



**HAL**  
open science

## **A hybrid knowledge extraction approach to support early concurrent engineering in the aerospace industry**

Eliott Duverger, Eric Levrat, Alexis Aubry, Rebeca Arista

### ► **To cite this version:**

Eliott Duverger, Eric Levrat, Alexis Aubry, Rebeca Arista. A hybrid knowledge extraction approach to support early concurrent engineering in the aerospace industry. 22nd IFIP International Conference on Product Lifecycle Management, PLM 2025, Jul 2025, Séville, Spain. <hal-05279680>

**HAL Id: hal-05279680**

**<https://hal.science/hal-05279680v1>**

Submitted on 23 Sep 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

# A hybrid knowledge extraction approach to support early concurrent engineering in the aerospace industry

Elliott Duverger<sup>1</sup> ✉<sup>[0009-0000-3344-7325]</sup>, Eric Levrat<sup>1</sup><sup>[0000-0001-5884-2938]</sup>,  
Alexis Aubry<sup>1</sup><sup>[0000-0002-7335-821X]</sup>, and Rebeca Arista<sup>2</sup><sup>[0000-0001-9382-1593]</sup>

<sup>1</sup> Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France  
{elliott.duverger,eric.levrat,alexis.aubry}@univ-lorraine.fr

<sup>2</sup> Airbus SAS, Blagnac 31700, France  
rebeca.arista@airbus.com

**Abstract.** Early concurrent engineering necessitates a high degree of knowledge to assess the impact of design decisions. Part of Knowledge-Based Engineering (KBE), recent methodologies employing ontologies have gained more interest over the past years due to their shared conceptualizations, knowledge representations, and reasoning capabilities. Merged with Model-Based Systems Engineering (MBSE) activities, they enable knowledge-based systems to be seamlessly integrated within Industry 4.0. The incorporation of KBE within MBSE is of major importance in complex engineering sectors with an increasingly competitive market, such as the aerospace industry. This paper leverages the application of a hybrid knowledge extraction approach to support cross-domain collaboration during the conceptual design phase of a new aircraft.

**Keywords:** Concurrent Engineering · Knowledge-Based Engineering · Ontology-Based Engineering · Model-Based Systems Engineering · Conceptual Design · Knowledge-acquisition

## 1 Introduction

Concurrent Engineering is a “*systematic approach to the integrated, concurrent design of products and their related processes, including manufacturing and support*” [19]. This principle promotes the consideration of elements from all the product lifecycle and is therefore closely linked to PLM approaches. The ability to anticipate issues in the domains of manufacturing and maintenance can significantly enhance the overall time, cost and quality of aircraft development. However, during the conceptual design phase, this approach relies heavily on expert knowledge to assess the impact of design decisions.

The ability to elicit, manage, and share engineering knowledge is key to meet the challenges of an increasingly competitive industry. Nevertheless, a substantial portion of knowledge remains concealed in the minds of experts or buried in databases.

Knowledge-Based Engineering (KBE) addresses the challenges associated with

the capture, formalization, management, and diffusion of knowledge within complex organizations. Nevertheless, the investments in KBE have not always yielded the anticipated outcomes, thereby hindering their adoption.

The democratization of ontologies in recent years has introduced Ontology-Based Engineering (OBE). OBE has emerged as a solution to address the conventional challenges associated with expert systems while also generating new opportunities to incorporate KBE with the industry digitization.

The remainder of this paper is structured as follows: Section 2 provides an overview of the related work addressing KBE with knowledge acquisition methods and processes. Section 3, details the method employed to capture knowledge supporting early concurrent engineering. The results of a preliminary implementation of the aforementioned method are presented in Section 4. Finally, discussions are held in Section 5 while Section 6 concludes this paper.

## 2 State of the art

### 2.1 Knowledge-Based Engineering

Knowledge-Based Engineering principles started to gain importance during the 90's. Two fundamental contributions of this era were CommonKADS [15] and MOKA [16] methodologies which paved the way for knowledge elicitation and formalization to support knowledge-based systems development.

Addressing concurrent product design and manufacturing, a web-based framework leverages knowledge bases to support human decisions in early stage [20]. More recently and within the aerospace industry, a structured approach named NOMAD [4] has been proposed to develop KBE systems. It ensures that domain knowledge is effectively captured, organized, and implemented to support engineering decision-making and problem-solving processes.

However, the lack of robust knowledge acquisition techniques, knowledge management strategies, and formalization methods was identified as an issue hindering Knowledge-Based Engineering [9, 8].

Nonetheless, the democratization of ontologies has strengthened knowledge-based methods by providing structured knowledge representation, automated reasoning, and cross-system interoperability [21]. The integration of ontologies within KBE thus introduced Ontology-Based Engineering (OBE) approaches. According to Arista et al. [2] "*OBE systems describe a particular and well-defined domain from different perspectives: classification, interoperability, behavior, and semantics*". With structured methods to capture, manage, and share domain knowledge, OBE enhances collaboration for complex system design.

Models for Manufacturing (MfM) is an OBE methodology that leverages an agnostic approach for knowledge reuse to support complex manufacturing system design in the aerospace industry [12]. MfM has been used at the core of Model-Based Systems Engineering (MBSE), incorporating shared conceptualizations, a common understanding and reasoning capabilities to support standardized systems engineering processes [7, 2].

However, Mas et al. [11] consider that OBE systems still hold some limitations, in particular regarding the use of various knowledge sources (e.g. domain experts, existing documentation, and industrial databases). In this paper, the authors have identified new technologies such as Big Data (BD) and Artificial Intelligence (IA) as important parts for knowledge acquisition.

## 2.2 Knowledge acquisition methods

The majority of knowledge acquisition methods consider humans as their primary source of knowledge [1, 8].

In the context of PLM and the data-driven manufacturing era, the applications of machine learning algorithms used for extracting valuable insights across life-cycles are continuously increasing [18, 14].

Regression analysis has been used to estimate the cost of aerospace components in the early conception stage to offer better cost understanding [13]. Also, Tang et al. [17] have employed surrogate models to generate aircraft performances as well as supply chain outputs. In this paper the authors employ a combination of Design of Experiments (DoE) and sensitivity analysis to extract key cross-domain drivers supporting decision-making in early design phase.

Random Forest (RF) models have been leveraged to support design space exploration of Turbine Rear Structures [5]. A highlight of this study is the extraction of design parameters' importance and *if-then* rules. Following a similar approach, Decision Trees were employed to extract production rules [1]. In this paper, the authors identify the need for a bridge between manual knowledge curation and automatic data-driven knowledge generation. Recently, Li et al. [10] have proposed a scalable knowledge capture blending human expertise with computational methods to improve knowledge management practices across various industries.

## 3 Hybrid knowledge acquisition method

A previous contribution has leveraged MfM to support early concurrent engineering [6]. In this paper, several ontological models are described in a MBSE environment as an agnostic approach to manage knowledge and interfaces between tools. To achieve, a hybrid knowledge acquisition method is delineated by employing formalized expert knowledge to drive machine learning studies.

The objectives of this method, as illustrated in Figure 1, are twofold: first, to strengthen interconnections between aircraft and manufacturing domains by formalizing cross-domain knowledge patterns, and second, to leverage these patterns as the foundation of data-driven activities.

The expert-centric knowledge formalization of cross-domain patterns allows the identification of key concepts and parameters through individual semi-structured interviews with experts and architects following a 5-step protocol (i.e., *contextualization, enumeration, extension, conclusion, and verification* [6]). This manual

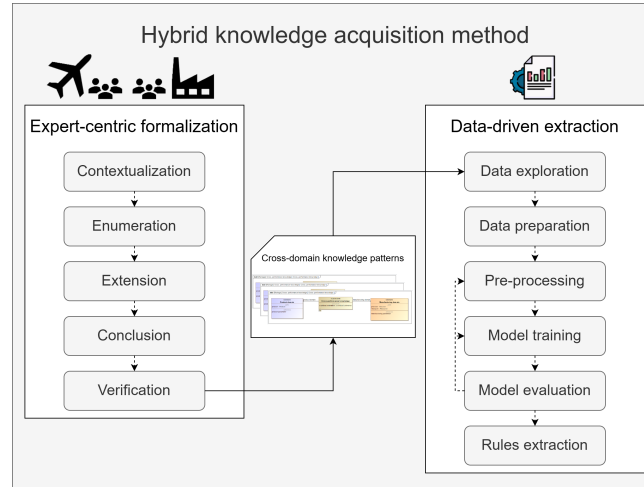


Fig. 1: Hybrid knowledge extraction approach

process aims to benefit from experts' experience and intuition to orient data-driven knowledge extraction on a reduced scope.

The data-driven knowledge extraction is grounded in the SIRUS algorithm [3] that exploits the inner structure of Random Forest models to generate a stable rule set which can then be integrated into the ontology. To achieve a proper engineering rule extraction, an iterative process of *data exploration*, *data preparation*, *pre-processing*, *model training*, and, *model evaluation* is identified.

By combining the intuition of domain experts with the learning scalability of large industrial databases, this approach is leveraged through a case study in the remainder of this paper.

## 4 Case study

### 4.1 Expert-centric formalization

Expert-centric formalization was conducted with several industrial experts and architects in order to identify key knowledge supporting early concurrent engineering. From these interviews, a consensus has emerged toward the workload estimation.

During the conceptual design phase, the cornerstone of concurrent engineering activities between product and manufacturing engineers revolves around workload estimation. While generating aircraft concepts, architects are tasked with timing the associated manufacturing processes. This allows for a preliminary dimensioning of the manufacturing system which can subsequently influence cross-domain trade-off activities toward a global optimum. However, the high degree of abstraction inherent to the early stages of development prevent engineers from feeling sufficiently confident in their estimations.

As illustrated in Figure 2, this cross-domain knowledge pattern has identified the product shape, weight, and, material in combination with the category of manufacturing process as key characteristics for workload estimation. This knowledge is regarded as a “*know-how*” engineering skill that is challenging to elicit using traditional techniques. It is reasonable to think that product components with similar features produced through identical manufacturing processes require a comparable amount of workload. Furthermore, the workload estimations performed during the early stage of co-architecture do not necessitate a high precision. Inexact predictions with a correct order of magnitude are considered sufficient for proceeding to further iterations loops. Data-driven studies on past program databases are developed in the next section for extracting insights that support engineers in their early co-architecture activities.

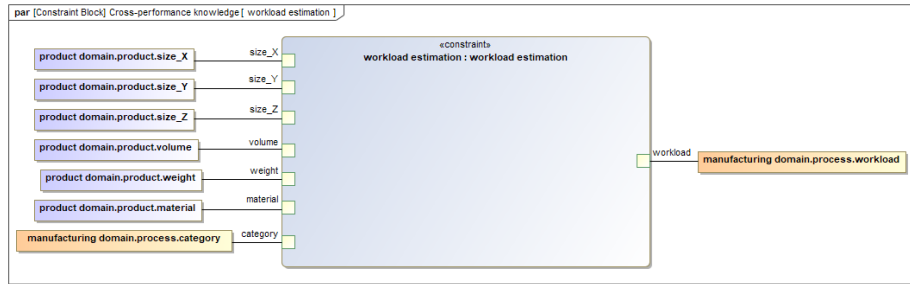


Fig. 2: Cross-domain knowledge instantiated for workload estimation

## 4.2 Data-driven study

### Data exploration

In the industry, real-world data is frequently not stored with knowledge capitalization considerations. Consequently, mostly heterogeneous, non-indexed and non-standardized sets of data are available. Often coded for operational activities and tools, the information is not readily interpretable by humans. Exploration and comprehension of the various databases is guided by the workload estimation pattern illustrated in *Figure 2*.

### Data preparation

The various selected datasets are transformed, joined, and filtered to obtain meaningful data in the desired format. With regard to the workload estimation, legacy databases, containing aircraft elements characteristics, are linked to operational databases describing their related manufacturing operations and workload time.

### Pre-processing

Once all the identified drivers have been retrieved and structured for AI training,

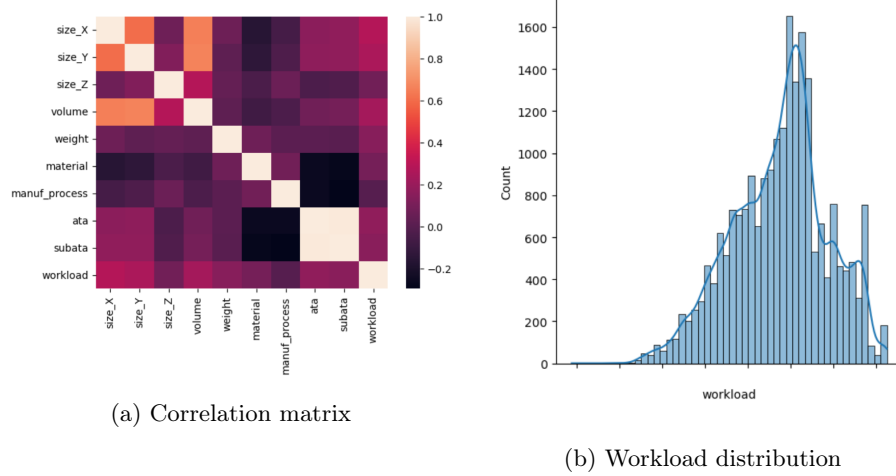


Fig. 3: Input dataset analysis

data visualization and analysis are conducted in order to gain a better understanding of the input dataset and perform appropriate cleaning activities. Correlation analysis is performed first as it presents two objectives: First, it identifies redundancies in the input features allowing dimensionality reduction, and second, it allows to discover important features that are not necessarily obvious. As illustrated in Figure 3a, significant correlation score were found, identifying ATA and subATA information as important features for workload estimation. It is imperative to ensure the balanced representation of the input dataset to avoid over-representation of certain values, which could be a source of bias during training. With the same regard, values that are under-represented are carefully filtered out as they may introduce unnecessary noise. The target value distribution, as illustrated in Figure 3b, is close to a Gaussian curve which does not necessitate further transformation. It should be noted, however, that the values have been removed for confidentiality reasons. Finally, categorical values undergo transformation to ensure comprehension by the model. The OneHotEncoding technique is selected as it avoids the undesirable hierarchy that arises from other encoding techniques with a cost in complexity. This solution is therefore appropriate for datasets with an acceptable number of distinct values.

### Model training

This knowledge extraction method is grounded in the interpretability of Random Forest (RF) models. Nevertheless, RF models need to be parameterized with information such as tree number, maximum depth, split criterion, etc... Through Grid Search cross-validation techniques, sets of predefined parameters are explored to find the combination yielding the best performances. Visualizing

the cross validation scores for specific parameters helps select an optimized configuration while avoiding unnecessary complexity. As illustrated in Figure 4, it can be observed that an increasing tree depth up to 30 improves the model while changes in the forest size exhibit minimal impact between 150 and 300 trees.

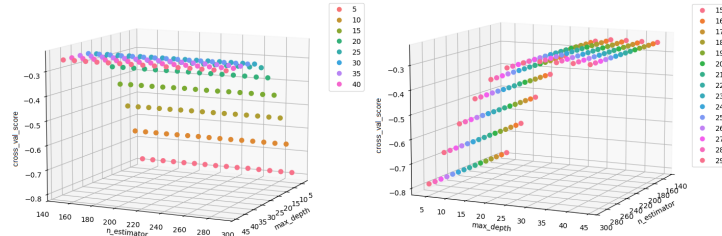


Fig. 4: Gridsearch analysis

**Model evaluation**

In order to ensure confidence in the learning process, various metrics can be selected. The most common metrics used to evaluate regression models are  $R^2$ , which describes the model’s ability to handle variability, and the Root Mean Squared Error (RMSE), which quantifies its performance in absolute terms. However, when considered individually, these metrics are not sufficient to evaluate a regression model. Instead, the combined assessment of high  $R^2$  and low RMSE is considered as an appropriate evaluation metric.

A validation sample, which has not been used during the training stage, is employed to evaluate the model capacity to generalize on new data. As illustrated in the Figure 5, the points represent the predicted workload while their measured value are on the red line. The vertical distance between a point and the red line thus depicts the error of the prediction. A trend delineated by a ma-

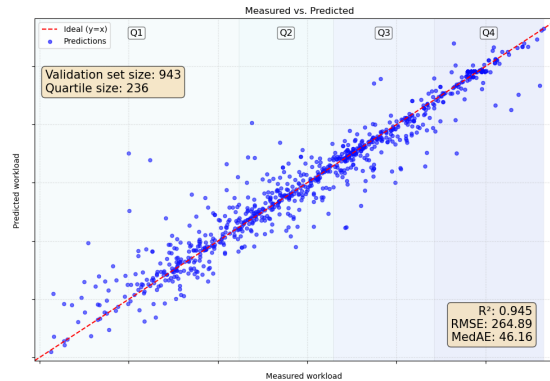


Fig. 5: Model evaluation: performance visualization

jority of points close to the red line indicate a good learning process. With  $R^2$  and RMSE scores of respectively 94% and 264, the model performances are considered satisfying. The next step is therefore to extract *if-then* rules from this model.

### Rules extraction

The SIRUS (Stable Interpretable RULes Set) algorithm has been developed by B enard et al. [3] to extract *if-then* rules from the inner structure of RF models. In this paper the authors leverage the split alignment of each tree composing the forest. By reducing the split values following vigintiles and boolean decompositions for respectively discrete features and categorical features, the estimation model keeps most of its predictive accuracy while being sufficiently stable for rule extraction. Afterwards, an analysis is conducted on each tree path to identify the most recurrent ones. As Table 1 describes the top occurring rules in the forest, it appears that the weight and the volume are the main characteristics for workload estimation.

nb occurrence	rule body	rule head
169	IF weight > 561.95	THEN workload = 429.56
128	IF weight ≤ 561.95	THEN workload = 196.19
98	IF volume ≤ 1968598007.86 AND weight > 561.95	THEN workload = 327.17
71	IF volume ≤ 400229969.33 AND weight ≤ 561.95	THEN workload = 169.36
65	IF volume ≤ 400229969.33 AND weight > 561.95	THEN workload = 212.82
65	IF weight > 707.41	THEN workload = 439.68
63	IF volume ≤ 729882523.49 AND weight ≤ 561.95	THEN workload = 186.65

Table 1: Extracted rules

## 5 Discussions and perspectives

A knowledge acquisition method, considering experts’ intuition and industrial databases as knowledge sources, was applied to extract knowledge supporting early concurrent engineering. Explicit if-then rules describing workload estimation knowledge were extracted for early manufacturing impact analysis. The formalism of these rules is compatible with the the expressiveness of ontological or MBSE languages providing the means to perform instantaneous cross-domain impact assessments during the conceptual design phase thereby supporting early concurrent engineering.

This approach was leveraged in a context of concurrent engineering between aircraft and its manufacturing system. Nevertheless, it can be applied for various kinds of knowledge relevant in all the product lifecycle. The hybrid nature of this acquisition method do not present scalability issues as the knowledge formalized from domain expert permit to reduce the scope of each data-driven study.

However, several concerns persist regarding the applicability of the captured knowledge. As the data-driven studies are based on historical data, its relevance remains related to a certain similarity threshold. Therefore, knowledge management guidelines should be defined for the knowledge relevance and maintenance considering technological and data evolution across the lifecycle.

## 6 Conclusion

Early concurrent engineering demands a high degree of knowledge to assess the impact of design decisions across the product lifecycle. However, this knowledge is often inaccessible, either concealed in the minds of experts or buried in databases. The need for systematic knowledge acquisition and reuse methods is therefore critical in complex engineering sectors such as the aerospace industry. A method has been applied to capture engineering rules that support cross-domain decision-making during the conceptual design phase. This hybrid approach has permitted the identification of key concepts and parameters which were formalized in a cross-domain knowledge pattern. This pattern has then been used as the foundation of data-driven activities resulting in the extraction of engineering knowledge articulated as if-then rules.

The combination of human-centric and data-driven techniques has benefited from the intuition of experts and the learning scalability of AI models. Incorporated in Ontology-Based Engineering methodology, this approach provides new opportunities for knowledge capitalization and strengthens the integration of Knowledge-Based Engineering with Model-Based Systems Engineering environments. The instantaneous execution of the captured knowledge would therefore support collaborative engineering processes between product and manufacturing domains during the conceptual design phase.

**Acknowledgments.** The present work is the result of a collaborative effort between Airbus and the University of Lorraine. The authors would like to express their profound gratitude to all involved colleagues for their invaluable support and contribution during the development process.

**Disclosure of Interests.** The authors have no competing interests to declare relevant to this article's content.

## References

1. Ali, M., Ali, R., Khan, W.A., Han, S.C., Bang, J., Hur, T., Kim, D., Lee, S., Kang, B.H.: A data-driven knowledge acquisition system: An end-to-end knowledge engineering process for generating production rules. *IEEE Access* **6** (2018)
2. Arista, R., Zheng, X., Lu, J., Mas, F.: An ontology-based engineering system to support aircraft manufacturing system design. *Journal of Manufacturing Systems* **68**, 270–288 (2023)
3. B enard, C., Biau, G., Da Veiga, S., Scornet, E.: Interpretable random forests via rule extraction. In: International conference on artificial intelligence and statistics. pp. 937–945. PMLR (2021)
4. Curran, R., Verhagen, W.J., Van Tooren, M.J., Van Der Laan, T.H.: A multidisciplinary implementation methodology for knowledge based engineering: Knomad. *Expert Systems with Applications* **37**(11), 7336–7350 (2010)
5. Dasari, S.K., Cheddad, A., Andersson, P.: Random forest surrogate models to support design space exploration in aerospace use-case. In: Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Proceedings 15. pp. 532–544. Springer (2019)

6. Duverger, E., Aubry, A., Levrat, E., Arista, R.: An ontology-based engineering methodology to support early concurrent engineering. *IFAC-PapersOnLine* (2025)
7. Hu, X., Arista, R., Zheng, X., Lentès, J., Sorvari, J., Lu, J., Ubis, F., Kiritsis, D.: Ontology-based system to support industrial system design for aircraft assembly. *IFAC-PapersOnLine* **55**(2), 175–180 (2022)
8. Kuegler, P., Dworschak, F., Schleich, B., Wartzack, S.: The evolution of knowledge-based engineering from a design research perspective: Literature review 2012–2021. *Advanced Engineering Informatics* **55**, 101892 (2023)
9. La Rocca, G.: Knowledge based engineering: Between ai and cad. review of a language based technology to support engineering design. *Advanced engineering informatics* **26**(2), 159–179 (2012)
10. Li, Z.: Leveraging computational algorithms for effective explicit and tacit knowledge capture: A hybrid approach combining expert interviews, machine learning, and data mining techniques. *Applied and Computational Engineering* **114** (2024)
11. Mas, F., Arista, R., Skrzek, M., Oliva, M., Morales-Palma, D., Szejka, A.L.: A framework for ontology-based engineering systems: Advances and open questions about knowledge capture and use in aerospace manufacturing. In: Dassisti, M., Madani, K., Panetto, H. (eds.) *Innovative Intelligent Industrial Production and Logistics*. pp. 166–174. Springer Nature Switzerland, Cham (2025)
12. Mas, F., Racero, J., Oliva, M., Morales-Palma, D.: A preliminary methodological approach to models for manufacturing (mfm). In: *Product Lifecycle Management to Support Industry 4.0: 15th IFIP WG 5.1 International Conference, PLM 2018, Turin, Italy, 2018, Proceedings 15*. pp. 273–283. Springer (2018)
13. Muia, T., Salam, A., Bhuiyan, N.: A comparative study to estimate costs at bombardier aerospace using regression analysis. In: *2009 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE (2009)
14. Oksuz Gurdal, B., Testik, O.M.: A framework for product life cycle management based digital twin implementation in the aerospace industry. *Applied Stochastic Models in Business and Industry* **41**(2), e70001 (2025)
15. Schreiber, G.: *Knowledge engineering and management: the CommonKADS methodology*. MIT press (2000)
16. Stokes, M., et al.: *Managing engineering knowledge: MOKA: methodology for knowledge based engineering applications*, vol. 3. Professional Engineering Publishing London (2001)
17. Tang, Z., Pinon-Fischer, O.J., Mavris, D.N.: Identification of key factors in integrating aircraft and the associated supply chains during early design phases. In: *14th AIAA Aviation Technology, Integration, and Operations Conference* (2014)
18. Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., Sui, F.: Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology* **94**, 3563–3576 (2018)
19. Verhagen, W.J.: *Concurrent Engineering in the 21st Century: Foundations, Developments and Challenges*. Springer International Publishing (2015)
20. Xie, S., Tu, P., Aitchison, D., Dunlop, R., Zhou, Z.: A www-based integrated product development platform for sheet metal parts intelligent concurrent design and manufacturing. *International journal of production research* **39**(17) (2001)
21. Zhang, Q., Zheng, S., Yu, C., Wang, Q., Ke, Y.: Digital thread-based modeling of digital twin framework for the aircraft assembly system. *Journal of Manufacturing Systems* **65**, 406–420 (2022). <https://doi.org/10.1016/j.jmsy.2022.10.004>