Role of spatial data uncertainty in executions of precision farming operations

Jere Kaivosoja
Role of spatial data uncertainty in executions of precision farming operations

Jere Kaivosoja

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Engineering, at a public examination held at the lecture hall M1 of the school on 13th December 2019 at 12 pm.

Aalto University
School of Engineering
Department of Built Environment
Supervising professor
Professor Kirsi Virrantaus, Aalto University, Finland

Thesis advisor
Dr. Jussi Nikander, Aalto University, Finland

Preliminary examiners
Associate Professor Lammert Kooistra, Wageningen University, Netherlands
Dr. Michael Norremark, Aarhus University, Denmark

Opponent
Dr. Michael Norremark, Aarhus University, Denmark
### Abstract

**Author**  
Jere Kaivosoja

**Name of the doctoral dissertation**  
Role of spatial data uncertainty in executions of precision farming operations

**Publisher**  
School of Engineering

**Unit**  
Department of Built Environment

**Series**  
Aalto University publication series DOCTORAL DISSERTATIONS 217/2019

**Field of research**  
Geoinformatics

<table>
<thead>
<tr>
<th>Manuscript submitted</th>
<th>Date of the defence</th>
<th>Permission for public defence granted (date)</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 May 2019</td>
<td>13 December 2019</td>
<td>23 October 2019</td>
<td>English</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monograph</th>
<th>Article dissertation</th>
<th>Essay dissertation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abstract**

The topic of this dissertation is the quality of spatial data related to arable farming work executions. There is a lot of uncertainty related to farming operations. Whether the field was treated evenly or with variable rate application, the performed field work is often inaccurate and vague. This is mainly because driving patterns in the field are not accurate and the optimal machinery input amount needs are uncertain. It is important to study this uncertainty in order to understand its impact on precision farming. This dissertation approaches the problem of uncertainty by applying geographical information quality evaluation and measurement methods to the evaluation of farm work execution. This makes it possible to exploit existing evaluation principles for spatial data in farm work execution cases. ISO 19157:2013 standard defines the key elements in spatial data quality.

By studying agricultural technology research case studies, the effect of the quality elements and the estimated overall accuracy level of the farming work execution in grain production were constructed. The different case studies integrated remote sensing technologies such as satellite images and drones with cameras and hyperspectral technologies together with different spatial farm data and farmer’s tacit knowledge. The case studies included also positioning errors based on dynamic GNSS positioning accuracy measurements and simulations and measured field work driving accuracies of farmers. The temporal quality was studied by developing methods how to apply real time data from external sources in ISOBUS environment.

The machinery driving lines were overlapping by 10 % on average. Accurate steering assistance can cut that in half but there are still remarkable overlaps especially in headland areas. The biggest difficulty is the optimization of the variable rate application levels meaning the thematic accuracy. The thematic accuracy was determined as the variation of different tasks conducted for the same purpose being 22 % on average. The temporal accuracy was completely a case dependent containing a response to the immediate rain forecasts or applicability of one month old satellite image. A single precision farming operation was estimated to benefit about 31 €/ha which was estimated to be only 23 % of the total precision farming benefit potential according to the variables in this work. The overall accuracy of spatial data inputs was estimated to be 61 % in relation to optimal treatment in the studied cases. This number indicates the quality of spatial data inputs to farm machinery. This uncertainty is large in contrast to typical attempted precision farming adjustments and defined machinery performance requirements. These results suggest that there is a need for better uncertainty management, before different precision farming applications can truly be developed and evaluated.

**Keywords** spatial data quality, GNSS, UAV, agriculture, uncertainty, precision farming


**ISSN (printed)** 1799-4934  
**ISSN (pdf)** 1799-4942

**Location of publisher** Helsinki  
**Location of printing** Helsinki  
**Year** 2019

**Pages** 150  
Acknowledgements

Eventually. Never have I doubted this, but sure there were always something more urgent or easier to grab on. This has been a quite freely drifting task while coping with innumerable amount of project applications, several managed projects and writing dozens of other publications in different topics. There are many people to thank, and I hereby wish to sincerely acknowledge at least most who has contributed to this work in one way or another. First, I take this opportunity of sending thanks to all my colleagues in who were working in Vihti and express my grateful regards to you. You introduced me to the agriculture. Pasi Suomi, Liisa Pesonen and Raimo Linkolehto, we have made a very long journey in precision farming research starting from the year 2003. You made this possible. I would also say special thanks to Raimo for your essential contribution to the appended articles.

Next I like to thank my Advisor Dr. Jussi Nikander. You really helped me with the structure and focus of the work repeatedly. It was trustful to work with you. I owe sincere thanks to my supervisor, Professor Kirsi Virrantaus, for activating me to this work in 2017 and for her valuable comments and words of encouragement during the writing. Her insightful suggestions and constructive comments have greatly improved this summary.

In this topic that combines agriculture and geoinformatics, I am highly grateful to my pre-examiners Dr. Michael Nørremark and Associate Professor Lammert Kooistra for taking the time to read my manuscript and providing helpful comments for improvements. I highly appreciate that Dr. Nørremark also agreed to act as an opponent.

I would like to thank the Spectrophotogrammetry team in Finnish Geospatial Research institute led by Dr. Eija Honkavaara. I would also thank my colleagues at Tampere, the cosy environment, interesting discussions and especially the peer pressure were something that pushed me through the last mile. I am also grateful to my current immediate manager Dr. Lauri Sikanen and all the six former ones, who all were positively supportive in LUKE (formerly known as MTT). I also thank my colleague Dr. Juha Backman for taking the time to read my manuscript to improve details before printing. I would also acknowledge all the others who contributed to the appended papers and all of my friends and sport teams keeping mind that I am responsible for the content of this dissertation.

I would like to thank my parents and my sisters. From my father, I learned to never give up. My mother made me to believe in myself and my two sisters Emilia and Kristiina showed me how to get things done. Last, I am sincerely grateful to my wife Anna for giving her best support and love during these years and my two kids, you mean everything to me and after all, you made me want to get this done.

Lempäälä, 7 November 2019
Jere Kaivosoja
List of Publications

This doctoral dissertation consists of a summary and of the following publications which are referred to in the text by their Roman numerals.

I. Kaivosoja, Jere; Pesonen, Liisa; Kleemola, Jouko; Pölönen, Ilkka; Salo, Heikki; Honkavaara, Eija; Saari, Heikki; Mäkynen, Jussi and Rajala, Ari; 2013. **Case Study of a Precision Fertilizer Application Task Generation for Wheat Based on Classified Hyperspectral Data from UAV Combined with Farm History Data.** SPIE Remote Sensing for Agriculture, Ecosystems, and Hydrology XV, 88870H (16 October 2013); doi: 10.1117/12.2029165

II. Kaivosoja, Jere; Jackenkroll, Marcus; Linkolehto, Raimo; Weis, Martin and Gerhards, Roland. 2014. **Automatic control of farming operations based on spatial web services.** Computers and Electronics in Agriculture 100: 110-115. https://doi.org/10.1016/j.compag.2013.11.003


Authors’s Contribution

Publication I, A Case Study of a Precision Fertilizer Application Task Generation for Wheat Based on Classified Hyperspectral Data from UAV Combined with Farm History Data. This publication studied advanced possibilities of exploiting classified raster maps originating from the hyperspectral data cube to produce an application task for a precision fertilizer application aiming for a higher yield. The author is responsible for all of the parts in this conference proceeding and has written 90% of it. All co-authors made overall comments about case study concept. Pesonen supported with the introduction section text and with the reference data collection plans. Pesonen, Kleemola and Rajala supported with the agronomic knowledge details in the overall text. Saari and Mäkynen provided the prototype hyperspectral instrument, related setups. Honkavaara corresponded the hyperspectral image data processing and mosaicking. Pölönen and Salo produced biomass and nitrogen classifications out of spectral datasets.

Publication II, Automatic Control of Farming Operations Based on Spatial Web Services. This study presented a task controller commanding ISOBUS sprayer and using external spatial data during a task execution. The author is responsible for main parts of this article and has written 80% of it. The co-authors Weis and Jackenkroll in the supervision of Gerhards were responsible for the external spatial web service and the related text sections including the final design of the first two figures. Co-author Linkolehto was responsible for the design and programming work of the task controller prototype developed and studied in this work. All authors contributed to the general level system architecture.

Publication III, GNSS Error Simulator for Farm Machinery Navigation Development. This study introduces a new tool for supporting the development of sensor fusions in steering and navigation automation by implementing Allan variance methodology and spatial autocorrelation determination to reveal positioning error characteristics based on dynamic tests in farming conditions. The presented simulator is capable of simulating autonomous GNSS positioning in bad conditions and in normal field conditions, sub-meter level differential (DGPS) positioning and nominally high accurate RTK (real time kinematics) GNSS positioning in various conditions. The author is responsible for all of the parts in this article and has written 100% of it. The co-author Linkolehto is responsible for data collection system design and construction and participated in data collection procedure.

Publication IV, Spatial Overlapping in Crop Farming Works. In this study, new data mining methods for finding out the average overwork percentages of regular grain farming practices within different complete fields were developed. Also by using spatial analyses, the sources of overlapping in spraying works were determined in order to be able to estimate the positive impact that can be gained with technology adaptation. The author is responsible for all of the parts in this article and has written 100% of it. The co-author Linkolehto is responsible for related data collection system design and construction.
Publication V, Different Remote Sensing Data in Relative Biomass Determination and in Precision Fertilization Task Generation for Cereal Crops. The main goal in this publication was to demonstrate in real conditions, how much difference there are in biomass estimations in contrast to actual fertilization task variations. The author is the main responsible for this article and has written 100% of it. The co-authors Näsi, Hakala, Viljanen and Honkavaara participated in data collection, managed the hyperspectral and photogrammetric data collection, processing and related pre-classifications.
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEV</td>
<td>Allan Standard Deviation</td>
</tr>
<tr>
<td>CAN</td>
<td>Controller Area Network</td>
</tr>
<tr>
<td>CTF</td>
<td>Controlled Traffic Farming</td>
</tr>
<tr>
<td>DOP</td>
<td>Dilution of Precision</td>
</tr>
<tr>
<td>DSS</td>
<td>Decision Support System</td>
</tr>
<tr>
<td>FMIS</td>
<td>Farm Management Information System</td>
</tr>
<tr>
<td>FN</td>
<td>Flicker noise</td>
</tr>
<tr>
<td>FTP</td>
<td>File Transfer Protocol</td>
</tr>
<tr>
<td>GCP</td>
<td>Ground Control Point</td>
</tr>
<tr>
<td>GM</td>
<td>First order Gauss Markov process</td>
</tr>
<tr>
<td>GML</td>
<td>Geography Markup Language</td>
</tr>
<tr>
<td>GNSS</td>
<td>Global Navigation Satellite System</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HP</td>
<td>High Performance</td>
</tr>
<tr>
<td>I-ECU</td>
<td>Electronic Control Unit of the Implement</td>
</tr>
<tr>
<td>LBS</td>
<td>Location Based Service</td>
</tr>
<tr>
<td>LPIS</td>
<td>Land Parcel Identification System</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NMEA</td>
<td>National Marine Electronics Association</td>
</tr>
<tr>
<td>NSSDA</td>
<td>National Standard for Spatial Data Accuracy</td>
</tr>
<tr>
<td>OGC</td>
<td>Open Geospatial Consortium, Inc.</td>
</tr>
<tr>
<td>OL</td>
<td>Outliers</td>
</tr>
<tr>
<td>PA</td>
<td>Precision Agriculture</td>
</tr>
<tr>
<td>PDM</td>
<td>Process Data Message</td>
</tr>
<tr>
<td>PF</td>
<td>Precision Farming</td>
</tr>
<tr>
<td>PID</td>
<td>Proportional-Integral-Derivative</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectral Density</td>
</tr>
<tr>
<td>RPAS</td>
<td>Remotely Piloted Aircraft System</td>
</tr>
<tr>
<td>RR</td>
<td>Rate Ramp</td>
</tr>
<tr>
<td>RGB</td>
<td>Red Green Blue</td>
</tr>
<tr>
<td>RTK</td>
<td>Real Time Kinematics</td>
</tr>
<tr>
<td>RW</td>
<td>Random Walk</td>
</tr>
<tr>
<td>TC</td>
<td>Task Controller</td>
</tr>
<tr>
<td>UAS</td>
<td>Unmanned Aerial System</td>
</tr>
<tr>
<td>UAV</td>
<td>Unmanned Aerial Vehicle</td>
</tr>
<tr>
<td>VGI</td>
<td>Volunteered Geographic Information</td>
</tr>
<tr>
<td>VRA</td>
<td>Variable Rate Application</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>VBS</td>
<td>Virtual Base Station</td>
</tr>
<tr>
<td>WFS</td>
<td>Web Feature Service</td>
</tr>
<tr>
<td>WMS</td>
<td>Web Map Service</td>
</tr>
<tr>
<td>WN</td>
<td>White Noise</td>
</tr>
<tr>
<td>WPS</td>
<td>Web Processing Service</td>
</tr>
</tbody>
</table>
1. Introduction .................................................................................. 1
  1.1 Background and motivation ...................................................... 2
  1.2 Structure of this dissertation ..................................................... 4
  1.3 Farm machinery field works ..................................................... 4
    1.3.1 Farm Management Systems .......................................... 4
    1.3.2 Farming regulations ...................................................... 5
    1.3.3 Precision Farming and Smart Farming ........................ 5
    1.3.4 Decision support systems .............................................. 6
    1.3.5 Remote sensing in agriculture ...................................... 7
    1.3.6 Remote sensing applications ........................................ 7
    1.3.7 Timing of the field operations ....................................... 9
  1.4 Input data for the farm machinery ............................................ 9
    1.4.1 Application tasks ......................................................... 10
    1.4.2 ISOBUS environment ................................................... 11
    1.4.3 Positioning of the machinery ...................................... 12
    1.4.4 Defined inputs ............................................................. 12
  1.5 Spatial input data in farm machinery ..................................... 13
  1.6 Uncertainty and quality in spatial data .............................. 14
  1.7 Related research .................................................................... 15
    1.7.1 Decision support and application tasks ....................... 15
    1.7.2 Temporal accuracy of spatial data in work task .......... 16
    1.7.3 Positioning error determination .................................. 16
    1.7.4 Spatial data uncertainty and quality research .............17
    1.7.5 Research topic background ........................................ 18
  1.8 Objectives and Research Questions ....................................... 19
2. Theoretical foundations .............................................................. 22
  2.1 Measuring spatial uncertainty ................................................ 22
    2.1.1 Spatial data quality ...................................................... 23
  2.2 Spatial data quality elements ............................................... 23
  2.3 Thematic accuracy .................................................................. 25
  2.4 Temporal quality ..................................................................... 25
Introduction

2.5 Positional accuracy ................................................................. 25
  2.5.1 Allan variance ............................................................... 26
  2.5.2 Spatial autocorrelation ................................................... 27
  2.5.3 Field coverage ............................................................... 27

2.6 Completeness ........................................................................ 28

2.7 Usability element .................................................................... 29

2.8 Logical consistency .................................................................. 29

2.9 Overall quality measure .......................................................... 29

3. Material and Methods ................................................................. 31
  3.1 General data collection methodology ...................................... 32
  3.2 The application task variations and the thematic accuracy.... 33
  3.3 Temporal quality of the machinery input data ....................... 34
  3.4 Positional accuracies ............................................................. 34
  3.5 Positioning and machinery location ....................................... 35
  3.6 Completeness ......................................................................... 37
  3.7 Usability .................................................................................. 37
  3.8 Overall quality and quality measures ..................................... 38

4. Results ........................................................................................ 40
  4.1 Developed tools ..................................................................... 40
  4.2 Thematic accuracy ................................................................... 42
  4.3 Temporal accuracy .................................................................. 45
  4.4 Positional accuracy ................................................................. 46
  4.5 Completeness .......................................................................... 50
  4.6 Usability ................................................................................... 51
  4.7 Overall quality ......................................................................... 51

5. Discussion .................................................................................... 55

6. Conclusions ................................................................................ 60

References ........................................................................................ 63
1. Introduction

Arable farming means growing crops in fields. Common arable crops in Finland are barley, wheat, oat and rye. The farming process goes as follows: first the bare field is typically ploughed, then cultivated and after that the seeds are planted in the soil. All of these operations are made with a tractor together with specific machinery. Fertilizers are added simultaneously with the planting, or afterwards during the crowing season. Pests and diseases are often controlled by spraying with chemicals done with the tractor-machinery combination. Crops are harvested in autumn with a combine harvester. The large-scale arable farming has been made possible by the machinery technology in the 20th century but the typical traditional farming has not changed much since the mid-twentieth century. The required labour hours per hectare are decreasing, machinery is getting bigger and smarter but the agricultural landscape is similar in practice. The big machinery treats large fields in a uniform which is not a sustainable way since there are lots of variations within-field (Heege, 2013). To do it better, knowledge about correct methods is required. However, information revolution in farming is progressing slowly. This is due the concrete existing machinery capabilities, data integration difficulties and knowledge gaps in the crop growth management.

According to FAO (Food and Agriculture Organization of the United Nations) (FAO, 2017), agriculture in 2050 will need to produce almost 50 percent more food than in 2013 and rapid technological development offers the prospect of meeting future food needs. Understanding the uncertainty related to the implementation of knowledge into farming processes is needed in order to define future directions in the technological development of arable farming: how well can we guide the farm machinery to perform the field operations?

The farming conditions between different fields and within-field can vary a lot. For example soil type, field slopes, water management, nutrition resources, seasonal weather conditions and used crop variety mainly causes these variations. Crop research is simply using small randomized field plot trials in order to minimize the effect of these variations. In real farming, a lot of knowledge is required to be able to practice arable farming in a cost-effective way at optimal time. Within-field management is called precision farming (PF), where different parts of the field are treated differently in order to fulfil known needs of each location in the field. This is strongly linked to geoinformatics science.

In addition to field variations, there is lots of spatial uncertainty in the quality of arable farming field work performances. Despite the aim is to threat the field evenly or by precision farming manner, the work result is often spatially imprecise: most commonly the machinery
driving lines are partially overlapping. This causes major difficulties in the field knowledge management and the development of farming processes especially because usually the imprecise field actions are not recorded.

Spatial data uncertainty is an umbrella term for spatial data errors, randomness, and vagueness aspects (Zhang and Goodchild, 2002). The uncertainty implies that there is something we are not sure of in spatial data and analysis due to various reasons, such as measurement errors, data generalization, ignorance of knowledge and incomplete representation of all factors in analysis (Li, 2017). The work on uncertainty modelling is required to deal with the inevitable impacts of uncertainty on decision-making (Zhang and Goodchild, 2002).

This dissertation approaches the farm machinery work result uncertainty problem from the spatial data quality perspective by studying different spatial quality elements in farm machinery works. The quality evaluations provide tools for measuring uncertainty. The research encompasses the main machinery independent factors that effect on the spatial quality of arable farming field works: driving accuracy and positioning quality, the quality of precision farming execution tasks, the temporal elements in the execution process and the overall spatial data input quality for farm machinery. These recognized factors and measures can give access to better economic estimations, machinery hardware and software development requirements and more efficient and smart precision farming development. The addressed elements are also completing the precision farming knowledge in order to better meet the crop research environment where external variables are controlled.

1.1 Background and motivation

Slow changes in arable farming are happening in terms of precision farming. The lack of appropriate sensors and configurations (Sylvester-Bradley et al., 1999) has long been identified as the main obstacle. Even there has been fast development with the sensor technologies there is still a big problem about what should and can really be measured. There is seldom a possibility to apply direct measurements. For example, when optimizing additional nitrogen fertilization amounts is the target, one could use information about plants existing nitrogen level, biomass amount, current heat sum and a forecast, existing nitrogen level in the soil, the level of other nutrients, the yield potential of the soil at the current location, water resources at current point and so on. The requested nitrogen amount is a composition of different variables (Diacono et al., 2012). The nitrogen fertilization for cereal crops does not follow any kind of generalized methodology that guarantees maximum nitrogen use efficiency (Raun et al., 2005). Often the precision farming actions are made based only on one or few data sources combined with indirect prediction models or farmer’s tacit knowledge. There is not much information about variations these are causing.

Other critical element in the both precision and traditional farming is the accuracy of the work that has been done. The inaccurate working decreases the positive effect of precision farming and increases the unwanted variations (Grift et al., 2009) within-field. The farmers tend to overlap their work about 10 % (Nieminen and Sampo, 1993, Griffin et al., 2005). The driving results can be improved (Lipinski et al., 2016) by adapting mostly GNSS (Global Navigation Satellite System) based steering assistance and automation systems. However, the true dynamic accuracy and affordability of those systems are not well known.

So, the field works are done rather inaccurately and the quality of precision farming application tasks are vague. These facts make it difficult to develop and truly evaluate the preci-
sion farming benefits and to determine the machinery and sensor development requirements. In Australia, Lawes and Robertson (2011) found that precision farming applications gave a 10 €/ha economic payoff. As a single action, Meyer-Aurich et al. (2010) measured precision nitrogen fertilization potential in Germany. Their data suggest gross economic potential of 2.6 €/ha for winter wheat. An average Finnish farm has only about 50 ha of arable land and in relation to that, these economic numbers are rather small. There is typically a much greater variation in the yield amount and quality (Waheed et al., 2006) in field trials so it would be essential to increase knowledge about the quality of data involved in farm work execution. So far there have not been any tools for this evaluation in the agricultural science domain. However, this data is basically all spatial data and thus it obeys the common rules of spatial data behaviour.

At the general level, spatial data is simply any data that has a location and an attribute data. More precisely, spatial data also known as geospatial data or geographic information is information about an object that can be represented by numerical values in a geographic coordinate system. It is a characteristic that describes the location, geometrics and topology of a spatial object. Spatial data is represented in a vector or raster format and is often accessed, manipulated or analysed through Geographic Information Systems (GIS). In spatial sciences or geographic information sciences, there are multiple tools and methods for data analysis.

If it is possible to get information on uncertainties related to spatial data and track the uncertainty propagation, the obtainable accuracy can be accessed (Goodchild, 1991). The key problem is the need for the effective identification of the sources of uncertainty (Zhang and Goodchild, 2002). Although spatial data users may not be aware of the uncertainty in all the datasets, it is critical to evaluate data quality in order to understand the validity and limitations of any conclusions based on spatial data (Li, 2017). Spatial data quality is an evaluation of the similarity between spatial data and geographical truth, including both positional truth and attribute truth (Li, 2017). The closer spatial data is to the truth, the higher is its quality. According to Laskey et al. (2010), systematic approach to data quality requires estimating and describing the quality of data as they are collected. This means recording data quality as metadata, managing the uncertainty propagation in data processing models and then exploiting uncertainty by the end user (Laskey et al., 2010). This dissertation is based on the spatial data quality and uncertainty research and standardization. The quality evaluations provide concrete measures about the uncertainty in farm machinery.

There is already a long history in spatial data quality research in many topics. The articles Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities (Devillers et al., 2010), Fundamentals of Spatial Data Quality (Devillers and Jeansoulin, 2010) and Elements of Spatial Data Quality (Guptill and Morrison, 1995) present the history behind. Researchers have already increasingly been exploring the relationship between data quality and decision-making processes in large and heterogeneous spatial data with complex lineages, whose quality may be undocumented, mixed, or unknown (Li et al., 2012). These notifications also suit well for the quality estimation of the farm machinery work execution.

There is a need for connecting the spatial data quality research together with the farm machinery work quality evaluations to be able to indicate the significance of different elements in the work execution. This approach will quantify the uncertainty related to spatial data inputs for farm machinery such as how good are the instructions for the machinery. When we combine that with the performance of the machinery itself, we can concretize the difference between field trials and entire field applications.
1.2 Structure of this dissertation

This dissertation consists of the summary part and five publications. In Chapter 1, the summary first introduces the topic and then Chapter 2 addresses the theoretical foundations of this dissertation. The following Chapter 3 presents the materials and methods used, and Chapter 4 presents the results that are significant for this dissertation from all included publications and summarise the overall picture of quality. These results are discussed in Chapter 5 and Chapter 6 concludes the dissertation. Five publications are appended in the latter part of the thesis. In the next steps, we introduce the concept of field works in arable farming, the spatial data in farming context, the definition of data inputs to the farm machinery, the uncertainty concept and then related research is introduced.

1.3 Farm machinery field works

In the crop farming, the driver aims to work evenly the entire field, first by circulating the field a few times near field boundaries to produce headlands and then by driving parallel straight lines back and forth. The positional accuracy of these driving lines is very important: it causes null or double treatments in the field producing unwanted effects in the field. The overlapping work increases driving distance, takes more time, increases soil compaction, consumes more farming inputs and increases environmental load. The overlapping work can be reduced by improving the driving accuracy and/or by controlling the implement and its sections more accurately. Different methods for improving the driving accuracy on consecutive rows have now been developed almost a century. Currently, these methods include different track marking devices, using permanent tracks (controlled traffic farming CTF, Yule, 1998), using an electronic steering assistance or an automatic steering system, typically by applying GNSS positioning based systems. These electronic steering automation methods have widely been developed (Mousazadeh, 2013). In the precision farming, the driving accuracy challenge is similar but on top of that, the work amount of the farm machinery is spatially adjusted. The CTF approach is not considered in this research. It requires a different strategic level approach where farm machinery working widths are equal or multiplied and soil conditions are suitable.

Besides the driving, the field operations on arable farming often require very data intensive and thorough planning. Changing conditions may cause various difficulties even when the operational plan is made with proper preparations. One of the key concerns of the farm managers as summarized by Sørensen et al. (2010) is that monitoring of field operations is time consuming and that there is a need for additional information and advanced technologies to manage monitoring and data acquisition online in the field. This has led to the Farm Management Systems (FMS) and new available technologies in close relation to the farming process. The productization for agricultural purposes is continuously developed.

1.3.1 Farm Management Systems

To be able to operate with digital farm data from a single interface, Farm Management Systems (FMS) also called as Farm Management Information Systems (FMIS) have widely been developed. Their idea is to support management decision making and compliance to management standards by means of storing and processing of strategic, tactical, operational and evaluation data together with up to date regulations and compliance standards for crop pro-
duction and animal husbandry. The integrated features between different FMSs can vary a lot. According to review of Tummers et al. (2019), the most dominant features in FMSs were financial management, reporting features, data acquisition, operation plan generation and crop management. Wolfert et al. (2017) concluded that data quality has always been a key issue in the FMS. Tummers et al. (2019) identified the current biggest obstacles with FMSs being lack of standardized data formats and system integration difficulties.

Typically many of the agricultural production principles and standards are already included in the FMS. However, the dynamic nature of agricultural production standards and principles requires that data is replaced and revised from external sources (Nikkilä et al., 2012a). The temporal accuracy of these datasets can be critical. It is still challenging to transfer regulations between IT systems and to provide means for an integration and interpretation of such rules in decision and management support tools. A conceptual model of a FMS suitable for automated compliance control is given by Sørensen et al. (2010) and Fountas et al. (2015) involving the acquisition of spatial and temporal data and the subsequent processing and inference for final decision support within the operations management and activity documentation aimed at external stakeholders.

1.3.2 Farming regulations

Arable farming is strictly regulated both at national and international level. These regulations have a direct impact on how the farmers plan and do their farming works. In the last three decades there has been an increase in the number of legal regulations to be conformed. The guidelines concern all farmers and are related to food safety and environmental acts like the fertilization of nutrients, the use of pesticides and seed types. There are also voluntary standards to show compliance to stricter requirements such as the EU Organic standard (EU Regulation 834/2007) or private industry standards e.g. GlobalGAP (2007). A higher price level for specialized productions or better food quality can be a driving factor for compliance to stricter standards. According to Nash et al. (2011), these agricultural standards are composed of a set of rules including the event of compliance, a definition of terms used and how compliance to the rule is to be assessed. Integration of these rules into an automated management procedure is required to provide a better spatial and temporal response. The compliance of regulations can vary during the work execution because for example a threat of rain. It is important to study the regulation exploitation especially from the temporal accuracy point of view.

1.3.3 Precision Farming and Smart Farming

The core idea of PF is to spatially and timely optimize the farming inputs to maximize the farming outcomes while reducing the environmental stress. The optimization is often difficult and a result of compromises. The key starting point is to detect the current within-field variation. Figure 1 presents a yield map for a field of about 20 ha from the year 2014. Data from the combine harvester is presented as points. The yield varies between 1 – 9 t/ha and there are different areas, where the yield was constantly higher or lower. This means that PF treatments can potentially be affordable at some level.
The representation of Figure 1 can introduce the complicated topic of agricultural data. Figure 1 represents yield amount measurements as points. The yield amounts are weighted averages measured in the combine harvester system about seven seconds after the actual cutting. The less than one-ton amounts are filtered out afterwards. The headland areas can be seen as low yield areas although the reason might be the used width of the cutting table. The yield values between the measurement points are not interpolated. The mapping challenges (Kaivosoja and Backman, 2011) and role of yield maps in precision farming are presented for example by Blackmore (2003).

The idea that combines all the relevant technologies in farm management and in PF is called Smart Farming. While PF goes into a single farming process one at the time, the Smart Farming covers the whole farming process. It is basically a framework that is needed to be able to successfully perform smart acts in the field or with livestock. In short, Smart Farming in the context of arable farming is a combination of information management, precision farming knowledge and farm machinery automation. Walter et al. (2017) presented that Smart Farming is the key for sustainable farming in the future.

### 1.3.4 Decision support systems

In agriculture, a decision support system (DSS) or (AgriDSS) (Lundström and Lindblom, 2018) is a software based system which helps farmers to solve complex issues related to crop production. With different inputs, the decision making processes are used for management, operations, planning, or different recommendations in agriculture. Good quality data are
crucial to support decision making processes and that is also essential in farm machinery works. DSS can be part of FMS, can be a separate application or service, or it can be partly integrated into other systems such as FMS. In the precision farming work execution, the DSSs prepare the application tasks based on different source data. The application task includes a plan for example to spread certain pesticide mixture 200 l/ha evenly to the entire field. If it includes spatial variation (PF task), it is called variable rate application (VRA) task. The task defines how much farm machinery should spread seeds, fertilizers or pesticides at certain location at certain time. A map based application task is pre-planned and a sensor based application task is calculated real-time. From the uncertainty point of view, the performance of DSS is essential.

1.3.5 Remote sensing in agriculture

Remote sensing is data collection about an object without making physical contact with it. It is used in numerous fields but in current usage, the term generally refers to the use of satellite-, aircraft- or drone-based sensor technologies to detect and classify objects on Earth based on propagated signals. According to Batini et al. (2017), photogrammetry as well as laser scanning (each airborne, terrestrial and mobile) are taken as part of remote sensing. The remote sensing is furthermore meant with the used algorithms, techniques to process the acquired data sets. Recently there has been a fast development and productization with drones, which are also called RPAS (remotely piloted aircraft system), UAV (unmanned aerial vehicles) or UAS (unmanned aerial system). The prices of UAV equipment have gone down and new sensor technologies are coming into markets. Also new, free and relevant satellite technology has become available for the environmental mapping. The relatively new Sentinel-2 satellites are providing useful data several times per week with up to 10 x 10 meter accuracy. Remote sensing data has widely been accepted as a PF source data. In farming, there is also a fourth platform for remote sensing: sensors can be mounted on the farm machinery. This is an application of mobile mapping, which is the process of collecting spatial data from a mobile vehicle. Such commercial tools in agriculture include Yara N-sensor and Trimble GreenSeeker. The work of Batini et al. (2017) provided a grouping of data quality dimensions in the remote sensing domain. These dimensions included spatial, radiometric and spectral resolution, temporal resolution, precision in the cluster accuracy, positional accuracy, thematic precision, temporal validity, data completeness, spatial redundancy, readability, accessibility and consistency. ISO/TR 19121:2000 (ISO, 2000) provide standard for imagery and gridded data and how this type of data should be handled.

1.3.6 Remote sensing applications

An additional fertilization is one of the most studied precision applications (Mulla, 2013). Nitrogen management is also essential in sustainable agriculture (Zhang et al., 2015) and as mentioned, there is no generalized methodology for maximum use efficiency (Raun et al., 2005). The crop yield typically responses sensitively to the fertilizer dosage. Insufficient dose of the nitrogen fertilizer leads to low amount and often also a low quality of the yield. Excess of nitrogen, on the other hand, causes a risk of lodging of the growth, causing yield losses. Unused nitrogen in the soil is prone to leaching throughout the growing period and after. The already developed precision nitrogen application methods for crops utilize an optical sensing
of the growth status during the growing season to determine how much additional nitrogen is 
needed in different areas of a field. The technique is typically based on visible and NIR (near 
infrared) measurement data which is calculated to Normalized Difference Vegetation Index 
(NDVI) map or its variants. The NDVI map indicates the relative amount of green mass in the 
field by calculating the relative difference between NIR and red light: \( \frac{(NIR-red)}{(NIR+red)} \). 
Typically, the greenness of the crop correlates with the nitrogen status. However, the method 
is not able to differentiate situations of low growing density with high nitrogen content from 
those of high growing density and low nitrogen content. Thus, generating nitrogen fertiliza-
tion plans from these data cannot be optimal in all of the situations. The economic benefits 
gained with precision nitrogen management have been around 5 % (Nissen, 2012).

Hyperspectral imaging for example was found to be an interesting method for obtaining 
separate biomass and nitrogen content data to improve spatial fertilizing planning 
(Honkavaara et al., 2013, Pölönen et al., 2013, Clevers and Kooistra, 2012). There are many 
ways how to apply index calculations with RGB (Red, Green, Blue) and multispectral data 
and the exact applied wavelengths can also vary. Dong et al. (2015) presented 28 chlorophyll-
related vegetation indices suitable to be applied with Sentinel-2 data and by simulation stud-
ies; they found out that incorporating red-edge reflectance (around 700 nm) improved the 
estimates for assessing vegetation growth rate and predicting crop productivity. Also, Hunt et 
al. (2017) noted that assessing red-edge detection could make a difference in determining 
nitrogen applications to potato. That is also what current commercial solutions support. The 
tractor mounted YARA N-Sensor five spectrometer detects the wavelengths of 550 nm, 650 
nm, 700 nm, 710 nm and 840 nm (Varco, 2010). A drone installed Parrot Sequoia multispec-
tral camera measures the wavelengths of RGB, 550 nm, 660 nm, 735 nm and 790 nm. The 
most accurate wavelengths of Sentinel-2 satellite are 490 nm, 560 nm, 665 nm, 842 nm with 
10 meter spatial resolution and 705 nm, 740 nm, 783nm, 865 nm, 1610 nm, 2190 nm with 20 
meter spatial resolution. The following Figure 2 shows the spectral curves of pure healthy 
wheat leaf and unhealthy wheat leaf measured with a leaf spectrometer CI-710. The spectral 
centres of Sentinel-2, Parrot Sequoia and Yara-N-sensor Five are also presented in Figure 2.

![Spectral reflectance of healthy and unhealthy wheat leaves and the spectral centre points of different remote sensing technologies (satellite, multispectral drone camera and tractor measurement system)](image-url)
These remote sensing methodologies are rather vulnerable. Especially the passive sensing that is typically used with satellites and drones produce reflectance information that is directly related to available sunlight. Pena-Yewtukhiw et al. (2015) found out that even the sensor output difference of 0.05 NDVI units could strongly affect the resulting nitrogen rate prescription, depending on the selected algorithms. Also image mosaics that are mandatory with UAV sensing may create large radiometric errors that effect on spectral vegetation indices (Rasmussen et al., 2016). However, for example in the work of Näsi et al. (2018), the development with pre-processing methods has gained a level where image mosaics are relatively error free within single bands.

These remote sensing applications produce source data for DSSs that prepare PF applications in the execution of Smart Farming. The main uncertainties are linked to the fact that how well the selected remote sensing approach suits to the measured phenomenon, how stable is the selected remote sensing methodology and how up-to-date information it produces. On top of information timing, the actual field work timing in agriculture is important.

## 1.3.7 Timing of the field operations

The correct timing of the farm machinery work execution is usually a challenge and can be very critical. Correct operational timing is one of the key goals (Jørgensen, 2018). The timing of the additional fertilization can be difficult for example. On the other hand, the fertilization date would be as late as possible so that the differences in vegetation and the growth potential can be better seen and mapped. On the other hand, if the fertilization is made too late for example with barley, it will only increase the protein content, which decreases the value of the barley and that is not wanted at all.

The timing in the farm machinery work executions is often a compromise due to existing weather conditions, long term growth conditions and available human and machinery resources including the fleet management. From data quality perspective, interesting is the time and date of the source data that will be used in the field operation. At the general level, more up-to-date data is better. That can be challenging due to required data collection methods, processing times or for example with satellite imagery when the latest images are not usable due to cloud covers. Assessing real time data during machinery work execution is limited, current standards do not support it.

Changing conditions like current rain and wind, pesticide alarms, weather forecasts, applicable matter content changes, working schedules, work applied by other working units, different risk analysis, information from aerial systems or advisory recommendations could have a huge impact on the application task rates.

## 1.4 Input data for the farm machinery

In this dissertation, the farm machinery consists of all the components that the machine manufacturer would commonly deliver. It includes the working implement hardware, the electric controller unit and the device descriptions. So this work is not a specific machine dependent. From that perspective, the main input data for farm machinery is the application task or VRA task. Figure 3 presents prototype farm machinery in action. The driving speed is about 10 km/h and the GNSS unit is on top of the tractor. Already seeded rows can be seen on the right side of the tractor. On the left side of the combine driller, the marker marks the
centre line to the next driving pass. A typical combine driller will adjust two different parameters, the amount of seeds and the amount of spring fertilizer; both can have separate application task maps. This prototype tractor in Figure 2 spreads seven different types of seeds or fertilizers.

![Prototype machinery in action adjusting simultaneously several different seeds/fertilizers via ISOBUS in 2017](image)

**Figure 3.** Prototype machinery in action adjusting simultaneously several different seeds/fertilizers via ISOBUS in 2017

There can also be variations within the working width of the machinery. Some machines especially boom sprayers can control different adjacent nozzles separately, this is called section control. Spin disk fertilizers have a certain Gaussian style work pattern where density near the sprayer is higher. The driving lines are planned to partially overlap each other. Spreading distance can be for example 40 metres while distance between driving lines is 20 metres.

### 1.4.1 Application tasks

The VRA task can be used for example for seeding, weed control, lime, different fertilizations, manure, soil cultivation intensity, fungicide and pesticide spraying. Figure 4 presents an example of a VRA map. The application task generation applies many methodologies from geoinformatics science; the whole chain including the acquisition, storage, processing production, presentation and dissemination of geoinformation. Currently, tools for using the application tasks are available but the principles for general task constructions for different applications do not exist.
The application task map is typically simplified in relation to source data so that it is suitable for machinery performance. In this case (Figure 4), the simplification has made in a way that the map is a 20 x 20 m grid in vector format. The simplification can be made in many ways and that will be discussed later in the Usability chapter 2.7. To adapt possible changes caused by VRA and positioning data in farming operation, it is necessary to be able to deliver a proper message to the Electronic Control Unit of the Implement (I-ECU). ISOBUS (ISO 11783) is a standardized way to do that.

1.4.2 ISOBUS environment

To control the implement in a standardized way, the idea of ISOBUS Task Controller (TC) has been introduced. ISOBUS has already gained a relatively large market share and is implemented by many manufacturers. ISO/FDIS 11783-10 (ISO, 2007) is an interface relating to communication at the software level between FMS and mobile implement control system using board computers (ISOBUS-TC). TC uses XML-based formats for communication with FMS, and Process Data Messages (PDM) via controller area network (CAN) bus to communicate with the I-ECU. TC executes data setup and machine configurations and takes care of the documentation of the work executed by the machinery. For the spatial working rate changes of the implement, TC uses an ISOBUS task map. In the standard, the ISOBUS task is structured as one task per one work. In practice, a planned and suitable task is selected from a drop list at the beginning of the work. Then the entire work is done according to it navigating inside the task map with GNSS.
1.4.3 Positioning of the machinery

The position of the machinery in the field and in the application map is interesting. The GNSS is widely used for the position estimation of agricultural machinery with applications such as steering assistance, automatic steering, semi-autonomous machinery or robotics. The principle of GNSS positioning method is to measure how much time it takes to retrieve messages from different satellites whose time is accurately known and locations are moderately known. When information from four satellites is retrieved, three coordinates and the time of the receiver can be calculated. More satellites improve the location estimation. There are several different GNSSs, such as GPS, GLONASS, BeiDou and Galileo. The combinations of these different systems can also be done to be able to improve the accuracy of the receiver's location.

The positioning quality depends also on the available GNSS satellite constellation and its variation, and the quality of signals from different satellites. Single signals include errors caused by atmospheric effects, multipath effects, ephemeris errors and clock errors (Rankin, 1994 and Vernotte et al., 2001). For many purposes, the accuracy and reliability of autonomous GNSS positioning has been improved by exploiting different correction systems of the GNSS positioning or with possible supporting measurement systems. The selection is often a balance between the expenses of the system and the expected accuracy and efficiency levels of the system. The farm machinery positioning accuracy and the accuracy of the recorded driven path are affecting the uncertainty of the work execution.

1.4.4 Defined inputs

Data inputs can be considered as standardized parameters handled together with device descriptions in the computer in the tractor cabin. The device descriptions are variables such as machinery type, machinery measures and dimensions, work pattern and amount calibrations. Figure 5 reveals data involved in the work amount decision making. The active parts are TC in the tractor, I-ECU in the machinery and the active hardware tool that does the work. The inputs to the TC are the precision application task map, already done work map to avoid overlapping work (the same working unit or other working units), the device description values and the coordinates of the current machinery location in the same coordinate system as the application task map. From those data, the TC forms the work parameter values. This is the core of this dissertation. With these parameters, the farm machinery executes the current farming work.
When these inputs are accumulated together with the machinery performance, we end up with an entire work performance. There are standards for this machinery performance, for example there are tray tests for spin-disk fertilizers, where the standard deviation of the work result should be less than 5%. Actual field tests have given poorer results. For example Lawrence and Yule (2007) found out that an average fertilization field distribution variation calculated over all test fields was 37.9%, The simulated results by Kaivosoja and Linkolehto (2012) gave similar type of results, but both of those included the entire field coverage in irregularly shaped fields and included the propagated uncertainties related to spatial input data. This dissertation examines the input data uncertainty also in contrast to this farm machinery performance. Will it be more profitable to improve data inputs to the machinery rather than invest in a more accurate and flexible machinery? To access this information, we need methods how to measure and evaluate the quality of the machinery inputs.

1.5 Spatial input data in farm machinery

The requirements for the farm machinery input data are that the values are optimal. The problem is that it is not known what values are right. Factors such as upcoming seasonal weather will effect on the optimal values, for example heavy rain may leach nutrients from slopes. We need to make compromises for optimal values. The following Figure 6 presents the problematic field of spatial data inputs in farm machinery work execution.
Figure 6. Progress of precision farming field work from a plan to the applied result

First there is a plain field map (a), next there is a planned VRA task (b) showing the simplified application rates for the machinery. This can be based for example on a classified remote sensing data. The next map (c) shows the actual driving lines of the farmer (the headlands are not presented to be able to keep the drawing simple). The next map (d) illustrates the overlapping that was caused by the driving lines (Publication IV Figure 5 and 6 shows a true overlapping map). The following map (e) shows the true applied application rates as colours and the last map (f) shows a true need. The difference between last two maps (e, f) in Figure 6 is visible and the situation is the same in real cases. As such, the map (e) presents the farm machinery spatial data inputs in a simplified way. When the true machinery performance (true spread patterns) is added to map (e), the actual machinery input can be presented.

The spatial data uncertainty in the executions of precision farming operations is basically the difference between maps (e) and (f) in Figure 6 combined with the fact that the map (f) itself is uncertain: the true need is not known. Thus the sources of the uncertainties are: the application task and how it was constructed including the temporal relevancy and the field navigation including the driving pattern to cover the entire field.

1.6 Uncertainty and quality in spatial data

Spatial uncertainty is the difference between the contents of a spatial database and the corresponding phenomena in the real world (Goodchild, 2008). As introduced, the term ‘uncertainty’ consists of errors, randomness and vagueness aspects (Zhang and Goodchild, 2002). Fisher (1999) and Devillers and Jeansoulin (2006) used a term ambiguity instead of randomness. Error is the difference between the presented value in contrast to true value. Error can be measured if the true values exist. Vagueness is linked to poor definitions: poor documentation and especially if objects are fuzzy (van Oort, 2006). Randomness or ambiguity arises due to disagreement on the definition of objects. According to Fisher (2000) many studies on spatial data quality are limited to the treatment of error, thereby ignoring the two other forms of uncertainty. The same observation can be made from the overview of spatial
data quality definitions where the measurement of error requires agreement on a clear definition of what is reality (van Oort, 2006). The ambiguity and vagueness deals with imprecision, where the true value does not exist (Fisher et al., 2006). So, spatial uncertainty can be described as either error or imprecision.

Statistics have traditionally been the solution for uncertainty (Madetoja, 2018), where accuracy and precision characterize the systematic and random components of error in measurements. Random errors cause variation and systematic errors can be removed with calibration. The topic of geostatistics provide statistical methods of modelling uncertainty with spatial data (Chilès and Delfiner, 2012), where different functions can be used to analyse the distribution of errors. However, poorly defined objects cannot be approached with a concept of error, but with the matters of vagueness and ambiguity that require different methods (Fisher et al., 2006).

According to Griffith et al. (2015), uncertainty analysis is still a separate topic from spatial analysis and only a few methods dealing with uncertainty have been adopted as a common practice or implemented in applications. However, issues of quality and uncertainty of spatial data have become important because data imperfections propagated through spatial analysis affect the decision-making process (Bielenka and Burek, 2019) and the different aspects of uncertainty and quality have been studied widely. These include books such as error propagation in GIS (Heuvelink, 1998), statistical approach in uncertainty modelling (Zhang and Goodchild, 2002), data quality in spatial data mining (Stein et al., 2009), different perspectives for internal and external spatial data quality (Devillers and Jeansoulin, 2006), visualization of spatial data quality to present uncertainty (Sun and Wong, 2010), developments of uncertainty modelling and quality control of spatial data (Shi et al. 2015) and. The concepts of uncertainty in spatial data and spatial data quality are similar, but the standards have evolved only for the quality (Bielenka and Burek, 2019). Szatmari and Pasztor (2019) concluded that there is a need to validate the uncertainty models and great attention should be paid to the assumptions that are made in uncertainty modelling. Applying standardized methods could provide systematic and a qualified approach to measure the uncertainty.

1.7 Related research

This chapter introduces recent studies from the relevant domains. Chapter 1.7.1 considers the application task, while the next chapter presents methodologies that have been found to improve data currentness, chapter 1.7.3. is about positioning error determination. The last two chapters introduce spatial data uncertainty research in other domains and the related research done by the author.

1.7.1 Decision support and application tasks

Different algorithms are part of DSS that is needed to be able to produce application maps. The implemented DSS in FMS produces for example an estimate for the required nitrogen fertilization need based on remote sensing data classifications, such as NDVI, but the DSS system requires calibration information about crop’s remaining nitrogen needs and responsiveness according to the predicted yield potential (Raun et al., 2005, Lukina et al., 2001) being an important factor for the nitrogen fertilization. Data from other sources are usually combined with remote sensing as inputs to DSS for determining the nitrogen application rates (Shanahan et al., 2008; van Evert et al., 2012). The amount of yield within-field can
vary a lot, e.g. Robertson et al. (2007) found in an Australian study that the standard deviation of in-field variation was between 200 – 1200 kg/ha. Depending on the protein content of the grain this means variation in nitrogen need from 5 to 35 kg/ha (meaning roughly up to 50€/ha). These variations may also correlate from year-to-year (Pesonen et al., 2010) and that might not correlate with the plant status in the middle of the growing season. Also, if some areas of the field have been treated differently, with different fertilization rates at seeding for instance, it will definitely distort the results in the plant status determination. This variation should be included in the DSS. In order to demonstrate the variation between different methodologies, case studies are needed where different typical classification methods, nitrogen content estimations and external data sources are used for the same additional fertilization task construction.

1.7.2 Temporal accuracy of spatial data in work task

There is a growing number of research comparing different dates of remote sensing data such as Vincini et al. (2014), where they compared Sentinel-2 data from three different growing stages of winter wheat canopies and Kumhalová and Matějková (2017) where they studied winter wheat and winter barley, and concluded that images acquired in later phenological phases conduct better results. Data processing may take from few minutes to a few days, even weeks, but other aspect is that how to adapt that data to the work execution.

Earlier research related to ISOBUS task controller and data transfer has had its focus on an XML-based transfer of data from the FMS to on-board devices and in a data dictionary of identifiers for process data variables and data elements (Nash et al., 2009b). Peets et al. (2012) studied collection and management of data acquired by ISO 11783 compliant and non-compliant on-the-go sensors, but their focus was also on data collection, not exploiting it during the work. The work by Iftikhar and Pedersen (2011) focused on the bi-directional data exchange between the farming devices including climate control and production monitoring equipment, the temperature monitoring sensor and the farming systems featuring agricultural advisory service, supplier, contractor and manufacturers. The solution focused on ISOBUS-available functions in an Agri-domain specific format. There has not been research on exploiting spatial web services during farming operation or implementing them in the ISOBUS environment. However, both our work and the work of Iftikhar and Pedersen (2011) have a solution on how to involve external data to a work execution process on-the-go. Also the current ISOBUS standard has a peer control section, where a machinery unit can control a second machinery unit.

With a demonstration of spatial web services, it could be possible to check and update automatically whether and how the relevant agricultural rules, regulations and best practices are still fulfilled. Rather than constructing and evaluating a single complete task, it would be more practical to evaluate separate spatial decisions which form the task in hand. Then decisions can be made based on available spatial and rule-based data sets.

1.7.3 Positioning error determination

Beside typical GNSS accuracy tests, machinery mobility and the changing environment set challenges for the GNSS inaccuracy definition. Min et al. (2013) made dynamic GPS tests with a tractor in citrus orchards and found that receivers performed differently under various orchard conditions. Pirti (2008) found that tree canopy on one side increased the standard
deviation of static position around 40%. Martin et al. (2017) determined that uncertainty of the raw GPS data is fairly well characterised by the Gaussian prediction region in static conditions, but the work of Stombaugh et al. (2002) showed that a static performance of receivers might not be even indicative of dynamic performance and there could be a need of a dynamic GNSS test standard. There are also many different GNSS positioning methods and correction principles. These different GNSS methodologies are presented in Publication III.

First we focus on positioning error studies. GNSS errors are often represented as white noise (Groves, 2008). However, a simple white noise is not always sufficient (Oksanen et al., 2005, Niu et al., 2014) since almost all of the GNSS positioning error sources are time or space correlated (Rankin, 1994, Bierman, 1995). Zandbergen (2008) also found that the positioning error of spatial data is not normally distributed. To determine the real GNSS positioning error used on a mobile machine, Bell (2000) evaluated the control system’s accuracy by the mean and standard deviation of the GNSS measured tracking error from the desired trajectory. That is a rather laborious approach when we are interested in multiple devices and conditions. There have been various GIS-based dynamic tests for different steering systems of the tractor (Kaivosoja, 2008) and there have been standard proposals for agricultural machinery, such as the dynamic testing of satellite-based positioning devices (ISO, 2010), and the testing of satellite-based auto-guidance systems during straight and level travel (ISO, 2012). However, proposed complicated test tracks were seen unbeneﬁcial to be assembled in multiple places. Nevertheless, these test standards would be a suitable basis for gathering data for the GNSS error measurements.

Other approach to understanding the positional error would be trying to regenerate it. Rankin (1994) constructed simulations which modelled GNSS error statistics for various receivers. The simulator had a model of GPS satellite orbits, which were used to create the dilution of precision (DOP) values that translated pseudo range errors to X-Y-Z errors which the simulator generated. The works of Oksanen et al. (2005) and Backman et al. (2010) introduced GPS noise simulations for agricultural tractors. In those studies, different state-space models were ﬁtted to approximate real measured errors. Backman et al. (2010) simulated also the number of visible satellites, which had an effect on the temporal error. Roberts et al. (2002), Niu et al. (2010), Khelifa et al. (2011) and Niu et al. (2014) showed that Allan variance is a feasible way to analyse the static GNSS positioning characteristics. Based on Allan variance interpretations, Niu et al. (2014) identiﬁed four dominant noise terms from the GNSS positioning error. It would be beneﬁcial to adapt Allan variance analyses to the error simulation in a dynamic environment of mobile machines.

1.7.4 Spatial data uncertainty and quality research

The variable technologies for spatial data acquisition and processing, data management tools and different platforms lead to a large amount of uncertain spatial data in relation to the real world. Therefore spatial data quality and uncertainty are an increasingly important issue in geographical information science with over two thousand publications after 1990 (Bielecka and Burek, 2019). The method for applying spatial data quality and uncertainty studies for data has mostly been adapted for accessing the quality of volunteered geographic information (VGI) by for example Girres and Touya (2010), Goodchild and Li (2012), Antoniou and Skopeliti (2015) and Senarathe et al. (2016).

There are some general applicability studies, such as Mobasheri (2013), which applied spatial data quality standards to study automatization possibilities for quality evaluations of geo-
graphic datasets, and Wang and Huang, (2007) which used earlier standards and presented a methodology for representing data quality rules. He at al. (2005) developed a framework for quality assessment and uncertainty handling in spatial data mining. Freire et al. (2014) applied thematic accuracy, completeness, and geometric quality to access quality of building determination from very high spatial resolution satellite imagery. They found quality assessment to be exhaustive since it involved the comparisons of extracted features against a reference data set.

McBratney et al. (1997) studied the uncertainty of yield and related variables in precision agriculture as measured by their spatial and temporal variances. Mõisja et al. (2018) applied spatial data quality standards and defined elements to study the mapping results of different fieldworkers and found differences in the rates of misclassification, omission, and commission errors. This work defines the spatial data quality elements in the executions of precision farming operations by following modern standards to estimate and demonstrate the role of uncertainty. To my knowledge, research on spatial data quality elements in agriculture has not been conducted before.

1.7.5 Research topic background

The foundations for this dissertation have long been built and the approaches in the related publications are a result of iterative works of the author. Challenges with the ISOBUS task execution were first been addressed in Suomi et al.( 2006), Pesonen et al. (2007), Ojanne et al. (2009), Kaivosoja et al. (2012), Nikkilä et al. (2012b) and Pesonen et al. (2012) and Nikkilä (2013). The overlapping work with farm machinery was introduced for example in Kaivosoja (2009) and Kaivosoja and Linkolehto (2012). The inaccuracies in field navigation and GNSS positioning were previously handled in Kaivosoja et al. (2006), Kaivosoja (2008), Kaivosoja and Backman (2011) and Kaivosoja and Linkolehto (2012). The overall work quality has been touched in Suomi et al. (2005), Kaivosoja (2009) and Kaivosoja and Backman (2011). The study of task variations had its first steps in Publication I, Pölönen et al. (2013), Honkavaara et al. (2012) and Honkavaara et al. (2013), where hyperspectral drone data was combined with spatial data from the farm management system. Comparisons with satellite imageries were first studied in Kaivosoja et al. (2017). The drone related studies have continued in the development of estimation methods (Näsi et al., 2018, Viljanen et al., 2018, Oliveira et al., 2019) and in the development of decision support systems (DSS) (Nikander et al., 2018).
1.8 Objectives and Research Questions

The main objective of this dissertation is to see how well we can guide the farm machinery to perform the field operations. The guiding is done by applying spatial data. This work finds and presents novel and concrete measures about the uncertainty and quality of farm machinery spatial data inputs in PF work execution. In other words, the objective is to develop new methodologies to determine the quality of the application tasks and the positioning of the machinery. This will show the extent of the spatial uncertainty in PF and reveals the reasons why gained benefits in PF activities have stayed relative low. The core idea is to detect spots in PF activities that need to be improved in the future and to present, how good data the farm machinery gets in order to work accordingly. This dissertation examines the spatial data quality elements in farm machinery work execution in new ways by developing methodologies and finally by summing them up by applying GIS standardization in order to reveal spatial data uncertainty. In the literature review, no similar studies were found. There are six main research questions (RQ1-6) of this dissertation and they are as follows:

RQ1. Which are the core elements of spatial data quality in farm machinery operations?

There are no existing tools on how to evaluate farm machinery input data in a systematic way thus it is hard to evaluate and guide future development needs. The fact that input data is spatial data gives an opportunity to apply methods from spatial sciences. The idea is to adapt spatial data quality research methods to the farm machinery input data quality determination to be able to recognize the relevant elements and to be able to evaluate the quality in the topic under study. This will build a framework for the rest of the research questions.

RQ2. How much there is variation between typical precision application tasks made for same purpose in a real case study?

This builds a geoinformatic processing framework for application task generation. Precision application tasks are typically based on remote sensing data. There are certain benefits and drawbacks with different methodologies based on drones and satellites. The application tasks are results of multiple geoinformatic processes. Even slight changes in the process parameters can have a great effect on the application task. The sensitivity of spatial data processing needs to be shown. In general, it would be important to show how different the precision farming application tasks would be based on different remote sensing data and processing methods, even when the intention for the tasks is the same additional nitrogen fertilization. Additional nitrogen fertilization is the most common variable rate application.

First we need to find out, how much there is variation of relative biomass estimations based on different remote sensing data obtained for the same purpose as it has been recognized as quite common practice. Then we need to find out how different are the fertilization application tasks made for the same purpose. Furthermore, it is also possible to use supporting data such as previous yield maps, fertilization amounts and farmer knowledge to adjust the application task. It is essential to show how much including or excluding single dataset could affect the results. This will reveal the impact of geoinformatic processing in relation to source data availability and quality.
RQ3. How multiple external spatial data sources could be integrated into the work execution process and how much they can influence to the application?

The date of the source spatial data on farm machinery input data can be very important. The ISOBUS standard defines that the application task is static. However, temporally accurate data could be exploited in many ways during the work execution. It would be beneficial to be able to demonstrate a good way how to integrate external and standardized data into the execution process.

This would give possibilities for increasing the temporal accuracy of the application tasks a lot and give options to be able to estimate the temporal aspect of the application task. By applying geographic and farm machinery communication standards, broad applicability can be expected. To take consider a long term effect in precision application tasks, source data from greatly different dates can be studied revealing the effect of the remote sensing data date in precision fertilization task generation.

RQ4. What is the structure of GNSS positioning error in mobile farming works and how much is the error? How to take account the local variations in positioning error?

GNSS positioning has a huge role in all types of farm work executions. It directly has an effect on the driving accuracy and the navigation accuracy over the application map. The GNSS positioning is affected by many different types of error sources. The error structure in mobile conditions is not studied much. Understanding the structure would make it easier to understand and estimate the spatial variations in positioning quality. A reliable data collection method for mobile GNSS error is needed and several case studies have to be driven in order to catch the real world phenomenon. The replication of the found errors would verify that most of the relevant components are taken account.

RQ5. How to spatially evaluate the accomplished driving performance in fields?

Farmers tend to overlap field work driving passes. This causes over treatment in the field and wastes supplies and time. Even the GNSS accuracy is related to driving accuracy, there are still many other elements involved. The amount of overlapping especially in the field scale is not well studied. It is important to quantitatively measure the actual overlapping in real conditions to be able to present the significance of it. The quantitative analysis would require lots of completed regular field works with data logging. It is also necessary to divide the overlapping into different components to be able to demonstrate how much different driving assistance or section controls could decrease the overall overlapping. The component determination requires the development of new methods in geoinformatics.

RQ6. What is the overall input data quality in farm machinery field works in real cases?

All of the earlier research questions (RQ2-RQ5) handle certain quality elements in farm machinery field work execution. There are not studies on how these elements interact. Summing up of those is needed to be able to demonstrate how low or high the actual quality can be, and how to guide the future research and development. The summing up can be done in many different methods, so it is necessary to adapt existing suitable standards as much as possible. Adapting the standards could reveal elements that are not included in the other research
questions. The quality itself can be rather discrete so it would be beneficial to translate result values to equations considering economic payoffs in precision farming processes.

The objectives for each question are: the theoretical foundations for input data quality (RQ1), the quality of task generation processes and source data (RQ2), the integration of external data sources (RQ3), determination of positioning errors (RQ4), field work overlapping determination (RQ5) and overall input data quality (RQ6). Publications I and V answers the RQ2, Publications I, II and V answers the RQ3, Publication III answers the RQ4 and Publication IV answers the RQ5. The overall spatial quality in RQ1 and RQ6 is handled in all of the Publications.
2. Theoretical foundations

This chapter introduces the theories and the scientific framework that is relevant to measuring the spatial data quality of farm machinery works. A book titled *Precision in Crop Farming: Site Specific Concepts and Sensing Methods: Applications and Results* (Heege, 2013) complements this dissertation from agronomic perspective but does not overlap. This work approaches its’ topic from Geographic information science (GIScience) discipline. It applies i.a. geoinformatics, geomatics, cartography, geodesy, geostatistics, spatial analysis, remote sensing and spatial data sciences. It also touches sciences such as photogrammetry, big data analytics and signal processing and applies them in an agronomic environment. The following chapters introduce the applied theories. The selected theories provide a basis for the research and give new insights into the agronomy domain. Next chapters will present the theoretical foundations for this dissertation.

2.1 Measuring spatial uncertainty

Uncertainty of measurement is defined in metrology (ISO, 1993): it is the parameter that is associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measured object. The measurement uncertainty is represented using standard deviation. However, uncertainty in geographic information relates to user and producer process and analysis and representation of geographic information (Jakobsson, 2006). Chiles and Delfiner (2012) showed that there is a growing number of modelling spatial uncertainty with geostatistical methods. Jakobsson (2006) and Devillers and Jeansoulin (2006) presented that spatial data quality approach is production centred, while uncertainty is data exploitation (customer) centred approach. According to Shi et al. (2015), when U.S. National Center for Geographic Information and Analysis NCGIA formulated its research agenda about measuring the uncertainty in spatial data in 1988, two initiatives were related to spatial data quality: 1) accuracy of spatial databases which emphasized the sources and management of spatial data uncertainty (Goodchild and Gopal, 1989) and 2) visualizing the quality of spatial information. As mentioned the standards have evolved only for the quality (Bielenka and Burek, 2019) since the multidimensionality of imprecision.
2.1.1 Spatial data quality

The book Devillers and Jeansoul (2006) presents the fundamentals of spatial data quality. Spatial data quality is the similarity between spatial data and both positional truth and attributes truth. The quality result is uncertain when the truth is not accurately known. The usefulness of quality is measured by its ability to reduce the uncertainty of a decision (Devillers and Jeansoul, 2006). Brimicombe (2003) showed that the annual number of articles on spatial data quality has strongly grown since 1987.

There are many terms related to spatial data quality. The term ‘accuracy’ means measurement of error and it is usually used especially with positional uncertainty (Li, 2017). Accuracy can be understood as performance characteristic for total error (Menditto et al., 2007). Often accuracy is a description of systematic component such as trueness (Madetoja, 2018).

Error can be measured, if the true value is exist. When the truth is not known, we use the term ‘imprecision’. Methods for modelling error are probabilistic in nature, while fuzzy set theory (Zadeh, 1965) and expert opinion among others are utilized with imprecision modelling. Zhang and Goodchild (2002) classified error sources as abstraction, data acquisition, geoprocessing and combining data sets. Goodchild (2004) stated that mapping is not considered as a measurement science due to transformations, generalization and subjective interpretations in data acquisition. This makes the error analysis difficult.

Several ISO Standards are related to data quality and these are being harmonized into a single standard called ISO 19157:2013 Geographic information - data quality (ISO, 2013). It establishes the principles for describing the quality of geographic data. The standard has four main scopes. It defines components for describing data quality, specifies the components and content structure of a register for data quality measures, describes general procedures for evaluating the quality of geographic data and establishes principles for reporting data quality. The ISO standard defines quality as “degree to which a set of inherent characteristics fulfils requirements” (ISO, 2013). These components and definition meet the requirement for the overall quality evaluation methodology for the farm machinery input data.

2.2 Spatial data quality elements

The ISO 19157:2013 standard says: “Data quality shall be described using the data quality elements. Data quality elements and their descriptors are used to describe how well a data set meets the criteria set forth in its data product specification or user requirements and provide quantitative quality information.” So also for the farm machinery input data, the quality elements would be important to be studied to be able to understand quality aspects.

The ISO standard continues: “When data quality information describes data that have been created without a detailed data product specification or with a data product specification that lacks quantitative measures and descriptors, the data element may be evaluated in a non-quantitative subjective way as a descriptive result for each element.” This gives some freedom in the environment where the spatial elements might be singular. Also the ISO standard says that data quality evaluation can be applied to data set series, a data set or a subset of data within a data set.

According to ISO 19157:2013, data quality comprises six elements: completeness, thematic accuracy, logical consistency, temporal quality, positional accuracy and usability. The elements and their subcategories are presented in Figure 7.
The completeness consists of commission and omission, thematic accuracy consists of thematic accuracy correctness, non-quantitative attribute correctness and quantitative attribute accuracy. The logical consistency consists of conceptual, domain, format and topological consistency. The temporal quality includes accuracy of time measurement, temporal consistency and temporal validity. The positional accuracy includes absolute external, relative internal and gridded data positional accuracy. The final element usability has no subcategories in ISO 19157:2013. These elements may also be characterized as producer data quality elements, because they can be used by the data producer (Yang et al., 2013).

Next we study the spatial data quality elements from the farm machinery work execution perspective. The work of Mõisja et al. (2018) used a similar, quality element selection approach for the study of quality of fieldworker mapping.

The thematic accuracy, positional accuracy and temporal quality are in a key role with farm machinery tasks. Also the role of the other elements: completeness, logical consistency and usability are presented and discussed in next sub chapters. Figure 8 interprets the different spatial data quality elements in farm machinery work execution based on Publications (I-V). The following chapters will introduce these elements.

Figure 8. Main spatial data quality elements in farm machinery work execution
2.3 **Thematic accuracy**

The attribute accuracy in general refers to the accuracy of the quantitative and qualitative information attached to each feature. It consists of three data quality elements (ISO, 2013): 1) classification correctness, 2) non-quantitative attributes correctness and 3) quantitative attribute accuracy. In farm machinery input data quality, it can be understood as the accuracy of the variable rate application map. There are equations such as Equation 1 (Greenwood et al., 1986):

\[
N_{\text{NEED}} = \left( \frac{5.35 \cdot \text{Biomass}^{-0.442}}{1000} \right) \cdot \frac{\text{Biomass} - \text{Nitrogen}}{1000}
\]

This calculates an assumption on the nitrogen fertilization need and can be a basis for one approach. As presented, there is no generalized methodology for maximum use efficiency (Raun et al., 2005). In this dissertation, different methods are studied on producing variable rate application tasks and calculate variations between them (RQ2 and RQ6, Publication I and V).

2.4 **Temporal quality**

The timely accuracy named as temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. Temporal accuracy is often equated with currentness (Thapa and Bossler, 1992). According to standard (ISO 19157:2013), temporal quality has three elements: 1) accuracy of time measurement, 2) temporal consistency (the order of events) and 3) temporal validity. With farm machinery, the temporal validity of data and DSS and task controller response capabilities corresponds to the temporal quality. Typically, the currentness in PF applications refers to the weather or to the growth status of the crops, such as Zadoks scale with 94 different codes (Zadoks et al., 1974) for cereals. The main codes are such as germination, seeding growth, tillering, stem elongation, booting, awn emergence, flowering, grain milk development, dough development and ripening. The temporal quality is handled in Publications II and V (RQ3 and RQ6).

2.5 **Positional accuracy**

The positional accuracy refers to the absolute and relative accuracy of the positions of geographic features. According to ISO 19157:2013, it also has three data quality elements: 1) absolute accuracy is the closeness of the coordinate values in a dataset to values accepted as or being true. 2) Relative accuracy is the closeness of the relative positions of features to their respective relative positions accepted as or being true and 3) positional accuracy of gridded data. In farm machinery works, the positional accuracy means the accuracy of the positioning system, accuracy of the machinery location and positional accuracy of the application map.

The application map accuracy often comes back to the positional accuracy. For remote sensing source data, the georeferencing accuracy is important. With UAV data, it depends on the amount and quality of control points. When external data is used in the application task generation, the positional accuracy of source data is important. For example, the positional accuracy of the tractor in the spring fertilization execution or the combine harvester's posi-
tional accuracy when previous yield maps have been done. The Sentinel-2 Mission Performance Centre MPC has quality requirements for absolute location uncertainty: 10 m standard deviations away from the mean without ground control points (GCP) and 6.3 m with GCPs. Roughly, this means that the target accuracy is about one pixel. The positional accuracy is related to the RQ4, RQ5 and RQ6 and it is handled in Publications III and IV. As tools for positional accuracy determination, next the Allan variance and Spatial autocorrelation and field coverage are presented.

2.5.1 Allan variance

The Allan variance is a two-sample variance (Lesage and Audoin, 1997) and is a time-based domain analysis technique originally developed to study the frequency stability of precision oscillators. The Allan variance analysis has been widely accepted as a method of identifying stochastic processes (Niu et al., 2014). A detailed definition and explanation of the Allan variance are presented in IEEE (2008). The Allan variance presents how much there is variation in average between clusters with certain length (time interval or frequency). In spatial statistics, the Variogram analysis is a quite similar approach with more dimensions. The equation for the estimation of Allan variance with the fixed $\tau$ estimator is:

$$\hat{\sigma}_{x}(\tau,M) = \frac{1}{2(M-1)} \sum_{i=0}^{M-2} (\bar{x}_{i+1} - \bar{x}_{i})^{2},$$

(2)

Where $M$ is the number of averaged clusters in data and $\bar{x}_{i}$ is the mean value of the clusters. To analyse the error characteristics, a log–log plot of the Allan deviation (ADEV) is used (Niu et al., 2014). The plot is called Allan variance plot and the ADEV is a square root of the Allan variance.

The examination of the Allan variance plot allows different noise components to be distinguished by the slope of the plot in particular time regions and in different magnitudes (Land et al., 2007). Gaussian white noise is characterized by a region of -1/2 slope, Random walk noise is associated with +1/2 slope and Flicker noise is shown as a region of 0 slope (IEEE, 2008 and El-Sheimy et al., 2008). The frequency spectrum of Flicker noise is linear in logarithmic space. The fourth found type of noise presented by Niu et al. (2014) was a First order Gauss-Markov (GM) process, which appears as raised points in the Allan variance plot (Niu et al., 2014). The GM processes are stochastic processes that follow both Gaussian and Markov processes. GM processes have an exponential autocorrelation function with variance and correlation time (Rankin, 1994). The equations for GM process are presented in Publication III. The correlation time in GM process has typically been around 600-3600 seconds with static GPS positioning (Cannon and LaChapelle, 1992). Borre and Tiberius (2000) found that GPS receiver’s noise at sampling rate below 1 Hz is independent and does not have any time correlation. The most possible reason for a time-correlated noise is the remaining tropospheric path delay (Howind et al., 1999). According to Rankin (1994), that can be modelled as the First GM process.

Although the Allan variance analysis can give great estimations in general noise determination, El-Sheimy et al. (2008) stated that Allan variance does not always determine a unique noise spectrum since the spectral characterization is limited. This put restrictions on what can be learned about a noise from the examination of its Allan variance. However, Niu et al. (2014) concluded that the Allan variance analysis method can be proposed for the GNSS po-
sitioning, with some conventional method, for example, PSD (power spectral density) as a supplement if necessary. The PSD describes how the squared value of a noise is distributed directly over the different frequencies. It shows at which frequencies variations are strong and which are weak. Niu et al. (2014) showed that the static GNSS noise types identified by the PSD and the Allan variance methods were sufficiently the same.

Still, even the Allan variance analysis could retrace the positioning error; there is the problem that error is greater near a high surrounding canopy. That mostly happens near the field boundaries. The greater error can be caused by the surrounding forests, buildings, hills, signal multipath sources, slopes in the field and different obstacles inside the field parcel.

2.5.2 Spatial autocorrelation

The spatial autocorrelation represents a degree of dependency among observations on a geographic space. Close objects are similar to each other. In positioning errors in machinery navigation, this means that parallel driving tracks in certain areas would have similar amounts of error. The spatial autocorrelation can be indicative of the local error behaviour. The classic spatial autocorrelation statistics compare the spatial weights with the covariance relationship at pairs of locations. There are two main approaches: raw-data approach and a matrix approach. In the raw data approach, the spatial structure takes the form of a free X and Y geographical coordinates of the samples. In the matrix approach, the spatial structure is in the form of a geographic distance matrix among locations. There are two traditional autocorrelation calculation methods: Geary’s C (Geary, 1954) and a Moran’s I (Moran, 1950). Geary’s C is more sensitive to local spatial autocorrelation and that can be estimated it to better indicate the spatial autocorrelation of error in mobile positioning. The equation of Geary’s C is:

\[ C = \frac{(N-1) \sum_{i} \sum_{j} \omega_{ij} (Z_i - Z_j)^2}{2W \sum_i (Z_i - \overline{Z})^2}, \]  

Where \( N \) is the number of measured points indexed by \( i \) and \( j \) and \( Z \) is the amount of error. Variable \( \overline{Z} \) is the mean error and \( \omega_{ij} \) is a matrix of spatial weights representing the neighbourhood and \( W \) is the sum of all of those spatial weights. The result of Geary’s C is between 0 and 2. Values lower than 1 demonstrate increasing positive spatial autocorrelation, whilst values higher than 1 illustrate increasing negative spatial autocorrelation. Implementation of spatial autocorrelation could provide the aspect of decreasing positional accuracy near field borders.

Spatial modellers commonly use the term unconditional Gaussian simulation to refer to the process of generating spatially correlated random fields (Begueria, 2010). The unconditional simulation does not honour the data values, but does replicate the mean and variance (Deutsch and Journel, 1998).

2.5.3 Field coverage

The other positional accuracy aspect understands the location where the farm machinery has been driven and where it is currently represents the positional awareness. The machinery location is the result of the driver’s navigation possibly assisted with semi- or fully steering automation. The absolute accuracy of these supporting systems is well understood but it is not well known that how much these systems could actually improve the farming results or how good the result really is in the field scale. Kvíz et al. (2014) found that the regular overlap
of passes done by the driver was in the range between 1% and 6% of the working width of the machine with the drivers who were aware of the test setup. They also found that value can be significantly minimized by utilizing precise guidance systems and the working width had a significant influence on the accuracy of field operation. They measured pass-to-pass deviations by measuring the distance between the tire tracks of two neighbouring passes. They concentrated on nominally straight driving parts. However, the shape and size of a field significantly affect the number of machinery passes (Galambosova and Rataj, 2011; Oksanen, 2013) complicating the even coverage of the work. There are no quantitative measures about the assisted or unassisted work accuracy in traditional farming in real field conditions. Shockley et al. (2011) simulated the overall automatic steering and assumed that the spraying overlap in unassisted crop farming would be between 9.5% and 19.1% with 16 meter working width.

Quantitative measures about farming success have not been previously done. An efficient methodology for collecting data is needed. A speed detection with a low-cost GPS has been accepted for several studies: Witte and Wilson (2004) studied 1 Hz GPS for speed determination and concluded that it was accurate enough for biomechanical and energetic studies especially in relatively straight courses and Keskin and Say (2006) concluded that low-cost GPS receivers can be confidently used to measure the ground speed in agricultural machinery operations.

2.6 Completeness

Completeness refers to the degree to which geographic features, their attributes and their relationships are included or omitted in a dataset. For example, a city map without any roads has a big completeness problem for many use cases. With farm machinery operations, there are typically no completeness problems. However, the completeness can refer to the task area: which parts of the field are included in the task and how well the field boundaries are digitized. There are some challenges on determining the actual field size. There are official field parcel borders in Finland that are annually corrected from digitized ortophotos. These borders define the official size of the field plots; those polygons do not take into account the surrounding or within-field ditches so the actual farming size is somewhat smaller. Also many fields are divided into different growing parcels. Slusarski and Justyniak (2017) studied the uncertainty of cadastre sizes determined by GNSS measurements or orthophoto. They found that there is a significant impact of the distance between polygon points on the accuracy of the area determination.

From the farm machinery point of view, there are two problems related to the field sizes. The correct field size should give an answer to how much fertilizers, seeds or pesticides are needed in the tank? Also machinery could refuse to spread in the areas that are not included in the task. On the other hand, unwanted areas that were included in the task for some reason such as ditches, dwellings, barns, electric poles etc. could arise as a heavy fertilization need in an automated process making. That could lead to remarkable errors on the requested tank fillings. The completeness is related to RQ1 and RQ6 and is handled in this summary.
2.7 Usability element

According to ISO 19157:2013 standard, the usability element shows how suitable the data set is for the identified usages. In our approach, the user is the Implement-ECU. The usability means how well the task characteristics meet the implement capabilities. This means that the spatial resolution and application rate variations of the task are realistic for the machinery adjustments and measures: driving speed, machinery working width and work pattern. There can be three main difficulties: 1) too complicated processing for the task controller, 2) the machinery is adjusting itself almost all the time and the adjustment delays or inaccuracies (for example with a proportional-integral-derivative controller PID) will have a great effect, 3) the entire working width work (up to 40 metre wide) is done according to the conditions at the current location of the GNSS unit.

These machinery parameters and characteristics are fully case dependent. Also, high detailed task maps can be simplified with many different methods, so in this thesis we only demonstrate the adjustments and map generalisations in real conditions. This work is related to Publication I and RQ6.

2.8 Logical consistency

The degree of adherence to logical rules of data structure is known as logical consistency. According to ISO (19157:2013) it has four data quality elements: 1) conceptual consistency, 2) domain consistency, 3) format consistency, 4) topological consistency. The consistency is typically used to specify conformance with certain topological rules (Kainz, 1995) such as polygons are bounded by non-intersecting lines or for example that a building is recorded as a building and it is marked as a polygon in a GML format in ETRS-TM35FIN coordinates.

Logical consistency has an important role in the chain between external data sources, FMS and ISOBUS task execution. Works such as agroXML (Schmitz et al., 2009), Agrovoc (2017), Nash et al. (2009a) have been working in this area. The logical consistency is outside the scope of this work since this dissertation does not cover data format level of individual cases. Recent projects such as Geospatial ICT infrastructure for precision farming operations management (GWAI, 2018) deals this kind of problem field, where OGC standardization meets ISOBUS standardization in concrete and practical way.

2.9 Overall quality measure

To sum up the different elements, the ISO 19157:2013 standard presents that the quality of dataset is measured using a variety of methods: “A single data quality measure might be insufficient for fully evaluating the quality of the data specified by a data quality scope and providing a measure of quality for all possible utilizations of a dataset. A combination of data quality measures can give useful information. Multiple data quality measures may be reported for the data specified by a data quality scope” (ISO, 2013). In overall quality estimations, this leads to set of values. However, to facilitate data set comparisons, it is necessary that the results in the data quality reports are expressed in a comparable way with common understanding.

For data quality basic measures, the ISO 19157:2013 defines two principle categories: counting-related and uncertainty-related data quality basic measures. In the context of our research, this is problematic. The counting-related measures are based on counting the errors
of correct items. The item should be correct or not and that is not suitable for the farm machinery tasks since the uncertainty of the ground truth. In the standard, the uncertainty-related basic measure has three very critical limitations: uncertainties are homogenous, the observed values are not correlated and the observed values have a normal distribution (ISO, 2013). Also the NSSDA used in the United States (FGDC, 1998) assumes the positional error of spatial data is normally distributed. These can be very limiting. So this dissertation tries to find the best methods of measuring the quality of different elements by not forced to focus on the existing standards at the measurements level. The last part in the results chapter will examine and summarize the quality measures from the standardization perspective.
3. Material and Methods

The main research strategy was to use case studies in order to develop new methods and evaluate the qualities involved in machinery work execution. Each publication included either farm machinery data collection or remote sensing data processing, all linked to the real world phenomenon in Finland. Table 1 summarizes five publications with the related research questions (the main focus on bold), core research methods and the scope of the publications.

<table>
<thead>
<tr>
<th>Publication</th>
<th>RQ</th>
<th>Core research methods</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>I, A Case Study of a Precision Fertilizer Application Task Generation for Wheat Based on Classified Hyperspectral Data from UAV Combined with Farm History Data</td>
<td>1, 2, 3, 6</td>
<td>Case study, Spatial classification, Spatial calculation, Geostatistics</td>
<td>Application tasks, Thematic accuracy, Precision fertilization</td>
</tr>
<tr>
<td>II, Automatic Control of Farming Operations Based on Spatial Web Services</td>
<td>1, 3, 6</td>
<td>Case study, Literature survey, Prototype construction</td>
<td>Application tasks, Thematic accuracy, Precision fertilization</td>
</tr>
<tr>
<td>III, GNSS Error Simulator for Farm Machinery Navigation Development</td>
<td>1, 4, 6</td>
<td>Case studies, Literature survey, Time series analysis, Simulations</td>
<td>GNSS error, Allan Variance, Spatial autocorrelation</td>
</tr>
<tr>
<td>IV, Spatial Overlapping in Crop Farming Works</td>
<td>1, 5, 6</td>
<td>Quantitative research, Case studies, Calculation, Spatial modelling</td>
<td>Spatial overlapping, Drivers performance</td>
</tr>
<tr>
<td>V, Different Remote Sensing Data in Relative Biomass Determination and in Precision Fertilization Task Generation for Cereal Crops</td>
<td>1, 2, 3, 6</td>
<td>Case studies, Spatial analysis, Spatial classification, Comparisons</td>
<td>Application tasks, Thematic accuracy, Precision fertilization, Remote sensing</td>
</tr>
</tbody>
</table>

The focuses of the different Publications in contrast to spatial data quality elements are presented in the following Figure 9.
The following chapters present the methodologies used in this dissertation. First a general data collection methodology is presented and after that, specific methodologies for answering the RQs are presented.

### 3.1 General data collection methodology

The Publications (II-IV) used the same data collection principle to collect farming data. In those tests, we collected both static and dynamic GNSS positioning data. In each test, we recorded NMEA (National Marine Electronics Association) -type GNSS-messages and depending on the case, we simultaneously collected data from different GNSS units or GNSS and farm machinery process data. In our principal dynamic test setup in Publication III, two high accurate RTK (real time kinematics) GNSS antennas were located on both sides of the tractor’s roof. The test unit was located in the middle (Figure 10). The true position of the test unit was determined based on the positions and orientation of two RTK receivers. The determined error was the difference between the measured true position and the recorded position of the test unit at the same time.

![Figure 10. Measurement setup showing the position of GNSS antennas (Adapted from Publication III)](image)
Data collection and synchronization were executed in National Instrument LabVIEW environment on a docked laptop used in our Cropinfra platform (Pesonen et al., 2012). The test drives were close to regular farming operations. All machinery data were recorded from ISO-BUS process data messages with 5Hz or 10Hz interval. For data collection in Publication III, we used seven different GNSS receivers in total, the details of those are presented in the publication. In Publications II-IV, the main GNSS data source was a low-cost Garmin 19x receiver. Table 2 shows a sample data of logged spraying operation. Table 2 presents the status of the implement: power take-off revolutions per minute (PTO), spraying valve status on/off (ON), lifting status sensor voltage representing implements posture (Lift. V) and the GNSS information simultaneously from the CAN bus with 5 Hz interval.

Table 2. Sample data logging of the spraying work

<table>
<thead>
<tr>
<th>Time</th>
<th>ON</th>
<th>PTO</th>
<th>Lift. V</th>
<th>LAT</th>
<th>LON</th>
<th>Speed</th>
<th>Direction</th>
<th>Elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>75757.2</td>
<td>0</td>
<td>311</td>
<td>54.4</td>
<td>60.450895</td>
<td>24.346639</td>
<td>2.304</td>
<td>335.07</td>
<td>83.02</td>
</tr>
<tr>
<td>75757.4</td>
<td>0</td>
<td>317</td>
<td>54.4</td>
<td>60.450895</td>
<td>24.346641</td>
<td>2.808</td>
<td>335.07</td>
<td>83.01</td>
</tr>
<tr>
<td>75757.6</td>
<td>1</td>
<td>320</td>
<td>54.8</td>
<td>60.450894</td>
<td>24.346644</td>
<td>2.916</td>
<td>335.07</td>
<td>83.01</td>
</tr>
<tr>
<td>75757.8</td>
<td>1</td>
<td>314</td>
<td>54.0</td>
<td>60.450892</td>
<td>24.346646</td>
<td>4.860</td>
<td>335.07</td>
<td>83.01</td>
</tr>
<tr>
<td>75758.0</td>
<td>1</td>
<td>317</td>
<td>54.4</td>
<td>60.450890</td>
<td>24.346648</td>
<td>4.680</td>
<td>335.07</td>
<td>83.00</td>
</tr>
<tr>
<td>75758.2</td>
<td>1</td>
<td>329</td>
<td>54.8</td>
<td>60.450888</td>
<td>24.346651</td>
<td>4.500</td>
<td>336.93</td>
<td>83.00</td>
</tr>
<tr>
<td>75758.4</td>
<td>1</td>
<td>346</td>
<td>54.0</td>
<td>60.450885</td>
<td>24.346653</td>
<td>4.852</td>
<td>338.21</td>
<td>83.02</td>
</tr>
<tr>
<td>75758.6</td>
<td>1</td>
<td>361</td>
<td>54.0</td>
<td>60.450883</td>
<td>24.346656</td>
<td>5.304</td>
<td>339.33</td>
<td>83.02</td>
</tr>
</tbody>
</table>

3.2 The application task variations and the thematic accuracy

For the thematic quality evaluations, we had two test setups (Publications I and V) which are presented next. Publication I exploits classified hyperspectral UAV data in order to construct the application task for precision fertilization in a single use case. Publication V focuses on passive remote sensing in precision fertilization. First, different relative biomass estimations based on different methodologies and dates were compared. Then nitrogen fertilization application tasks were generated based on different reasonable parameters.

In Publication V, the test area was about 20 ha cereal crop field in Vihti. We compiled different remote sensing data taken from the field. These included six datasets from Sentinel-2 satellite, three datasets from professional UAV system and three consumer level UAV datasets. Then we used the classified remote sensing data and farmer’s knowledge to produce precision nitrogen fertilization tasks. All the different application tasks were planned in order to have three similar nitrogen fertilization levels. Then we compared these maps between each other, with 36 biomass samples and we also calculated the correlation between biomass amount and nitrogen content from the vegetation. To evaluate the effect of the usage of other data sources, we demonstrated possible task variations by applying previous yield maps, vegetation samples, farmer’s know-how and commercial software classification.

The Publication I focused on producing fertilization maps based on classified biomass and nitrogen maps based on hyperspectral UAV data in the summer 2012. The hyperspectral data pre-processing methodologies are described in the paper Honkavaara et al. (2012). The test
field was around five hectares having a different spring fertilization levels. Dry vegetation and nitrogen content estimations were produced in the work of Pölönen et al. (2013). We used additional data to support the task construction. We used a yield potential map and spring fertilization map. The methodologies behind were described in the Publication I. Then we applied three different methods in order to calculate the nitrogen fertilization need. To evaluate these classifications, we calculated 2 m x 2 m grid over the test field, and calculated image information from each dataset (previous yield maps, spring fertilization, biomass (hyperspectral classification), nitrogen (hyperspectral classification), NDVI map, additional fertilization amount, yield map). Then we calculated correlations to the yield map and yield increment in relation to mapped biomass estimation.

### 3.3 Temporal quality of the machinery input data

Publications II and V determined the temporal quality. To estimate the remote sensing date effect, we used the methodologies presented in the previous application task variations chapter and in the Publication V. To be able to study the real time adjustments we constructed a pesticide spraying case study presented in Publication II. There we included services capable of providing WFS (web feature service) and WMS (web map service) data according to OGC (Open Geospatial Consortium) standards and a task controller capable of exploiting them and communicating with machinery by using ISOBUS. We also used an FTP (File transfer protocol) server to provide weather information. As a background for this TC development are the prototypes of ISOBUS compatible TCs (Miettinen et al., 2006, Ojanne et al., 2009) that was developed for our research purposes. We constructed a Research Task Controller equipped with ISOBUS compatible process data message capabilities operating in a LabVIEW-environment. We used Geodata server at University of Hohenheim providing data such as the variable rate application task, field boundaries, ground water (SYKE, 2012), and classified flood risk data. The GeoServer application published them via WFS for the TC’s user interface as vector oriented GML-files. The weather and forecast information were located in an FTP server as an XML file. The test drives were carried out as simulations with 5 Hz frequency.

### 3.4 Positional accuracies

The positional accuracy methodologies are originally presented in Publications III. The publication evaluates the dynamic GNSS positioning error in arable farming environment and constructs an error simulator.

According to previously and in Publication III described GNSS data collection methods, different dynamic GNSS error data were collected. The following Table 3 shows the measured mean positioning error of the main five tests for latitude and longitude coordinate for each setup.
Table 3. Mean error and standard deviation of the error in the test drives (in metres)

<table>
<thead>
<tr>
<th>GNSS</th>
<th>Mean error (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomous extreme (LAT)</td>
<td>0.417</td>
</tr>
<tr>
<td>Autonomous extreme (LON)</td>
<td>1.543</td>
</tr>
<tr>
<td>Autonomous (LAT)</td>
<td>0.598</td>
</tr>
<tr>
<td>Autonomous (LON)</td>
<td>1.528</td>
</tr>
<tr>
<td>High Performance DGPS (LAT)</td>
<td>0.101</td>
</tr>
<tr>
<td>High Performance DGPS (LON)</td>
<td>0.309</td>
</tr>
<tr>
<td>RTK Float (LAT)</td>
<td>0.065</td>
</tr>
<tr>
<td>RTK Float (LON)</td>
<td>0.034</td>
</tr>
<tr>
<td>RTK (LAT)</td>
<td>0.007</td>
</tr>
<tr>
<td>RTK (LON)</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Next we calculated the Allan standard deviation plots for the test drives. Based on calculated Allan Variances, we calculated the slopes between different cluster times. The slope values are determined as $\Delta \log(y)/\Delta \log(x)$ of calculated Allan deviations for each cluster time. Then we manually pointed out the closest noise centres by examining both absolute values and separate Allan variance plots that revealed the overall characteristics. In addition to noise terms discovered by Niu et al. (2014) in static tests, we discovered rate ramps and outliers from the dynamic data.

From the spatial data quality perspective, the GNSS error reconstruction was needed to be able to ensure that we are interpreting the errors sufficiently. We used the found noise terms for the simulation constructions by using maximum of five separate noise terms for a single simulation. The tuning process had two cost functions for the optimization: 1) the standard deviation of the differences in percentage between real and simulated data in separate cluster times 1-128, 2) the average of the differences. The maximum accepted value for both of them were 10% calculated from the real values. We used PSD calculations (Publication III) only for the evaluation of our results. On top of these presented noise characteristics the spatial dependencies required development of methodologies.

We calculated the spatial autocorrelation values (Geary’s C and Moran’s I) for four different datasets. Then we constructed a thematic map approximating the spatial heterogeneity. We used a method called an unconditional Gaussian simulation to construct a spatially correlated random field. The calculated field was multiplied by 0.888 to adjust the mean value of the random field to be 1. We used this field as a 25 x 25 m grid map.

3.5 Positioning and machinery location

Publication IV studies the driving accuracy in the fields by analysing 140 different complete field works. The field works were regular farming works recorded in 5 Hz. For each selected work, we calculated the travelled distance while implement was working. This distance multiplied by the working width was compared with the detected size of the field producing the overwork percentage of the work under the study. This corresponds to a minimum overlapping percentage. We used data from a 16-metre sprayer (92 drives), combine driller (22 drives), combine harvester (8 drives) and a roller, a cultivator and a spin disk fertilizer (each
six drives) representing a wide scale of farm machinery. There were three different farmers driving and they planned their driving path according to traditional farming practices. To validate our findings, we compared one combine drilling work driven with an automatic steering system (AGI-4 TopDock) with a cm level positional accuracy in 2015 to identical work made without steering assistance in 2014. The field plot, the driver and the machine were the same in both cases. The overwork amount of a single work was the total worked area divided by the measured. The field size determination is presented later in the Completeness chapter. For the statistical analysis of data, we calculated the overwork percentage of each data and then calculated an overall average and the standard deviation of the overwork amount.

Different overlapping elements for the spraying work were determined to be able to find out the meaning of the determined overwork. We used a buffer zone method to (Figure 11) see the characteristics of the overlapping work.

Figure 11. Visualized double and triple overlapping work (adapted from Publication IV, Figure 2) on a field map
Based on the visual analysis, we evaluated that the detected overwork consist of five main different types of working inaccuracies: 1) overlapping the parallel driving lines, 2) overlapping last driving line, 3) working outside the field boundaries (spraying ditches), 4) overlapping before and after the headland turns 4) headland overlap because of the gentle enter/exit angle (over 45 degrees). Then we developed simple methods, to separate these elements from mapped data, these methods are described in Publication IV.

### 3.6 Completeness

The spatial data quality completeness in farm machinery operations considers the field coverage uncertainty. The study of this is presented in Publication IV. We defined the actual field area to be the spatial work coverage of the annual sowing work based on our data logging. We calculated the coverage by using a buffer zone method for driving lines (Kaivosoja, 2008), where conscious measurement points forms a vector line. Then for each vector, a surrounding polygon was calculated with the distance corresponding to the working width of the machine. In practice the outer buffer zone line of the headland driving path formed the field boundary. Sowing data from the different years were compared and the biggest area was selected to represent the actual size of the field.

### 3.7 Usability

In Publication I, we processed the nitrogen need map suitable for farm machinery. The resolution of the task should correspond to the working width of the machine and the amount adjustments possible and relevant to the machine. First we sampled the nitrogen need map to 5 metre pixel size and then we applied a simple contouring method (Smith, 2013) with a 9 x 9 window Gaussian smoothing. We adjusted the contouring intervals to be 5.5 to have a reasonable amount of zones. These steps are presented in Figure 12. Then we combined the contour lines with the existing field boundaries. We added a rounded average amount based on nitrogen need map for each zone surrounded by the contours.

![Figure 12. Two simplification steps in the fertilization application task generation process](image)

To detect the effect of the usability improvement on the machinery, we calculated the difference that the simplification process made.
3.8 Overall quality and quality measures

As the basis for the overall quality estimations, we first measured a common yield variation to have a sort of base line or ground truth. The yield variation was originally presented in Publication V. The next step was to decide and select the best suitable measures for each spatial data quality element in farm machinery operations from the ISO 19157:2013 standardization point of view. The Annex D in the standard provides and defines a list of standardized data quality measures. Next we analysed them and select suitable methods for farm machinery data input perspective.

**Thematic accuracy:** the suggested misclassification matrix is not suitable, since there are no correct or true values in application tasks. We can only compare different datasets with each other. The differences are not true/false since variable rates are hierarchical.

**Temporal accuracy:** for real time data, the most suitable is value domain conformance or non-conformance. Output is Boolean (true indicates that an item is or is not in conformance with its value domain), there are no existing measures for presenting the inaccuracy that old remote sensing data produces. For this, we simply calculate a mean value of differences between classified satellite images from different dates.

**Positional accuracy:** the mean error and standard deviation are the most applicable method in the standard. On top of this, we propose Allan standard deviation plots and spatial autocorrelation representation. There are many proposed methods in the standard, such as “Absolute circular error at 90 % significance level of biased data” or “uncertainty ellipse” but those do not fit our data.

**Completeness:** the rate of missing items is the most suitable for our purpose, in our case it is missing area or excess area. An example is: 10 % (The data set has 10 % less houses than the universe of discourse).

**Usability:** the suggested measures for usability are true/false –type, but in our case, the usability for machinery is complicated and depending entirely on machinery. For this, we calculate an example of how data simplification affects the task data.

As the **overall quality measure**, we will present the spatial data quality elements diagram implemented with the relative quality measures for the recognized elements. To concretize the elements, this dissertation studies a theoretic Potential improvement of implementing different precision farming actions. This leads to a methodology of combining different quality elements in a relative simple equation representing the quality element relations and potential advantages in the work execution process.

A standard deviation quantifies the amount of variation of a set of data values. However, the standard deviation equation squares the differences emphasizing the large differences. When separate quality elements are multiplied together, a simple mean deviation could produce more stable and realistic numbers.

When we are concentrating on the difference between PF application and even rate application, we assume in this dissertation that their average rate is the same in a typical case. Then if the standard deviation of the PF application task is higher than the standard deviation of the adjusted phenomenon in the field, the result of the PF application may be harmful: the result is worse than what it would be with even rate application. Thus, the standard deviation between variable rates and true values should be smaller in a case where the usage variable rate will be beneficial.
If we have a regular field with ten different equal sized application zones, the following Figure 13 demonstrates the difference between different application rates and imaginary true values. The Y-axis shows true values in the field, the x-axis shows the application rates for variable rate, even rate and true value applications. The even rate is 100, and the PF variable rates are 75, 100 and 125. One of the application zones is misclassified. This will also mean that it is uncertain whether the variable rate is better than even rate.

Figure 13. Three theoretical fields (variable rate, even rate and true rate) each with ten application zones

In this case (Figure 13), there was 73% correct values with the even rate, and 84% of correct values with the variable rate (mean difference in relation to total amount). Of course, the real true values are not known. The standard deviation of the True values in Figure 13 was approximately 30%.

When the work overlapping amount is connected, it can be used as a multiplier. If the work overlapping for example is 15%, it means that 15% of the area got 200% treatment, while remaining 85% got 100% (in even rate applications). Thus the entire field got 1.15 times the requested amount on average. The economic payoff in this dissertation means the added value in euro corresponding to a better yield (bigger and/or better quality) and decreasing of matter inputs and decreasing fuel consumption. Other elements are not considered.
4. Results

The first results chapter 4.1 presents the concrete tools that were built and the following chapters reveal the results related to research questions one by one.

4.1 Developed tools

Two software tools were developed under the scope of this dissertation: 1) task controller commanding the ISOBUS sprayer and using external spatial data (Publication II), 2) Dynamic GNSS error simulator (Publication III). The first one related to temporal accuracy, and the second one is related to the positional accuracy. The task controller demonstrated how to joint OGC standards and ISOBUS standards in the real time farm machinery work execution to reveal the practical possibilities of temporal accuracy. The GNSS simulator process guided and evaluated the realistic GNSS error reconstruction to be able to understand the phenomenon of dynamic GNSS positioning errors.

Figure 14 (Figure 3 in Publication II) presents the interface for the task controller constructed in LabView platform. It has six different screens. The main user interface shows the VRA task and a background map, the current VRA value, the risk of rain, wind speed and direction for the decision support, the status of tractor and CAN-bus, the driving direction and information about the current work rate and applied amount and the estimated status of the sprayer’s tank. The other screens work preparations and detailed communication information from external sources and from the machinery.
The GNSS error simulator was constructed in Matlab Simulink environment. The simulator consists of five main parts. The generalized block design for latitude is presented in Figure 15. The generalized block design for longitude is similar.

![Figure 14. User interface of the developed Research-TC (Publication II)](image)

![Figure 15. Generalized block design of the constructed error simulator (adapted from Publication III)](image)
More detailed presentation is in Publication III. First are the data input blocks, which include error free coordinates, setup parameters, defined noise components and the grip map based on the spatial heterogeneity. The setup parameters define the requested level of accuracy and reliability. The user can define the following parameters: offset, standard deviation, correction reliability, positioning reliability and randomness. The second block set selects the exact noise parameters for the simulation. The reliability test block determines if the simulation drops out to the poorer mode. The simulation selection block selects the correct parameters from the Table (Publication III, Table 6) based on the selected class. The third block set simulates the errors. The main block design of different noise types is presented in the Publication III, Figure 8. Next, in the third block set the simulator sums up the separate simulations and multiplies the result of the selected noise to match it to the requested standard deviation taking into account the found difference between coordinates. The offset value is then weighted and added separately to both coordinates. The following fourth block set is the spatial autocorrelation weighting. This subsystem multiplies the amount of error according to the used map. The effect of error is distributed between latitude and longitude by giving a Gaussian random weight between 0.8 and 1.2. New random weight was calculated in every 30 seconds having effect only to longer cluster times. The last component of the error simulator is the erroneous coordinate output. The built-in delay is one measurement according to the simulation frequency. This is caused by the spatial autocorrelation determination. A complete simulation result made for a single seeding work can be seen in Publication III, Figure 12.

4.2 Thematic accuracy

This chapter presents the results related to RQ2. First we compared different fertilization tasks. The different fertilization tasks presented in the following Figure 16 were only based on the remote sensing data (Sentinel-2, Professional UAV, Commercial UAV).
The correlations between different datasets are presented in Publication V, Table 1. The average correlation of all data was 0.66, the same day consumer UAV correlation was 0.90 and the average correlation of the four most common remote sensing data for biomass mapping was 0.72 and the difference between them on average was 22 % representing 0.78 quality.
level. Those selected maps were marked with X in Figure 16; Sentinel-2 NDVI, Professional UAV NDVI, Tractor mapping and Commercial UAV VARI. Next the demonstrative task maps which combined other data to remote sensing are presented (Figure 17). The methodologies are presented accurately in Publication V:

1. Previous yield maps and Sentinel-2 NDVI data
2. Commercial UAV with RGB-camera and DroneDeploy classification
3. NDVI classification from professional UAV and supervised classification
4. Commercial UAV with RGB-camera with farmer teaching
5. Sentinel-2 VARI, 2.7 and applied Nitrogen Rate Calculator results
6. Vegetation sample interpolation

![Figure 17](image)

Figure 17. Different fertilization tasks based on remote sensing and external data (adapted from Publication V, Figure 6)

The nitrogen fertilization tasks were clearly deviating even though the average application rates were equalized. A random selection between these methodologies would probably lead to a PF task worse than an Even rate application. In this work, we assume that the selection of the suitable methodology is done with expert knowledge.

To further study the complexity of the source data and correlations, the correlations for different datasets and the yield map were calculated in Publication I and are presented in Table 4. The sum of seasonal fertilizations had 0.71 correlation to the final yield. The classified biomass map had -0.79 correlation to the yield increment: the lowest yield areas had the highest increment. When this was correlated to the absolute yield increment (kg/ha), the correlation was -0.27. None of these common methods estimated the yield increment sufficiently.
Table 4. Correlation between different datasets and yield map and estimated yield increment

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Yield map</th>
<th>Correlation to yield increase %</th>
<th>Corr. to yield increase kg/ha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of previous yield maps</td>
<td>0.17</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>yield potential map</td>
<td>-0.09</td>
<td>-0.23</td>
<td>-0.20</td>
</tr>
<tr>
<td>Nitrogen need</td>
<td>-0.11</td>
<td>0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>spring fertilization amount</td>
<td>0.34</td>
<td>-0.14</td>
<td>-0.06</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.00</td>
<td>-0.23</td>
<td>-0.13</td>
</tr>
<tr>
<td>Nitrogen map</td>
<td>0.37</td>
<td>-0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>Biomass map</td>
<td>0.30</td>
<td>-0.79</td>
<td>-0.27</td>
</tr>
<tr>
<td>Yield map</td>
<td>1.00</td>
<td>0.32</td>
<td>0.84</td>
</tr>
<tr>
<td>Additional fertilization</td>
<td>0.64</td>
<td>0.23</td>
<td>0.55</td>
</tr>
<tr>
<td>Sum of seasonal fertilizations</td>
<td>0.71</td>
<td>0.16</td>
<td>0.56</td>
</tr>
<tr>
<td>Yield increase %</td>
<td>0.32</td>
<td>1</td>
<td>0.77</td>
</tr>
<tr>
<td>Absolute yield increase (kg/ha)</td>
<td>0.84</td>
<td>0.77</td>
<td>1</td>
</tr>
</tbody>
</table>

So expert knowledge is required in order to be able to construct relevant applications.

4.3 Temporal accuracy

This chapter considers the temporal accuracy related to RQ3. The prototype task controller was constructed in Publication II. As the case study, the precision pesticide spraying case was used. The spraying amount output in the prototype task controller commanding ISOBUS sprayer used the following Equation 4 (Publication II):

\[
Amount = \frac{AR \times VRA}{100} \times GW \times R
\]  

Where \( AR \) is the decided overall application rate, \( VRA \) is the percentage value of the VRA task map in current location, \( GW \) is the ground water area map, \( R \) is the binary status of rain. If it rains, the output value is zero; the situation is the same if the machinery is in a ground water area. In practice, real-time external rain information shuts down the entire spraying process. These demonstrated methodologies on how to adapt spatial web services into the machinery work execution process in real time.

In Publication V, different remote sensing data was used to produce precision fertilization tasks. The satellite images presented in Figure 15 were analysed in terms of the effect of date difference. The correlation with Sentinel-2 based NDVI maps with 37 day difference was 0.88. After the biomass estimations were transformed to the nitrogen application rates in a straightforward way, the difference between conducted fertilization amounts based on the satellite images from different dates was 24% (Figure 18).
4.4 Positional accuracy

This chapter addresses the RQ4 and RQ5. The Allan standard deviation plots for determined dynamic GNSS positioning error and the simulated errors are presented in Figure 19. The units of the time axis $X$ is 0.1 seconds, and the unit of the $Y$-axis is $10^{-8}$ degrees in WGS84 coordinates to correspond to the operational values. The combined spatial errors with different positioning methods were: Autonomous positioning in poor conditions: 1.60 m, autonomous positioning: 1.64 m, differential positioning: 33 cm, RTK positioning in Float mode: 7 cm, RTK positioning: 8 mm. It can be seen from the plots that the simulated curves are imitating the corresponding measured ones accurately.

Figure 19. Comparison of measured and simulated data: Allan standard deviation plots of the test drives and simulation results (Adapted from Publication III, Figure 9)
More detailed results of the simulations and the verifying power spectral density comparisons are presented in Publication III. The similarity of these curves (Figure 19) support the assumption that the dynamic GNSS error consist of different noise components that we were able to determine. Our simulation varied 6.4% on average compared with complete test drives. The cost function limit of 10% decided in the simulator construction was fulfilled. The parameters of the noise terms in the simulator for the different types of positioning quality are presented in Appendix I.

The spatial autocorrelation tool used a map for multiplying the error. The map that caused the spatial heterogeneity effect is presented in Figure 20. The standard deviation of this map was 0.13 and the multiplier varied from 0.7 to 1.3. Publication III (Figure 12) shows how this background multiplier map was used in practice. In principle, the map is scaled to the work area extend, and the error is multiplied according to the temporal location of the tractor or simulated location on the map.

![Simulated error multiplier map named A-Map with spatial autocorrelation (Publication III, Figure 6)](image)

The next part considers the driving accuracy related to RQ5. The driving accuracy results were originally presented in Publication IV. First the spraying works from four different years made by two experienced drivers were studied. The evaluated 92 spraying work drives made in 17 different fields revealed 15.9% overlap on average (Table 5) and the average standard deviation of these drives was 9.0%. The average overlap of 22 combine driller drives was 7.7% and the standard deviation was 2.1%. The measured harvesting work was overlapping 1.7%. The roller and cultivator works produced large overlap with large standard deviation since the nature of the work, where different parts of the field are worked with different intensity.
depending on the soil type and conditions. The determined overwork of the spin disk fertilization was significantly smaller than the overwork of the spraying with the equal working width. The following Table 6 shows the details of the machinery that was used, the determined overwork and standard deviation amounts and the measured average overlaps in metres.

Table 5. Calculated average overwork and the standard deviation of different field works (Adapted from publication IV, Table 2.)

<table>
<thead>
<tr>
<th>Machine</th>
<th>Machine type</th>
<th>Width metres</th>
<th>Drives</th>
<th>Overwork %</th>
<th>Std. %</th>
<th>Overlap metres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardi Twin Track</td>
<td>Sprayer</td>
<td>16.0</td>
<td>92</td>
<td>15.7</td>
<td>9.0</td>
<td>2.54</td>
</tr>
<tr>
<td>Junkkari Maestro</td>
<td>Combine driller</td>
<td>4.0</td>
<td>22</td>
<td>7.7</td>
<td>2.1</td>
<td>0.31</td>
</tr>
<tr>
<td>Sampo Comia</td>
<td>Combine harvester</td>
<td>4.2</td>
<td>8</td>
<td>1.7</td>
<td>3.3</td>
<td>0.07</td>
</tr>
<tr>
<td>Kire</td>
<td>Roller</td>
<td>5.2</td>
<td>6</td>
<td>59.1</td>
<td>13.7</td>
<td>3.07</td>
</tr>
<tr>
<td>Amazone BBG Carrier</td>
<td>Cultivator</td>
<td>3.0</td>
<td>6</td>
<td>19.0</td>
<td>7.5</td>
<td>0.57</td>
</tr>
<tr>
<td>Bögballe DZ Trend</td>
<td>Spin disk fertilizer</td>
<td>16.0</td>
<td>6</td>
<td>9.5</td>
<td>8.8</td>
<td>1.52</td>
</tr>
</tbody>
</table>

Next the overlapping components are presented. In the following Figure 20, the dominant green colour represents successful spraying; all other colours are representing different types of work overlapping. In this field, pass-to-pass overlapping was less than 5 % being about 75 cm in practice with a 16 metre wide sprayer boom. The presented overlapping types in Figure 21 are as follows:

1) overlapping the parallel driving lines
2) overlapping last driving line of the session
3) working outside the field boundaries (spraying ditches)
4) overlapping before and after the headland turns
5) headland overlap because of the gentle enter/exit angle (over 45 degrees)
Figure 21. Classified overlapping of the performed spraying work in the large field (adapted from Publication IV, Figure 6)

Different overlapping classes for spraying are shown in Table 6 for three different fields and on average for all the 17 fields that were studied. The successful class named ‘OK’ in Table 6 presents the area that got a single treatment according to the plan. The inaccurate parallel driving caused 10.3 % overlap on average in the spraying while the headland overlapping was less than 2 %.
<table>
<thead>
<tr>
<th>Field (ha)</th>
<th>OK (%)</th>
<th>1 (%)</th>
<th>2 (%)</th>
<th>3 (%)</th>
<th>4 (%)</th>
<th>5 (%)</th>
<th>Gaps (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8</td>
<td>77.6</td>
<td>18.5</td>
<td>-0.9</td>
<td>-0.2</td>
<td>1.6</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>5.4</td>
<td>88.5</td>
<td>8.8</td>
<td>1.1</td>
<td>0.0</td>
<td>1.3</td>
<td>0.8</td>
<td>0.0</td>
</tr>
<tr>
<td>22.4</td>
<td>83.4</td>
<td>8.1</td>
<td>2.6</td>
<td>0.5</td>
<td>3.4</td>
<td>1.7</td>
<td>0.1</td>
</tr>
<tr>
<td>ave. 5.4</td>
<td>85.3</td>
<td>10.1</td>
<td>1.4</td>
<td>0.0</td>
<td>1.7</td>
<td>1.4</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The single automatic steering test comparison revealed that the overwork amount without automatic steering was 8.3 % (parallel passes 57) and with automatic steering was 4.3 % (54 parallel passes). Theoretical calculations reveal that a plain steering assistance with 15 cm pass-to-pass accuracy would decrease the overall spraying work overlapping amount from 15 % to about 7 %. A section control would decrease the overlapping up to 3-5 %, depending on the amount of controllable sections. Roughly, the steering assistance cuts the overlapping in half.

### 4.5 Completeness

This chapter gives a partial answer to RQ6. The sowing areas (seeded area) from different years were compared in Publication IV. We used 38 sowing works in total. The average standard deviation of these buffered areas for the same fields was 0.03 ha. The determined field sizes were 0.14 ha smaller than the official field polygons in average, representing 3 % total official area of our test fields. This means that our test fields were 3 % smaller than they were in official Land Parcel Identification System (LPIS). Figure 22 shows all the determined area similarities in percentage. To clarify the image, a large 25.7 ha field with 99.2 % similarity (0.20 ha difference) is not presented. With this dataset, the field size did not correlate with the difference to the official field size meaning that the shape complexity of the field might be more important.

![Figure 22. Official field size compared to area that was worked each year](image)
4.6 Usability

The fertilizer application task calculated in Publication I is shown in Figure 23 side by side with the ‘not simplified’ application task. The difference between this task and data it was simplified from was 23.2 % on average. In other words, after the application task map was modified to machinery friendly (spatial and value simplification), the task values changed 23.2 % on average.

![Figure 23](image)

**Figure 23.** Source map and constructed application task for the nitrogen fertilization

4.7 Overall quality

This chapter combines the overall quality stated in RQ1 and RQ6. But first, an average yield variation was calculated. The average of the yield amount variation in the test fields was 32.7 % (Publication V). The largest test field’s yield had a variation of 23.3 % and the average of them is 28 % representing the expected variation in the field later in this work. These values give an idea on how much there is variation within-field under regular farming practices. Next we compiled the spatial data quality components. Figure 24 illustrates the effect of different elements.

![Figure 24](image)

**Figure 24.** The effect of different spatial data quality elements
The presented values are as follows: Completeness: 97\% since the official field sizes are 3\% too large. Thematic accuracy: 78\% since there was 22\% difference between common classifications for the same purpose. The logical consistency was ok. The temporal quality was 0\% or 100\%, according to the existence of rain information, other aspect is that temporal quality is between 100-76\%, because the satellite images from the different dates conducted a 24\% difference in application rates. The positional accuracy was 99-60\% when done with GNSS steering assistance (RTK error 8 mm plus the headland turnings) and 1.6 m autonomous GNSS error is 40\% of the width of the combine driller (4 m). Other option for positional accuracy is 92-84\%, because the overlapping rates of the farmer in regular drives were between 7.7-15.6\%. The usability element was 89\%, since data changed 23\% while it was converted to machinery friendly in the case study and that was split in half because of the fact that the thematic accuracy and the usability element may be linked and when multiplied together the errors could be exaggerated.

In general, these elements are cumulative from the overall quality perspective. If these quality element measures are multiplied, the overall quality is between 78\% and 31\%. A sophisticated example result would be: 0.97*0.78*1.00*1.00*0.90*0.89 = 61\%. That would indicate the quality of spatial data inputs in farm machinery work execution representing the accuracy level of the inputs.

To retrace more concrete values, the following Equation 5 for the potential improvement of the precision farming work execution was constructed in this dissertation:

\[
E = \frac{(A+I)}{2} \left( UMT (Q - 1 + C) \right) + S (D - 1) (F + IM),
\]

Where:

- \(E\) = Potential improvement of implementing PF technologies €/ha
- \(A\) = Estimated typical yield profit 1 000€/ha as a result of 5 000 t/ha typical average yield and the prize of the grain in the markets
- \(I\) = Cost of inputs €/ha according to current prices
- \(U\) = Estimated usability level, the map simplification process (usability element)
- \(M\) = Machinery performance, assumed to be 0.98 indicating how well the machinery performs (for example spreads as it should spread; uncertainty of the work pattern)
- \(T\) = Temporal accuracy: considered to be 100\% (1.0)
- \(Q\) = Quality of the application map (thematic accuracy) (Publication I and V)
- \(C\) = Existing field and yield variation, 0.28 variation (conducted from Publication V)
- \(D\) = Measured overlap rate: double application rates (Publication IV)
- \(S\) = Steering assistance performance: approximately 0.5 (conducted from Publication II and IV)
- \(F\) = Measured fuel consumption €/ha (conducted from Publication IV)

The completeness is not considered here since it can be estimated to be influencing only the overall input needs, not to the PF performance as such. Also the quality of temporal accuracy element is considered to be 100\% in this equation. More detailed explanations of the equation 5 are presented in Appendix II. The following Table 7 shows the results of the usage of the Equation 5.
Table 7. Estimated payoffs of precision farming task execution

<table>
<thead>
<tr>
<th></th>
<th>Cultivation</th>
<th>Startup fertilization</th>
<th>Seeding</th>
<th>Additional fertilization</th>
<th>Weed Control</th>
<th>Pesticide Control</th>
<th>Harvest</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Map quality (Q)</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>-</td>
</tr>
<tr>
<td>2. Field variation (C)</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
<td>-</td>
</tr>
<tr>
<td>3. Estimated increase</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>4. Overwork (O)</td>
<td>1.353</td>
<td>1.077</td>
<td>1.095</td>
<td>1.157</td>
<td>1.157</td>
<td>1.017</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5. Fuel €/ha (F)</td>
<td>10.84</td>
<td>3.88</td>
<td>2.92</td>
<td>2.63</td>
<td>2.63</td>
<td>8.81</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>6. Usability (U)</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.9</td>
</tr>
<tr>
<td>7. Inputs €/ha (I)</td>
<td>0</td>
<td>200</td>
<td>150</td>
<td>70</td>
<td>30</td>
<td>30</td>
<td>0</td>
<td>69</td>
</tr>
<tr>
<td>8. PF €/ha</td>
<td>26.2</td>
<td>31.4</td>
<td>30.1</td>
<td>28.0</td>
<td>27.0</td>
<td>27.0</td>
<td>26.2</td>
<td>28.0</td>
</tr>
<tr>
<td>9. Steering assistance €/ha</td>
<td>1.9</td>
<td>7.5+0.2</td>
<td>5.7</td>
<td>3.4</td>
<td>2.5</td>
<td>2.5</td>
<td>0.1</td>
<td>24</td>
</tr>
<tr>
<td>10. €/ha (E)</td>
<td>28.1</td>
<td>38.9</td>
<td>35.8</td>
<td>31.4</td>
<td>29.5</td>
<td>29.5</td>
<td>26.2</td>
<td>219</td>
</tr>
<tr>
<td>11. If usability=1</td>
<td>+3</td>
<td>+4</td>
<td>+4</td>
<td>+3</td>
<td>+3</td>
<td>+3</td>
<td>+3</td>
<td>+24</td>
</tr>
<tr>
<td>12. If thematic=1</td>
<td>+96</td>
<td>+115</td>
<td>+110</td>
<td>+103</td>
<td>+99</td>
<td>+99</td>
<td>+96</td>
<td>+718</td>
</tr>
<tr>
<td>13. If 1/2 fuel</td>
<td>+5</td>
<td>+2</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+4</td>
<td>+16</td>
</tr>
</tbody>
</table>

The rows at the Table 7 are: 1) Quality of the application map, representing the found thematic accuracy in contrast to the application $Q$, 2) The estimated field variation $C$, 3) the estimated realistic yield increment based on first and second columns, 4) overwork amount in the specific application (Publication IV) $O$, 5) Measured fuel consumption during the task €/ha $F$, 6) the estimated relative cost of map simplification meaning the Usability $U$, 7) the cost of application inputs €/ha such as seeds and fertilizer $I$, 8) the estimated benefits of the PF actions €/ha without taking account the investments, 9) the estimated benefits of steering assistance or automation (€/ha), 10) the estimated total benefits €/ha. 11) if the usability would be 100 %, how much more would be the payoff, 12) If the thematic map ($Q$) is 100 %, how much more would the payoff be, 13) if the machinery fuel consumption would be cut in half, how much more would the economic payoff be. The start-up fertilization and seeding are often done simultaneously with the combine driller, so part of the columns are combined.

The variables $C$, $M$ and $Q$ are overall and general cautious estimations, not case specific measures. Tuning them based on real experiment data could make great differences in individual operational cases in precision farming. The concrete benefits out of each task would be as follows:
**Cultivation:** work intensity in relation to classified soil; bigger yield

**Start-up fertilization:** Precision fertilization rates based on existing nutrient levels; bigger yield, better quality

**Seeding:** Variable seeding density and depth in relation to spatial growth potential; better respond to yield

**Additional fertilization:** Fertilization based on crop growth; bigger yield, better quality

**Weed control:** Weed detection, better growth conditions; bigger yield

**Pesticide control:** Pest mapping; bigger yield, better quality

**Harvest:** Grain quality map (silage or malt); better quality

The yield increment potentials are cautious especially with weed and pesticide control, where potentially 90% of the inputs may be reduced with correct technologies. When the average results of the Table 7 are considered, the linear equations of $E = 466Q - 332.15$ and $E = 466C - 99.147$ can be approximated. These show the effect of the expected yield variation $C$ and the Thematic quality $Q$. These will show how much euros will be expected with certain yield potential expectations with the selected values.

The estimated average yield profit was 31€/ha. If all PF actions would be taken, the economic payoff was estimated to be 219 €/ha and the cumulative increase would be 1.36. That would mean a 5900 kg/ha yield (instead of 5 000 kg/ha) and 18% less input matter costs. The cumulative increase in the yield (based on each successful PF execution) is not taken into account in the economic payoff rows meaning that a single PF execution does not assume that other PF executions are also done. Similarly, the decrease of the field variations caused by the other PF actions is not taken into account (in the optimal PF case, $C$ in harvest column would be 0).

The last two rows of the Table 7 shows that having an optimal quality application map would be much more profitable than cutting the tractor fuel consumption in half inside the fields.
5. Discussion

The usage of data quality approach to measure uncertainty was feasible. As this dissertation demonstrated, the common spatial data quality elements can fit the farm machinery input data quality estimations, where thematic accuracy, temporal quality, positional accuracy and usability are the most valid elements. The completeness refers to the spatial field coverage of the application task. The thematic accuracy means the accuracy of the application task values or the variation between them because there is no proper ground truth. The temporal quality means how up-to-date data is used for the machinery adjustments. The positional accuracy means the awareness of the machinery location in relation to application tasks and parallel driving lines. The final usability element refers to the task map simplification for the machinery.

The difference between tasks was 22 % on average with most common data and with the simplest application task generation method, where relative biomass estimation was turned directly into the fertilization amount. The presented correlation of the tasks based on different remote sensing data was 0.66. When different source data and methodologies were applied, the results were completely deviating (Figure 17) and there were no meaningful correlations. All the tested remote sensing methods managed to estimate the relative differences of biomass. Similarly, Pena-Yewtukhiw et al. (2015) found that even a slight difference in the single task generation parameter could produce a great difference in the end. When we used other parameters for task generations, the differences were large even when based on visual estimations. Comparisons to the vegetation samples had very poor results, there was hardly a correlation. This might be caused by inaccuracies in the collected sample area, sample point positional accuracy, local vegetation and soil variations, spatial accuracy of remote sensing data. The correlations for different datasets and the yield map demonstrate that a fruitful fertilization plan would be very difficult to produce. Even in a single season examinations, the correlations were very low. The sum of fertilizations had 0.71 correlation to the final yield, this means that if there were more fertilizers, there was also more yield, but that was maybe the only remarkable note. Very interesting was that the classified biomass map had a -0.79 correlation to the relative yield increment (%). This could tell something about the hyperspectral biomass determination or about the growing conditions during that season: high biomass in the middle of the summer was not growing well anymore in the late summer.

If the VRA map quality element in the Equation 2 (which sums up the economic payoff) is replaced by the optimal values, the overall payoff increases from 219 €/ha to 937 €/ha. This implicates the significance of the application task quality and the potential of PF applications.
Publication II and chapter 4.3 presented a prototype on how to improve especially temporal accuracy in decision making during the field operation. For the ISOBUS standardization, this kind of task execution would require changes. First, the OGC-related standards as the map format should be enabled and second, the TC should be able to fuse multiple tasks to a single command as presented. However, this kind of solution could be utilized in many ways. Some examples of beneficial data sources are listed below:

- On-board sensors
- Weather and forecast service (rain, wind, temperature, heat sum)
- Disease pressure services
- Real time remote sensing data services
- The work of other working units
- Sensors on the field (infestation, moisture, etc.)
- Sensitive environment (ground water, neighbouring plants and crops)
- Machine parameters (external measurements and calibrations)
- Other external risks (pests, heavy rains, flood, fire)
- Other location based services (LBS) (e.g. neighbouring information)

The optimal timing for the additional fertilization would have been in the middle of July, but even the Sentinel-2 NDVI-map in early June estimated visually correctly the relative biomass. The visual study of Figure 18. shows that Sentinel-2 data from 2. June and 9. July are giving very similar information. This is indicating that these NDVI-level differences can be spotted even in a very early stage of growth at least in this case. The measured difference was 24%. From the other perspective, the temporal information can be binary to the farming process: machinery output is null if there is rain detected. However, many of the possibly real-time delivered information could be like this, such as information about infestations, pests, and other machinery works. Services such as predictive disease models may have an operable temporal span from hours to days.

With vegetation mapping and remote sensing, the plants growth status changes all the time, and the changes are increasing while the crops are growing. For example there are 10 main stages and 94 different codes in the Zadoks scale (Zadoks et al., 1974). This means that the crop goes in the new stage almost weekly. There could be an assumption that remote sensing could give more accurate information, if the sensing took place at the same stage or maybe even at the same code, when the planned field actions are taken. The within-field variations of the growth status is also a challenge, like when flag leaf sheath opening is happening in some parts of the field, that could exaggerate the differences in greenness.

The structure of GNSS positioning error was presented in Table 3, Figure 19 and Appendix I. The mean error of dynamic GNSS positioning with different GNSS correction methods was what expected. The errors varied from 160 cm to 1 cm. The Allan variance plots and regenerations gave good results and we were able to demonstrate that dynamic GNSS positioning error consist of five noise components, based on: White Noise, Random Walk, Flicker Noise, Gauss Markov Process, Outliers and Rate Ramps supported with the spatial autocorrelation effect. Unfortunately, we also found that there can be relatively huge differences within even a single test drive if it is split. This means that the actual noise components can be very temporal and can vary a lot. Our developed spatial autocorrelation effect managed to have similar type of long term variation. However we only manage to show how the spatial autocorrelation
can be taken account. It would require more tests to determine a more generic way how the spatial heterogeneity affects the positioning and how it can be scaled. In our simulations, the spatial heterogeneity map influenced equally the different noise components. In practice, the case might be different. The spatial autocorrelation value could also have had an effect on the reliability setup values by increasing the possibility of correction loss.

We managed to construct a simulator which can simulate GNSS errors according to found Allan variance plots. The overall PSD slopes were also close to each other verifying the correct trend. The constructed errors seem to model the actual error fairly well. However, there was a great difference between the Allan variances of first and second halves of the measurement data (17.3 % on average, Publication III). Also the difference between uncorrected GPS positioning and uncorrected GPS positioning in bad conditions was about 26 %. This means that there is a great variation between different drives, and the Allan variances are highly depending on the temporal positioning conditions during the tests. Our cost function limit for parameter optimization (less than 10 %) was much smaller meaning that we are simulating the real test drives accurately. These figures are indicating that we can replicate our test drives, but tuning the variation and tolerances of simulator components to the universalizable level would require massive amount of test drives. However, the randomizing parameters in the simulating process made a great variation in total even without changing the input parameters. Especially the spatial heterogeneity map affected really in the longer cluster times, where the Allan variance differences between test drives were much greater. This means that taking into account the spatial autocorrelation improves the realism of the simulation. In total, the difference between test drives and the difference and variation within simulations were close to each other meaning that the simulator is acting very realistically.

The assumption that farmers tend to overlap their work about 10 % of the implements width matched very well to our finding of the spraying line overlapping of 10.3 %. This value included the overlapping while driving the headlands and overlapping caused by obstacles in the field. Kvíz et al. (2014) found that the regular overlap of passes was in the range between 1 % and 6 %. When we examined our pass-to-pass spraying drive accuracy, we got similar values: 3.5 % and 4.7 %. This would indicate that using steering automation in the middle of the field as they usually are applied would not remove all the overlapping. The overlapping with spin disk fertilization was almost half compared with the spraying. This could happen because the driver is not trying to avoid the gaps more than the overlapping. These results also put the possibility to apply controlled traffic CTF in a new light.

Based on our overwork classification results, 69 % of the overlapping work was made in parallel driving. The usage of GNSS assistance would easily cut the amount of overlapping in half. The section control was calculated to decrease the amount of overlap by 30 %.

The measured over consumption of pesticides was 16 %, seeds 8 % and fertilizers was 10 %. The case is similar with time and fuel consumption. The field size did not correlate with the amount of overlap. The automatic steering comparison in combine drilling was made in a single field.

This dissertation demonstrated the quality of farm machinery input data by using spatial data quality standards. There are still uncertainties that we were not able to tackle, mostly those are related to source data of farming machine applications that were not in our focus, such as seeding and startup fertilization and the knowledge of the existing variation in the field. Those uncertainties will propagate in the complete field management. Basically, this work only demonstrated guidelines with single cases about the quality. Datasets in this dis-
sertation were mostly based on different case studies. It can be difficult or impossible to draw general conclusions out of them. However, there have not been any better or more precise methods dealing with the overall spatial data quality or uncertainty in this subject. The estimated yield potential values in Table 7 are rather rough estimations and are based mostly on geographical data management findings and personal farmer discussions. Those could need some iterative considerations in future research. However, the related equation and Table 7 shows a practical way how to apply quality measures. The payoff of individual tasks (additional fertilization (1.05) fits the level of 5% (Nissen, 2012) in existing applications.

The missing ground truth in many cases and/or small amount of measurement data together with highly case dependent demonstrations weakens our conclusive results, but most importantly; this dissertation managed to present an overall view of the area under study. According to the findings in Publications (I-V), the spatial data quality in farm machinery work execution is 61% accurate. The key problems that are causing it are: 1) the precision application task; there are no best ways to do it and the machines should be able to do it accurately and in detail, 2) the positional accuracy especially when the driver does not use any steering guidance systems.

All these case studies were made in Finland. There can be some elements that are exaggerated because of that. In Finland, the field sizes are relative small and the shape of the fields are very irregular making it difficult to drive wisely in the field. The GNSS positioning conditions are somewhat poorer than in more southern latitudes. The illumination conditions for the remote sensing instruments are a bit different than on average because of the low angle of the sun. Despite all of those, the general trends and relative quality levels are assumed to be similar.

To be able to put and evaluate the different quality elements together, the determined quality elements were presented as percentages. This might not be very informative with individual cases since it depends on the starting values and the values that they are compared. However, for the single separate cases, the concrete values are presented in the results section especially in the different Publications (I-V).

Table 7 does not suit well for the conclusive estimations where all the PF actions would have been taken, since each execution will increase the yield but also decrease the variation in the field. If those were linearly taken account, all the PF actions in the theoretical case would produce 5900 kg/ha yield which is feasible.

Accurate positioning, machinery section controls and headland automation will significantly increase the work quality. From PF perspective, if the quality of positional accuracy is fixed, a proper field monitoring can be done. That would mean that it is possible to know, what has been done in each location at the field and that the inputs are more controlled (for example less double treatments). This would make the field less complicated which makes it easier to develop the PF applications and make the field level applications respond better to the field trial and split plot results. Still new and massive field trials are needed in order to understand the relation between different quality factors of crops and their spectral reflectance. The suggestion for future PF research is to first have an overall data collection infra, such as Cropinfra (Pesonen et al. 2012, Backman et al. 2019), on top of that there is a need to improve data quality with accurate positioning and headland automation and then combine the entire system with field trial infrastructure, where the knowledge for thematic and temporal accuracy is gained and beneficial PF applications can be designed. A missing part of this chain will leave propagating uncertainties.
As the FAO (2017) presented the need for producing 50 percent more food (48 % if Sub-Saharan Africa and South Asia is excluded), resolving spatial data uncertainty could help to meet the demand. Solving uncertainty related to thematic accuracy is the key and in the end it requires managing the seasonal weather forecast in relation to the knowledge of crop growth and soil properties. Technological solutions such as drones capable of continuously measuring the growth status and drones capable of spreading nutrients and pesticides non-destructively and on demand together with smart mixed cropping environment could provide possible tools. However, the unpredictability of weather may always leave some uncertainty to the elements.
6. Conclusions

This work focused on determining the spatial data quality elements in farm machinery input data to retrace the involved uncertainty. We used spatial data quality standards to present thematic, temporal, positional, completeness and usability elements in input data. The thematic quality is very poor and complicated. There are no straight forward ways to determine accurate precision application rates, such as additional fertilization amounts, even within very similar datasets, the thematic variation was 22 % on average. The temporal quality is critical if the work process requires real-time data. In some cases even one month old satellite image could provide sufficient information. The positional quality consists of two main elements: the positional accuracy and the driving accuracy in the field. The farmer can easily overlap 10 % of the working width. The completeness aspect focuses on the accuracy of the field size, which proved to be 3 % smaller in practice compared with the official sizes. The usability here means that how can the machinery use the application data. This means that the variable rate application needs to be simplified so that machinery adjustments and working width are realistic. This simplification changed the application task for 23 %.

According to our measurements, our test fields had a 30 % yield variation in average. There should be plenty of room for precision farming activities and there should clearly be visible variations in within-field. Those relative differences were easy to determine with different remote sensing methods, but there is huge step needed to use these biomass variations in a consistent way. Just picking up a drone or a free satellite image would possibly not give a sufficient knowledge for the additional fertilization.

The finding that a representative value of the overall spatial data input accuracy was 61 % is understandable but alarming. It represents the farming inputs: if the 100 kg/ha of fertilization was the aim, on average the input was 61 kg/ha or 164 kg/ha, even before the actual machinery input uncertainty is added. There are lot of things in regular and precision farming that still needs to be improved. According to my current understanding, about half of the problems are caused by the lack of biological knowledge, but the other half is caused by the current technologies.

The result that our fields were in average 0.14 ha smaller than the official field borders was remarkable for the farmer’s bureaucracy. The farmers have to inform all of their field plots with 0.01 ha accuracy. This problem can be compared with living area of a house vs. the house footprint. The farmer cultivates the net size of the field, but the actual size includes ditches. The situation is the same for all the Finnish farmers.
The economic payoffs were at the expected levels. The presented equation gives a tool to exploit found quality levels in a practical way. So, what regular farmer can do? Investing proper steering assistance will make a solid improvement. There are numerous tempting possibilities to produce precision application tasks, but the farmer should really know what he/she is using. In arable farming, the future weather has the last word. In the precision farming, the lack of sensors is not an issue anymore, it is the lack of knowledge how to use it; the lack of smart decision support systems. The summer 2017 was wet and cold in Finland, the grass was growing nicely and the amount of available nitrogen was limiting the growth. The summer 2018 was relatively dry and hot in Finland. The amount of nitrogen hardly affected on the grass yield while the lack of water resources was significantly limiting the growth. The occurrences of this kind emphasize the prediction needs in contrast to small parameter tuning in ICT environment.

The spatial data quality elements were not a starting point of the different Publications. Rather than that, it was a natural consequence when the different found items that formed the elements were combined. This can be seen as a complexity of the results in contrast of the pure quality evaluation, but on the other hand it was necessary due to a continuous learning process. One way to continue this work would be finding the measured values in different PF applications instead of using estimated values in some of the variables in the Equation 2. Other possibility to continue this work is to measure or simulate the real machinery work inputs in the field. The framework for such approach exists, the work Kaivosoja and Linkolahti (2012) already simulated the spatial variations of spin disc fertilization work distribution that was combining 2D tray test results informed by the machinery manufacturer and the collected driving data. The uncertainties presented in this dissertation propagate in the actual work mapping and thus it was seen fair to manage those uncertainties before. Combining the machinery input quality at this stage could have misled the focus too much to the machinery performance.

Although this work was about the uncertainty of machinery input data, basically this work presented the knowledge gap between the results of field trials in contrast to results with entire field actions. To continue with the uncertainty studies, next steps would be to update the existing field variations and then study the uncertainty related to thematic accuracy with all the farm machinery applications.
References


He, B., Fang, T., Guo, D., 2004. Quality assessment and uncertainty handling in spatial data mining, Proc. 12th Int. Conf. on Geoinformatics, Geospatial Information Research: Bridging the Pacific and Atlantic, University of Gävle, Sweden, 7-9 June 2004


Jakobsson, 2006. On the future of topographic base information management in Finland and Europe, Aalto University publication series Doctoral Dissertations, Maanmittauslaitoksen julkaisuja, 101


Kumhálová, J., Matějková, S., 2017. Yield variability prediction by remote sensing sensors with different spatial resolution. International Agrophysics, 31(2) 195-202


APPENDIX I

Noise components for the simulations (Adapted from Publication III, Table 6)

The noise components for a single noise simulation are presented in one vertical line. The unit of update rate is 10 Hz.

<table>
<thead>
<tr>
<th>Noise term:</th>
<th>Autonomous GPS poor lat</th>
<th>Autonomous GPS poor lon</th>
<th>Autonomous GPS lat</th>
<th>Autonomous GPS lon</th>
<th>DGPS lat</th>
<th>DGPS lon</th>
<th>RTK poor lat</th>
<th>RTK poor lon</th>
<th>RTK lat</th>
<th>RTK lon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update rate</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Update probability</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Reset time</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>334</td>
<td>44</td>
<td>29</td>
<td>31</td>
<td>51</td>
<td>52</td>
<td>2.4</td>
<td>1.4</td>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>Noise type</td>
<td>RW</td>
<td>WN</td>
<td>WN</td>
<td>FN</td>
<td>WN</td>
<td>WN</td>
<td>WN</td>
<td>WN</td>
<td>WN</td>
<td>WN</td>
</tr>
<tr>
<td>Update rate</td>
<td>59</td>
<td>150</td>
<td>640</td>
<td>10</td>
<td>2</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Update probability</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Reset time</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>350</td>
<td>310</td>
<td>690</td>
<td>104</td>
<td>16</td>
<td>16</td>
<td>1.4</td>
<td>1.3</td>
<td>3</td>
<td>2.5</td>
</tr>
<tr>
<td>Noise type</td>
<td>GM</td>
<td>GM</td>
<td>GM</td>
<td>RW</td>
<td>WN</td>
<td>FN</td>
<td>WN</td>
<td>WN</td>
<td>WN</td>
<td>FN</td>
</tr>
<tr>
<td>Update rate</td>
<td>2000</td>
<td>100</td>
<td>3200</td>
<td>63</td>
<td>19</td>
<td>440</td>
<td>6</td>
<td>6</td>
<td>35</td>
<td>150</td>
</tr>
<tr>
<td>Update probability</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reset time</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>570</td>
<td>150</td>
<td>620</td>
<td>62</td>
<td>137</td>
<td>68</td>
<td>25</td>
<td>6</td>
<td>2.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Noise type</td>
<td>GM</td>
<td>GM</td>
<td>GM</td>
<td>RW</td>
<td>GM</td>
<td>WN</td>
<td>GM</td>
<td>RN</td>
<td>RN</td>
<td>FN</td>
</tr>
<tr>
<td>Update rate</td>
<td>5800</td>
<td>390</td>
<td>190</td>
<td>1020</td>
<td>1200</td>
<td>130</td>
<td>350</td>
<td>511</td>
<td>890</td>
<td></td>
</tr>
<tr>
<td>Update probability</td>
<td>0.5</td>
<td>0.01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reset time</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>350</td>
<td>195</td>
<td>110</td>
<td>190</td>
<td>55</td>
<td>-3</td>
<td>-3.5</td>
<td>7.9</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Noise type</td>
<td>OL</td>
<td>WN</td>
<td>GM</td>
<td>GM</td>
<td>GM</td>
<td>RR</td>
<td>RR</td>
<td>RR</td>
<td>GM</td>
<td></td>
</tr>
<tr>
<td>Update rate</td>
<td>2900</td>
<td>1000</td>
<td>950</td>
<td>500</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Update probability</td>
<td>0.1</td>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reset time</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2900</td>
<td>100</td>
<td>115</td>
<td>8</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise type</td>
<td>OL</td>
<td>OL</td>
<td>GM</td>
<td>RR</td>
<td>RR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RW=Random Walk  GM=Gauss Markov Process
WN=White Noise  RR=Rate Ramp
APPENDIX II

Equation (5)

\[
E = \frac{(A+I)}{2} \left( UMT (Q - 1 + C) \right) + S (D - 1) (F + IM)
\]

\(E\)=Estimated euros, payoff €/ha
\(A\)=Typical yield profit, 1000€/ha
\(Q\)=Map quality, about 0.78. (temporal quality)
\(C\)=Variation in the field, estimated to be 0.28 (Publication V)
\(I\)=Typical cost of machinery inputs, 0-200 €/ha
\(U\)=Usability, the map simplification for the machinery, 0.89
\(D\)=Double treatment, the measured overlap rate, about 1.10
\(M\)=Machinery accuracy, 0.98
\(S\)=Steering assistance performance, 0.5 (cut the overlapping in half) Publication IV
\(F\)=Fuel consumption, 2.6-10.8 €/ha Publication IV
\(T\)=Temporal accuracy 1.0

a) Half of the estimated benefit is more yield, half of it is less machinery inputs
b) Usability \((U)\) smoothes the effect of application task, it’s not an overall multiplier
c) Machinery accuracy \((M)\) smoothes the effect of application task but also weakens the machinery work in driving
d) \(Q\)-1+C is positive when the map quality is high and the field variation is high. Not simplified equation is \(C \cdot (1/C) \cdot (Q-1-1)\). The outcome is negative, when the \(Q+C\) is less than 1, if \(Q=0.78\) and \(C=0.2\) for example. Meaning that the task is deviating from the true values more than there is variation in the field. In these cases, an even rate application would be more beneficial
e) Using the steering assistance cut the overlapping in half (Publication IV). It is possible to replace this \(S\) with more precise values from Publication III and IV if for example a specific automatic steering is considered
Role of spatial data uncertainty in executions of precision farming operations

Jere Kaivosoja