4th Conference on Production Systems and Logistics

Data-driven Prediction of Internal Turbulences in Production Using Synthetic Data

Jan Schuhmacher[1,2,3], Thomas Bauernhansl[1,2]

[1] *Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Nobelstrasse 12, 70569 Stuttgart, Germany*

[2] *University of Stuttgart, Institute of Industrial Manufacturing and Management IFF, Nobelstrasse 12, 70569 Stuttgart, Germany*

[3] *Reutlingen Center Industrie 4.0, Fraunhofer IPA & ESB Business School (Reutlingen University), Alteburgstraße 150, 72762 Reutlingen, Germany*

Abstract

Production planning and control are characterized by unplanned events or so-called turbulences. Turbulences can be external, originating outside the company (e.g., delayed delivery by a supplier), or internal, originating within the company (e.g., failures of production and intralogistics resources). Turbulences can have far-reaching consequences for companies and their customers, such as delivery delays due to process delays. For target-optimized handling of turbulences in production, forecasting methods incorporating process data in combination with the use of existing flexibility corridors of flexible production systems offer great potential. Probabilistic, data-driven forecasting methods allow determining the corresponding probabilities of potential turbulences. However, a parallel application of different forecasting methods is required to identify an appropriate one for the specific application. This requires a large database, which often is unavailable and, therefore, must be created first. A simulation-based approach to generate synthetic data is used and validated to create the necessary database of input parameters for the prediction of internal turbulences. To this end, a minimal system for conducting simulation experiments on turbulence scenarios was developed and implemented. A multi-method simulation of the minimal system synthetically generates the required process data, using agent-based modeling for the autonomously controlled system elements and event-based modeling for the stochastic turbulence events. Based on this generated synthetic data and the variation of the input parameters in the forecast, a comparative study of data-driven probabilistic forecasting methods was conducted using a data analytics tool. Forecasting methods of different types (including regression, Bayesian models, nonlinear models, decision trees, ensemble, deep learning) were analyzed in terms of prediction quality, standard deviation, and computation time. This resulted in the identification of appropriate forecasting methods, and required input parameters for the considered turbulences.

Keywords

Data-driven Prediction; Probabilistic Forecasting; Turbulences; Synthetic Data; Flexible Production;

1. Introduction

Not only since the outbreak of the Corona pandemic, have companies been subject to numerous unplanned events or turbulences that hinder economic production [1], [2]. A difference is made between internal turbulences originating in the company itself, such as resource failures and missing parts at the workplace, and external turbulences originating outside the company, such as delivery failures and delivery delays [3].

DOI: https://doi.org/10.15488/13438
ISSN: 2701-6277

The capacity for change describes the ability of a production system to respond to external and internal changes. According to *Erlach et al.* [4], a distinction must be made between short-term turbulences, medium-term trends, and long-term disruptions. Due to the short-term effect of turbulences requiring a short-term (low reaction time) and reversible action in the production system, the change capability of flexibility fulfills these requirements best [4]. The built-in flexibility of flexible production systems with defined boundaries (flexibility corridors) allows production systems to deal with turbulences without having to change the structure of the system, experiencing only minor losses in terms of time, effort, costs, or performance [5] [6]. The detection and prediction of turbulences offers huge potential for target-optimized handling of turbulences using existing flexibility corridors [7]. For example, using the operation and routing flexibility of flexible production, a predicted internal turbulence of a production resource failure or process delay can be countered by allocating the production order to an alternative production resource [8].

To address the challenge of generating the required database for identifying suitable probabilistic forecasting methods, an experimental research design has been applied to predict internal turbulences [9]. Starting with analyzing the state of science of data-driven forecasting and synthetic data generation, a simulation study to generate synthetic data has been conducted based on a minimal system with defined turbulence events. Then, the synthetic data was used for a comparative study of data-driven probabilistic forecasting methods to predict internal turbulences by splitting the data into training data for the forecasting models and test data to validate the forecasting models.

2. Analysis of the state of the science

In the following, the state of science in the fields of data-driven forecasting and synthetic data generation is analyzed to set the theoretical base for the simulation study and comparative study of data-driven probabilistic forecasting methods and to highlight the limitations of existing approaches.

2.1 Data-driven forecasting

Forecasting methods can be model-driven or data-driven [10], [11]. Model-driven approaches use classical statistical methods and assume that the data-generating process is known and that an explicit stochastic behavior can be assumed. In data-driven approaches, the data-generating process is considered unknown, and algorithms are used to meet specific analysis or prediction goals based on input data [11]. For the prediction of turbulences, a data-driven approach was chosen, since the data-generating, turbulence-triggering process is considered unknown.

A probabilistic forecast represents an estimate of the corresponding probability of all possible future outcomes of a random variable. Unlike single-value forecasts, such as time series averages, or quantile forecasts, probabilistic forecasts represent a probability density function considering future quantities or events [12], [13]. For predicting turbulence events in production systems, these probabilistic forecasting methods are particularly suitable since they allow the determination of the probabilities of occurrence of all turbulence events considered, taking into account the selected parameters and not only one particular event. By using simulators, a probabilistic forecast can also be carried out with deterministic forecast methods by varying the attribute values and/or combining different forecasting scenarios, as in the case of ensemble forecasts.

Major types of data-driven probabilistic forecasting methods are ensemble methods, deep learning methods, methods based on Bayes' theorem, and methods using regression [7].

Standard ensemble methods are random forests and gradient boosted trees. They combine many weak deterministic predictors (decision trees in the case of random forests and gradient boosted trees) to create a superior deterministic predictor. To turn ensemble methods into probabilistic forecasting methods, the mixing or combining process is adapted to consider probabilities instead of a single aggregated outcome [12].

Deep learning methods are based on a multi-layer feedforward artificial neural network trained with stochastic gradient descent [14]. The input data is processed through the neuronal network and afterwards the determined output is compared to the desired output [15]. Deep learning methods are probabilistic because the gradient descent converges faster when the loss function reflects a probabilistic forecast [12].

The method of the naïve Bayes classifier is a probabilistic classification method based on the Bayes' theorem. The main assumption is that given the value of the class, the value of any attribute is conditionally independent of the value of any other attribute [16], [17]. Although this independence of input attributes and target attributes is given only in rare cases, this method often allows good (prediction) results even in cases where this condition is not fulfilled [16], [17].

Methods using regression, such as logistic regression, support vector machines and fast large margin, can also be applied for probabilistic forecasts. They provide particularly good results when there are strong relationships or correlations between a target categorical variable and one or more input variables [15].

A major challenge is identifying the suitable forecasting method for a specific application. A common approach to selecting an appropriate forecasting method for a particular use case is to analyze the state of the art and science to determine which forecasting methods have provided good results in the past for that use case. For probabilistic forecasting of internal turbulences, no existing prior work could be identified in the analysis of the state of the art and science. Discussions with experts as well as literature sources led to the result that a parallel test of different forecasting methods (considering forecast quality and computation time) and a variation of the input parameters (add, replace or remove) must be performed [18], [19]. However, data-driven forecasting methods require a large database that is not always available in industrial practice. Synthetic data generation offers huge potential to build up the required database of different input parameters.

2.2 Synthetic data generation

Synthetic data is used when real data is difficult to access, real data sets are not detailed enough or not of the right scale (amount of data), or if real data sets are incomplete [20]. Synthetic data can be divided into the three main types of dummy data, rule-based generated synthetic data, and synthetic data generated using artificial intelligence [21].

Dummy data is randomly generated data created, for example, by means of a mock data generator. Due to the random data generation, features, relationships, and statistical patterns of the original data are lost, so the representativeness of dummy or mock data is minimal compared to the original data [21].

Rule-based generated synthetic data can be generated by defining specific rules for generating the data, e.g., by using simulation tools [21], [22]. The use of simulation methods to generate (or "grow") and analyse synthetic data to gain specific insights is part of the research field of data farming [23], [24]. This requires the definition of all features, relationships, and statistical patterns to be reproduced in the generated synthetic data. Therefore, the data quality of the synthetic data strongly depends on the quality of the predefined set of rules. Nevertheless, rule-based synthetic data generation is useful, especially when no data are available (yet) [21].

Another way to generate synthetic data is to use artificial intelligence algorithms. This involves training the AI model with real data to learn appropriate features, relationships, and statistical patterns of the data set. The trained AI algorithm then generates entirely new data points and models them to replicate the training data's features, relationships, and statistical patterns [21]. This AI-based generation of synthetic data can be used when sufficient data is available for training the AI algorithms.

Since no real data on production systems' internal turbulences is available, but the effects of turbulence on the production system are known, a rule-based approach for the simulative generation of synthetic data is pursued to generate synthetic data based on a simulation study described in the next section.

3. Simulation study for the generation of rule based synthetic data

Following VDI 3633, seven steps must be considered to perform a simulation study [25]. The procedure starts with the task definition, followed by the systems analysis, model formalization, implementation, execution of simulation experiments, and analysis of the results. In parallel, the phases of data collection and data preparation have to be considered. In the following, the major steps of the simulation study preparation, implementation of the simulation model, and generation of synthetic data are described.

3.1 Preparation of the simulation study

For simulating an existing or planned system, its transformation into a model that can be experimented with is required. According to VDI 3633 [25], the mapping accuracy of the model should not be as detailed as possible but detailed enough to meet the specific objective. For the generation of synthetic data based on turbulence scenarios of a production system, a minimal system with reduced structural complexity was defined by reducing the number and variants of elements, relations, and properties in the system and by applying the modeling principles of abstraction and representation [26], [27], [28].

The minimal system model is based structurally and process-wise on an actual work system for the assembly of a product at two alternative workstations supplied with pre-picked component sets from the warehouse via automated guided vehicles (AGV). For the considered work system, there is an insufficient database regarding quantity, quality, and granularity of real process and turbulence data. By transferring this working system into a simulation model, synthetic data is generated as input data for the investigation of suitable data-driven forecasting methods. It is also used for predicting turbulences and their effects on process and throughput times.

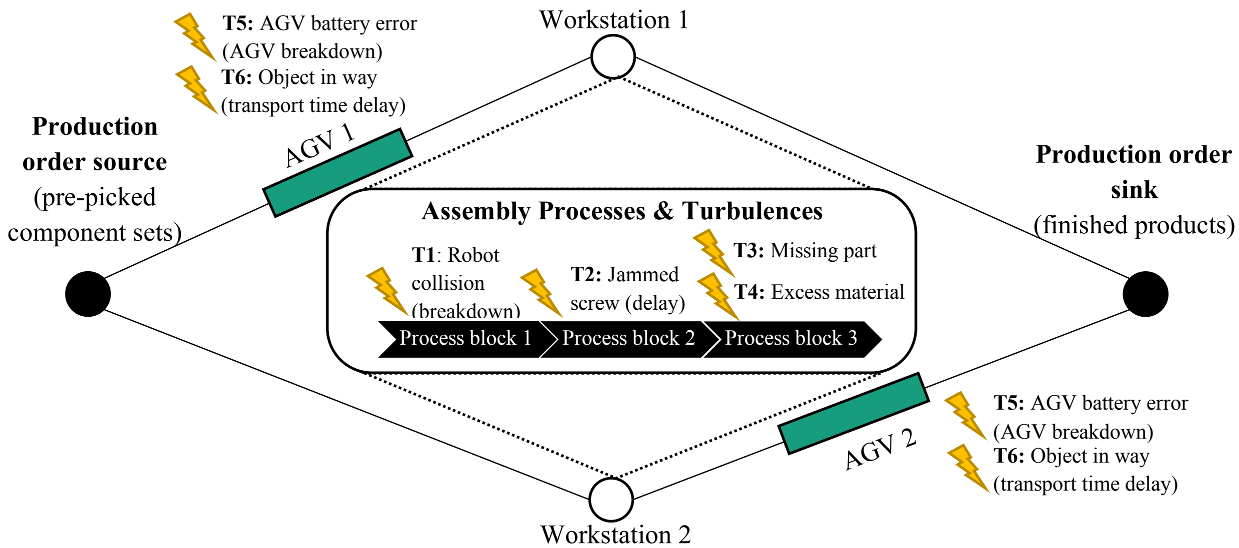


Figure 1: Concept of defined minimal system with turbulence events (T)

The minimal system (see Fig. 1) consists of one source (entry point for production orders in the form of pre-picked component sets into the system), two production resources (2 alternative assembly workstations), two intralogistics transport resources (2 alternative AGVs), and one sink (exit point of completed production orders). For the model implementation, the turbulences (T1-T6) to be considered were defined, specified in terms of their effect within the minimal system, and located in the model according to their locations of occurrence. The results of this analysis are summarized in the following section, together with the implementation of turbulence modeling.

3.2 Implementation of the simulation model

Previous work has already shown that negative effects of internal turbulences on lead time and delivery reliability can be countered by measures of intralogistics' autonomous control using the flexibility corridors of flexible production systems [29]. To model the defined minimal system, a multi-method simulation with agent-based modeling of the transport resources' autonomous behavior and event-oriented modeling of the turbulence scenarios with stochastic turbulence events is required. MATHLAB/Simulink, SimPy and AnyLogic were investigated as potentially suitable tools for this multi-method simulation [30], [31]. The simulation tool AnyLogic was selected due to its comprehensive method library for material flow simulations and agent modeling [31].

3.2.1 Modelling of resource breakdowns (T1 and T5)

The technical availability as a measure of a technical system's ability to meet the requirements placed on it is a major measure in the context of resource breakdowns [32]. The technical availability of production resources, like industrial robots, is specified over 99 % [33], [34] and the technical availability of AGVs is indicated with 98 % [35]. In industrial practice, also mean time between failures (MTBF) [33], [34] and mean time to repair (MTTR) [33] are important parameters for system availability. The MTBF for industrial robots is on average 40,000 hours (approx. 4.5 years) and more [33], [36]. Since these very high MTBFs would result in very long periods by the time, a purely technical failure of the corresponding production and logistics resources occurs, resource failures caused by human error are assumed for T1 and T5. Thus, for T1, a collision of the robot manipulator and the assembly fixture caused by the human operator is assumed resulting in the assembly fixtures' replacement and the robot's recalibration. For T5, a missed replacement of the AGV battery by the logistics employee is assumed. So, a warning message on the AGV fleet manager dashboard is ignored by the employee. For the modeling, a failure probability of 1 % for T1 and 3 % T5 has been assumed. Due to the lack of actual data, a uniform distribution as proposed by [37], [38] is assumed for the MTTR to solve the resource breakdowns.

3.2.2 Modelling of Process delays (T2 and T6)

For process-related delays or lead time delays in production systems with unknown delay reason, exponential distributions are usually assumed [39], [40]. For process times with some time dispersion, usually triangular [37], uniform [38], [41], [42], [43], and normal distributions [42], [44] are used in simulation models. A triangular distribution is frequently used for service and operation times where not enough samples are available for a more specific distribution, but some information about minimum, maximum, and mean values are available [37], [45]. Therefore, a narrowly set triangular distribution was used for modeling the target process times of assembly (without turbulence) with a certain tolerance due to the insufficient database of real data. For T2, as an assembly-related process delay with the specific reason "jammed screw", a triangular distribution was chosen because the error removal process (similar to the actual assembly process) requires loosening and re-screwing the jammed screw. This approach of splitting the process time and delay goes in line with *Wiendahl* [46], who states that turbulence occurs when the current actual value of a parameter deviates significantly from its mean value. For the representation of a time delay of the transport process due to an unknown obstacle on the transport path (T6), an exponential distribution was assumed in the model based on the literature [39], [40].

3.2.3 Modelling excess supply and missing part (T3 and T4)

For T3 and T4, no common comparative values could be found in the literature. The probabilities of occurrence of the turbulence were, therefore, modeled with assumed probabilities of 5 % (T3) and 4 % (T4).

Like the other process times in the model, the process times for the required picking, transport, and waiting times for the subsequent delivery of the missing material (T3) or return of the excess material were represented in the simulation model by a triangular distribution with minimum, maximum, and mean values.

The considered turbulences, turbulence attributes, and considered values and distributions as described above are summarized in table 1.

Table 1: Turbulences, turbulence attributes, distributions, and probabilities of the minimal system

Turbulence	Turbulence attributes	Distribution and probabilities
T1	Probability of occurrence	Probability: 1 %
	Mean time to repair (MTTR)	Uniform distribution [37], [38]
T2	Process delay	Triangular distribution [37], [45]
T3	Probability of occurrence	Probability: 5 %
	Process delay	Triangular distribution [37], [45]
T4	Probability of occurrence	Probability: 4 %
T5	Probability of occurrence	Probability: 3 %
	Mean time to repair (MTTR)	Uniform distribution [37], [38]
T6	Delay of transport time (AGV)	Exponential [39], [40]

3.3 Generation of synthetic data

By executing the simulation model, synthetic data for more than 5,000 production orders for 20 working days with two shifts was generated. All turbulences considered are independent of each other so that several of these turbulences can occur randomly during the processing of a production order within the defined minimal system. The results of the probabilistic forecasting comparing different forecasting models based on the generated synthetic data are described in the following chapter.

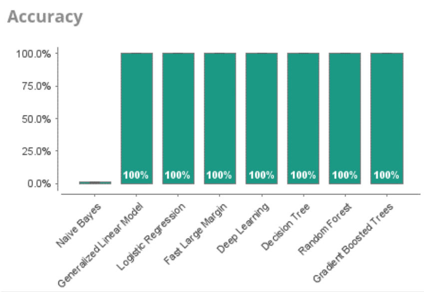
4. Probabilistic forecasting

To perform comparative studies of the selected forecasting methods and models, the data analytics tools RapidMiner, SPSS, Python (Pandas), R (CRAN-Project), and Minitab were analyzed based on the previous work of *Flückiger* [47]. The tool "RapidMiner" was chosen for the comparative investigation of the different probabilistic forecasting methods due to the available method library, the low effort for data import and the creation of the (forecasting) models as well as the high usability.

The synthetic data has been divided into training data (60 %) for training the data-driven prediction models and test data (40 %) for determining the prediction accuracies of the models. The investigated models and methods of different types (Naïve Bayes, generalized linear model, logistic regression, fast large margin, deep learning, decision tree, random forest, gradient boosted trees) were compared in terms of prediction accuracy for the respective turbulence, standard deviation and computation time. The results of the probabilistic forecasting are described in more detail on the two examples of the prediction of T1 as a production-related turbulence and T6 as an intralogistics-related turbulence.

4.1 Probabilistic forecasting of T1

The comparison of the investigated probabilistic forecasting methods to predict the occurrence of T1 of a resource breakdown due to a robot collision caused by the worker shows that most of the investigated forecasting methods lead to a 100 % accuracy to forecast the turbulence based on the generated synthetic training data (see Fig. 2). A major reason for this high forecasting accuracy of this turbulence (apart from the idealized data quality of the synthetic data) is that the occurrence of this turbulence is highly correlated to an increase of overall lead time. More precisely the robot defect has a negative impact on process block 1 where this turbulence is affecting the actual processing time of the production order. The probabilistic forecasting based on the naïve Bayes model leads to an accuracy of less than 1% since the main assumption of this model is that the value of any attribute is conditionally independent of the value of any other attribute [16], [17]. Considering the accuracy, standard deviation, and total computational time, the generalized linear model (GLM) led to the best overall performance for the forecasting of this turbulence considering the seized workstation, the process times of the workstations, and the overall lead time of the production order. By detecting this turbulence, the process-related delay can be predicted probabilistically and, using the flexibility corridors of flexible production, the negative effects (e.g., extended lead times) of this turbulence on subsequent orders can be reduced. For example, subsequent orders can be preventively allocated to an alternative production resource by tapping the routing flexibility of flexible production systems.



Weights by Correlation

Attribute	Weight
processtep1	0.989
leadtime	0.902
processtep3	0.025
processtep2	0.005
seizedws	0.005

Model	Accuracy	Standard Deviation	Gains	Total Time	Training Time (1,000 ...
Naive Bayes	0.9%	± 0.2%	-2,808	13 s	12 ms
Generalized Linear Model	100.0%	± 0.0%	26	12 s	21 ms
Logistic Regression	100.0%	± 0.0%	26	12 s	27 ms
Fast Large Margin	100.0%	± 0.0%	26	14 s	12 ms
Deep Learning	100.0%	± 0.0%	26	24 s	58 ms
Decision Tree	100.0%	± 0.0%	26	12 s	11 ms
Random Forest	100.0%	± 0.0%	26	36 s	16 ms
Gradient Boosted Trees	100.0%	± 0.0%	26	29 s	34 ms

Figure 2: Major results of probabilistic forecast T1

4.2 Probabilistic forecasting of T6

For the prediction of T6 of a transport time delay of an AGV due to an object on the driving path, the comparison of the probabilistic forecasting methods showed again that the GLM is the most favorable forecasting method, with a forecasting accuracy of 100 % and fastest overall computation time (see Fig. 3).

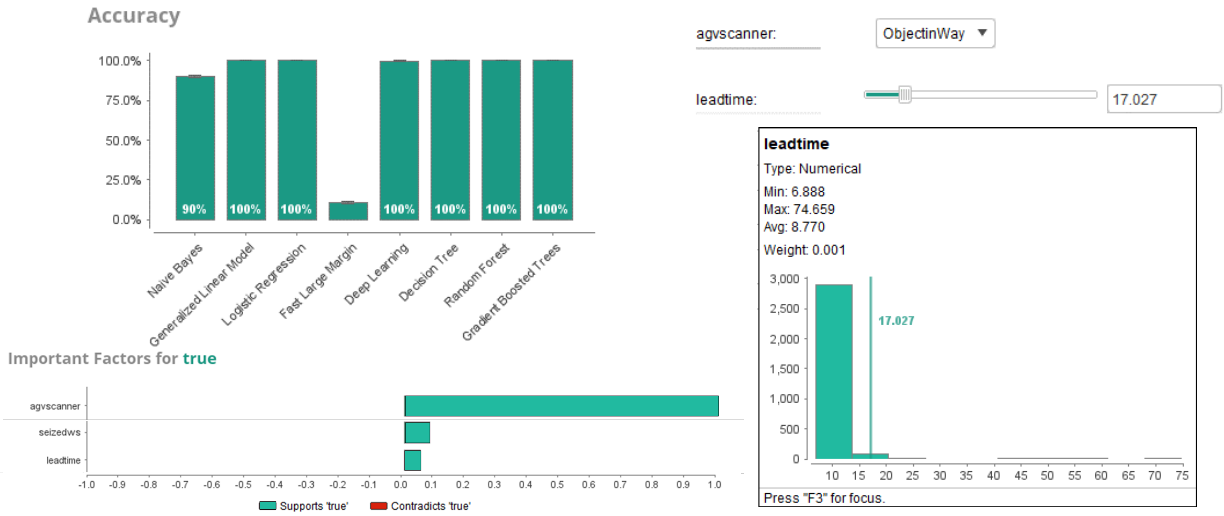


Figure 3: Major results of probabilistic forecast T6

As input parameters for this prediction AGV scanner alerts, the seized workstation, the process times of assembly as well as the overall lead time have been considered. Since not all input parameters correlate with the occurrence of T6, the fast large margin model leads to a poor forecasting accuracy of 11 % that can be increased to over 99 % by limiting the input parameters to the relevant factors. The analysis shows that the most important factors for the occurrence of T6 are the notification of the AGV scanner "object in way" and the seized workstation serviced by the AGV. The probabilistic forecast states that in the most likely case the lead time of the affected production order is expected to increase from an average of 8.770 minutes to 17.027 minutes. This probabilistic information can be used, for example, to transport subsequent orders to their destination on an alternative transport route using the material handling flexibility of the flexible production system, or to allocate subsequent production orders to alternative production resources (routing flexibility) which are not blocked by an obstacle.

4.3 Results of the comparative analysis

For the comparative analysis (also see table 2) of probabilistic, data-driven forecasting methods, different combinations of input parameters and forecasting models have been tested to predict the defined turbulences, as described above on the example of the turbulences T1 and T6. The comparative analysis using synthetic training data to train the prediction models and synthetic test data to validate the models showed that for most of the turbulences considered, the GLM is particularly suitable in terms of prediction accuracy and computational time. Nevertheless, some of the other forecasting methods showed comparable forecasting results. Therefore, a parallel comparison of different forecasting methods for the envisaged data-driven forecasting of turbulences based on real production will be maintained.

Table 2: Overview of the used input parameters to forecast turbulences

Turbu-lence	**Input parameters for forecast**							**Forecasting model with highest accuracy**
	AGV Battery warning	AGV scanner alert	Seized work-station	Lead time	Process time step 1	Process time step 2	Process time step 3	
T1			X	X	X	X	X	GLM (100 %)
T2			X	X	X	X	X	Decision tree (97.8 %)
T3			X	X	X	X	X	GLM (100 %)
T4			X	X	X	X	X	Naïve Bayes (96.2 %)
T5	X		X	X	X	X	X	GLM (100%)
T6		X	X	X	X	X	X	GLM (100%)

5. Conclusion

It has been shown and validated that a data-driven, probabilistic prediction of internal turbulences can be achieved through a simulative generation of synthetic data. In a further step, the findings regarding the identified models are to be validated with real data from the Werk150 – The Factory of the ESB Business School (Reutlingen University). The focus must be on the acquisition of manual process times for predicting possible assembly process delays since these frequently occur in practice and, therefore, offer great potential for reducing the negative influences of this turbulence on the lead time and delivery reliability of production orders. In the future, more potential arises from the inclusion of digital twins, which can provide additional input parameters for probabilistic forecasting of technical defects (e.g., through wear models of the corresponding resources) and other stochastic derivations in the production system. Future work is also intended to investigate how much forecasting methods can contribute to the exploitation of flexibility corridors in combination with autonomous control of intralogistics and, thus, to increase flexible production systems' resilience.

References

[1] Gruszczynski, L., 2020. The COVID-19 Pandemic and International Trade: Temporary Turbulence or Paradigm Shift? European Journal of Risk Regulation 11 (2), 337–342.

[2] Kupers, R. (Ed.), 2014. Turbulence: A corporate perspective on collaborating for resilience.

[3] Wiendahl, H.-H., 2006. Auftragsmanagement im turbulenten Umfeld – Teil 1 – Anforderungen. WT Werkstatttechnik online 96, 183–189.

[4] Erlach, K., Berchtold, M.-A., Gessert, S., Kaucher, C., Ungern-Sternberg, R., 2022. Flexibilität, Wandlungsfähigkeit und Agilität in der Fabrikplanung. ZWF 117 (7-8), 442–447.

[5] ElMaraghy, H.A., 2005. Flexible and reconfigurable manufacturing systems paradigms. International Journal of Flexible Manufacturing Systems 17 (4), 261–276.

[6] Andersen, A.-L., ElMaraghy, H., ElMaraghy, W., Brunoe, T.D., Nielsen, K., 2018. A participatory systems design methodology for changeable manufacturing systems. Intern. Journal of Production Research 56 (8), 2769–2787.

[7] Hill, R., Berry, S., 2021. An Introduction to Industrial Analytics, in: Hill, R., Berry, S. (Eds.), Guide to Industrial Analytics. Springer International Publishing, Cham, pp. 3–20.

[8] Schuhmacher, J., Hummel, V., 2022. Training Module on Probabilistic Forecasting of Production and Intralogistics Turbulences within Learning Factories. SSRN Journal.

[9] Leavy, P., 2022. Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches. Guilford Publications.

[10] Breiman, L., 2001. Statistical modeling: The two cultures. Statistical science 16 (3), 199–231.

[11] Freitag, M., Kück, M., Ait Alla, A., Lütjen, M., 2015. Potentiale von Data Science in Produktion und Logistik: Teil 1 – Eine Einführung in aktuelle Ansätze der Data Science. Industrie 4.0 Management 31 (5), 22–26.

[12] Vermorel, J., 2020. Probabilistic Forecasting. https://www.lokad.com/probabilistic-forecasting-definition.

[13] Gneiting, T., Katzfuss, M., 2014. Probabilistic Forecasting. Annual Review of Statistics and Its Application. 1 (1), 125–151.

[14] H2O.ai, 2021. Deep Learning (Neural Networks). https://docs.h2o.ai/.

[15] McCarthy, R.V., McCarthy, M., Ceccucci, W., Halawi, L., 2019. Applying predictive analytics. Springer, 205 pp.

[16] RapidMiner, 2021. Operator Reference Guide - Naive Bayes. https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/bayesian/naive_bayes.html.

[17] Maimon, O.Z., Rokach, L., 2005. Decomposition methodology for knowledge discovery and data mining: Theory and applications. World Scientific, Hackensach, NJ, 323 pp.

[18] Kieu, L.-M., Malleson, N., Heppenstall, A., 2020. Dealing with uncertainty in agent-based models for short-term predictions. Royal Society open science 7 (1), 191074.

[19] Lloyd, D.J.B., Santitissadeekorn, N., Short, M.B., 2016. Exploring data assimilation and forecasting issues for an urban crime model. European Journal of Applied Mathematics 27 (3), 451–478.

[20] Mannino, M., Abouzied, A., 2019. Is this Real?, in:Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. ACM, New York, NY, USA, pp. 549–561.

[21] Syntho, 2022. Was sind synthetische Daten? https://www.syntho.ai/de/what-is-synthetic-data/.

[22] AnyLogic, 2023. Synthetic data. https://www.anylogic.de/features/artificial-intelligence/synthetic-data/.

[23] Lechler, T., Sjarov, M., Franke, J., 2021. Data Farming in Production Systems - A Review on Potentials, Challenges and Exemplary Applications. Procedia CIRP 96, 230–235.

[24] Sanchez, S.M., 2020. Data Farming. ACM Transactions on Modeling and Computer Simulation 30 (4), 1–30.

[25] Verein Deutscher Ingenieure, 2014. Simulation of systems in materials handling, logistics and production: Fundamentals. Beuth Verlag GmbH, Berlin ICS 03.100.10.

[26] Kuzgunkaya, O., ElMaraghy, H.A., 2007. Assessing the structural complexity of manufacturing systems configurations. International Journal of Flexible Manufacturing Systems 18 (2), 145–171.

[27] Windt, K., Philipp, T., Böse, F., 2008. Complexity cube for the characterization of complex production systems. International Journal of Computer Integrated Manufacturing 21 (2), 195–200.

[28] Swarup, S., 2019. Adequacy: What Makes a Simulation Good Enough?, in: 2019 Spring Simulation Conference (SpringSim). Tucson, AZ, USA. 29.04.2019 - 02.05.2019. IEEE, pp. 1–12.

[29] Schuhmacher, J., Hummel, V., 2021. Development of an Intelligent Warehouse and Autonomously Controlled Intralogistics Scenario to Investigate the Mastering of Short-Term Turbulences at Werk150. SSRN Journal, 1–6.

[30] SimPy, 2020. Discrete event simulation for Python. SimPy. https://simpy.readthedocs.io/en/latest/.

[31] AnyLogic, 2022. Multimethod Simulation. https://www.anylogic.de/use-of-simulation/multimethod-modeling.

[32] Signoret, J.-P., Leroy, A., 2021. Reliability Assessment of Safety and Production Systems. Springer International.

[33] Automationspraxis, 2012. Roboter machen weniger Fehler. Automationspraxis. https://automationspraxis.industrie.de/allgemein/roboter-machen-weniger-fehler/.

[34] publish-industry Verlag GmbH, 2020. Ein Roboter für alle Einbaulagen. publish-industry Verlag GmbH. https://www.industr.com/de/ein-roboter-einbaulagen-2468562.

[35] Balyo, 2022. AGV Installation. https://www.balyo.us/success/your-savings/how-do-you-generate-savings.

[36] Raviola, A., 2020. Diagnostics and Prognostics of Industrial and Collaborative Robots. https://www.dimeas.polito.it/content/download/2772/18711/file/35_Raviola_a.pdf.

[37] Grigoryev, I., 2021. AnyLogic in Three Days: Modeling and Simulation Textbook, 5th ed.

[38] Yu, A.J., Seif, J., 2016. Minimizing tardiness and maintenance costs in flow shop scheduling by a lower-bound-based GA. Computers & Industrial Engineering 97, 26–40.

[39] Gudum, C.K., 2002. On the distribution of lead time delays in supply chains. København.

[40] Mathematics Stack Exchange, 2022. statistics - What density distribution could be used for time delays? https://math.stackexchange.com/questions/3146378/what-density-distribution-could-be-used-for-time-delays.

[41] Aramon Bajestani, M., Beck, J.C., 2015. A two-stage coupled algorithm for an integrated maintenance planning and flowshop scheduling problem with deteriorating machines. Journal of Scheduling 18 (5), 471–486.

[42] Dodin, B., 1996. Determining the optimal sequences and the distributional properties of their completion times in stochastic flow shops. Computers & Operations Research 23 (9), 829–843.

[43] Wang, L., Zhang, L., Zheng, D.-Z., 2005. A class of hypothesis-test-based genetic algorithms for flow shop scheduling with stochastic processing time. The International Journal of Advanced Manufacturing Technology 25 (11-12), 1157–1163.

[44] Heydari, M., Mahdavi Mazdeh, M., Bayat, M., 2013. Scheduling stochastic two-machine flow shop problems to minimize expected makespan. 10.5267/j.dsl, 163–174.

[45] Fairchild, K.W., Misra, L., Shi, Y., 2016. Using Triangular Distribution for Business and Finance Simulations in Excel. Journal of Financial Education 42 (3-4), 313–336.

[46] Wiendahl, H.-H., 2011. Auftragsmanagement der industriellen Produktion. Springer Berlin Heidelberg.

[47] Flückiger, I., 2015. Data Analytics in Organisations and Business: 5.5 Selection of Software Tools. ETH Zürich. https://stat.ethz.ch/education/semesters/as2015/analytics/Data_Analytics_2015-11-25.pdf.

Jan Schuhmacher (*1987) has been a research associate at the Fraunhofer Institute for Manufacturing Engineering and Automation IPA since 2017 and doctoral student at the Institute of Industrial Manufacturing and Management IFF at the University of Stuttgart since 2020. His major field of research is the design and optimization of changeable, autonomously controlled intralogistics systems.

Thomas Bauernhansl (*1969) has been the director of the Institute of Industrial Manufacturing and Management IFF at the University of Stuttgart and the Fraunhofer Institute for Manufacturing Engineering and Automation IPA since 2011. He is a member of the WGP, the Scientific Society for Production Technology as well as Deputy Chairman of the Steering Committee of Allianz Industrie 4.0 BW.