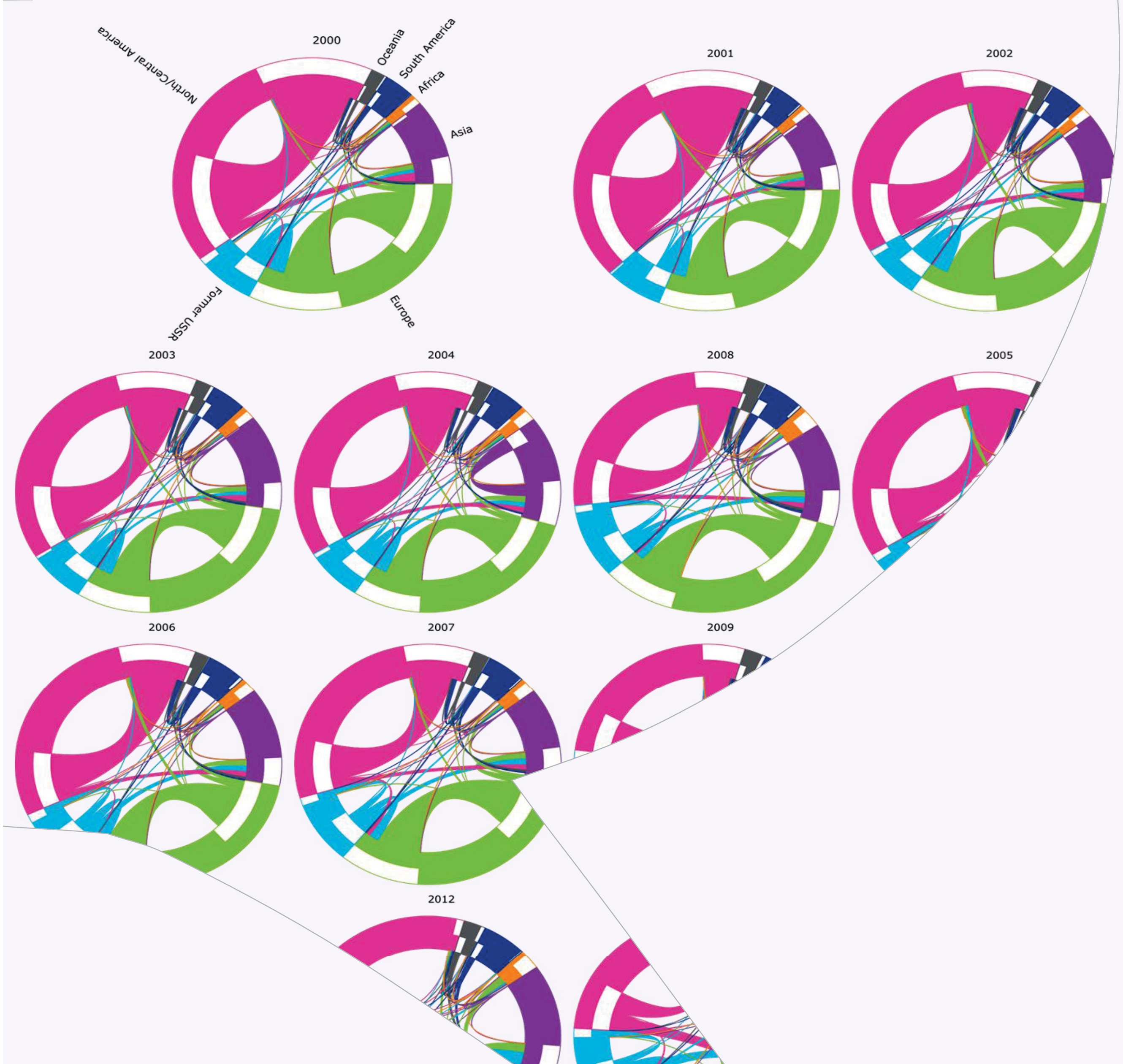

Natural resources and
bioeconomy
studies 74/2016

Model for international trade of sawnwood using machine learning models

Tuomas Nummelin and Riitta Hänninen

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Natural resources and bioeconomy studies 74/2016

Model for international trade of sawnwood using machine learning models

Tuomas Nummelin and Riitta Hänninen

Natural Resources Institute Finland, Helsinki 2016

ISBN: 978-952-326-341-3 (Print)
ISBN: 978-952-326-342-0 (Online)
ISSN 2342-7647 (Print)
ISSN 2342-7639 (Online)
URN: http://urn.fi/URN:ISBN:978-952-326-342-0

Authors: Tuomas Nummelin and Riitta Hänninen
Publisher: Natural Resources Institute Finland (Luke), Helsinki 2016
Year of publication: 2016
Cover photo: Tuomas Nummelin
Printing house and publishing sales: Juvenes Print, http://luke.juvenesprint.fi

Abstract

Tuomas Nummelin and Riitta Hänninen

Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland

The study tests the potential of machine learning models to analyse and forecast global bilateral trade flows of soft sawnwood by countries. The empirical trade flow data including annual import and export quantities and prices of soft sawnwood from 2000 to 2014 is obtained from FAOStat. We compare forecasting results from three methods, which can be classified as machine learning models: support vector machines (SVM), Neural networks and Random forests that is an ensemble of decision trees.

The significant changes in the global sawnwood markets in the North America, Asia and Western Europe after the financial and economic crises of 2008-2009 raise the question to update the modelling of trade flows. The information on the global trade flow developments are important for decision makers involved in strategic planning and forecasting at the European Union-, national- or industry level. The aim of this study is to test methods previously quite rarely applied in the forest sector market modelling, but which could be helpful for analysts to visualize and examine a large amount of bilateral trade data to have a view on ongoing changes and to assess next years' developments. The results also support The Finnish Forest Sector Economic Outlook, which are published biannually by the Natural Resources Institute Finland (Luke). As an example we present empirical results for Finland and for some of its main competitor countries and export destinations.

Keywords: coniferous sawnwood, international trade, predictive modelling, machine learning

Contents

1. Introduction

Sawnwood industry and the wood products sector as a whole is an important source of income and employment especially in the regional economies in many of the European Union countries, like Finland. From the Finnish forest industries' export income, wood products industry accounted for about 20 percent in 2015 (Luke database). The sector has also a key role in filling the EU and Finnish national climate targets. It is significant in mobilizing wood resource flows from forests also to the other forest based industries as well as to bioenergy production. The by-product flows from sawmills to the other industries, like pulp, paper, carton board and bioenergy are important. Consequently, these industries are affected by the sawnwood market developments.

Previous international and domestic research focusing on sawnwood market developments, demand and international trade flows are relatively scarce. From the few studies between 2000 and 2016 can be mentioned e.g. Simanusong & Buongiorno 2001, Kangas & Niskanen 2003, Mutanen 2006, Hänninen et al 2007, Jonsson 2010, Hietala et al. 2013, Hurmekoski et al. 2015, and Buongiorno 2016. The UNECE's Forest Sector Outlook studies for different global regions, are examples on the long-term forecasts at a larger scale including demand for sawnwood and other forest industry products (e.g., UNECE/FAO EFSOSII for Europe).

Sawnwood markets and business environment of the traditional forest industry countries of the boreal zone has changed in many ways in the last decade and changes will continue in future. Among important changes have been increasing supply of sawnwood from Eastern European countries with low production costs, the introduction of restrictions to roundwood availability imposed through forest conservation programs (e.g., Hänninen and Kallio 2007) and increased Russian export tariffs on roundwood (Solberg et al. 2010). A large change in the European markets was the founding of the European Monetary Union (EMU) in 1999. At that time, it was publicly seen, that the currency union will have mostly positive effects in the participating countries, especially from elimination of the speculation related to currencies. However, also speculation was directed at whether this could cause severe adaptation problems to, for example, Finnish forest industry firms for which exchange rate devaluations have traditionally had an important role.

The global financial crisis caused a significant drop in global sawnwood consumption in 2009, which has had important effects on global sawnwood trade flows. In North America and Western Europe, the recovery of economic growth has remained slow after the crisis, and the activity in housing construction, that is the main end-use sector of sawnwood, is has been recovering relatively slowly in Europe and North America. As a result the main European producers and exporters, Finland, Sweden, Germany and Russia have directed their export supply to North Africa and Middle East as well as to Asia. In Asia Japan is an important export area and the role of China has grown in importance as an export destination to the European and North American sawnwood producers during the last few years.

The past changes in the sawnwood markets increase the importance to continuously update the modelling of the markets and to search new applicable tools to help to detect the ongoing development and possible signs of shifts in bilateral trade flows that could be signals for future changes in market structures. In the present study, our focus is to test a tool to analyse and predict the markets at short term, 1-2 years ahead. The results also support the Finnish Forest Sector Outlook Studies (e.g. 2016), which are produced biannually by the Natural Resources Institute Finland (Luke).

Potential of machine learning models to analyse and predict annual trade flows of soft sawnwood between countries in a relatively simple and still rigorous way was examined in the study. The aim was to test the approach using a complete dataset of bilateral trade statistics obtained from the FAOStat database. Three methods were tested, that can be classified as machine learning models: support vector machines (SVM), Neural networks and Random forests, which is an ensemble of decision trees. The models were tested by the prediction power for 2 years ahead.

2. Previous studies

The previous multi country studies on forest sector markets, international trade, cross country comparisons and consumption have commonly applied econometric modelling and panel estimation (e.g. Simanusong & Buongiorno 2001, Hänninen et al 2007, Hurmekoski et al. 2015, Buongiorno 2016). The studies of Buongiorno (2016) and Kangas and Niskanen (2003) can be mentioned examples from gravity model applications. Partial equilibrium models have largely been applied in forest sector studies to formulate long term scenarios and to analyse policy effects (e.g. Kallio et al. 2006, Bolkesjo et al. 2006). Global forest products market model GFPM (see Buongiorno et al. 2003) was applied e.g., by Turner et al. (2005) to analyse the US policy effects on international forest products trade flows and Buongiorno (2015) to predict global timber production and forest products prices. Chang and Gaston (2014) applied dynamic spatial equilibrium model to analyse and forecast aggregated trade flows of Canadian sawnwood. Disaggregation by species was made by using a proxy of price. Previous studies are dominated by the modelling of historical developments and presenting long-term scenarios. Short-term approach has been scarce, and from the studies on softwood lumber trade can be mentioned Buongiorno et al., 1984 and Hetemäki et al. (2004).

It is known that economic modelling and forecasting always includes uncertainties, and forecast failures are more the rule than the exception. For forecasting, there are many methods available, guessing, extrapolation, time-series and econometric models, large spatial equilibrium models, agent based computational economics, machine learning models, etc. The purpose of the analyses, the resources available for it and the practical applicability determine the kind of models that are applied. In economic analyses and forecasting, econometric and time-series models are important tools. Basing on the economic theory they present how the economy functions, they can also provide forecasts and policy advice. However, it is also true that econometric models suffer from forecasting failures. They may be misspecified, and the economy may have been subject to structural changes, that have been the case also in the sawnwood markets during the last decade or so.

Buongiorno (1996) discusses the difficulties related to the different model types. Difficulties in the econometric modelling are related to structural changes and statistical problems in parameter estimation due to e.g. data quality and quantity. Large forest sector models are not necessarily better in managing with structural changes, and they suffer from high complexity and a need for number of assumptions and estimated elasticities.

Traditionally, the main assumptions in modelling and forecasting are, that 1) the model represents the economy well, and 2) the structure of the economy will remain relatively unchanged during the forecast period. Recent research, however, suggest a loosening of these strict assumptions. It is admitted that models are incorrect in unknown ways, that the economy is complicated and changes over time, and that the data used in estimation is inaccurate (Hendry and Ericsson, 2003). In short term approach, economic theory may not always be very helpful, and models not based on the economic theory may perform better than the theory-based models. Also pooling of forecasts may be useful because biases are averaged (Hendry and Clements, 2003). So, with the increasing importance of short-term forecasts for forest sector decision-making, an important future challenge for the research is to develop tools for it by formulating new models and adapting new techniques to the forest sector. However, when applying forecasts in policy making, it must be strongly stressed, that it is important to understand forecasting uncertainty (e.g. Hänninen 2004). The use of models always needs additional expert knowledge in formulating future views.

The present study tests the machine learning models to analyse and predict annual bilateral trade flows of soft sawnwood between countries in a relatively simple way. Our emphasis is in practical analysis, visualization of trade flow data and testing short-term forecasting methods rather than in a theoretically well specified model or long term predictions. In modelling, we apply variables commonly used in previous studies of forest sector trade flows.

2.1. Global demand for coniferous sawnwood

In 2014, about 300 million cubic metres of coniferous sawwood was consumed globally (Faostat). North America is the largest consumption region, followed by Europe, Asia, former USSR, South America, Africa and Oceania (see Figure 1). At the regional level, Asia and Africa are clearly net importers of coniferous sawnwood. In Africa, domestic production covered about 20 per cent of apparent consumption and in Asia the respective percentage was about 60 in 2014 (Faostat). Instead, Europe is net exporter, and the share of the apparent consumption from domestic production has increased significantly during the 2000's. In 2014 it was 140 per cent from domestic production. For the European sawnwood producers this development has meant tightening competition for markets shares.

Changes in the global sawnwood consumption have been significant during the research period in North America, Asia and Western Europe. The effects of the global financial and economic crises led to the level of coniferous sawnwood demand falling in 2008-2009. In North America and Western Europe, the economic growth has remained low after 2009, and consequently the recovery of housing construction, that is the main end-use sector of sawnwood, has been slow in many countries. From the end use of wood products about 70 per cent is related to construction (Hänninen et al. 2007b), which includes different items, such as structural and non-structural frames, interior and outdoor products, window frames and doors, floors, roof trusses, etc. At least in Europe, sawnwood markets can be characterized as path dependent. The building practises has been traditional in construction for a long time, and possible changes in consumption structures depend on how new consumption patterns develop (Hurmekoski 2015).

In housing construction, the largest end-use sector of coniferous sawnwood is the single-family homes, the construction of which has suffered most from the economic recession. Wood is used also for wooden high-rise construction in North America and Europe, and for example in Finland one important goal of the Finnish Bioeconomy Strategy (2014) is to promote the use of wood in high-rise construction. Other end-uses of coniferous sawnwood are in furniture-, wood packaging and pallet industry. In addition, the use of substitutes, such as wood based panels affect the demand for coniferous sawnwood in certain end-uses.

A small drop of consumption was also seen in Asia in 2008, where Japan, that is a large importer of sawnwood, has suffered from a stagnant development of housing construction. After 2009 up to 2014 consumption growth has been high in Asia, which can be detected from Figure 1 (lilac inner circles). This is for the most part due to the demand increase in China, which imports coniferous sawnwood especially for its furniture industry. Also Africa shows a continuing growing trend in consumption (orange inner circles in Figure), which comes from increasing demand in certain Northern African countries, especially Egypt.

Changes in regional demand lead changes in trade flows, which in principle, also reflect competition ability between the exporter regions and countries. In 2000 trade flows of sawnwood totalled 118 million m3, total value of trade was 23,4 billion USD (2010 USD) and in 2014 103 million m3, total value of trade was 26.3 billion USD (2010 USD). The average price of sawnwood was 200 USD/m3 (2010 USD) in 2000 and 255 USD/m3 (2010 USD) in 2014.

2.2. Regional and country specific trade flows

According to Figure 1, the distribution of trade between regions has changed a lot during the period of 2000 - 2014. The most evident features are reductions in the North American and European trade shares in the total global trade, and increments especially in Asian, the former Russian and African shares. In North America, internal trade has been decreasing during 2000-2014; imports from Canada dropped to the USA that has suffered from low construction activity after the economic recession. As a result Canada has directed sawnwood exports from the USA to Asia (widening pink flow in Figure 1), where China as a "world factory" needs growing amounts of raw material for its furniture- and other wood products industry. Figures 2 and 3 the increment in China's trade. Imports to China have grown also from Europe. It is also worth of noticing, that a large part of the sawnwood imported to China returns back to the USA and Europe in the form of further processed products, like Chinese furniture and other wood products.

Figure 1. Trade flows, consumption and production of sawnwood in 2000-2014. Inner circle white arc is exports for certain area, colour filed arc is consumption in same area. Outer circle white arc is imports for that area and colour filled arc is production.

European share of global trade has decreased during the period from 2000 to 2014, and there is also reduction in trade inside Europe. Due to the weak demand growth in European markets, exports to overseas regions, Africa and Asia, are clearly increased. The region of the former USSR has gone through large political and economic changes after the collapse of the Soviet Union in 1991, which led to a severe collapse also in sawnwood production, domestic consumption and exports. Recovering of the sector has continued during the 2000's and Russian Federation competes for market shares in Western Europe, Asia and Northern Africa (blue flows in Figure 1) with large producers, Sweden, Finland and Germany. For Russia, China has become a large single export market between 2000 and 2014 (Figures 2 and 3). From the European producer countries, Sweden and Germany have maintained their shares in global trade quite well, but the shares of Finland, France and Austria have decreased.

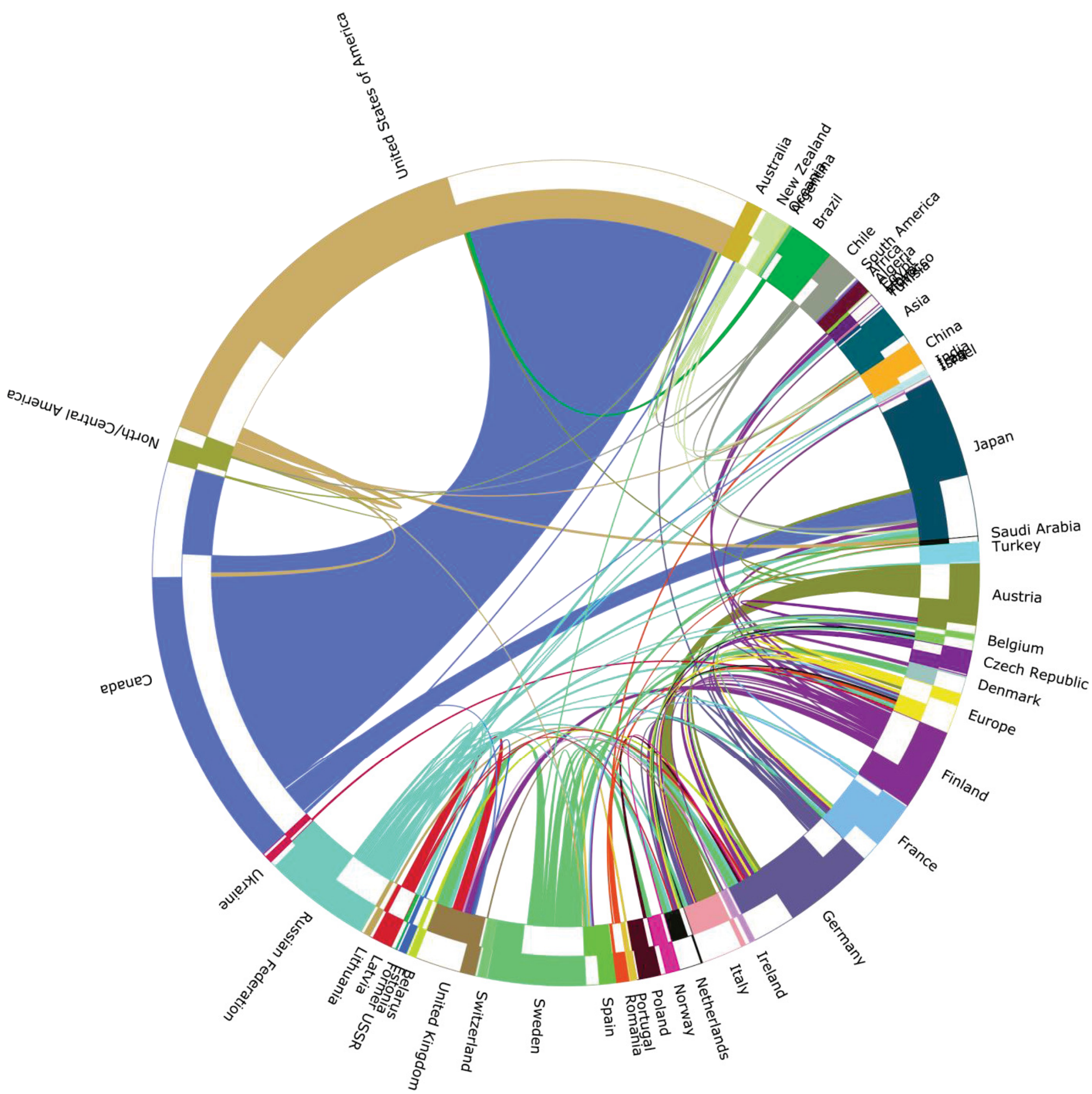

Figure 2. Sawnwood trade flows, production and consumption in 2000. Inner circle white arc is exports for certain country, colour filed arc is consumption in same country. Outer circle white arc is imports for that country and colour filled arc is production. For clarity only 10 % of largest trade flows are shown which covers 90 % of total trade flow. Note: In this picture all areas like Europe mean aggregate for the rest of not named countries in that region.

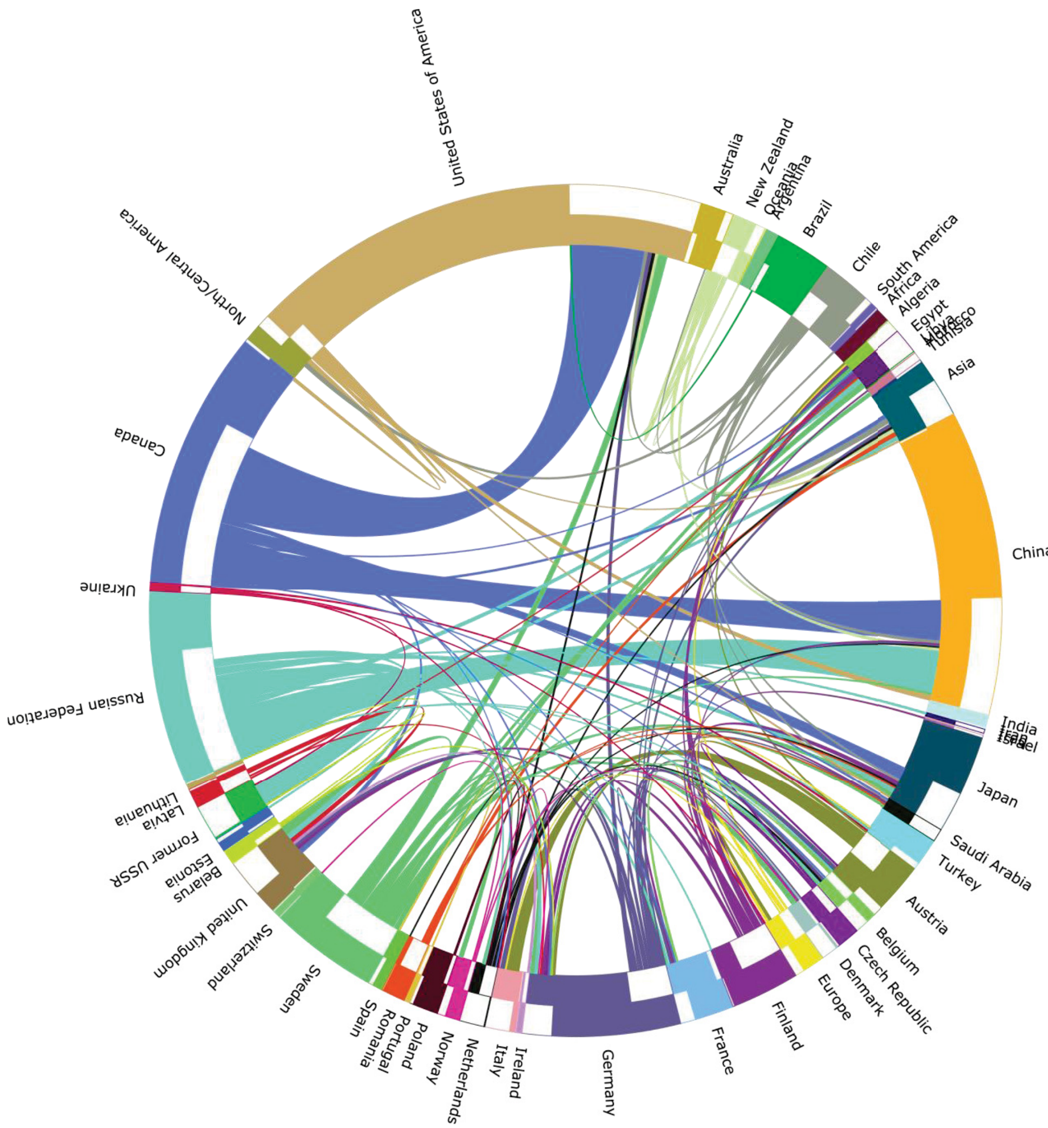

Figure 3. Sawnwood trade flows, production and consumption in 2014. Inner circle white arc is exports for certain country, colour filed arc is consumption in same country. Outer circle white arc is imports for that country and colour filled arc is production. For clarity only 10 % of largest trade flows are shown which covers 87.8 % of total trade flow. Note: In this picture all areas like Europe mean aggregate for the rest of not named countries in that region.

3. The model

Statistical models are most often applied in modelling and forecasting forest sector imports and exports. Instead, we compare three methods, which can be classified as machine learning models. A major difference between the two approaches is that machine learning techniques emphasize forecasting and statistical models emphasize inference. Machine learning models have a capability to handle large data sets and recognize implicit dependencies and relationships in the data learning to adapt their behaviour.

In modelling, we apply variables commonly used in previous studies on forest sector demand and trade flows (e.g. Kangas and Baudin 2003, Hietala et al. 2013, Hurmekoski 2015, Buongiorno 2015, 2016). Demand for forest products is commonly modelled and demand elasticities estimated (see Kangas and Baudin 2003) using time series data and the general form:

$$Consumption = f(P_t, GDP_t, C_{t-1}), \quad (1)$$

where apparent consumption of a certain product is explained by its price, gross domestic product (GDP) and consumption with one lag. It is assumed basing on the economic theory, that income growth increases consumption as well as price decline, and as a result of price increase consumption goes down. Current consumption is also dependent on last period’s development. The derivation of the demand functions in the previous studies is commonly assumed to be based on the cost minimization behaviour of a representative firm.

The basic import demand and export supply framework for a certain product were presented e.g., by Kangas and Baudin (2003):

$$Demand = f(P_d, P, X), \quad (2)$$
$$Supply = f(P_d, P_x, Z), \quad (3)$$

where the import demand (2) can be explained by the price of domestically produced product (P_d), the price of imported product (P) and a vector of demand shifters (X), e.g. gross domestic product (GDP), population, etc. It is assumed basing on the economic theory, that a rise (a drop) in the import price decreases (increases) imports, and income (GDP) growth increases imports. The price effect of domestic production (P_d) depends on the substitutability or complementarity between domestic and imported products.

The export supply (3) can be explained by the export price (P_x) of the product, the price (cost) of domestic production (Pd) and supply shifters (Z), e.g. production capacity. It is assumed that a rise (a drop) in the export price increases (decreases) exports and a rise (a drop) in unit costs decreases (increases) exports.

In the trade flow studies basing on the Armington (1969) theory, trade patterns are explained only by changes in relative prices (e.g. Hietala et al. 2003) between competitive producers or countries. The theory assumes imperfect substitution between products from different countries of origin and identical and constant substitution elasticities between the pairs of countries.

Exchange rate variables are included in the previous models (e.g. Hietala et al. 2013). Exchange rate changes affect competition in the markets. For example, results for Finland indicate that sawnwood exports to the United Kingdom have been greatly affected by the currency movements between Finnish currency and British pounds (Hänninen 1998, Hietala et al. 2013).

The availability of data as well as practical applicability of the model affected also the empirical modelling and variables chosen in the present study. In the present study, we assume that the trade of sawnwood is similar process between all the countries. The ad hoc model for bilateral trade in quantities (X_{ij}) includes the following explanatory variables in machine learning model estimation:

$$X_{ij} = f\big(Y_{ij}, GDPCap_i, GDPCap_j, Aind_i, Aind_j, i, j, Q_i, Q_j, C_i, C_j, Ix_i, Ix_j, Ex_i, Ex_j, Ix\ share_j, Ex\ share_i, P_{ij}, \mathrm{tb}, \mathrm{JPY/USD}, \mathrm{CNY/USD}, \mathrm{GPB/USD}, \mathrm{SEK/USD}, \mathrm{RUB/USD}, \mathrm{EUR/USD}\big), \quad (4)$$

The explanatory variables in the machine learning models are also depicted in Figure 4.

- The dependent variable is quantity (m^3) of bilateral trade flow (X_{ij}).
- Previous quantity of bilateral trade flow (Y_{ij}).
- The source country (exporter) (country index) i and the target country (importer) (country index) j are described with general information like geographical area (e.g. Europe) (AInd). We use these indices to describe local trade conditions, which might differ greatly in the different geographical regions.
- The general economic activity country is included in model with GDP per capita of the country (GDPCap). The economic activity of country is impacting both demand and supply. Availability of annual forecasts from different sources, e.g. International Monetary Fund (IMF) for GDP is good almost every country and it helps practical usage of the model.
- The more specific aspects of the exporter and importer are the production (Q) and consumption (C) of sawnwood, total imports (Ix) and total exports Ex) of sawnwood. These describe the specific conditions of sawnwood trade in these countries.
- The specifics of bilateral trade (partnership) are captured with export (Ex share) and import shares (Ix share) of trade flows and real unit costs (P) paid for bilateral trade flows. These describe how important specific trade is for exporter and for importer.
- The foreign exchange rates were included to depict general condition in the World economic markets. We used six exchange rates to capture information about possible competitive advantage of significant currencies in the trade of sawnwood. Due to the fact that research on currency effects in sawnwood or other forest product trade is largely missing we included the currencies of the regions (euro area) or countries (USA, Canada, Sweden, Russia, Japan and China), that are globally important sawnwood exporters or importers.
- The years between trades describes how many years have passed since the bilateral (tb). This measure is partly establishment of trade relations and common pattern of trade.

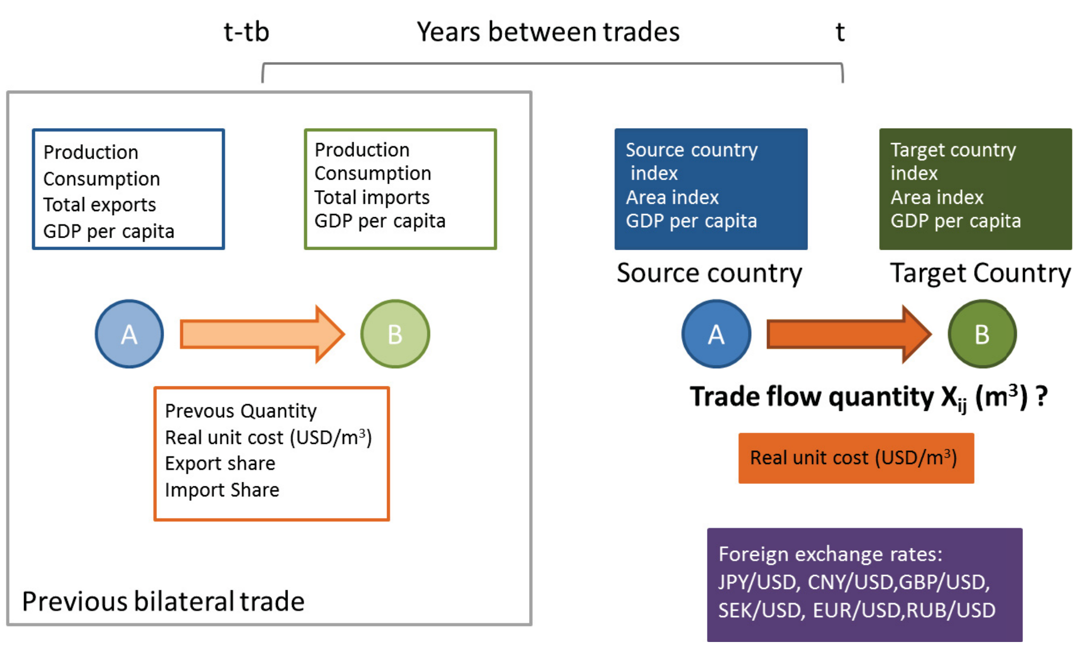

Figure 4. The idea of the model.

4. Data

The study uses annual data between the years 2000 and 2014. The data for production, exports and imports of coniferous sawnwood comes from the FAO databank. The annual apparent consumption for each country was calculated using the data for production, imports and exports. The FAO data includes Forestry Trade Flows data and, Forestry Production and Trade data (http://faostat3.fao.org/home/E). In the FAO databank sawnwood is determined as

"Wood that has been produced from both domestic and imported roundwood, either by sawing lengthways or by a profile chipping process and that exceeds 6 mm in thickness. It includes planks, beams, joists, boards, rafters, scantlings, laths, boxboards and "lumber", etc., in the following forms: unplanned, planed, end-jointed, etc. It excludes sleepers, wooden flooring, mouldings (sawnwood continuously shaped along any of its edges or faces, like tongued, grooved, rebated, V-jointed, beaded, moulded, rounded or the like) and sawnwood produced by resawing previously sawn pieces."

It must be noticed, that the commodity group contains various items with different price levels. Also the species compositions of trade flows between countries differ. This means that the average unit prices applied in the study due to the lack of better price data, do not necessarily describe very well the actual prices of the sawnwood trade between countries. The volume is reported in cubic metres solid volume. In our study, we use a list of countries from the FAO databank to model individually the sawnwood trade flows, which we are interested in, and aggregate the rest of the countries to regional aggregates.

The foreign exchange rates included in the modelling are among the major currencies used to trade forest products (UNECE 2016, p.3): JPY/USD, CNY/USD, GPB/USD, SEK/USD, RUB/USD and EUR/USD. Japan (JPY) and China (CNY) are large importers of sawnwood and Britain (GPB) is important in the European region. Russia (RUB) and Sweden (SEK) are globally large exporters. In the Euroarea, Finland and Germany are large exporters and Germany is also an important export destination, for example, for Finnish sawnwood. The exchange rates are presented in relation to the US dollar because the values of the trade flows are in US dollars in the FAO database. The statistical sources are the Federal Reserve's database about foreign exchange rates (http://www.federalreserve.gov/datadownload/). For the RUB/USD the source is the OECD database.

The economic activity of certain country is described by Gross domestic product (GDP) per capita. The original data includes GDP at market prices constant USD 2010 and total population of the country in question. The source of the data is World Bank's World's Development Indicators (http://databank.worldbank.org/data/home.aspx).

The deflator applied to deflate all the monetary valued data in the study is the United States GDP deflator with 2010 as the base year. The original FAO value data being in current dollars, we decided to use a deflator describing the development of the US dollar purchasing power. Similar decision on the deflator was also made in the study of Buongiorno (2016), who modelled global forest products trade flows.

Table 1. Data of the study.

Variable & symbols	Description	Source
Exports of sawnood	m3, between countries	http://faostat3.fao.org/home/E
Imports of sawnwood	m3, between countries	http://faostat3.fao.org/home/E
Production of sawnwood	m3, all countries	http://faostat3.fao.org/home/E
Import & export prices of sawnwood	USD/m3, import and export unit values for all countries	Calculated by dividing total values by quantities
Apparent consumption (AC)	m3, all countries	AC=production+imports-exports
Exchange rates	JPY/USD, CNY/USD, GPB/USD, SEK/USD, RUB/USD, EUR/USD	http://www.federalreserve.gov/datadownload/ for RUB/USD: the OECD database
GDP per capita	At market prices constant USD 2010, all countries	http://databank.worldbank.org/data/home.aspx
GDP deflator	The US GDP deflator, base year 2010. Used for deflating all monetary data.	http://databank.worldbank.org/data/home.aspx

We used data from 2000-2012 to train model and data from 2013-2014 were used to test models forecasting performance. The training data consisted 14949 observations and test data consisted 2410 observations.

In the following we look relations between model variables concerning the whole data. Figure 5 shows the Spearman's rank correlations between variables. We see some expected associations as if unit costs increase then trade flow quantity decreases. The strong positive correlation between quantity and previous quantity is expected and negative correlation of years between trades.

Figure 5. Spearman's rank correlation matrix for model features. Note variables are shorten to fit picture.

The correlation between foreign exchange rates is also expected, because those are highly entangled in market.

Positive correlations between target country GDP per capita and target country total imports and consumption are as expected by the economic theory. GDP growth increases consumption and total imports of sawnwood. However, with respect to the bilateral trade flows (Quantity) the respective positive correlation remains quite low.

4.1. Preprocessing the data

Next we will look what transformations are made to data prior modelling. The foreign exchange rates are calculated as yearly averages from monthly data. Also the exchange rates are converted to the same form (currency/USD). The GDP per capita is calculated using GDP at market prices in constant 2010 USD and total population.

The trade flows used in the models were combined in the following way. The FAO's trade flow data have the information about the trade between countries A and B in both directions, which means that country A reports that it is exporting a certain quantity of sawnwood with a certain total value in USD to country B. Likewise country B reports that it is importing a certain quantity of sawnwood from a certain country A with a certain total value in USD. These two flows should represent the same trade happening between countries A and B. The quantities and the total values of these bilateral flows should be quite close to each other, in general. Note, that the total value for imports should include some costs related to transportation, tariffs, etc. and quantities should be the same. However there are many cases were the numbers does not match. Therefore we decided to use the maximum quantities and values of imports and exports to represent the trade flows between countries. After this, total values in USD are deflated using the US GDP deflator and then real unit price is calculated for a trade flow. Export and import shares are calculated to each trade flow between a pair of countries for each year. The grand total imports and exports for country were also calculated.

The years between trades in bilateral pairs are calculated and the values of explanatory values in previous trade are calculated.

5. Predictive models tested - Machine learning methods

We used R statistical software in computations. We used caret package to do training and tuning for models. The following packages were also used in modelling Neuralnet, RSNNS, kernlab, qrnn, quantregForest, RandomForest. For parallel computations packages doParallel and doSNOW were used. For general data analysis and manipulation tidyverse package was used.

5.1. Models

Support vector machines (SVM) are the extension to support vector classifiers which are in sense extension of perfectly separating hyperplane classifier, see for detailed description Hastie et al. (2009) or James et al. (2013). The SVMs allow developing a non-linear decision boundary for classification using extended feature space. The computational feasibility of SVMs is made possible with a "kernel trick" which allows to calculation of certain inner product with efficient manner Cortes and Vapnik (1995). Features of feature space (explanatory variables are commonly called features in machine learning literature) are only affecting via inner product to computation of support vector classifier. We used three different kernel functions linear, polynomial and radial (see e.g. Hastie 2009). Those are common choices among SVM literature. The support vector machines can be also used in regression which was introduced by Drucker et al. (1997).

Random forests Breiman (2001) is an ensemble of decision trees. Output of random forest is a combined output of single decision tree outputs in the forest e.g. for classification majority vote of trees and for regression average of tree outputs. In random forest each tree is grown with a random subset of training data and with random subset of features (explanatory variables) to reduce overfitting and bias. Quantile regression forests Meinshausen (2006) are extensions to Random forests providing information about the full conditional distribution of response variable. In contrast of random forest's conditional mean of response variable. Quantile regression forest can be seen as a generalization of random forests.

Other machine learning method which we test and uses an ensemble of decision trees which is extreme gradient boosting with regression tree ensemble (https://xgboost.readthedocs.io/en/latest/). Where the ensemble of decision trees is grown using specific version of gradient decent optimization method. Both ensemble based approaches utilize boosting. Boosting is powerful learning idea that combining outputs of many "weak" classifiers to produce powerful "committee" or ensemble to produce prediction Hastie et. al. (2009).

Neural networks (or artificial neural networks) are large collection models and learning methods. Neural net-works try to mimic biological neural networks in idea. The three key components in neural networks are net-work topology, activation rule and learning rule. Network topology defines which variables are involved in network and how. The activation function describes how multiple inputs are combined as output. The learning rule defines how the network learns weights for variables in activation function based on the training data. In a sense neural networks take linear combinations of inputs and then model the target as non-linear function of these linear combinations Hastie et al. (2009). Neural networks can be used in both regression and classification. Quantile regression neural network Cannon (2011) is application of neural network to quantile regression.

5.2. Tuning parameters for machine learning models in training data

The tuning of parameters in machine learning models was done using 10-fold cross validation in training data set (years 2000-2012).

In the following we show the tested parameters for tuning of machine learning models.

- For support vector with linear kernel we used tested following cost parameters 0.1, 0.5, 1, 10, 100

- For support vector with radial kernel we tested grid with the cost parameters 0.5, 1, 10, 100, 1000, or 10000 and sigma parameter for radial kernel 0.001, 0.01, 0.1, 0.8 or 1.6. Totalling 30 different parameter combinations.
- For quantile regression forest for tune parameter was number of randomly selected predictors when making a cut in while building trees. The tested parameter values were 120,142,160 and number of trees was held constant 2001. Another setup was number of randomly selected predictors were 10, 20, 30,40,50,60 and number of trees was 1000.
- For random forest tune parameter was number of randomly selected predictors when making a cut in while building trees and number of trees. The tested parameter values were 20, 40, and 60 for selecting predictors and for number of trees 501, 1001, 1501.
- For neural network we test couple of different kind of structures. One tested structure was three hidden layers with following parameter grid. Number of neurons in the 1st hidden layer 12, 24, and 48, in 2nd hidden layer 5, 10, 15 and in 3rd 3,9,12. Totalling 27 different parameter combinations. Second tested structure was three hidden layers with following parameter grid. Number of neurons in the 1st hidden layer 20, 25,30, in 2nd hidden layer 20,30,40 and in 3rd 12,25,35. Totalling 27 different parameter combinations. Third tested structure was three hidden layers with following parameter grid. Number of neurons in the 1st hidden layer 50, 80,110, in 2nd hidden layer 30, 60, 90 and in 3rd 20, 40, 60. Totalling 27 different
- For extreme gradient boosting with regression trees we tested following parameters in grid : gamma(minimum loss reduction required to make a split) was held constant 0, eta (step size shrinkage used in update) was 0.3, 0.4 max depth (maximum depth of a tree) was 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, min child weight (minimum sum of instance weight needed in a child) was held constant 1, nrounds (number of boosting iterations) were 50,100,150,200,250,300,350,400,450,500, colsample by tree (subsample ratio of columns when constructing each tree) was 0.6, 0.8. The grid size was in total 400 different parameter combinations.

6. Results

6.1. Best machine learning models

The performance of machine learning models was measured by Root Mean Square Error (RMSE). The RMSE of the models varied a quite a lot, see Figure 6. The tuning of model parameters seems to enhance the performance of all models. Although the building of a good model seems to be more art than exact science. Also computational burden of models varied quite a bit ranging from couple of seconds to hours to train model (Laptop with Intel Core i7-3720 QM 2.60GHz with 16 Gb RAM and 64-bit Windows 7). This seemed little bit random that some models took unexpectedly long to train. In some cases this might be due to problems in gradient decent optimization. In practical analysis and the model updating, relatively fast training times are advantages.

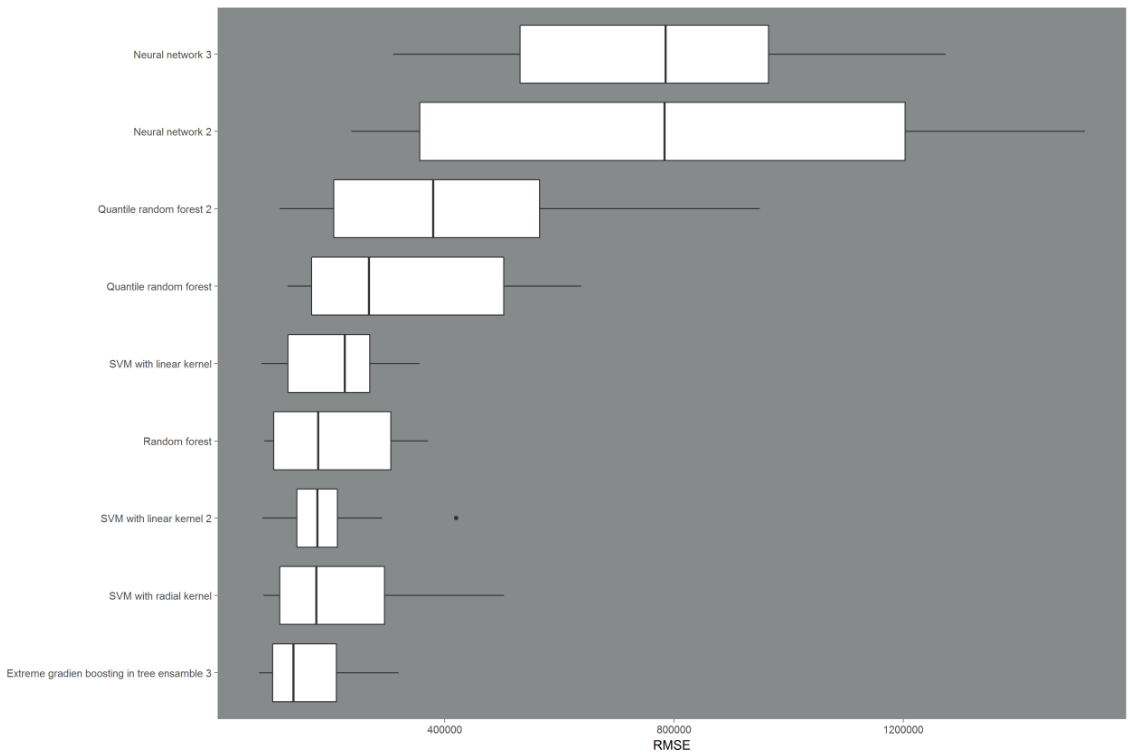

Figure 6. Resampling results for different machine learning models for 10-fold cross validation in training data 2000-2012. The vertical line in the Boxplot is median. The box extends from 25th percentile to 75th percentile and whiskers extends to 1.5 * IQR (Inter quartile range).

The performance of artificial networks (Figure 6) seems inconsistent and seems that the structure of tested networks might have not been the best possible. Support vector regressions performed quite consistently and had quite good results. The best performer in training data was model using extreme gradient boosting for regression tree ensemble.

In the test data 2013-2014, the performance of models is shown in the following table Table 2.

Table 2. RMSE for machine learning models in test data (years 2013-2014).

Model	RMSE
Neural network 3	2902354683.5
Neural network 2	537535.6
Quantile random forest 2	368252.7
Quantile random forest	330377.6
Extreme gradient boosting in tree ensemble 3	159976.0
SVM with radial kernel	158463.0
Random forest	153002.2
SVM with linear kernel 2	132514.4

The results in Table 2 show that top performing models are the same as in training. However there is slight change of order which model is the best. The magnitudes of RMSE are in line with results in training data, which indicate that models work ok.

Tables 3 and 4 show the performance of different machine learning models in respect of error (observed bilateral trade flow quantity - predicted quantity) distributions in training and test data. All the models make some quite large mistakes in both datasets.

Table 3. Summarising statistics for error distributions of models in test data.

Model	Minimum	25% percentile	median	75% percentile	Maximum	Share of negative errors	Share of positive errors
Neural network 3	-2.90e+09	-2.90e+09	-2.90e+09	-2.90e+09	-2.89e+09	1.0000	0.000
SVM with radial kernel	-2.32e+06	-6.85e+04	-2.86e+04	7.26e+03	4.43e+06	0.7110	0.289
Neural network 2	-4.00e+04	-3.16e+04	-1.56e+04	2.18e+03	1.37e+07	0.7380	0.262
SVM with linear kernel 2	-1.14e+06	-6.34e+04	-1.05e+04	3.84e+04	4.09e+06	0.5330	0.467
Quantile random forest 2	-1.17e+05	1.05e+02	1.21e+03	1.18e+04	1.08e+07	0.0589	0.941
Quantile random forest	-1.48e+05	9.90e+01	1.12e+03	1.06e+04	1.10e+07	0.0610	0.939
Extreme gradient boosting in tree ensemble 3	-5.69e+05	-1.55e+03	8.32e+02	5.77e+03	6.35e+06	0.4000	0.600
Random forest	-6.62e+05	-1.88e+03	-3.79e+02	2.04e+02	4.71e+06	0.7070	0.293

A bit surprising is that for some (e.g. Quantile random forest, Quantile random forest 2, Neural network 3) models do systemic errors in estimations i.e. the most errors have the same sign.

Table 4. Summarising statistics for error distributions of models in training data.

Model	Minimum	25% percentile	median	75% percentile	Maximum	Share of negative errors	Share of positive errors
Neural network 3	-2.90e+09	-2.90e+09	-2.90e+09	-2.90e+09	-2.85e+09	1.000	0.000
SVM with linear kernel 2	-1.44e+07	-4.42e+04	-1.77e+04	9.90e+03	7.83e+06	0.669	0.331
SVM with radial kernel	-6.02e+06	-4.56e+04	-1.37e+04	1.76e+04	1.70e+07	0.616	0.384
Neural network 2	-4.00e+04	-1.38e+04	-3.41e+03	1.39e+04	4.77e+07	0.562	0.438
Quantile random forest 2	0.00e+00	5.70e+01	7.55e+02	7.66e+03	1.75e+07	0.000	1.000
Quantile random forest	0.00e+00	5.30e+01	6.91e+02	6.89e+03	1.75e+07	0.000	1.000
Random forest	-3.45e+06	-1.69e+03	-2.67e+02	-1.42e+01	5.97e+06	0.766	0.234
Extreme gradient boosting in tree ensemble 3	-8.93e+04	-3.05e+03	-9.47e+01	2.59e+03	1.55e+05	0.511	0.489

Partly this systemic “over” estimation can be explained with the fact that individual bilateral trades has been in the training data (see figures 2 and 3) more focused on few large trade flows. The trade flows are more fragmented in the test data (2012-2014) and this leads to smaller individual quantities.

Parameters for the best tuned models were the following.

- For (Extreme gradient boosting in tree ensemble 3) extreme gradient boosting with regression trees the parameters were gamma(minimum loss reduction required to make a split) was 0 held constant while training, eta (step size shrinkage used in update) was 0.4, max depth (maximum depth of a tree) was 4, min child weight (minimum sum of instance weight needed in a child) was 1 held constant in training, nrounds (number of boosting iterations) was 400, colsample by tree (subsample ratio of columns when constructing each tree) was 0.8.
- For (SVM with radial kernel) support vector regression with radial kernel, the parameters were the cost parameter 10 and sigma parameter for radial kernel was 0.001.
- For (SVM with linear kernel 2) support vector regression with linear kernel the cost parameter was 100.
- For (Random forest) random forest parameter were 40 for number of randomly selected predictors when making a cut in while building trees and 1501 for number of trees.
- For (SVM with linear kernel) support vector regression with linear kernel the cost parameter was 0.01.
- For (Quantile random forest) quantile regression forest for tune parameter were 160 number of randomly selected predictors when making a cut in while building trees and number of trees was 2001.
- For (Quantile random forest2) quantile regression forest for tune parameter were 60 number of randomly selected predictors when making a cut in while building trees and number of trees was 1000.
- For (Neural network 2) neural network the network structure was 5 neurons in the 1st hidden layer and 0 in 2nd and 3rd hidden layers.
- For (Neural network 3) neural network the network structure was 24 neurons in the 1st hidden layer and 5 neurons in 2nd hidden layer and 12 neurons in 3rd hidden layer.

6.2. Example predicted bilateral trade flows from the models

We look how the top models performed predicting trade flows in detail in test data. We specifically look two major export markets for Finland, Germany and Japan.

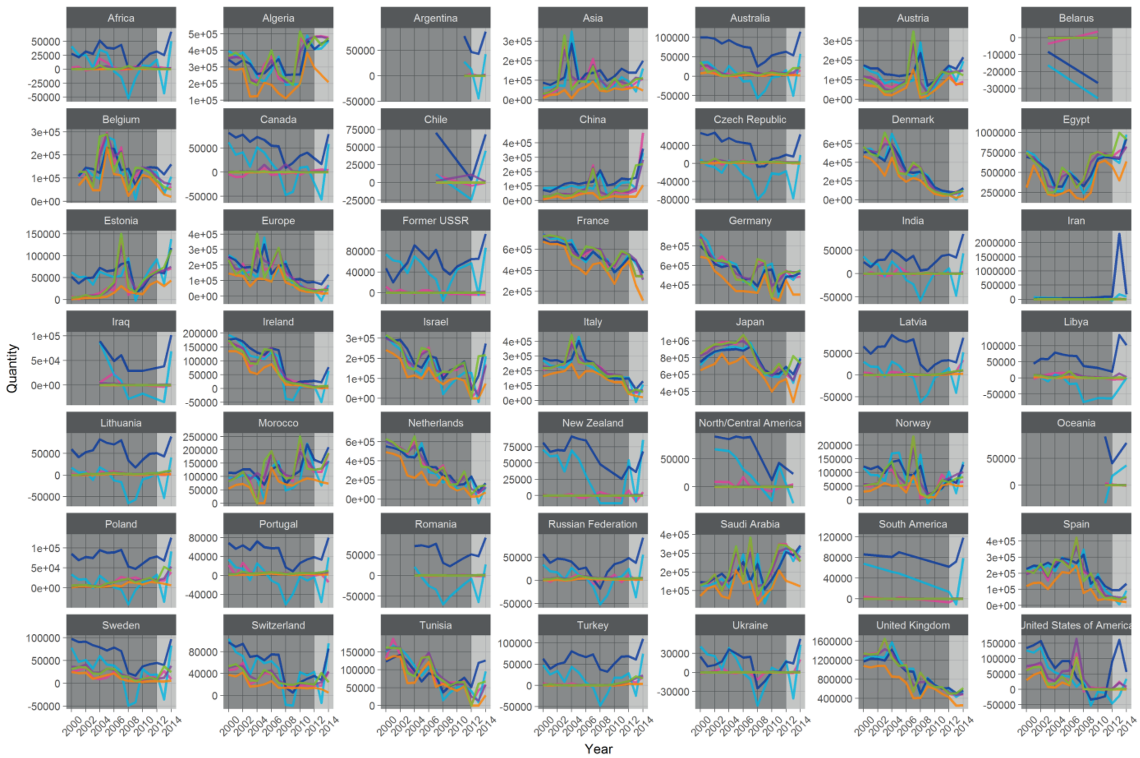

Figure 7. Finland's exports of sawnwood. Darker background is for training data. Lighter background is for test data. Green line is the actual traded quantity (m^3). Colours are for different models, pink is Extreme gradient boosting in tree ensemble 3, orange is Quantile random forest, purple is Random forest, light blue is SVM with linear kernel 2 and dark blue is SVM with radial kernel.

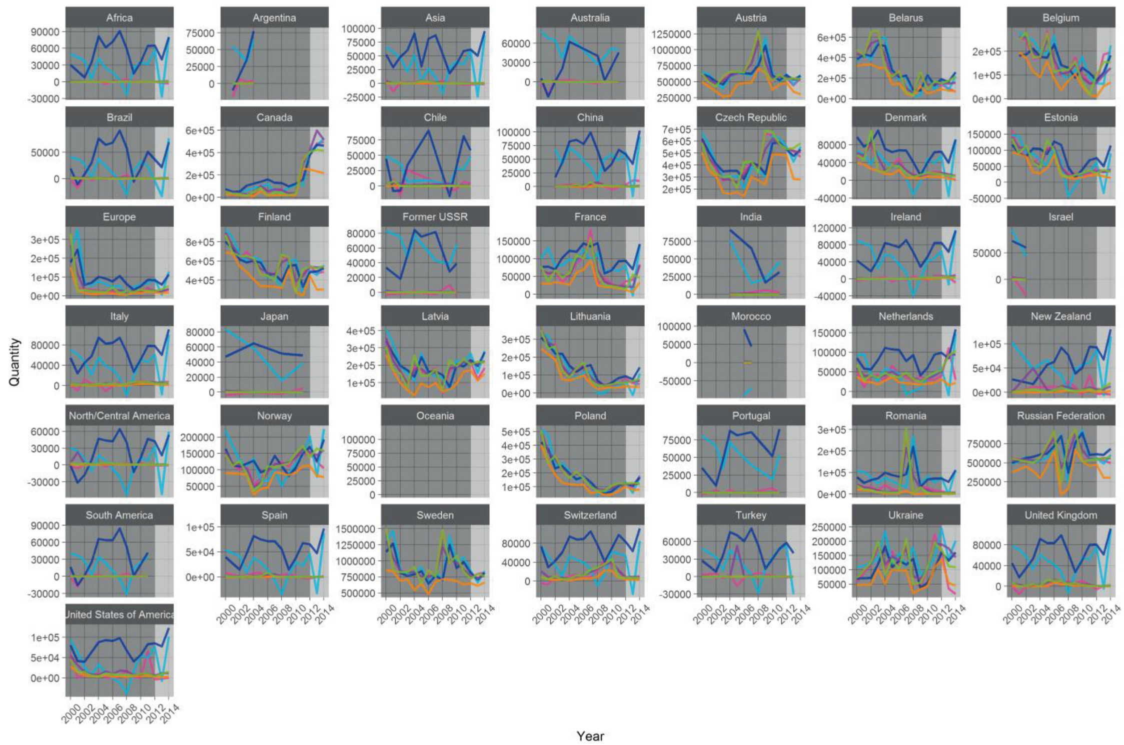

Figure 8. Germany's imports of sawnwood. Darker background is for training data. Lighter background is for test data. Green line is the actual traded quantity (m^3). Colours are for different models, pink is Extreme gradient boosting in tree ensemble 3, orange is Quantile random forest, purple is Random forest, light blue is SVM with linear kernel 2 and dark blue is SVM with radial kernel.

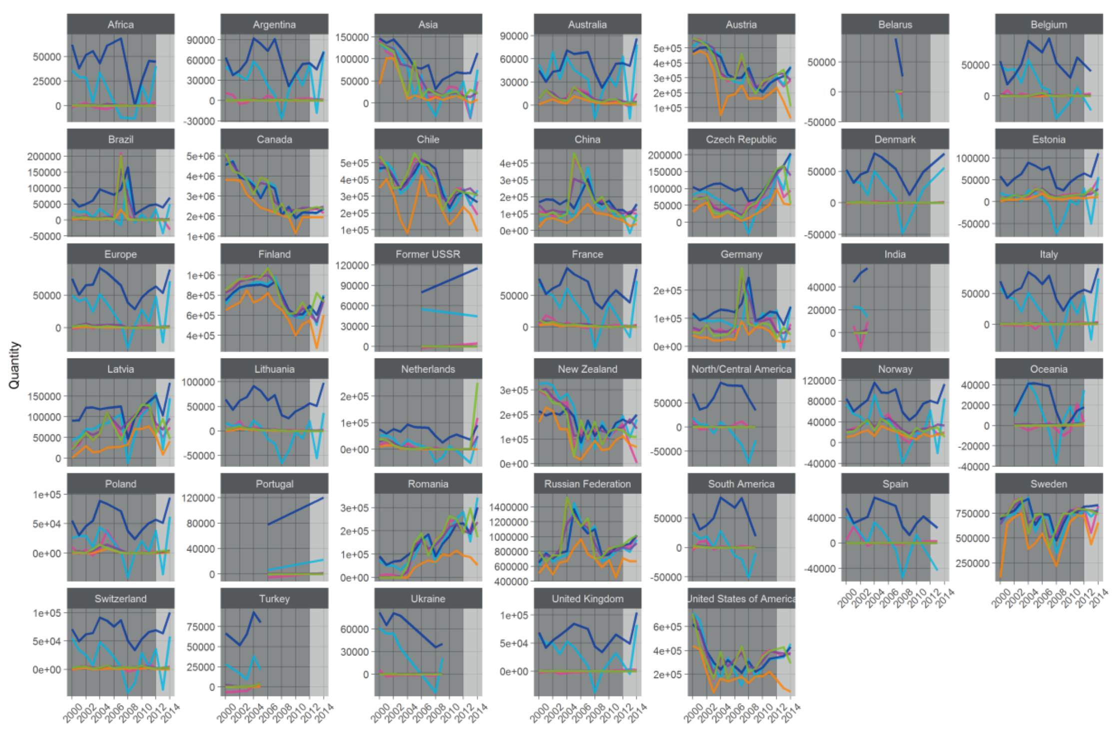

Figure 9. Figure 9 Japan's imports of sawnwood. Darker background is for training data. Lighter background is for test data. Green line is the actual traded quantity (m^3). Colours are for different models, pink is Extreme gradient boosting in tree ensemble 3, orange is Quantile random forest, purple is Random forest, light blue is SVM with linear kernel 2 and dark blue is SVM with radial kernel.

In general when we look the how well the models are able to predict bilateral trades flows all the best models seems to work reasonably. However there is clear evidence in figures 7, 8 and 9 that the all the best models lack predictive quality in some bilateral trades. This can be seen easily in predictions of Quantile random forests where some of the bilateral trades are predicted to be almost constant (orange horizontal line in figures). Also other models seem time to time miss the direction of change in trade flows. This might indicate that the explanatory variables are unable to pick this change. It can be also seen that the bilateral trade flows with more data points in training data are better predicted in test data i.e. the trade partnership has been historically strong.

6.3. Sensitivity analysis of Extreme gradient boosting in tree ensemble

The sensitivity of the Extreme gradient boosting in tree ensemble 3 for changes of certain input parameters is tested. It is good practise to test sensitivity of model. In here we test set of input parameters to see how sensitive our model is in changes in those parameters, more precise how bilateral traded quantity (dependent variable) is affected by change of certain model parameters, see table 5.

Table 5. Parameters for sensitivity analysis.

Parameter	Minimum alue	Maximum value	Test grid values
unit cost of sawnwood (USD/m^3)	0	2000	$\{(10n)_{n=0}^{25}, (50m)_{m=6}^{40}\}$
importer's GDP per capita (USD)	600	100000	$\{(1684.75n + 600)_{n=0}^{59}\}$
Years between trades (Year)	-1 (-1 meaning that new bilateral trade flow)	20	$\{(n)_{n=-1}^{20}\}$
Quantity of previous bilateral trade (m^3)	0	48 * 10^6	$\left\{\begin{matrix}(3222.11n)_{n=0}^{9}, \\ (6499.929m + 29000)_{m=0}^{14}, \\ (1408235q + 120000)_{q=0}^{34}\end{matrix}\right\}$
	0	1	$\{(0.05n)_{n=0}^{20}\}$
Exchange rate SEKUSD	5	15	$\{(0.25n + 5)_{n=0}^{40}\}$
Exchange rate EURUSD	0.5	1.5	$\{(0.0256n + 0.5)_{n=0}^{39}\}$
Exchange rate GBPUSD	0.3	1	$\{(0.0179n + 0.3)_{n=0}^{39}\}$
Exchange rate RUBUSD	20	40	$\{(0.513n + 20)_{n=0}^{39}\}$
Exchange rate CNYUSD	5	10	$\{(0.128n + 5)_{n=0}^{39}\}$
Exchange rate JPYUSD	70	140	$\{(1.795n + 70)_{n=0}^{39}\}$
Previous importer's production (m^3)	0	70 * 10^6	$\left\{\begin{matrix}(5263.158)_{n=0}^{19}, \\ (368750m + 150000)_{m=0}^{24}, \\ (4.3 * 10^6 q + 10 * 10^6)_{q=0}^{14}\end{matrix}\right\}$
Previous exporter's production (m^3)	0	70 * 10^6	$\left\{\begin{matrix}(5263.158)_{n=0}^{19}, \\ (368750m + 150000)_{m=0}^{24}, \\ (4.3 * 10^6 q + 10 * 10^6)_{q=0}^{14}\end{matrix}\right\}$
Previous importer's consumption (m^3)	0	110 * 10^6	$\left\{\begin{matrix}(89473.68)_{n=0}^{19}, \\ (750000m + 2 * 10^6)_{m=0}^{24}, \\ (5.7 * 10^6 q + 30 * 10^6)_{q=0}^{14}\end{matrix}\right\}$

All the other parameters in model were at mean values for that bilateral trade while varying parameter in interest. Example of model sensitivity calculation in in formula (5a, 5b), where unit cost sensitivity is calculated.

$$X_{ij} = f\left(\bar{Y}_{ij}, \overline{GDPCap}_i, \overline{GDPCap}_j, \overline{A\imath nd}_i, \overline{A\imath nd}_j, i, j, \bar{Q}_i, \bar{Q}_j, \bar{C}_i, \bar{C}_j, \overline{Ix}_i, \overline{Ix}_j, \overline{Ex}_i, \overline{Ex}_j, \overline{Ix\ share}_j, \overline{Ex\ share}_i, P_{ij}, \overline{\mathrm{tb}}, \overline{\mathrm{JPY/USD}}, \overline{\mathrm{CNY/USD}}, \overline{\mathrm{GPB/USD}}, \overline{\mathrm{SEK/USD}}, \overline{\mathrm{RUB/USD}}, \overline{\mathrm{EUR/USD}}\right), \qquad (5a)$$

$$P_{ij} \ni \left\{(10n)_{n=0}^{25}, (50m)_{m=6}^{40}\right\} \qquad (5b)$$

The quantity is estimated every point in given in formula (5b). This gives the estimate for quantity's sensitivity for change of unit costs.

In figure 10 shows results of sensitivity analysis, how sensitive is the predicted quantity is to change of specific model parameter. We see that a change in quantity of previous bilateral trade is the variable that model is the most sensitive. The other parameters which are clearly sensitive in model are unit costs and previous import share. On the other hand the parameters which cause least variation in quantity are exchange rates and years between trades. Next we look those three most sensitive parameters in more detailed manner.

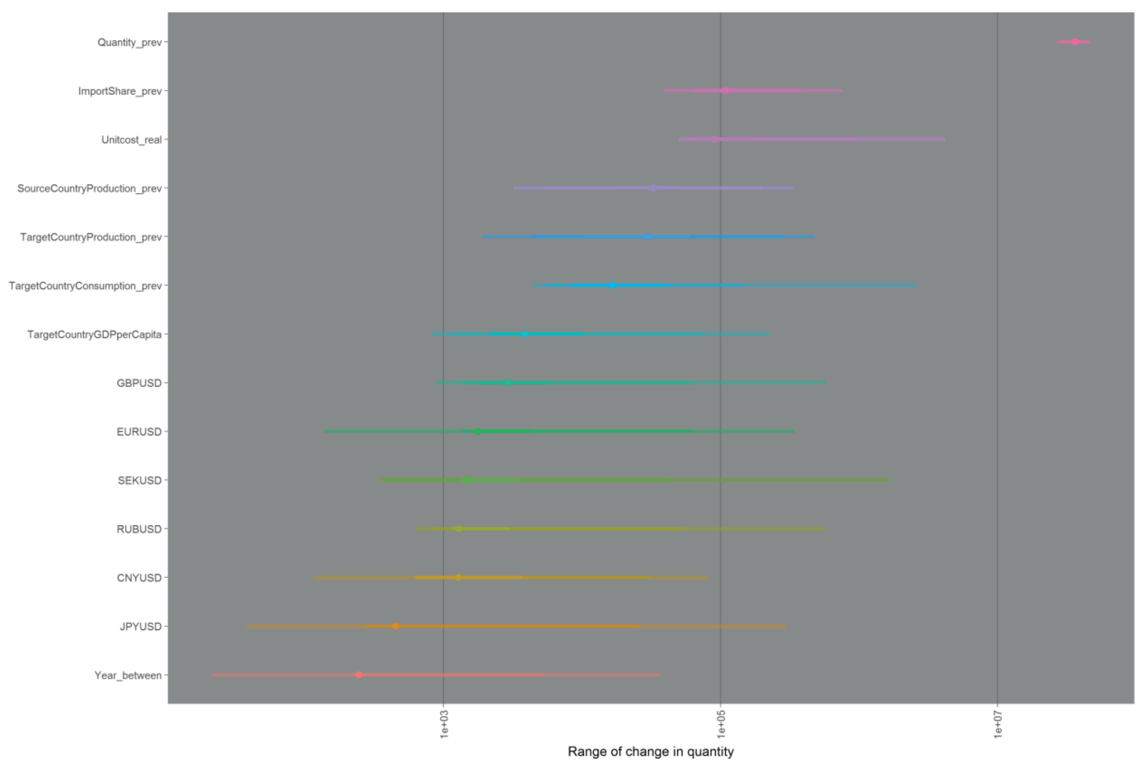

Figure 10. The range how much predicted bilateral quantity varies in respect to each parameter. Point depicts median range, the most transparent line (widest) is from minimum range to maximum range, the next less transparent line is from 2.5% percentile to 97.5% percentile and the most opaque line is from 25% percentile to 75% percentile i.e. middle 50% of distribution mass.

The large variation in ranges of exchange rates could be a result of a fact that certain exchange rate affects only a certain set of bilateral trades. In contrast the unit costs affect all bilateral trades.

In figure 11 we see the price sensitivity of trade flows for importer countries.

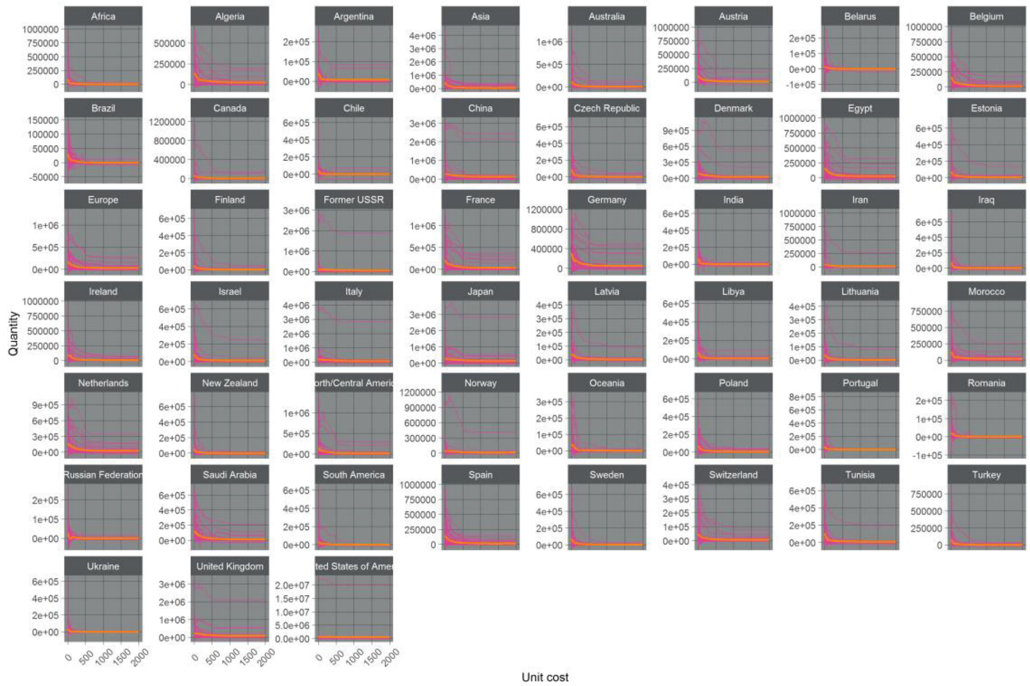

Figure 11. The sensitivity analysis for real unit costs for sawnwood. All the other predictors kept on average values in bilateral trades. The results are show for all importer countries. Each pink line shows sensitivity of quantity for price change in bilateral trade. Orange LOESS (Cleveland 1981) smoothed sensitivity for certain importer country.

We see that price sensitivity of countries varies a quite lot between import partners and importers. In figure 12 we see the sensitivity of price aggregated in importer country. It is interesting to see the profiles of the country sensitivity to price have different shapes.

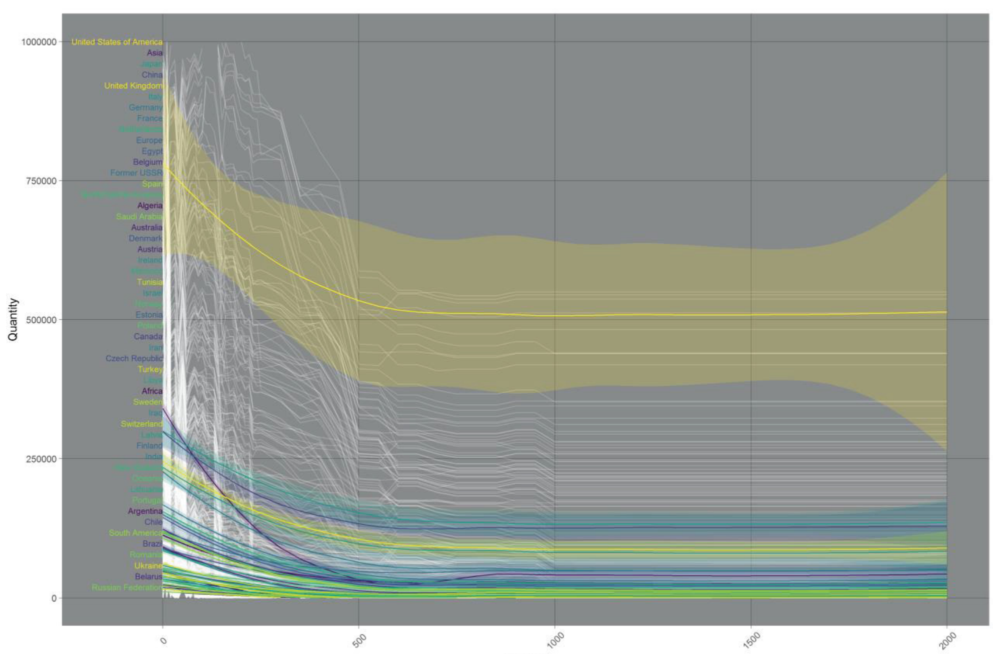

Figure 12. Sensitivity analysis for effects of changes in real unit cost of sawnwood in Extreme gradient boosting in tree ensemble 3. All the other predictors kept on average values in bilateral trades. The transparent white lines show the individual bilateral trade sensitivities. Coloured lines LOESS (Cleveland 1981) smoothed sensitivity for certain importer country.

The levels were individual countries end (stabilise) reflect the overall consumption in that country. It is also notable that the decrease of all the lines is deeper when unit costs are low. This indicates that the price is more effective attribute impacting traded quantity in lower prices. This might also indicate that when price goes over certain threshold then the commodity is regarded as luxury good and the price is not affecting to consumption.

Next we look quantity of previous bilateral trade, see figure 13. We can see clearly that the increasing amount of trade in history increases the predicted quantity. It is also interesting to note that there seems to steps in individual trades. This might be result from path dependency in trade when certain trade partners have been chosen then it is more practical to stick with those. If we compare the sensitivity of price and previously traded quantity we see that price stabiles earlier than previously traded quantity.

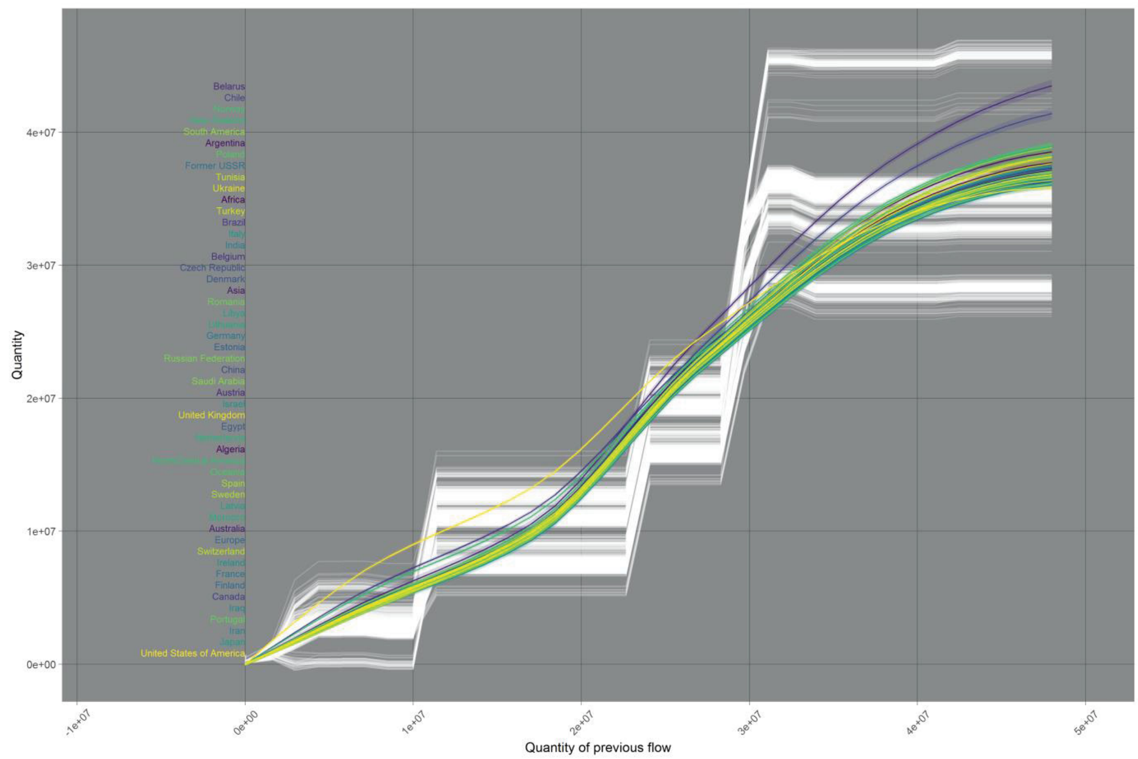

Figure 13. Sensitivity analyses for effects of changes in quantity of previous bilateral trade in Extreme gradient boosting in tree ensemble 3. All the other predictors kept on average values in bilateral trades. The transparent white lines show the individual bilateral trade sensitivities. Coloured lines LOESS (Cleveland 1981) smoothed sensitivity for certain importer country.

Next we look the sensitivity of previous import share, see figure 14.

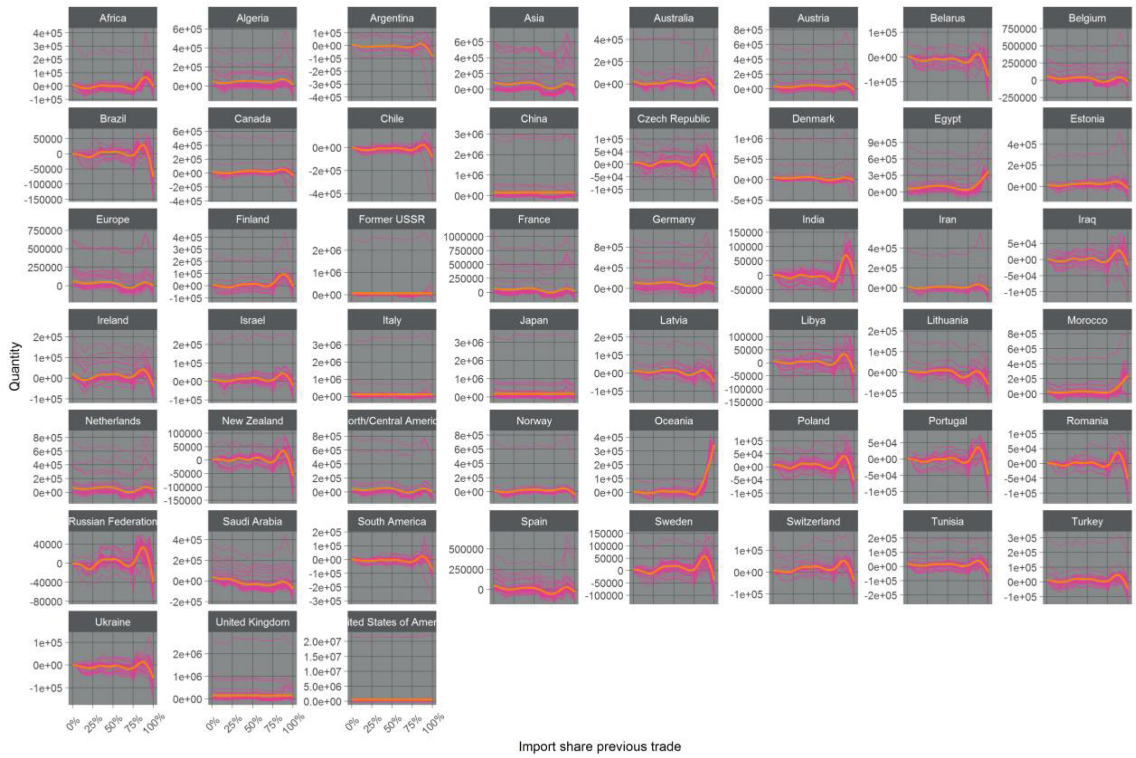

Figure 14. The sensitivity analysis for previous import shares. All the other predictors kept on average values in bilateral trades. The results are show for all importer countries. Each pink line shows sensitivity of quantity for import share change in bilateral trade. Orange LOESS (Cleveland 1981) smoothed sensitivity for certain importer country.

We see in figure 14 that the effect that previous import share has on quantity looks quite different for each importer country. Also the rate of change is different depending on the in which place of scale the share was previously. In general it can be said that if the share increases till approximately to 80% it has an increasing effect to quantity after 80% import share the effect is decreasing. This might indicate the idea of main trade partners focusing on those and then supporting partners.

In contrast the figure 15 shows the sensitivity SEKUSD exchange rate. We see that there are some changes in bilateral trades but in general level everything looks calm.

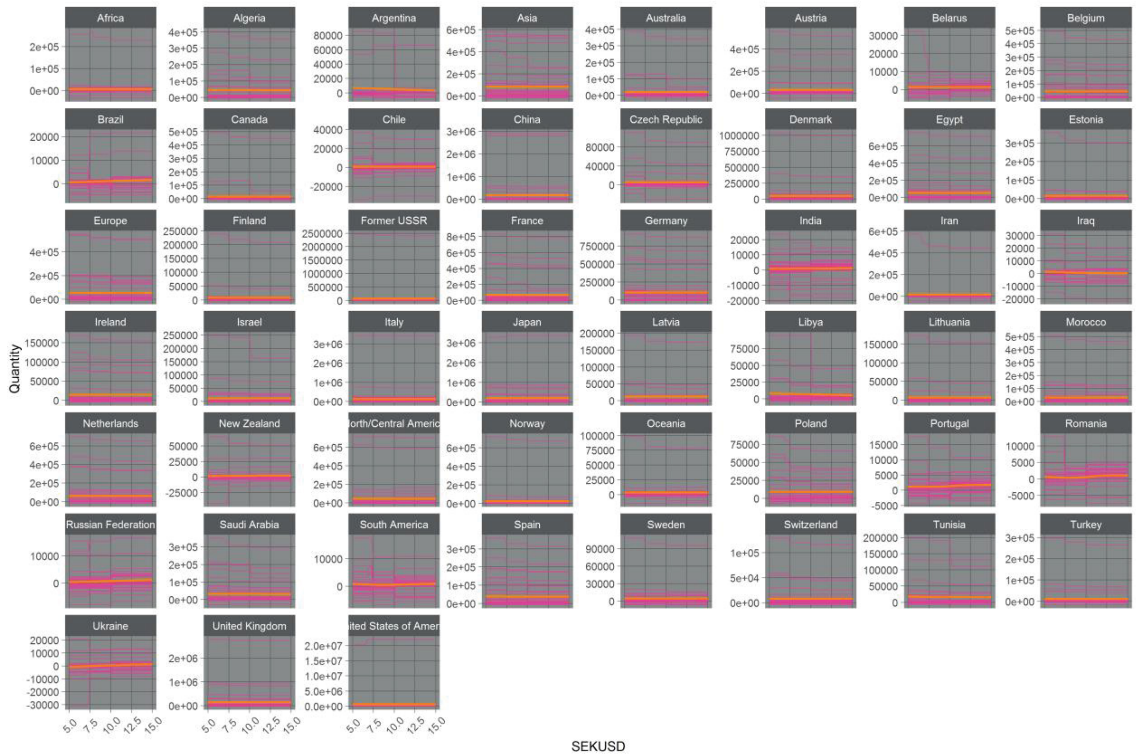

Figure 15. The sensitivity analysis for previous import shares. All the other predictors kept on average values in bilateral trades. The results are show for all importer countries. Each pink line shows sensitivity of quantity for import share change in bilateral trade. Orange LOESS (Cleveland 1981) smoothed sensitivity for certain importer country.

In conclusions we can say that the model seems to perform broadly as expected. The effects of certain parameters were on the ball park like price and previous quantity. On the other hand the minor or indistinguishable impact of exchange rates was a bit unexpected. However, it is natural, that all exchange rates including that of Swedish Krone affect only a certain set of bilateral trades.

6.4. Which features are important in making of predictions in the Extreme gradient boosting in tree ensemble

In this section we look which explanatory variables are important in Extreme gradient boosting in tree ensemble 3 model. We see in the figure 16 the top important explanatory variables. Quantity of previous trade seems to dominate importance of variables clearly. Follow-up variables are unit costs, previous consumption of sawnwood in importer country are not surprising. The GDP per capita of exporter is a bit more surprising.

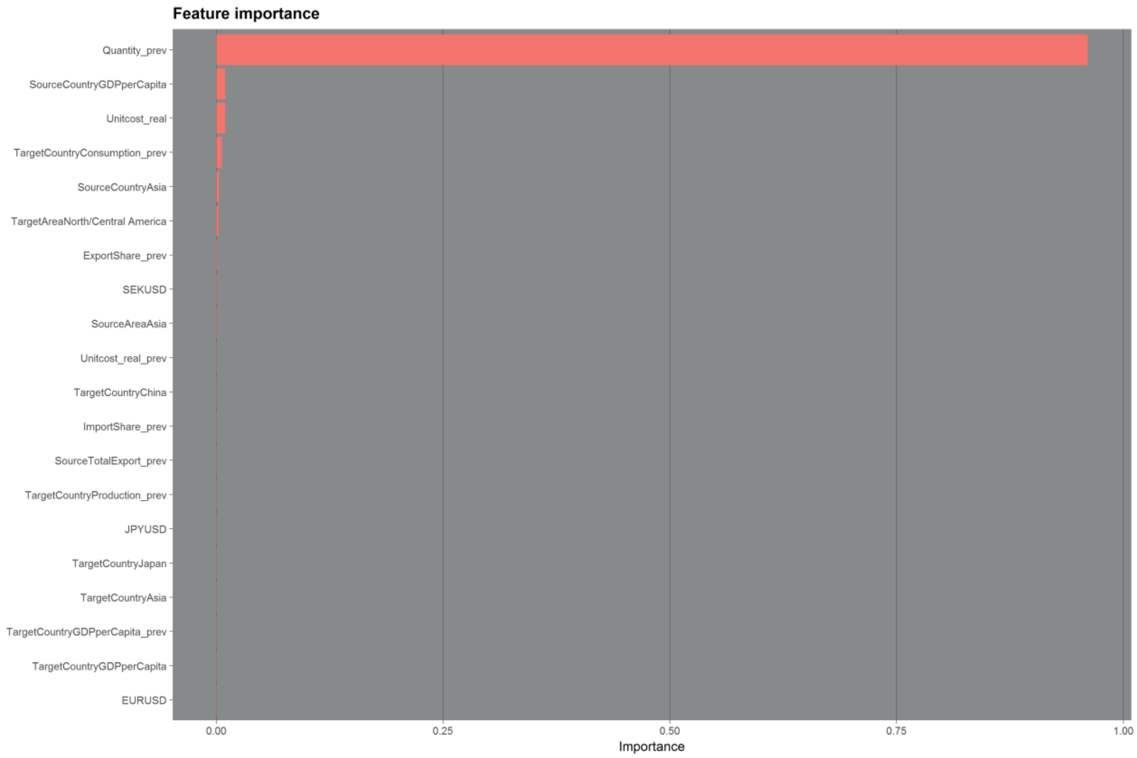

Figure 16. Importance of explanatory variables in Extreme gradient boosting in tree ensemble 3 model. Showing only the top 20 most important variables.

In figure 17 we could see an example of two regression trees from ensemble of trees in Extreme gradient boosting in tree ensemble 3 model.

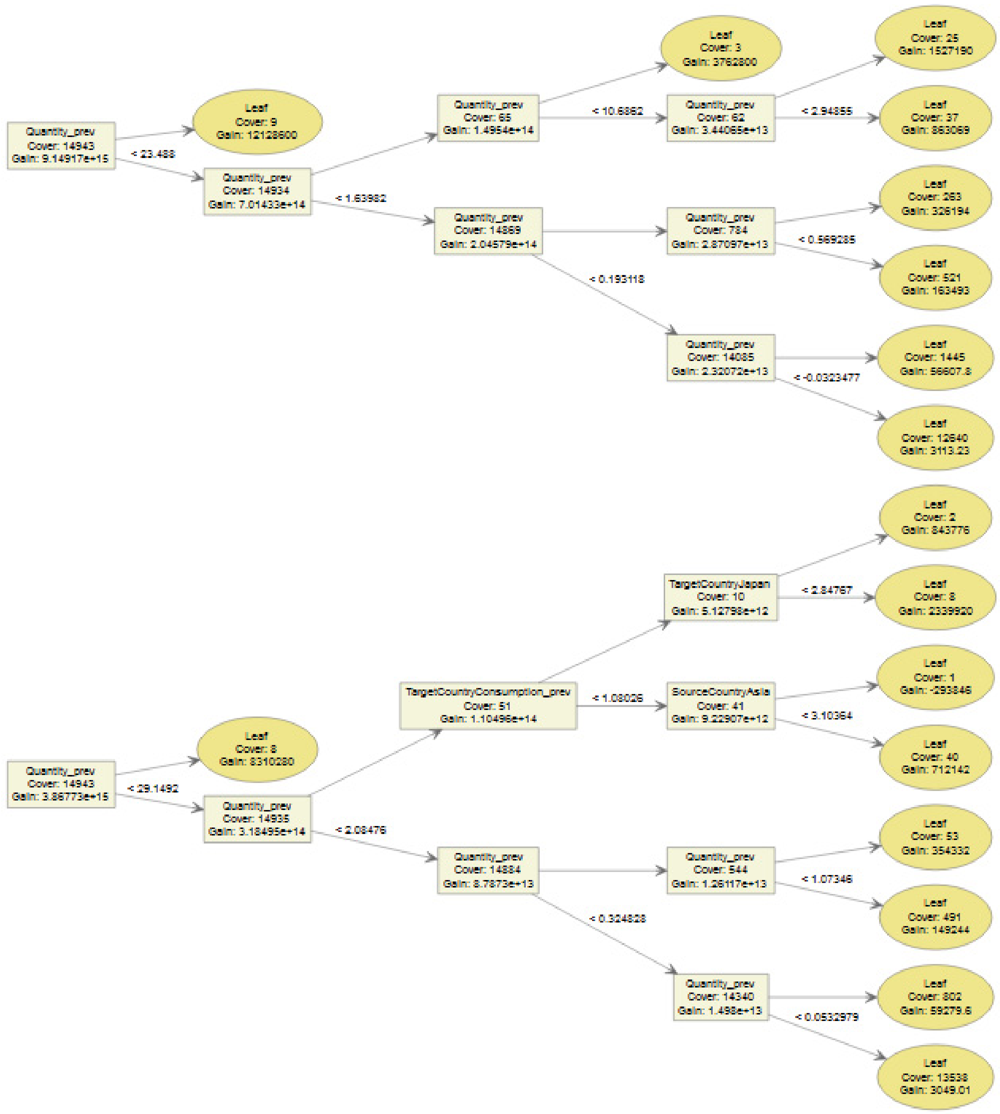

Figure 17. An example of decision trees in ensemble of Extreme gradient boosting in tree ensemble 3 model.

We can also look which are the common features in certain positions in trees in ensemble. We know that the deepness of trees is determined by parameter in model which is the same for all trees in the model. This helps to illustrate trees in compact manner. In figure 18 top six explanatory variables in each position in trees.

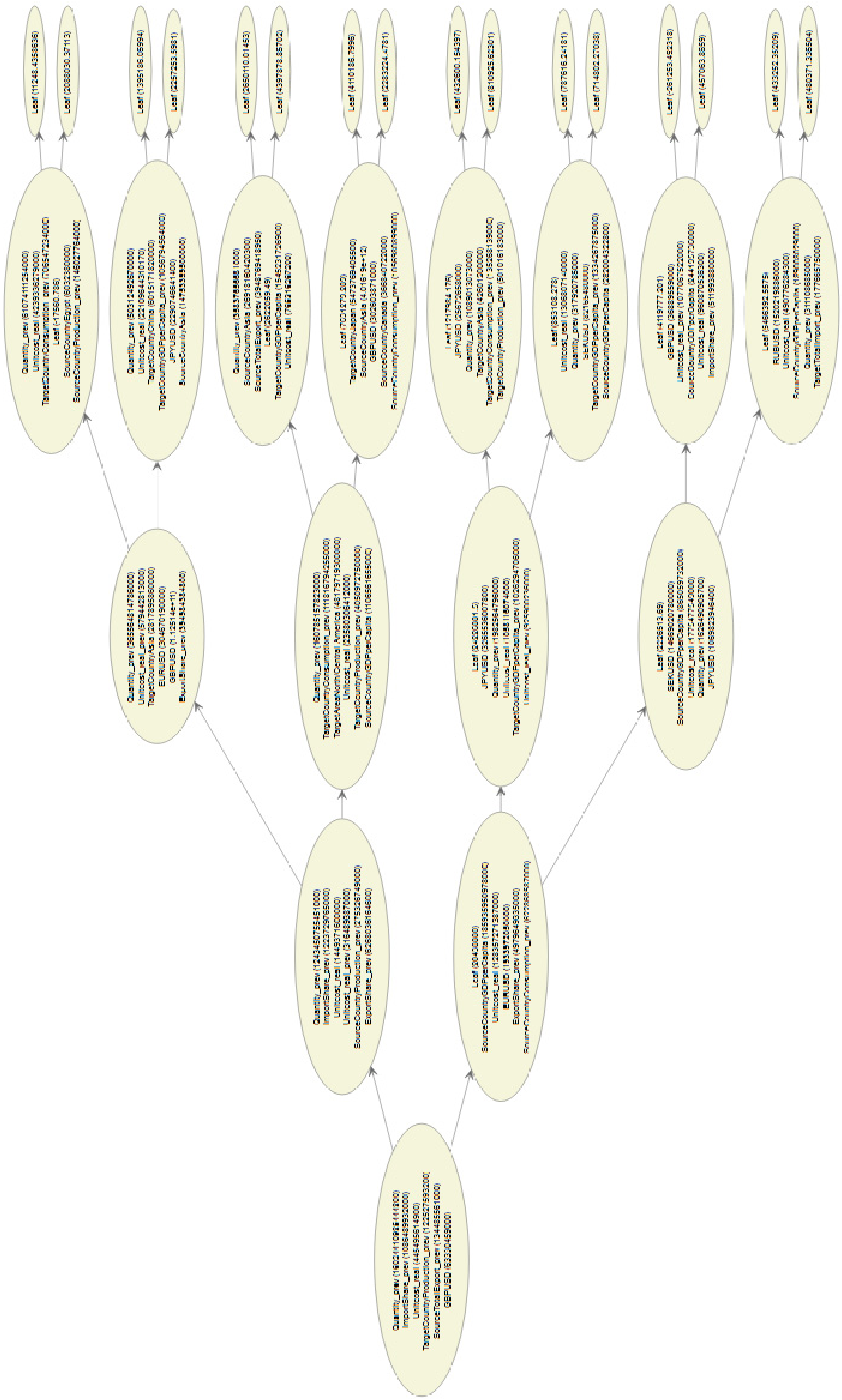

Figure 18. Top six explanatory variables in each tree position.

An example how changes in explanatory variables are reflected in bilateral trade flows between Finland and Japan can be seen in Figure 19. The changes in variables are reflect as assumed based Figure (5) and as economic theory predicts e.g. price decreases imports increases. However in the period of financial and economic crisis 2008-2009 those assumptions previously mentioned might not hold. Also the changes in explanatory variables seem to affect predicted outcome.

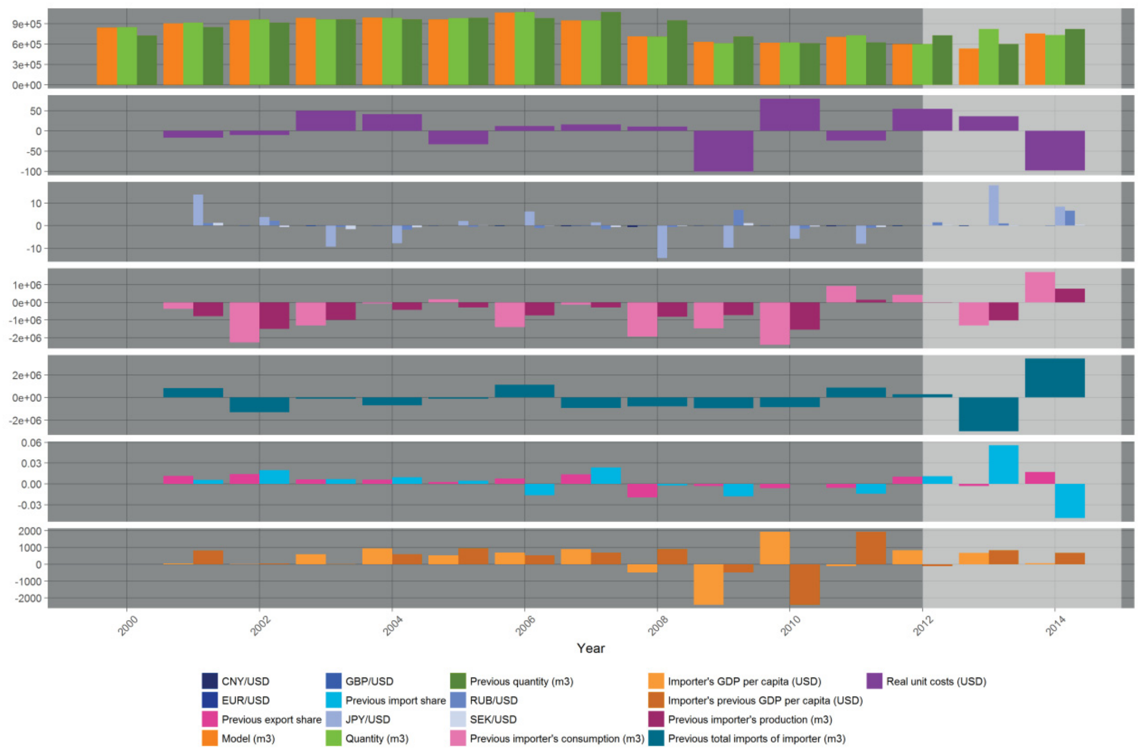

Figure 19. How the change in explanatory variables is seen in bilateral trade flow. Bilateral trade flow on Finland to Japan. The explanatory variables are differentiated with previous year.

7. Conclusions

We found some promising machine learning methods suitable for predictive analysis of the global sawnwood trade flows. The explanatory variables used in these models were quite easily available and this could make this kind of approach feasible in practical terms. Also we found out that the quality of data is imperative for success in machine learning models.

We used annual bilateral trade flow data by countries available in the FAOSTAT databank. In our case the quality of country specific data caused some hindrance while modelling. In some cases, there were quite large differences in quantities and values of the import and export countries'' reported values for bilateral trade, i.e. imports to country A from country B did not match with the respective exports from country B to country A. In these cases we used the flow of larger volume and value to represent the bilateral trade. Although we could have removed countries for which data was not with highest quality, we decided that we are willing to accept this trade of in quality to gain more coverage over the different countries. The coverage of all the countries in FAO's database is one key strengths of these models. We can use the models to predict trade flows for almost any country in the world. Even though the predictions in some bilateral partners is more uncertain than some other is this gives good option to analyse "what if scenarios" and possible new markets.

For technical point of view the computational requirements for some of the tested machine learning models were quite large. Even our model and dataset were not so huge. This might be partly resulting from poor implementation or that the method does not suit this kind of data well. Although the best machine learning methods were fast and consistent.

The choice of explanatory variables was based on the previous studies modelling the demand and trade of coniferous sawnwood. We also assumed that the trade of sawnwood is similar process between all the countries in the data. A part of the predictive power of the model results from the dynamics of the model. Some lagged variables were strong explanatory, e.g. previous bilateral traded quantity. These give interesting insight to dynamics of trading. It seems that the trading between established partners is more stable than with new partners.

To improve the models a disaggregated trade flow data would be needed. The FAO data for coniferous sawnwood is highly aggregated, and the product group includes various products of different values. Also data about end-uses of coniferous sawnwood in different countries would improve the quality of models. However, information and research is missing on the end uses of sawnwood.

In applying the model in practical analysis, it is important to update the models when new data is available. In formulating of future views, additional expert knowledge is needed to assess the effects of possible changes in the operation environment on the sawnwood markets. These changes may be related, for example to trade- or other policies and to unpredictable fluctuations in the economic developments. The changes may affect sawnwood markets also at a very short term, in some months ahead. In further studies, it would be necessary to try formulate models basing on monthly data, which could be available at least for certain countries.

This study revealed some dynamics of global trade flows of sawnwood. It also demonstrated that machine learning method could be used as practical way to do predictive analysis in this field. The resulting models can be used to support expert opinion and develop "what if analyses". The variables used as explanatory variables form a set of variables which are reasonably easily available to these models to be practical to use.

References

Breiman L., 2001, Random forest, Machine Learning, 45:5-32

Bolkesjo, T.F., Tromborg, E., and Solberg, B., 2006. Bioenergy from the forest sector: economic potential and interactions with timber and forest products markets in Norway. Scandinavian Journal of Forest Research 21 (2), 175-185.

Buongiorno, J. 1996. Forest sector modelling: A synthesis of econometrics, mathematical programming, and system dynamics methods. International Journal of Forecasting, vol. 12: 329-343.

Buongiorno, J. 2016. Gravity models of forest products trade: applications to forecasting and policy analysis. Forestry 2016:0, 1–10. doi:10.1093/forestry/cpw005

Buongiorno, J. 2015. Global modelling to predict timber production and prices: the GFPM approach. Forestry 2015; 88, 291-303, doi:10.1093/forestry/cpu047

Buongiorno, J., Brannman, J., and Bark, T. 1984. Econometric versus univariate and bivariate time-series forecasts: The case of softwood lumber imports. Forest Science Vol. 30, No. 1: 194-208.

Buongiorno, J. Zhu, S., Zhand, D., Turner, J., and Tomberlin, D. 2003. The Global Forest Products Model: Structure, Estimation, and Applications. Academic Press, Amsterdam, 288 pp.

Cannon, A. J., 2011, Quantile regression neural network: Implementation in R and application to precipitation downscaling, Computers & Geosciences 37, 1277-1284

Chang, W.Y. & Gaston, C. 2014. The competitiveness of Canadian osfrtwood lumber: a disaggregated trade-flow analysis. Canadian Journal of Forest Research 44:1494-1506 (2014). dx.doi.org/10.1139/cjfr-2014-0058.

Cleveland, W. S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots, Journal of the American Statistical Association. 74 (368): 829–836

Cortes, C., Vapnik, V., Support-Vector Networks, 1995, Machine Learning, 20, 273-297

Drucker, H., Burges, C., J., C., Kaufman, L., Smola, A., Vapnik, V., 1997, Support Vector Regression Machines, in Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press

Finnish Bioeconomy Strategy 2014. http://www.bioeconomy.fi/facts-and-contacts/finnish-bioeconomy-strategy/

The Finnish Forest Sector Economic Outlook 2016. Available only in Finnish: Viitanen J. & Mutanen, A. (eds.) Metsäsektorin suhdannekatsaus 2016-2017. Luonnonvara- ja biotalouden tutkimus 49/2016. 73 p. Luonnonvarakeskus, Helsinki 2016.

Friedman, J., Hastie, T. and Tibshirani, R., Additive Logistic Regression: A Statistical View of Boosting, 2000, The Annals of Statistics, Vol. 28, No 2, 337-407

Hastie, T., Tibshirani, R., Friedman, F., 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction Second Edition, Springer

Hendry, D.F., Clements, M.P., 2003. Economic forecasting: some lessons from recent research. Economic Modelling 20,301–329.

Hietala, J., Hänninen, R. & Toppinen, A. 2013. Finnish and Swedish sawnwood exports to the UK market in the European Monetary Union regime. Forest Science 59(4): 379-389.

Hänninen, R. 2004. Econometric models in forest sector forecasting. Editorial. Journal of Forest Economics 10(2004): 57-59.

Hänninen, R., Toppinen, A. & Toivonen, R. 2007. Transmission price changes in sawnwood and sawlog markets of the new and old EU member countries. European Journal of Forest Research 126 (1) 11-120.

Hänninen R., Toppinen A., Verkasalo E., Ollonqvist P., Rimmler T., Enroth R., Toivonen, R. (2007). Puutuoteteollisuuden tulevaisuus ja puurakentamisen mahdollisuudet. [The future of the wood products industry and the opportunities of wood construction]. Working Papers of the Finnish Forest Research Institute 49.

Hurmekoski, E., Hetemäki, L., & Linden, M. 2015. Factors affecting sawnwood consumption in Europe. Forest Policy and Economics 50(2015)236-248.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013, An Introduction to Statistical Learning, Springer

Jonsson, R. 2010. Modelling wood products demand, supply and trade. In Mantau, U.(Ed.) EUwood-Real Potential for changes in growth and use of EU Forests. Methodology report. Hamburg, Gemany, pp. 33-45.

Kallio, A.M.I., Moiseyev, A., Solberg, B. 2006. Economic impacts of increased forest conservation in Europe: A forest sector model analysis. Environmantal Science & Policy 9(5), 457-465.

Kangas, J. & Niskanen, A. 2003. Trade in forest products between European Union and the Central and Eastern European access candidates. Forest Policy and Economics 5 (2003) 297–304.

Meinshausen. N., 2006, Quantile Regression Forest, Journal of Machine Learning Research 7, 983-999

Simangunsong, B.C.H., Buongiorno, J., 2001. International demand equations for forest products: a comparison of methods. Scand. J. For. Res. 16 (2), 155–172.

Toppinen, A. & Kuuluvainen, J. 2010. Forest Sector modelling in Europe - the state of the art and future research directions. Forest Policy and Economics 12(2010) 2-8.

Turner, J.A., Buongiorno, J:, Zhu, S., and Prestemon J.P., 2005. The US forest sector in 2030: markets and competitors. Forest Products Journal 55 (5), 27-36. UNECE/FAO 2011. European Forest Sector Outlook Study (EFSOSII). UNECE Timber Committee- FAO European Forestry Comission. www.unece.org.

UNECE 2016. Forest Products Annual Market Review 2015-2016. ECE/TIM/SP/40. United Nations, Geneva. 141 p.

luke.fi

Natural Resources Institute Finland
Latokartanonkaari 9
FI-00790 Helsinki, Finland
tel. +358 29 532 6000