UNIVERSITY OF HELSINKI
FACULTY OF AGRICULTURE AND FORESTRY

Analysing the reliability of forecast information provided by UNECE member States

Master's thesis

Markus Stolze

May 2019

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Agriculture and Forestry	Department of Forest Sciences

Tekijä — Författare — Author

Markus Mikael Stolze

Työn nimi — Arbetets titel — Title

Analysing the reliability of forecast information provided by UNECE member States

Oppiaine —Läroämne — Subject

Forest Products Management

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	May 2019	59

Tiivistelmä — Referat — Abstract

The purpose of this master's thesis is to evaluate the reliability of forest products forecast information produced by United Nations Economic Commission for Europe member States. The study also aims to answer which dimensions of data quality are the most important when producing these predictions

This study is carried out as quantitative research and it focuses on the predictions made by the 27 member States, produced between 2002 and 2017. This research aims to find out what methods are used by different member States and which methods produce the most reliable results. This research also aims to find out if there are any differences in reliability when assessing different product flows (removals, production, exports or imports) of the various products analyzed.

There were clear differences visible between different products in the results of this research. In some products, almost all member States had managed to produce reliable predictions, while for others majority of member States didn't manage that. There were also differences between member States and some were clearly more reliable than others. The biggest factor affecting reliability was volume: for most parts, bigger volumes meant more reliable predictions. Production and removals were more reliable product flow than imports or exports. This is due to the nature of imports and exports, as they are more easily affected by outside impacts.

Although all member States were able to be sorted into four groups based on how different product flows looked like, no clear patterns were visible when observing how different member States produce predictions. Almost all of the interviewed representatives of member States reported that they were using almost or exactly the same methods to produce predictions.

Avainsanat — Nyckelord — Keywords

Forest industry, forest products, forecast, predictions, UNECE, FAO, FAOstat, data quality, reliability

Säilytyspaikka — Förvaringsställe — Where deposited

HELDA/E-thesis [ethesis.helsinki.fi/en]

Muita tietoja — Övriga uppgifter — Further information

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Maatalous-metsätieteellinen tiedekunta	Metsätieteiden laitos

Tekijä — Författare — Author

Markus Mikael Stolze

Työn nimi — Arbetets titel — Title

UNECE:n jäsenmaiden tekemien metsäteollisuuden ennusteiden luotettavuuden analysointi

Oppiaine —Läroämne — Subject

Metsäteollisuuden markkinointi ja johtaminen

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master’s thesis	Toukokuu 2019	59

Tiivistelmä — Referat — Abstract

Tämän pro gradu -tutkielman tarkoituksena on selvittää, kuinka luotettavia Euroopan Talouskomission jäsenmaiden tekemät metsäteollisuuden ennusteet ovat. Tutkimuksessa halutaan myös selvittää, mitkä tekijät vaikuttavat tutkimuksessa käytettävän datan laatuun.

Tutkimus on määrällinen ja siinä keskitytään 27 jäsenmaan tekemiin ennusteisiin 12 tuotteen osalta, vuosien 2002 ja 2017 välillä. Tutkimuksessa selvitetään, millaisia keinoja ennusteiden tuottamiseen käytetään eri jäsenmaissa ja onko tietyt tavat ennusteiden tuottamiseen parempia kuin toiset. Samalla selvitetään, onko viennin, tuonnin, tuotannon ja harvennusten välillä eroavaisuuksia luottavuuden osalta, sekä miten luotettavuus vaihtelee eri tuotteiden osalta.

Tutkimuksessa havaittiin selviä eroavaisuuksia eri tuotteiden välillä. Osassa tuotteista lähes kaikki jäsenmaat olivat onnistuneet tekemään luotettavan ennusteen, kun taas joissain tuotteessa ainoastaan muutama jäsenmaa onnistui hyvin. Myös eri jäsenmaiden välillä oli selviä eroavaisuuksia. Suurin vaikuttava tekijä jäsenmaiden tekemien ennusteiden luotettavuuteen oli määrä: mitä suuremmat määrät tuotteita käsiteltiin, sitä varmemmin ennusteen luotettavuus pysyi hyvänä. Tuotanto ja harvennukset olivat varmemmin luotettavia, kuin tuonti ja vienti, jotka reagoivat helpommin ulkopuolisiin muutoksiin.

Vaikka jäsenmaat voitiin jakaa neljään ryhmään tulosten perusteella, ei selviä eroavaisuuksia havaittu keinoissa miten eri jäsenmaat tuottavat ennusteita. Vaan lähes kaikki haastatellut jäsenmaan edustajat kertoivat käyttävänsä lähes tai täysin samoja keinoja ennusteiden tuottamiseen.

Avainsanat — Nyckelord — Keywords

Metsäteollisuus, metsätuotteet, ennuste, UNECE, FAO, FAOstat, datan laatu, luotettavuus

Säilytyspaikka — Förvaringsställe — Where deposited

HELDA/E-thesis [ethesis.helsinki.fi/en]

Muita tietoja — Övriga uppgifter — Further information

Table of contents

List of Figures

List of Tables

Acknowledgements

First of all, I'd like to express my gratitude to my supervisor, Mrs. Jaana Korhonen from University of Helsinki. She offered guidance throughout the process of this research and pointed me to the right direction when I was lost. I'm truly grateful for her assistance.

I would also like to thank Mr. Alex McCusker and Mr. Florian Steierer from United Nations Economic Commission for Europe for providing the opportunity to conduct this research. Mr. McCusker has also been providing essential help with data used in this research and has made much needed suggestions during this process.

Last but not least, I want to thank United Nations Economic Commission for Europe (UNECE) and the Food and Agriculture Organization (FAO) for funding and believing in this study. It would have not been possible without the valuable support of the UNECE and FAO.

Helsinki, Finland
May 13th 2019

Markus Stolze

1. INTRODUCTION

We live in a world where data are gathered everywhere and all of the time. Why are we doing this and for what? In a modern world, data is defined to be "information, especially facts or numbers, collected to be examined and considered and used to help decision-making" (Cambridge Dictionary 2019). More and more different kinds of data are available and these data are used to make different decisions. We understand that bad data can lead to bad decisions and good data is necessary for making informed decisions. However, what are good data and what make some data bad or unreliable? In order to have a better understanding of what are good data, it needs to be analysed.

Before the difference between good and bad data can be discussed, there has to be an understanding of how data quality can be measured. In order to measure data quality, the dimensions that affect data quality need to be identified. Only after that, can measuring of data quality start. Data quality is a well-researched area and it is possible to determine the dimensions of good data. It is also essential to find the right ways to measure data quality. Using information provided by previous studies, this study outlines the key characteristics of good data, and how the quality of data can be assessed. This is followed by an empirical examination of forecast data provided by member States of the United Nations Economic Commission for Europe (UNECE). This includes assessing differences in variation and reliability of different kind of forecasts for different products.

The forecast information in question focuses on forest products in UNECE region. The main goal of these predictions (the term predictions is used to cover both current year estimates and next year forecasts) is to give information about trends in the forest sector before actual data is available. Predictions are produced by official correspondents from each member state. These predictions have been collected since the 1960s and have been available in a database since 2002, which generates the question: how accurate are these predictions? This research is aiming to answer that question and gain better understanding of what makes some data better than others and how does this affect predictions for future forest products markets.

1.1 Objective of thesis

This research focuses on forecast information provided by member States of the UNECE. The objective of this research is to have an understanding about the reliability and quality of forecast information provided by UNECE member States. This research focuses on forest products that are the most produced and traded. Data on four product flows are used in this research: imports, exports, production and removals. The aim is to see how reliable forecast data provided by member States are, and if there are commonalities between different countries. It is assumed, that member States are producing their predictions differently and therefore to investigate, if there are some methods that are better than others for producing predictions. A comparison was also made to see if simply repeating last year's data would provide more accurate prediction than an actual forecast information.

More specifically, this study aims to answer following questions:

What are the main dimensions of data quality, when producing forest product predictions?

Are forecast data about forest product markets presented by UNECE reliable? And if so, how reliable?

1.2 What is the United Nations Economic Commission for Europe?

The United Nations Economic Commission for Europe (UNECE) was founded 1947 and its "major aim is to promote pan-European economic integration". 56 States are members of the UNECE. The member States include the countries of Europe, as well as North America (Canada and United States), Central Asia (Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and Uzbekistan) and Israel. All countries are located in the northern hemisphere. This region covers over 47 million square kilometres, 17% of world population and more than 40% of world's forests. This means, that it is a major source of wood and forest products, its members States account for about 60% of industrial roundwood produced globally. Thus, understanding the volumes harvested converted to forest products and traded is important. (UNECE 2019a)

Before this study goes deeper into predicting future forest products production, consumption and trade, the study will outline which dimensions are important for data quality, when making these predictions and how to differentiate between reliable and unreliable data.

2. THEORETICAL PERSPECTIVE

What is good data? With the large quantities of data used in this research, it is valid to determine what make some data better than others. Modern data quality is usually defined by how usable the data is, also known as “fitness for use” (Chen et al. 2013). Juran (1989) has claimed that data are high quality if they fit the intended use of purpose. With Juran’s (1989) theory in mind, one set of data could be considered as high quality for some, but not for others. Wang and Strong’s (1996) study of data quality is frequently cited in research related to data quality (Scopus 2018)). Their research has four different categories in data quality: intrinsic data quality, contextual data quality, representational data quality and accessibility. Under each category, there are also 15 dimensions that further refine data quality. However, the categories are meant to include what is in the dimensions, since that way they are more usable in real life applications.

Wang and Strong’s (1996) research was one of the first that focused on data quality and still remains a basis for the definition of data quality. The baseline for their research was to “develop a framework that captures the aspects of data quality that are important for data consumers” (Wang and Strong 1996). Data consumers are those who are using the data and, in most cases, the same people who store the data. Wang and Strong (1996) used a two-stage survey, where they came up with the most important factors related to data quality. Their four categories and 15 dimensions are displayed in figure 1 below.

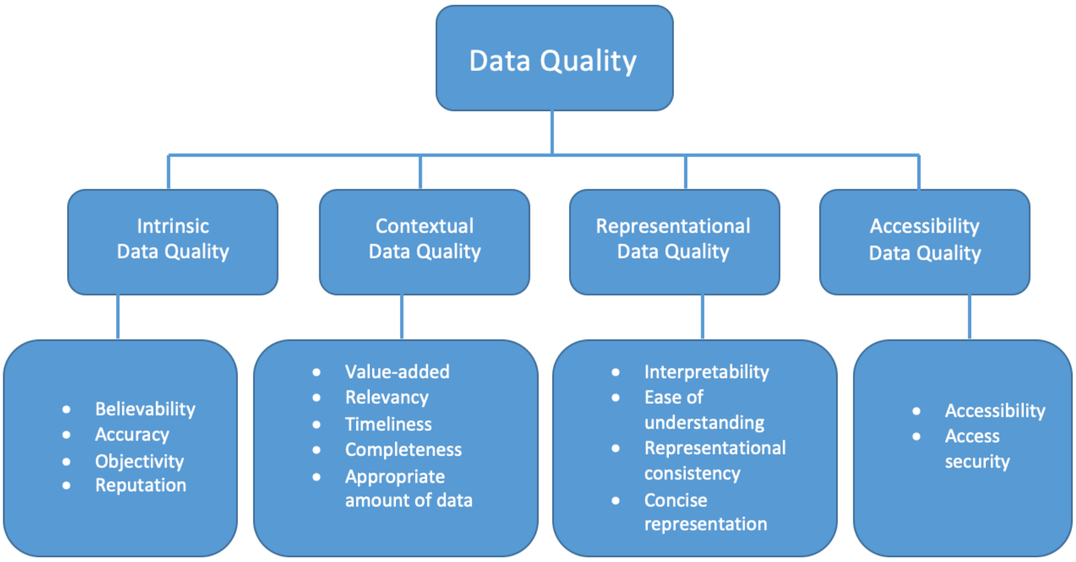

Figure 1. A Conceptual framework of data quality (Wang and Strong 1996).

Each of the four categories is supposed to combine dimensions that are linked to them. In this way it is easier to use this framework, when there is no need to consider each dimension alone, but the whole category instead. Therefore, intrinsic data quality is combining believability, accuracy, objectivity and reputation. Contextual data quality is combining value-added, relevancy, timeliness, completeness and appropriate amount of data. Representational data quality is a merge of interpretability, ease of understanding, representational consistency and concise representation. Finally, accessibility data quality is a mix of accessibility and access security. Wang and Strong (1996) also rated all the dimensions from the most important to the least, with the most important dimensions for data quality being believability, value-added, relevancy, accuracy and interpretability. It should be pointed out, that the first two categories focus on quality of data itself and latter two on consumer usability.

While Wang and Strong's (1996) study of data quality is highly cited, there are other studies that have their own data quality frameworks. For example, Bovee et al. (2003), Liu and Chi (2002) and Huang et al. (2012). These studies are often used as a data quality framework. Within these four different studies, there are over 35 dimensions between them, that attempt to define data quality. However, accuracy, completeness, interpretability, relevancy and timeliness are common to all four of these studies. This research, tries to take dimensions that are used in most frameworks of previous studies and see, if there are commonly used dimensions that would make a usable combination when evaluating forest product predictions.

Bovee et al. (2003) researched producing a framework for assessing overall information quality. The research suggests that availability of information is no longer a strategic advantage, but quality of information is. They define quality of information similarly to the way Wang and Strong (1996) did with fitness for use. Bovee et al. (2003) framework is based on following requirements: i) accessibility of information; ii) interpretability of information; iii) relevancy of information and; iv) integrity of information. Without meeting all of these requirements, the information in question would be considered as being deficient.
Accessibility is an essential criterian, since if data can't be accessed by users, when or where they need it, other dimensions related to quality are irrelevant. Information must be intelligible and meaningful for its user. And yet again, if either of these requirements are not met, all other qualities are irrelevant. However, whether information is unintelligible or meaningless is highly related to particular users. For example, simply not understanding a foreign language might make it impossible to read for some, but not for others. The third requirement, relevance, requires information to "be relevant to our domain and purpose of interest in a given context" (Bovee et al. 2003). Relevancy also includes whether or not the information is current enough. In many cases updating information often enough is limited due to the cost of doing so. The last criteria, integrity, implies freedom from flaws, mistakes or any other problems related to having wrong information available. Information is also expected to be accurate, so it is usable for its users. Information is also expected to be consistent and complete.

Liu and Chi (2002) have created their own framework for data quality and at first they notice that data quality measurement model might change as the use of data changes. Instead of an empirical view of data quality, such as Wang and Strong (1996), Liu and Chi (2002) take a theoretical view of the situation. The justification for a theoretical view is based on the limitation of researcher's own experience, which might have a negative effect on empirical and intuitive approach. Both empirical and intuitive approaches lack theoretical structure on how a certain attribute is received and represented. The theoretical approach also has the upside of justifications;

based on why and how attributes are grouped into certain forms. Liu and Chi's (2002) concept has three components: there is different definitions for measurement of quality, quality of data in earlier stages positively impact later data and there is increasing order for different views of data quality. Their model is presented as pyramid with application quality in the top, presentation quality below, organization quality third and collection quality at the bottom (Figure 2 below). The concept is that, "one measure of data quality at a lower level is useful to measure the quality of many sets of data at a higher level" (Liu and Chi 2002). The higher the level, the more theories it has to satisfy, so that theory can determine the data quality. Each four levels include various other attributes, which are common with other frameworks. Collection quality consists of: accuracy, objectivity, completeness, integrity of the collector, clarity and other collection theory-specific attributes. Organization quality includes collection quality as lower level and other attributes such as reliability of data clerk, consistency, storage efficiency, retrieval efficiency, navigability and organization theory –specific qualities. Next level is presentation quality, which includes two lower levels, and the following attributes: faithfulness, neutrality, interpretability, formality, semantic stability and presentation theory-specific qualities. Application quality, yet again with all the lower levels included, has following attributes: ease of manipulation, timeliness, privacy, security, relevancy, appropriate amount of information and application theory-specific attributes. According to Liu and Chi (2002), application quality is the level with the most frequent problems, which then makes using the data in question hard.

Figure 2. Evolutional data quality (Liu and Chi 2002).

One of the newer research on data quality is from Huang et al. (2012), who made a research focusing on data quality with genome annotations (Huang et al. 2012). They were able to have professionals on that specific field to answer questions about data quality and what they think are the most important dimensions affecting data quality. As they are asking opinions of others, their research is defined as an empirical study. Huang et al. (2012) created five data-quality constructs, with 2-5 dimensions each. Five constructs in their research were in order from the most important to the least important: accuracy, accessibility, usefulness, relevancy and security. Rating of the constructs were rated in the context of their field of profession. It is pointed out in the research (Huang et al. 2012), that due open nature of sharing in medical community, security is not associated with accessibility and could therefore be ranked low. Majority of the dimensions under each constructs are similar than in Wang and Strong's (1996) research, although some are named differently.

Four data quality studies, as presented above, have a lot in common. There are usually four categories and those categories include four dimensions on average. Majority of the dimensions are same or at least very similar. There are, however, different approaches for these studies. Three approaches that are often used in data quality: an intuitive, a theoretical and an empirical approach. Bovee et al's (2003) research has an intuitive approach, Liu and Chi's (2002) has a theoretical approach and both

Huang et al. (2012), as well as, Wang and Strong (1996) can be categorized to the empirical approach. The initiative approach is selected when researcher has a feeling, usually based on past experiences, which attributes are important for the study. This is the most used approach in data quality (Wang and Strong 1996). The theoretical approach is selected when the focus is on the process of how data are manufactured. Using theoretical approach provides a set of data that is useful for the research in question. Empirical approach is often selected when data is based on evidence and real-world events, rather than facts or feelings. Wang and Strong (1996) point out that only empirical approach manages to capture the voice of data consumers. Liu and Chi (2002) claim that intuitive and empirical approaches usually create confusing definitions for basic data quality attributes and they claim that theoretical approach reduce these problems.

Theories presented earlier have also been implemented in practice. For example, Kovac et al. (1997), have introduced a framework for data quality. The model provides consistent measurements for data quality and improvement for data handling process. They wanted to develop data quality in practical environment to have better data quality. It is based on Wang and Strong's earlier framework, where Timeliness + Reliability + Accuracy = Quality (TRAQ). The TRAQ model has two vital objectives. First, it must provide consistent measurement of data quality and delivery reliability, which should display delivery process and the external client view of delivery process. Having all this, it grants management a possibility to appraise performance of specific clients. The second objective is to have improvements for the delivery process repeatedly. This system was designed to produce repeated improvements for the delivery process. In the end, Kovac et al. (1997) claimed that TRAQ provided massively benefits for the business in question. TRAQ model has inspired many other models for defining data quality framework, such as RUMBA, which stands for reasonable, understandable, measurable, believable and achievable.

Data quality in forest sector have also been studied before. Kallio et al. (2018) have made a research, which focuses on issues most seen in data related to forest sector. They use data from FAOSTAT, which is closely related to UNECE and their database. Therefore, their research is good starting point when thinking about problems related to data quality in forest industry. According to them, it is not surprising, that there are inconsistencies in the data. It is still important to collect data, so modelling the production process and material streams in the forest sector would become more accurate. This would allow to have improved estimates for wood use coefficients (Kallio et al. 2018). They also point out that collecting reliable data even for the main mechanical products is a challenge, let alone the by-products with smaller quantities. In some cases, measurements might not be collected at all, which produces even more problems. There are also big challenges with measurement errors, such as converting solid cubic meters to loose cubic meters (Kallio et al. 2018). Even with all the problems with data quality in the forest sector, there is still a lot of potential. These statistics provide important data for wide range of users from business, policy makers and scientific analytics. Kallio et al. (2018) notice that some regions are better than others with producing data quality and that often poor data quality is related to problems with illegal logging and corruption. In conclusion, it is important to be cautious when using forest product data (or any data for that matter).

There are clearly some dimensions that are used in the majority of the data quality studies. For example, accuracy, completeness, consistency, interpretability and relevancy are featured in all four studies (Wang and Strong 1996, Liu and Chi 2002, Bovee et al. 2003, Huang et al. 2012). All together 36 different dimensions were introduced in the four researches presented above, where only 14 were in more than one research. Next, this study will take a closer look at the dimensions of data quality, used by previously presented studies and which dimensions are the most important for this research with forecast data.

2.1 Data quality dimensions selected for this research

On the table 1 below are all the dimensions, that were featured in at least two of the researches out of four introduced above. To be accurate, some dimensions were similar, but not quite close enough to be considered identical.

Table 1. The most popular dimensions in researches introduced above. Modified from (Rantala 2016).

Dimension	Wang & Strong (1996)	Bovee et. al. (2003)	Liu & Chi (2002)	Huang et al. (2012)	Featured in researches
Accuracy	x	x	x	x	4/4
Complete-ness	x	x	x	x	4/4
Consistency	x	x	x	x	4/4
Interpretabil-ity	x	x	x	x	4/4
Relevancy	x	x	x	x	4/4
Accessibility	x	x		x	3/4
Appropriate amount of data	x		x	x	3/4
Timeliness	x	x	x		3/4
Believability	x			x	2/4
Ease of ma-nipulation			x	x	2/4
Objectivity	x		x		2/4
Reputation	x			x	2/4
Security			x	x	2/4
Value-added	x			x	2/4

This research will focus on 9 out of 14 dimensions presented above. This selection was done in order to have relatively small number of dimensions for this research, which allows selected dimensions to have meaningful impact. Selecting all 14 above was too many dimensions for this research, since with fewer dimensions there can be a better focus on the predictions. It seems logical to select the five dimensions, that are featured in all the researches: accuracy, completeness, consistency, interpretability and relevancy. All five dimensions are crucial for the forecast data analysed in this research: it has to be accurate, completed, consistent, interpretable and relevant. In addition, the appropriate amount of data, accessibility, timeliness and believability are selected as important dimensions for this research. As stated previously, Wang and Strong's (1996) research is used as a baseline for this study, so it is good to note that all four categories that are featured in selected dimensions.

Intrinsic data quality is the first category from Wang and Strong's (1996) research, where believability and accuracy are selected. Accuracy is fairly self-evident, since this study is analysing forecasted data. Accuracy is the most important aspect of predictions, as it is the main goal. Believability is a crucial dimension when forecast data is checked: can a country have 100% increase on a production? Is this value believable or is there a simple mistake with a decimal place when producing forecasts? Wang and Strong's (1996) category itself also holds objectivity and reputation, but these are not featured in this research. Reasoning behind this is that predictions produced by UNECE's member States are not objective, since they are produced by representatives of said member state. Also, reputation is dismissed since all predictions are handled with similar expectations. In other words: all forecasts are equal.

Contextual data quality is the second category of Wang and Strong's (1996) research. It consists of value-added, relevancy, timeliness, completeness and appropriate amount of data. This is important category, since four of nine dimensions chosen are from this category. The value-added dimension is the only featured from this category that is not mentioned. Value-added is defined as giving you a competitive edge and adding value to your operations. While this is extremely important, it doesn't add anything else that other dimensions don't already do when thinking about forecast data. The dimensions featured from this category are significant for this research: relevancy, timeliness, completeness and appropriate amount of data. Relevancy is defined in Wang and Strong's (1996) research as applicable, relevant, interesting and usable. Those all are things a good prediction should aim for and therefore it is selected as a dimension for this research. Timeliness is a crucial dimension for prediction, since there is a clear window of time when predictions are usable. Producing them too early makes them very inaccurate and produced too late makes them useless, if actual values are already available. Completeness and appropriate amount of data are similar dimensions, but both have their uses. With completeness, the data has enough depth and scope of information contained in the data that is big enough. Appropriate amount of data is useful in this research so that clear trends can be seen: if a country produces predictions only every third year, trends aren't visible since the analysis only assesses those countries with data available at an annual basis.

The third category in Wang and Strong's (1996) research is representational data quality, which includes following dimensions: interpretability, ease of understanding, representational consistency and concise representation. From this category only interpretability is selected, as it is featured in all the researches introduced. It is vital since it makes sure that the data in question can be explained: if data can't be explained, there is no use for it. The other three dimensions: ease of understanding, representational consistency and concise representation, are not featured in any other researches introduced. While they are useful, they don't add too much after interpretability. All three are already, to some extent, included in interpretability.

The fourth category in Wang and Strong's (1996) research is accessibility data quality. There are only two dimensions: accessibility and access security. Accessibility is included in this research and it can be defined as having good accessible and up-to-date data. Accessibility is a dimension, that only becomes important when there is a problem with it. As long as everything works as expected, access to the data is not a prioritized. However, without it, there is no way of using the data. While access security is certainly an important aspect, it does not play a major role in this research. All the data used in the analysis is publicly available for everybody and therefore security is not a concern. In table 2 below are all nine dimensions, which are used in this research to ensure good data quality and their definitions by Wang and Strong (1996).

Table 2. Data quality dimensions in this research and their definitions.

Dimension	Definition by Wang and Strong (1996)
Accessibility	Accessible, retrievable, speed of access and up-to-date
Accuracy	Data are certified error-free, accurate, correct, flawless, reliable and errors can be easily identified
Appropriate amount of data	The amount of data
Believability	Believable
Completeness	Breadth, depth and scope of information contained in the data
Consistency	Continuously presented in the same format, consistently represented and formatted
Interpretability	Interpretable
Relevancy	Applicable, relevant, interesting and usable
Timeliness	Age of data

2.2 Effect of data quality in forest product predictions

How do all these dimensions affect predictions analysed in this research? There are some dimensions that have more importance for users only, such as timeliness and other dimensions that affect the quality of data itself, such as accuracy. While each has different use for this research, they are all important. This chapter goes through nine dimensions specified in table 2 above and specifies how the quality of data can be defined using them. Instead of going over general data quality, this study is focusing on predictions used later in our analysis and seeing what the specific qualities of this data are.

The first dimension that is taken a closer look at is accuracy. It is arguable one of the most important aspect of data quality in this case, since it is considered as the objective of prediction. Representatives of member States are trying to make them as accurate as possible. In the definition of Wang and Strong (1996) accuracy has also "errors can be easily identified". When producing data, a small mistake could have a massive effect on data, but if mistakes are easily identified, it makes it a lot easier to fix said mistake. This is also useful for the users of predictions: even if a mistake slips by the producer, it can be still identified as mistake for users. When a number doesn't make sense, it is usually a mistake. This brings us to the second dimension: believability. There is much same as in accuracy, as predictions are expected to be believable. If a country has a production increase of 200% for a single product, it is not believable. There would have to be prior information about plans of new production or larger scale of harvest, that any production could grow in such a rate. There is an exception to this with these specific predictions: products with very small production or trade volumes can have a 200% growth in percentage terms since already small changes in the absolute figures cause huge changes in percentage terms. These cases are problematic when measuring reliability of predictions in percentage: the difference in most cases is not meaningful, with error of 1.9, but makes certain prediction seem unreliable. This problem has been taken into consideration later, when comparing the predictions also with absolute numbers in addition to percentage values.

Next two dimensions, appropriate amount of data and completeness, are closely related to each other. Appropriate amount of data is the first dimension to be used in this research. The threshold was set as two out to three possible data were provided (66.6%). Member State meeting or exceeding the threshold were included in the assessment of the study since a smaller number would not be sufficient for this analysis. Countries reporting a product with 0 quantity were included when counting the amount of data. With data, it's not only, that there are enough data. Data have to be completed and well thought out, which brings out the next dimension: completeness. As defined earlier, completeness includes breadth, depth and scope of information contained in the data (Wang and Strong 1996). It is possible to fill out form for prediction and not think about if there is all the potential knowledge. To help with this task, UNECE prefills the questionnaires with data from previous years. This way correspondents are left with easier task to completing the task. Completeness comes down to making the data have all the information possible, which is crucial when aiming for the best possible reliability of predictions.

Consistency of data is extremely important for this research, since this research is analysing 15 years of predictions. If a prediction is made in one way earlier and completely different next year, it most likely will affect the results. There is also another aspect for consistency, as there are predictions from nearly 30 different countries: they have to represent predictions consistently, so they can be compared with other countries. This also affects people from UNECE, since they have to make all forms understandable, so all different member States will understand how to fill those. Predictions are also made for two years at time, so both years need to be consistent with each other.

Relevancy is a dimension that is fairly close to consistency, as well as completeness. As relevancy is defined as "applicable, relevant, interesting and usable", it becomes even more important (Wang and Strong 1996). Information in relevant data has to be usable, so no unwanted or unneeded information should be part of forecasts. Relevant information might also be something that is only rumoured to happen, as this study is analysing forecasts that are made for a next year as well. If there is a plan, that is not yet confirmed, but possible, it could be relevant for a prediction.

Accessibility, interpretability and timeliness are little different the other six dimensions as outlined in table 2. They have very little to do with the quality of data itself and more with how users can benefit from the data. Accessibility is essential for users of these predictions, since if nobody can access them, what is the point of producing them? Accessibility also includes speed of access and data to be retrievable, which should not be a problem in a modern world with fast internet widely available. Accessibility is also linked with interpretability, since predictions have to be in a formatted in a way, that users can access them. This has been solved by having all of the predictions in Microsoft Excel and available in UNECE's website. Interpretability includes representing forecasts in language, that is widely known – English. All products are coded similarly in all UNECE's forms, which also helps users, as these codes are easily checked. Timeliness, or age of data, is logical dimension to include in this research. Predictions are made before actual values are available, to represent what most likely will happen. There is on average window of 9 months or 21 months, depending on which prediction is used, when they are usable. After actual values are out, predictions have no value for anybody. Therefore, it is also important that predictions are produced when they are valuable for users and also being available for use.

Now that there is a good understanding on how data quality is constructed, there will be a closer look on what predictions are included in this research and how they are going to be analysed. In later parts on this research data quality will be analysed and determinates how predictions have managed to fill the requirements and expectations set to them.

3. METHODS AND BACKGROUND OF THE RESEARCH

3.1 Data used in the research

In this research there are four main sets of data: estimates, forecasts, repeated data and historical data. Two of these are predictions: estimates and forecasts. Estimate is a set of data, that is made during a year for that specific year. Usually estimates are made in September and at that time there are usually preliminary actual data for first six months of that year. Forecasts are made at the same time as estimates, but instead for the following year. For example, during September 2018 forecasts were made for year 2019, without the full knowledge how 2018 even turned out in the end. Historical data are usually gathered around six months after year has ended and it provides the "real" numbers, which are used when comparing accuracy of predictions. Historical data are revised, if new information is provided later. Historical data can be changed even a number of years after reference year has passed.

The last set of data is repeated data, which is created using historical data from previous year. This is produced for this study and not by member States. However, it is treated as prediction for purposes of this study. It is a set of data that is created only from previous year, for example: historical data from 2017 is used to create repeated data for 2018. There is only one set of historical data being created and it is from previous year. Historical data, used to create repeated data, have been taken from database in September 2018. This means, that historical data used in this analysis might not be the same as it would have been, when predictions of previous years were produced.

All this in mind, estimates are expected to be more accurate and reliable than forecasts. With preliminary data already presented for the first six months, it is easy to understand why this is expected to be true. Doing forecasts over a year ahead makes it impossible to react to new trends. Beating repeated data is a clear benchmark for estimates and it is interesting to see if are more reliable than repeating previous' years data.

3.2 Member states included in research

Not all UNECE member States have provided forecast information on forest products. In order to make sure that enough information was provided, member States with more than 66.6% of possible forecast and estimate data have been included in this research. Out of 1116 possible data points, in order to reach 66.6% mark, 746 or more data points are required. In the table below, are member States, which have provided enough data and therefore will be included in this research. Total number of member States included in this research is 27. Out of these countries, two (Canada and the United States) are in North America, The Russian Federation spans from Europe to Asia and the rest are from Europe. Member States included in the research and response rates of estimates and forecasts are presented in the table 3 below.

Table 3. Countries included in the research.

Country	*Response rate*	*Country*	*Response Rate*
Poland	*100.0%*	*Germany*	*94.6%*
Estonia	*99.8%*	*Austria*	*94.5%*
Sweden	*99.6%*	*Latvia*	*94.0%*
Switzerland	*99.5%*	*Lithuania*	*93.5%*
Cyprus	*98.6%*	*Spain*	*86.7%*
Netherlands	*98.2%*	*France*	*85.7%*
Slovakia	*97.8%*	*Romania*	*82.6%*
United Kingdom	*97.8%*	*Serbia**	*82.4%*
Turkey	*97.5%*	*Norway*	*82.0%*
Russian Federation	*96.9%*	*Slovenia*	*79.6%*
Czech Republic	*96.7%*	*Ireland*	*79.5%*
Croatia	*96.4%*	*Italy*	*67.7%*
United States	*95.5%*	*Canada*	*67.5%*
Finland	*94.6%*		

Note: Serbia includes Serbia and Montenegro's data prior to 2005.

In total 28 UNECE member States have provided some, but not enough to meet 66.6% or none of the data and therefore can't be included in this research, are following: Albania, Andorra, Armenia, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Denmark, The F.Y.R of Macedonia, Georgia, Greece, Hungary, Iceland, Israel, Lichtenstein, Luxembourg, Kazakhstan, Kyrgyzstan, Malta, Monaco, Montenegro, Portugal, Republic of Moldova, San Marino, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

3.3 Products included in the research

This analysis includes 12 products, four of which are removals of logs from forests and remaining 8 production of forest products. Four products that are removals: coniferous saw logs and veneer logs, non-coniferous saw logs and veneer logs, coniferous pulpwood and non-coniferous pulpwood. From these four products, only volume of harvested logs is measured. 8 forest products selected are following: coniferous sawn wood, non-coniferous sawn wood, plywood, particle board (including OSB), OSB, fibreboard, wood pulp, paper and paperboard. From these products three flows are measured: production, export and import. In addition, also average volume of each product is presented in the table. This is calculated from all 27 of the member States included in the research and it provides an understanding of which products are bigger than others in volume. A more detailed description of products is listed below in table 4.

Table 4. Products included in this research and average volumes of each product.

Product	***JFSQ-code***	***HS2012-code***	***Average volume (x 1,000)***
Coniferous sawlogs and veneer logs	*1.2.1.C*		*19,867 m3*
Non-Coniferous sawlogs and veneer logs	*1.2.1.NC*		4,011 m3
Coniferous pulpwood	*1.2.2.C*		8,954 m3
Non-Coniferous Pulpwood	*1.2.2.NC*		4,731 m3
Coniferous sawn wood	*5.C*	*4407.10*	4,722 m3
Non-Coniferous sawn wood	*5.NC*	*4407.21/22/25/26/27/28 /29/91/92/93/94/95/99*	690 m3
Plywood	*6.2*	4412.31/32/39/94/99	470 m3
Particle board (including OSB)	*6.3*	44.10	1,368 m3
OSB	*6.3.1*	*4410.12*	498 m3
Fibreboard	*6.4*	*44.11*	659 m3
Wood Pulp	*7*	*47.01/02/03/04/05*	2,125 mt
Paper and Paperboard	*10*	*48.01/02/03/04/05/06/08/09/10, 4811.51/59 48.12/13*	4,235 mt

3.4 Structure of analysis

The aim of this analysis is to see how accurate the estimates, forecasts and repeated data are compared to actual non-repeated data. This comparison was achieved by calculating data with following formula:

$$\frac{(x-y)}{y} * 100 = a$$

In this formula, x is forecasted data, y is value from historical data and a is the result. Structuring the formula this way made it possible to have a clear understanding of how predictions compared to an actual value. As comparisons were done in percentages, results were multiplied by 100% of the value so they would be either negative or positive. A negative value would show that prediction is smaller than actual value and therefore underestimated. In contrast, positive value shows that prediction is bigger and therefore overestimated. After having each value calculated for estimates, forecasts and repeated data, averages were counted for 17 different categories. The table 5 below shows all the categories and what product flows they include.

Table 5. Products and product flows included in the research

Products	*JFSQ-Code*	*Flows*
All products		*All flows*
*All products**	*Ex. 1.2.1.C/NC, 1.2.2.C/NC*	*Only exports*
*All products**	*Ex. 1.2.1.C/NC, 1.2.2.C/NC*	*Only imports*
*All products**	*Ex. 1.2.1.C/NC, 1.2.2.C/NC*	*Only production*
All logs	*1.2.1.C/NC, 1.2.2.C/NC*	*Only harvests*
Coniferous sawlogs and veneer logs	1.2.1.C	*Only harvest*
Non-Coniferous sawlogs and veneer logs	1.2.1.NC	*Only harvest*
Coniferous pulpwood	1.2.2.C	*Only harvest*
Non-Coniferous Pulpwood	1.2.2.NC	*Only harvest*
Coniferous sawn wood	5.C	*Exports, Imports and production*
Non-Coniferous sawn wood	5.NC	*Exports, Imports and production*
Plywood	6.2	*Exports, Imports and production*
Particle board (including OSB)	6.3	*Exports, Imports and production*
OSB	6.3.1	*Exports, Imports and production*
Fibreboard	6.4	*Exports, Imports and production*
Wood Pulp	7	*Exports, Imports and production*
Paper and Paperboard	10	*Exports, Imports and production*

In addition to percentage of actual value, also absolute differences have been counted for all the categories. This gives a good perspective, since some percentage differences were massive, but absolute values were minimal. It is good to understand, that not every country has provided data from every product, product flow and year.

However, majority of countries have been regular with missing data: if it is missing in 2002 it usually is still missing in 2017.

3.5 Sorting of countries

As this research aims to know, if the size of a country's forest industry has any effect of how well they produce forecasts, 27 countries in this research were divided into three groups: big, medium and small countries. The criteria for sorting were simple: average unit volume of all products and all product flows between 2002 and 2017. Since some products are in cubic metres and others in metric tonnes (see table 4) no units can be used in this process. At first, average number of each product flow was calculated and then grand average of all products and product flows. The line for categories were following: less than 500 for small, between 500 and 2,500 for medium and over 2,500 for big countries. In the end, each category is fairly well balanced: there are 8 big countries, 11 medium countries and 8 small countries.

The 8 big countries are: the United States of America, Canada, the Russian Federation, Germany, Sweden, Finland, France and Poland, as listed in table 6 below. They all have big numbers in removals and production, but not necessary in exports or imports. This group is the only one with bigger exports, than imports, which suggest that they are exporting a lot of produced goods to other countries. The big countries contain the major players of forest product markets in UNECE region. The United States is the biggest country in UNECE region, as they are the biggest importer, producer and harvester. Only export volumes are not the biggest in the region. Canada is the second biggest country by average volume, due to its massive number of removals, production and exports. Canada is the biggest exporter in the UNECE region, as they export a big part of the harvested roundwood to the United States. The Russian Federation has nearly identical numbers of removals as Canada, but clearly smaller numbers in all other product flows. Germany, France and Poland have similar structure in product flows: harvests are the biggest but all the remaining product flows are similar to each other. Sweden and Finland have a similar scale between different product flows: big exports and production, small imports and massive removals.

Table 6. Countries with average volume of all products more than 2,500.

	Average exports volume	*Average imports volume*	*Average production volume*	*Average removals volume*	*Average volume of all products*
United States	*3 104.65*	*8 199.12*	*32 225.40*	*90 002.79*	***33 382.99***
Canada	*8 369.82*	*919.08*	*13 100.00*	*39 582.23*	***15 492.78***
Russian Federation	*3 204.55*	*370.14*	*6 790.56*	*39 559.75*	***12 481.25***
Germany	*3 537.17*	*3 220.33*	*7 655.77*	*10 921.13*	***6 333.60***
Sweden	*3 213.24*	*352.05*	*5 155.51*	*16 346.60*	***6 266.85***
Finland	*2 773.50*	*202.63*	*4 512.21*	*11 962.18*	***4 862.63***
France	*1 210.41*	*1 644.09*	*3 247.36*	*6 734.16*	***3 209.00***
Poland	*627.63*	*767.35*	*2 154.08*	*7 548.82*	***2 774.47***
Average	***3 255.12***	***1 959.35***	***9 355.11***	***27 832.21***	***10 600.45***

The 11 medium countries are: Austria, Spain, Turkey, the United Kingdom, Czech Republic, Italy, Romania, Latvia, Norway, Slovakia and the Netherlands. The medium category have an average production volume between 500 and 2500, presented in more detail in table 7. Many medium countries have a one or two big product flows, usually one of them being removals, exports or imports, but do not have as big of a market share as bigger countries do or not with as many product flows as above. For example, the United Kingdom is a major importer of European forest products but is exporting and producing a lot less than France or Germany. A big part of the medium group has a big volume of removals, but the rest of the product flows are minimal, such as Czech Republic, Portugal and Latvia.

Table 7. Countries with average product volume between 500 and 2500.

	Average exports volume	*Average imports volume*	*Average production volume*	*Average removals volume*	*Average volume of all products*
Austria	*1 675.50*	*582.12*	*2 519.42*	*3 341.80*	***2 029.71***
Spain	*681.49*	*1 000.89*	*1 848.27*	*3 123.29*	***1 663.49***
Turkey	*147.02*	*593.37*	*1 926.01*	*3 542.07*	***1 552.12***
United Kingdom	*245.59*	*2 321.95*	*1 495.99*	*1 979.43*	***1 510.74***
Czech Republic	*548.50*	*348.97*	*960.98*	*3 578.04*	***1 359.13***
Italy	*562.46*	*2 126.84*	*1 943.60*	*474.20*	***1 276.77***
Romania	*564.04*	*158.43*	*1 041.89*	*2 344.41*	***1 027.19***
Latvia	*449.99*	*93.86*	*589.98*	*2 578.32*	***928.04***
Norway	*340.00*	*233.61*	*789.13*	*2 092.18*	***863.73***
Slovakia	*259.49*	*156.41*	*531.66*	*1 932.06*	***719.91***
Netherlands	*523.87*	*1 136.61*	*421.80*	*186.96*	***567.31***
Average	***573.57***	***773.80***	***1 206.01***	***2 205.81***	***1 189.80***

The 8 small countries are: Estonia, Switzerland, Lithuania, Croatia, Ireland, Slovenia, Serbia and Cyprus. The average volumes of these countries are presented in table 8. The order of average product flows is same as in the medium category, with removals being the biggest, followed by production, imports and exports. Many of these are smaller than average European countries, both in population and land-area covered, and therefore it's not a surprise that they also produce less than the medium category. Estonia and Lithuania have similar structure as they have a bigger product flow of removals than in the other three product flows. Switzerland, Slovenia and Ireland are all pretty evenly matched as structure of product flows go but volumes are simply not as big as in the medium category.

Table 8. Countries with average product volume less than 500.

	Average exports volume	*Average imports volume*	*Average production volume*	*Average removals volume*	*Average volume of all products*
Estonia	*170.46*	*135.90*	*283.17*	*1 347.96*	***484.37***
Switzerland	*256.14*	*291.69*	*498.59*	*872.43*	***479.71***
Lithuania	*132.57*	*150.89*	*252.64*	*1 179.16*	***428.82***
Croatia	*125.27*	*107.22*	*195.70*	*808.52*	***309.18***
Ireland	*180.84*	*152.55*	*256.00*	*608.63*	***299.51***
Slovenia	*204.65*	*185.98*	*224.71*	*579.65*	***298.75***
Serbia	*43.28*	*129.22*	*131.03*	*320.93*	***156.11***
Cyprus	*0.06*	*35.18*	*0.69*	*1.47*	***9.35***
Average	***139.16***	***148.58***	***230.32***	***714.84***	***308.22***

3.6 Sorting countries based on how the figures look like

Countries were sorted into three categories based on the average product volume, which allowed these countries to be compared with other countries. In this comparison three distinct patterns were noticed, which multiple countries shared. These patterns were found in all three categories and most of the countries were divided into new groups. As a reminder, in this study categories are used when referring to size of country: small, medium or big. Groups are used to describe sorting of countries based on figures of results. These terms should not be mixed. The logic behind this is to figure out if these countries share similar tools or ways of producing predictions.

This comparison was done visually with graphs from five main product flows: all products, exports, imports, production and removals. Visual analysis provides a unique possibility to observe clear trends that might go unnoticed with other approaches. Visual analysis shouldn't be used as a substitute for statistical analysis, but rather as an additional way of doing observation (Garcia and Mendonca 2004). This is exactly what have been done in this research, with statistical analysis as a basis for visual analysis. Visual representation offers agility and adaptability to data analysis (Garcia and Mendonca 2004), which allows to make efficient analysis in a shorter time span. This is important as there are thousands of data points. However, there will be a statistical analysis with individual products, when trying to identify which products are more accurate than others.

For this, series with absolute values were used with average difference from each year. This was done due percentage difference was effected in many cases with small quantity changes, which had major effect on average percentages. Using absolute values, change of 2 to 0.2 wouldn't make a noticeable difference, whereas in percentage difference it would show as 100% drop. In order for countries to be considered into a group three or more out of potential, five main product flows had to match the definitions of group in question. Four main groups, the definitions for groups and which countries are divided in those groups are presented in table 9.

Table 9. Sorting countries into groups based on results of production reliability.

	Group 1:	**Group 2:**	**Group 3:**	**Group 4:**
Characteristics	All three series are similar. Estimates and repeated data show only minor differences.	Estimate is clearly the best series and overall really close to the actual value and overall the best out of all data.	Series don't fit in either previous groups. There are clear patterns visible with usually a spike during a financial crisis and a drop year or two after. None of the series are clearly better than others.	A mix of two or three previous groups. One or two product flows might suggest a group, but other product flows are not clearly in that group. On average hard to say which prediction is the most accurate.
Countries	Croatia, Cyprus, Czech Republic, Latvia, the Netherlands, Poland, Turkey	Austria, Finland, Lithuania, Norway, Sweden, the United Kingdom	Canada, Estonia, France, Italy, Serbia, Spain, Switzerland, the United States	Ireland, Romania, Russian Federation, Slovakia, Slovenia, Spain

Out of 27 countries, 21 were placed into first three groups. The remaining six countries were left to the fourth group. These countries included Ireland, Romania, the Russian Federation, Slovakia, Slovenia and Spain. Some countries indicated with one or two product flows, that they could have been placed into one of the other three group. However, requirement of three or more was not met and therefore they were placed into fourth group.

One objective of this research was to find out if there are similarities between countries in same group. In order to find out, selected countries from each country would be interviewed. Interviewing people from each country who produce the forecast is made with e-mail. Another option was interviewing face-to-face via Skype but traditional face-to-face was not an option since travelling to each country is rather expensive and time consuming. Using e-mail was chosen over other options, because it provides each person to have enough time for answering and possibly consulting more people from their team. Since these forecasts have been made over 15 years, it is likely that more than one person from each country has been part of making predictions. Questionnaires that were sent for selected representatives of selected member States consist of 6 main questions. These questions are designed to find out if there are any specific mechanics or tools that are used, as well as when estimates and forecasts are produced, who they are made for and who are involved in this process. The following questions were asked:

- What tools or methods do you use for producing the predictions?
- Who is involved in generating the predictions?
- Timing of forecasting:
    - When do you usually produce predictions for the rest of the on-going year?
    - When do you produce predictions for the following year?
- Who are the primary users of predictions you produce?
- Do you have any specific goals in your mind when you produce the predictions?
- Any other comments?

Two of the selected countries were unavailable: United States and Latvia. Responsible person from the United States had retired and was therefore not available. Earlier contact from Latvia had left this position and new person was not yet appointed. In addition, responsible persons from Sweden and Switzerland didn’t answer in given time. To have more answers, this questionnaire was also sent to Norway and Czech

Republic, who didn't answer and to France and Serbia, who did answer. In total, there are 7 countries who sent answers: Austria, Estonia, Finland, France, Netherlands, Poland and Serbia. So, answers from these countries are included and analysis is formed to see if there are anything in common or clear patterns between countries.

3.7 Evaluating different data quality dimensions

A survey was made in March 2019 in a meeting of the ECE/FAO Team of Specialist on Forest Products Statistics held in Geneva (UNECE 2019b), where representatives of member States, as well other people working closely with this data, were asked to evaluate different dimensions of data quality. The objective of this survey was to find out which of the dimensions selected earlier was the most important when producing forest product predictions. Each person was presented with qualities and the definitions of these qualities are as defined by Wang and Strong (1996). Then they were asked to rank qualities from 1 to 9, ranking the most important as 1. Each person had also a possibility to comment on all the elements affecting data quality. These qualities were presented earlier in table 2.

4. RESULTS

Before the financial crisis there was a long period of growth in the global economy. According to World Bark, in 2006, the economy growth of the whole world was 4.29% on average (World Bank 2019a). At first, the financial crisis of 2007-2009 had a clear impact on estimates and forecasts. Before financial crisis, all predictions are usually smaller than the actual outcome and after 2007 forecasted data are significantly larger than actual outcomes. This is logical since the financial crisis came as a surprise for many companies and countries. The World Bank has tracked economic growth since 1961 and 2009 is the only year when the average economic growth of the whole world was negative (World Bank 2019a). Of course, some countries in UNECE region were affected with negative economic growth for more than one year. At most cases forecasts are similar to estimates and repeated data, but one year behind. Overestimating of forecasts appear in the graphics as positive values, since it is bigger than the actual outcome. This is presented in the Figure 3 below.

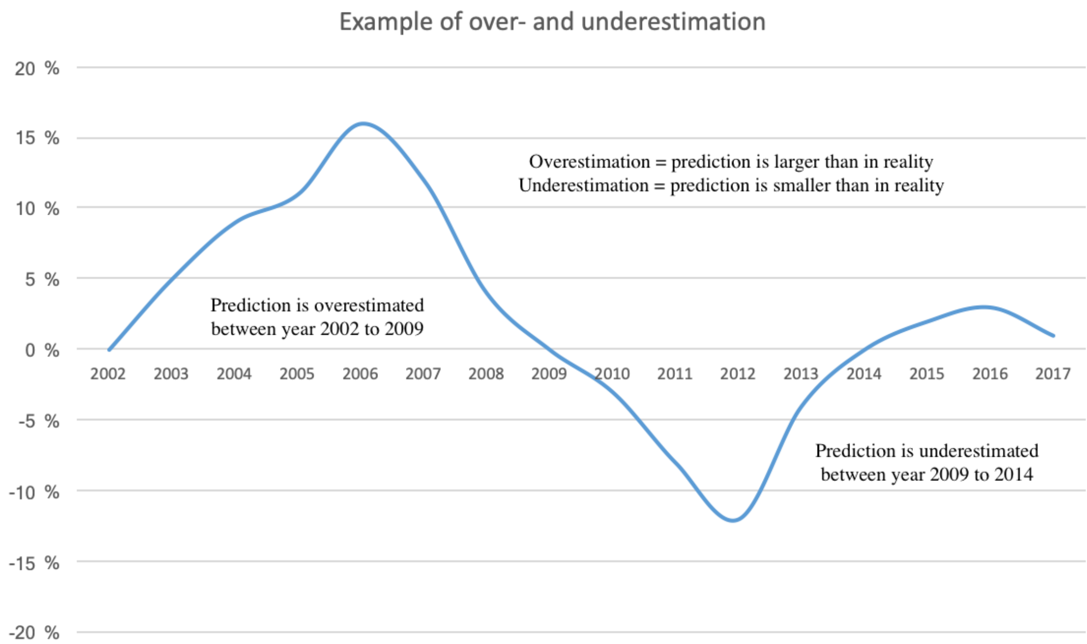

Figure 3. Explanation of over- and underestimation in results.

The economy continued to grow after the financial crisis, although growth was slower in some countries. For example, in 2012, in addition to 2009, the growth in the European Union was negative (World Bank 2019b).

Financial crisis had an even bigger effect on a country's major importers and exporters of forest products. For example, differences from actual outcomes of the export and production estimates of Finland and Sweden were clearly higher than European average difference (% outcome). This is due to not being able to see the financial crisis before it happened. These countries were also hit fairly hard with the change that people had with forest products. Paper was used less than ever as habits of people transformed from traditional paper to electronic services. In 2007 the apparent consumption of paper and paperboard in UNECE region was 207,696 tonnes and it dropped by 14.5% in two years to just 177,526 tonnes in 2009 (UNECE 2010).

4.1 Grouping of countries by prediction quality

Each member State was grouped into one of four different groups outlined earlier in table 9. This was done with visual analysis based on how different product flows looked like. The characteristics of the first group were distinctive. All three charts, estimate, forecast and repeated usually stayed really close together. In some cases, forecast might make same "moves" one year behind the other two. It is hard to say which chart is the best since they usually stay so close together. Figure 4 is an example graph of Poland with all products and product flows between 2002 and 2017. Group 1 consist of 7 countries: Poland, Latvia, Croatia, Netherlands, Cyprus, Czech Republic and Turkey. Out of these countries, the only ones with any problems is Cyprus and Turkey. Cyprus is a really small country, with minimal volumes in forest products. Out of five product flows, only all the products and imports had meaningful volumes. Therefore, the Cyprus was placed with only 40% match from product flows. Turkey have some problems that are unique for Turkey alone: series seem to be underestimated for most parts, but different charts seem to stay together.

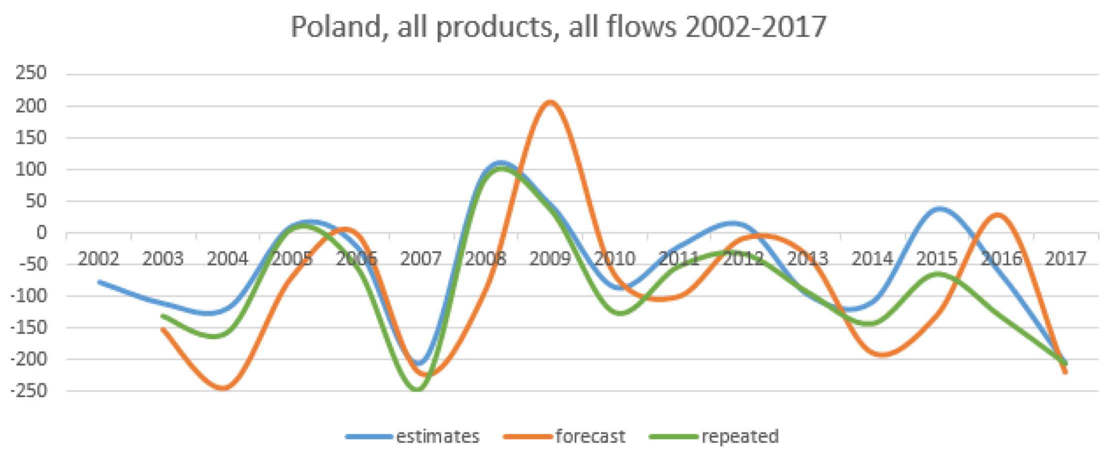

Figure 4. An example of group 1: All of Poland's products and product flows from 2002 to 2017.

Note: Figure presents absolute values of all products and therefore no unit can be used (see Table 4).

Countries in the second group have one surprising feature of near perfect estimates. Forecast and repeated might not be so close to the actual outcomes but every year the estimates are nearly perfect. Finland's products and product flows between 2002 and 2017 is an example of this, as seen in Figure 5. Group 2 consists only 6 countries: Finland, Sweden, Norway, Austria, United Kingdom and Lithuania. While United Kingdom and Lithuania weren't perfect match, but still clearly fit into this group. In this group estimate was clearly the best out of any forecast data and something that all countries should aim for.

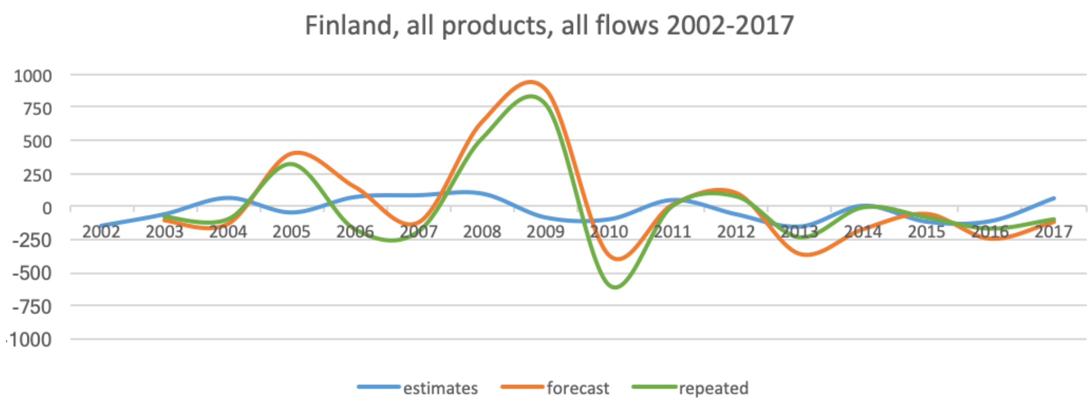

Figure 5: An example of group 2: All of Finland's products and product flows from 2002 to 2017.

Note: Figure presents absolute values of all products and therefore no unit can be used (see Table 4).

Group 3 was the biggest group and consists of 8 countries: France, Estonia, Italy, Spain, Switzerland, United States, Serbia and Canada. In this group the figures look as you would expect, with clear difficulties between 2008 and 2011. There is clearly a spike around 2008, which suggest that forecasts overestimated the volumes. Around 2010 there is a clear drop, which suggest that volumes were underestimated. With most countries, there is not big changes after the year 2012 and all three product flows are usually pretty good from there on. In this group, estimates are usually the best but not good enough to fit into group 2, where estimates are nearly perfect.

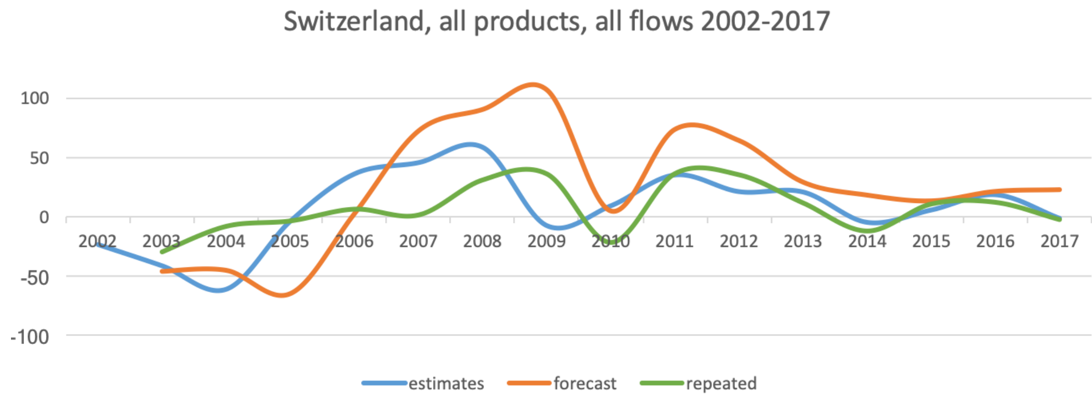

Figure 6: An example of group 3: All of Switzerland's products and product flows from 2002 to 2017.

Note: Figure presents absolute values of all products and therefore no unit can be used (see Table 4).

After sorting countries into groups, three countries from each group were contacted and asked how they produce forecast information. The selection of countries was done with an aim to pick those that would match the definition of group as well as possible and pick countries with different product volumes. In the end, answers from Finland, France, Poland, Austria, Netherlands, Estonia and Serbia were received and analysed.

4.2 Analysing the answers

For easier comparison, each answer was cut down to 5 sections: category of country (big, medium or small), methods and tools used, number of people involved in the process, primary users of these forecasts and any specific goals, as well as other comments. Three big countries were part of this questionnaires: Finland, France and Poland, two medium countries: Austria and Netherlands. and two small countries: Estonia and Serbia. This was planned so that a more even view of all member States could be achieved.

As for methods and models, there are some elements that are common in all member States but there are also differences. Most of the member States use outside experts for knowledge, which then should help when making better estimates and forecasts. Outside experts include trade associations, economists and other professionals who work in the forest product industry. In some cases, these numbers are compared to previous year's numbers and this should provide a benchmark. When these predictions are made, usually around August or September, preliminary data from quarter 1 and quarter 2 are available and used to make better estimates. Answer from France specified a formula, where data from Q1-2 is compared with a previous year's data of Q1-2. Formula behind was, that previous year's Q1-2 value was divided by the whole year's data and this percentage value was used to multiply this year's Q1-2 value to match whole year's data. Finland has published a presentation where they go through the methods of making predictions for forest products (Natural Resources Institute Finland 2019). Their starting point is that estimates and forecasts are demand-driven: demand creates exports, which promotes production, that will affect domestic roundwood markets and roundwood imports, which should rise the stumpage prices. In the end the demand will create also employment and profitability for all parties involved. This is essential when producing predictions, as demand will have an effect on all parties, products and product flows involved. Answer from Poland also specified that create a regression analysis with the statistical data collected.

The number of people involved in the process is fairly similar for all countries. Usually there are one or two persons who are producing the predictions but they receive help from outside experts or from other colleagues from the same organization. Finland was the only country, where more than 2 roles were specified, as there were in total 5 people involved in the process. In some cases, national statistical office or some other unit of the ministry is helping with the analysis. France was the only country, where predictions are produced only for UNECE, as they want to improve before producing them for French administration. In other countries, predictions are produced also for other shareholders. For example, the Ministry of Environment and national market outlooks use these predictions.

Other comments that came up in the answers were that collecting the knowledge from outside experts offers a good possibility to stay in touch with them and having a good feel what is going on in the national markets. Answer from Estonia mentioned that they usually do rather conservative predictions, whereas in Serbia they want to produce as reliable predictions as possible. In Poland they wanted to highlight, that they rely on statistical data and the experience of experts to produce as trustworthy predictions as possible.

In the end, most of the countries use similar methods to produce estimates and predictions. While there are some models and formulas, in most cases predictions are produced fairly freely. Outside experts from the industry are consulted a lot, which is essential when gathering information. In addition to outside experts, Q1-2 numbers of production, removals and trade are available when producing estimates and forecasts, which provides an important help. There doesn't seem to be any pattern between how predictions are made and countries from same size group or categories based on how different product flows look like (see table 9). This isn't all that surprising, as producing predictions seem to be done with opinions of many rather than strict science.

4.3 How elements of data quality affect predictions

As explained earlier in section 3.7, people who work with forest product predictions were asked to evaluate nine dimensions of data quality. In total 14 people took part in survey, where 11 were directly involved producing forest product predictions. 5 out of 9 elements were ranked as the most important from at least 1 person: accuracy, believability, completeness, consistency and relevancy. Overall, consistency was ranked the most important element of data quality, followed by believability and relevancy. More detailed results are presented in table 11 below.

Table 10. Correspondent evaluation of data quality dimensions.

Rank	Dimension	Average ranking
1	Consistency	3.29
2	Believability	4.07
3	Relevancy	4.32
4	Completeness	4.68
5	Accuracy	4.89
6	Timeliness	5.39
7	Accessibility	6.00
8	Interpretability	6.25
9	Appropriate amount of data	6.43

Consistency was ranked the most important dimension only 2 times out of possible 14. However, the lowest ranking it got was 6 by two answers and 50% of answers had consistency in the top 3 dimensions. This meant, that consistency had average ranking of 3.29, which was clearly the best. It was followed by believability, which had an average ranking of 4.07. Believability was ranked inconsistently with all; some people ranked it the most important (4 times) and by one person the least important dimension. All in all, believability remains in the second most valued dimension in the questionnaire and it was ranked in the top 3, 50% of the time. The third highest ranked dimension was relevancy, with average ranking of 4.32. It was ranked the most important twice and in the top 3 only 5 times. Much like believability, relevancy was ranked both the most and the least important dimension of data quality. The fourth highest ranked dimension, completeness, was ranked on average 4.68 and it is arguably one of the more interesting dimensions in this questionnaire. Based on rankings, it seems to be ranked either very close to the top or very close to the bottom. It has the top 3 ranks 8 times, which is more than any other dimension, but also in the bottom 3 ranks 4 times. This means, that only two people ranked completeness in the middle. Completeness is followed by accuracy with average ranking of 4.89. Accuracy, the fifth highest ranked dimension, was ranked the most important twice and in the top 3, 50% of the time. It has very similar situation with completeness, as it seems to be ranked either in the top or in the bottom, with only 1 person ranking accuracy in the middle. It is fair to say that these two dimensions are controversial. They are both either ranked near the top or near the bottom and very rarely in between. The sixth dimension, based on ranking, is timeliness, with average rank of 5.39. It was ranked mostly in the middle, with 50% of rankings being between 4 and 6. None of the participants of this survey ranked it the most nor least important dimension. The seventh dimension, accessibility, had average ranking of 6.00. It seemed to be one of those dimensions that some people valued close to the top and others rated in the least important (4 times). It was followed by the eighth dimension, interpretability, with an average ranking of 6.25. It was rated only once in the top 3, and six times in the bottom 3. The least important dimensions based on this survey was the ninth dimension, the appropriate amount of data, with average ranking of 6.43. It was rated in the bottom 3 almost 65% of the time and only three times in the top 5. The last three dimensions, timeliness, interpretability and appropriate amount of data, are different to first six, as they mainly affect users of data. As this questionnaire was answered mainly by those who are producing predictions, dimensions more closely for users are not as highly valued.

Range of rankings were relatively close, as the ranking of different dimensions are between 3.29 and 6.43. Five dimensions were ranked the most important by at least one person, spreading answers broadly. On the other hand, six dimensions were valued the least important by at least one person, which supports previous observation. One reason for this might be, that some dimensions are quite similar to other dimensions in this survey. There are differences between each dimension, but each participant might comprehend them differently. All in all, the answers provide good understanding which dimensions of data quality are valued over other qualities. And when producing predictions, consistency, believability and relevancy are regarded as the most important dimensions.

4.4 Comparison of different products

Now that there is an understanding in how reliable predictions between member States are, what about products? To have an answer to this question, the difference for each country from every year and product was calculated. In order to show a percentage difference, all values used in the calculations are changed to absolute values. This was done because, as presented before, in some years the predictions have been overestimated and in others underestimated. If these values would not be absolute they would, in some cases, counter each other out. Values presented in tables of this chapter are all percentage differences from the actual value. This means, that the smaller the percentage is, the better. If the value is 9.2%, the prediction has been 90.8% same as historical value.

There are clearly products with different success rates, as displayed in Table 12 below. Every product has a more accurate estimate than forecast, which was expected as estimates are done one year before forecasts. When looking overall average numbers, estimates are over 10% more accurate than forecasts. However, repeated data is better than estimates in 7 out of 12 product averages between 2002 and 2017, and on average 1.1% more accurate. This is unexpected based on how predictions are made, as learned earlier when discussing about styles of making these. Since estimates are usually based on last year's data and then updated with new knowledge, they should be more accurate.

Table 11. All products and estimates, forecast and repeated %-difference from actual outcome.

Product description	Product code	Estimate	Repeated	Forecast
Paper and paperboard	10	10.0%	8.6%	15.4%
Coniferous saw logs and veneer logs	1.2.1.C	12.4%	11.5%	16.8%
Coniferous sawn wood	5.C	12.6%	13.4%	21.3%
Particle board (including OSB)	6.3	13.5%	14.7%	21.1%
Fibreboard	6.4	18.4%	25.5%	37.2%
Coniferous pulpwood	1.2.2.C	19.6%	15.3%	25.1%
Non-Coniferous pulpwood	1.2.2.NC	22.5%	18.4%	29.8%
Plywood	6.2	25.1%	25.4%	39.8%
Non-Coniferous saw logs and veneer logs	1.2.1.NC	27.5%	22.0%	36.0%
Non-Coniferous sawn wood	5.NC	28.1%	18.0%	46.9%
Wood Pulp	7	53.4%	71.0%	70.9%
OSB	6.3.1	56.0%	42.2%	69.1%
	Averages	**24.9%**	**23.8%**	**35.8%**

When looking more closely at different products, a clear trend with all products is visible: estimates and repeated data has usually similar success rate and forecast is noticeably worse. Only in one product, wood pulp, is forecast better than either estimate or repeated data.

The most accurate product is paper and paperboard which had the smallest difference between predictions and actual value in all three sections. Estimate was on average 10% off from actual value, while repeated was 15.4% and repeated data was clearly the best: only 8.6% from actual value. The second best product was coniferous saw logs and veneer logs, with estimate 12.4%, repeated 16.8% and repeated only 11.5% off from actual value. Coniferous sawn wood has nearly as good estimate as 1.2.1.C with 12.6%, but forecast is noticeably worse since it is down to 21.3%, while repeated is steady at 13.4%. From there it goes slowly downwards, but only last two products, wood pulp and OSB are clearly more unreliable than others. These differences are usually caused by predictions from few, which are way off. In many cases also paired with small quantities, where a drop of 2 metric tonnes could cause a 200% difference.

As discussed before, some countries with small volumes could have major percentage differences from the actual values and that might not be meaningful. To get around this problem, Table 13 below takes a look at all of the products, but this time only estimates. There are also percentages without 5 least accurate countries of that specific product and also only the most reliable 5 countries of that specific product.

Table 12. Estimates and average percentage difference to actual value.

Product description	Product code	Average difference (%)	Without 5 least accurate countries	Only the most accurate 5 countries
Paper and Paperboard	10	10,0 %	6.1%	2.8%
Coniferous sawlogs and veneer logs	1.2.1.C	12,4 %	9.3%	4.2%
Coniferous sawn wood	5.C	12,6 %	9.6%	4.4%
Particle board (including OSB)	6.3	13,5 %	10.5%	5.9%
Fibreboard	6.4	18,4 %	12.1%	7.3%
Coniferous pulpwood	1.2.2.C	19,6 %	13.1%	5.0%
Non-Coniferous Pulpwood	1.2.2.NC	22,5 %	16.3%	7.6%
Non-Coniferous sawlogs and veneer logs	1.2.1.NC	27,5 %	11.5%	6.5%
Plywood	6.2	27,8 %	19.4%	10.5%
Non-Coniferous sawn wood	5.NC	28,1 %	14.5%	7.1%
Wood Pulp	7	53,4 %	22.2%	3.8%
OSB	6.3.1	56,0 %	30.4%	8.5%
	Average	**24.9%**	**14.6%**	**6.1%**

Including results only from the 22 most accurate countries of that specific product is done, when viewing results without 5 countries with the least accurate prediction. This should provide us a better understanding of how well different products are comparing with each other. This means, that not necessarily same countries are always excluded, as it depends on the results. Similar method is used, when only the most accurate countries are presented. Here can be seen, that when 5 least accurate countries are not included, it evens out all products. Average percentage is down from 24.9% to 14.6% when not including 5 least accurate countries of that specific product. While this could be misleading, when taking away results from 5 countries out of total 27, it also provides a better view of how different products are actually doing. Obviously all products have better percentage difference to actual values than before, but it especially makes products with worst estimates look better. The difference with OSB has increased more than 25%, from 56.0% to 30.4% and wood pulp has even better increase of quality with percentage difference down from 53.4% to 22.2%. Another noticeable improvement is non-coniferous saw logs and veneer logs, where the percentage difference has improved from 27.5% to 11.5%.

While the order of products does not change significantly when ignoring bottom 5 countries, it does when only looking at the numbers from the 5 countries with most accurate results. Paper and paperboard are yet again the most reliable product, but after that comes wood pulp, which has been near to bottom previously. This comes down to having the biggest producers and exporters of pulp in that top 5, which leads to having steady estimates in nearly every year and therefore only 3.8% difference from year to year. Another interesting product is plywood, where top 5 percentage is 10.5%, making it the least accurate of all products. When looking more closely, it seems that all countries are really close together, meaning that there are no outstanding estimates, neither in good or bad.

Table 13. Forecasts and average percentage difference to actual value.

Product description	**Product code**	Average difference (%)	Without 5 least accurate countries	Only the most accurate 5 countries
Paper and Paperboard	10	15.4%	9.7%	5.3%
Coniferous sawlogs and veneer logs	1.2.1.C	16.8%	13.8%	8.7%
Particle board (including OSB)	6.3	21.1%	16.3%	9.5%
Coniferous sawn wood	5.C	21.3%	16.5%	9.7%
Coniferous pulpwood	1.2.2.C	25.1%	17.9%	7.2%
Non-Coniferous Pulpwood	1.2.2.NC	29.8%	20.9%	10.2%
Non-Coniferous sawlogs and veneer logs	1.2.1.NC	36.0%	17.7%	8.0%
Fibreboard	6.4	37.2%	17.2%	10.9%
Plywood	6.2	39.8%	19.4%	11.2%
Non-Coniferous sawn wood	5.NC	46.9%	24.5%	11.6%
OSB	6.3.1	69.1%	41.9%	12.6%
Wood Pulp	7	70.9%	30.9%	7.0%
	Average	**35.8%**	**20.6%**	**9.3%**

In table 14, similar values have been calculated for forecasts. Average percentage difference of all products and countries is 35.8%, but without 5 least accurate countries in each specific product the percentage difference is down to 20.6%. When looking values without 5 least accurate countries, the order of products seems to stay pretty similar. All products have better success rate as before, but more importantly the percentage values are more reasonable. For example, fibreboard is on average 37.2% off from actual value, but without 5 least accurate countries it is only 17.2% off. While 17.2% is far from perfect, it's starting to be more reasonable error rate for forecasts. This is the case with almost every product and only two products, OSB and wood pulp are above 25% difference on average.

When looking at the most accurate 5 countries only, it is similar to estimates. The percentage difference without 5 least accurate countries was 20.6% and it is down to just 9.3% with 5 most accurate countries. The order of products changes more with when there are only the 5 most accurate countries, comparing to having results without 5 least accurate countries. Yet again, paper and paperboard is the most accurate with 5.3% difference and wood pulp is again the second most accurate with 7.0%. However, the percentage differences are close together with all products, ranging from 5.3% to 12.6%.

Table 14. Repeated and average percentage difference to actual value.

Product description	**Product code**	Average difference (%)	Without 5 least accurate countries	Only the most accurate 5 countries
Paper and Paperboard	10	8.6%	5.6%	3.5%
Coniferous sawlogs and veneer logs	1.2.1.C	11.5%	9.5%	5.8%
Coniferous sawn wood	5.C	13.4%	9.8%	5.5%
Particle board (including OSB)	6.3	14.7%	11.4%	6.7%
Coniferous pulpwood	1.2.2.C	15.3%	11.4%	5.7%
Non-Coniferous sawn wood	5.NC	18.0%	12.8%	7.6%
Non-Coniferous Pulpwood	1.2.2.NC	18.4%	12.6%	5.5%
Non-Coniferous sawlogs and veneer logs	1.2.1.NC	22.0%	10.9%	5.3%
Plywood	6.2	25.4%	14.8%	8.4%
Fibreboard	6.4	25.5%	10.5%	6.8%
OSB	6.3.1	42.2%	25.2%	9.1%
Wood Pulp	7	71.0%	19.8%	4.4%
	Average	**23.8%**	**12.9%**	**6.2%**

Table 15 above presents same calculations done for repeated data as for estimates and forecasts before. Average percentage of difference is down from 23.8% to 12.9%

when not including 5 least accurate countries. This is, once again, a massive improvement when evaluating the success of repeating old data. One of the most noteworthy product is wood pulp, with a massive improvement from 71.0% to 19.8%. However, this doesn't change the order of products significantly, as all the remaining products have clearly better rate of success without the bottom countries.

When looking about success rates with top 5 countries of each product, can something interesting be seen. While all products are relatively close together, ranging only from 3.5% to 9.1% and therefore having an average of 6.2%, meaning that for the first time, estimates are better than repeated data. Although estimates are only slightly better, 6.1% to 6.2% of repeated, it is worthwhile to point out that this is the only time this has happened when comparing the averages of all products. As stated previously, beating repeated data is the benchmark for all the predictions of data and this is why it is so crucial.

As this study has proved, paper and paperboard are clearly the most accurate products. It has the smallest percentage of error of all products and in all three categories of predictions. When looking at estimates, 25 out of 27 countries had less than 20% difference between actual values and estimates, while 19 out of 27 had less than 10% average difference. And in forecasts 23 countries of out possible 27 with less than 20% difference, while in repeated 26 of 27 manages to reach that. So, almost all countries manage to get all predictions really close to actual values and there aren't any countries, where any prediction would be way off. But why is that? There are multiple reasons for this, but it comes down to product discussed here: paper and paperboard. It is an important product with big volumes. It's not as sensitive to changes with quick schedule as other products might be, since production is usually planned far ahead and new capacity requires planning.

One of the biggest factors is volume, since products with smaller volumes are more vulnerable to changes. This does not only affect paper and paperboard, but also other products. As presented earlier in Table 4**,** different products have a variety of volumes, many products with big volumes are close to the top in this comparison as well. There is, however, one product that doesn't seem support this theory: wood pulp. Volumes of wood pulp are fairly large, but for some reason it doesn't seem to translate into ability to make accurate predictions. However, when taking a closer look at countries for that specific product, it does seem that countries with massive volumes of production, exports or imports are close to the top with good average percentages. Unlike paper and paperboard, the quality of predictions is dropping quite rapidly after top 10 countries. This would ultimately support the theory that bigger volumes should mean better estimates and forecasts, as well as reliable repeated data.

4.5 Comparison of different countries

Below in Table 16 can be seen each of the 27 countries with percentage differences for each three prediction groups. In previous chapter showed that out of 12 products, repeated data was more accurate 7 times. The trend is similar with individual countries as well, which is expected as same data set is used for this chapter as with products. In fact, out of 27 countries estimate is the most accurate in 11 times and repeated data is better in 16 times. Yet again, forecasts are never the best prediction out of three, but in few cases the forecast is better than either estimate or repeated.

Table 15. Countries and percentage difference for each product flow compared to actual value. (Continues to following page.)

Member State	estimate	repeated	forecast
Austria	6.3 %	7.2%	9.3%
Poland	7.8%	7.7%	11.4%
France	9.5%	7.1%	18.3%
Germany	9.8%	9.2%	11.9%
United States	9.9%	7.4%	13.9%
Canada	13.3%	7.2%	15.6%
Italy	14.0%	15.6%	22.5%
Lithuania	14.0%	16.1%	21.8%
Spain	14.8%	18.1%	32.9%
United Kingdom	15.4%	16.8%	30.9%
Netherlands	17.4%	14.6%	21.1%
Turkey	18.0%	14.7%	22.6%
Finland	19.4%	16.9%	21.8%
Czech Republic	21.1%	10.5%	26.4%
Estonia	22.2%	23.9%	39.9%
Latvia	22.5%	36.0%	38.2%

Sweden	25.4%	12.9%	22.4%
Switzerland	28.2%	26.1%	33.4%
Serbia	29.0%	29.3%	42.5%
Russian Federation	29.9%	14.1%	33.6%
Croatia	31.4%	34.0%	38.3%
Romania	36.5%	34.5%	50.8%
Cyprus	42.2 %	56.6%	89.9%
Slovakia	44.6%	39.2%	77.9%
Norway	45.8%	34.9%	65.7%
Slovenia	48.3%	87.4%	77.3%
Ireland	83.3%	50.0%	88.2%

There is one country that does not support the theory that having big volumes makes more accurate predictions: the Russian Federation. There is a clear reason for this and it is slightly due the structure of this research. Historical data, which is used in this research can be revised later if new information about it is provided. This means, that in some cases the comparison of estimate or forecast is not made with the historical data that would be produced during that year but rather with historical data that is revised several years later. With the Russian Federation, all the historical data of removals has been revised later. Therefore, estimates and forecasts of removals (products 1.2.1.C, 1.2.1.NC, 1.2.2.C and 1.2.2.NC) produced doesn't look very accurate. This problem makes the average error rate look worse than it should.

4.6 Comparison of different sections

Now that there is an understanding how different products and countries compare with each other, how does different predictions compare with each other? To find out how well estimates, forecasts and repeated data compare, the best prediction was counted for each product and each country. As some countries were missing some products, all products do not add up to 27. Below in Table 17 are each product with scores for all three prediction groups. Estimates were the most accurate in 138 cases out of possible 321, while forecasts were the most accurate in 11. For the remaining 172 times, repeated data were the most accurate, making it the most accurate 53.6% of the time.

Table 16. How different predictions compare with each product.

Product description	**Product code**	Estimate	Repeated	Forecast
Coniferous sawlogs and veneer logs	*1.2.1.C*	16	11	0
Non-Coniferous sawlogs and veneer logs	*1.2.1.NC*	11	14	2
Coniferous pulpwood	*1.2.2.C*	14	12	0
Non-Coniferous Pulpwood	*1.2.2.NC*	6	18	1
Coniferous sawn wood	*5.C*	13	14	0
Non-Coniferous sawn wood	*5.NC*	9	18	0
Plywood	*6.2*	7	20	0
Particle board (including OSB)	*6.3*	14	9	4
OSB	*6.3.1*	10	15	2
Fibreboard	*6.4*	13	13	1
Wood Pulp	*7*	14	12	1
Paper and Paperboard	*10*	11	16	0
	Total	**138**	**172**	**11**

When taking a closer look, particle board becomes an interesting product, since 4 times forecasts were the most accurate. As learned earlier in table 12 forecast error rate for particle board was only 21.1%, which wasn't too far from the estimates and repeated data. This means, it is the product with the smallest number of most accurate repeated data as well.

It is good to point out, that while in plywood predictions, repeated data is better 20 times, there isn't such a big difference in the error rate. Looking back to Table 12 with error rates, estimates actually have smaller percentage difference to actual value than repeated data. This is surprising, since error rates are impacted how well different predictions are doing. When taking a closer look to numbers, it does seem that all countries have really similar error rates in estimates and repeated data. The same could not be said about product non-coniferous sawn wood, where repeated data are better 18 times out of 27. Error rates of repeated data and estimates are clearly different. Percentage difference with estimate is 28.1% comparing to 18.0% of repeated, it does clearly affect also how many times repeated data is better than estimates.

Table 17. How different predictions compare with each country.

	estimate	repeated	forecast
Austria	8	4	0
Canada	4	7	1
Croatia	2	10	0
Cyprus	4	5	1
Czech Republic	5	7	0
Estonia	6	6	0
Finland	8	3	1
France	5	7	0
Germany	6	5	1
Ireland	2	8	1
Italy	6	5	1
Latvia	9	3	0
Lithuania	9	3	0
Netherlands	3	8	1
Norway	5	7	0
Poland	5	7	0
Romania	1	9	2
Russian Federation	3	9	0
Serbia	8	4	0
Slovakia	4	8	0
Slovenia	5	6	1
Spain	9	3	0
Sweden	3	9	0
Switzerland	2	10	0
Turkey	6	6	0
United Kingdom	7	4	1
United States	3	9	0
Total	**138**	**172**	**11**

In the Table 18 above are same numbers, but other way around. This table is displaying what is the best prediction group in each country. From 27 countries, estimates were the most accurate 9 times, while repeated data were more accurate 16 times. In two cases, both estimates and repeated had similar numbers with 6 products each. This is similar to percentage differences, where estimates were the most accurate in 11 countries. However, in Romania forecasts were the most accurate two times out of possible 12, but more surprisingly estimates were the most accurate only with 1 product. The surprising part is, that in Romania both estimates and repeated data have very similar percentage difference from actual values but this doesn't seem to have an impact here.

Other countries with clear majorities in one prediction groups are the United States, Switzerland, Sweden, Russian Federation and Croatia, who have 9 or more products where repeated data is the most accurate. On the other hand, in Spain, Lithuania and Latvia estimates are the most accurate 9 times out of possible 12 products. In addition, Serbia, Finland and Austria have 8 products where estimates are the most accurate.

Lithuania, Croatia and Switzerland have similar situation with Romania, as results here suggest that it would be one-sided, but as seen earlier percentage results are actually very even. Both Switzerland and Croatia have 10 products, where repeated data are more accurate. However, in Croatia estimates are indeed on average more accurate than other predictions, although both repeated and estimates are close. In Switzerland repeated data are on average better, but only so slightly 26.1% to 28.2% and that makes it surprising that numbers in Table 18 are so flop-sided. In Lithuania the situation is similar to Switzerland: while estimates are slightly better, it comes as a surprise that 9 out of 12 products have more accurate estimates.

To conclude, if repeated data was the benchmark for each country, most of them have failed to meet that requirement. Most of the time estimates of each country are at least close to repeated data, but on average can't reach that level.

4.7 Comparison of different product flows

As highlighted earlier in Table 5, there are in total of four different product flows used: imports, exports, production and removals. Now that it is known which products and countries are the most accurate with their predictions, it is interesting to find out if there are any clear differences between different product flows. As it can be expected, there are clear differences between each product flow. This is section presenting all product flows with graphs and visual analysis, as well as looking at numbers. As mentioned before, there is one clear flaw in this approach: one big drop for one country could make a clear impact even when measuring totals for 27 countries.

Production is the most accurate in all three prediction categories and overall there are not any major problems in any year. Below in Figure 7 presents average percentage difference from the actual values. On average estimates are 12.5% from the actual values, while forecasts are 19.0% and repeated data are 10.1%. As can be seen from the figure 7, the years between 2008 and 2011 are well above the averages. This is due financial crisis as discussed earlier. After 2013 the percentage error is clearly smaller than before and there doesn't seem to be any movement for either way in any of the three prediction groups.

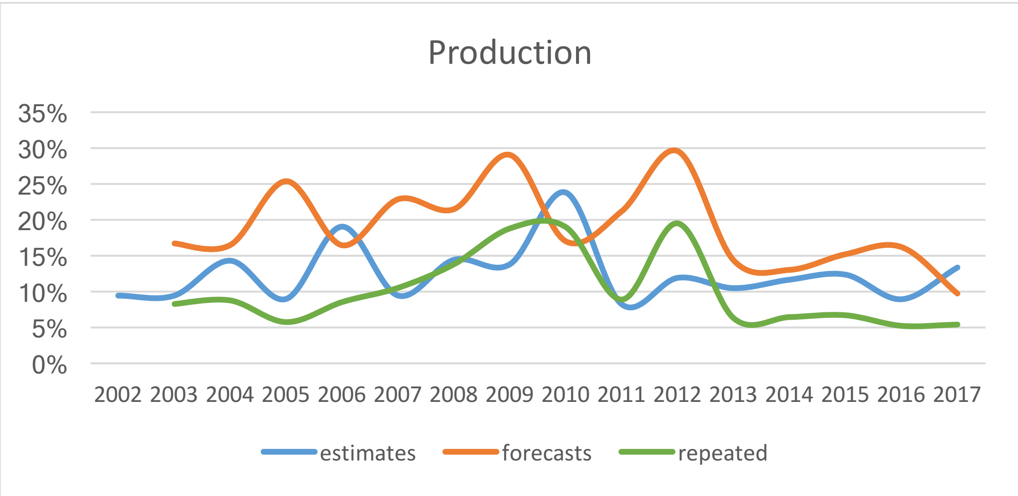

Figure 7. Progression of estimates, forecasts and repeated in production between 2002 and 2017.

Removals have percentage error of 14.6% for estimates, 20.0% for forecasts and 15.3% for repeated data, making it the second most accurate. Again, financial crisis has a clear impact on how accurate each prediction is. Things have turned back normal after 2011, although there seems to be a clear spike in 2013. However, this is caused by one country, where volume of removals dropped dramatically over 1000% in a year. In fact, if all the data of 2013 would be ignored, repeated data would be more accurate than estimates.

Overall removals do seem to be an accurate product flow and majority of countries have success on making predictions. This is presented in Figure 8 below. By 2017 all three predictions are less than 10% off from actual values, which is the best year. Reasons for good results are big volumes, where small drops don't make big changes in totals and planning that goes to harvest: countries usually plan volumes of harvest well ahead to match production. Another important reasons for success of removals is that people who are responsible for harvest often share their knowledge with correspondents.

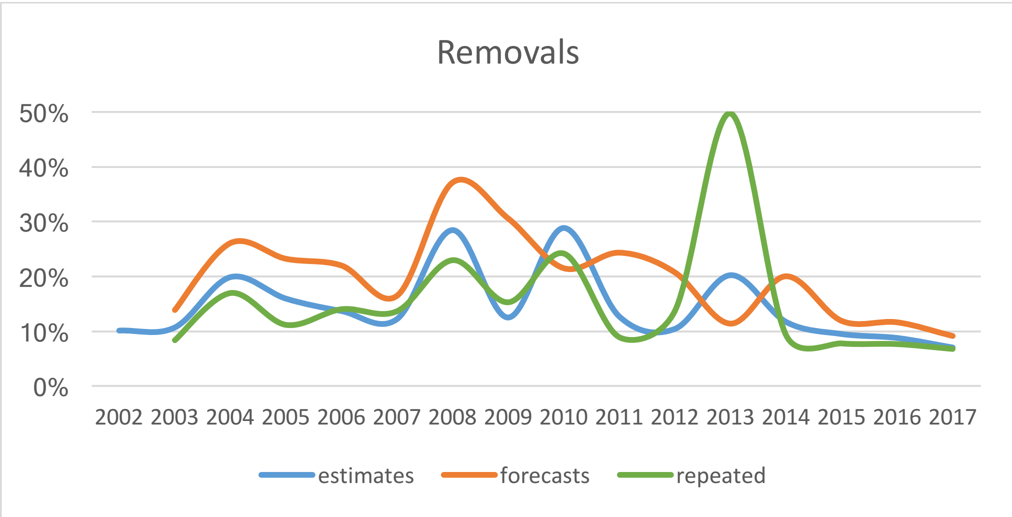

Figure 8. Progression of estimates, forecasts and repeated in removals between in 2002 and 2017.

The third best product flow, imports, is fairly close to removals, as shown in Figure 9. Estimates of imports are on average 16.6% off from actual value, while forecasts are 24.9% and repeated data are just 13.4%. When looking at the figure, it is clear that financial crisis has had a big effect on the imports. Both forecasts and repeated data make a massive spike in 2009, but it takes just a year to adjust forecasts. As estimates are produced during the year, they are not so far off. Compared to production and removals, it seems that quality of all three predictions of imports are similar for each year and especially after 2011, there is no real change. However, the quality of predictions never reaches the levels previous product flows have managed.

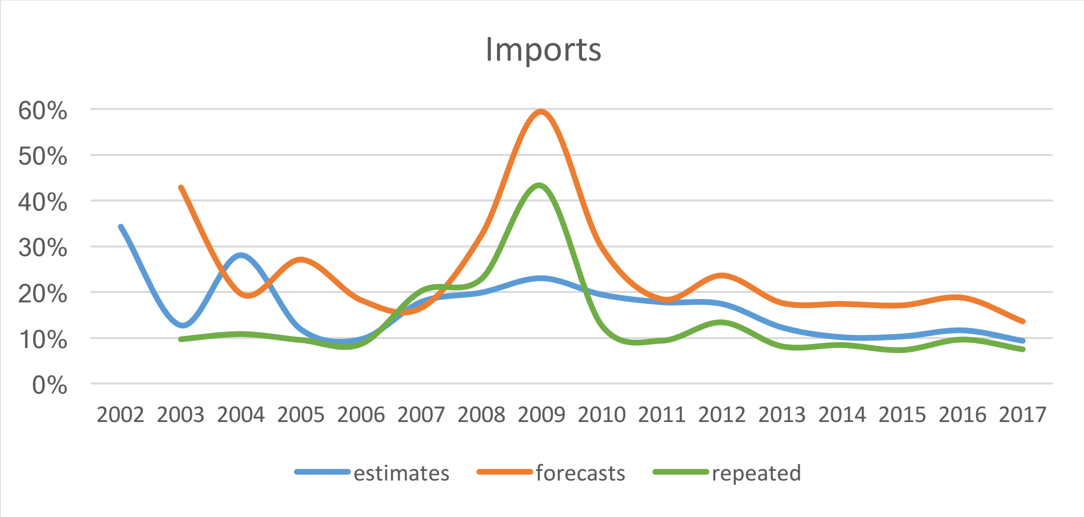

Figure 9. Progression of estimates, forecasts and repeated in imports between in 2002 and 2017.

Out of the four product flows, exports are the least accurate. Estimates are the most accurate of three predictions with 36.1% average difference from actual values, while forecasts are 58.6% off and repeated data 43.3%. As presented in Figure 10, in 2007 repeated data makes bigger spike than in any other product flows and reaches close to 250% difference. This spike is, depending on prediction, around years 2007 and 2008, which is one year earlier than in imports. This means that countries are declining with imports only after exports are dropping. While of course some countries have reasonable small differences in all three predictions, majority of countries are not particularly accurate. Much like in imports, estimates react faster than the other two predictions.

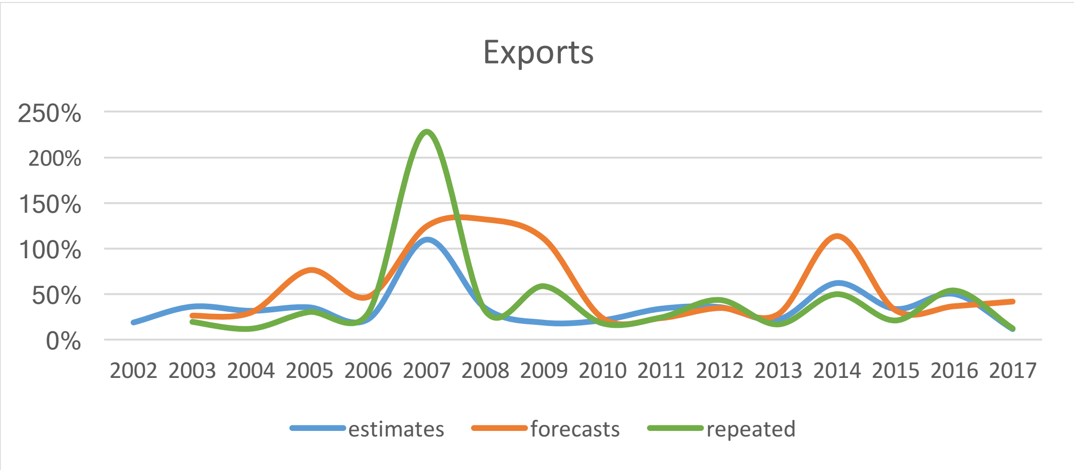

Figure 10. Progression of estimates, forecasts and repeated in exports between in 2002 and 2017.

As was expected, there are clear differences between product flows. Production and removals are more accurate than imports and exports. Kandilov (2007) points out, that exchange rates do have an effect on trade. This means that problems with exchange rates would have bigger impact on exports and imports, rather than production and removals. As discussed earlier in the methods on producing forecasts, they are demand-driven (Natural Resources Institute Finland 2019). So, at first, there has to be demand, which is fulfilled by exports and only after that there is more production. Therefore, exports would be hit first if demand is lowered. In this chain of demand-driven interest, importing raw material comes after production and this could have an effect on why in imports the spike is one year after exports.

4.8 Comparison of product flows and products together

This study has already proved that there are clear differences between product flows and products but how does this translate into one single product? This segment will go through each product and compare how accuracy of predictions are affected by different product flows. Only 8 forest products are included in this segment, as logs (products 1.2.1.C, 1.2.1.NC, 1.2.2.C and 1.2.2.NC) are only measured in one product flow: removals. Measured product flows are exports, imports and production. Some products are lacking numbers from different countries, especially in production. This is logical, as certain products are not produced in every country. In order to have a better understanding about the accuracy of each product, there are also values without the lowest 5 countries in that specific prediction, product and product flow. Chapter 4.7 outlined that production is more accurate than imports and exports, but are there any exceptions to this between different products?

The most accurate estimate was with paper and paperboard and below in Table 19 is paper and paperboard with each product flow separated. This highlights an interesting fact, as countries with lowest accuracy are taken away. It seems, that with exports, there are some countries that have very high error percentage, but as it is taken away, exports are similar to imports. Production has the lowest error rate and there isn't big change when ignoring 5 countries with the lowest accuracy. With exports and imports repeated data is more accurate than estimates but estimates are more accurate with production.

Table 18. Paper and paperboard with each product flow separated.

Product 10	Estimate		Repeated		Forecast	
Paper and paperboard	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	80.9%	9.1%	19.8%	8.4%	193.7%	12.1%
Imports	10.7%	8.9%	8.7%	7.4%	13.7%	12.0%
Production	5.9%	4.7%	6.6%	5.4%	9.3%	7.5%

All three product flows are in similar structure in estimates, forecasts and repeated data are very close for coniferous sawn wood (Table 20). In each production flow, production is over 50% smaller than imports or exports. Estimates and repeated are also similar to each other, but forecasts are noticeable worse. Compared to paper and paperboard, there are no big massive drops in accuracy when five countries with lowest accuracy are dropped off, but still noticeable drop.

Table 19. Coniferous sawn wood with each product flow separated.

Product 5.C	Estimate		Repeated		Forecast	
Coniferous sawn wood	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	19.4%	13.8%	22.6%	13.1%	31.0%	20.0%
Imports	20.1%	13.1%	20.1%	14.0%	28.6%	20.9%
Production	9.5%	6.9%	9.4%	7.6%	14.0%	10.8%

Separated product flows in particle board are relatively similar to the previous table with coniferous sawn wood. Error rates of particle board are presented in Table 21 beneath. Both have better than average error rates and ignoring five countries with lowest accuracy does not improve the error rate all that much. What is interesting is that estimates are clearly more accurate than repeated data in all three product flows.

Table 20. Particle board (including OSB) with each product flow separated.

Product 6.3	Estimate		Repeated		Forecast	
Particle board (incl. OSB)	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	18.5%	14.3%	23.7%	16.3%	31.6%	22.3%
Imports	18.5%	14.2%	17.8%	14.5%	26.0%	21.4%
Production	10.9%	8.8%	23.7%	16.3%	18.5%	14.9%

Exports on fibreboard seems to be a difficult product to predict, as presented below in Table 22. In all three sections, error rates on exports are high. This error rate is fixed when five countries with the lowest accuracy are ignored but remain clearly higher than in products presented above. Repeated data is better than other predictions, both with and without ignoring five countries with lowest accuracy.

Table 21. Fibreboard with each product flow separated.

Product 6.4	Estimate		Repeated		Forecast	
Fibreboard	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	41.4%	18.0%	52.8%	15.3%	88.6%	24.4%
Imports	18.3%	15.7%	16.0%	13.9%	25.3%	20.8%
Production	16.0%	12.8%	14.3%	11.9%	22.0%	18.1%

Plywood is similar product to fibreboard and this also shows in the results. Below Table 23 display results on how accurate predictions on plywood are doing, with product flows separated. The important thing to realize is, that there are at least 5 countries, that don't produce plywood and therefore ignoring five countries with lowest accuracy doesn't prove the error rate. This could also be the reason why production seems to be hardest to predict, when comparing averages to imports.

Table 22. Plywood with each product flow separated.

Product 6.2	Estimate		Repeated		Forecast	
Plywood	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	101.7%	20.2%	31.5%	18.6%	86.1%	27.2%
Imports	21.4%	15.8%	20.7%	16.7%	31.8%	25.4%
Production	24.2%	24.2%	39.8%	18.8%	75.0%	75.0%

All product flows for non-coniferous sawn wood seems to be hard to predict, as is demonstrated in Table 24. Estimates and repeated data have similar error rates and forecasts are quite close at production. It has clearly lower accuracy than coniferous sawn wood, although it is uncertain why. Ignoring five countries with lowest accuracy does improve the results significantly but at this point, it is unclear if volume is the only reason for above average error rate.

Table 23. Non-coniferous sawn wood with each product flow separated.

Product 5.C	Estimate		Repeated		Forecast	
Non-coniferous sawn wood	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	53.0%	16.9%	86.2%	30.2%	25.7%	18.5%
Imports	24.9%	18.2%	37.0%	28.6%	23.3%	18.6%
Production	23.3%	11.4%	37.7%	16.6%	20.5%	10.2%

Next, there is wood pulp. Table 25 below shows that wood pulp has the lowest accuracy in exports in all three categories. Average error rates are by far the worst and even when five countries with lowest accuracy are ignored, the situation doesn't turn into anything particularly great. Production is the most accurate product flow out of three, while still worse than average, it is not that bad. Estimates and repeated data are fairly close to each other.

Table 24. Wood pulp with each product flow separated.

Product 7	Estimate		Repeated		Forecast	
Wood pulp	Average of all countries	W Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	170.7%	59.3%	202.8%	40.8%	277.3%	52.1%
Imports	42.5%	14.7%	30.1%	12.8%	65.7%	20.2%
Production	18.1%	14.0%	17.0%	14.0%	26.7%	19.1%

The last individual product observed more closely is OSB, results are below in Table 26. Part from exports in wood pulp, OSB is the product with lowest accuracy. Exports are the product flow with lowest accuracy and imports are in middle of exports and production, which matches all the products. OSB has the same situation as plywood above, where OSB is not produced in every UNECE member state and therefore, ignoring five countries doesn't improve the production numbers.

Table 25. OSB with each product flow separated.

Product 6.3.1	Estimate		Repeated		Forecast	
OSB	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries	Average of all countries	Without 5 least accurate countries
Exports	102.4%	52.1%	79.1%	42.7%	104.7%	69.8%
Imports	59.2%	23.5%	31.9%	22.4%	73.0%	31.7%
Production	17.9%	17.9%	21.3%	21.3%	24.4%	24.4%

Overall, separating all the product flows did not make a big difference. As already discovered, production is more accurate product flow compared to exports or imports and this was proven in the tables above. Paper and paperboard are the most accurate product, apart from exports. Even the exports part is arguable, as ignoring five countries with lowest accuracy makes paper and paperboard the most accurate product. Even without ignoring five countries, average error rate of 5.9% for estimates on production across all the UNECE member States is really impressive. Production numbers for coniferous sawn wood was the only product to come even close to that.

5. DISCUSSION AND CONCLUSION

Representatives of countries were asked which dimensions out of the most used in data quality literature. Three most important dimensions based on the answers were consistency, believability and relevancy. These dimensions were followed by completeness, accuracy and timeliness. Rankings between believability and accuracy were relatively close. The least important dimensions were accessibility, interpretability and appropriate amount of data. Overall, five out of nine dimensions were ranked the most important by at least one answer. These results mean that people who work with forest product predictions value consistent data, that is believable, completed and relevant.

It does seem, that volumes of forest product in member state does matter when producing predictions. Member states with bigger volumes were, on average, better than member States with smaller volumes. As discussed earlier, there are exceptions to this theory, but overall there is strong indication for this. When looking 15 most accurate countries based on estimates, 12 of them are also in the list of 15 biggest countries based on average product volume. It comes down to numbers, where volume makes changes smaller: drop of 2 from 200 to 198 is meaningless, just 1%. If the drop of 2 is from 2.5, it is down by 80%. Therefore, bigger volume provides possibility to have small changes and it won't show as clearly as in countries with smaller volumes.

There doesn't seem to be any differences in which category of forecast country would go based on the size of the country. All four categories were evenly filled with countries from different categories. There were member States from all three categories, sorted by size of product flows, in each group. These groups were based on how different results of member States looked like. Size of country didn't have noticeable difference on how forecasts are produced and all countries that took part in the interviews used similar methods and tools when producing forecasts. These methods included consulting experts across the industry for information about trends in the market and potential investments. As expected, preliminary data from Q1 and Q2 is used when creating predictions, since they are available when predictions are produced during September.

As for future forest product predictions, there are no easy ways to guarantee complete accurate predictions. As just presented, many countries use similar methods when producing predictions. Using data available, discussing experts around the industry and using experience from previous years allows predictions to be as close to the reality as possible. Going back and checking how well the previous year has turned out gives a good perspective if those predictions were as accurate as hoped. Each UNECE member state has their own specialties as forest industry produces, exports and imports different volumes in each country and this should be taken into consideration when reading the results. Data quality of UNECE member States should improve and I believe they are in a good position to do so, due to the close co-operation, exchange of information and guidance provided by member States.

In this research, four different product flows were measured and analysed: production, imports, exports and removals. Out of those product flows production has the highest accuracy, which could be traced back to predictability and planning that goes into production. Removals had the second highest accuracy, which indicates that in order of having production going as planned, removals of wood are also planned ahead. This planning becomes more difficult after production is done, as the accuracy of imports indicates. Exports are even less accurate than imports and both are vulnerable to outside effects. This could be seen during financial crisis, where accuracy of exports and imports were less than ideal.

Out of the 12 products analysed, paper and paperboard have the highest accuracy. All three predictions in production are great and really good in imports, only slightest drop in accuracy is in exports. Paper and paperboard are followed by coniferous saw logs and veneer logs, as well as coniferous sawn wood. All three products have high volumes and have very few countries with low accuracy predictions. Two products with the lowest accuracy were wood pulp and OSB, both with around 50% error rate on average. It came down to all countries struggling to at least some length and some having really low accuracy.

6. LIMITATIONS

Comparison between different countries and division into groups was done with visual analysis. There is a clear possibility of human error with this approach, as something important could be overlooked or go unnoticed. Visual analysis was chosen due to the large number of data. With visual analysis, it was possible to go through many graphs in a reasonable amount of time. Number behind of graphs had already been valued during the mathematical part of analysis and they were estimated to be correct. Using visual analysis will provide a possibility for new research, where different approaches can be used.

Time series analysis is another approach, which could provide new information about data used in this research. Due lack of time and prior knowledge of the time series analysis, it was not used in this research. It would, however, had provided more information about changes happening year-on-year basis. It would be highly recommended method for future research on the same subject,

A significant problem with this research is with the data. Historical data used in this research had been updated later, as new information about it was known. This means that comparing predictions to data that have been updated five years afterwards, makes them look worse than they would have been during the time when they would have been used. The bigger problem comes with repeated data and the fact that it was created using data that have been updated afterwards. In this analysis the repeated data looks more accurate than it would, if un-updated data would have been used and it creates unfair comparison to other predictions. If this research is re-done, un-revised historical data (i.e. the data that were available at the time of the questionnaire) should be used to create repeated data. It might also be interesting to compare a forecast (two years ahead) with data repeated for two years ahead.

7. SOURCES

Bose, R. 2009. Advanced analytics: opportunities and challenges. Industrial Management & Data Systems. vol. 109, no. 2, pp. 155-172.

Bovee, M., Srivastava, R.P. & Mak, B. 2003. A conceptual framework and belief-function approach to assessing overall information quality. International Journal of Intelligent Systems. vol. 18, no. 1, pp. 51-74.

Cambridge dictionary 2019. Definition of data. Accessed on 7.2.2019. Available on: https://dictionary.cambridge.org/dictionary/english/data

Chen, Y., Zhu, F. & Lee, J. 2013. Data quality evaluation and improvement for prognostic modeling using visual assessment based data partitioning method. Computers in Industry. vol. 64, no. 3, pp. 214-225.

Garcia and Mendonca 2004. Visual Analysis of Data from Empirical Studies. Accessed on 27.2.2019. Available at: https://www.researchgate.net/profile/Maria_Cristina_Oliveira4/publication/228890391_Visual_analysis_of_data_from_empirical_studies/links/53cfb9200cf25dc05cfb14b3/Visual-analysis-of-data-from-empirical-studies.pdf

Huang, H., Stvilia, B., Jörgensen, C. & Bass, H.W. 2012. Prioritization of data quality dimensions and skills requirements in genome annotation work. Journal of the American Society for Information Science & Technology. vol. 63, no. 1, pp. 195-207.

Juran, J.M. 1989. Juran on Leadership for Quality: An Executive Handbook. New York: The Free Press.

Kallio et al. 2018. On the Reliability of International Forest Sector Statistics: Problems and Needs for Improvements.

Kovac, R., Lee, Y.W. & Pipino, L. 1997. Total Data Quality Management: The Case of IRI. IQ, pp. 63.

Liu, L., & Chi, L. 2002. Evolutional Data Quality: A Theory-Specific View. IQ. pp. 292-304.

Natural Resources Institute Finland, 2019. Forest Sector Short-Term Foresight Market Research in Luke. Available at: http://jukuri.luke.fi/handle/10024/542827

Rantala M 2016. Data quality analysis in industrial maintenance; Theory vs. Reality. Lappeenranta University of Technology.

Scopus 2018. Citations of "Beyond Accuracy: what data quality means to data consumers". Accessed on 5.11.2018. Available at: https://www.scopus.com/record/display.uri?eid=2- s2.0-0002918344&origin=resultslist&sort=r f&src=s&st1=.+Beyond+Accuracy%3a+What+Data+Quality+Means+to+Data+Cons umers.&st2=&sid=5018f040baec873a164eb2ad7548fbd6&sot=b&sdt=b&sl=76&s=TITLE-ABS- KEY%28.+Beyond+Accuracy%3a+What+Data+Quality+Means+to+Data+Consume rs.%29&relpos=0&citeCnt=1782&searchTerm

UNECE, 2010. Forest products annual market review 2009-2010. Available at: http://www.unece.org/fileadmin/DAM/timber/publications/sp-25.pdf

UNECE 2019a. About UNECE: Mission. Accessed on 28.04.2019. Available at: http://www.unece.org/mission.html

UNECE 2019b. Sixth Meeting of the ECE/FAO Team of Specialists on Forest Products Statistics. Accessed on 10.05.2019. Available at: https://www.unece.org/index.php?id=50820

Wang, R.W. & Strong, D.M. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. vol. 12, no. 4, pp. 5-33.

World Bank, 2019a. Economic growth in 2006. Accessed 8.2.2019, available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2017&name_desc=false&start=1961&type=points&view=chart&year=2017

World Bank, 2019b. Economic growth in European Union 2002-2017. Accessed in 19.2.2019, available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?end=2017&locations=EU&start=2002