

Comparison of ARIMA, SVR and ANN models including exogenous variables for short-term freight demand forecasting at a large national food distributor

Juliana Schneider

February 4, 2019

1 Introduction

While technological progress in the past decades has led to a significant decrease in relative emissions per tonne-kilometre in truck haulage, this achievement is compensated by an increase in freight transportation (24). In fact, absolute emissions of carbon dioxide rose by 20 % between 1995 and 2017, while freight transportation in general increased by 64% between 1991 and 2016 (25)

Between 2010 and 2030, according to (6), transport volumes for road haulage (in millions of tons) will face an estimated increase of 17% while its transport performance (in billion kilometres) will increase by about 39%. Thus, about 80% and 74% of predicted growth until 2030 for transport volumes and performance respectively will be due to road haulage.

Traditionally, freight transportation demand models are concerned with either vehicle or commodity movements, focusing on origin-destination matrices and geographical scope (e.g. urban, regional, international (20)). Accordingly, the question of how many passengers and how much of different kinds of commodities is transported by which type of vehicle on which route is a task of interest for political institutions such as the German Federal Ministry of Transport and Infrastructure (Bundesministerium für Verkehr und Infrastruktur, BMVI) as it is essential for the planning of infrastructure based on transportation demand.(6)

Furthermore, in order to comp with the enormous impact of road haulage on climate, it is a major task for the government to improve and adapt infrastructures as well as strategic planning to transportation demands (5).

However, considering the growing demand for freight transportation, not only politics but also companies share an interest in the modeling of freight transportation demand (23).

What's more, according to (11), freight forecasting can help improve profitability as empty or imbalanced trips are reduced. Furthermore, by planning efficiently, the drivers' quality of life and work-life-balance is enhanced and employee turnover is reduced. This, of course, again reduces a company's expenses, since a high turnover rate imposes costs by the need to constantly hire and train new employees.

Another advantage of sophisticated freight demand forecasting is the enhancement of customer satisfaction through shorter delivery times, higher reliability and adapted pricing.

Unfortunately, little advancement has been made in the forecasting of short-term freight demand for truckage companies so far. This thesis now aims at building a bridge towards the objective of providing a useful, and applicable model to users.

By comparing three different kinds of methods - a rather classic time series approach and two machine learning models - combined with the incorporation of exogenous variables provided by public sources, first advice for future forecasts for this business case shall be given as well as a framework for further research.

After a brief description of the use case at hand, an overview of existing approaches, the theoretical background for variable selection and statistical methods used in this thesis is provided, followed by an introduction to the business case, the truckload company providing the data and a description of these data. Afterwards, the procedure of variable selection, model establishment and model testing is described in detail. Finally, the results achieved for this business case are reported and discussed.

2 Use Case: Short-term forecasting of freight weight for road haulage - on the example of NAGEL group

2.1 NAGEL group

2.2 The data

2.3 External data

hier nur kurz beschreiben, dass es externe Daten geben wird.

3 State of the Art in the Modeling of Freight Transportation Demand

NOCH USE OF EXTERNAL DATA BESCHREIBEN

First of all, while browsing literature on the topic of freight transport demand, the sheer amount of different aspects to take into account and consequently the various different quantities under observation in different papers was striking. Common measures in freight demand modeling are: tonne-kilometres, mode choice, freight volume, origin-destination-matrices, and so on.

As the goal of this thesis is to establish a model to predict short-term freight demand for a single user, comparability and therefore scalability of the target variable is not necessary. Hence, the target variable "weight" included in the data provided by NAGEL is not altered, but of course, in further works it might be of use to modify the target variable, e.g. to value/weight-ratios.

As initially mentioned, freight transport modeling has so far been concerned with the so-called four-step transport models consisting of trip generation (number of trips from or to origin), trip distribution (destination of trips from origin), mode choice, and trip assignment (choice of path or route between origin and destination), adopted from traditional passenger transport modeling(8; 22).

However, as there are differences in freight demand and passenger demand modelling, adaptations have been proposed by (8). Among other things, the first step in the four-step model - trip generation - should not count the number of trips starting from a certain region of origin, but count the amount of freight (in tonnes) delivered from this origin.

In order to construct an entire freight demand model system, the separate measures of all four steps need to be connected, e.g. through value/weight ratios.

However, the objective of this thesis is not to establish an entire freight model system but it is rather concerned with questions that would belong to step one.

Despite the adaptations proposed by (8), not many other advances have been made. In fact, while passenger transport is a largely studied field due to government regulations, this does not hold true for freight demand (20). (8) further states that within step one, models so far have only used aggregate data, where time series analysis models, with and without exogenous variables, have already been applied to short-term forecasting.

(20), besides distinguishing between aggregate and disaggregate models, classify into international, intercity and urban freight transportation. Here, at first glance, the category of an intercity model seems suitable, but in this category, freight transportation demand is either modeled as a utility max-

imisation task or an inventory based model for mode choice and production decision, which again is not the scope of this work.

Meanwhile, since, as mentioned above, most literature so far is concerned with aggregate data, within the use case at hand the focus is on the analysis of disaggregate data. Apart from this, further parts of the four step model are not of interest, i.e. neither other parts of step one (e.g. I/O-models) nor other steps, which is why the research on existing literature was expanded to include more specific studies.

(21) used artificial neural networks to predict the demand for weather-sensitive retail products.

(13) predicted container throughput using Dynamic Factor Analysis and ARIMAX models.

(15) established hybrid ARIMA and ANN models based on DWT decomposition for general time series analysis. (17) used ARIMA to predict full truckload transportation prices.

(18) applied SARIMA models to forecast the demand in a beverage supply chain.

(12) combined grey models, ANN and Support Vector Machines to model full truckload volume. (2) use both univariate and multivariate ARIMA models, examining the impact of new links, new destinations and lower fares on air transport demand at the Port of Reggio Calabria. (14) build a hybrid of ARIMA and SVR for general long term time series prediction. (16) suggest Ensemble Neural Network models to yield more accurate results in time series forecasting.

As is evident, all of the papers mentioned above applied models to either measures similar to but not exactly weight or time series analysis in general. Several authors proposed hybrid models consisting of combined versions of the basic models to be compared further below.

Regardless of the lack of literature exactly matching the target of this thesis, the combined information gained from all of these papers provides reasonable background to be supportive of it. Furthermore, this scarcity in literature proves the necessity of a primary approach to adress the objective of the use case at hand.

Hence, it is important to first retrace the reasoning behind the choice of rather basic versions of ARIMAX, SVR and ANN models for the prediction of freight weight demand, which the following section is dedicated to.

4 Artificial Neural Networks

Although seemingly new in the world of statistical tools, Artificial Neural Networks have been invented many decades ago and the enthusiasm for them

HAS experienced several revivals throughout the years (12). The idea stems from human brains (26) and the way their neurons learn, store and process information together (Balkin). Although this sounds like a very complex, ambitious venture, NNs are "simply a parameterized non-linear function that can be fitted to data for prediction purposes" (19). The process, although very much data-driven (26) and sensitive to the training data (1), it is not non-parametric (9).

can be fitted to data for prediction purposes" (19). Nowadays, with the advance of fast and high-performance personal computers, basically anyone can fit a neural network to their data (12).

During these different stages of popularity, ANNs have been developed further. Refined architectures allow the researcher to adapt a neural network to their problem's structure, starting from rather simple one-layer feedforward algorithms to more complex long-short-term-memory multi-layer-perceptron NNs. (QUELLE)

Here, a short overview shall be given before discussing the choice of type of NN for this specific case.

4.1 The general structure of NNs

A NN typically consists of (at least) three layers: One input, one hidden and one output layer. ANNs learn by example (Balkin; 12): Neurons in the input-layer receive the input from the data and pass them on to the hidden layer, where information is processed and then passed to the output layer, which usually consists of one single neuron and is equivalent to the target value. (QUELLE)

Layers - and neurons - are connected by activation, also called transfer functions, which multiply the output received by the preceding neuron by a weight and pass this new output on to the next layer and so on.

Lastly, a learning algorithm is applied to estimate the parameters in the network. When fed some training data, the network's output is compared to the actual output value of the target variable through a loss function. This error is then passed backwards through the layers to adjust the transfer functions' weights.

Multilayer perceptrons are the most popular type of (feedforward) NN (Balkin; 4)

The most common learning algorithm is backpropagation; others are e.g. Levenberg-Marquardt or

4.2 Network size and structure

A crucial step in modelling NNs is to define the network size and structure, i.e. define the number of layers and nodes (neurons), in order to accurately depict the complexity of the structure of the data and avoid overfitting or underfitting as well as keep computation times low (?).

There is no exact way to determine the number of nodes in each layer, although a lot of previous work has been dedicated to this problem (?). Of course, it depends on the problem at hand.

The number of input neurons may be equal to the number of input variables in one's problem, but in a time series, the number of input neurons is not evident (26). It is already clear at this point that the number of output neurons is one, as we are interested in one continuous target variable and as is recommended by (?).

However, the number of hidden layers and the number of neurons within each hidden layer still has to be determined.

As the objective of this thesis is to present simple models to be contrasted with each other, one hidden layer should suffice for the moment. Now, the number of hidden neurons has to be established by trial and error - since there is no clear rule - as is done by most researchers (?).

4.3 Activation functions

Activation functions - also known as transfer functions - introduce nonlinearity by mapping the input node to the output node(26). They must be differentiable (QUELLE)

Generally, the activation a of a neuron n_i is given by

$$a_i = f(n_i) = f(\sum w_{ij} * x_j) \quad (1)$$

where $f(n_i)$ is the transfer function, w_{ij} is the connection weight between node j and node i , x_j the input signal from the node j , and a_i is the output of the neuron i (4). Among the most popular are sigmoid, hyperbolic tangent, sine or cosine, and linear functions.

$$f(x) = (1 + \exp(-x))^{-1} \quad (2)$$

sigmoid or logistic. Sigmoid functions can take values between 0 and 1

$$f(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x)) \quad (3)$$

Hyperbolic tangens (tanh) function can take values between -1 and 1.

$$f(x) = \sin(x) \quad (4)$$

or

$$f(x) = \cos(x) \quad (5)$$

sine / cosine can take values between -1 and 1.

$$f(x) = x \tag{6}$$

linear can take values between $-\infty$ and ∞ .

(26) point out that for target variables with continuous values, the use of a linear activation function in the output node is advisable. It must be noted that HIER SEKUNDÄRQUELLE COTTRELL ET AL IN ZHANG 1998 linear activation functions may not be applied on data with trends; hence differencing may be necessary before feeding the data to the network.

Since most papers reviewed for this thesis mention the use of sigmoid or logistic transfer functions in the hidden nodes and linear activation in the output nodes, it seems a reliable *modus operandi*.

4.4 An overview over types of Artificial Neural Networks

(9) counts feedforward ANNs, Jordan ANNs, Elman ANNs (the latter two being types of RNNs); and Multicurrent Networks as most relevant for time series analysis. According to (Balkin), RNNs are better suited for time series data than FNNS; in fact, bad results in an RNN may hint that there is not time dependency in the data at all. However, a feedforward NN in its simplest form reduces to an $AR[p]$ -model whereas an RNN may reduce to an $ARMA$ -model.

Due to their ability to approximate any (linear or nonlinear) function underlying the data arbitrarily closely by a nonlinear function, MLPs and RBFNNs classify as universal function approximators (9). They can reduce to an $AR[p]$ -model with an appropriate function that is estimated with a learning algorithm.

Considering the vast number of different types of NNs as well as their general flexibility, it is not a trivial task to choose a suitable model, specify it, and train it (16).

5 Model Choice

There are several ways to model time series data, some of which have been well established for many decades already, while others have recently emerged in the space of popular methods as well. ARIMA models are an example of the former kind of methods, while SVM and ANN are examples of the latter kind. (1)

(14) state ARIMA, SVM and ANN as the three major methods used in time series analysis.

| | ARIMA | SVR | ANN |
|------------|--|---|---|
| | | Models linear as well as nonlinear relationships (Han 2014, Adhikari 2014) | models nonlinear relationships |
| | | Only few data required (Wu 2008) | variable selection |
| | | Variable selection | useful for long-term* time series (Han 2014) |
| Pro | Efficient (Adhikari 2014) | Able to find global maximum (Han 2014) | non-parametric and nonlinear modeling skill, flexible and data-driven nature, good generalization ability, and reasonably good accuracy. (Adhikari 2014) |
| | | strong generalization ability (Adhikari 2014) | RNN capable of including time dependencies (Adhikari 2014) |
| | | faster convergence speed (Wu 2008) | |
| | | | May lead to overfitting (Deng 2005) |
| | | | May only find local maximum (Deng 2005, Adhikari 2014) |
| Con | Models linear relationships between variables – accuracy may be afflicted (Adhikari 2014) | Training times increase drastically with respect to the number of training samples (Deng 2005) | Sensitive for preselected model architecture and data fed to the model (Adhikari 2014) |
| | Easy to be disturbed by noise (Wu 2008) | | Computational cost RNN > FNN (Adhikari 2014) |

ANN are particularly useful as a data-driven method, if there are data but difficulties in describing the process behind them (26). In the case at hand, as there is not much literature exactly fitting the purpose of my research question, ANNs might be a convenient approach to handle the data.

All three of the aforementioned methods may prove suitable to forecast freight weight, as they each have advantages. Before going into detail about each method and specifying each model according to the data, in this section, ARIMA, SVR and ANN shall be directly contrasted in order to give a brief justification as to why these three were selected to be compared in this thesis.

Later, a conceptual choice of the specific models for each method (e.g. choice of kernel, type of NN) shall be made before specifying them exactly according to the data in the application.

6 Variable selection

A well-defined model should include all relevant explanatory variables that contribute to the variance found in the target value and exclude irrelevant, noisy variables. This way, the data and the process constituting it can be better understood, computational time and the so-called curse of dimensionality can be reduced while maintaining a high prediction accuracy (7). The curse of dimensionality describes the problem arising in a high dimensional (i.e. containing many variables) space, where one needs very large samples to accurately depict said space (QUELLE). Variable selection is furthermore important as irrelevant data may cause poor generalization when given new data (7).

A first selection of variables should be given by experts. (QUELLE) Next to the researchers' expertise, there exist several variable selection (also called feature selection) methods that quantify the explanatory value of each possible variable and include and / or exclude them in the final model by a prespecified selection criterion. Directly comparing all 2^N possible subsets of variables is computationally unfeasible, which is why usually metaheuristics are applied. The most common types of metaheuristic variable selection

methods shall be described in the following, as well as the choice process for a variable selection method for this case.

In their study on the influence of national, regional and local economic indices on freight volume for different segments of the truckload industry in the USA, (11) first examined correlation matrices before using stepwise multiple regression to draw a selection from the large set of initial variables collected. (21) suggest using ensemble methods such as Random Forest to take advantage of several different machine learning methods. (10) used Spearman rank correlation to select external variables to predict railway freight volume.

Albeit the growing number of available methods for variable selection, neither is there one gold standard that is generally useful, nor is there one method especially advisable for each kind of data (21).

References

- [1] Adhikari, R. (2015). A neural network based linear ensemble framework for time series forecasting. *Neurocomputing*, 157:231–242.
- [2] Andreoni, A. and Postorino, M. N. (2006). Time series models to forecast air transport demand: A study about a regional airport. *IFAC Proceedings Volumes*, 39(12):101–106.
- [Balkin] Balkin, S. Using recurrent neural networks for time series forecasting.
- [4] Benkachcha, S., Benhra, J., and El Hassani, H. (2015). Seasonal time series forecasting models based on artificial neural network. *International Journal of Computer Applications*, 116(20):9–14.
- [5] Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit (2018). Klimaschutzbericht 2018: zum aktionsprogramm klimaschutz 2020 der bundesregierung: Referentenentwurf.
- [6] Bundesministerium für Verkehr und Infrastruktur (2014). Verkehrsverflechtungsprognose 2030.
- [7] Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28.
- [8] de Jong, G., Gunn, H., and Walker, W. (2004). National and international freight transport models: An overview and ideas for future development. *Transport Reviews*, 24(1):103–124.
- [9] Dorffner, G. (1996). Neural networks for time series processing. *Neural Network World*, 6:447–468.

- [10] Feng, F., Li, W., and Jiang, Q. (2018). Railway freight volume forecast using an ensemble model with optimised deep belief network. *IET Intelligent Transport Systems*, 12(8):851–859.
- [11] Fite, J. T., Don Taylor, G., Usher, J. S., English, J. R., and Roberts, J. N. (2002). Forecasting freight demand using economic indices. *International Journal of Physical Distribution & Logistics Management*, 32(4):299–308.
- [12] Gao, S., Zhang, Z., and Cao, C. (2011). Road traffic freight volume forecast using support vector machine combining forecasting. *Journal of Software*, 6(9).
- [13] Intihar, M., Kramberger, T., and Dragan, D. (2017). Container throughput forecasting using dynamic factor analysis and arimax model. *PROMET - Traffic&Transportation*, 29(5):529–542.
- [14] João F. L. Oliveira and Teresa Bernarda Ludermir (2014). Iterative arima-multiple support vector regression models for long term time series prediction. In *ESANN*.
- [15] Khandelwal, I., Adhikari, R., and Verma, G. (2015). Time series forecasting using hybrid arima and ann models based on dwt decomposition. *Procedia Computer Science*, 48:173–179.
- [16] Kourentzes, N., Barrow, D. K., and Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, 41(9):4235–4244.
- [17] Miller, J. (2018). Arima time series models for full truckload transportation prices. *Forecasting*, 1(1):121–134.
- [18] Mircetic, D., Nikolicic, S., Maslaric, M., Ralevic, N., and Debelic, B. (2016). Development of s-arima model for forecasting demand in a beverage supply chain. *Open Engineering*, 6(1):14.
- [19] Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, 86(3):373–378.
- [20] Regan, A. and Garrido, R. (2001). Modelling freight demand and shipper behaviour: state of the art, future directions. *Travel Behaviour Research*.
- [21] Taghizadeh, E. (2017). Utilizing artificial neural networks to predict demand for weather-sensitive products at retail stores. *CoRR*, abs/1711.08325.
- [22] Transport and Infrastructure Council (2016). Australian transport assessment and planning guidelines: T1 travel demand modelling.

- [23] Tsekeris, T. and Tsekeris, C. (2011). Demand forecasting in transport: Overview and modeling advances. *Economic Research-Ekonomska Istraživanja*, 24(1):82–94.
- [24] Umweltbundesamt (2018a). Emissionen des verkehrs.
- [25] Umweltbundesamt (2018b). Fahrleistungen, verkehrsaufwand und modal split.
- [26] Zhang, G., Eddy Patuwo, B., and Y. Hu, M. (1998). Forecasting with artificial neural networks. *International Journal of Forecasting*, 14(1):35–62.