical implementation. Latent variable models offer very good performance for some similar problems and therefore three different methods were selected for comparison.

Modeling subsection describes 5 different models for analysing expert opinion data. The models included are briefly described in literature review and more thorough mathematical description is given in this section. The models are then compared and assessed in section 3 based on the reliability measures described in this section. Reliability of the methodology is mainly discussed from sensitivity viewpoint but high level schemes for verification and validation are also introduced.

Finally, a practical implementation of the tool developed as part of this study is described. The implementation ties together data collection, its preprocessing, analysis and visualisation of TRL estimates.

2.2 Requirements specification

An important part of the tool development is the definition of requirements. As presented before, there are multiple different methods for conducting technology futures assessment or technology readiness assessment. Each of these methods have features that are advantageous for some purposes but disadvantageous for other purposes. First of all, there is a clear distinction between quantitative and qualitative methods. In the context of ObservatoryNANO project, most focus is on qualitative research and quantitative research methods should be designed so that they support the qualitative research efficiently and economically.

Most part of the research related to the work in ObservatoryNANO project is qualitative research based on interviews and text based sources. This part of the research cannot and should not be replaced by any means as it provides the readers grounds for their decision making. The role of quantitative research considered in this work is supportive. Quantitative results as such should not be the basis of any decision making but they can be and should be used to summarise the findings of the analyst work. Moreover, figures and estimates provided by the quantitative analysis can be used for efficient representation of the research findings in, for instance, presentations.

Another important aspect of the quantitative research is its reliability. The amount of data is seldom the same in case of different studies. Data collected from different sources for foresight purposes is typically contradictory and therefore the goal of the quantitative analysis is to give the best possible estimate given the data together with a reliability estimate of results.

The requirements presented here are summarised in the listing below.

- Data collection
  - Method shall collect data economically
  - Method shall collect data from a large group ($n \geq 100$) of relevant experts
  - Data collected shall include TRL estimates, technology impact estimates and estimates of when to reach a TRL
- Analysis
  - Method shall take into account the reliability of data entry
  - Method shall estimate the reliability of the results
  - Method shall be robust, i.e. not sensitive to small variations in data
- Results
  - Method shall express the results for each technology in two dimensions: TRL and technology impact
  - Method shall represent the results in visual form
  - Numeric results behind the visual results shall be available to the user
  - Numeric reliability estimates shall be available to the user

The main requirements for the methods are clearly contradictory with each other. On one hand, data should be collected from as many experts as possible but on the other hand there is only a limited number of relevant experts in the world. The quality of results will improve only to a certain point when the number of experts invited to participate is increased. At some point, inviting a larger number of experts does not improve the quality of the results because the quality of experts drops. Moreover, increasing the number of experts invited to participate increases the costs of conducting TRL assessment. Therefore, the number of experts that can be invited to participate in the study is limited by total number of relevant experts, the quality of experts and resources available for conducting the study.

2.3 Data

2.3.1 Collection and description

An integral part of the development of methodology for this study is collection of expert opinion data. There are several different methods for data collection suggested in futures research including interviews, surveys, Delphi and the use of expert reports and publications. The selected approach for this study is a web questionnaire because of its simple implementation and relatively low cost of the process.

The data collection procedure for collecting expert data is called sampling, which means selecting a subgroup of a population to be invited to participate in the study. A population can be defined as an aggregate of all the elements sharing a common set of characteristics and comprising the universe for the purpose of the decision making problem, e.g. technology assessment. Information about a population may be obtained by either sampling or by taking a census, which means involving the whole population in the study. Malhotra (2004) summarises the conditions when sampling should be used instead of census. Generally, sampling is more economic of two and census is only favourable when the population is small. In case of technology assessment, it is not possible to know every researcher or company working with a certain technology and therefore the census approach is not possible in practice.

Malhotra (2004) divides the sampling process in five steps: definition of population, determining the sample frame, selecting a sampling technique, determining the sample size and executing the sampling. Target population is the collection of elements, in this case experts, that possess information sought by the researches, which in this study means knowledge on technological development stage. Therefore the relevant expert population for technology assessment are all researchers and people employed by companies or other organisations that possess information about technological development stage of technology or technologies being assessed.

A sampling frame is a representation of the elements of the target population and consists a list or set of directions for identifying the target population. For instance, a sampling frame might be a telephone book from which the researcher chooses the numbers to be dialled. It is often possible to compile a list of population elements but using this kind of list may lead to sampling frame error, which should be taken into account when analysing the results (Malhotra, 2004).

Sampling techniques can be generally classified in two classes, nonprobability and probability sampling. In nonprobability samples, the researcher subjectively chooses the elements whereas in probability sampling the sampling units are selected by chance. Probability sampling techniques have many favourable properties and they allow the calculation of confidence intervals and other statistics by nature. On the other hand, they require a precise definition of target population and a general specification of sampling frame. By the nature of this study, it is generally difficult to determine the whole population for technology assessment and therefore nonprobability sampling techniques are more convenient. Moreover, it is likely that a realistic target population is at maximum thousands of people and therefore every expert identified can be included in the study. (Malhotra, 2004)

The most relevant nonprobabilistic sampling method is therefore judgement sampling, in which the population elements are selected based on the judgement of the researcher. The researcher chooses the elements exercising judgement or expertise with a belief that they are representative of the population of interest. Common examples for using judgement sampling listed by Malhotra (2004) include testing market potential of a new product or expert witnesses used in court, which closely resemble the requirements of technology assessment.

The data was collected by two email questionnaires sent to groups of relevant experts on the technology areas studied: (1) printed electronics, (2) optical interconnects and (3) universal memory. The study on printed electronics focuses on manufacturing technology capabilities for printing nanoscale features. The study on optical interconnects focuses on different nanostructures that can be used for building on-chip and chip-to-chip optical interconnect systems to replace traditional electrical interconnects. The focus reports related to this data can be found on the ObservatoryNANO project website (ObservatoryNANO, 2010b,c).

The population for each study was the people working with or studying the selected technology area. This includes professors and researchers in universities and research organisations as well as people employed by companies developing the selected technologies (typically CEOs, CTOs, managers or senior staff working in research and development). Most of the invited people were based in Europe because of the ObservatoryNANO project's focus on the European situation. The sample of invited people was gathered from sources such as company databases or publicly available information in the Internet.

Table 3 contains the number of invited people, number of respondents and the corresponding response rate for the surveys conducted. Based on several sources (Kaplowitz et al., 2004; Sheehan, 2001; Hamilton, 2009), the response rates of 26% and 31% can be considered very good when taking into account the length of the questionnaires. The response rates in the reference studies range from 13% to 33%. The response rates are also very good from economic point of view because they basically indicate that getting one more response to the questionnaire requires inviting of 3-4 new people to take part. The experts invited also used a fair share of their time to answer the questionnaire. They used roughly 25 minutes to answer (by mean and median), which means that the total expert contribution was between 10 to 20 hours depending on the respondent count. This much contribution is often very costly to obtain by other means. However, it is likely that the response rates and other figure for these kind of surveys are very case-specific and the results cannot not yet be considered as a general guideline.

Table 3: Response rates for questionnaires conducted

Questionnaire	**Invited**	**Clicked**	**Responded**	**Bounced**	**Un-subbed**	**Responded/ Clicked**
Printed electronics	110	46%	26%	6%	1%	56%
Optical interconnects	85	53%	31%	7%	0%	58%
Universal memory	264	31%	18 %	6%	0%	58%

List below gives an example of the data types collected by the questionnaires. The structure given is used in the printed electronics case. One part of the data contains the expert opinion related to several claims related to TRLs, time in years to reach the TRLs, technology impact as well as several open-ended questions. The focus in this study is the quantitative data and therefore the open-ended questions are not considered in more detail.

- Expert performance
  - Expert's own assessment of his expertise
    {Not familiar, Know how the technology works, Have been working with or studying the technology}
- Expert opinion
  - Expert opinion on several claims related to TRLs
    {Strongly disagree, Moderately disagree, Moderately agree, Strongly agree}
  - Expert opinion on time to reach TRLs
    {Now, 1-3 years, 4-7 years, Later, Never}
  - Expert opinion on several claims related to technology status
    {Very bad, Bad, Neutral, Good, Very good}
  - Several open-ended questions

Expert opinion is measured on an ordinal scale from strongly disagree to strongly agree. Neutral option was omitted in the questionnaire because it would not provide any extra information to the assessment. Neutral opinion was allowed in the assessment of technology status because it is possible that a technology offers neither negative or positive impact on some property (e.g. performance) but has a significant impact on some other property (e.g. cost). In addition to expert opinion, the respondents were asked to assess their own level of expertise on a three level ordinal scale.

2.3.2 Interpretation and analysis

Before any analysis can be performed, the interpretation of data must be defined. As suggested by Stevens (1946), stretching the strict rules of mathematical methods allowed for ordinal data a bit may lead to fruitful results. The basic structure of data collected is a group of ordinal responses for a number of questions regarding the TRL status of a certain technology. Each question corresponds to one level of TRL and there are 1-4 questions for each TRL.

This is the case in data collected for assessing the status of printed electronics nanoscale manufacturing methods, which is used as a sample for later development of the expert opinion aggregation model. The printed electronics dataset is the most complex of the data collected and the same principles apply in the analysis of other datasets as well. This applies also for assessing the impact of technologies because a similar ordinal scale is used.

The final goal of the analysis in this study is to draw a conclusion on the TRL of technologies assessed. As the proposed TRL assessment scheme consists of multiple questions per one level of TRL, the expert opinions need to be combined to form the final conclusion. Moreover, it is important for the analyst to know something more, preferably the probability distribution, about the expert opinion collected because there is seldom a yes or no answer in the assessment. Probabilistic modeling is also necessary for estimating the uncertainties related to the process. Based on this reasoning, several schemes for interpreting the TRL data collected were considered.

The simplest interpretation for the data is one presented in table 4 in the form of basic statistics. Even though their use cannot be argued mathematically, they offer a compact presentation of the dataset. For calculating the basic statistics, the following mapping was used between the ordinal categories and integers: strongly agree=4, moderately agree=3, moderately disagree=2, strongly disagree=1. By a simple visual inspection of the table, it can easily be stated that gravure printing and flexography printing are nowhere near of reaching even TRL 1 because both their median and mean score lie under the average value of 2.5. On the other hand, all claims related to nanoimprint lithography on TRL 1 seem to indicate that all of the requirements for this TRL would already be reached.

Drawing conclusions directly from descriptive statistics such as mean or median leads to a clearly biased estimate, though. Fundamentally, the data collected is multinomial because there are four categories of possible responses and the frequency of answers observed for each category is known for each question. In the very basic multinomial interpretation, a probability of experts agreeing of reaching a certain TRL can be calculated by a simple multiplication. Table 5 contains the frequencies

Table 4: Basic statistics for printed electronics data. Barplots of category frequencies for each technology can be found in Appendix A.

	Nanoimprint lithography				Ink-jet printing			
	median	mean	stdev	count	median	mean	stdev	count
Fundamental 1	3	2,8	0,9	27	3	2,8	1,0	29
Fundamental 2	3	3,2	0,9	27	3	2,6	0,9	29
Fundamental 3	3	2,7	0,9	27	2	2,5	1,0	29
Applied 1	3	2,7	1,0	27	2	2,5	0,9	29
Applied 2	3	2,8	1,1	26	2,5	2,5	1,0	28
Applied 3	2	2,4	0,9	25	2	2,4	1,0	27
Applied 4	2	2,2	0,9	24	2	2,1	1,0	26
Prototype 1	2	2,2	0,8	24	2	2,4	0,9	26
Prototype 2	1,5	1,8	0,9	24	2	2,1	1,0	25
Prototype 3	1,5	1,7	0,8	24	1,5	1,6	0,7	24
Markets 1	4	3,6	0,7	23	3	2,9	1,1	26
Markets 2	1	1,7	1,0	24	1	1,4	0,8	21
Markets 3	1,5	1,8	0,9	24	1,5	2,0	1,2	22
	Gravure printing				Flexography printing			
	median	mean	stdev	count	median	mean	stdev	count
Fundamental 1	2	2,2	1,2	21	2	2,2	1,2	22
Fundamental 2	2	2,3	1,1	21	2	2,3	1,0	22
Fundamental 3	2	2,1	1,0	21	2	2,1	0,9	22
Applied 1	2	2,1	1,0	20	2	2,1	0,9	21
Applied 2	2	2,0	1,1	20	2	2,0	1,0	21
Applied 3	2,5	2,4	0,8	18	2	2,2	0,9	19
Applied 4	1,5	1,7	1,0	18	2	1,6	0,8	19
Prototype 1	2	2,3	1,0	19	2	2,1	0,9	20
Prototype 2	2	2,1	1,1	19	1,5	1,8	1,0	20
Prototype 3	1	1,8	1,0	17	1	1,6	0,7	18
Markets 1	4	3,0	1,3	17	3	2,8	1,1	19
Markets 2	1	1,6	1,0	16	1	1,5	0,7	16
Markets 3	1	1,5	0,9	14	1	1,6	0,8	16

for each category of TRL 1 (Fundamental research) for the Nanoimprint lithography technology. Direct probability calculation suggests that the probability of an expert strongly agreeing with each claim is only 2%. The probability of an expert agreeing strongly or moderately to each claim is 33%. Therefore one cannot draw a conclusion that TRL 1 would be reached because the average value of responses for each TRL 1 related question is above the average value of 2.5.

Table 5: Sample data for TRL 1 of Nanoimprint lithography in printed electronics

	Strongly agree	Moderately agree	Moderately disagree	Strongly disagree
Fundamental 1	22%	44%	26%	7%
Fundamental 2	41%	44%	7%	7%
Fundamental 3	22%	37%	33%	7%

By the reasoning above, it seems necessary that the multinomial format of data is retained throughout the process to achieve unbiased estimates in the modeling point of view. This implies that the ordinal nature of data is retained, too. For the final analysis of TRL and related uncertainties, the probabilities assigned for each TRL related question need to be combined to conclude the probability of reaching the specific TRL. In order to achieve this, the scheme in table 6 is suggested. This kind of interpretation is not needed for subjective logic as the framework contains operators for combination of beliefs of independent (or dependent) items.

The suggested scheme is based on the assumption that individual TRL related items would be independent. This is of course not the case but it greatly simplifies the analysis and it would be very difficult to model the dependencies between the TRL claims. The independence between claims should be considered in the questionnaire design.

The scheme for agreement consists of four different groups based on the number of experts agreeing with the claims. The highest category consists of the events where all three items are agreed and the lowest category consists of the events where all three items are disagreed. The uncertainty is taken into account in a similar mapping that classifies responses based on the certainty of the respondent.

Table 6: Suggested TRL interpretation scheme. The pattern describes ordinal category frequencies for: strongly agree-moderately agree-moderately disagree-strongly disagree. The sum of probabilities exceeds 100% due to rounding.

(a)

Agreement	Patterns	Probability
3 items	3-0-0-0	2%
	2-1-0-0	9%
	1-2-0-0	15%
	0-3-0-0	7%
	total	33%
2 items	2-0-1-0	6%
	1-1-1-0	17%
	0-2-1-0	12%
	2-0-0-1	2%
	1-1-0-1	5%
	0-2-0-1	3%
	total	45%
1 item	1-0-2-0	6%
	0-1-2-0	8%
	1-0-1-1	3%
	0-1-1-1	4%
	1-0-0-2	0%
	0-1-0-2	1%
	total	22%
0 items	0-0-3-0	1%
	0-0-2-1	1%
	0-0-1-2	0%
	0-0-0-3	0%
	total	2%

(b)

Uncertainty	Patterns	Probability
3 items	0-2-1-0	12%
	0-3-0-0	7%
	0-0-3-0	1%
	0-1-2-0	8%
	total	28%
2 items	1-2-0-0	15%
	1-1-1-0	17%
	0-1-1-1	4%
	1-0-2-0	6%
	0-2-0-1	3%
	0-0-2-1	1%
	total	46%
1 item	2-1-0-0	9%
	2-0-1-0	6%
	0-1-0-2	1%
	1-1-0-1	5%
	1-0-1-1	3%
	0-0-1-2	0%
	total	24%
0 items	3-0-0-0	2%
	2-0-0-1	2%
	0-0-0-3	0%
	1-0-0-2	0%
	total	4%

2.4 Modeling

2.4.1 Multinomial and weighted multinomial models

Multinomial model for TRL data is very simple and is for most parts already described in data interpretation subsection. The main idea in the multinomial model is that the probability of each ordinal category is its relative frequency among the response data to one item. The multinomial data frame for item i can be defined as

$$m_i = \{f_{i1}, f_{i2}, \ldots, f_{iK}\} = \{\frac{n_{i1}}{N}, \frac{n_{i2}}{N}, \ldots, \frac{n_{iK}}{N}\} \tag{10}$$

where K denotes the number of categories, n the number of responses per category per item and N the total number of responses per item. For convenient comparison of performance between different models, an expected value or a point estimate can be calculated to the multinomial data frame:

$$E[m_i] = \sum_k k \cdot f_{ik}. \tag{11}$$

This is different from the true expectation operator for multinomial data and is used only for comparing of model performances. Another approach for comparing model performances is comparing the shares of experts agreeing and experts disagreeing, which better retains the ordinal nature of data.

The basic multinomial model does not allow any weighing of experts to be done. To apply weighing, the basic model can be modified by including weights in the calculation of category frequencies:

$$mw_i = \{f_{i1,w}, f_{i2,w}, \ldots, f_{iK,w}\} = \{\frac{\sum_j w_{ij} I_{ij1}}{\sum_j w_{ij}}, \frac{\sum_j w_{ij} I_{ij2}}{\sum_j w_{ij}}, \ldots, \frac{\sum_j w_{ij} I_{ijK}}{\sum_j w_{ij}}\}, \tag{12}$$

where the indicator function I_ijk gets value one if judge j has responded to item i by the category k and is otherwise zero. Parameter w_ij denotes the judge and item specific weight of each judge and is used weighing the scores and normalising the results. Multinomial expectation function in eq. 11 works for weighted multinomial model, too. The practical comparison between models is performed in R and implementation is done according to equations 10 and 12.

2.4.2 Judge-specific model

All multi-rater latent variable models discussed here are based on the assumption that judge j observes the true value of the item Z_i with an additive judge-specific error ϵ_{ij}. The quantity $t_{ij} = Z_i + \epsilon_{ij}$ denotes the perception of the judge, which is assigned in one of the ordinal categories c by the following equation:

$$\gamma_{j,c-1} < t_{ij} \leq \gamma_{j,c}. \tag{13}$$

The variable $\gamma_{j,c}$ denotes the judge-specific category cutoff for category c. For the model presented in Johnson and Albert (2001), extreme cutoffs are defined $\gamma_{j,0} = -\infty$ and $\gamma_{j,K} = \infty$.

Based on category cutoffs, the following likelihood function can be written for the data:

$$L(\mathbf{Z}, \{t_{ij}\}, \gamma, \{\sigma_j^2\}) = \prod_{i=1}^{n} \prod_{j=1}^{J} \frac{1}{\sigma_j} f(\frac{t_{ij} - Z_i}{\sigma_j}) I(\gamma_{j,y_{ij-1}} < t_{ij} \leq \gamma_{j,y_{ij}}), \tag{14}$$

where f denotes the standard normal distribution density function and I the indicator function that gets the value 1 when the perceived value t_{ij} lies in between the categories and 0 when it does not.

Rater variances σ_j^2 are assumed to be independent by Johnson and Albert (2001) and their prior distribution when f is the standard normal distribution is defined as an inverse-gamma density:

$$g(\sigma_j^2, \lambda, \alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)}(\sigma_j^2)^{-\alpha-1}\exp(-\frac{\lambda}{\sigma_j^2}),\ \alpha, \lambda > 0. \tag{15}$$

Category cutoffs γ are assumed to be independent and distributed uniformly a priori. The prior for true values of items Z_i is the standard normal distribution denoted by ϕ. The joint prior density distribution in the model by Johnson and Albert (2001) is therefore given by:

$$g(\mathbf{Z}, \gamma, \{\sigma_j^2\}) = \prod_{i=1}^{2} \phi(Z_i, 0, 1) \prod_{j=1}^{J} g(\sigma_j^2, \lambda, \alpha) \tag{16}$$

and the joint posterior density is formed by combining equations 14 and 16 into:

$$g(\mathbf{Z}, \{t_{ij}\}, \gamma, \{\sigma_j^2\}|\mathbf{y}) \propto L(\mathbf{Z}, \{t_{ij}\}, \gamma, \{\sigma_j^2\}g(\mathbf{Z}, \gamma, \{\sigma_j^2\})) \tag{17}$$

Due to the complexity of the model, its joint posterior probability cannot be evaluated directly and an MCMC sampling scheme is needed. As the sampling procedure is relatively complex and not in the main focus of this study, it is not presented in detail. The complete scheme including MATLAB source code can be found in Johnson and Albert (2001). The R implementation of the model strictly follows the original MATLAB source code.

2.4.3 Intraclass correlation based model - "ICC"

Mast and Wieringen (2004) base their model for bounded ordinal data on the intraclass correlation coefficient (ICC, Eq. 5). Instead of defining category cutoffs, they employ a much simpler approach by two maps, $LRD(Z)$ and $LDR(k)$, between the latent variables $Z \in \mathbb{R}$ and ordinal variables $k \in \mathbb{D}$, where $\mathbb{D}$ is assumed a finite set whose categories are labeled $1, 2, \ldots, a$. The maps are defined in the following equations:

$$LRD(Z) = \left\lceil \frac{a \exp Z}{1 + expZ} \right\rceil \tag{18}$$

$$LDR(k) = \log \frac{k - 1/2}{a - k + 1/2} \tag{19}$$

Based on the log-likelihood function, estimates for true latent values Z_i and the measurement error σ_e^2 can be established by maximisation:

$$\hat{Z}_1, \ldots, \hat{Z}_n, \hat{\sigma^2_{e,ml}} = \arg\max \sum_{i=1}^{n} \sum_{k=1}^{a} n_{ik} \log \Phi(A(+)) - \Phi(A(-)). \tag{20}$$

Function Φ denotes the cumulative standard normal distribution and function A is defined by:

$$A(\pm) = \frac{LDR(k \pm 1/2) - Z_i}{\sigma_e} \tag{21}$$

The R implementation of the ICC method is done according to equations specified in this subsection and Mast and Wieringen (2004).

2.4.4 Rater behaviour based model - "Ratings"

Another possible approach for latent variable modeling is suggested by Ho and Quinn (2008), who use a model with fixed category cutoff points but include rater critical and discriminative behaviour in the model. In the model y_{ij}^{obs} denotes the observed judgements that belong to one of the ordinal categories. Corresponding latent variables are defined by

$$y_{ij}^{*} = \alpha_j + \beta_j \theta_i + \epsilon_{ij}. \tag{22}$$

Parameters α_j take real values and are used to model the central point of rater j scale. If α_j of a rater j is larger than α of other raters, the rater j gives on average higher estimates for values of items. Similarly, if α_j is lower, the rater tends to give lower estimates on average. Parameter θ_i denotes the latent value of item i and parameter β_j the rater's ability to be able to discriminate between different values of items. Raters with high discrimination value are sensitive to small differences in item values whereas raters with low discrimination value estimate the item values close to random.

A Bayesian MCMC approach is used for fitting the model that has the likelihood function of:

$$p(\mathbf{Y}^{obs}|\alpha, \beta, \theta, \gamma) = \prod_{i,j} \left\{ \Phi(\gamma_{y_{ij}^{obs}} - \alpha_j - \beta_j \theta_i) - \Phi(\gamma_{y_{ij}^{obs}-1} - \alpha_j - \beta_j \theta_i) \right\}, \tag{23}$$

where Φ denotes again the standard normal distribution function. Prior distributions used by Ho and Quinn (2008) are α_r $N(1,1)$, β_r $N(-5,20)$ truncated to the positive half of real line, uniform distribution for γ and standard normal distribution for θ_p. After running the simulation and obtaining MCMC samples, ordinal category probabilities are their relative proportion among all MCMC samples. R implementation for the ratings method is provided as an R library named "Ratings" by the authors of Ho and Quinn (2008).

2.4.5 Subjective logic

Subjective logic is based on representing beliefs by a vector of four components: belief, disbelief, uncertainty and relative atomicity. Mathematically the belief information of event x is defined by $\omega_x = (b(x), d(x), u(x), a(x))$. The use of subjective logic for TRL assessment with ordinal data requires the definition of a mapping between the ordinal data and subjective logic framework. The mapping used for this study is visualised in figure 8 and is based on ordinal data received by experts, expert evaluation of his own expertise and analysts evaluation of an expert's expertise.

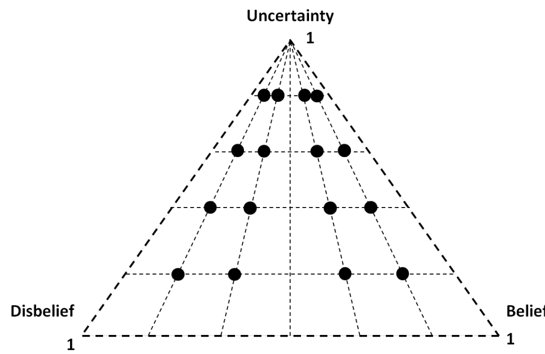


Figure 8: Mapping of TRL data to subjective logic framework. The dots represent the effect of uncertainty in the values of belief and disbelief. When uncertainty approaches 1 in the top corner of the triangle, the numerical differences in belief values between different categories become smaller. Therefore opinions with high uncertainty possess less information than opinions with low uncertainty.

In case of 4 category ordinal scale, the line between belief and disbelief can be divided in 5 equally long segments. The centre points of these segments denote the different ordinal categories expect the line that is in the middle of belief and disbelief, which corresponds to the omitted neutral category. By this definition, the probabilities assigned for the belief of ordinal categories are: 0.9 for strongly agree, 0.7 for moderately agree, 0.3 for moderately disagree and 0.1 for strongly disagree. The belief of omitted neutral category would be 0.5. By definition, disbelief in subjective logic is $d(x) = 1 - b(x)$.

As the data acquired by the questionnaire does not include an uncertainty value for experts' assessments of each item, the uncertainty needs to be based on estimated expert value. For this, a simple scheme is suggested where both expert's own and analysts assessment of expertise, on 0-2 integer scale, are summed to get a 0-4 integer scale. This score is directly mapped to uncertainty values by: $u(0) = 1$, $u(1) = 0.8$, $u(2) = 0.6$, $u(3) = 0.4$ and $u(4) = 0.2$. This gives a linear, monotonously increasing certainty value as the function of expertise. To meet the requirement of belief vector components summing to one, belief and disbelief scores need to be multiplied by $(1 - u(x)$. To summarise the procedure, a level 3 expert moderately disagreeing with an item would get a belief vector of $\omega_x = (b(x), d(x), u(x), a(x)) = (0.18, 0.42, 0.4, 0.5)$. The base rate or relative atomicity is always set to 0.5 because no prior information on any item or expert is assumed.

As a result of applying the procedure described above, there is a list of belief vectors for each TRL assessment item. To reach a consensus between experts, subjective logic offers a consensus operator combining conflicting beliefs (Jøsang, 2002). This operator is defined between opinions of agents A and by with beliefs, $\omega_x^A = (b_x^A, d_x^A, u_x^A, a_x^A)$ and $\omega_x^B = (b_x^B, d_x^B, u_x^B, a_x^B)$. The common denominator is defined $\kappa = u_x^A + u_x^B - u_x^A u_x^B$, which is assumed to be larger than zero throughout this study.

$$\begin{aligned}
b_x^{A,B} &= (b_x^A u_x^B + b_x^B u_x^A)/\kappa \\
d_x^{A,B} &= (d_x^A u_x^B + d_x^B u_x^A)/\kappa \\
u_x^{A,B} &= (u_x^A u_x^B)/\kappa \\
a_x^{A,B} &= \frac{a_x^A u_x^B + a_x^B u_x^A - (a_x^A + a_x^B) u_x^A u_x^B}{u_x^A + u_x^B - 2u_x^A u_x^B}
\end{aligned} \tag{24}$$

The consensus operator is both commutative and associative, which means that the same consensus between multiple opinions is reached by any order of application of the operator.

Calculation of a single TRL value requires the combination of different questions related to the TRLs. As presented earlier in the subsection 2.3, the questionnaire consists of 13 TRL related questions per each technology which are distributed as: 3 questions per TRL 1, 4 questions per TRL 2, 3 questions per TRL 3, 2 questions per TRL 4 and one question per TRL 5. Beliefs about independent individual TRL questions can be combined by the multiplication operator in subjective logic (Jøsang, 2010). The operator is defined as $\omega_{x \wedge y}$ for the belief frames $\omega_x = (b_x, d_x, u_x, a_x)$ and $\omega_y = (b_y, d_y, u_y, a_y)$:

$$\begin{aligned}
b_{x\wedge y} &= b_x b_y + \frac{(1-a_x)a_y b_x u_y + a_x(1-a_y)u_x b_y}{1 - a_x a_y} \\
d_{x\wedge y} &= d_x + d_y - d_x d_y \\
u_{x\wedge y} &= u_x u_y + \frac{(1-a_y)b_x u_y + (1-a_x)u_x b_y}{1 - a_x a_y} \\
a_{x\wedge y} &= a_x a_y.
\end{aligned} \tag{25}$$

Consensus and multiplication operators allow the combination of TRL related beliefs by calculating the belief of experts agreeing with all claims related to a single TRL. This is equal to performing the operation $\omega_{x\wedge y\wedge z}$, where x, y and z are the TRL related claims for which the belief vector is achieved by the consensus operator. In order to find out the belief for events such as $\omega_{x\wedge y\wedge \bar{z}}$, which corresponds to a situation where TRL items x and y have been reached but z has not. The negation or complement operator, $\omega_{\neg x}$ needed here is defined (Jøsang, 2001):

$$\begin{aligned}
b_{\neg x} &= d_x \\
d_{\neg x} &= b_x \\
u_{\neg x} &= u_x \\
a_{\neg x} &= 1 - a_x
\end{aligned} \tag{26}$$

2.5 Reliability

2.5.1 Sensitivity analysis

There are many views on how to measure reliability of a decision analysis tool such as the one suggested in this study. However, in this case two measures are by far more important than the others. Firstly, the analyst does not have any prior knowledge on how many respondents there will be for the questionnaire. Moreover, it is usually even unknown how many experts there are regarding the technologies to be assessed. Therefore, the robustness of the decision support models is the most important individual measure for comparing model performances. Robustness in practice means that the model should vary as little as possible as the function of number of respondents. The robustness is measured as the model's sensitivity to data available.

Clemen (1996) defines sensitivity analysis as finding out what inputs of the model really matter in the final decision. As all the data in this study is considered equal, we can only compare the models on how they perform when different sets of data is used. The comparison is done using two different schemes: using a random set of data sampled from the whole set of data both using and not using replacement. The former is called bootstrap sampling. The best model in robustness sense is the one that has least variation in the function of experts available to the study.

The reasoning behind the suggested scheme is to find out how much the results could change if more experts had responded to the study. The comparison is performed by item level and the performance of best available method is then compared on final decision level as the decision making is based on item values.

Two measures for sensitivity analysis will be used in the results section of this study. Mean Absolute Percentage Error (MAPE) compares the estimates to the best estimates achieved by each method and presents the result in percentages. The use of percentages is essential because the use of different methods lead to different estimates and therefore the absolute values as such would not results in a fair comparison. MAPE can be defined by:

$$MAPE = \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right|, \tag{27}$$

where A_i denotes the actual value for item i and F_i the forecasted value. There are some weaknesses in MAPE, most notably its inability to handle zero actual values and the fact that there is no upper limit to the percentage error, which might lead to highly skewed error estimates. The estimates in this study will be limited between zero and one and most of the items to be estimated will be considerably higher than zero.

To assess the robustness in case of biased estimators, the methods will be compared against each other also using Median Absolute Deviation or MAD. MAD is a measure of variability, which does not take into account the reference value of the data as MAPE does. MAD is defined by:

$$MAD = median_i(|X_i - median_j(X_j)|), \tag{28}$$

which corresponds to taking the median from the absolute deviation of the data from its median. The use of median values makes MAD robust to outliers and therefore a favourable choice for this study.

2.5.2 Verification and validation

Validation of the results is not straightforward because there is naturally no reference data available. There are in practice two strategies for validation of results, by using all available cues of TRL status to manually validate the results or by asking the a group of experts their opinion on the final results.

Validation using cues means trying to find publicly available information such as news articles and scientific publications to get cues on the TRL. These might include for example news of companies starting to manufacture a new product (i.e. market entry phase), R&D organisations successfully demonstrating a component or prototype (i.e. applied research or prototype phase). This kind of validation is always case specific and relies solely on the information available. It can be effectively used for rejecting the results because if completely contradicting information is available, it is likely that the TRL estimates are incorrect. Searching news articles and browsing scientific publications or patents is very time consuming and difficult for a non-technology expert making this validation scheme sometimes difficult to apply.

Another strategy for validation is to ask the experts what they think about the results. The experts may or may not be the same as in the assessment phase. This approach closely resembles the Delphi methodology, where subsequent assessment rounds are performed to reach a group consensus. The expert validation gives experts a chance to revise their opinion and indicate if there is something very wrong with the results. Even though the expert review is widely used for the validation of academic studies, it may still provide incorrect information for TRA purposes. The experts may not always be aware of the latest advancements and would then be likely to argue that the assessment is incorrect. It is also possible that two or more experts give contradictory comments in the review phase. In conclusion, there is no perfect way for validating and no method can guarantee that the estimates would be indisputably valid.

Both publicly available information and expert reviews were successfully used during the process of TRA. Findings related to validation are presented in section 3 and discussed in 4.

2.5.3 Estimating and reducing uncertainty and bias

As discussed in literature review, there are multiple sources for uncertainty and bias in the process of Technology Readiness Assessment. The uncertainty included in the process can be reduced by collecting large amount of data and building a model that tries to resolve the uncertainty as much as possible. From this perspective, one possible solution would be to define Technology Readiness Levels in such detail that the experts can give their estimate in binary statements of either true or false. This decomposition would first of all make TRL related claims observable variables, which would lead to a simpler assessment problem for the experts. A simple process is essential for reducing expert originating motivational and cognitive biases.

The main challenge is that defining very detailed claims of each TRL is a difficult task for anyone. It is also very difficult to create a technology independent questionnaire that consists of dozens of claims. Another approach for designing the questionnaire is to leave it as simple as possible and rely on each experts ability to understand and assess technology readiness as a whole. This kind of approach would require the expert to give only one assessment of TRL per technology resulting in a very simple statistical model where TRL would be assigned to the category which got the most expert votes.

In conclusion, two extreme approaches are possible in choosing the modeling approach for TRA. The first tries to minimise expert originated biases by defining an exhaustive questionnaire for TRA. The challenge is that defining such questionnaire is a difficult task and will probably introduce analyst originated motivational bias due to the analyst not fully understanding the technologies being assessed. The other approach where a minimal questionnaire is designed tries to eliminate any model originated biases by using as simple model as possible. This relies on the experts in determining the TRL and requires clear definition of TRLs to be available to the experts. As more room is left for experts own assessment, it is likely that there will be a greater amount of motivational bias as well as cognitive bias originating from the expert assessment.

Other sources of biases present are more clearly related to the data collection procedure. The selection of respondent population is crucial to the assessment. It should be large enough to cover all relevant experts but compact enough not to include any people that do not understand the area under study. Moreover, the questionnaire invitation and structure need to be so attractive that the invited experts feel it is useful for them to participate. Response rate is a good indicator for the quality of results. A very low response rate (e.g. below 10%) indicates that there may be something wrong with the questionnaire as the experts are not interested in taking part. A faulty questionnaire will naturally lead in very biased estimates of the TRL.

Most of the subjects relevant for TRA require a global perspective. Especially in case of information and communication technologies, experts are distributed all over the world most notably in Europe, Asia and America. The English language is a natural choice for the study but it needs to be noticed that people have very different levels of proficiency in English. Very simple language should be used to avoid bias originating from different interpretations of questions. In extreme cases people understand the question in opposite ways rendering it totally useless.

It is likely that the experts suffer from wishful thinking when assessing technologies they are personally working with. If TRA is used for instance in European Union level funding decision, it is obvious that many experts would answer in such way that could increase their chances in getting more funding. One solution is to assign low weights to experts considered suffering from wishful thinking, which is of course very difficult to know. Somewhat easier task is to try weighing experts based on estimates on their expertise level. Weighing tries to ensure that opinions from best experts would affect the results most but it is difficult to define who is a high quality expert and who is not. A simple weighing scheme is present in the weighed multinomial model as well as in subjective logic and it can be used to assign weights on experts by any ranking considered appropriate.

2.6 Implementation

The practical implementation of the expert elicitation process consists of the following steps. First, some desk research is required for the analyst to become familiar with the subject. It is crucial to understand the big picture of technologies to be assessed in order to design a valid questionnaire. The knowledge on current state-of-the-art and most potential emerging technologies is especially needed. In the questionnaire design phase, the analyst needs to define what is the set of technologies to be assessed and what are the most relevant factors corresponding to the technology impact. The questionnaire is designed based on the findings in the desk research based and then executed by a web survey tool.

Analysis of the questionnaire data is typically done in either the web survey tool or in a spreadsheet application. For this study, a simple tool is developed to perform all data preparation, analysis and reporting work semi-automatically based on analysts commands. As the questionnaire structure is different each time, a fully automatic solution cannot be achieved. The implementation follows the guidelines set in requirements specification in the beginning of this section. The visualisation provided by the tool is in figure 12 in section 3.

The tool developed is a front-end for loading the data, performing the analysis and saving the results for later time. It also contains the analysis engine allowing an easy extension of new analysis or visualisation methods. The main advantages of the tool are to reduce bias from individual analysts actions and offer a consistent framework for visualising the results. Moreover, as the data is encapsulated by the tool there is a smaller chance for altering or losing the work in progress by accident. In conclusion, the tool developed makes the process much less vulnerable to analyst originated biases, reduces the amount of work needed and in general results in a leaner process.

3 Results

3.1 Comparison

In total 8 methods are compared in this section. They are a basic multinomial model (Basic), a weighed multinomial model (Weighed), three subjective logic models (SL Basic, SL Averaging, SL Medium) with different parameters, a judge-specific latent variable model (Judge-specific), ICC based latent variable model (ICC) and rater behaviour based latent variable model (Ratings). The basic subjective logic model assumes that expert opinions are in the extreme ends of the belief-disbelief scale and therefore it closely resembles the basic multinomial model. The averaging subjective logic model distributes the expert opinion uniformly between belief and disbelief and as a results provides estimates that are heavily biased towards neutral values, thus the name "averaging". The SL Medium model is in between these two in terms of the model behaviour.

The basic multinomial model is used as a reference for the comparisons. The basic model does not make any assumptions on the data and it is not able to include any additional data in calculating the TRL score. All other methods try to benefit from additional data either from the analyst (weighed multinomial and subjective logic) or from the latent structure of data and larger data set for performing the analysis (latent variable models). For instance, the basic model uses data from 28 raters for assessing the score of an individual TRL item whereas latent variable models are able to use data from roughly $28 \times 13 \times 4 = 1456$ individual ratings (experts,TRL items and technologies).

Table 7 reveals that from the methods compared, averaging subjective logic and ratings are the least discriminating and the weighed multinomial model is the most discriminating. This means that 65% of the scores given by the averaging subjective logic model are closer to the neutral or indecisive value than the scores given by the basic model. To give a practical example, a model assigning all scores to the neutral value would get the most robust results with highest bias. Similarly the averaging subjective logic clearly biases the scores towards the neutral value because of the handling of uncertainty in the model. Therefore, on average the scores given by the averaging subjective logic model are 4 percentage points closer to the neutral value than the scores given by the basic model. Full table of results used for calculating the scores for comparison can be found in Appendix D.

The latent variable models, Ratings and ICC, are also less discriminating than the basic model. The only model that is more discriminating than the basic model is the weighed multinomial. If the comparison is taken on the level of individual technologies (Tab. 8), one can see that the pattern is the same for each technology. This is because the weighed multinomial assigns larger weights for experts that work with a certain technology and these experts are more likely to give answers with more confidence and tend to favour the technology they are working with.

Table 7: The effect of models in comparison to the basic multinomial model. The first row indicates the share of individual TRL scores being closer to neutral opinion. The second row contains the average effect of the model: e.g. -4% means that on average subjective logic scores were 4 percentage points closer to neutral than with the basic model. Notable deviations from expected value labeled by exclamation marks.

Measure (unit,expected)	Weighed	SL Averaging	SL Medium	SL Basic	Ratings	ICC	Judge-specific
Scores closer to neutral (%, 50%)	21 (!)	65 (!)	48	31 (!)	67 (!)	54	54
Median effect (pp,0%)	2	-3 (!)	1	6 (!)	-2	0	0
Average effect (pp,0%)	2	-4 (!)	1	5 (!)	-2	-1	0
St dev of the effect (pp,0%)	5	7 (!)	7 (!)	8 (!)	6 (!)	5	3

The comparison on the level of individual technologies also reveals that the methods tend to favour gravure and flexography printing. In case of weighed multinomial, this is probably due to the fact that they are the least known technologies regarding their ability to print nanoscale features. Therefore the amount of highly skilled experts is lower and high weights are assigned on them. The same behaviour, where gravure and flexography get more extreme scores, can be seen for all other methods, too. This is clearly an adverse results because it means that two technologies that get the most extreme scores with the basic method get less extreme scores with other models. The more advanced methods therefore tend to average the scores especially when there is a clear indication of an extreme value. In practice this would lead to a situation where an item that gets 100% respondents agreeing a priori would get significantly lower score after applying the more advanced models because the models would think this as a faulty score that is not likely given rest of the data.

Table 8: Results broken down on the level of individual technologies show that nanoimprint lithography is drawn most aggressively towards neutral values. Weighed multinomial model on the other hand provides more extreme results for each technology. ICC and Judge-specific models provide the most robust results in broad sense. Notable deviations from expected value labeled by exclamation marks.

	[pp]	Weighed	SL Averaging	SL Medium	SL Basic	Ratings	ICC	Judge-specific
NIL	Average	1	-8 (!)	-3	2	-7	-2	0
	Median	0	-11 (!)	-4	-1	-6	-2	0
	St dev	4	8 (!)	8 (!)	9 (!)	5	5	0
INK	Average	2	-3	0	4	-2	-2	0
	Median	2	-3	2	5	-2	-2	0
	St dev	5	8 (!)	8 (!)	8 (!)	6	5	3
GRA	Average	3	-1	4	8 (!)	1	0	1
	Median	2	0	5	8 (!)	2	1	0
	St dev	3	4	4	4	6	4	3
FLE	Average	4	-3	1	4	-2	1	1
	Median	5	-1	0	5	-2	2	0
	St dev	5	8 (!)	8 (!)	10 (!)	6	5	3

3.2 Reliability, robustness and sensitivity

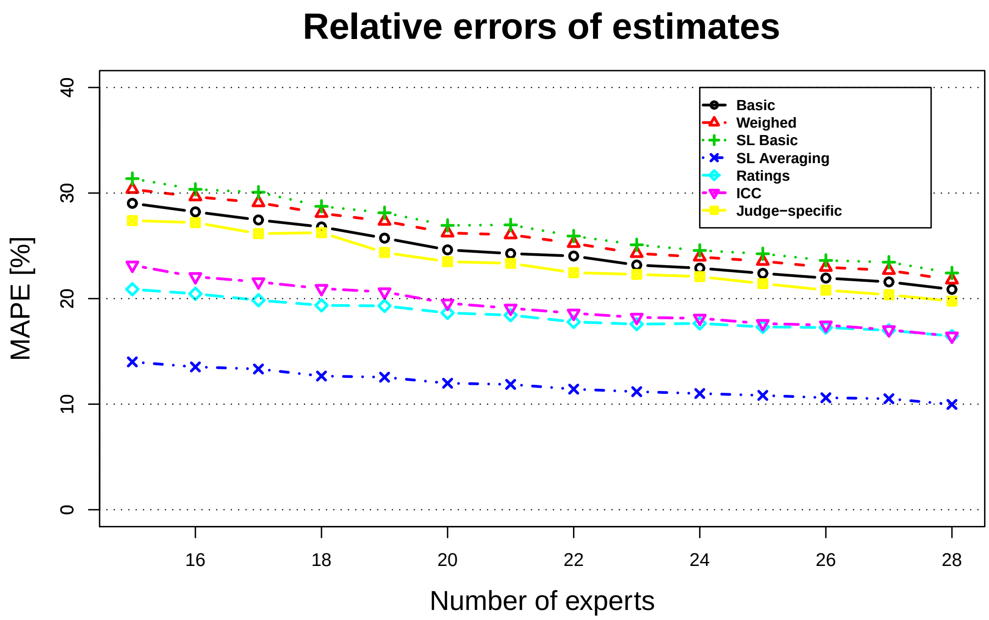
A view to the reliability of methods is got by comparing them in terms of deviation from the best available estimate as well as their variability. The deviation from the best available estimate, as a function of number of respondents, is calculated for each method separately. The reference value, i.e. the best available estimate, is the median estimate with the full population of respondents. Using bootstrap sampling, 1000 samplings for an increasing number of respondents (15 to 28) are drawn from the full dataset for comparing the methods. All estimates are then compared to the reference value and an average relative error (in percentages) is plotted in figure 9a.

Tests were executed at the same time with same samples for all methods apart from "Judge-specific". A sample size of 1000 was used, which seems to be large enough for the results to converge close to their final values. Tests for the "Judge-specific" method were run independently from others with 500 samples and 500 MCMC iterations for each sample. The convergence of MCMC iterations was only inspected visually because it is not in the focus of this thesis. The results are very well in line with other methods and do not indicate that further inspection would be required for MCMC convergence. Execution of 500 samplings for "Judge-specific" method and 1000 samplings for other methods takes a full weekend to finish even on the most powerful student computer provided by the Aalto University.

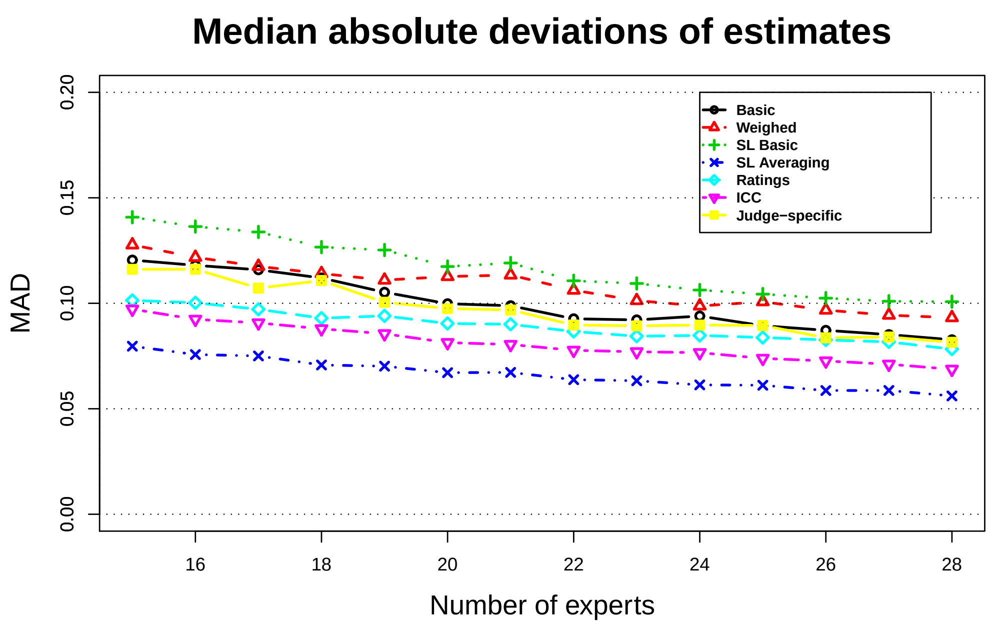
Figure 9a shows that there are clear differences in the robustness of the methods. This is closely linked with the type of the methods, as the three least averaging methods perform the worst. This is because the basic methods do not make any assumptions on the data, which increases the variance of results. From robustness viewpoint, latent variable models Ratings and ICC, achieve similar performance even though their implementations differ significantly. Finally, the subjective logic implementing a classification for expert uncertainty seems to be the most robust of the compared methods. This, however, comes with a price of the method being the most biased as well.

Another view to robustness is achieved by calculating the Median Absolute Deviation (MAD) for the bootstrapped estimates. The MADs in figure 9b show a similar pattern that is achieved for average errors. This concludes the point that the more advanced methods for aggregating expert opinion have clearly an effect in the robustness of results achieved.

Based on results introduced earlier in this section, subjective logic seems the most convenient method for expert opinion aggregation in this case. None of the more advanced latent variable models proved to be significantly better than subjective logic in terms of reliability, robustness and sensitivity measures used. Moreover, subjective logic provides very convenient operators for combining beliefs from various sources of information and merging of different statements by e.g. the multiplication operator. The performance of ICC is perhaps the largest surprise in the comparison. It is the simplest from latent variable models, results in good robustness and relatively low bias. The reason for choosing subjective logic over ICC for



(a) Median Absolute Percentage Error (MAPE)



(b) Median Absolute Deviation (MAD)

Figure 9: Median Absolute Percentage Error (MAPE) (a) and Median Absolute Deviation (MAD)(b) for estimates of methods compared. Clear groups with different performance levels are present in figure (a) whereas the results in (b) are more dispersed. Both measures, however, lead to same conclusion of methods' performances.

full TRL estimates is that subjective logic can be modified more easily for different questionnaire settings and the performance difference is very small between the two methods.

Based on the TRL data interpretation scheme suggested in section 2, the final belief mass in favour of a certain TRL level to be reached is equal to the average of belief masses in favour of sub-TRLs have been reached according to the equation below. In the equation a denotes the number of items agreed on the left hand side and i each item the TRL consists of.

$$b(3a) + \frac{2}{3}b(2a) + \frac{1}{3}b(1a) = \frac{b(i1) + b(i2) + b(i3)}{3} \tag{29}$$

Figure 10 represents the average error in case of TRL 1 as a function of the number of respondents. TRL 1 consists of 3 items and it is easy to notice that the average errors are in fact approximately one third of the average error of item specific errors. Moreover, when three items have been composed into the TRL estimator, the difference in relative errors between the methods get smaller. It is very likely that this effect results from the correlation between the items. In case of TRL 1, all the items describe actions in fundamental research phase, which are interlinked with each other up to some extent. This linkage or correlation between items results in lower variance in the final results.

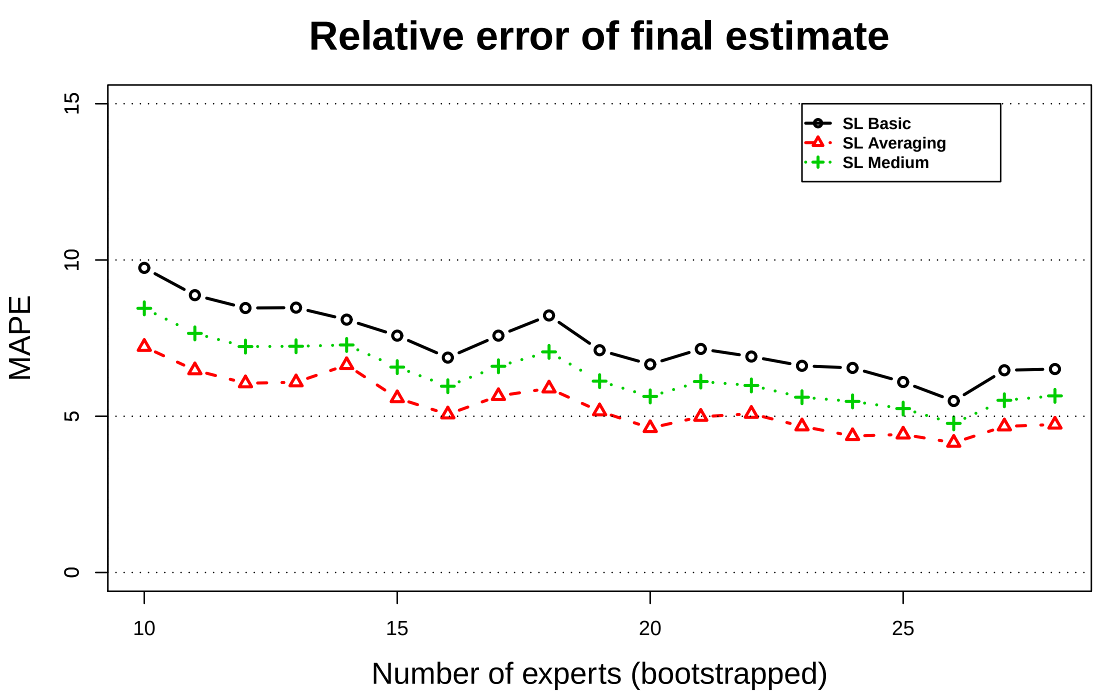
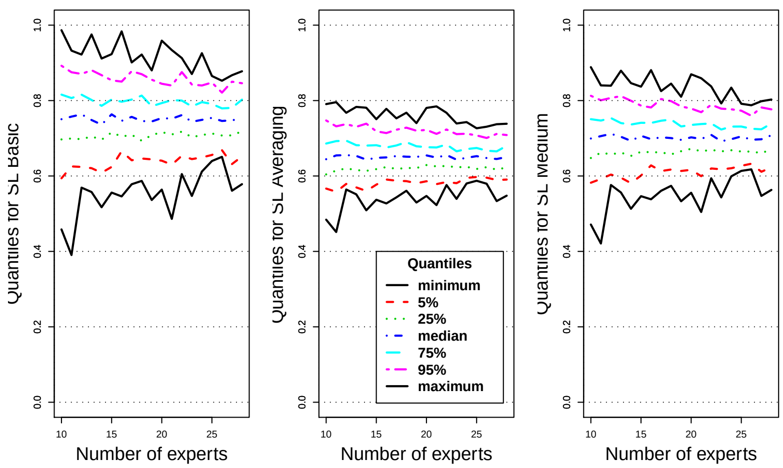
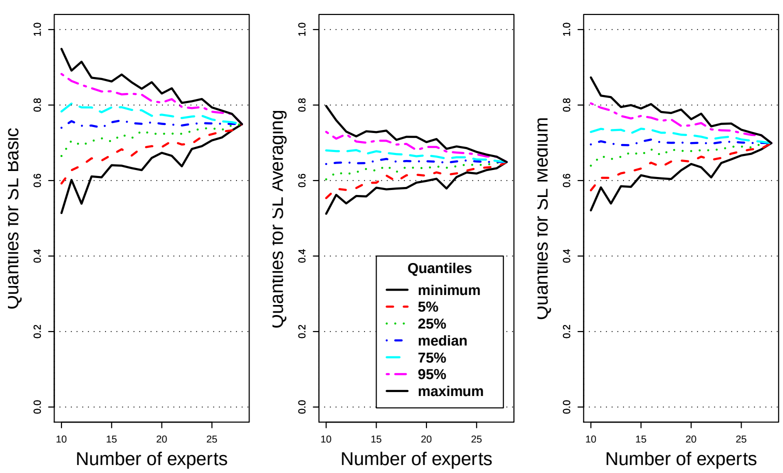


Figure 10: MAPE for TRL 1 probability of Nanoimprint lithography as function of number of respondents. Sampling with replacement (bootstrap), number of samples 100.

Even though the variance of the results is smaller in the final estimates, there are still significant differences between different methods. Figure 11 shows that the averaging subjective logic framework in terms of deviation of results estimates the probability of TRL 1 at 65%. On the other hand, the SL Basic with most deviations would suggest the probability of TRL 1 being approximately 75%. Based on these results of the final estimates, using multiple items to define the levels of TRL is a way to decrease variability of the final results but it does not guarantee that there would be no differences between the analysis methods used.



(a) Sampling with replacement (Bootstrap)



(b) Sampling without replacement

Figure 11: Distribution of the final estimate of TRL 1 for Nanoimprint lithography. (a) Sampling with replacement (bootstrap), (b) sampling without replacement. Number of samples 100. The most averaging subjective logic variant has least variation whereas the basic subjective logic has most variation.

3.3 Final estimates

Figure 12 represents the final results obtained by the tool implemented. Based on the evidence reported earlier in this section, subjective logic was selected as the aggregation method because it is simple to implement and provides adequate performance. The same method is used for aggregating both TRLs and technology impacts taking into consideration that 4 level ordinal scale was used for TRLs and 5 levels for the impact. Estimates of individual impacts are listed in table 9 suggesting that the nanoimprint lithography has clearly the highest potential of succeeding in future.

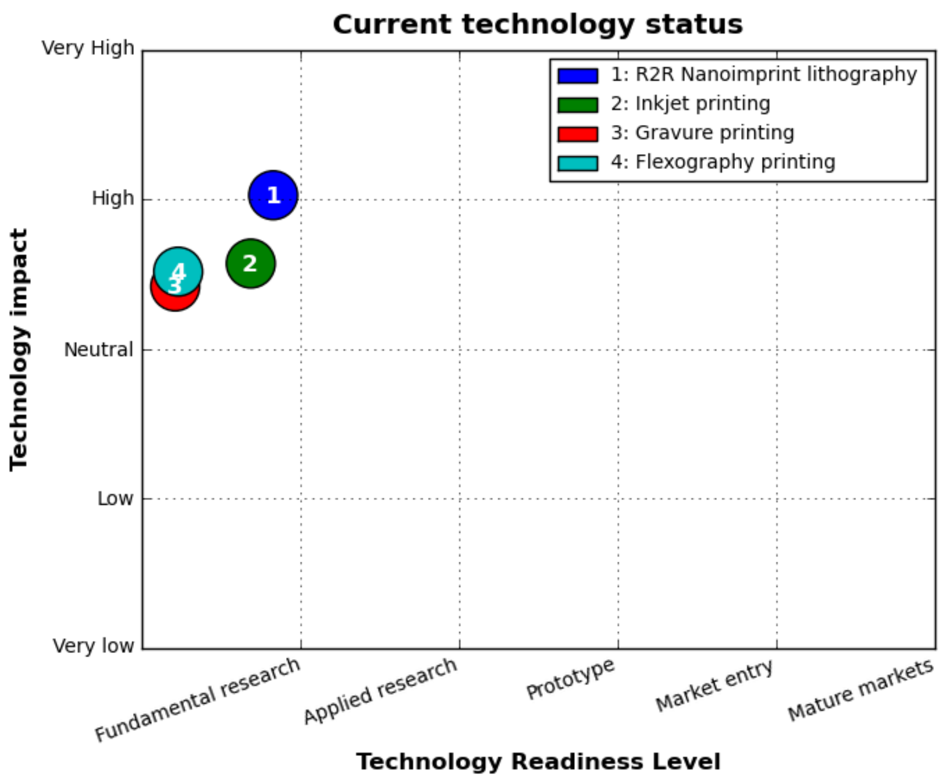


Figure 12: Final estimates obtained by the tool. Analysis method used is subjective logic

In case of TRLs, achieving the final estimates requires definition of a probability level when a certain TRL level is reached. Figure 11 shows that different analysis methods suggested different probabilities for TRL 1 to be reached in case of Nanoimprint lithography technology and whether TRL 1 can be assigned for NIL technology depends on the limit. The motivation for using expert elicitation scheme in the first place is to obtain as objective estimates for the TRL status as possible. It is very difficult to know, which one of the experts possesses correct information on the TRL status and therefore it is also difficult to set the decision limit when a

Table 9: Impact values for nanoscale printing technologies for the moment being and at estimated time of market entry. Resulting impacts are plotted in figure 12

[%]	Novel features	Performance	Cost effectiveness	Scalability	EHS	Average
Now						
NIL	78	73	71	81	75	76
INK	77	51	63	64	66	64
GRA	51	59	63	63	60	59
FLE	41	60	73	75	58	62
Market entry						
NIL	89	86	84	87	81	85
INK	77	70	73	72	70	72
GRA	64	71	76	73	57	68
FLE	62	69	74	71	55	66

level of TRL should be reached. Based on practical consideration, the decision limit should lie somewhere between 70% and 90%. Limits above 90% would result in each technology obtaining TRL 0 just because there is a small number of experts who do not really know what is really going on in the field. Setting the limit so high would mean a highly negative approach. On the other hand, a limit of 50% would mean that only half of the probability mass is in favour of reaching a TRL leading to highly positive and thus biased estimates.

The effect of decision limit setting can be seen in table 10. The effect is huge as 10% difference in the decision limit can result in one level difference in the TRL estimate. One possible explanation based on findings in ObservatoryNANO (2010c) can be that nanoimprint lithography and ink-jet printing are the most common printed electronic technologies used. Therefore, if all the issues of printing nanoscale features can be solved, these technologies are immediately ready for prototype production thus leading to rapid development in the technology readiness. This is not the case with gravure and flexography printing with which it is unclear if they can ever be used for printing nanoscale features. Nevertheless, the choice of the decision limit itself is very difficult and can lead to highly biased results. The results here also indicate that technology readiness is not a strictly linear process. Linearity, however, is assumed in this work to make analysis and visualisation of results simpler.

Table 10: TRL level estimates on different decision limits. The limit setting greatly affects the results, which need to be considered when analysing the results.

Limit	50%	60%	70 %	80 %	90%
NIL	2.5	2.5	1.9	0.9	0.8
INK	2.7	1.9	0.9	0.8	0.7
GRA	0.5	0.5	0.4	0.3	0.3
FLE	0.6	0.5	0.4	0.4	0.3

The decision tool was also used for TRA of universal memory technologies, which are a group of non-volatile random access memory technologies (ObservatoryNANO, 2010a). The questionnaire had similar structure to ones used for printed electronics and optical interconnects with the difference of TRL assessment consisting of only one question per TRL. Some of the universal memory technologies have already entered the market and therefore there is clearly more variation in the TRL levels assigned to different technologies. Corresponding results of the universal memory TRA are in table 11.

In case of universal memory (table 11), there is much less sensitivity in the results than in the printed electronics case. It is clear that decision limit of 90% is too high. There is also little difference between results based on decision limit of 50% and 60% indicating that both are too low values for the limit. According to the table, four of the technologies (MRAM, FeRAM, PCRAM and SONOS) are already in the market. This correctly represents the reality and can be validated from multiple sources in the media and was confirmed by a number of experts in the peer-review of the ObservatoryNANO project Universal Memory briefing (ObservatoryNANO, 2010a). Based on this reasoning, the correct decision limit in this case is somewhere close to 70%.

Table 11: TRL level estimates on different decision limits. The limit setting greatly affects the results, which need to be considered when analysing the results.

Limit	50%	60%	70 %	80 %	90%
MRAM	5.0	4.9	4.7	4.0	1.0
QD RAM	3.0	2.8	2.7	2.6	2.0
FeRAM	5.0	5.0	5.0	5.0	4.0
PCRAM	4.8	4.7	4.6	3.9	1.0
RRAM	3.5	3.4	2.9	2.8	0.9
CNT RAM	3.4	3.3	3.3	3.0	0.9
Racetrack	3.6	2.9	1.9	0.9	0.8
Millipede	0.5	0.4	0.4	0.3	0.3
CBRAM	3.6	3.5	3.4	3.4	0.9
SONOS	5.0	4.9	4.8	3.9	3.0
NEMS RAM	5.0	5.0	3.9	3.8	0.9

Given the tables 10 and 11, it is not clear where the decision limit of reaching a certain TRL should be set. There is no correct value for the limit and it is very likely that best results will be achieved by tuning the parameter case specifically because the data set properties vary between different questionnaires and respondent populations. Tuning the decision limit brings another source of bias in the analysis process. The effect of different parameter values is briefly discussed here but a more detailed discussion would require more experience for applying the tool in real world TRA exercises.

3.4 Validation of results: case universal memory

The most recent application of the tool and process in practice at time of writing this thesis is the ObservatoryNANO universal memory briefing (ObservatoryNANO, 2010a). In total 51 experts in various areas of emerging memory technologies were contacted during the process and the composed results can be seen in figure 13. The information collection process included quantitative questions regarding TRL, technology impact and time estimates. A large amount of qualitative information was also collected.

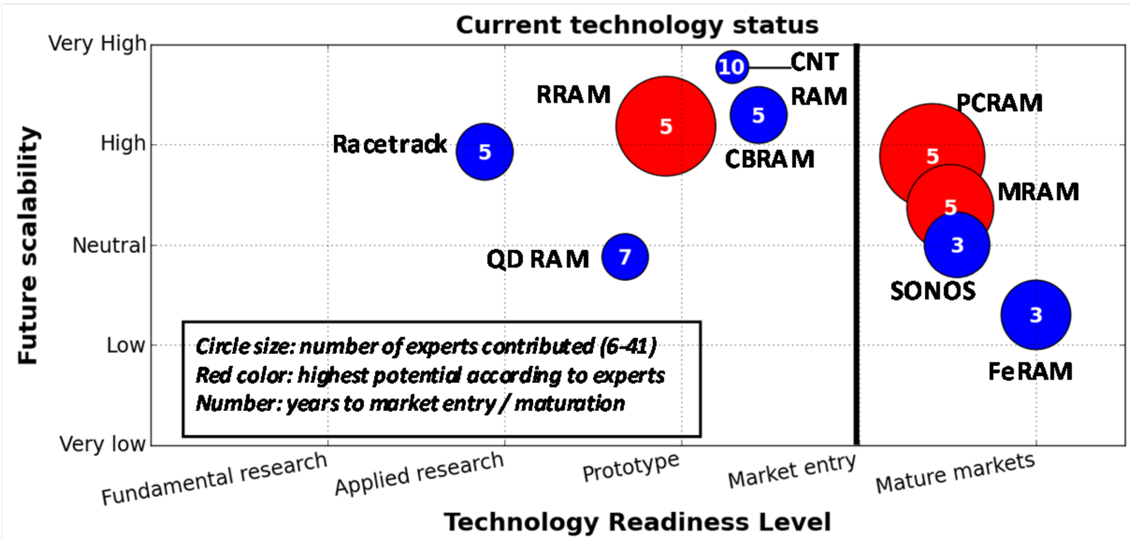


Figure 13: Final estimates obtained by the tool for Universal Memory technologies. Original figure provided by the tool slightly modified manually in order to fit it in report template. Circle size denotes the number of respondents to the questionnaire, which is directly related to reliability of the results. Therefore the conclusion for RRAM, PCRAM and MRAM can be considered most reliable. Red color used shows what are the most potential technologies according to qualitative feedback from the experts. The number inside circles tells the median estimate of time to market entry or maturation.

The validation of results was performed by both expert peer review and validation by publicly available cues. Four technologies, namely PCRAM, MRAM, SONOS and FeRAM, are claimed to lie in between market entry and mature markets according to the experts contacted. This can be validated by public search engines and press releases from semiconductor companies developing memory technologies. For instance, Fujitsu company reports on its website that it has already delivered some 100 million FeRAM devices, which concludes the point. None of the experts in the review process argued against the TRL suggestion regarding these mature technologies. Hewlett-Packard and Hynix have reported their plans to sell RRAM end-user products by the end of 2013, which is reasonable close to the expert assessment[3]. According to the news article and other news on the subject, the RRAM is clearly in the prototype phase.

[3]EETimes: "HP, Hynix to commercialize the memristor" (August, 2010)

3.5 Other means for estimating technology readiness level

The questionnaire used for nanoscale printed electronics TRL assessment also included a question where experts were asked to indicate a time estimate of when they expect a certain TRL to be reached 4.3. A 5 point scale is used with levels corresponding to TRL already reached (Now), TRL will be reached in short term (1-3 years), TRL will be reached in medium term (4-7 years), TRL will be reached in long term (Later) or TRL will never be reached (Never). The last option is given for the experts to indicate that the technology will never mature up to a certain level due to e.g. limitations in the technology.

Table 12: Estimating TRL by year estimates. Experts gave their answer on scale {Now, 1-3 years (Short), 4-7 years (Medium), Later (Long), Never}. The scale is also transformed to numerical values and means are calculated for the values. TRL estimated from expert given year estimates gives slightly more optimistic results than the detailed claims.

		TRL 1	TRL 2	TRL 3	TRL 4	TRL 5
NIL	Mode	Now	Short	Short	Medium	Long
	Median	Now	Short	Short	Medium	Long
	Mean	[1.0, 2.9]	[1.3, 3.7]	[1.7, 4.6]	[3.3, 7.0]	[5.9, 10.1]
INK	Mode	Now	Short	Short	Medium	Long
	Median	Now	Short	Short	Medium	Long
	Mean	[0.6, 2.1]	[1.5, 4.0]	[1.8, 4.7]	[3.3, 6.9]	[5.5, 9.5]
GRA	Mode	Now	Short	Short	Medium	Long
	Median	Now-Short	Short	Medium	Medium	Long
	Mean	[0.6, 2.5]	[1.3, 3.8]	[2.9, 6.4]	[4.0, 7.9]	[6.7, 11.0]
FLE	Mode	Short	Short	Short	Medium	Long
	Median	Short	Short	Medium	Medium	Long
	Mean	[0.5, 2.6]	[1.6, 4.6]	[2.9, 6.4]	[3.7, 7.6]	[6.4, 10.8]

Table 12 presents this data in terms of mode, median and mean. Mode and median represent the original categories whereas the mean is calculated from lower and upper limits of the categories. Based on the table, it can be concluded that nanoimprint lithography, ink-jet printing and gravure printing either have reached TRL 1 or are very close to reaching it. Flexography printing will most probably reach TRL 1 in 1-3 years if time. Moreover, according to the time estimates it is quite clear that none of the technologies have reached TRL 2 or any of the higher TRLs. It will take more at least 4 years for any nanoscale printing method to reach the market and almost 10 years to reach the mature market phase.

Means calculated from the category limits give somewhat controversial results when compared to modes and medians. For instance, the mode and median state that flexography printing will be the last technology to reach TRL 1 but mean indicates that it would be among the first ones. The means seem to be well in line with modes and medians in case of TRL 2 and TRL 4.

As a conclusion, year estimates can be used for estimating TRLs up to some extent but the largest challenge is that estimating reliability of the results becomes much more difficult. On the other hand, the year estimates provide perhaps the most important results of the whole TRL assessment scheme as they provide an insight in what will happen in the future. They can be used for assessing the TRLs, too, but decision on the assessment scheme should be made case by case. Use of year estimates for TRL estimation was tested in practice for the optical interconnects case (ObservatoryNANO, 2010b) and it proved to work reasonably well.

4 Discussion

4.1 Process

The process for TRA has been considered from multiple points of view during the course of this thesis. The background of this work lies in various attempts in industry and academia to understand technology readiness and its assessment. On the other hand, expert opinion elicitation and aggregation along with its advantages, limitations and uncertainties have been considered. As a result, a process and a tool for expert elicitation for Technology Readiness Assessment was developed.

The work from NASA and DoD for evaluating the TRL shows that the scheme has practical relevance and the fact that TRLs have been adopted by multiple organisations around the world (including the ObservatoryNANO project) proves that there is a need for such concept. All the organisations so far have been using TRLs as an internal tool where the assessment work is done by an individual or a small group. In the ObservatoryNANO project, TRLs are used for reporting technological advancements to the general public. The general process for producing reports for the project have included desk research, expert interviews and questionnaire and therefore it is sensible to try to improve the quality of the process and make it more fluent by designing a process for expert elicitation.

Apart from executing expert interviews and questionnaires, there was no prior knowledge on how the experts would react to the web questionnaires and how to analyse the data. A practical approach was selected using a set of simple email based questionnaires for obtaining the data. This kind of approach is reasonable because in general expert opinion data is very costly to obtain. As a surprise, the reception of questionnaires and feedback from experts was very positive. Moreover, the response rates reaching 30% can be considered very high for web questionnaires on any standards. Many of the respondents replied to the invitation with gratitude and they were also eager to see the results when published.

Despite the warm welcome, it is by no means easy to design and execute such surveys. Even though much consideration is put in questionnaire design and understanding the technology, there is a chance of asking wrong questions. Therefore, understanding the technology is crucial and the requirement cannot be bypassed by developing the process further. The worst case scenario for questionnaire design would be that it cannot be used at all because irrelevant or faulty questions were asked because of limited technological knowledge. Language and cultural issues also play a role in the data collection. The scope of questionnaires is world-wide and the language skills and culture vary from expert to expert. The questions should be so unambiguous that every respondent understands them in the same way. This requirement is very hard to satisfy and one has to rely on the assumption that majority of the respondents got the question correctly.

It is also very difficult to obtain a statistically valid sample, which needs to be taken into account when analysing and interpreting the results. The amount of responses received should be at least on the level achieved in this study: from 30 to 50 experts. This of course depends on the number of technologies to be assessed, the more technologies the more experts need to be reached. As the response rates achieved were above 10%, the number of respondents can be quite easily increased by inviting more people to participate in the study. There is of course a limit to this because there is no reason to invite non-experts to join the study and therefore the sample size cannot be increased infinitely.

The results and findings of this study can be considered from qualitative and quantitative points of view. The former includes the whole process as a way for collecting and analysing expert opinion and reporting the TRL status based on multiple sources of information. Previously, the process for producing reports for the ObservatoryNANO project has included desk research, expert interviews and case specific questionnaires. The process developed here partially answers to each part by providing a lot of qualitative information on the TRL status, which could only be received by extensive expert interviews. Quantitative data is also obtained and this data can be used for creating visualisations to support other parts of the TRA. Similarly to the Delphi process, the quantitative data cannot be considered as a statistical representation of the technology status and therefore one needs to be careful on how to interpret the results.

4.2 Methodology

Altogether three approaches for questionnaire design and TRL assessment were presented in sections 2 and 3. The most thorough follows the guidelines of NASA/DoD's TRA process where each TRL consists of several items, which need to be achieved before a certain TRL can be reached. This approach was applied for analysing nanoscale printed electronics manufacturing technologies. It is mostly used in development of the analysis framework presented in this study.

A simpler approach was employed in the case of optical interconnects consisting of bare time estimates provided by the experts. This results in a much shorter questionnaire still making proper execution of TRA possible. The largest disadvantage is that the experts only provide a time estimate without any indications of their confidence, which makes estimation of uncertainty more difficult. TRA framework developed in this study does not currently utilise time estimates but their analysis is rather simple in a spreadsheet program.

The final approach simplified the original TRA scheme by using only one question per TRL, i.e. whether the TRL has been reached or not. A 7-point ordinal scale was used and the analysis was performed with the framework developed.

In conclusion, there is no clear answer to how the TRL status should be asked from the experts. According to table 3 in section 2, the respondents behaved similarly in each case and approximately 55% of the experts who clicked the invitation link also responded to the questionnaire. The differences in the ratio of "Clicked" invitees is dependent on the population and attractiveness of the invitation letter. These findings suggest that the questionnaire design can be altered case specifically according to expectations from the TRA results. For a general overview, it may be enough to use a simple questionnaire but in more complex cases more thorough approach can be applied.

The TRA process developed by NASA/DoD is essentially linear and results in binary estimates of TRL. This is understandable when the TRA is used for risk analysis and a certain risks are closely related to TRLs. For forecasting purposes that are a central part of this study, greater flexibility could be allowed in assigning the TRL. This could in practice mean assigning a probability for each TRL and visualising the information effectively. The visualisation is, however, supposed to be very simple and including the distributions could end up in a messy result difficult to use in practice. The process as such does not take into consideration the fact that mature technologies also develop all the time. The development can be modelled by technology impact but currently the model is only a static representation of the technology status. An estimate for TRL development is included in the visualisation and future impact estimates were asked in the printed electronics questionnaire but the experts found it difficult because the definition of a reference level is not unambiguous.

4.3 Future research

The process developed in this thesis deals with multiple disciplines. Background for the assessment framework includes technology readiness assessment by NASA/DoD, general technology assessment theories from multiple authors, decision analysis theory and elicitation and aggregation of expert opinion. Based on the findings of this study, there is no clear need for studying different analysis methods further unless more effort is put in the data collection first.

Before executing the questionnaires for the three cases presented earlier, it was very uncertain if a web questionnaire is a suitable tool for collecting expert opinion for TRA. It turned out that a large amount of both quantitative and qualitative data can be collected in such way and it is possible to find out the technological status by the process developed here. The simplest way for achieving higher quality results would be increasing the amount of experts taking part in the process and further improving the questionnaire invitation letter and structure.

To increase the amount of expert contribution the questionnaires could be used as part of a Delphi process for instance in an international conference. This would, of course, require a significant effort from the analyst, which is in contrast with the requirements specified in section 2. Another more likely scenario could be introducing periodical screening of certain technology areas, where the same group of experts would provide their feedback in e.g. 6 or 12 month intervals. This kind of approach would add a temporal dimension in TRA. Thus, one possible theme for future research would be how to improve the reliability of the analysis by getting the experts more involved in the process.

Another viable goal for future research would be to study the effect of variations in the questionnaire design to the final results. This could include both questionnaire structure and data type to be collected. Three different structures were introduced in this study and further research could try to address what are the factors in the questionnaire structure affecting response rate and what kind of questionnaire has the best balance between response rate and quality of responses. Another important part in the questionnaire design is also the model and corresponding data collection of expert uncertainty.

The subjective logic framework, for instance, allows the use of continuous scales for both belief and uncertainty for each individual item. Therefore, it would be favourable if the questionnaire allowed the experts to give estimates of their level of certainty for each individual item. One possible method for gathering this kind of information could be the Visual Analog Scale (VAS) used in psychometrics and implemented in some web survey platforms. Such method for data collection could potentially lead to simpler questionnaire structure capable of collecting more data than the current solution. Visual Analog Scale is not, however, implemented in the survey platform used and therefore its application was not possible for this thesis. Currently, most commercial survey platforms do not have an implementation for VAS which furthermore limits its applicability.

In conclusion, the framework and tool for technology readiness assessment have proven to be useful for the purpose they are designed. The process provides independent information to support traditional qualitative analysis, which definitely helps the TRA process even though the statistical significance cannot be guaranteed. This is in line with the Delphi method that is the closest counterpart in technology assessment methodology. Based on practical experience from applying the tool in real world exercises of printed electronics, optical interconnects and universal memory, the methodology works as it was intended to. According to expert reviews and manual validation of results from various sources, it seems that the estimates represent the current technology status rather accurately.

References

Aaltonen, M. (2009). Evaluation and organization of futures research methodology - version 3.0.

Altunok, T. and Cakmak, T. (2010). A technology readiness levels (TRLs) calculator software for systems engineering and technology management tool. *Advances in Engineering Software*, 41(5):769 – 778.

Bhushan, B., editor (2006). *Springer handbook of nanotechnology*. Springer,, New York.

Britt, B. L., Berry, M. W., Browne, M., Merrell, M. A., and Kolpack, J. (2008). Document classification techniques for automated technology readiness level analysis. *J. Am. Soc. Inf. Sci. Technol.*, 59(4):675–680.

Clemen, R. T. (1987). Combining overlapping information. *Management Science*, 33(3):373–380.

Clemen, R. T. (1996.). *Making hard decisions: an introduction to decision analysis*. Duxbury Press,, Belmont, CA.

Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19:187–203(17).

Daneshkhah, A. R. (2004). Uncertainty in probabilistic risk assessment: A review. A BEEP working paper, The University of Sheffield.

DoD (2009). *Technology Readiness Assessment (TRA) Deskbook*. Department of Defense.

European Space Agency (ESA) (2009). What is technology? `http://www.esa.int/SPECIALS/Technology/SEMYSRWPXPF_0.html`. [Online; Accessed 20-October-2010].

Galbraith, C., Ehrlich, S., and DeNoble, A. (2006). Predicting technology success: Identifying key predictors and assessing expert evaluation for advanced technologies. *The Journal of Technology Transfer*, 31(6):673–684.

Garthwaite, P. and Dickey, J. (1985). Double- and single-bisection methods for subjective probability assessment in a location-scale family. *Journal of Econometrics*, 29(1-2):149–163.

Glenn, J. C. and Gordon, T. J. (2009). *Futures Research Methodology Version 3.0*. The Millennium Project, 3.0 edition.

Graettinger, C. P., Garcia, S., Siviy, J., Schenk, R. J., and Syckle, P. J. V. (2002). Using the technology readiness scale to support technology management in the dod's atd/sto environments - a findings and recommendations report conducted for army cecom. Technical report, Carnegie Mellon Software Engineering Institute.

Hamilton, M. B. (2009). Online survey response rates and times. `http://www.supersurvey.com/papers/supersurvey_white_paper_response_rates.pdf`. [Online; Accessed 20-October-2010].

Hassall, J. (1999). Methods of analysing ordinal/interval questionnaire data using fuzzy mathematical principle.

Ho, D. E. and Quinn, K. M. (2008). Improving the presentation and interpretation of online ratings data with model-based figures. *The American Statistician*, 62(4):279–288.

James, A., Choy, S. L., and Mengersen, K. (2010a). Elicitator: An expert elicitation tool for regression in ecology. *Environmental Modelling & Software*, 25(1):129 – 145.

James, A., Choy, S. L., and Mengersen, K. (2010b). Elicitator: An expert elicitation tool for regression in ecology. *Environ. Model. Softw.*, 25(1):129–145.

Johnson, V. E. and Albert, J. H. (2001). *Ordinal Data Modeling (Statistics for Social and Behavioral Sciences)*. Springer, 1st ed. 1999. corr. 2nd printing edition.

Jøsang, A. (2001). A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, 9(3):279–311.

Jøsang, A. (2002). The consensus operator for combining beliefs. *Artificial Intelligence*, 141(1-2):157–170.

Jøsang, A. (2010). Subjective logic. `http://persons.unik.no/josang/papers/subjective_logic.pdf`. Unpublished draft [Online; Accessed 20-October-2010].

Jøsang, A. and Knapskog, S. J. (1998). A metric for trusted systems. *Proceedings of the 21st National Security Conference.*

Kaplowitz, M. D., Hadlock, T. D., and Levine, R. (2004). A Comparison of Web and Mail Survey Response Rates. *Public Opin Q*, 68(1):94–101.

Kuhnert, P., Hayes, K., Martin, T. G., and F., M. M. (2009). Expert opinion in statistical models. *18th World IMACS / MODSIM Congress.*

Kynn, M. (2008). The heuristics and biases bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171:239–264(26).

Lipscomb, J., Parmigiani, G., and Hasselblad, V. (1998). Combining expert judgment by hierarchical modeling: An application to physician staffing. *Management Science*, 44(2):149–161.

Liu, I. and Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 14(1):1–73.

Malhotra, N. K. (cop. 2004.). *Marketing research: an applied orientation*. Prentice Hall,, Upper Saddle River, N.J.

Mankins, J. C. (1995). Technology readiness levels. http://www.hq.nasa.gov/office/codeq/trl/trl.pdf. [Online; Accessed 20-October-2010].

Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica*, 65(9-10):1216 – 1223.

Martin, H. and Daim, T. (2007). Technology roadmapping through intelligence analysis: Nanotechnology. pages 1613 –1622.

Mast, J. D. and Wieringen, W. V. (2004). Special issue measurement system analysis for bounded ordinal data. *Quality and Reliability Engineering International*.

Merriam-Webster Online Dictionary (2010a). "science". http://www.merriam-webster.com/dictionary/science. [Online; Accessed 20-October-2010].

Merriam-Webster Online Dictionary (2010b). "technology". http://www.merriam-webster.com/dictionary/technology. [Online; Accessed 20-October-2010].

Meyer, M. A. and Booker, J. M. (1991.). *Eliciting and analyzing expert judgement: a practical guide*. Academic Press,, London.

Morris, P. A. (1974). Decision analysis expert use. *Management Science*, 20(9):1233–1241.

Morris, P. A. (1977). Combining expert judgments: A bayesian approach. *Management Science*, 23(7):679–693.

Nolte, W. L., Kennedy, B. C., and Dziegiel, R. J. J. (2003). Technology readiness level calculator. www.dtic.mil/ndia/2003systems/nolte.ppt. [Online; Accessed 20-October-2010].

ObservatoryNANO (2010a). Briefing no.4: Universal memory. http://www.observatorynano.eu/project/document/3357/. [Online; Accessed 20-October-2010].

ObservatoryNANO (2010b). Ict sector focus report: Nanophotonics for optical communications. http://www.observatorynano.eu/project/filesystem/files/ObservatoryNanoFocusReport_OpticalInterconnects.pdf. [Online; Accessed 20-October-2010].

ObservatoryNANO (2010c). Ict sector focus report: Printed electronics. http://www.observatorynano.eu/project/filesystem/files/ObservatoryNanoFocusReport_PrintedElectronics.pdf. [Online; Accessed 20-October-2010].

ObservatoryNANO (2010d). Observatorynano factsheets. http://www.observatorynano.eu/project/document/3319/. [Online; Accessed 20-October-2010].

Ouchi, F. (2004). A literature review on the use of expert opinion in probabilistic risk analysis. Policy Research Working Paper Series 3201, The World Bank.

Pearl, J. (1990). Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning*, 4:363–389.

Poole, C. P., Jones, F. J., and Owens, F. J. (2003). *Introduction to Nanotechnology*. John Wiley & Sons, Inc., New York, NY, USA.

Porter, A. L., Ashton, W. B., Clar, G., Coates, J. F., Cuhls, K., Cunningham, S. W., Ducatel, K., van der Duin, P., Georgehiou, L., Gordon, T., Linstone, H., Marchau, V., Massari, G., Miles, I., Mogee, M., Salo, A., Scapolo, F., Smits, R., and Thissen, W. (2004). Technology futures analysis: Toward integration of the field and new methods. *Technological Forecasting and Social Change*, 71(3):287–303.

Roco, M. (2001). International strategy for nanotechnology research and development. *Journal of Nanoparticle Research*, 3(5-6):353–360.

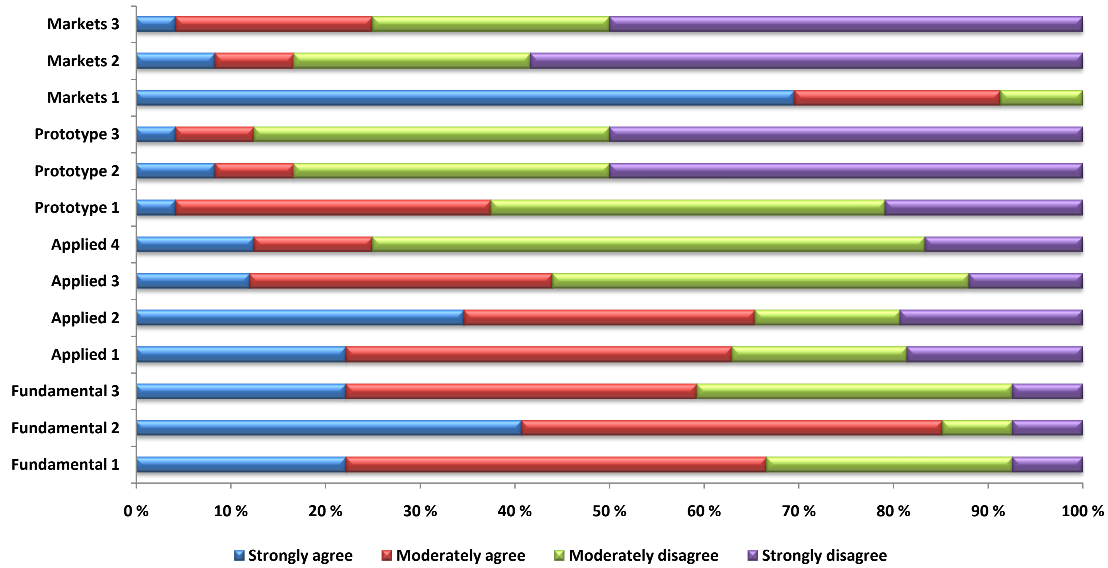
Sheehan, K. (2001). E-mail survey response rates: A review. http://jcmc.indiana.edu/vol6/issue2/sheehan.html. [Online; Accessed 20-October-2010].

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

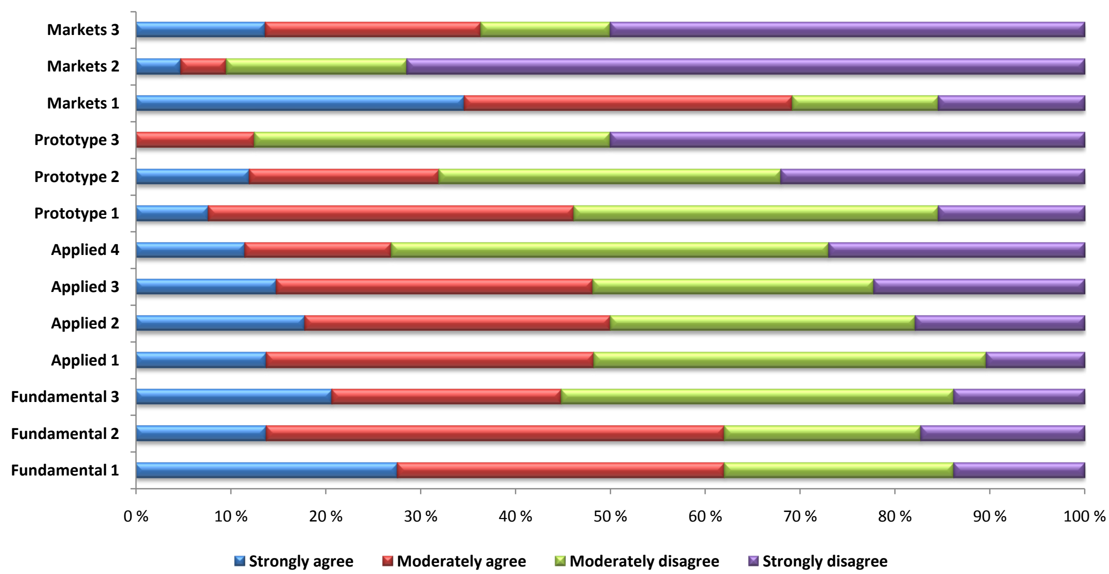
UK MOD (2010). Policy, information and guidance on the technology management aspects of uk mod defence acquisition. http://www.aof.mod.uk/aofcontent/tactical/techman/index.htm. [Online; Accessed 20-October-2010].

Winkler, R. L. (1996). Uncertainty in probabilistic risk assessment. *Reliability Engineering & System Safety*, 54(2-3):127 – 132. Treatment of Aleatory and Epistemic Uncertainty.

Appendix A

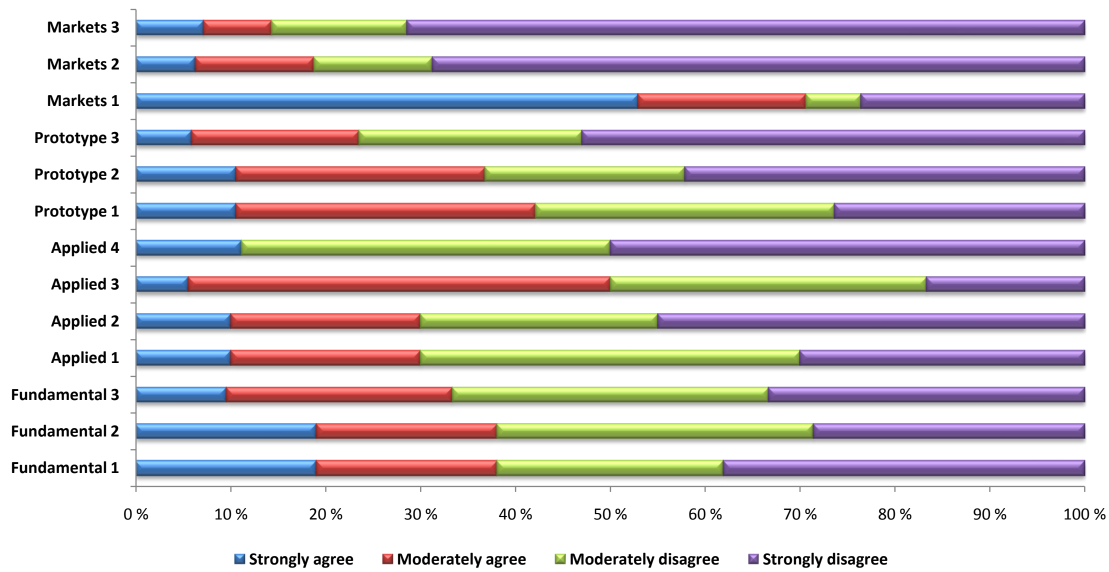


(a) Nanoimprint lithography

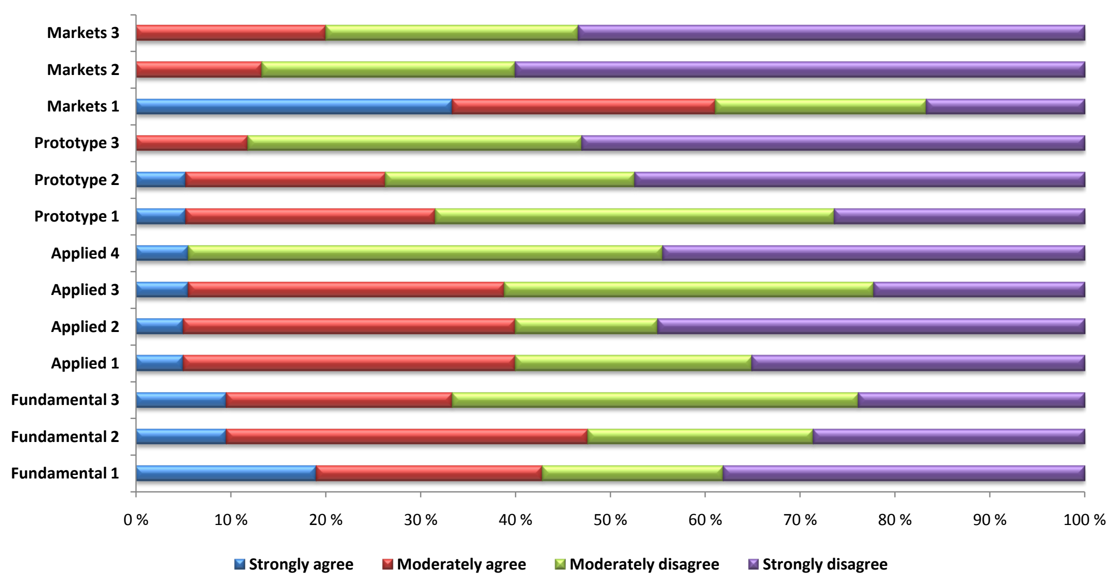


(b) Ink-jet printing

Figure A.1: Data on TRL related claims for printed electronic manufacturing technologies: Nanoimprint lithography (a) and ink-jet printing (b)



(a) Gravure printing



(b) Flexography printing

Figure A.2: Data on TRL related claims for printed electronic manufacturing technologies: gravure printing (a) and flexography printing (b)

Appendix B

Table B.1: A simple taxonomy of Futures Research Methods (Glenn and Gordon, 2009)

Method	**Quantitative**	**Qualitative**	**Normative**	**Exploratory**
Agent Modeling		X		X
Causal Layered Analysis		X		X
Chaos and Non-Linear Systems	X			X
Cross-Impact Analysis	X			X
Decision Modeling	X			X
Delphi Techniques		X	X	X
Econometrics and Statistical Modeling	X			
Environmental Scanning		X		X
Field Anomaly Relaxation		X		X
Futures Polygon	X	X	X	X
Futures Wheel		X	X	X
Genius Forecasting, Vision, and Intuition		X	X	X
Interactive Scenarios		X	X	X
Morphological Analysis		X	X	
Multiple Perspective		X	X	X
Participatory Methods		X	X	
Prediction Markets	X		X	
Relevance Trees		X	X	
Robust Decisionmaking	X			X
Scenarios	X	X	X	X
Science and Technology Roadmapping	X	X	X	X
Simulation-Gaming		X		X
State of the Future Index	X	X	X	X
Structural Analysis	X	X		X
Substitution Analysis				
Systems Modeling	X			X
Technological Sequence Analysis		X	X	
Text Mining		X	X	X
Trend Impact Analysis	X			X
Visioning		X	X	
Wild Cards	X	X		X

Appendix C

Printed electronics questionnaire with the following manufacturing technologies assessed:

- Roll-to-roll nanoimprinting
- Inkjet printing
- Gravure printing
- Flexography printing

1. How familiar are you with the following technologies?
    - I am (or have been) working with or studying this technology
    - I know how the technology works
    - I am not familiar with this technology
2. Fundamental research
    (a) Physical laws and phenomena behind this technology for printing nanoscale features are known?
    - Strongly agree
    - Moderately agree
    - Moderately disagree
    - Strongly disagree

    (b) Benefits of this technology in nanoscale printed electronics for ICT applications are known?

    (c) Critical elements (functions/components/materials/inks) of this technology enabling ICT applications have been identified?
3. Applied research: Component and system level proof-of-concept
    (a) Critical elements (functions/components/materials/inks) are available for printing proof-of-concept nanoscale features for ICT applications?

    (b) This technology has been demonstrated in printing proof-of-concept nanoscale features for ICT applications?

    (c) End user requirements for the ICT applications have been defined?

    (d) Performance of proof-of-concept models is on required level?
4. Prototype
    (a) Final operating environment and external interfaces for ICT applications have been defined?

    (b) Prototype products with all required functionalities have been printed with this technology?

    (c) Performance of prototype products in the final operating environment is on required level ?
5. Market phase
    (a) There is an ICT market need for nanoscale electronics printed with this technology?

    (b) First products for ICT applications containing nanoscale features printed with this technology have entered the market?

    (c) Multiple vendors are providing products printed with this technology?
6. Please indicate the year when you expect the listed Technology Readiness Levels to be reached
    (a) Fundamental research: Basic elements and technology concept
    - Now

- 1-3 years
- 4-7 years
- Later
- Never

(b) Applied research: Component and system level proof-of-concept

(c) Prototype: Functional prototypes built

(d) Markets: Market entry

(e) Markets: Mature markets

7. What is the current status of technologies listed compared to the current state-of-the-art printed electronics manufacturing technologies?

    (a) Manufacturing process enables novel features in nanoscale printed electronics ICT applications
        - Very good
        - Good
        - Neutral
        - Bad
        - Very bad

    (b) Performance (e.g. speed, linewidth, yield, reliability) of the manufacturing process

    (c) Cost effectiveness of the manufacturing process

    (d) Scalability of the manufacturing process

    (e) Environment, Health and Safety (EHS) aspects of the manufacturing process

8. Please estimate the status of technologies listed when they hit the market compared to the current state-of-the-art printed electronics manufacturing technologies?

    (a) Manufacturing process enables novel features in nanoscale printed electronics ICT applications

    (b) Performance (e.g. speed, linewidth, yield, reliability) of the manufacturing process

    (c) Cost effectiveness of the manufacturing process

    (d) Scalability of the manufacturing process

    (e) Environment, Health and Safety (EHS) aspects of the manufacturing process

9. What kind of ICT applications could need printed nanoscale features? Comments?

10. Are there any other significant impacts or issues related to technologies assessed?

11. Please, give general feedback on technologies assessed or the questionnaire here

Appendix D

Table D.1: Full results of method comparisons for the printed electronics data

		Basic multinomial	Weighed multinomial	SL Averaging	SL Medium	SL Basic	Ratings	ICC	Judge-specific
1	NIL	67 %	68 %	63 %	66 %	70 %	61 %	63 %	66 %
	INK	62 %	67 %	65 %	70 %	75 %	56 %	61 %	65 %
	GRA	38 %	36 %	33 %	29 %	26 %	36 %	38 %	40 %
	FLE	41 %	32 %	32 %	28 %	24 %	36 %	38 %	43 %
2	NIL	85 %	89 %	72 %	82 %	91 %	73 %	78 %	85 %
	INK	62 %	63 %	57 %	61 %	65 %	53 %	53 %	61 %
	GRA	38 %	38 %	45 %	40 %	36 %	42 %	42 %	35 %
	FLE	45 %	50 %	47 %	48 %	49 %	41 %	40 %	43 %
3	NIL	59 %	63 %	60 %	62 %	64 %	60 %	60 %	59 %
	INK	45 %	46 %	54 %	52 %	50 %	46 %	51 %	43 %
	GRA	33 %	29 %	33 %	28 %	23 %	36 %	34 %	32 %
	FLE	32 %	23 %	31 %	25 %	19 %	38 %	35 %	30 %
4	NIL	63 %	68 %	63 %	68 %	74 %	58 %	57 %	63 %
	INK	48 %	52 %	56 %	57 %	58 %	48 %	51 %	46 %
	GRA	30 %	30 %	35 %	29 %	24 %	33 %	35 %	32 %
	FLE	38 %	33 %	35 %	34 %	33 %	35 %	32 %	40 %
5	NIL	65 %	78 %	75 %	83 %	91 %	60 %	64 %	65 %
	INK	50 %	57 %	58 %	61 %	65 %	40 %	50 %	44 %
	GRA	30 %	29 %	29 %	26 %	22 %	25 %	28 %	22 %
	FLE	38 %	29 %	29 %	26 %	24 %	26 %	28 %	31 %
6	NIL	44 %	44 %	48 %	47 %	47 %	49 %	48 %	44 %
	INK	48 %	54 %	51 %	53 %	54 %	47 %	46 %	48 %
	GRA	50 %	50 %	51 %	55 %	58 %	43 %	45 %	47 %
	FLE	37 %	45 %	50 %	53 %	56 %	43 %	38 %	34 %
7	NIL	25 %	28 %	44 %	37 %	31 %	39 %	38 %	26 %
	INK	27 %	30 %	44 %	41 %	38 %	36 %	34 %	31 %
	GRA	11 %	10 %	21 %	13 %	5 %	17 %	21 %	13 %
	FLE	5 %	0 %	18 %	9 %	1 %	17 %	16 %	7 %
8	NIL	38 %	36 %	45 %	41 %	38 %	39 %	38 %	37 %
	INK	46 %	50 %	47 %	49 %	51 %	41 %	45 %	44 %
	GRA	42 %	37 %	39 %	34 %	29 %	34 %	41 %	41 %
	FLE	30 %	29 %	40 %	36 %	32 %	34 %	33 %	28 %
9	NIL	17 %	17 %	31 %	24 %	17 %	24 %	19 %	17 %
	INK	32 %	36 %	43 %	39 %	36 %	35 %	34 %	32 %
	GRA	37 %	35 %	39 %	37 %	35 %	28 %	32 %	35 %
	FLE	25 %	25 %	34 %	32 %	30 %	26 %	21 %	22 %
10	NIL	13 %	16 %	30 %	23 %	17 %	22 %	16 %	13 %
	INK	13 %	12 %	25 %	19 %	12 %	22 %	15 %	14 %
	GRA	24 %	16 %	21 %	15 %	8 %	19 %	22 %	25 %
	FLE	11 %	5 %	19 %	11 %	3 %	19 %	13 %	12 %
11	NIL	91 %	88 %	80 %	83 %	86 %	78 %	93 %	91 %
	INK	69 %	78 %	68 %	74 %	81 %	68 %	66 %	65 %
	GRA	71 %	72 %	68 %	72 %	77 %	57 %	70 %	67 %
	FLE	63 %	71 %	67 %	73 %	79 %	58 %	63 %	59 %
12	NIL	17 %	20 %	30 %	26 %	21 %	21 %	15 %	17 %
	INK	10 %	0 %	13 %	7 %	1 %	14 %	8 %	10 %
	GRA	19 %	12 %	16 %	12 %	9 %	14 %	14 %	19 %
	FLE	13 %	6 %	14 %	9 %	4 %	14 %	11 %	13 %
13	NIL	25 %	25 %	37 %	34 %	30 %	25 %	20 %	25 %
	INK	36 %	33 %	31 %	31 %	31 %	30 %	28 %	38 %
	GRA	14 %	7 %	14 %	9 %	4 %	16 %	12 %	14 %
	FLE	19 %	17 %	25 %	21 %	17 %	20 %	15 %	19 %