

## Methodological Assessment of Data Suitability for Defect Prediction

DOI: 10.12776/QIP.V24I2.1443

Peter Schlegel, Daniel Buschmann, Max Ellerich, Robert H. Schmitt

Received: 2020-04-03 Accepted: 2020-06-29 Published: 2020-07-31

### ABSTRACT

**Purpose:** This paper provides a domain specific concept to assess data suitability of various data sources along the production chain for defect prediction.

**Methodology/Approach:** A seven-phase methodology is developed in which the data suitability for defect prediction in interlinked production steps is assessed. For this purpose, the manufacturing process is mapped and potential influencing variables on the origin of defects are identified. The available data is evaluated and quantified with regard to the criteria relevancy, completeness, appropriate amount of data, accessibility and interpretability. The individual assessments are then visualized in an overview, gaps in data acquisition are identified and needs for action are derived.

**Findings:** The research shows a seven-phase methodology to systematically assess data suitability for defect prediction and identify data gaps in interlinked production steps.

**Research Limitation/implication:** This research is limited to the analysis of contextual data quality for the use case of defect prediction. Other data analytics applications or processes outside of manufacturing are not included.

**Originality/Value of paper:** The paper provides a new approach to identify gaps in data acquisition by systematically assessing data suitability for defect prediction and deducting needs for action. The accuracy of predictive defect models is then to be improved by the subsequent optimization of the data basis.

**Category:** Research paper

**Keywords:** predictive quality; defect prediction; failure prediction; data suitability; data quality

## 1 INTRODUCTION

The implementation of Industry 4.0 shapes the competition of manufacturing companies on global markets (Bal and Erkan, 2019). Those who want to stay ahead of their competition for the long term are required not only to record all available information securely and in real time, but also to process it in order to be able to analyze it precisely and continuously (Uhlemann et al., 2017). Having the right information available at the right time is an enormous challenge. Therefore, it is indispensable to recognize patterns in the data stream, to learn from them and to be able to derive the right predictions for the company, the processes and the products (Brecher et al., 2017). This applies in particular to the prediction of product defects in interlinked production steps (Eger et al., 2018). In a data-centric view of the entire production chain, there is great potential for the predictive identification of defect causes, the derivation of suitable measures and thus the reduction of defect costs. Without continuous data acquisition, it is not possible to trace correlations resulting from workpiece handling in different process steps (Ghimire et al., 2015). Classical tools such as Statistical Process Control mostly consider individual process steps and are therefore unsuitable for this application (Škulj et al., 2013).

The vision of the Internet of Production (IoP) describes a real-time, secure availability of information at any time and any place (Brecher et al., 2017; Pennekamp et al., 2019). Precise and continuous data analysis, pattern recognition for prediction and, based on this, reliable decision-making should support production systematically and sustainably.

The IoP infrastructure shown in Figure 1 consists of four underlying layers (Brecher et al., 2017):

- The *Raw Data* layer as well as the raw data access via the respective application software.
- A *Middleware+* for the administration of the data access on different proprietary systems.
- The *Smart Data* layer for the generation of knowledge based on the *Digital Shadow* and the *Smart Expert* layer on which the domain-specific usage of the aggregated knowledge takes place.

The term *Digital Shadow* is defined as the sufficiently precise representation of the processes in production, development and adjoining areas with the purpose of creating a real-time-capable evaluation basis for all relevant data (Bauernhansl et al., 2016). The relevant representation refers specifically to a smaller scope of data than that contained in the raw data, since only data relevant to the application case are passed on (Brecher et al., 2017).

In order to implement models for predicting product defects, a systematic selection of the relevant data in the sense of the *Digital Shadow* must first take place. For this systematic selection, manufacturing companies currently lack

knowledge regarding their own data quality (DQ) and the suitability of collected data for this specific application (Schuh et al., 2019). Previous research work in this context has mainly focused on general description models and metrics for the evaluation of DQ and not on the context-related development of methods for data evaluation (Wang and Strong, 1996; Batini et al., 2009; Zaveri et al., 2015).

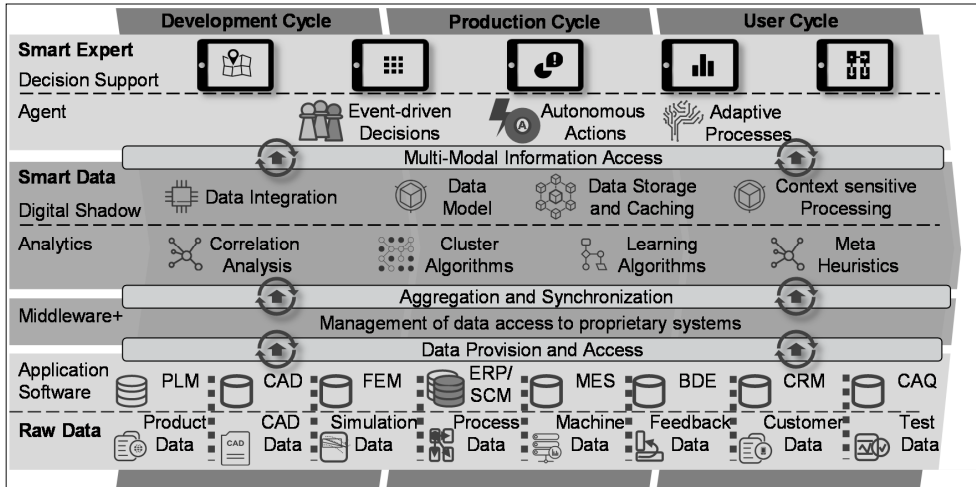


Figure 1 – Infrastructure of the Internet of Production  
(Source: Based on Brecher et al. (2017))

## 2 RELATED WORK

Based on the deficits mentioned, a systematic evaluation of the data suitability for the application case of defect prediction is carried out within the framework of the methodology presented. The main objective of the developed methodology is the creation of transparency regarding data suitability of existing data sources, the identification of data gaps and the derivation of need for action.

### 2.1 Implementation of Defect Prediction

In literature, a variety of data-driven methods and strategies for the implementation of defect prediction in manufacturing can be found. Eger et al. (2018) describe the *ForZDM methodology*, which expands single process boundaries of pre-existing Zero Defect Manufacturing (ZDM) approaches towards a production line perspective. This makes it possible to contrast and counter defects before, during and after their emergence through the integration of multi-level system modelling, big data analysis, Cyber Physical Systems and real-time data management (Eger et al., 2018). Wang (2013) presents a general framework of ZDM and explains how to apply Data Mining approaches to manufacture the products with zero-defect. The developed framework has a modular structure and consists among others of the main components fault prognosis, fault diagnosis as well as the subsequent correction and compensation

(Wang, 2013). Lieber et al. (2013) describe a methodical framework based on data mining for predicting the physical quality of intermediate products in interlinked manufacturing processes in the context of a rolling mill case study. Other approaches to defect prediction also do not include specific modules for their systematic assessment of the available database (Arif, Suryana and Hussin, 2013; Kao et al., 2017; Wuest, Irgens and Thoben, 2013; Schmitt and Deuse, 2018).

From literature, it becomes clear that the implementation of methods for defect prediction is often addressed in the context of ZDM strategies. Implications for the quality of the data basis, which result from the examination of interlinked production steps, are considered little or not at all. The presented frameworks contain modules for data preparation and feature extraction, but no methodology for the context-related evaluation of the data basis.

## **2.2 Assessment of Data Quality**

As mentioned in the introduction, a wide range of general methods and techniques already exist for the assessment of DQ. The earliest work in the area of DQ assessment is published by Wang and Strong (1996). In their work, they present a conceptual framework of DQ including 15 dimensions within four categories. Batini et al. (2009) provide a systematic and comparative description of methodologies that help the selection, customization, and application of DQ assessment. Cai and Zhu (2015) analyze the data characteristics of the big data environment, present quality challenges faced by big data, and formulate a hierarchical DQ framework from the perspective of data users. Zaveri et al. (2015) conducted a comprehensive survey on the assessment of linked DQ and identified 16 quality dimensions that have been studied in the literature. The work unifies and formalizes commonly used terminologies across papers related to DQ and provides a comprehensive list of 18 quality dimensions and 69 metrics (Zaveri et al., 2015). These dimensions were classified into four categories: (i) Accessibility, (ii) Intrinsic, (iii) Contextual, and (iv) Representational. Gürdür, El-khoury and Nyberg (2018) base their analysis on the work of Zaveri et al. (2015) and present a study that explains and applies a DQ assessment methodology as a post-integration phase for linked enterprise data. The authors examine a case study from the automotive industry using the linked enterprise data approach to integrate data from different development tools (Gürdür, El-khoury and Nyberg, 2018). Ardagna et al. (2018) propose a methodology to build a DQ adapter module, which selects the best configuration for a context-aware DQ assessment based on the user main requirements: time minimization, confidence maximization, and budget minimization.

The literature contains a large number of general methods and metrics for evaluating DQ. However, there is no methodology for evaluating the database in the context of defect prediction. Regardless of the formal quality of the data for the application of suitable prediction models, there is uncertainty in many



companies as to whether the information required for defect prediction is available in digital form at all. From this, the need for research is derived to develop a methodology to assess the suitability of the existing database and to identify gaps in data acquisition. The research hypothesis derived from this is:

*H: The accuracy of a model for defect prediction can be improved by a systematic assessment of data suitability and the subsequent elimination of the identified data gaps.*

### **3 PREDICTIVE MODELING OF PRODUCT DEFECTS**

Predictive modeling is applied to a range of techniques that find relationships between a target variable and the other variables in the data set (Wang, 2013). For example, such functions could be classification, value prediction or association rules (Raudys, 2001; Backhaus et al., 2016). In the given context, the target variable is a specific product defect that is identified by quality control in production. The other influencing variables include the parameters of individual process steps in the production chain as well as recorded sensor and meta data of disturbance and environmental variables. The objective of defect prediction is to predict the occurrence of certain defect types based on process parameters, sensor and meta data of the manufacturing process.

Usually, all available data sources are included in the database and then informative and non-redundant features are extracted (cf. Figure 2). Feature extraction converts raw data into informative features that efficiently represent the information relevant for analysis (Rawat and Khemchandani, 2017). The more accurately the features represent the data set, the faster and more accurately the predictive model will work (Kacprzyk et al., 2006). The features created are iteratively tested and optimized during the modeling process, which allows the dimension of the input data to be reduced (Liu and Motoda, 2008). For a successful feature extraction a suitable pre-processing of the data as well as a high contextual and formal DQ of the raw data is necessary. While the evaluation of formal DQ (e.g. accuracy, consistency etc.) is excluded from this research, the objective is to evaluate the context-related data suitability with a special focus on the degree of information content of the data.

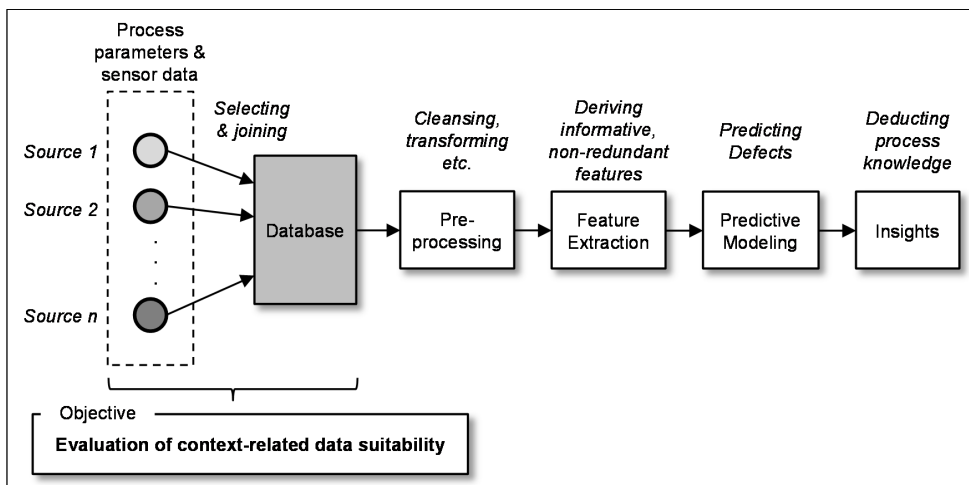


Figure 2 – Localization of the Research Objective in the Predictive Modelling Process

#### 4 METHODOLOGY

DQ is usually understood as a multi-dimensional concept. The dimensions represent the views, criteria, or measurement attributes for DQ problems that can be assessed, interpreted, and possibly improved individually. By assigning scores to these dimensions, the overall DQ can be determined as an aggregated value of individual dimensions relevant in the given application context.

The basis for the developed methodology for evaluating the context-related data suitability for defect prediction is a slightly modified framework for evaluating data quality according to Wang and Strong (1996), which defines 15 DQ dimensions in four categories. Of the 15 dimensions depicted in Figure 3, the following five are addressed within the methodology of this work: *relevancy*, *completeness*, *appropriate amount of data*, *accessibility* and *interpretability*. The two criteria *value-adding* and *timeliness* will not be considered further in the context of the developed methodology for the time being. Whether the available data or the resulting features actually add value can only be determined after modeling. The criterion *timeliness* and therefore the requirements for the speed with which the data is available depends strongly on the real-time requirements of the respective predictive model and thus on the application case in the production. Therefore, no general evaluation can be made. Since this work is a purely context-related evaluation of DQ, the more accurate term *data suitability* instead of data quality is used from here on.

This methodological approach provides companies with guidance on how to gain insight into the information content of the data collected and its suitability for defect diagnosis and prediction. The need for increased transparency and understanding of the data basis is due to the fact that initiated data analyses in

practical industrial utilization often bring only limited gain in knowledge. The methodical approach describes a step-by-step process that aids stakeholders in discussing, defining and identifying gaps in data collection through systematic uncovering and visualization.

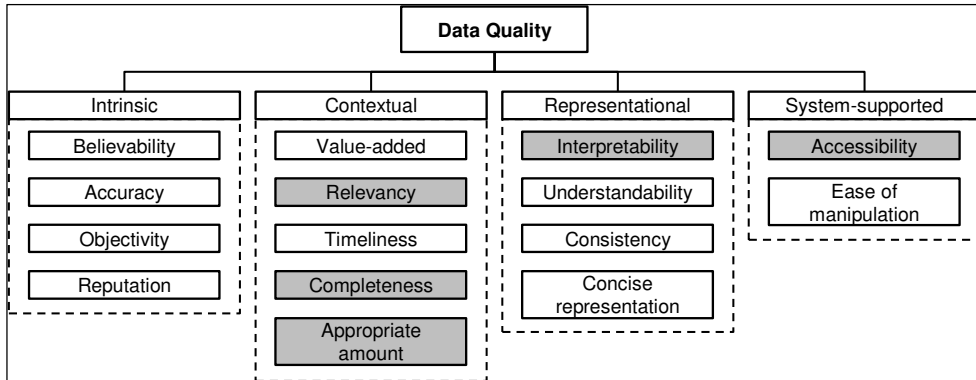


Figure 3 – Framework of 15 Quality Dimensions (Source: Based on Wang and Strong (1996), Hildebrand et al. (2015) and Zaveri et al. (2015))

#### 4.1 Seven-phased Approach

Based on the five identified criteria, a multi-phased methodology for evaluating data suitability is presented in the following. The implementation requires stakeholders from different disciplines to contribute both process and information technology expertise. The methodology consists of the following seven phases:

- Phase 1: Process mapping
- Phase 2: Identification of potential influencing variables (relevancy)
- Phase 3: Evaluation of completeness
- Phase 4: Evaluation of the amount of data
- Phase 5: Evaluation of accessibility
- Phase 6: Evaluation of interpretability
- Phase 7: Identification of data gaps and derivation of need for action

Parallel to the stepwise approach of the methodology, the results are documented in an overview of the framework for visualization purposes (Figure 4).

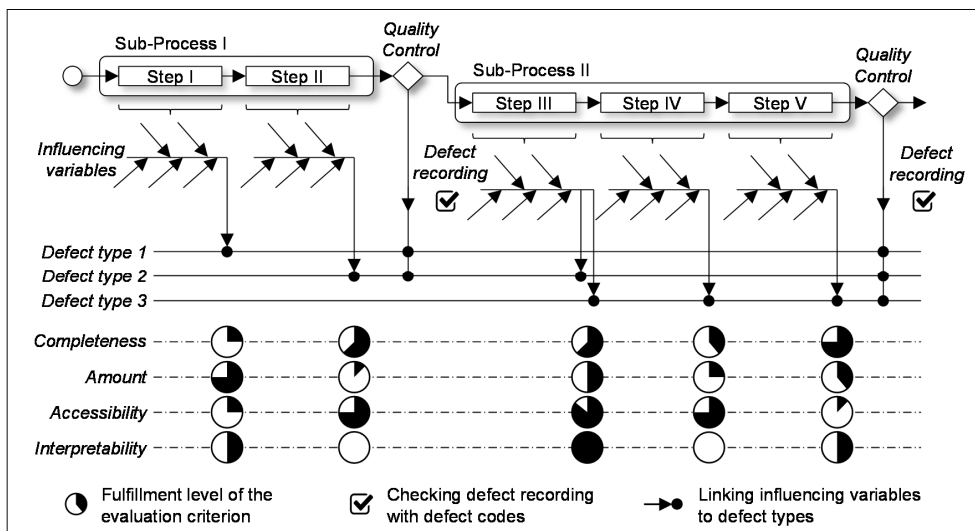


Figure 4 – Framework Visualization of the Methodology

#### 4.1.1 Phase 1: Process Mapping

Previously existing documentation of the manufacturing process is reviewed and a profound understanding of the process is built up within the stakeholder team. In workshops, the framework for the development of a defect prediction model is set and system boundaries as well as target variables are defined. In the context of defect prediction, the target variables are documented defect types with defined defect codes. For subsequent modeling, it can be helpful to restrict the focus to particularly frequent defect types in order to reduce effort. Subsequently, the process is visualized as a functional flow chart in which the essential sub-processes of production are mapped (Figure 4).

#### 4.1.2 Phase 2: Identification of Potential Influencing Variables (Relevancy)

In the next step, the expert team identifies potential influencing variables of the individual process steps, which could affect a specific product characteristic and thus the defect type. This is done completely detached from recorded data and thus represents only the existing process knowledge. A procedural model for the identification of quality-relevant influencing variables has already been elaborated in previous publications (Schmitt et al., 2016; Schmitt et al., 2020). The model in question consists of four steps: the identification of quality characteristics (here: defect types), quality-relevant processes, quality-relevant information and quality-relevant sources of information. The identified potential influencing variables are systematically mapped in a cause-effect diagram and assigned to the specific defect type by linking them to the defect line (Figure 4). Whether the identified influencing variables are reflected in recorded process data is then checked in the further course of the methodology.

#### 4.1.3 Phase 3: Evaluation of the Completeness

The existing database is then evaluated in regard to its degree of coverage of the potential influencing variables identified in phase 2. This evaluation of the criterion *completeness* is carried out separately for each individual process step from a data-centric point of view. The measure of *completeness* for individual influence variables  $I_i$  is defined as follows:

$$Q_{complete}(I_i) := \begin{cases} 1 & \text{if corresponding data is recorded} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The evaluation of *completeness* of the data basis for entire process steps  $S_j$  results from the arithmetic mean over all potential influencing variables of the process step.

$$Q_{complete}(S_j) := \frac{\sum_{i=1}^n Q_{complete}(I_i)}{n} \quad (2)$$

The determined value of  $1 \geq Q_{complete}(S_j) \geq 0$  is subsequently visualized as Harvey Ball and included in the methodology's framework (Figure 4) in the process overview. In addition to evaluating the *completeness* of recorded data in relation to individual process steps, it is also checked whether defect data recording has been implemented for individual products with defined defect codes (Figure 4).

#### 4.1.4 Phase 4: Evaluation of the Amount of Data

In phase 4, the amount of available data is evaluated for every individual process step. For this purpose, the storage duration of the data recording is considered, i.e. to what extent historical data is available for the training of a predictive model. Five evaluation levels are introduced for this purpose:

- No storage of historical data
- Storage of historical data for one week
- Storage of historical data for one month
- Storage of historical data for twelve months
- Storage of historical data for > twelve months

In addition, the evaluation of this criterion is affected by the heterogeneity of the parameter variations in the data sets. On top of the parameter variation itself, special focus is placed on the occurrence of the considered defect types in the recording period. This is due to the fact that for a successful training of a predictive model a sufficiently large number of defects must have occurred in the recording period. This ensures that a balanced training data set can be created without the application of over or under sampling techniques. The overall evaluation of the criterion is then visualized in the form of Harvey Balls and placed in the methodology's framework (Figure 4).

#### 4.1.5 Phase 5: Evaluation of Accessibility

In phase 5, the IT systems and their underlying databases in which the identified data is stored are reviewed with regard to data accessibility. The range goes from centrally stored SAP databases to decentrally distributed data in local systems on the shop floor. The focus of the evaluation criterion lies primarily on the possibilities for data export from the individual storage systems. Five evaluation levels are introduced for this purpose:

- Data export is not possible
- Manual screening and transfer of data
- Manual data export via storage device
- Semi-automated export via network
- Automated export via network

Afterwards, the evaluation of data accessibility for single process steps is again visualized in the form of Harvey Balls and placed in the methodology's framework (Figure 4).

#### 4.1.6 Phase 6: Evaluation of Interpretability

The evaluation of the interpretability of the available data takes place in phase 6. For this purpose, the extent to which the collected data can be clearly assigned to a specific product and thus to occurring defects is evaluated. The data integration and therefore the linkability of data from different production steps to a certain product is made possible by the available meta information. If such a linkability via meta data is not given, the information contained in the data can be neither interpreted regarding its influence on the defect cause nor be modeled later. A total of four levels are distinguished in the evaluation:

- The recorded data cannot be assigned to individual products and thus to defects that occur.
- The assignment of recorded data to predefined product types can be implemented.
- The assignment to individual products can be implemented indirectly, e.g. via batch assignment or the time stamp.
- The recorded data can be directly assigned to individual products, e.g. via a product ID.

The results of the evaluation carried out are also visualized and entered in the methodology's framework (Figure 4).

#### 4.1.7 Phase 7: Identification of Data Gaps and Derivation of Need for Action

After the production process has been mapped, potential influencing factors have been derived and the suitability of existing data has been evaluated on the basis of defined criteria, the main fields of action are systematically identified. The degree to which Harvey Balls are filled in serves as an indicator for the greatest gaps in data suitability. The elimination by means of suitable measures is then prioritized. First, measures with low personnel and financial expenses are to be addressed in order to achieve the so-called quick wins. However, in principle the strategy with which the gaps in data acquisition are to be closed depends strongly on the respective target variable. This can, for example, be characterized by a particularly frequent type of defect after a certain process step.

## 5 OUTLOOK

In future work the presented methodology will be elaborated in more detail as well as extended by further features. For phase 2, the identification of potential influencing variables, a flexible and process-independent meta-model for manufacturing data is currently being created to simplify the identification of influencing variables and their mutual relationships within the manufacturing process. Furthermore, the *understandability* of the data in the form of formal DQ criteria such as data format and data structure is going to be included, which is not particularly addressed in the current draft. The main focus of future work lies on the validation of the methodology on the basis of a real use-case from the field of high volume sink production. Before and after the implementation of the methodology, available data will be pre-processed, features will be extracted and predictive models for defect prediction will be created. The accuracy of these models for defect prediction will be compared before and after applying the methodology in order to verify the formulated research hypothesis: *The accuracy of a model for defect prediction can be improved by a systematic assessment of data suitability and the subsequent elimination of the identified data gaps.*

## 6 CONCLUSION

The systematic approach, in which the individual phases are completed step by step, makes it possible to evaluate data suitability for defect prediction with regard to various defined criteria. The continuous documentation and visualization of the evaluations allow a quick overview of the data gaps and thus the need for action along the production process. By applying the methodology before the actual modelling, the user receives information about the data suitability at a level of detail, which he does not receive by Machine Learning models with a black box character. The methodology therefore supports a systematic selection of the relevant data in the sense of the *Digital Shadow*. It is postulated that the accuracy of predictive models can be increased by using this systematic evaluation of data suitability and the elimination of identified gaps.

Since the initial effort is rather high, especially for many sub-process steps, it can be useful to limit the focus to certain sub-processes with a high number of occurring defects. In order to train a capable prediction model, formal DQ criteria are relevant as well as the applied content criteria.

## ACKNOWLEDGEMENTS

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

## REFERENCES

- Ardagna, D., Capiello, C., Samá, W. and Vitali, M., 2018. Context-aware data quality assessment for big data. *Future Generation Computer Systems*, [e-journal] 89, pp.548-562. DOI: 10.1016/j.future.2018.07.014.
- Arif, F., Suryana, N. and Hussin, B., 2013. A Data Mining Approach for Developing Quality Prediction Model in Multi-Stage Manufacturing. *International Journal of Computer Applications*, [e-journal] 69(22), pp.35-40. DOI: 10.5120/12106-8375.
- Backhaus, K., Erichson, B., Plinke, W. and Weiber, R., 2016. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-662-46076-4.
- Bal, H.Ç. and Erkan, Ç., 2019. Industry 4.0 and Competitiveness. *Procedia Computer Science*, [e-journal] 158, pp.625-631. DOI: 10.1016/j.procs.2019.09.096.
- Batini, C., Capiello, C., Francalanci, C. and Maurino, A., 2009. Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, [e-journal] 41(3), pp.1-52. DOI: 10.1145/1541880.1541883.
- Bauernhansl, T., Krüger, J., Reinhart, G. and Schuh, G., 2016. *Wgp-Standpunkt Industrie 4.0*. Frankfurt am Main: Wissenschaftliche gesellschaft für produktionstechnik.
- Brecher, C., Klocke, F., Schmitt, R. and Schuh, G. eds., 2017. *Internet of Production für agile Unternehmen: AWK Aachener Werkzeugmaschinen-Kolloquium 2017*. Aachen, Germany, 18-19 May 2017. Aachen: Apprimus Verlag.
- Cai, L. and Zhu, Y., 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, [e-journal] 14(2), pp.1-10. DOI: 10.5334/dsj-2015-002.



- Eger, F., Coupek, D., Caputo, D., Colledani, M., Penalva, M., Ortiz, J.A., Freiburger, H. and Kollegger, G., 2018. Zero Defect Manufacturing Strategies for Reduction of Scrap and Inspection Effort in Multi-stage Production Systems. *Procedia CIRP*, [e-journal] 67, pp.368-373. DOI: 10.1016/j.procir.2017.12.228.
- Ghimire, S., Melo, R., Ferreira, J., Agostinho, C. and Goncalves, R., 2015. Continuous Data Collection Framework for Manufacturing Industries. In: I. Ciuciu, ed. 2015. *On the move to meaningful internet systems: OTM 2015 workshops, Lecture Notes in Computer Science*. Cham, Heidelberg, New York, Dordrecht, London: Springer. pp.29-40. DOI: 10.1007/978-3-319-26138-6\_5.
- Gürdür, D., El-khoury, J. and Nyberg, M., 2018. Methodology for linked enterprise data quality assessment through information visualizations. *Journal of Industrial Information Integration*, [e-journal] 15, pp.191-200. DOI: 10.1016/j.jii.2018.11.002.
- Hildebrand, K., Gebauer, M., Hinrichs, H. and Mielke, M. eds., 2015. *Daten- und Informationsqualität: Auf dem Weg zur Information Excellence*. Wiesbaden: Springer. DOI: 10.1007/978-3-658-09214-6.
- Kacprzyk, J., Gunn, S., Guyon, I., Nikraves, M. and Zadeh, L.A. eds., 2006. *Feature extraction: Foundations and applications, Studies in Fuzziness and Soft Computing*. Berlin, Heidelberg, Springer. DOI: 10.1007/978-3-540-35488-8.
- Kao, H.-A., Hsieh, Y.-S., Chen, C.-H. and Lee, J., 2017. Quality prediction modeling for multistage manufacturing based on classification and association rule mining. *MATEC Web of Conferences*, [e-journal] 123(9), p.00029(2017). DOI: 10.1051/mateconf/201712300029.
- Lieber, D., Stolpe, M., Konrad, B., Deuse, J. and Morik, K., 2013, “Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning. *Procedia CIRP*, [e-journal] 7, pp.193-198. DOI: 10.1016/j.procir.2013.05.033.
- Liu, H. and Motoda, H., 2008. *Computational methods of feature selection, Chapman & Hall / CRC data mining and knowledge discovery series*. Boca Raton: Chapman & Hall/CRC.
- Pennekamp, J., Glebke, R., Henze M., Meisen T., Quix, C., Hai, R., Gleim, L., Niemietz, P., Rudack, M., Knape, S., Epple, A., Trauth, D., Vroomen, U., Bergs, T., Brecher, C., Bührig-Polaczek, A., Jarke, M. and Wehrle, K., 2019. Towards an Infrastructure Enabling. In: The Institute of Electrical and Electronics Engineers (IEEE) IEEE Industrial Electronics Society (IES), *Proceedings of the 2nd IEEE International Conference on Industrial Cyber-Physical Systems (ICPS 2019)*. Taipei, Taiwan, 6-9 May 2019. IEEE. DOI: 10.1109/ICPHYS.2019.8780276.
- Raudys, Š., 2001. *Statistical and Neural Classifiers: An Integrated Approach to Design, Advances in Pattern Recognition*. London: Springer. DOI: 10.1007/978-1-4471-0359-2.

Rawat, T. and Khemchandani, V., 2017. Feature Engineering (FE) Tools and Techniques for Better Classification Performance. *International Journal of Innovations in Engineering and Technology*, [e-journal] 8(2), pp.169-179. DOI: 10.21172/ijiet.82.024.

Schmitt, J. and Deuse, J., 2018. Similarity-search and Prediction Based Process Parameter Adaptation for Quality Improvement in Interlinked Manufacturing Processes. In: The Institute of Electrical and Electronics Engineers (IEEE), *IEEE International Conference on Industrial Engineering and Engineering Management*. Bangkok, Thailand, 16-19 December 2018. IEEE. DOI: 10.1109/IEEM.2018.8607361.

Schmitt, R.H., Ellerich, M., Schlegel, P., Ngo, Q.H., Emonts, D., Montavon, B., Buschmann, D. and Lauther, R., 2020. Datenbasiertes Qualitätsmanagement im Internet of Production. In: W. Frenz, ed. 2020. *Recht und Technik: Handbuch Industrie 4.0*. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-662-58474-3\_25.

Schmitt, R.H., Ngo, Q.H., Groggert, S. and Elser, H., 2016. Datenbasierte Qualitätsregelung. In: R. Refflinghaus, Ch. Kern, and S. Klute-Wenig, eds. 2016. *Qualitätsmanagement 4.0 – Status Quo! Quo vadis? - Bericht zur GQW-Jahrestagung 2016*. Kassel: Kassel University Press. Ch. 6. DOI: 10.19211/KUP9783737600859.

Schuh, G., Rebentisch, E., Riesener, M., Ipers, T., Tönnies, C. and Jank, M.-H., 2019. Data quality program management for digital shadows of products. *Procedia CIRP*, [e-journal] 86, pp.43-48. DOI: 10.1016/j.procir.2020.01.027.

Škulj, G., Vrabič, R., Butala, P. and Sluga, A., 2013. Statistical Process Control as a Service: An Industrial Case Study. *Procedia CIRP*, [e-journal] 7, pp.401-406. DOI: 10.1016/j.procir.2013.06.006.

Uhlemann, T.H.-J., Schock, C., Lehmann, C., Freiberger, S. and Steinhilper, R., 2017. The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems. *Procedia Manufacturing*, [e-journal] 9, pp.113-120. DOI: 10.1016/j.promfg.2017.04.043.

Wang, K.-S., 2013. Towards zero-defect manufacturing (ZDM)—a data mining approach. *Advances in Manufacturing*, [e-journal] 1(1), pp.62-74. DOI: 10.1007/s40436-013-0010-9.

Wang, R.Y. and Strong, D.M., 1996. Beyond Accuracy. What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, [e-journal] 12(4), pp.5-33. DOI: 10.1080/07421222.1996.11518099.

Wuest, T., Irgens, C. and Thoben, K.-D., 2013. Analysis of Manufacturing Process Sequences, Using Machine Learning on Intermediate Product States. In: C. Emmanouilidis, M. Taisch and D. Kiritsis, eds. 2013. *Advances in Production Management Systems, IFIP Advances in Information and Communication Technology*, Berlin, Heidelberg: Springer. Vol. 398. DOI: 10.1007/978-3-642-40361-3\_1.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S. and Hitzler, P., 2015. Quality assessment for Linked Data. A Survey. *Semantic Web, [e-journal]* 7(1), pp.63-93. DOI: 10.3233/SW-150175.

---

## ABOUT AUTHORS

**Peter Schlegel** – (P.S.) Laboratory for Machine Tools and Production Engineering WZL of RWTH Aachen University, Aachen, Germany, Research Associate, e-mail: p.schlegel@wzl.rwth-aachen.de, Author's ORCID: 0000-0003-0906-4644.

**Daniel Buschmann** – (D.B.) Laboratory for Machine Tools and Production Engineering WZL of RWTH Aachen University, Aachen, Germany, Research Associate, e-mail: d.buschmann@wzl.rwth-aachen.de, Author's ORCID: 0000-0001-8363-1331.

**Max Ellerich** – (M.E.) Laboratory for Machine Tools and Production Engineering WZL of RWTH Aachen University, Aachen, Germany, Head of Department, e-mail: m.ellerich@wzl.rwth-aachen.de, Author's ORCID: 0000-0001-9113-0203.

**Robert Heinrich Schmitt** – (R.H.S.) Laboratory for Machine Tools and Production Engineering WZL of RWTH Aachen University, Prof., Aachen, Germany, e-mail: r.schmitt@wzl.rwth-aachen.de, Author's ORCID: 0000-0002-0011-5962.

## AUTHOR CONTRIBUTIONS

P.S. – conceptualization, methodology and writing; D.B. – methodology and validation; M.E., R.H.S. – review and editing.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).