This paper investigates the productivity prediction for aircraft final assembly lines. Various approaches are proposed and the performance of each approach is compared and discussed. The predictive models established in this paper can be divided into three groups. The first group is a simulation model, built by practitioners with extensive insights into the plant layout, material handling, manufacturing process and resource allocation. The second group consists of three representative regression models, including linear, polynomial and exponential regression. The last one is based on machine learning, including multilayer perception, gradient boost regression tree and random forest.

Compared with typical light industrial products, aircraft product has a large number of modified types in order to meet the customers’ personalized requirements. In addition, the management of aircraft final assembly lines depends on different productivity ranges and managers often need to verify within a given tolerance the performance of the predictive models, which is not covered by previous studies. In light of the above, this paper provides a comprehensive comparison among different modelling approaches. A real aircraft final assembly line including three aircraft types is adopted to illustrate the feasibility of each approach.

The advantages and limitations of different approaches in terms of their efficiency, precision and generalization ability are presented. The aim of this paper is to provide a practical guidance to the choice of suitable approaches and datasets, as well as data processing techniques. The conclusion drawn in this paper can be applicable to other customer order dependent and multi-product assembly problems.
Dear Editor,

I am sending you our manuscript entitled “Productivity Prediction in Aircraft Final Assembly Lines: Comparisons and Insights in Different Productivity Ranges” by Tengfei Long, Yuan Li, and Jun Chen. We would like to have the manuscript considered for publication in the Journal of Manufacturing Systems.

Accurate productivity prediction is a primary concern in the aircraft final assembly line management because aircraft productivity can vary greatly. As a highly customized product, aircraft products are usually made into a large number of different modified types, which further increases the complexity in productivity prediction. In recent years, various approaches such as simulation, regression analysis, and machine learning models have been applied as promising approaches and achieved significant results in evaluating the productivity. However, researchers prefer to use a small-scale assembly line as the study cases and the number of related features tends to be small. There is a lack of systematic comparisons among all kinds of approaches in the literature as well. Moreover, for aircraft final assembly that has a large number of modified aircraft types, the final assembly lines of these modified types may not provide enough historical data to train a new predictive model. In light of these problems, we propose a systematically comparison among three groups of predictive models including seven approaches.

This paper presents the advantages and limitations of different approaches under different scenarios in terms of their efficiency, precision and generalization ability, providing a guideline for the selection of predictive models for the final aircraft assembly lines. Based on the engineering application requirements, the performance of these models will be compared and discussed in different productivity ranges as well. A potential relationship between prediction accuracy and the number of historical data distributed over different ranges is pointed out. Furthermore, two more final assembly lines of other modified aircraft types will be proposed to verify the generalization ability of these predictive models. More interestingly, we discover that machine learning approaches show a tendency to surpass the simulation approach in generalization ability, prompting a need for further research.

The subject classifications of the manuscript and the proposed associate editor are below:

**Subject classifications:**

20: Manufacturing systems design & operations;

80: Manufacturing information systems.

**Proposed associate editor:** Lihui Wang, Editor-in-Chief

**Proposed reviewers:**

Dr. Qian Zhang, Liverpool John Moores University, Q.Zhang@ljmu.ac.uk

Dr. Maysam Abbod, Brunel University London, maysam.abbod@brunel.ac.uk

Dr. Eve Zhang, Loughborough University, Y.Zhang@lboro.ac.uk

Pro. Yiming Rong, Southern University of Science and Technology, rongym@sustech.edu.cn

Pro. Qinglong An, Shanghai Jiao Tong University, qlan@sjtu.edu.cn

The research presented in our manuscript is supported by the National Key Research and Development Program of China (Grant Nos. 2019YFB1707501).

No potential conflict of interest was reported by the authors.

We look forward to hearing from you.

Yours sincerely,

Tengfei Long
Dear Editors and Reviewers,

I am very glad to receive the good news. Here are the suggestion from the Editor-in-Chief:

“The only remaining concern is its poor relevance to JMS. Normally, readers of JMS would like to see that newly published papers are closely related to the published JMS literature so that your paper finds a home in this publishing venue. Upon reviewing your 76 listed references, no JMS publication can be found which indicates a poor relevance to JMS. I wonder why your manuscript should be published in JMS if no related work has been reported in JMS.”

After a carefully review of the reference, up to 25% (19 references) of the 76 references have been replaced by the relevant ones published in JMS. Appendix A include all those replaced reference for your information. The selection of those references were based on their relevance to this paper as can be broadly categorized in the following groups.

- Relevance to the field studied in this paper, i.e. the aircraft final assembly lines, such as [1] and [5].
- Relevance to the adopted methods in this paper, the JMS also has a meaningful guidance, such as:
  - Simulation Method (7 references).
  - Regression Analytic Method (5 references).
  - Machine learning Method (5 references).
  - Other Method (2 references).
- The paper reviewed and discussed some traditional methods published earlier in JMS, and newly published articles as well, such as:
  - Published earlier (4 references published in 20th century).
  - Published in the year between 2001 to 2019 (4 references).
  - Newly published (11 references published in 2020 or 2021).
- Review and research articles published in JMS are also covered in this paper, such as:
  - Review articles (3 references i.e. [29], [34] and [36]).
  - Research articles (other 16 references).
- Topics related to manufacturing systems in the JMS article are also presented, such as:
  - Digital Twins (5 references, i.e. [29-31], [35], [56] and [73]).
  - Ensemble modelling and other data enhancement approaches (3 references, i.e. [65], [72] and [75]).

Thanks for your time and efforts in reviewing this paper and your valuable suggestions.
Appendix A: References published by the JMS


Highlights

• A systematically comparison of three groups of productivity predictive models including seven approaches is introduced.
• The performance of predictive models is discussed in different productivity ranges.
• Two more modified aircraft types are proposed to verify the generalization ability.
• The results provide a guideline for the selection of predictive models in aircraft final assembly lines.
Productivity Prediction in Aircraft Final Assembly Lines: Comparisons and Insights in Different Productivity Ranges

Tengfei Long a; Yuan Li a; Jun Chen *,b;

a Northwestern Polytechnical University, Xi’an 710072, PR China;
b School of Engineering and Materials Science Queen Mary University of London, Mile End Road, London, E1 4NS, United Kingdom;

Abstract

This paper investigates the productivity prediction for aircraft final assembly lines. Various approaches are proposed and the performance of each approach is compared and discussed. The predictive models established in this paper can be divided into three groups. The first group is a simulation model, built by practitioners with extensive insights into the plant layout, material handling, manufacturing process and resource allocation. The second group consists of three representative regression models, including linear, polynomial and exponential regression. The last one is based on machine learning, including multilayer perception, gradient boost regression tree and random forest.

Compared with typical light industrial products, aircraft product has a large number of modified types in order to meet the customers’ personalized requirements. In addition, the management of aircraft final assembly lines depends on different productivity ranges and managers often need to verify within a given tolerance the performance of the predictive models, which is not covered by previous studies. In light of the above, this paper provides a comprehensive comparison among different modelling approaches. A real aircraft final assembly line including three aircraft types is adopted to illustrate the feasibility of each approach.

The advantages and limitations of different approaches in terms of their efficiency, precision and generalization ability are presented. The aim of this paper is to provide a practical guidance to the choice of suitable approaches and datasets, as well as data processing techniques. The conclusion drawn in this paper can be applicable to other customer order dependent and multi-product assembly problems.

Key words: Productivity prediction; Aircraft product; Assembly line; Generalization ability; Machine learning.

1. Introduction

Due to a large number of involved manufacturing resources, a wide variety of labors’ professions, complex production scheduling and high uncertainty, the management of aircraft final assembly lines puts forward high requirements on situational awareness [1]. As aircraft final assembly is directly dependent on customer orders, the range of productivity can vary greatly [2]. In the most extreme case, an aircraft assembly line may have no customer orders for the entire year or more and would therefore be moved to a maintenance mode. In other cases, with a startling rise in customer orders, the assembly line must run at its full capacity to elevate 

* Corresponding author. E-mail address: jun.chen@qmul.ac.uk; Phone: +44 (0)20 7882 8873; Webpage: https://www.sems.qmul.ac.uk/staff/jun.chen#
productivity to the maximum limit. The inevitable occurrence of extreme orders results in an uneven distribution of historical data across productivity, which creates lots of uncertainty for productivity predictions. In addition, the storage cost caused by excess productivity and the compensation cost caused by insufficient productivity are both much higher than the costs of other industrial products [3]. With the implementation of the Industry 4.0, more elaborate controls and precise productivity evaluations are required, which cannot be met by the current practice. Therefore, accurate productivity prediction is pressing and becomes a primary concern of the aircraft final assembly line management.

As a highly customized product, aircraft products are usually made into a large number of different types, which further increases the complexity in productivity prediction [4]. In the past, based on the extrapolations of working hours for each operation, practitioners can calculate a standard mean production time to predict the productivity with no further discrimination among different aircraft types. With the maturity and application of Virtual Reality technology, working hour measurement based on the virtual environment becomes more readily available, which allows practitioners to obtain more accurate working hours of different aircraft types with more efficiency and lower cost [5]. Therefore, differences between aircraft types should no longer be ignored in future. However, currently modified types often do not accumulate enough historical data to train independent predictive models. Therefore, the problem requires a predictive model capable of not only having a higher prediction accuracy on a certain type, but also relatively accurate in prediction for other types.

In the study of productivity prediction problems, many conventional performance indicators, such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), are employed to evaluate the performance of predictive models. However, these indicators cannot reveal the full picture for the predictive models of some special cases of aircraft final assembly lines. For example, many studies evaluate performance in an averaged manner, but it seems more valuable to explore the variation of predictive accuracy across the productivity for aircraft final assembly lines which have an extremely varied productivity. Furthermore, small prediction errors are sometimes allowed in practice [6]. Evaluating the performance of a model within different error thresholds can be useful information to managers in deciding the usage of different models.

In light of the above, this paper compares different predictive-modeling methods ranging from those based on professional domain knowledge, to statistical approaches, and to machine learning methods, incorporating more performance indicators, with the aim to more rigorously assess the performance of different modelling approaches. The remaining of the paper is organized as follows. State-of-the-art approaches are reviewed in Section 2 with a focus on their advantages and disadvantages on the productivity prediction in aircraft final assembly lines. The datasets of an aircraft final assembly line are introduced in Section 3. Section 4 introduces performance indicators and seven different approaches, which are tested and compared in Section 5. Section 5 also presents insights into the efficiency, precision and generalization ability of each approach. Finally, conclusions are drawn in Section 6.

2. Literature review

The productivity predictive models for aircraft final assembly lines are studied extensively in the last 30 years. Existing approaches can be broadly categorized into three groups: (1) simulation models based on the queuing theory, (2) regression models based on statistical
approaches, and (3) machine learning methods.

2.1. Simulation Methods

The queuing theory provides a fundamental understanding of assembly lines. The method usually defines certain objects characterizing corresponding manufacturing elements in the assembly lines, associated with some basic rules that often serve as constraints, length of time and step size. Early examples include those by Gutjahr and Nemhauser [7], Bulgak and Sanders [8], Lachampt F and Perron R [9].

Since then, various manufacturing elements such as labors were taken into account in simulation models. Buzacott [10] considered the impact the labor levels on productivity in simulation, indicating that different labors combination and production strategies can significantly affect productivity. MacDuffie [11] studied the role of labor management for the automobile assembly line in improving productivity, and pointed out that labors, as a unique resource, can significantly improve productivity of the assembly line and product quality through reasonable organization and management. Wang et al. [12] established a simulation model taking each movable labor into account in the assembly line for evaluating the work efficiency of labors.

To better simulate the production of real assembly lines, more manufacturing elements were synthesized in simulation models, such as assembly line balance [13], machines energy consumption [14], material flow [15] and operating cost [16]. Wei [17] established a simulation model containing many environment factors like room temperature and lighting. Leung [18] enumerated machines’ states such as stock-out, blocking and starvation that can be considered in a simulation model. Furthermore, many researchers like Azadeh et al. [19] integrated simulation models and heuristic algorithms to optimize the assembly line, in which the simulation model plays a role as a productivity evaluation tool. The physical parameters of materials are also taken into account in the productivity estimates. Stavropoulos et al. [20] considers the properties of different materials, like the friction, during the assembly.

Based on the study in [21], it shows that the simulation based on the queuing theory has certain advantages and offers practicality in the analysis of the assembly line. Many simulation software platforms, such as Plant Simulation [22], Witness [23] and Quest/Delmia [24], have been developed and widely used in the productivity prediction, which enable engineers to predict the productivity that are closer to real scenarios [25]. In the management of A400m aircraft, Airbus analyzed the productivity of the final assembly line by using the simulation software, combining with the assembly process data stored in industrial Digital Mock-Up [26]. Yuan [27] studied the productivity prediction in the assembly line of complex products, and obtained the optimal solution of aeroengine assembly line balance by Plant Simulation software. Mas et al. [28] studied the conceptual designs of aircraft assembly lines. Product, process and resource are considered in this simulation platform, which enable industrial software systems to develop a knowledge-based application prototype.

Digital Twins are considered as the next generation of simulation [29], which enable the simulation model completely synchronized with the real states of the assembly lines. Among the technologies of Digital Twins, the interface and interaction technology are considered as an important addition to simulation methods [30]. Before the maturity of Digital Twins, simulation methods are often time consuming to run and not always practical. The obstacles of applications
and the solutions offered by Digital Twins are as follows:

(1) With the increasing number of manufacturing elements being taken into account, the simulation modelling process requires increasing professional domain knowledge of the assembly lines [31]. It would be extremely demanding to set up parameters such as the probability distribution of working hours of each operation [32], the law of machine failure [33] and differing speeds of Automatic Guided Vehicles (AGVs) to handle various materials [34]. With the help of Digital Twins, these parameters can be efficiently set up in simulation models in a more realistic way. For example, the working hour of personnel operations is considered to contain a great deal of uncertainty. The Virtual Reality, considered as an early integration technology of Digital Twins, has greatly simplified the parameter setting of simulation models [35].

(2) In the case of time constraint, simulation often fails to meet the efficiency requirements. Nowadays simulation models play a role as a productivity evaluation tool for other optimization algorithms [36]. Optimization algorithms will provide a very large number of alternative resource allocation schemes for the simulation model to predict their productivity. However, constant modification of simulation models requires considerable time and higher maintenance cost [37]. On the basis of data interacting technology in Digital Twins, model updating can be carried out in a more efficient and low-cost way, which opens the door for the verification of a large number of alternative resource allocation schemes.

In light of the above, the predictive models based on simulation is widely used in many practical applications. Simulation models can be adopted to predict the productivity in the aircraft final assembly line and their performances will be comprehensively compared with other widely applied approaches in this paper.

2.2. Regression Analytic Methods

Regression analysis is a statistical method of analyzing data to understand the correlation between variables and the direction and strength of the correlation.

Linear regression (LR) assumes that there is a linear relationship between the dependent variable and independent variables, and establishes a model to solve this relationship. Early examples of LR include those in [38-40]. Ganesharajah [41] established a LR model between productivity and AGVs’ fleet size in a single loop layout of the assembly line. Mangal [42] associated the productivity with manufacturing elements in a Bosch assembly line based on the LR algorithm. Evaluated against the actual data from the assembly line, it shows that the predictive model led to a high prediction accuracy, which won the award of Kaggle challenge. Akpınar [43] established a multiple linear regression model, describing the relationship between operations’ cycle time and productivity in a vehicle assembly line.

Investigating the performance of LR algorithms in productivity prediction revealed two preconditions for LR to achieve the desired accuracy: (1) Each independent variable (i.e., labor, operation, material, etc.) must be linear with respect to the dependent variable (i.e., productivity). (2) The sample size of the training dataset must be large enough to overcome the possible influence of noise in the sample in order to obtain a smaller error [44]. With the advent of more complex manufacturing systems, the limitations of linear regression are gradually brought into attention. For aircraft final assembly lines, which is a nonlinear complex assembly system [45], there are complex coupling effects among the manufacturing elements. Therefore,
LR may not offer a satisfactory accuracy in the productivity prediction for the aircraft final assembly lines.

Since every continuous function defined on a closed interval can be uniformly approximated as closely as desired by a polynomial function, polynomial regression is widely used to describe the nonlinearity in assembly lines. As also recommended by Hurrion and Birgil [46], the polynomial functional form was chosen since it is a common form for the development of regression based metamodels. The author pointed out that the key to improving the accuracy of PR is to predetermine the appropriate polynomial degree. Dengiz et al. [47] used a PR algorithm to optimize batch sizes in a real printed circuit board assembly line which was considered as a Just-in-Time (JIT) system. The author believes that the non-linear effect of inventory time on productivity can be reduced as much as possible based on JIT concept. In the selection of the polynomial degrees, the author used an iterative way to successively increase the degree by comparing the accuracy of productivity prediction. As also noted in the theoretical research of Rawlings [48], the application of productivity prediction using PR should pay attention to the following issues: (1) Careful design of the polynomial degrees of each independent variable; otherwise, it is likely to result in overfitting or underfitting. (2) A relatively high demand on the amount of data compared with LR in order to reduce the test error.

Exponential regression (ER) is often widely used in assembly lines because of the memoryless property of exponential functions [49]. Rodríguez [50] adopted the same JIT concept as in [40] in planning the assembly line, assuming that the relationship between productivity and consumption of materials should conform to the law of the exponential distribution. Rodríguez emphasized that the relationships between variables are the main reasons for using different types of regressions. Di Maio [51] used the Relevance Vector Machines and ER to predict the residual time of materials in the assembly line. Compared with the LR model in his paper, the ER model gives a more accurate prediction. Based on the historical data of two automobile assembly lines, Inman [52] also pointed out the exponential relationship between productivity and various processing features. Although a number of studies have demonstrated the exponential relationship between productivity and many manufacturing elements of assembly lines, ER is not widely applied to the productivity prediction of aircraft final assembly lines. The main reason lies in the fact that the relationship between productivity and factors related to labor cannot be perfectly described through the exponential distribution and aircraft assembly lines are a typical type of highly labor-dependent manufacturing systems [53]. According to Diaz et al. [54], the manual assembly may reach 65% of the total production time, accounting for more than 20% of the total manufacturing cost.

For a long time, regression models and simulation models are both employed in the productivity prediction for aircraft final assembly lines [55]. In general, when engineers need to make more rigorous predictions, computational resources are requested in order to run simulation. When the prediction has a tight constraint on time and computational resource, the regression models are more frequently used because of their efficiency.

The reasons for adopting the above three regression methods in this paper are as follows: (1) LR is one of the most fundamental and important regression methods; (2) Because of its approximation ability, PR can describe nonlinearity in assembly lines; (3) Because of its memoryless property, ER can describe the continuous and independent occurrence of events in assembly lines, like the material generation.
2.3. Machine Learning Methods

Due to the requirement of domain knowledge in building simulation-based models, and the difficulty in updating such models as frequently as the assembly lines do, there is an increasing interest in applying machine learning which shows promising results for various assembly line prediction tasks [56]. In what follows, the predictive models with a special emphasis on machine learning applied to assembly lines are discussed.

Researchers have adopted various machine learning models and different features of assembly lines to predict various indicators such as cost, productivity and quality. Zhang and Fuh [57] studied the machine learning approach to predict the cost based on a developed backpropagation neural network model. The testing results based on their approach showed better performance as compared with the regression approaches. The proposed approach utilized 21 cost-related features in a small-scale assembly line. Haouani et al. [58] studied Multilayer Perceptron (MLP) implementation for modeling of the control system in assembly lines. In this implementation, more manufacturing elements have been extracted as features of the neural network for the productivity predictive model. Chen and Yang [59] designed an assembly line using a hybrid approach for the neural network modeling incorporating a stochastic local search. They used a backpropagation algorithm to generate the neural network models and refined the models using the simulated annealing approach in order to obtain the optimal resource allocation scheme with respect to the objective of the design.

When training and testing the machine learning models, data generated by simulation models were often used. Sabuncuoglu and Touhami [60] did an experimental investigation of machine learning modeling of an assembly line. The simulation model of an assembly line with several operations is developed to train a productivity predictive model based on MLP. They demonstrated that the productivity predictive model based on MLP can be effectively used. Chan and Speeding [61] developed a decision tree model, trained and tested by the dataset produced by a simulation model of the assembly line. They use the decision tree model and response surface plot as an off-line decision support tool.

Comparing these predictive models based on machine learning with simulation or regression models, insights into the strength of machine learning were observed. Kilmer et al. [62] studied the use of MLP as a modeling technique for the cost prediction of an assembly line. The author extracted four features and established two MLPs to estimate the total cost and its variance. The results showed that, the neural network model is competitive in accuracy when compared to simulation. Once trained, it can operate in nearly real time. Lee and Shaw [63] applied the neural net approach to real-time process sequencing. The author pointed out that the practical benefit of the machine learning approach is that the neural network incrementally learns the sequencing knowledge and can apply the knowledge for sequencing a set of jobs on a real time basis. When compared to genetic algorithms, the neural network shows the superiority of efficiency using only less than 0.2% of the computational time needed by genetic algorithms. Li J et al. [64] studied the Random Forest approach in the productivity predictions in shale gas. The results showed better prediction productivity than the other machine learning approaches in accuracy. Fulya etc. [56] established MLP models to predict the productivity, with 15 features representing the capacity of 15 buffers of an ideal assembly line. The authors compared the performance of the machine learning model with the PR and ER approaches and drew a conclusion that PR and ER tended to overestimate the productivity.
To summarize, many approaches of productivity prediction have been developed to explain the relationship between features and productivity. However, few studies deal with the aircraft final assembly line. Moreover, most studies have the following limitations: (1) they used a small-scale assembly line as the study cases. In this case, the number of related features tends to be small. As Haouani et al. [58] pointed out, it is important to extract relevant features for machine learning models. However, the diversity of manufacturing elements in aircraft final assembly lines presents a significant hurdle for feature extraction. (2) In order to illustrate the performance of the constructed models, the researchers need to compare it with other approaches. However, there is a lack of systematic comparisons among all kinds of approaches in the literature. (3) For aircraft final assembly that has a large number of modified aircraft types, the final assembly lines of these modified types may not provide enough historical data to train a new predictive model. Therefore, ensuring the accuracy of the productivity prediction of these modified aircraft types has become a new requirement for the productivity predictive models.

Therefore, this paper adopts three machine learning algorithms, which are MLP, GBRT and RF. MLP, as a representative supervised learning algorithm in machine learning, is not only seen in the most common application scenarios, but also widely followed by other algorithms in terms of structure, principle and flow. Ensemble methods, such as GBRT and RF, shows a high tolerance for imbalanced data and a high generalization ability in many papers [65]. However, the imbalanced data and multiple aircraft types mentioned in Section 1 have not received enough attention in previous studies. Therefore, these two approaches are also adopted in this paper for comparison.

Based on the above discussion and the engineering application requirements, we attempt to systematically compare three groups of productivity predictive models including seven approaches. The main contributions of this paper are as follows:

(1) The productivity predictions for aircraft modified types in the final assembly lines is investigated in this paper. In the past, practitioners often use a standard mean productivity as a surrogate for modified types. One of the main reasons behind such a practice is that modified types often do not accumulate enough historical data to train independent predictive models. In this paper, productivity prediction of modified types is based on the assumption that if the generalization ability of data driven models is sufficiently high, those models can be used for productivity prediction of modified types. Two types of modified aircraft are adopted to illustrate the feasibility of this method.

(2) A more comprehensive evaluation methodology for productivity predictive model adopted in this paper. In most of the previous literatures, the adopted performance indicators could not reveal the full capabilities of predictive models. Therefore, it is difficult to decide which of those models are suitable for the aircraft productivity management. In this paper, the performance of these models is compared and discussed not only in terms of four conventional performance indicators, but also under different productivity ranges, predictive error thresholds and aircraft modified types.

(3) We also highlight the following aspects as a practical guidance for practitioners to choose appropriate modelling methods, data sets and data processing techniques. 1) The cause and impact of imbalanced distribution of data within each productivity range on productivity predictions are highlighted and discussed. 2) The relationship between the modification degree
of a modified type and the generalization performance of predictive models are discussed.

3. Data

Modeling methods of this paper include two categories: Firstly, for simulation-based modelling approaches, domain knowledge of the assembly line organization is required, including the plant layout, process operation and material handling data etc. Secondly, for the regression and machine learning approaches, historical data such as labor allocation, buffer allocation, and AGV number etc. is required.

3.1. Data for Simulation

The simulation model mainly includes four sub-models: plant layout, material handling, operation process and resource allocation. The data used in the simulation model will be divided into four categories according to the division of sub-models. The roles, descriptions and relationships of data in these sub-models are introduced below.

3.1.1. Plant layout data

The layout information will be used in all space-related calculations in simulation. The plant layout sub-model has a rectangular area boundary of 240 m long and 66 m wide, containing all the spatial information inside this boundary. The layout data in a simulation model includes four types: station layout, source layout, buffer layout and path layout as summarized in Table 1. Sources serve as starting points for different materials, and their location information will be defined. The same goes for buffers as transfer points, paths as routes between points, and stations as the end points. The spatial information contained in plant layout sub-model will provide the basis for distance calculation in the material handling sub-model. These calculated distances will be furtherly combined with the speed attributes in resource sub-model to generate relevant time to further complete the calculation of productivity.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>station_layout</td>
<td>When constructing the model, the number of stations, the position and area of each station will be considered. The station will perform all operations required for final assembly of aircraft.</td>
</tr>
<tr>
<td>source_layout</td>
<td>Sources is the starting point for all materials to enter the final assembly system. The location of each source will be recorded here.</td>
</tr>
<tr>
<td>buffer_layout</td>
<td>Buffer is the place where all materials are transferred.</td>
</tr>
<tr>
<td>path_layout</td>
<td>The path used by labor or AGVs to carry different materials.</td>
</tr>
</tbody>
</table>

3.1.2. Operation process data

Operation data is used to calculate job flow in simulation. The operation process sub-model
contains a topology of 3927 operations, which means the sub-model does not contain any spatial information. All the information in the sub-model can be divided into two types: operations’ attributes and inter-operation relationships. Operation attributes refers to beginning and finishing conditions of each operation. Parameters such as resource and material demand are the beginning conditions, while the working hours of each operation refers to the finishing conditions. The release of resources and destruction of materials after finishing the operation are also defined. The inter-operation relationship refers to the precedence constraints that must be met in the assembly task execution process in the final assembly lines. Based on the operation topology, the sub-model can calculate the specific time under a specified throughput. In general, it includes the relationships among operations and attributes within an operation as summarized in Table 2.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>operation_network</td>
<td>Operation networks include serial and parallel relationships between jobs and other constraints.</td>
</tr>
<tr>
<td>operation_time_needed</td>
<td>The time required for each operation.</td>
</tr>
<tr>
<td>operation_material_needed</td>
<td>The materials required for each operation.</td>
</tr>
<tr>
<td>operation_tool_needed</td>
<td>The equipment required for each operation.</td>
</tr>
<tr>
<td>operation_labor_needed</td>
<td>The labors required for each operation.</td>
</tr>
</tbody>
</table>

3.1.3. Material handling data

The material handling sub-model covers the relevant logic of the material flow. The mass of materials is used to solve the remaining capacity of AGVs and buffers. The frequency of materials is coordinated with the order quantity, which determines the total amount of task for the material handling. The loading and unloading times of materials are included in the distribution time of materials together with the carrying time. The required data for this sub-model is summarized in Table 3.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>material_frequency</td>
<td>Based on statistical information, the frequency and quantity of various materials entering the assembly system.</td>
</tr>
<tr>
<td>material_mass</td>
<td>The mass of different kinds of material. If the material is measured by quantity, then the default value will be set as 1.</td>
</tr>
<tr>
<td>material_loadtime</td>
<td>Time requirements for loading on different buffers.</td>
</tr>
<tr>
<td>material_unloadtime</td>
<td>Time requirements for unloading on different buffers.</td>
</tr>
</tbody>
</table>

3.1.4. Resource allocation data

All kinds of resources constitute the main constraint of the aircraft final assembly line. In
fact, the allocation of resources is also the primary factor in management of the aircraft final assembly lines. In the simulation model adopted in this paper, four types of resources are involved, namely AGVs, tools, labors and buffers. The speed, capacity and quantity of AGVs are closely related to the carrying time in the material handling sub-model. The labors are divided into 17 professions. Labors in each profession can only perform certain operations and their speed also vary with professions. The capacity of buffers affects the transfer time of materials from the material handling sub-model to the operation process sub-model. The required data is summarized in Table 4.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGV_number</td>
<td>Number of AGVs in the assembly line.</td>
</tr>
<tr>
<td>AGV_speed</td>
<td>Speed of various types of vehicle.</td>
</tr>
<tr>
<td>AGV_capacity</td>
<td>Capacity of various types of vehicle.</td>
</tr>
<tr>
<td>Tool_number</td>
<td>Number of different type of tools.</td>
</tr>
<tr>
<td>labor_number</td>
<td>Number of labors in the profession.</td>
</tr>
<tr>
<td>labor_speed</td>
<td>Speeds of different labors.</td>
</tr>
<tr>
<td>labor_capacity</td>
<td>Capacity of different labors.</td>
</tr>
<tr>
<td>buffer_capacity</td>
<td>Capacity of various types of buffers.</td>
</tr>
</tbody>
</table>

3.2. Data for regression and machine learning approaches

In order to ensure productivity prediction accuracy, one should comprehensively consider relevant features that may affect prediction. In this study, we use up to 4 types of features, aiming to provide a sufficient set of features for the productivity prediction. These features are divided into four categories, including labor factors, tool factors, buffer factors and AGV factors, as indicated in Table 5. These four categories of features are the most important factors of production involved in aircraft final assembly.

- Labors: There are 17 professions in the final assembly lines. The 17 inputs refer to the number of labors in the corresponding profession. The type of these inputs is integer.
- Tools: There are 13 kinds of tools. The 13 inputs refer to the number of tools in the corresponding type. The data type is integer.
- AGVs: There are 4 AGVs in charge of material handling. The 4 inputs refer to the speed of the corresponding AGV. The data type is double.
- Buffers: There are 12 buffers in the final assembly lines. The 12 inputs refer to the capacity of the corresponding buffer. Buffer 1-4 contain standard components. Their capacity is measured by weight, provided by an electronic sensor. The data type is double. Buffer 5-8 contain customized components like cables and conduits. Buffer 9-12 contain aircraft bulks like engines and wings. The data type is integer.

Carefully comparing the data in Sections 3.1 and 3.2, we can investigate the difference in data requirements between the knowledge-based and data-driven models. The most obvious difference lies in the number of features. The data requirements of data-driven models can be regarded as a simplified version of those of the knowledge-based models. As the domain knowledge behind the features required by data-driven models is hidden, practitioners do not need to check all the parameters involved in the management of the aircraft final assembly line. Building predictive models requiring fewer features may lose certain fine details, but provide
simplification. More details of the data adopted by the predictive models are shown in the Appendix A.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_i$</td>
<td>The amount of labors in profession $i$, where $i=1, 2, \ldots, 17$.</td>
</tr>
<tr>
<td>$P_j$</td>
<td>The amount of tools in type $j$, where $j=1, 2, \ldots, 13$.</td>
</tr>
<tr>
<td>$S_m$</td>
<td>The speed of $m$th AGV, where $m=1, 2, \ldots, 4$.</td>
</tr>
<tr>
<td>$C_k$</td>
<td>The capacity of buffer $k$, where $k=1, 2, \ldots, 12$.</td>
</tr>
</tbody>
</table>

3.3. Productivity Division and Modified Types

According to the managerial purpose for productivity control, the productivity is divided into five regions, which respectively represent the five stages of productivity control: extremely low, low, normal, high and extremely high. When comparing the performances of different approaches, the accuracy in different productivity divisions, i.e., ranges, will be used. By comparing the performance of prediction models across different productivity ranges, practitioners will be able to better understand the reliability of predictions across these ranges offered by different modelling approaches.

For the above purpose, we choose the aircraft final assembly line as the study case in this paper for the following reasons.

- The historical data has been accumulated for 24 years. Unlike processing parameters such as materials’ surface temperature and machines’ internal vibration, which are difficult to collect, the collection of resource quantity is relatively simple. The managers have already recorded the resource allocation schemes since the final assembly line was put into operation. With the help of information systems, all the historical data was collected and managed in the Enterprise Resource Planning (ERP) system. Up to now, a total of 2616 historical data were recorded, which enables practitioners to build a data-driven productivity predictive model.

- Multiple types of aircraft are produced in this final assembly line. Therefore, it represents a typical case in aircraft final assembly and has practical engineering significance. There are three types of aircrafts in the final assembly line: two types of modified types and one basic type. Type Basic is an interregional transport aircraft, which is widely used in transporting goods. Type A refers to aircraft that have installation of various radio and navigation equipment according to customer requirements. Type B has been modified more radically, and is often customized by hospitals to have medical facilities for use in emergency situations. Extensive modifications in Type B have been made inside aircraft to for emergency assistance and medical personnel. A total of 2331 historical data samples of Type Basic were recorded, while Type A and Type B have 138 and 147 data samples respectively. Therefore, Type A and Type B do not have enough historical data to develop their own predictive models.

- Another striking feature is the imbalance of the dataset, where one class is represented by a large number of cases and the other one is represented by only a few cases. The ubiquity of this phenomenon in aircraft assembly lines, as well as the upsurge in algorithmic research on the imbalance problem, make this dataset from the aircraft final
assembly line extremely interesting. Hence, the effect of this imbalance on the performance of productivity predictive models warrants a further study. The data distribution in different productivity ranges and its effect will be illustrated and discussed in Section 5.3.

4. Predictive Models and Performance Indicators

In this section, seven predictive models are introduced for predicting the productivity, and seven performance metrics are adopted to access and compare modelling performances.

4.1 Models

In this study, we introduce seven models for productivity prediction, including linear regression (LR), polynomial regression (PR), exponential regression (ER), multilayer perceptron (MLP), gradient boosted regression trees (GBRT) and random forest (RF).

4.1.1. Simulation

The simulation model was built by practitioners. In the design and planning period, managers begin to use the simulation model for productivity prediction of the aircraft final assembly line. It largely preserves the internal structure and parameters of the assembly line. In addition, the parameters of the simulation model are set as close as possible to the reality of the assembly line. A manufacturing system dynamics model is developed to simulate the effects of layout, distribution, resources and processes on the production capacity of aircraft assembly lines. This model is able to explore the consequences of various parameters of resource allocation. It is noteworthy mentioning that, the model can also verify the impact of other additional parameters on productivity, such as different operation sequences [66]. However, in order to fairly compare the performance of the simulation method with those of other methods, the simulation model will focus on predicting the productivity under different resource allocations.

As shown in Fig. 1, there are four sub-models, which are plant layout, operation process, material handling and resource allocation, collectively calculating the total assembly time and productivity. Besides presenting the relations between the four sub-models, a simulation clock and an event list are used for recording the delivery time and operation time, whose relationship is calculated based on the start and end time on the event list, instead of in a purely additive way.

Fig. 1. Sub-system diagram of productivity predictive simulation model
(1) Plant layout sub-model

The plant layout sub-model involves all the distance calculations in material flow, mainly including: distance from source to buffer and from buffer to station. Any distance calculation needs to provide four types of information: the start location, the end location, the transfer location and the handling route.

- Calculation of the starting points: Firstly, based on the material_needed data of each operation in the Operation Process sub-model, the Bill of Material (BOM) is generated in the event list. Secondly, the coordinates of the corresponding source will be called one by one based on the BOM. Thirdly, the BOM will be converted into a list of starting point locations.

- Calculation of the end points and transfer points: The calculation method for the starting point is also applicable to the ending point. Combining with the station_layout and buffer_layout data, the BOM will be converted into a list of ending point and transfer point locations.

- Calculation of the handling route: Based on the location lists of all start, transfer, and end points, each material handling task will also be generated and the event list will be updated with the contents of the material handling tasks.

(2) Material handling sub-model

The material handling sub-model is responsible for the calculation of the delivery time. Firstly, based on the AGV_num data of Resource Allocation sub-model, the simulation clock can combine the handling tasks on the event list to assign specific tasks to each AGV. Secondly, the delivery time of each handling task will be calculated on the basis of the distance information and AGV_speed data.

(3) Operation process sub-model

The operation process sub-model involves all the calculations of operation time. The operation time includes three types: loading time, working hour and unloading time. Similar to the task assignment of AGV, each labor, tool is also assigned tasks in the event list. This process requires the labor_needed, tool_needed data in the Operation Process sub-model and the labor_number, tool_number in the Resource Allocation sub-model.

- Calculation of the loading time: After the completion of material handling process, this paper considers three operation start conditions, which are material, tool and labor. On the basis of the event list, the loading time will be calculated for each operation by finding the maximum waiting time for the three types of resources. The waiting time of material is determined by the capacity of the buffer and the number of materials.

- Calculation of the working hour: In this case, the working hour is a constant parameter set by the practitioners before. The influence of different assembly sequences or resource quantities on working hour is not further discussed in this paper. During processing, labors and tools are occupied and cannot be used by other operations.

- Calculation of the unloading time: In the unloading process, not only the unloading time is added into the operation time, but also the release of resources such as labors and tools. Resources such as labors and tools are re-incorporated into the resource allocation sub-model, updated in the event list to meet other requirements of uncompleted operations and recorded in the simulation clock. The subsequent behaviors of materials are further processed according to the definition: some materials are destroyed, and
other materials continue to flow in the next operations.

On the basis of the above calculations, the start and end times of each operation are recorded in the simulation clock, and the total operation time for completing one single aircraft product can be obtained under a specific resource allocation.

(4) Resource allocation sub-model:

This sub-model focuses on the resource pools that comprise these four types of resources and the roles they play in the above three sub-models.

- **AGVs**: The parameters include speed, capacity, and quantity. The AGV_number and AGV_capacity data affects the delivery time of material handling. In the event list, the allocation of specific delivery tasks will be different due to the change of the quantity and capacity of AGVs. Although the speed parameters do not directly affect the allocation of delivery tasks in the event list, the material handling time will also be affected once the distances calculated by the Plant Layout sub-model are fixed. After the completion of the simulation model, the quantity and capacity of AGVs will be fixed, but the speed will remain adjustable to verify the productivity under different allocations.

- **Labors and tools**: Labors for the aircraft final assembly line are organized in different professions. There are many labor parameters such as speed and capacity. In this paper, there will be a fixed number of 17 professions and the impact of different numbers of labors in each profession on productivity will be verified. The same applies to tools, which will have 13 fixed types, with a variable number of tools under each type. They will be called by different operations based on tool_needed and labor_needed data. The priority of these calls will be based on the operation_network data.

- **Buffers**: Once completed in the plant layout of the aircraft final assembly lines, the cost and cycle of change is high. Therefore, the position of buffers does not change in this paper once the parameters are set. However, buffer capacity is frequently adjusted by final assembly line managers to adapt to production. The capacity of buffers directly determines the start time of each handling task in the simulation clock. Only when the buffer capacity is greater than or equal to the mass of materials to be handled, can the handling task be performed by the AGV.

4.1.2. **Linear Regression**

The concept of LR is that there is a relationship between an independent feature and a dependent one. If the two variables move in the same direction, then there is a positive relationship between the two variables, otherwise negative.

The general linear regression formula is as follows:

\[ \hat{y} = \alpha_{t} X + \beta \]

where \( \hat{y} \) is the predicted productivity; \( X = \{x_{1}, x_{2}, \ldots, x_{f}\} \) is the vector of input features; \( \alpha_{t} \) is the coefficients vector; \( \beta \) is the bias value, subject to tuning in order to minimize the defined cost function, e.g. the ordinary least squares as:

\[ \min \frac{1}{2} \sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2} \]
where $y_i$ and $\hat{y}_i$ are the $i$th actual and predicted productivity respectively, and $n$ is the number of data samples.

4.1.3. Polynomial Regression

PR is a technique of regression analysis, where the relationship between the dependent feature and the independent features is described as a function of the independent variables of certain polynomial degrees. LR is a special case of PR, in which the polynomial order equals one. This method is beneficial for describing curvilinear relationships. However, it may easily be over-fitted. Therefore, one should carefully select the polynomial orders for PR. The model with the $d$th degree polynomial regression is constructed as:

$$\hat{y} = \beta + \sum_{i=1}^{d} \alpha_i X^i$$

(3)

where $\alpha_i$ is the corresponding coefficients vector to the $i$th polynomial degree. Similar to LR, the coefficients of the PR model are determined by minimizing the cost function, e.g. the ordinary least squares.

4.1.4. Exponential Regression

Exponential Regression model assumes that the relationship between the dependent and independent features follows an exponential function. The model of the exponential regression can be defined as:

$$\hat{y} = a \exp(\beta + \sum_{i=1}^{d} \alpha_i X^i)$$

(4)

4.1.5. Multilayer perceptron

Artificial Neural Network allows learning from available data [67]. As a class of feedforward Artificial Neural Network, MLP consists of at least three layers of nodes: an input layer, a number of hidden layers and an output layer. The nodes are connected between each layer with different weights. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. For instance, the commonly used three-layer MLP model is constructed as:

$$\hat{y} = (XW_h + b_h)W_o + b_o = XW_hW_o + b_hW_o + b_o$$

(5)

where $W_h$ and $b_h$ are the weight and bias of the hidden layer respectively. $W_o$ and $b_o$ are the weight and bias of the output layer respectively. MLP trains the network via different optimization and search algorithms, including the widely adopted back propagation algorithm [68].

4.1.6. Gradient Boost Regression Tree

GBRT is a flexible non-parametric statistical learning technique for regression. It builds the model in a stage-wise fashion, and allows optimization of an arbitrary differentiable cost function [69]. Like other boosting methods, gradient boosting combines weak learners into a single strong learner in an iterative manner. At each stage $s$, there exists a current predictive
model $F_s$ that can be further improved by adding an estimator $h$, $F_{s+1}(X) = F_s(X) + h$.

This implies that the selection of estimator $h$ can be expressed as:

$$h(X) = y - F_s(X)$$  \hspace{1cm} (6)

4.1.7. Random Forest

RF is a predictive model made up of many decision trees [70]. RF can usually generate good results even without tuning the hyperparameters [71]. Each tree in a RF learns from a set of randomly sampled data during the training process, a technique known as bagging. Notice the samples would be repeatedly applied in a single tree, so that the entire forest will have lower variance without increasing the bias. The predicted productivity is obtained through averaging predictions of each decision tree in RF. First, a small subgroup of independent features is randomly selected. Next the node is split with the best feature among the small number of randomly selected features. After splitting, a new list of eligible features is chosen arbitrarily. This process continues until the tree is completely grown. Ideally, in every terminal node there will be only one observation. As the number of features increases, the eligible feature set will be quite different from node to node. Nevertheless, significant features finally appear in the tree and their respective success in prediction will lead to more reliability.

4.2 Performance Indicators

In this section, we provide seven prediction performance metrics to evaluate and compare the abovementioned predictive models. The first four indicators are widely used and can directly describe the performance without discussing the distinctiveness of features, data scale and domain particularity. The fifth indicator is based on the experience and practice of the aircraft final assembly lines, which is inspired by the current management practice based on error tolerance. A set of error thresholds is therefore used to analyze the proportion of prediction cases under different error thresholds.

In addition, although some performance analysis methods do not take the form of indicators, they are adopted to be fully reveal the capability of productivity prediction models.

4.2.1. Accuracy

The accuracy of predictive models measures the relative percentage difference between the predicted and actual productivity. It is defined as:

$$\text{accuracy} = (100 - 1/n \sum_{i=1}^{n} (y_i - \hat{y}_i) / y_i )\%$$  \hspace{1cm} (7)

where $n$ is the number of data samples.

4.2.2. $R^2$

$R^2$, namely the coefficient of determination, denotes the variation in the dependent variable explained by the independent variables. The value of $R^2$ is no more than 1, which is the best possible value. Notice it could be negative when a model tries to fit nonlinear functions.
to sampled data.

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \frac{1}{n} \sum_{i=1}^{n} y_i)^2}
\] (8)

4.2.3. Mean Absolute Error

Mean Absolute Error (MAE) is defined to measure the average absolute deviations between the predicted and actual productivity. It is formulated as

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|
\] (9)

4.2.4. Root Mean Squared Error

Similar to MAE, Root Mean Squared Error (RMSE) is also an important metric to evaluate the model performance. By squaring the errors, it gives a greater weighting to data points with a larger error. It is characterized as

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\] (10)

4.2.5. Prediction accuracy within 5%, 10% and 15%

In practical productivity prediction of aircraft final assembly lines, small prediction errors are tolerable. Therefore, it is more meaningful to provide a metric that indicates the prediction accuracy within a set of predefined error thresholds to gain a full picture of the productivity of the predictive models. In this study, we set the thresholds as 5%, 10% and 15% respectively.

5 Computational results and discussions

5.1. Experimental setup

The computational experiments are conducted on a laptop with Intel Core i7-7820HQ CPU at 2.90 GHz and 32 GB of RAM. The predictive models are all implemented in Python using the Scikit-Learn library with default parameters.

To fairly compare different predictive models, the performance indicators presented in the tables below are the average values for ten independent runs. For each run, the training dataset is randomly selected as 70% of the historical data from Type Basic. The testing datasets include three parts: the remaining 30% of the historical data from Type Basic, the historical data of Types A and B.

5.2. Performance of different predictive models

The results of testing performance on Type Basic are provided in Tables 6 and 7. Overall, simulation outperforms all other modelling approaches. However, simulation spends much more time than other approaches. When there is requirement to evaluate a large number of pre-planned solutions before production, the real-time performance of simulation is often not sufficient.

Apart from the simulation method, RF and GBRT outperform other modelling approaches.
Once trained, both of them can be run in real-time. RF also has a very similar performance on the training and testing datasets. Carefully comparing the error between the prediction made by the RF and GBRT models and the actual values during testing, RF provides more predictions within the 5% and 10% error thresholds. To sum up, RF is a good choice for some productivity prediction situations when there are strict requirements in real-time performance (under 2 sec) and high prediction accuracy (i.e., more predictions within the 5% or 10% error thresholds).

The testing performances of the PR models are varied. PR has most predictions on the testing set within the 5% error threshold; on the other hand, PR has least predictions within the 10% error thresholds, indicating that PR is not a suitable alternative for the situations when the error is required to be within 10%. Considering other indicators, the overall performance of PR is not as good as that of RF and GBRT. LP has the poorest results, and its performances difference in predicting the training and testing sets are the most significant compared with other approaches, indicating that the relationship between independent features and productivity cannot be described perfectly by linear correlation.

### Table 6 Training performance of 6 predictive models on Type Basic

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>R²</th>
<th>MAE</th>
<th>RMSE</th>
<th>Error&lt;5%</th>
<th>Error&lt;10%</th>
<th>Error&lt;15%</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>81.22%</td>
<td>0.54</td>
<td>3.39</td>
<td>3.09</td>
<td>43.51%</td>
<td>70.52%</td>
<td>83.95%</td>
<td>15.46 sec</td>
</tr>
<tr>
<td>PR</td>
<td>83.52%</td>
<td>0.69</td>
<td>3.28</td>
<td>3.02</td>
<td>45.72%</td>
<td>72.75%</td>
<td>88.65%</td>
<td>18.28 sec</td>
</tr>
<tr>
<td>ER</td>
<td>81.57%</td>
<td>0.67</td>
<td>3.34</td>
<td>3.04</td>
<td>45.03%</td>
<td>63.90%</td>
<td>90.83%</td>
<td>18.96 sec</td>
</tr>
<tr>
<td>MLP</td>
<td>83.72%</td>
<td>0.71</td>
<td>3.13</td>
<td>3.01</td>
<td>43.69%</td>
<td>68.63%</td>
<td>81.55%</td>
<td>43.54 sec</td>
</tr>
<tr>
<td>GBRT</td>
<td>85.76%</td>
<td>0.75</td>
<td>2.94</td>
<td>2.93</td>
<td>52.56%</td>
<td>73.60%</td>
<td>90.42%</td>
<td>49.27 sec</td>
</tr>
<tr>
<td>RF</td>
<td>85.28%</td>
<td><strong>0.76</strong></td>
<td><strong>2.93</strong></td>
<td>2.94</td>
<td>52.20%</td>
<td><strong>74.32%</strong></td>
<td>89.82%</td>
<td>40.44 sec</td>
</tr>
</tbody>
</table>

### Table 7 Testing performance of 7 predictive models on Type Basic

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>R²</th>
<th>MAE</th>
<th>RMSE</th>
<th>Error&lt;5%</th>
<th>Error&lt;10%</th>
<th>Error&lt;15%</th>
<th>Run Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simu</td>
<td>94.65%</td>
<td><strong>0.87</strong></td>
<td><strong>2.86</strong></td>
<td>1.96</td>
<td><strong>84.50%</strong></td>
<td><strong>92.50%</strong></td>
<td><strong>94.00%</strong></td>
<td>243.08 sec</td>
</tr>
<tr>
<td>LR</td>
<td>74.75%</td>
<td>0.52</td>
<td>3.32</td>
<td>3.11</td>
<td>40.50%</td>
<td>58.42%</td>
<td>67.33%</td>
<td>0.35 sec</td>
</tr>
<tr>
<td>PR</td>
<td>78.98%</td>
<td>0.79</td>
<td>3.24</td>
<td>3.09</td>
<td>53.72%</td>
<td>55.22%</td>
<td>77.45%</td>
<td>0.55 sec</td>
</tr>
<tr>
<td>ER</td>
<td>77.15%</td>
<td>0.78</td>
<td>3.33</td>
<td>3.11</td>
<td>43.49%</td>
<td>66.62%</td>
<td>81.12%</td>
<td>0.47 sec</td>
</tr>
<tr>
<td>MLP</td>
<td>80.09%</td>
<td>0.77</td>
<td>3.17</td>
<td>3.09</td>
<td>51.10%</td>
<td>62.29%</td>
<td>79.50%</td>
<td>1.48 sec</td>
</tr>
<tr>
<td>GBRT</td>
<td>82.67%</td>
<td>0.82</td>
<td>3.16</td>
<td>3.07</td>
<td>52.36%</td>
<td>68.85%</td>
<td>86.51%</td>
<td>1.76 sec</td>
</tr>
<tr>
<td>RF</td>
<td>82.87%</td>
<td>0.81</td>
<td>3.14</td>
<td>3.04</td>
<td>52.45%</td>
<td>75.01%</td>
<td>86.17%</td>
<td>1.58 sec</td>
</tr>
</tbody>
</table>

5.3. Performance in different productivity ranges

As shown in Table 8, by investigating the performance of different productivity ranges, firstly we found that although the simulation model has the best performance, it is unable to make good predictions in the extremely low productivity range. The simulation model shows a rapid improvement in prediction as productivity increases. This observation is echoed by the assembly line managers for two possible reasons. First, for the extremely low productivity range, the randomness of features has a greater impact on productivity, while mass production greatly reduces the impact of randomness. Second, the construction of simulation model is based on domain knowledge. Different from other approaches that learn from the historical data within the extremely-low productivity range, the construction of the simulation model takes little advantage of the available data or prior knowledge in that range.
Table 8 Accuracy in different productivity ranges on Type Basic

<table>
<thead>
<tr>
<th>Model</th>
<th>EL</th>
<th>Low</th>
<th>Normal</th>
<th>High</th>
<th>EH</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simu</td>
<td>74.78%</td>
<td>99.08%</td>
<td>98.12%</td>
<td>85.16%</td>
<td>94.46%</td>
<td>94.65%</td>
</tr>
<tr>
<td>LR</td>
<td>70.21%</td>
<td>78.00%</td>
<td>79.05%</td>
<td>73.76%</td>
<td>82.24%</td>
<td>77.40%</td>
</tr>
<tr>
<td>PR</td>
<td>75.06%</td>
<td>79.64%</td>
<td>81.20%</td>
<td>78.83%</td>
<td>83.34%</td>
<td>80.53%</td>
</tr>
<tr>
<td>ER</td>
<td>72.49%</td>
<td>79.60%</td>
<td>78.00%</td>
<td>77.43%</td>
<td>84.71%</td>
<td>78.32%</td>
</tr>
<tr>
<td>MLP</td>
<td>77.26%</td>
<td>84.67%</td>
<td>84.23%</td>
<td>83.75%</td>
<td>87.15%</td>
<td>80.65%</td>
</tr>
<tr>
<td>GBRT</td>
<td>80.63%</td>
<td>85.94%</td>
<td>89.10%</td>
<td>82.58%</td>
<td>88.79%</td>
<td>83.49%</td>
</tr>
<tr>
<td>RF</td>
<td>81.03%</td>
<td>85.49%</td>
<td>85.93%</td>
<td>83.80%</td>
<td>88.95%</td>
<td>82.83%</td>
</tr>
</tbody>
</table>

* The productivity is divided into five regions, which respectively represent the five stages of productivity control: extremely low (EL), low, normal, high and extremely high (EH).

As shown in Fig. 2, Machine learning methods including MLP, GBRT and RF maintain a steady performance of accuracy across all productivity ranges. However, there is a significant difference in their prediction accuracy across different productivity ranges. This difference may be driven by the uneven distribution of historical data across different productivity ranges. By investigating the data distribution shown in Fig. 3, the amount of historical data in the EL range is far less than the number of data in other ranges. However, even for the EL range, MLP, GBRT and RF models can still make better predictions than the simulation model. The phenomenon of the unbalanced datasets observed in this paper deserves more attention in future studies (see discussions in Section 6), because it is ubiquitous to all aircraft final assembly lines.

Comparing the three regression models, including LR, PR and ER, PR has a higher accuracy except in the EH range. While ER has the highest accuracy in the EH range among the three regression models. Furthermore, the accuracy of the predicted values of ER and LR in different productivity ranges fluctuates more dramatically, indicating that the relationship between productivity and independent features may follow different laws under different productivity ranges.

Fig. 2. The accuracy in different productivity ranges on Type Basic
Fig. 3. The distribution of Type Basic’s historical data in different productivity ranges

5.4. Generalization ability for different aircraft final assembly lines

The results of test performance on Types A and B are provided in Tables 9 to 11. Overall, the same model performs better on Type A than it does on Type B in terms of accuracy, indicating that the difference between these two types and Type Basic is the dominant factor impacting the generalization performance of the predictive models.

Simulation outperforms other predictive models in generalization ability when the accuracy metric is used. The reason may be that although there are differences among Types A, B and Basic, the simulation model keeps as much professional domain knowledge of resources, materials and processes as possible which are largely applicable in other types. However, simulation does not give the best performance in the drop ratio on Types A and B. The drop ratio defines the performance degradation when using the models built for Type Basic for Types A and B prediction. Results confirm there exists a larger technological difference between Types B and Basic than that of Types A and Basic.

Table 9 Accuracy of 7 predictive models among 3 types

<table>
<thead>
<tr>
<th>Model</th>
<th>Type Basic</th>
<th>Type A</th>
<th>Drop ratio</th>
<th>Type B</th>
<th>Drop ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simu</td>
<td>94.65%</td>
<td>89.59%</td>
<td>5.35%</td>
<td>73.58%</td>
<td>22.26%</td>
</tr>
<tr>
<td>LR</td>
<td>74.75%</td>
<td>68.11%</td>
<td>8.88%</td>
<td>47.83%</td>
<td>36.01%</td>
</tr>
<tr>
<td>PR</td>
<td>78.98%</td>
<td>75.64%</td>
<td>4.23%</td>
<td>59.77%</td>
<td>24.32%</td>
</tr>
<tr>
<td>ER</td>
<td>77.15%</td>
<td>73.03%</td>
<td>5.34%</td>
<td>50.87%</td>
<td>34.06%</td>
</tr>
<tr>
<td>MLP</td>
<td>80.09%</td>
<td>74.11%</td>
<td>7.47%</td>
<td>68.39%</td>
<td>14.61%</td>
</tr>
<tr>
<td>GBRT</td>
<td>82.67%</td>
<td>71.59%</td>
<td>13.40%</td>
<td>68.06%</td>
<td>17.67%</td>
</tr>
<tr>
<td>RF</td>
<td>82.87%</td>
<td>74.52%</td>
<td>10.08%</td>
<td>68.72%</td>
<td>17.07%</td>
</tr>
</tbody>
</table>

The following discussion will exclude simulation models and focus only on predictive performance on Type A. PR outperforms other predictive models in terms of the accuracy and drop ratio. It can also be seen that on Type A, all methods except LP can achieve more than 70% accuracy, indicating the feasibility of the data-driven approaches on the types that have little difference from the basic type. Carefully comparing the error between the prediction made by the predictive models and the actual values during testing, GBRT makes most predictions within the 5% error threshold. However, GBRT does not give accurate predictions on Type A across all performance indicators. To sum up, PR is a good choice for productivity prediction on Type A when there are strict requirements in real-time performance (under 1 sec) and relatively high prediction accuracy (less than 15%).
Theoretically speaking, RF indeed show a higher generalization ability in terms of the accuracy. On the other hand, MLP has the least drop ratio and the second highest accuracy. MLP, GBRT and RF are the methods that achieved more than 68% in accuracy, indicating that in the case of productivity prediction with a relatively slightly lower precision requirement, the machine learning approaches seem feasible. Although in terms of accuracy, they appear to be inferior to simulation methods, for the final assembly lines of more divergent types like Type B, there seems to be a tendency to catch up with or even outperform the simulation method.

### Table 11 Testing performance of 7 predictive models on Type B

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>R²</th>
<th>MAE</th>
<th>RMSE</th>
<th>Error&lt;5%</th>
<th>Error&lt;10%</th>
<th>Error&lt;15%</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simu</td>
<td>73.58%</td>
<td>0.53</td>
<td>3.08</td>
<td>3.10</td>
<td>52.27%</td>
<td>86.36%</td>
<td>90.28%</td>
<td>243.22 sec</td>
</tr>
<tr>
<td>LR</td>
<td>47.83%</td>
<td>0.30</td>
<td>5.29</td>
<td>3.58</td>
<td>12.00%</td>
<td>39.50%</td>
<td>51.04%</td>
<td>0.34 sec</td>
</tr>
<tr>
<td>PR</td>
<td>59.77%</td>
<td>0.48</td>
<td>3.99</td>
<td>3.32</td>
<td>28.79%</td>
<td>56.86%</td>
<td>69.41%</td>
<td>0.55 sec</td>
</tr>
<tr>
<td>ER</td>
<td>50.87%</td>
<td>0.33</td>
<td>5.11</td>
<td>3.56</td>
<td>17.71%</td>
<td>52.59%</td>
<td>71.19%</td>
<td>0.47 sec</td>
</tr>
<tr>
<td>MLP</td>
<td>68.39%</td>
<td>0.48</td>
<td>3.12</td>
<td>3.26</td>
<td>29.24%</td>
<td>58.33%</td>
<td>76.29%</td>
<td>1.48 sec</td>
</tr>
<tr>
<td>GBRT</td>
<td>68.06%</td>
<td>0.38</td>
<td>3.16</td>
<td>3.27</td>
<td>22.19%</td>
<td>53.95%</td>
<td>78.64%</td>
<td>1.76 sec</td>
</tr>
<tr>
<td>RF</td>
<td>68.72%</td>
<td>0.49</td>
<td>3.11</td>
<td>3.24</td>
<td>30.80%</td>
<td>64.24%</td>
<td>78.84%</td>
<td>1.65 sec</td>
</tr>
</tbody>
</table>

In conclusion, the simulation method performs the best in generalization ability. The reason may be that the simulation model itself retains the process flow, factory layout and other professional content. Under the premise of low real-time requirement, the simulation method is the first choice. Apart from the simulation method, PR showed a better performance on Type A, while machine learning methods showed a lower drop ratio on both Types A and B. RF has a better generalization ability on Type B, indicating for types of larger difference from Type Basic, machine learning is a competent solution. Theoretically speaking, RF indeed shows a high tolerance for noisy data in many papers [72]. This characteristic, which implicitly creates multiple joint features and can solve nonlinear problems, is worthy of further exploration and application.

### 6. Conclusions

In this article, we introduce a more comprehensive performance comparison for seven different productivity predictive models. Some guidelines are provided for the selection of predictive models in term of modified types, productivity ranges and error thresholds. Issues such as imbalanced historical data pertaining to the aircraft final assembly line productivity predictions that require further investigations are also pointed out.
(1) For productivity predictions of aircraft final assembly lines of multiple types, the feasibility of the investigated predictive methods in this paper is shown using three aircraft types. The knowledge-based simulation method has advantages in predicting the modified types, because the model retains domain knowledge that is applicable across different types. This performance advantage over other predictive models is more obvious for the types of less modification from the basic type. For predictions of more extensively modified types, ensemble methods such as RF show a tendency to surpass the simulation approach.

(2) This paper demonstrates the necessity of a comprehensive performance evaluation for productivity predictive models. There are limitations in evaluating the performance of predictive models by using only conventional performance indicators presented in an averaged manner. As a matter of fact, there is no predictive model that can completely outperform the others across all productivity ranges. The simulation method, even though it outperforms all other methods in terms of its averaged performance, does not provide good performance in the extremely low range. Therefore, it is necessary for practitioners in the aircraft final assembly lines to make sensible choice from different predictive models for different productivity ranges. Similarly, if time budget is insufficient, it is important to figure out the tolerance for prediction error in order to choose the right data-driven models. For example, PR can make most predictions within a relatively small error (5%) threshold, but the overall accuracy is not the best. On the contrary, GBRT is able to perform well if the tolerance is big (e.g. 15%). However, it performs the worst within smaller error thresholds.

(3) The problem of imbalanced dataset in the aircraft final assembly lines is further discussed. This paper points out the causes of the imbalanced dataset by studying the distribution of historical data in different productivity ranges. The imbalance dataset is due to large variation in aircraft productivity and the existence of extreme orders. The impact of this imbalance on the performance of the predictive models is discussed.

The study in this paper paves several potential ways for future research.

(1) The predictive models in this paper can be integrated into the multi-objective optimization framework to further optimize the plant layout, operation process, material handling and resource allocation. Data Twins, as a new development trend of the simulation method, can enable the site situation in aircraft final assembly lines to match with the simulation model continuously, leading to a real-time feedback for the implementation of optimization scheme [73].

(2) The interpretability of productivity predictive models should be further enhanced. This paper makes full reference to the management methods in aircraft final assembly lines, such as productivity ranges, error thresholds and modified types. However, little knowledge can be gained from the predictive model itself to provide guidelines to the management. Thus, one of the future works is to introduce models such as the fuzzy rule-based systems [74] that are interpretable into productivity predictions. The aim is not only to provide a more accurate estimation but also interpret the effect of resources on productivity from the predictive model structure.

(3) The problem addressed in this paper also imposes a big challenge for data-driven models, especially the lack of historical data in some productivity ranges. In such a case, transfer learning [75] can be used to leverage the advantage offered by the types that do have sufficient data. Another important approach that may contribute towards addressing this problem is
granular computing. Granular computing is an important technique for determining optimal granularity under unbalanced data sets. Granular computing is a theory of computation that simulates human thinking and reasoning by processing information at different levels of precision [76]. In the productivity predictive model, granular computing method can be used to train the approximators by the data granularity so that the training data is granulated at different levels, which improves the performance for imbalanced datasets.
**Acknowledge**

The authors would like to thank the National High-tech Research and Development Program of China (2019YFB1707501) for their financial support for this project, as well as the anonymous reviewers for their comments and suggestions to improve the manuscript.
References


Appendix A. The 46 inputs and 1 output of the historical data

<table>
<thead>
<tr>
<th>Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Min.</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Types</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
</tr>
<tr>
<td>Index</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>Inputs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>28</td>
<td>30</td>
<td>153.0</td>
<td>153.0</td>
<td>153.0</td>
<td>153.0</td>
<td>40</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>Min.</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Types</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Double</td>
<td>Double</td>
<td>Double</td>
<td>Double</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
</tr>
<tr>
<td>Index</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>Inputs</td>
<td>Buffer 08</td>
<td>Buffer 09</td>
<td>Buffer 10</td>
<td>Buffer 11</td>
<td>Buffer 12</td>
<td>Tool 01</td>
<td>Tool 02</td>
<td>Tool 03</td>
<td>Tool 04</td>
<td>Tool 05</td>
<td>Tool 06</td>
<td>Tool 07</td>
</tr>
<tr>
<td>Max.</td>
<td>24</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>25</td>
<td>12</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Min.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Types</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
</tr>
<tr>
<td>Index</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Inputs</td>
<td>Tool 08</td>
<td>Tool 09</td>
<td>Tool 10</td>
<td>Tool 11</td>
<td>Tool 12</td>
<td>Tool 13</td>
<td>AGV 1</td>
<td>AGV 2</td>
<td>AGV 3</td>
<td>AGV 4</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Max.</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>28</td>
<td>30</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Min.</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Types</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Double</td>
<td>Double</td>
<td>Double</td>
<td>Double</td>
<td>Double</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

* Corresponding author. E-mail address: jun.chen@qmul.ac.uk; Phone: +44 (0)20 7882 8873; Webpage: https://www.sems.qmul.ac.uk/staff/jun.chen#
### Appendix B. Figure Descriptions

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Figure Title</th>
<th>Figure Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sub-system diagram of productivity predictive simulation model</td>
<td>Figure 1 shows the four sub-models, their elements, and relationships among these elements in the productivity predictive simulation model constructed in this paper.</td>
</tr>
<tr>
<td>2</td>
<td>The accuracy in different productivity ranges on Type Basic</td>
<td>Figure 2 shows the accuracy change of the 7 productivity predictive models with the change of productivity. The figure is presented separately with 5 productivity ranges. The X-axis represents the actual productivity, and the Y-axis represents the prediction accuracy. It is worth pointing out that the Y-axis range in each chart is different. The 7 predictive models are distinguished by different colors.</td>
</tr>
<tr>
<td>3</td>
<td>The distribution of Type Basic’s historical data in different productivity ranges</td>
<td>Figure 3 shows the distribution of historical data across the 5 productivity ranges, which are extremely low, low, normal, high and extremely high. The amount of historical data within these five ranges is marked.</td>
</tr>
</tbody>
</table>
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: