AN EVALUATION FRAMEWORK FOR PUBLICATIONS ON ARTIFICIAL NEURAL NETWORKS IN SALES FORECASTING

Sven F. Crone
Department of Management Science
Lancaster University, UK
crone@bis-lab.com

Patrick C. Graffeille
Institute of Information Systems
University of Hamburg, Germany
patrick@graffeille.de

Abstract. Although artificial neural networks (ANN) promise superior performance in forecasting theory, they are not yet an established method in business practice. The vast degrees of freedom to parameterize ANNs have lead to countless heuristic approaches to simplify modeling, training, network selection and evaluation implemented with varying success. Consequently, a systematic evaluation is required in order to identify successful heuristics and derive sound guidelines to ANN modeling from publications. As each forecasting domain imposes different heuristics for classification or point prediction on specific datasets, a literature review is conducted, identifying 2538 publications within the domain of ANN forecasting but only 32 of them applicable to the domain of sales forecasting. The identified publications are evaluated through a framework regarding their validity and reliability in experiment design and documentation, in order to promote superior publications, derive recommendations for future experiments and possibly identify gaps in current research and practice.

I. INTRODUCTION

Artificial Neural networks (ANN) have achieved state-of-the-art performance in non-linear prediction, classification and clustering in management science [1-3]. In forecasting, theoretical research on NN receives unabated interest, leading to successful applications of time series and explanatory forecasting in various problem domains including demand planning [4-7]. ANNs promise attractive features to business forecasting, being a data driven learning machine, permitting universal approximation [8] of arbitrary linear or nonlinear functions from examples without a priori assumptions on the model structure, often outperforming conventional statistical approaches of ARIMA- or exponential smoothing- methods. Despite their theoretical capabilities, ANN are not an established forecasting method in business practice. Scepticism on ANN persist through mathematical complexity and missing transparency, in addition to inconsistent research findings and pessimistic reports on their performance [9, 10], in part due to a trial-and-error modelling process [11].

ANNs offer vast degrees of freedom in the modelling process, requiring a multitude of interdependent decisions on parameter-settings to assure valid and reliable performance. Since a complete enumeration of all parameter combinations often induces prohibitively high computation time, various heuristic modelling approaches, empirical guidelines, rules of thumb and simple tricks have been proposed, suggesting alternative approaches to determine the architecture, guide the training process and select appropriate models to minimize the objective function [10, 12-15], each with varying success. Unfortunately, no single heuristic has demonstrated its ability to deliver valid and reliable forecasts on arbitrary datasets. Consequently, the task of modelling ANNs for a particular prediction problem is considered as much an art as a science [9, 10]. In addition, each heuristic should be evaluated specifically within its application domain in order to derive robust recommendations for valid and reliable predictions. However, not all experiments have been conducted with the same scientific rigor, often questioning results and performance along with the heuristic proposed.

Following, we propose an evaluation framework for publications on ANN in business forecasting, to allow a distinction between superior publications prepared with necessary scientific rigor in experiment design, execution and documentation from less objective results. The framework aims to evaluate the validity and reliability of a publication through the degree of documentation and it adherence to essential prerequisites to produce valid results. We focus our analysis on point-predictions of customer demand or sales as needed in all operational systems dedicated to demand planning, inventory management or advanced planning in industry, wholesale and retail. The framework is applied to 32 publications on sales forecasting, identified from a systematic literature review and evaluation of 2568 papers in the general domain of forecasting with ANNs, in order to evaluate and identify publications with limited scientific rigor leading to suboptimal recommendations and heuristics.

Following a description of the Literature review in section 2, we develop our evaluation framework regarding the reliability and validity of each publication. Section 4 assesses the individual publications within the evaluation framework. Conclusions are given in section 5.

II. LITERATURE REVIEW

A variety of Literature reviews on business applications of ANNs have been published [10, 11, 16-18]. Following Adya and Collopy's survey and evaluation [11] we conduct a citation analysis at the ISI web of science (review conducted 04/2004)[1]. Our keyword search for neural networks in forecasting[2] yielded 2568 publications, demonstrating an unabated interest an continuous increase of publications on theoretical developments and practical applications since 1988 in a wide range of disciplines, including weather, biological processes, mathematical series and other non-business applications. Fig.1 gives an overview by year.

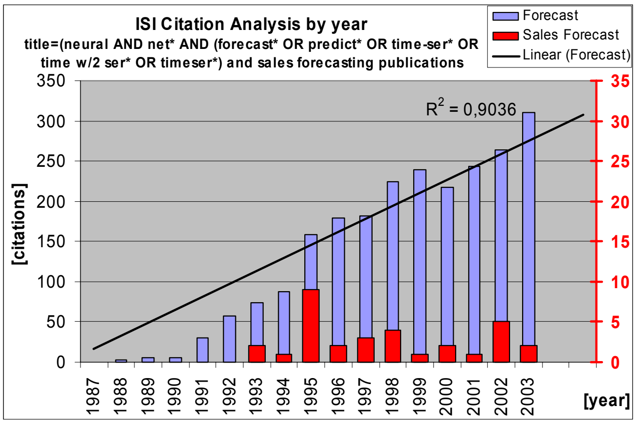

Fig.1: ISI citation analysis on neural networks for forecasting by year

However, significant differences exist in modelling suitable architectures for business forecasting for each domains, depending on the forecasting objective, application and dataset. Firstly, a variety of ANN publications focus on nominal predictions for classification tasks, which induce different modelling heuristics of little use to derive point predictions through the metric predictions necessary for sales forecasting. In addition, sales forecasting often implies noisy time series with only a few years of monthly or weekly observations, due to structural breaks from launches, relaunches, new product introductions, listing and delistings etc. as opposed to long time series with a high signal-to-noise ratio often found in engineering, physics or artificial time series. As specific heuristics have been developed for distinct applications, we limit our analysis to publications to the sub-domain of sales forecasting, in order to derive robust recommendations for valid and reliable experiments.

As a keyword search does not allow a valid separation of publications, we manually evaluated and eliminated all publications related to non-business and non-sales applications and nominal predictions or classifications, e.g. of marketing response rates, market shares, index prices, electrical load forecasting etc. and different application domains in order to derive homogeneous recommendations.

We identified additional studies through a bottom-up follow up of citations and additional searches by author name

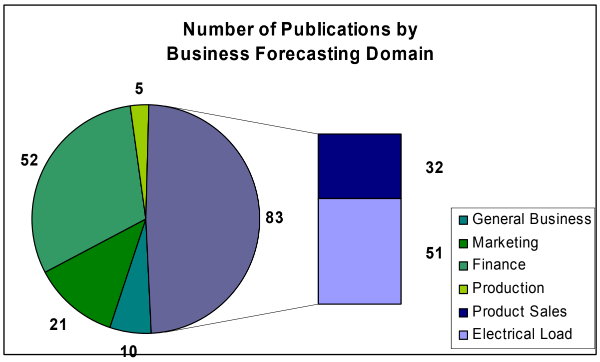

Fig.2: Distribution of Publications in the Business Forecasting Domain

and research group identified in IEEE Xlpore, ScienceDirect, Citeseer, Proquest and ABI/Inform. Our search yielded a total of 171 publications on time series or causal point predictions in business forecasting, in the categories of general publications on point predictions including examples from business, sales, marketing, production and finance. In the domain of sales forecasting, only 32 publications focussed on sales forecasting of discrete products or services [6, 7, 19-48] instead of electrical load forecasting etc. This limited number of publications dedicated to the use of ANN in business forecasting for demand planning verifies previous studies [11] and explains the large attention the sparse publications in the field receive. For example, the International Journal of Forecasting ranked publications on NN as the second and fourth most requested papers in 2001 and 2002 [49].

III. EVALUATION FRAMEWORK

Following, we propose an evaluation framework for the application of ANN in business forecasting based upon theoretical findings and empirical knowledge to allow a distinction between superior publications prepared with necessary scientific rigor in experiment design, execution and documentation from less objective results. In addition, we aim to integrate principles from the extensive experiences and publications on forecasting competitions in the business forecasting domain [50-52]

The derived framework may serve a variety of purposes. Firstly, the systematic evaluation of previous publications on NN for sales forecasting identifies those publications with superior quality in order to enhance progress in ANN applications on their findings and to extract mutual best practice from the group of leading modelling approaches. Also, it may direct further research into the systematic re-evaluation, confirmation and refinement of superior heuristics and the re-publication of prominent but biased studies, to prevent further reuse of suboptimal designs, e.g. [53-55]. In addition, it may help to identify gaps in current research and practice through dedicated research on common or neglected problems. Moreover, users of ANNs with limited experience could use the experiment design of those studies identified superior as a guideline for their individual modeling process, avoiding common mistakes based upon suboptimal and misleading publications. Finally, the framework may serve as a guideline for robust experimental designs based on “forecasting principles” for data pre-processing, architecture selection, training and documentation of ANN experiments [50].

The framework aims to evaluate the validity and reliability of a publication in two dimensions: an objective degree of reliability through documentation and a subjective degree of validity based upon the application of verified empirical knowledge. We follow a simple, unweighted scoring model, giving points for reliable documentation and application of rules to ensure validity, to derive a classification matrix of reliable versus unreliable and valid versus invalid publications. The dimensions are described below in detail.

A. Degree of reliability through documentation

In order to disseminate experiences and knowledge on modeling ANNs for business forecasting the experiments must be reliable. To assure reliability, published experiments should allow their unbiased interpretation, reevaluation, reconstruction and consequently consistent results in re-simulation through sufficient documentation. Following, we consider reliability to be the degree, to which independent experiments using the same network, datasets, parameters and modeling heuristics lead to consistent results limiting measurement or documentation errors. This necessitates a sufficient documentation of all free parameters and heuristics applied in model building, training and application.

We developed 5 categories with 30 criteria to sufficiently characterize an ANN forecasting experiment

1) Data Selection and Preprocessing
- a) *Data type & origin documented*
- b) *Uni- (time series) or multivariate data (causal)*
- c) *Removal of trend and/or seasonality from data**
- d) *Scaling of Data**
- e) *Amount of observations used for training*
- f) *Amount of observations used for validation*
- g) *Amount of observations used for testing*

2) Network Architecture
- *a) Number of input units used for all architectures*
- *b) Number of hidden units used for all architectures*
- *c) Number of output units used for all architectures*
- d) *Number of layers of nodes used from in- to output*
- e) *Transfer function used in hidden layers*
- f) *Output function used in output layer*
- g) *Shortcut connections (input → output layer) used**
- h) *Iterative t+1 or multiple-step-ahead forecasts*
- i) *Architecture (in-hidden-out) of best network*

3) Network Training
- a) *number of lagged & explanatory variables*
- b) *Number of iterations / epochs of data presentation*
- c) *Training method used for parameterization*
- d) *Objective / error function used for minimization*
- e) *(Early) stopping criteria used in training **
- f) *Number of re-initializations for successive training*
- g) *Size of learning rate*
- h) *Size of momentum term*

4) Network Architecture Selection
- a) *Method used to find optimal network **
- b) *Method to find the "optimum" number of input units*
- c) *Heuristic to determine ratio input to hidden units**

5) Network Evaluation
- a) *Training Error and/or Validation error*
- b) *Test error for generalization*
- c) *Robustness of errors through initializations*

Each criterion is attributed one point if reported and documented. For attributes allowing alternatives or choices, e.g. in preprocessing data through optional scaling or trend-removal, criterion is marked with * and a point is accredited if a method was used and the name or description of the method used was given as well as stating that no scaling had taken place, with no point given if scaling was not mentioned at all. During evaluation, no assumptions are made; e.g., if an author reports using the backpropagation learning algorithm, we do not presume use of a logistic transfer function as originally proposed by [56]. In order to gain all points, the learning rule used and the transfer function must be reported separately. The maximum score achievable is 30 points for complete documentation of the experiments.

B. Degree of validity through experiment design

As a second criterion, we attempted to determine the degree of validity of the experimental setup and results for the field of business forecasting in sales or demand planning. We consider a publication as valid based upon its adherence to empirical rules firmly established in ANN modeling and business forecasting, such as comparison to alternative forecasting methods, use of hold-out data for the final evaluation etc. We derive 2 categories of 14 criteria, allowing a maximum of 14 points.

1) Validity of Forecasting Experiment Setup

a) *Comparison of ANNs with established methods*
Relative performance between competing ANN models must be evaluated against standard statistical methods predominant in the empirical domain, such as naïve, exponential smoothing or ARIMA methods [11].

b) *Use of real time series data*
In order to evaluate the performance of heuristics in real world problems we need to evaluate representative time series. Synthetic data does not ensure equivalent complexities or properties as generated by real markets.

c) *Public availability of data material*
If the data is not publicly available, even if rescaled to adhere to privacy regulations, a verification of the results is impossible. Therefore, an evaluation of heuristics on public data may further it's validity.

d) *Use of alternative non-quadratic error measures*
Different error measures induce different biases in evaluation forecasting performance. Consequently, various error measures should be used for evaluation, while quadratic error measures should be avoided.

e) *Out-of-sample evaluation of ANN performance*
Data must be split up in distinct sets for parameterization and ex ante testing, requiring three datasets of training, validation and test data for early stopping [11]. Despite lack of empirical rules the training set should consist no less than 50% of the data [23, 57]

f) *Evaluation of multiple time origins*
Due to the variance in data, different models may be selected depending on the chosen time origin to split

training, validation and test dataset. Consequently, performance at different time origins must be averaged.

g) *Evaluation of various forecasting horizons*
In case of a forecasting horizon requiring more than a *t+1* forecast, the forecasting performance should be evaluated by an iterative *t+1* forecast based upon previous predictions or a simultaneous multiple step ahead forecast *t+1, t+2,…, t+n.*

2) Validity of Neural Network Experiment Setup

a) *Systematic evaluation of different ANN topologies*
Different ANN topologies instead of a single, arbitrarily chosen architecture should be evaluated, as the ANN design significantly impacts its performance. Evaluating and selecting the best ANN represents the dominant problem in ANN modeling.

b) *Scaling of data prior to ANN training*
Scaling input data towards the upper and lower bound of the transfer function is recommended to facilitate convergence in training [7, 58, 59].

c) *Use of more than 50 observations in testing*
Even for stationary time series training and evaluation of the ex-post performance on less than 50 observations questions the statistical validity and reliability due to the high sampling variability from noise in real data series[11]. Also, the issue of overparameterization must be considered [23, 60]

d) *Use of multiple initializations for each ANN model*
In order to account for randomized starting weights and local search methods for training each ANN model must be initialized, trained and evaluated a couple of times to allow robust network selection.

e) *Convergence of ANN model*
The ANN needs to demonstrate robust minimization of the objective function over the training set and the validation set to show that it has continuously learned to underlying generator in the data. [11]

f) *Generalization of ANN model*
The ANN needs to demonstrate robust minimization on the hold-out test data in addition to convergence, to demonstrate lack of overfitting and its ability to extrapolate values outside the learned datasets [11].

g) *Stability of ANN results*
The criterion of stability evaluates the robustness of the trained model through evaluation on different test datasets from re- or multiple sampling [11].

The criteria selected for the degree of validity impose certain subjectivity to our scoring model. However, most criteria represent proven or established recommendations.

IV. APPLICATION OF THE EVALUATION FRAMEWORK

We applied the evaluation framework developed above to the 32 publications on sales forecasting using ANNs. The results proved disappointing. Only 6 publications (19%) achieved the highest score of 61-70% reliability, with half the publications achieving less than or exactly 50% reliability. On the other hand, only 1 publication (3%) achieved less than 30% reliability, showing a large number of partially documented experiments. This confirms a common criticism in ANN modeling that most authors do not document their studies well enough, prohibiting a remodeling of experiments or successful ANN architectures in similar problems. Table 1 gives an overview of the frequencies encountered.

Tab. 1: Degree of reliability by class and frequency.

Summery - Degree of documentation

Class	*Frequency*	*Cumulative %*	*Class*	*Frequency*	*Cumulative %*
0%	0	0,00%	60%	8	25,00%
10%	0	0,00%	50%	7	46,88%
20%	1	3,13%	30%	6	65,63%
30%	6	21,88%	70%	6	84,38%
40%	4	34,38%	40%	4	96,88%
50%	7	56,25%	20%	1	100,00%
60%	8	81,25%	0%	0	100,00%
70%	6	100,00%	10%	0	100,00%
80%	0	100,00%	80%	0	100,00%
90%	0	100,00%	90%	0	100,00%
100%	0	100,00%	100%	0	100,00%

Analyzing the degree of validity we find similar results. Over 80% of all publications (84%) did not reach half of the potential score. Even worse, no publication achieved more than 60% of all points, questioning the scientific rigor applied in modeling ANN for sales forecasting. The results are displayed in Table 2.

Tab 2: Degree of validity by class and frequency

Summery - Degree of validity

Class	*Frequency*	*Cumulative %*	*Class*	*Frequency*	*Cumulative %*
0%	0	0,00%	30%	9	28,13%
10%	0	0,00%	40%	8	53,13%
20%	2	6,25%	50%	8	78,13%
30%	9	34,38%	60%	5	93,75%
40%	8	59,38%	20%	2	100,00%
50%	8	84,38%	0%	0	100,00%
60%	5	100,00%	10%	0	100,00%
70%	0	100,00%	70%	0	100,00%
80%	0	100,00%	80%	0	100,00%
90%	0	100,00%	90%	0	100,00%
100%	0	100,00%	100%	0	100,00%

In combining both evaluations, we derive a portfolio matrix to determine the overall quality of publications in ANN for sales forecasting, as displayed in fig. 3.

In addition to the findings above, a correlation between reliable and valid publications becomes evident, identifying that those publications with a high degree of documentation also took common recommendations and best practices in modeling into account. Consequently, the best publications [19, 61, 62] may be preferred as an introduction for sales forecasting problems and to derive modeling recommendations.

Some of our findings seem particularly worth mentioning. The percentages given refer only to the number of publications who did report the criteria in question. Most publications focus on standard network architectures: 28

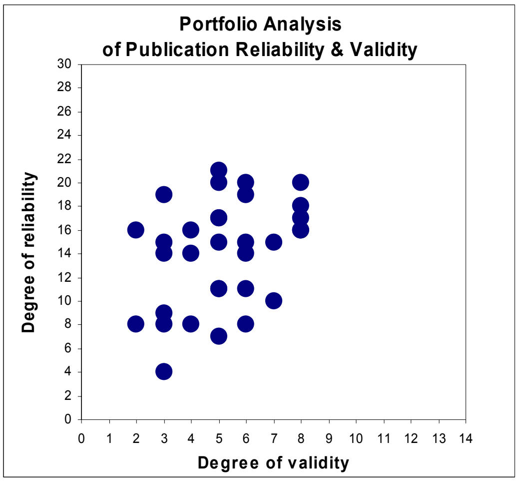

Fig.3: Portfolio Analysis of the evaluation framework

authors (100%) used at least one three-layered multilayer perceptron in one of their experiments, with 21 publications (84%) applying the standard backpropagation learning rule and 29 authors (94%) using only 1 output unit and one-step-ahead forecasts. In internal processing, 14 authors (93%) preferred the sigmoid transfer function in hidden nodes with 50% of the authors using the sigmoid function as an output function, 42% using the linear function and 8% applying the hyperbolic tangent. Interestingly, although 24 authors (96%) constructed and compared more than one ANN topology, only 13 (41%) reported their best network topology, questioning the relevance of the published results for others.

In 26 publications (81%) the ANN results were compared to other forecasting methods, such as ARIMA and regression models, as a benchmark, with ANNs outperforming the other methods in 24 publications (92%). However, dominance of ANNs over conventional methods may not be derived from this, especially as two of the three best studies in reliability and validity show only mixed results.

Considering preprocessing of data, only 15 (47%) of the publications reported scaled data, with only 3 publications (30%) removing trend and seasonality, 6 (60%) removing trend and/or seasonality for some experiments and 1 (10%) stating that trend and seasonality were not removed on purpose. 13 authors (40%) reported that the data used is publicly available. Most surprisingly, only 4 publications (13%) reported the essential splitting of data into training, validation and test set for their experiments, significantly questioning their validity.

Most authors evaluated their predictions by only one error measure, with no error measure dominating for the ex-post evaluation. The mean-squared error, mean square percentage error, rooted mean square error, mean absolute error, mean absolute percentage error and other error measures were used, despite obvious shortcomings of using only one error measure, and especially squared error measures, despite a substantial discussion in the business forecasting domain. Most disappointingly, no publication reported the ANNs ability to generalize and only 2 (6%) reported that their results were stable without documenting to what extent.

While it is not surprising, that most ANN research originating in the disciplines of Physics, Engineering and biological Neuroscience are unaware of basic principles in Business Forecasting [50], leading to invalid publications, most publications in the domain of sales forecasting itself seem to omit research findings from the application domain.

V. CONCLUSION

We developed an evaluation framework to judge the overall quality of publications in the field of business forecasting with an evaluation in the domain of sales forecasting for demand planning. Our analysis demonstrates that most publications lack the necessary degree of documentation as well as scientific rigor to allow valid and reliable analysis of competing ANN architectures, learning schemes and heuristics. The majority of the relevant publications do not allow a re-simulation from the published experiment design, questioning their findings. In addition, most publications show a lack of validity, not following necessary policies when modeling and implementing ANNs.

Combining the findings of reliability and validity, a correlation of the degree of rigor used in documentation with the degree of validity becomes evident, hinting that some authors follow scientific approach to publishing ANN experiments while others endanger the progress in research and application of ANNs for forecasting through suboptimal experiments.

For future publications, we recommend an increased emphasis on documenting models, modeling process, experiment design and results as well as taking common knowledge from the domain of ANN modeling and business forecasting into consideration. We hope that our framework may contribute as a guide to future experiment designs, disseminate knowledge between the domains and to pinpoint publications with high validity and reliability for future applications and research.

VI. REFERENCES

[1] L. C. Jain and V. Vemuri, *Industrial applications of neural networks*. Boca Raton: CRC Press, 1999.

[2] B. Kappen and S. Gielen, *Neural networks : best practice in Europe : proceedings of the Stichting Neurale Netwerken Conference 1997, Amsterdam, the Netherlands, 22 May 1997*. Singapore ; River Edge, N.J.: World Scientific, 1997.

[3] P. J. G. Lisboa, B. Edisbury, and A. Vellido, *Business applications of neural networks : the state-of-the-art of real-world applications*. Singapore ; River Edge, N.J.: World Scientific, 2000.

[4] F. M. Thiesing, *Analyse und Prognose von Zeitreihen mit neuronalen Netzen*, Als Ms. gedr ed. Aachen: Shaker, 1998.

[5] F. M. Thiesing and O. Vornberger, "Forecasting Sales Using Neural Networks Computational Intelligence," presented at

Proceedings International Conference, 5th Fuzzy Days, Dortmund, Germany, 1997.
[6] F. M. Thiesing and O. Vornberger, "Sales forecasting using neural networks," presented at Neural Networks,1997., International Conference on, 1997.
[7] Z. Tang and P. A. Fishwick, "Feed-forward Neural Networks as Models for Time Series Forecasting," *ORSA Journal on Computing*, vol. 5, pp. 374-386, 1993.
[8] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, vol. 2, pp. 359 - 366, 1989.
[9] C. Chatfield, "Neural Networks - Forecasting Breakthrough or Passing Fad," *International Journal of Forecasting*, vol. 9, pp. 1-3, 1993.
[10] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35-62, 1998.
[11] M. Adya and F. Collopy, "How effective are neural networks at forecasting and prediction? A review and evaluation," *Journal of Forecasting*, vol. 17, pp. 481-495, 1998.
[12] A. Lapedes, R. Farber, and Los Alamos National Laboratory., "Nonlinear signal processing using neural networks prediction and system modelling." Los Alamos, NM: Los Alamos National Laboratory, 1987.
[13] A. S. Weigend, "Predicting Sunspots and Exchange Rates with connectionist Networks," in *Nonlinear modeling and forecasting*, M. Casdagli and S. Eubank, Eds. New York, 1992, pp. 395-432.
[14] C. de Groot and D. Wurtz, "Analysis of univariate time series with connectionist nets: A case study of two classical examples," *Neurocomputing*, vol. 3, pp. 177-192, 1991.
[15] G. E. P. Box and G. M. Jenkins, *Time series analysis; forecasting and control*. San Francisco,: Holden-Day, 1970.
[16] K. Krycha, *A comparison of standard methods with a neural network approach to forecast univariate time series*. Wien: Institut für Höhere Studien, 1992.
[17] B. K. Wong, T. A. Bodnovich, and Y. Selvi, "Neural network applications in business: A review and analysis of the literature (1988-1995)," *Decision Support Systems*, vol. 19, pp. 301-320, 1997.
[18] B. K. Wong, V. S. Lai, and J. Lam, "A bibliography of neural network business applications research: 1994-1998," *Computers & Operations Research*, vol. 27, pp. 1045-1076, 2000.
[19] K. P. G. Alekseev and J. M. Seixas, "Forecasting the air transport demand for passengers with neural modelling," presented at Proceedings VII Brazilian Symposium on Neural Networks, 2002.
[20] I. Alon, M. Qi, and R. J. Sadowski, "Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods," *Journal of Retailing and Consumer Services*, vol. 8, pp. 147-156, 2001.
[21] A. P. Ansuj, M. E. Camargo, R. Radharamanan, and D. G. Petry, "Sales forecasting using time series and neural networks," *Computers & Industrial Engineering*, vol. 31, pp. 421-424, 1996.
[22] N. Baba and H. Suto, "Utilization of artificial neural networks and GAs for constructing an intelligent sales prediction system," presented at 2000 IJCNN, IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000.
[23] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, *Multivariate Analysemethoden*, vol. 10. Berlin, 2003.
[24] C.-W. Chua and G. P. Zhang, "A comparative study of linear and nonlinear models for aggregate retail sales forecasting," *International Journal of Production Economics*, vol. 86, pp. 217–231, 2003.
[25] J. Faraway and C. Chatfield, "Time series forecasting with neural networks: A case study," in *Research report 95-06 of the statistics group*. University of Bath, 1995, pp. 1-21.
[26] J. Faraway and C. Chatfield, "Time series forecasting with neural networks: A comparative study using the airline data," *Applied statistics*, vol. 47, pp. 20, 1998.
[27] H. Hippner, "Langfristige Absatzprognosen mit Neuronalen Netzen in der Automobilindustrie," in *Betriebswirtschaftliche Anwendungen des Soft Computing*, J. Biethahn, A. Hönerloh, J. Kuhl, M.-C. Leisewitz, V. Nissen, and M. Tietze, Eds. Wiesbaden, 1998, pp. 81-96.
[28] I. Jagielska, "A neural network model for sales forecasting," presented at 1993 Artificial Neural Networks and Expert Systems, New Zealand, 1993.
[29] J. J. Jiang, M. S. Zhong, and G. Klein, "Marketing category forecasting: An alternative of BVAR - Artificial neural networks," *Decision Sciences*, vol. 31, pp. 789-812, 2000.
[30] T. Kolarik and G. Rudorfer, "Time Series Forecasting Using Neural Networks," presented at Proceedings of the international conference on APL, Antwerp, Belgium, 1994.
[31] J. H. L. Kong and G. P. M. D. Martin, "A backpropagation neural network for sales forecasting," presented at 1995 ICNN, IEEE International Conference on Neural Networks, 1995.
[32] R. J. Kuo, P. Wu, and C. P. Wang, "An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination," *Neural Networks*, vol. 15, pp. 909-925, 2002.
[33] R. J. Kuo and K. C. Xue, "A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights," *Decision Support Systems*, vol. 24, pp. 105-126, 1998.
[34] R. J. Kuo and K. C. Xue, "Fuzzy neural networks with application to sales forecasting," *Fuzzy Sets and Systems*, vol. 108, pp. 123-143, 1999.
[35] J. T. Luxhoj, J. O. Riis, and B. Stensballe, "A hybrid econometric-neural network modeling approach for sales forecasting," *International Journal of Production Economics*, vol. 43, pp. 175-192, 1996.
[36] K. Nam and T. Schaefer, *Forecasting international airline passenger traffic using neural networks*, 1995.
[37] K. Nam, J. Yi, and V. R. Prybutok, "Predicting airline passenger volume," *The Journal of Business Forecasting Methods & Systems*, vol. 16, pp. 14-16, 1997.
[38] H. L. Poh and T. Jasic, "Forecasting and analysis of marketing data using neural networks: a case of advertising and promotion impact," presented at Artificial Intelligence for Applications, 1995. Proceedings., 11th Conference on, 1995.
[39] T. Ragg, W. Menzel, W. Baum, and M. Wigbers, "Bayesian learning for sales rate prediction for thousands of retailers," *Neurocomputing*, vol. 43, pp. 127-144, 2002.
[40] F. M. Thiesing, U. Middelberg, and O. Vornberger, "Parallel Back-Propagation for Sales Prediction on Transputer Systems," presented at Proceedings World Transputer Congress '95, Harrogate, UK, 1995.
[41] F. M. Thiesing, U. Middelberg, and O. Vornberger, "A Neural Network Approach for Predicting the Sale of Articles in Supermarkets," presented at 3rd European Congress on

Intelligent Techniques and Soft Computing, Aachen, Germany, 1995.
[42] F. M. Thiesing, U. Middelberg, and O. Vornberger, "Abverkaufsprognose mit paralleler Backpropagation," in *Fuzzy-Neuro-Systeme '95*. TH Darmstadt, 1995, pp. 373-380.
[43] F. M. Thiesing, U. Middelberg, and O. Vornberger, "Short Term Prediction of Sales in Supermarkets," presented at ICNN'95, Perth, WA, 1995.
[44] F. M. Thiesing and O. Vornberger, "Time Series Prediction by Neural Networks," in *Power Xplorer User Report*: Parsytec GmbH, 1995, pp. 33-36.
[45] F. M. Thiesing and O. Vornberger, "Abverkaufsprognose im Supermarkt mit Neuronalen Netzen.," in *Betriebswirtschaftliche Anwendungen des Soft Computing*, J. Biethahn, A. Hönerloh, J. Kuhl, M. C. Leisewitz, V. Nissen, and M. Tietze, Eds.: Vieweg, 1998, pp. 115-126.
[46] S. Thornassey, M. Happiette, and J. M. Castelain, "Three complementary sales forecasting models for textile distributors," presented at Systems, Man and Cybernetics, 2002 IEEE International Conference on, 2002.
[47] D. H. F. Yip, E. L. Hines, and W. W. H. Yu, "Application of artificial neural networks in sales forecasting," presented at Neural Networks,1997., International Conference on, 1997.
[48] M. Yohda, M. Saito-Arita, A. Okada, R. Suzuki, and Y. Kakemoto, "Demand forecasting by the neural network with discrete Fourier transform," presented at Data Mining, 2002. ICDM 2002. Proceedings. 2002 IEEE International Conference on, 2002.
[49] Elsevier, "Top requested papers of the International Journal of Forecasting," vol. 2003. online www, http://www.elsevier.com/homepage/sae/econworld/econbase/intfor/frame.htm: Elsevier, 2003.
[50] J. S. Armstrong, "Principles of forecasting: a handbook for researchers and practitioners," in *International series in operations research & management science ; 30*. Boston ; London: Kluwer Academic, 2001, pp. xii, 849.
[51] R. Fildes, M. Hibon, S. Makridakis, and N. Meade, "Generalising about univariate forecasting methods: further empirical evidence," *International Journal of Forecasting*, vol. 14, pp. 339-358, 1998.
[52] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *International Journal of Forecasting*, vol. 16, pp. 451-476, 2000.
[53] T. Hill, M. O'Connor, and W. Remus, "Neural network models for time series forecasts," *Management Science*, vol. 42, pp. 1082-1092, 1996.
[54] J. Faraway and C. Chatfield, "Time series forecasting with neural networks: A comparative study using the airline data," *Journal of the Royal Statistical Society Series C-Applied Statistics*, vol. 47, pp. 231-250, 1998.
[55] W. Remus and M. O'Connor, "Neural networks for time-series forecasting," in *Principles of forecasting: a handbook for researchers and practitioners*, J. S. Armstrong, Ed. Boston ; London: Kluwer Academic, 2001, pp. 245-258.
[56] D. E. Rumelhart, J. L. McClelland, and University of California San Diego. PDP Research Group., *Parallel distributed processing : explorations in the microstructure of cognition*. Cambridge, Mass.: MIT Press, 1986.
[57] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35-62, 1998.
[58] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, vol. 10, pp. 215-236, 1996.
[59] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling and Software*, vol. 15, pp. 101-124, 2000.
[60] Z. Tang, C. De Almeida, and P. A. Fishwick, "Time series forecasting using neural networks vs. Box-Jenkins methodology," *Simulations*, vol. 57, pp. 303-310, 1991.
[61] J. Faraway and C. Chatfietd, "Time series forecasting with neural networks: A comparative study using the airline data," *Applied statistics*, vol. 47, pp. 20, 1998.
[62] J. Faraway and C. Chatfield, "Time series forecasting with neural networks: A case study.," in *Research report 95-06 of the statistics group*. University of Bath, 1995, pp. 1-21.

---

[1] The ISI web of science includes 25 million citations from the Science Citation Index Expanded, Social Sciences Citation Index and Arts & Humanities Citation Index from 1980-2003, excluding limited publications from 2004. As no relevant publications are expected prior to the 1986 development of the MLP and the following surge in research interest the ISI-WoS coverage seems representative.

[2] The search term was [title=(neural AND net* AND (forecast* OR predict* OR time-ser* OR time w/2 ser* OR timeser*))] The relevant wildcard were considered.