2nd Conference on Production Systems and Logistics

Pre-Selection Of Suitable Regression Methods For The Determination Of Interactions And Forecasts In Global Production Networks

Günther Schuh[1], Andreas Gützlaff[1],
Julian Ays[1], Niklas Rodemann[1], Miha Podbreznik[1]
[1]*Laboratory for Machine Tools and Production Engineering WZL of RWTH Aachen University, 52074 Aachen, Germany*

Abstract

The locations of many manufacturing companies are distributed globally. This has led to the development of historically grown global production networks whose structure is often very complex, not transparent and influenced by many factors. The high number, as well as the volatility of the influencing factors and dependencies in the network additionally, complicate the network configuration. As a result, adaptation needs and optimization possibilities are recognized too late or not at all. In order to enable early recognition of saving potentials, active monitoring and analysis of changes and dependencies of the influencing factors on the production network is needed. The necessary consideration of a multitude of influencing factors requires further tools to be manageable by the network planner. Therefore, databased methods can be used as support for the forecast and the determination of dependencies of influencing factors. In other research fields, regression analysis is an established method for a databased analysis. This paper focuses on the use of regression analysis in global production networks. It is essential for an accurate analysis, to choose the right regression method out of the many different types in existence. A systematic literature review is conducted to establish an overview of regression methods used in other research fields. A search strategy is developed and implemented and the key findings of the literature review are derived and evaluated. In the second step, a new approach for the pre-selection of suitable regression methods for the determination of interactions and forecasts in global production networks is proposed.

Keywords

Regression Analysis; Global Production Network; Influencing Factors; Network Design

1. Introduction

Globally located production sites are interconnected with material, information and financial flows [1,2]. The challenge for companies is to manage their interconnected network of locations effectively and efficiently [2]. The growing complexity of global production networks complicates the management of the network and its quick adaptation to changes [3]. To achieve transparency and subsequently to overcome the complexity, an understanding of the network is of importance. Global production networks are characterized and defined by a multitude of influencing factors [3]. These influencing factors include internal data (e.g. production cost, production capacity, etc.) and external data (e.g. changes in local taxes, inflation, competitor's action, etc.) [4,5]. With a better understanding and forecast of influencing factors, predictions about the whole network can be made. This is why data-based analysis in global production networks is gaining importance [6]. A possibility for a data-based forecast of influencing factors is the use of regression analysis. Regression analysis is a data mining class [7], used for not only predictions and forecasts, but also the determination of causal relations between dependent and independent variables [8]. The distinctive differentiation between regression methods is in linearity [7,9,10]. Regression is widely researched and implemented in other research areas, e.g., it has been used in forecast models for economic and financial models [11], as evidence in court cases and as support for legislation [8]. It is the most widely used statistical approach in medical research [12]. COE ET AL. present the parallels between global production networks and the global financial network and propose an integration of finance into the global production network research [13]. Because of its characteristics and a proven track record in other research areas, this paper investigates which regression methods can be used and how to select a suitable regression method in global production networks. Regression methods can then be used to determine the interactions and forecasts in the network, e.g., forecasts of costs or determination of interdependencies between production sites.

2. State of the art

This chapter presents the state of the art approaches in the determination of interactions and forecasts in global production networks. Until now, the main influencing factors have already been identified and described. ABELE ET AL. identify and describe the main influencing factors in global production networks [14]. Many other authors develop catalogues of influencing factors such as NEUNER [15]. In this context, some authors focus specifically on methodologies for the network configuration. UDE develops a decision support method for the configuration and evaluation of global production networks [16]. REUTER ET AL. develop a method to evaluate performance differences between manufacturing sites, the correlations between key performance indicators and site characteristics are analysed [17]. Some publications propose and introduce databased analysis and support for decision making in global production networks. VERHAELEN ET AL. develop a methodology to support product allocation in global production networks, with the use of different data mining methods [18]. SCHUH ET AL. present a software tool for the configuration and optimization of global production networks [19]. GÖLZER ET AL. investigate the configuration and design of global production networks using Big Data [20]. MOURTZIS ET AL. propose the configuration of global production networks based on smart decision making [21]. HOCHDÖRFER ET AL. use a clustering analysis on product portfolios for the reduction of the planning complexity in global production networks [22]. MOSER ET AL. approach flexible migration planning in global production networks with the use of the Markov decision process [23]. A further step with the use of regression analysis in global production networks was made by RITTSTIEG [24]. The author investigates the cause-effect relationships of a few influencing factors with multivariant and univariant linear regression [24]. The results are not compared to other regression methods and the author does not offer a selection methodology for different regression methods [24]. TREBER presents the increase of transparency in production networks with the improvement of disruption management through increased exchange of information [25]. For the analysis, the author uses three different regression models [25]. In conclusion, the literature review indicates that there is no scientific publication investigating the potentials and the use of regression analysis in global production networks. Neither is there an established method for the selection of suitable regression models in global production networks. From numerous applications in other fields, it will be investigated which regression methods have potentials for use in global production networks.

3. Systematic analysis of suitable regression methods

To answer the research question, a systematic literature review methodology according to BRAMER ET AL. is used [26]. Systematic literature review is a method, that is effective, when used to answer a broad research question, because it represents a proven tool for summarizing many primary studies on a predefined object of investigation [26,27].

As described in the methodology, according to BRAMER ET AL., firstly a search strategy was developed, then tested, optimized, and finally reiterated [26]. Every step of the systematic search strategy was documented. As the database for the search, Web of Science was selected, since its database is, in comparison to other databases, more extensive in the fields of engineering and natural science [28]. In the next step, base search keywords were defined, expanded with synonyms, and connected with Boolean operators. To achieve the most accurate thematic fit and to limit the number of possible results, only the abstracts of the publications were used in the search. The search term was tested and optimized with multiple iterations. The final search term was:

$$AB = (regression\ AND\ (analysis\ \ OR\ method *)\ AND\ (forecast * \\ OR\ prognosis\ \ OR\ interrelation))$$

In the search, no timespan was defined, and all publication years were considered. The number of search results before filtering was 33.521.

A filtration and refinement process was defined (see Figure 1). In the first search, the used parameters were open access of publications and English as the language of the publications. After this step, the number of search results was reduced to 14.558. In the second step the results were filtered by the following Web of Science categories to achieve a thematic fit: “mathematics interdisciplinary applications”, “computer science artificial intelligence”, “computer science interdisciplinary applications”, “operations research management science”, “computer science information systems”. The initial results were evaluated, and the number of results was 343 publications, with the average citation of the publication cited being 7,43. To further reduce the number of the results and to ensure a high impact-factor, the publications were filtered according to the number of citations. The number of citations in the final selection is 7 overall citations or higher. This yielded 97 results. The list of publications was checked for errors and one publication was eliminated from the evaluation, because of unavailability. After the filtration and elimination process, the final number of evaluated publications is 96 (see Figure 1).

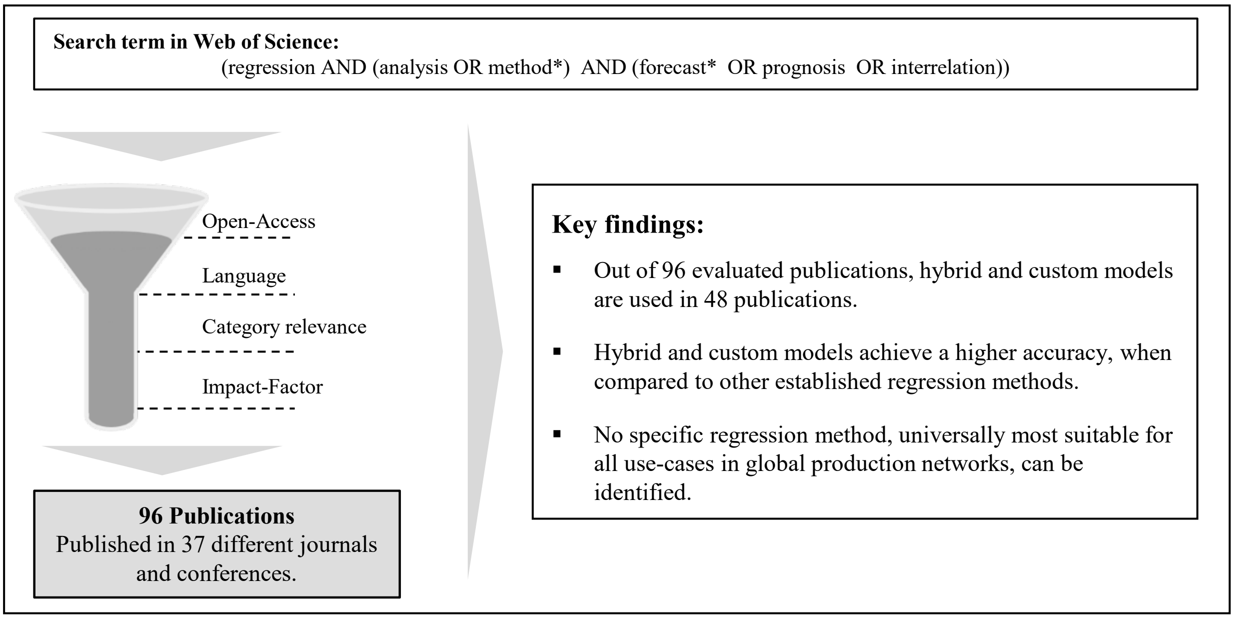

Figure 1: Overview of the systematic literature review

The publication year, of evaluated publications, ranges from the year 2003 to 2020. The total sum of citations is 2202 and the average citation is 22.94 per publication, ranging from 7 citations to 206 citations. The publications were published in 37 different journals and conferences.

Literature evaluation

The 96 evaluated publications mostly focus on the implementation of the models and methods. One publication discusses how to choose the best forecast model. Other 95 publication either describe the development of a new model or present the implementation of a model. Figure 2 shows how many times a model is used in the evaluated publications, in some, multiple methods were used and presented. In the following section, methods and models used in more than one reviewed publication are presented and a few use-cases are highlighted.

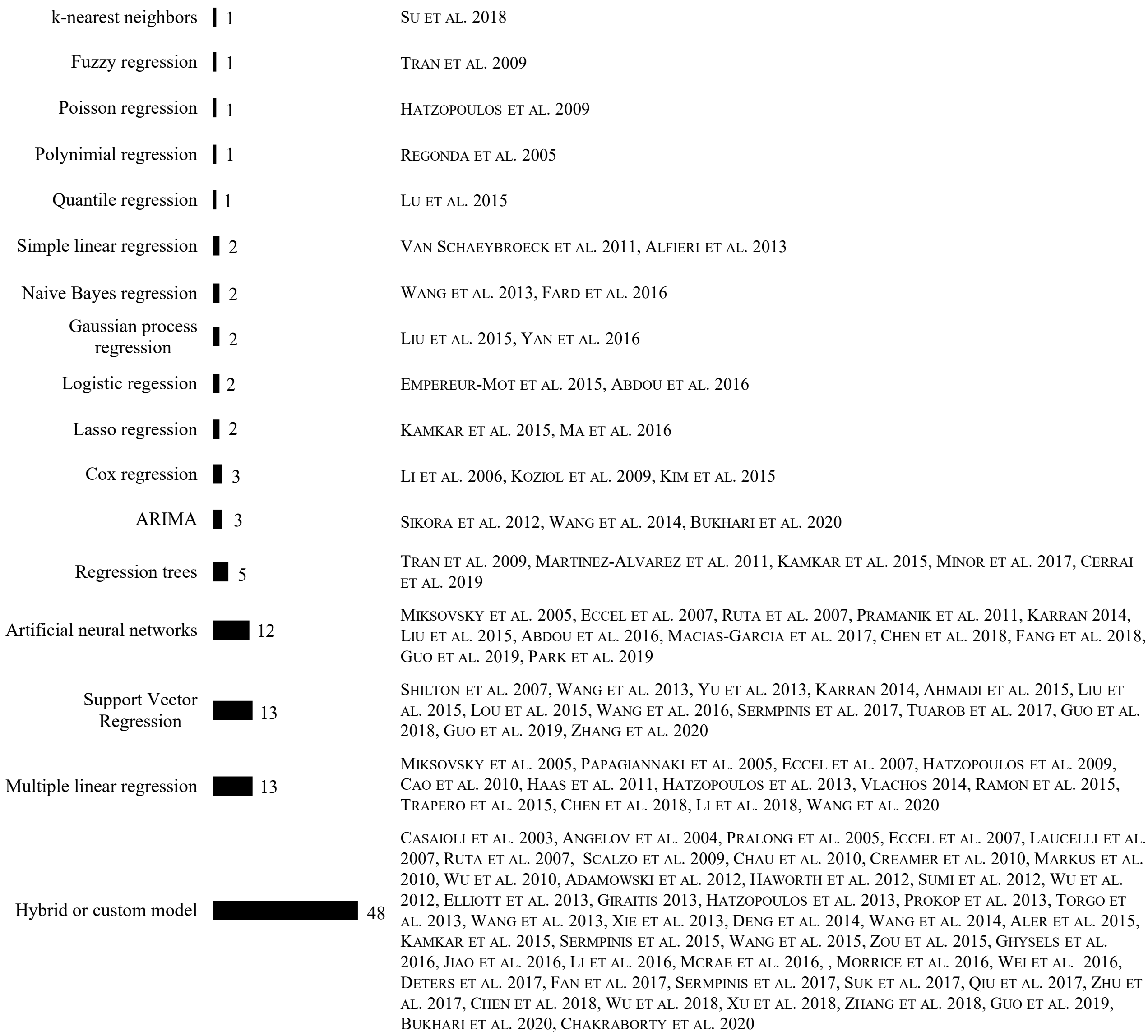

Figure 2: Summarized number of times the regression methods used in evaluated publications (N=96)

Hybrid or Custom Model

Hybrid or custom models are used in 44 publications. A hybrid model using both artificial neural networks (ANN) and support vector regression (SVR) proved a significant improvement in comparison to the use of a conventional ANN in a case for daily rainfall forecasting [29]. The hybrid model achieved the best result with filtering with single spectrum analysis. It was shown, that enumeration was more effective than correlation analysis since it can consider the nonlinear dependence between inputs and outputs [29].

In financial market forecasting, a new hybrid model was developed by combining autoregressive fractional integrated moving average (ARFIMA) and long-short term memory (LSTM) [11]. LSTM networks are a special kind of artificial neural networks. The prediction accuracy of the hybrid model was tested against autoregressive integrated moving average (ARIMA), ARFIMA and generalized regression neural network (GRNN) independently and showed 80% accuracy improvement [11]. With the high accuracy and reliability shown in the reviewed publications, hybrid and custom models offer the possibility to analyse any use-case in global production networks, since the model is adapted to the specific problem and is highly accurate.

Multiple and simple linear regression

A simple regression model is a linear regression model with a single response variable y and a single independent variable x. Multiple linear regression is an expansion of the simple linear regression and is used when more than one independent variable is given. [30]

A general multivariate regression model is a linear model consisting of k independent variables and n response variables, it can be represented using a matrix [31]. Linear regression can offer a simple approach to the analysis of influencing factors in global production networks, when dealing with simple and homogenous linear datasets. Linear methods are generally more susceptible to outliers but can be made more robust by regularization parameters [10].

Support Vector Regression

The support vector machines can be generalized to be used for regression. SVR is effective in real-value function estimation. It is a supervised learning method and it uses a symmetrical loss function, which equally adjusts high and low misestimates. The margin of tolerance is symmetrically placed around the function [32]. SVR have an excellent generalization capability with high accuracy of forecasts [32]. Due to its characteristics, it could be used to analyse influencing factors in global production networks with a complex and inhomogeneous dataset.

Artificial Neural Networks

The development of artificial neural networks was motivated by the information processing of the human brain [30]. An artificial neural network consists of neurons and connectors, which build a network. The classes of networks are mainly distinguished by the different network topologies and connection types, such as single-layer, multilayer, feed-forward or feed-back networks [30]. RUTH ET AL. use a robust ensemble of neural network regressors with smoothing of the output signal [33]. This composite of neural networks was tested against others in multiple competitions. Artificial neural networks are a possible method for the analysis of complex influencing factors and large datasets in global production networks.

Regression Trees

The regression tree can be considered a type of tree-based algorithm, with a continuous response variable [30]. In a publication, the M5P algorithm is used to create regression trees for improved earthquake prediction [34]. The M5P algorithm is a binary regression tree model where the last nodes are the linear regression functions that can produce continuous numerical attributes. It is a further development of the M5 algorithm introduced by QUINLAN [35]. This tree algorithm can handle high dimensionality tasks. The regression trees method could offer a use case in the event prediction of influencing factors in global production networks.

ARIMA

The autoregressive integrated moving average (ARIMA) model is mostly used for time series forecasting models. WANG ET. AL use the ARIMA model for precipitation simulations [36]. ARIMA models are widely used to calculate monthly time series with yearly variations. The authors slightly modify the model to account for inter-monthly variations, which are, as the authors suggest, often ignored. This modification improves the accuracy by 21%. It is concluded, that the ARIMA approach can be further improved with the use of neural networks and support vector machines [36]. This approach could be used for the forecast of influencing factors in global production networks since the model is fitted to time series data to better understand the data or to forecast data points.

Cox Regression

The Cox proportional-hazards model is a regression model, commonly used in medical research for the investigation of the association between the time, a specified event takes to happen, and other variables. The Cox regression approach has drawbacks since it does not enable the identification of interactions between variables [37]. KIM ET AL. propose an improved framework as an improvement to the traditional Cox model [38]. Although most of this method's use cases are in the field of medical research, it could, with some modifications, be used in global production networks.

Lasso Regression

Least absolute shrinkage and selection operator (Lasso) is a regression method commonly used to model the predicted risk of a likely outcome. It has often been used in areas, where the number of potential predictors is large relative to the number of observations [39]. Lasso regression has been shown to outperform many standard regression methods in some settings. MA ET AL. introduce a new framework to forecast retail sales using a multistage Lasso regression [40]. The improved framework results in significantly higher accuracy. This method could be of use in global production networks when large influencing factor constructs are used.

Logistic Regression

The logistic regression uses a logistic function to model a binary output variable [41]. The logistic regression model uses logistic regression to predict the odds of the binary outcome [41]. EMPEREUR-MOT ET AL. use logistic regression to calculate activity probabilities in clinical virtual screening methods [42]. In global production networks, this method could be used, for the odds prediction of event occurrence.

Gaussian Process Regression

Gaussian process regression is used, to profit from the normal distribution. Its property as a machine learning method allows automatic model building based on observations. The result is a probability distribution of possible interpolation functions and the solution with the highest probability [43]. A Gaussian process regression model has been introduced for forecasting the time series of wind energy [44]. The possibility of the forecast of influencing factors makes it a potential method for use in global production networks.

Naive Bayes Regression

Naive Bayes is usually used for classification tasks and is often more reliable than more sophisticated classification methods [45]. Naïve Bayes can also be used for regression. FARD ET AL. introduce new models with high accuracy for event prediction in longitudinal data [46]. Due to the high accuracy shown in the reviewed publications, naive Bayes regression could, when adapted, be used for the analysis of influencing factors in global production networks.

Key Findings

The key findings presented in Figure 1 were derived from the review of the publications. Other research areas use different regression methods for different use-cases and specific data sets. There is no specific method, that prevails in the number of uses. Hybrid and custom models are used in 48 publications. These differentiate themself from each other in the combination of models used and the structure of the model as well as the result expected. The highest accuracy achieved models, which were adapted and customized to a specific use-case or problem, because of their adaptation to the specific dataset and the expected result. Hybrid and custom models, which were compared in reviewed publications with other established methods, achieved a higher forecast and prediction accuracy. Because of different data sets, different influencing factors and different use-cases in global production networks, a universally best method for use in global production networks cannot be identified. Due to the different data sets of influencing factors, different regression methods could be most suitable for the analysis, with hybrid and custom models being the most accurate and most effort-intensive to implement.

4. Selection of suitable regression methods

As identified in section 3, there is no universally best regression method in global production networks. The analysed methods are suitable for different use cases depending on the data basis and the result requested. As accuracy is not the only criterion to be considered, a structured approach to the selection of a suitable model should be implemented.

This paper proposes a novel methodical approach to the pre-selection of regression methods (see Figure 3). The pre-selected regression methods, reviewed in the systematic literature review, are used as support.

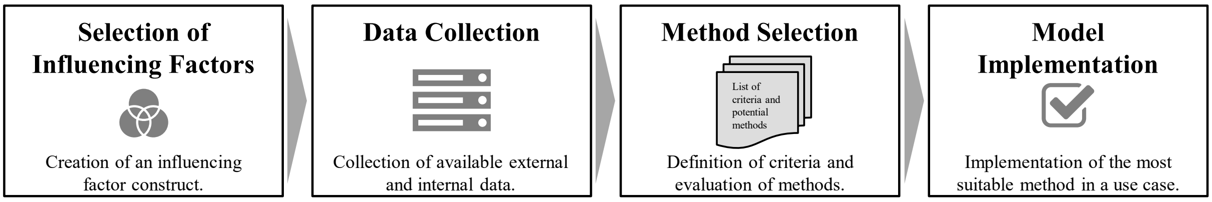

Figure 3: Proposed methodology for the method pre-selection

In the first step, all available internal and external data should be collected needed to describe a construct consisting of influencing factors, which were selected beforehand. Subsequently, the data is screened and reviewed and afterwards prepared for the analysis. If the collected data is homogenous, and there are little changes and outliers in data, and errors in forecasts are of little importance, a simple linear method or a method most convenient for the user can be chosen. Otherwise, a list of potential methods should be created and evaluated and structurally judged by experts according to criteria for a specific use-case. The criteria of YOKUMA ET AL. can be used as a guideline. YOKUMA ET AL. analyse two studies, where expert opinion, on the criteria for the selection of forecasting models, was examined and summarized in the following ranked list (from most to least significant): accuracy, cost savings from improved decisions, ease of interpretation, flexibility, ease of using available data, ease of use, quickly provided forecast [47].

Although accuracy is listed as the most important in the expert opinion, other criteria are nevertheless relevant and should be considered together in the selection and the development of new methods and models. For example, the aspects of convenience, market popularity, corporate guidelines and relative track record can be considered [48]. The criteria for the selection should be specific, weighted and can differentiate from use-case to use-case. In the creation of the list of potential methods, section 3 of this paper can be used as a guide, since these are the most relevant methods in other research areas in recent years. The best-rated model can be used, if the decisions made according to the analysis do not critical and standard model is sufficient. Otherwise, a new custom algorithm should be developed and tested against already established methods for accuracy and other selected criteria. The hybrid or custom models, used in reviewed publications, achieved, when compared to other algorithms, the highest accuracy. Potentially the highest accuracy in global production networks can be achieved with such a model. In the last step, the selected method should be implemented. The result of the implementation are forecasts and the relations between dependent and independent variables in the network.

5. Discussion and further research

The findings from the systematic literature review suggest, that the highest accuracy for regression models is achieved with hybrid or custom models, which are developed for a specific problem. The publications, that used standard regression methods and models mostly modified them and improved them for the specific use-case. One of the key findings is that there is not a universally accurate and best regression method in global production networks. The accuracy depends on many variables. One of them is available data, the more data there is available, the more accurate the forecast can be [49]. The newly developed and modified hybrid and custom models were validated with the standard regression methods and achieved a significant improvement of accuracy. As a next step, a new approach to the selection of regression methods in global production networks is proposed.

To apply regression methods in the context of global production networks, further research on the validation of the proposed methodology should be conducted. The influencing factors that have the most impact on global production networks have to be identified, characterized and defined. As a next step, a data mining model based on different regression analysis models for global production networks should be developed and implemented.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

References

[1] Schuh, G., Prote, J.-P., Gützlaff, A., Rodemann, N., 2019. Gestaltung von Produktionsnetzwerken: Verknüpfung von datengetriebenen Analysen mit Expertenwissen zur Konfiguration globaler Produktionsnetzwerke. ZWF 114 (1-2), 7–10.

[2] Arndt, T., Lemmerer, C., Biegler, C., Sihn, W., Lanza, G., 2017. Steuerung globaler Produktionsnetzwerke: Entwicklung eines Standortrollenmodells zur dynamischen Bewertung von Gestaltungsmaßnahmen. wt Wekstattstechnik online 107 (4), 241–246.

[3] Váncza, J., 2016. Production Networks, in: Produ, T.I.A.f., Laperrière, L., Reinhart, G. (Eds.), CIRP Encyclopedia of Production Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–8.

[4] Ferdows, K., 2018. Keeping up with growing complexity of managing global operations. IJOPM 38 (2), 390–402.

[5] Lanza, G., Ferdows, K., Kara, S., Mourtzis, D., Schuh, G., Váncza, J., Wang, L., Wiendahl, H.-P., 2019. Global production networks: Design and operation. CIRP Annals 68 (2), 823–841.

[6] Kwon, O., Lee, N., Shin, B., 2014. Data quality management, data usage experience and acquisition intention of big data analytics. International Journal of Information Management 34 (3), 387–394.

[7] Cios, K.J., Pedrycz, W., Swiniarski, R.W., Kurgan, L.A., 2007. Data mining: A knowledge discovery approach. Springer, New York, NY, 606 pp.

[8] Sykes, A.O., 1993. An Introduction to Regression Analysis. Coase-Sandor Institute for Law & Economics (20).

[9] Otte, R., Otte, V., Kaiser, V., 2004. Data Mining für die industrielle Praxis. Hanser, München, 389 pp.

[10] Runkler, T.A., 2010. Data Mining: Methoden und Algorithmen intelligenter Datenanalyse. Vieweg + Teubner, Wiesbaden.

[11] Bukhari, A.H., Raja, M.A.Z., Sulaiman, M., Islam, S., Shoaib, M., Kumam, P., 2020. Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. IEEE Access 8, 71326–71338.

[12] Jain, R.K., 1985. Ridge regression and its application to medical data. Computers and Biomedical Research 18 (4), 363–368.

[13] Coe, N.M., Lai, K.P.Y., Wójcik, D., 2014. Integrating Finance into Global Production Networks. Regional Studies 48 (5), 761–777.

[14] Abele, E., Dervisopoulos, M., Liebeck, T., 2006. Herausforderungen globaler Produktionsnetzwerke: Wie Unternehmen ihr Produktionsnetzwerk optimal aufstellen. wt Werkstattstechnik online 96 (4), 219–225.

[15] Neuner, C., 2009. Konfiguration internationaler Produktionsnetzwerke unter Berücksichtigung von Unsicherheit. Zugl.: Bayreuth, Univ., Diss., 2009, 1. Aufl. ed. Gabler, Wiesbaden, 316 pp.

[16] Ude, J., 2010. Entscheidungsunterstützung für die Konfiguration globaler Wertschöpfungsnetzwerke: Ein Bewertungsansatz unter Berücksichtigung multikriterieller Zielsysteme, Dynamik und Unsicherheit, 201 pp.

[17] Reuter, C., Prote, J.-P., Stöwer, M., 2016. Evaluation of Performance Differences between Manufacturing Sites. Procedia CIRP 50, 342–347.

[18] Verhaelen, B., Thomas, K., Häfner, B., Lanza, G., Schuh, G., 2017. Potenziale datenbasierter Produktallokationen, in: Wunder, T. (Ed.), CSR und Strategisches Management, vol. 114. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 96–100.

[19] Schuh, G., Potente, T., Kupke, D., Varandani, R., Hausberg, C., 2012. An Evolutionary Approach for Global Production Network Optimisation. Procedia CIRP 3, 382–387.

[20] Gölzer, P., Simon, L., Cato, P., Amberg, M., 2015. Designing Global Manufacturing Networks Using Big Data. Procedia CIRP 33, 191–196.

[21] Mourtzis, D., Doukas, M., Psarommatis, F., 2015. A toolbox for the design, planning and operation of manufacturing networks in a mass customisation environment. Journal of Manufacturing Systems 36, 274–286.

[22] Hochdörfer, J., Laule, C., Lanza, G., 2017. Product Variety Management Using Data-Mining Methods – Reducing Planning Complexity by Applying Clustering Analysis on Product Portfolios, 593–597.

[23] Moser, E., Stricker, N., Liebrecht, C., Hiller, A., Ziegler, M., Lanza, G., 2016. Preparation of Papers for IFAC Conferences & Symposia: Migration Planning for Global Production Networks using Markovian Decision Processes. IFAC-PapersOnLine 49 (12), 35–40.

[24] Rittstieg, M. Einflussfaktoren der Leistungsfähigkeit von Produktionsstandorten in globalen Produktionsnetzwerken. Dissertation, 1. Auflage ed., 266 pp.

[25] Treber, S., 2020. Transparenzsteigerung in Produktionsnetzwerken: Verbesserung des Störungsmanagements durch verstärkten Informationsaustausch, Karlsruhe, 263 pp.

[26] Bramer, W.M., Jonge, G.B. de, Rethlefsen, M.L., Mast, F., Kleijnen, J., 2018. A systematic approach to searching: an efficient and complete method to develop literature searches. Journal of the Medical Library Association : JMLA 106 (4), 531–541.

[27] Stamm, H., Schwarb, T.M., 1995. Metaanalyse. Eine Einführung. Zeitschrift für Personalforschung / German Journal of Research in Human Resource Management 1 (9).

[28] Mongeon, P., Paul-Hus, A., 2016. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 106 (1), 213–228.

[29] Chau, K.W., Wu, C.L., 2010. A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. Journal of Hydroinformatics 12 (4), 458–473.

[30] Montgomery, Douglas C. & Peck, Elizabeth A. & Vining, G. Geoffrey. Introduction to Linear Regression Analysis (Wiley Series in Probability and Statistics).

[31] Hui, E.G.M., 2019. Learn R for Applied Statistics. Apress, Berkeley, CA, 254 pp.

[32] Awad, M., Khanna, R., 2015. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. Apress Springer Science+Business Media, New York, NY, 248 pp.

[33] Ruta, D., Gabrys, B., 2007. Neural Network Ensembles for Time Series Prediction. International Joint Conference on Neural Networks.

[34] Martínez-Álvarez, F., Troncoso, A., Morales-Esteban, A., Riquelme, J.C., 2011. Computational Intelligence Techniques for Predicting Earthquakes: Martinez-Alvarez, F; Troncoso; A; Morales-Esteban, A; Riquelme, J.C; 6679, 287–294.

[35] Quinlan, J.R., 1992. Learning with Continuous Classes. AI, 343–348.

[36] Wang, H.R., Wang, C., Lin, X., Kang, J., 2014. An improved ARIMA model for precipitation simulations. Nonlin. Processes Geophys. 21 (6), 1159–1168.

[37] Gui, J., Moore, J.H., Kelsey, K.T., Marsit, C.J., Karagas, M.R., Andrew, A.S., 2011. A novel survival multifactor dimensionality reduction method for detecting gene-gene interactions with application to bladder cancer prognosis. Human genetics 129 (1), 101–110.

[38] Kim, D., Li, R., Dudek, S.M., Ritchie, M.D., 2015. Predicting censored survival data based on the interactions between meta-dimensional omics data in breast cancer. Journal of biomedical informatics 56, 220–228.

[39] Ranstam, J., Cook, J.A., 2018. LASSO regression. British Journal of Surgery 105 (10), 1348.

[40] Ma, S., Fildes, R., Huang, T., 2016. Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. European Journal of Operational Research 249 (1), 245–257.

[41] LaValley, M.P., 2008. Logistic regression. Circulation 117 (18), 2395–2399.

[42] Empereur-Mot, C., Guillemain, H., Latouche, A., Zagury, J.-F., Viallon, V., Montes, M., 2015. Predictiveness curves in virtual screening. Journal of cheminformatics 7, 52.

[43] Qing Shi, J., Choi, T., 2011. Gaussian process regression analysis for functional data. Chapman & Hall, Boca Raton, 196 pp.

[44] Yan, J., Li, K., Bai, E., Yang, Z., Foley, A., 2016. Time series wind power forecasting based on variant Gaussian Process and TLBO. Neurocomputing 189, 135–144.

[45] Frank, E., Trigg, L., Holmes, G., Witten, I.H., 2000. Technical Note: Naive Bayes for Regression. Machine Learning 41 (1), 5–25.

[46] Fard, M.J., Wang, P., Chawla, S., Reddy, C.K., 2016. A Bayesian Perspective on Early Stage Event Prediction in Longitudinal Data. IEEE Trans. Knowl. Data Eng. 28 (12), 3126–3139.

[47] Yokuma, J., Armstrong, J., 1995. Beyond accuracy: Comparison of criteria used to select forecasting methods. International Journal of Forecasting 11 (4), 591–597.

[48] Armstrong, J.S., 2009. Selecting Forecasting Methods. SSRN Journal.

[49] Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55 (10), 78–87.

Biography

Prof. Dr.-Ing. Dipl.-Wirt. Ing. Günther Schuh (*1958) holds the Chair of Production Systems at the Laboratory for Machine Tools and Production Engineering WZL at RWTH Aachen University, is a member of the board of directors of the Fraunhofer Institute for Production Technology IPT and director of the Research Institute for Rationalization (FIR) at the RWTH Aachen. He is the founder of Schuh & Co. Group and the e.GO Mobile AG. Prof. Schuh is a member of several supervisory boards and boards of directors.

Andreas Gützlaff (*1989) studied Business Administration and Engineering specializing in Mechanical Engineering at the RWTH Aachen University. He is Chief Engineer and Head of the Production Management Department at the Laboratory for Machine Tools and Production Engineering WZL at RWTH Aachen University.

Julian Ays (*1990) studied Business Administration and Engineering specializing in Mechanical Engineering at the RWTH Aachen University and Tsinghua University. He is Team Lead Global Production in the Production Management department at the Laboratory for Machine Tools and Production Engineering WZL at RWTH Aachen University.

Niklas Rodemann (*1992) studied Business Administration and Engineering specializing in Mechanical Engineering at the RWTH Aachen University. Since 2018, he has been a research assistant in the Production Management department at the Laboratory for Machine Tools and Production Engineering WZL at RWTH Aachen University. His research focuses on the configuration of global production networks.

Miha Podbreznik (*1995) studied Mechanical Engineering specializing in Production Engineering at the RWTH Aachen University. He is a Student Research Assistant at the Laboratory for Machine Tools and Production Engineering WZL at RWTH Aachen University in the Global Production group in the Production Management department.