Digital maturity models: comparing manual and semi-automatic similarity assessment frameworks

Bruno Cognet

Systems Engineering Department,
ETS, Montreal, Quebec, Canada
Emails: bruno.cognet.1@ens.etsmtl.ca

Jean-Philippe Pernot*

LISPEN, Arts et Métiers Institute of Technology,
HESAM Université, Aix-en-Provence, France
Email: jean-philippe.pernot@ensam.eu
*Corresponding author

Louis Rivest

Systems Engineering Department,
ETS, Montreal, Quebec, Canada
Email: louis.rivest@etsmtl.ca

Chritophe Danjou

Département de Mathématiques et Génie Industriel,
Polytechnique Montréal,
Montreal, Quebec, Canada
Email: christophe.danjou@polymtl.ca

Abstract: The fourth industrial revolution is forcing companies to define their digital strategy, making it imperative that they assess their digital maturity as a basis for improvements. As a result, a variety of maturity models have emerged. However, it can be difficult to identify which one is most appropriate. This paper introduces a new methodology to compare a manual and a semi-automatic framework for assessing the similarity of digital maturity models. It allows identifying the most adequate framework for comparing maturity models. Both frameworks have been designed to identify correspondences between KPIs. The analysis of the matches and the obtained results are then used to tune the semi-automatic framework. The proposed comparison methodology has been validated using two digital maturity models and shows that the semi-automatic framework provides good results in a very efficient manner. Several insights have been derived and will help to develop a new maturity model.

Keywords: Industry 4.0; smart manufacturing; digitalisation; maturity models; comparison framework; semi-automatic comparison; similarity assessment.

Reference to this paper should be made as follows: Cognet, B., Pernot, J-P., Rivest, L. and Danjou, C. (2021) 'Digital maturity models: comparing manual and semi-automatic similarity assessment frameworks', *Int. J. Product Lifecycle Management*, Vol. 13, No. 4, pp.291–316.

Biographical notes: Bruno Cognet received his Engineering degree in Mechanical, specialising in design and development from the Institut National des Sciences Appliquées (Lyon, France), in 2020. At the same time, he completed a research Master's degree (MSc A) from 2018 to 2020 at the Ecole de Technologie Supérieure (Montreal, Canada). This research led him to focus on the digital transformation of companies and their evaluation through maturity indicators.

Jean-Philippe Pernot is a Full Professor of Mechanical Engineering at Arts et Métiers Institute of Technology, France. For more than 15 years, he has acquired a solid experience in coordinating, managing and contributing to national and international research projects. He is currently a member of regional and national boards, and he is an expert for the European Commission for the evaluation of H2020 projects. His research focuses on computer-aided design, geometric modeling, reverse engineering and machine learning to support the digital transformation in the scope of the Industry 4.0.

Louis Rivest became a Professor at the Ecole de Technologie Superieure in Montreal, Canada, after spending a few years in the aerospace industry. He obtained his PhD from the Ecole Polytechnique de Montreal, in 1993, and Bachelor in Mechanical Engineering, in 1988. His research centres on the models, methods, tools and processes supporting product development. His teaching and research activities thus relate to CAD, PLM and Digital transformation.

Chritophe Danjou became a Professor at École Polytechnique de Montréal, Canada, in 2018, after completing a post-doctoral fellowship at École de Technologie Supérieure in Montreal. Previously he received his PhD in Advanced Mechanics in 2015, and an Engineering degree in Mechanical Systems in 2012 from the Université de Technologie de Compiègne, France. His research and teaching focus on the implementation and adoption of Industry 4.0 in factories and enterprises.

This paper is a revised and expanded version of a paper entitled 'Towards a Novel Comparison Framework of Digital Maturity Assessment Models' presented at PLM 2019 Conference, Moscow, Russia, 8–12 July 2019.

1 Introduction

Manufacturing companies feel increasing pressure to adopt the Industry 4.0 paradigm to evolve and remain competitive on the world market (Taisch et al., 2018; Oztemel and Gursev, 2020). To support small and medium-sized enterprises (SMEs) with their digital transformation, seen as a pillar of Industry 4.0, several maturity models have been developed for evaluating the level of digital maturity of individual companies (Mittal et al., 2018; Scremin et al., 2018). The results of these evaluations are then used to design and set up digital transformation plans. However, there are many different approaches to assess the digital maturity of a company, and it can be difficult to identify the most appropriate option as they do not always focus on the same set of criteria.

The current literature shows that each maturity model has its specific objectives, benefits and challenges. Therefore, a suitable solution may lie in the partial exploitation of some of their core advantages, making it imperative to clearly identify where the existing maturity models match and where they differ (Cognet et al., 2019). It must be noted that most digital maturity models do indeed share some common characteristics and goals. For example, most of them use a set of questions to evaluate certain criteria that are grouped in dimensions, and possibly sub-dimensions. However, once formalised, criteria are not always evaluated in the same way. In some cases, an evaluation is performed by asking the user to self-assess the levels, or by using some black box mechanisms to compute a score from multiple user-specified answers. Regarding the questionnaires, some approaches refer to self-assessment, whereas others are focused on guided assessment without an automatic benchmarking system.

Maturity models are usually organised in sets of questions and answers. Some models identify "items" (Schumacher et al., 2016), while others use "factors" (Samaranayake et al., 2017) or "variables" (Schwer et al., 2018) that define digital maturity. When evaluating the digital maturity of a company, the so-called criteria can be considered as KPIs for monitoring the status of a company; this is the terminology that is used in this document. Hence, comparing two maturity models is conducted so as to identify the level of correspondence between two sets of KPIs.

The final objective of our work is to propose an approach to systematically compare existing digital maturity models. The comparison is based on the identification of matches between the digital maturity indicators of various maturity models. In this context, the objective of this paper is to propose a methodology to evaluate two frameworks for comparing digital maturity models and their indicators (considered KPIs). One framework is called 'manual', and the other, 'semi-automatic'. The two frameworks aim to assess the distance between two maturity models. The methodology aims to adjust and validate the semi-automatic comparison framework to efficiently identify potential matches between the maturity indicators of several models. Both the manual and the semi-automatic frameworks have the same goal: to help experts to match the KPIs. This paper aims to demonstrate how two distinct frameworks for comparing two digital maturity models can be compared and how the semi-automatic one can be fine-tuned in order to produce similar results to the expert-driven manual one. This paper's contribution is to explain how the overall methodology was developed. The outcome of applying the overall maturity assessment is intended to be useful to organisations that need to determine which maturity model to use.

The specific contributions of this paper are threefold:

1. A new methodology for comparing two similarity assessment frameworks, one manual and the other semi-automatic, that are themselves used to compare two maturity models. The comparison is performed at the level of keywords, which are associated with KPIs related to questions/answers that have been taken from questionnaires and categorised by dimension and sub-dimension. At this abstraction level, the work is conducted in a common space (the space of the keywords) to obtain comparable results. It shows that moving from the frameworks to more tangible data, i.e., the matching matrices in our case, can help to more efficiently compare them while introducing ad-hoc criteria.

2 A list of the KPIs reverse engineered from two digital maturity models. These are of direct interest to companies, which can incorporate them in their own KPI grids and use them to perform their own digital maturity assessments, and to consulting firms wishing to develop their own maturity models.

3 An analysis of the similarities (matches) found between the two frameworks to support the fine-tuning of the semi-automatic framework. This latter framework is much more efficient than the manual one and, once fine-tuned, can be reused to work directly on the comparison of a large number of maturity models (not included in this paper).

The comparison of the two frameworks is composed of several steps: the reverse engineering of the KPIs from existing models, the definition of keywords, the manual/semi-automatic identification of the matches, analysis of the matches found in the two frameworks, and the fine-tuning of the semi-automatic framework. The two similarity assessment frameworks, and the methodology used to compare them, have been tested with two well-known digital maturity models, IMPULS (Lichtblau et al., 2015) and PwC - PricewaterhouseCoopers – (Geissbauer et al., 2016).

This paper is organised as follows. After a summary of the background literature in Section 2, an overview of the whole comparison methodology is presented in Section 3. The manual and semi-automatic similarity assessment frameworks are introduced in Sections 4 and 5, respectively. The methodology used to compare the two frameworks and enhance the semi-automatic one is defined in Section 6. Section 7 discusses the results obtained following the proposed methodology validated on two maturity models, IMPULS and PwC. The last section concludes this paper and discusses the next steps.

2 Background literature

In general, the term "maturity" refers to a "state of being complete, perfect, or ready" (Simpson and Weiner, 1989). Maturity models are tools used to identify the best practices for the transformation of an organisation (Schumacher et al., 2016). They provide a structured approach to initiate and accompany short-term operational projects, as well as medium-term tactical changes and long-term strategic change (Felch et al., 2019).

Currently, there is a variety of digital maturity models to support companies in their digitalisation activities. Their common goal is to assess the digital maturity of an organisation and to provide an indication of the actions required to increase the maturity level. Existing studies have reviewed most of the common maturity models, in general (Wendler, 2012), as well as digital maturity models in particular (Rossmann, 2018).

According to the above and other related studies, the most common features for maturity models include their incorporation of maturity dimensions (usually 3–7 dimensions that are descriptive of the maturity to be assessed, and which are often divided into more detailed maturity criteria, descriptive of the related maturity dimensions), maturity levels and related maturity descriptions (Akdil et al., 2018). Maturity dimensions, in general, can be divided into three broader categories: maturity of people/culture (e.g., skills, capabilities), processes/structures, and objects/technology (such as ICT tools). A recent literature review-based conceptual paper related to the broad concept of digital maturity (Rossmann, 2018) demonstrates that in current digital maturity studies, digital maturity has included aspects that can be divided into eight capability dimensions (i.e., broad digitalisation-related maturity categories): strategy, leadership, business model, operating model, people, culture, governance, and technology. It should be understood that the digital transformation of organisations concerns all their activities, and that digital technologies can be implemented to support this transformation (Salkin et al., 2018).

Several papers in the literature compare existing maturity models. Some authors perform an overall comparison of digital maturity models according to the following criteria: maturity level, dimensions, and scope of the study (Akdil et al., 2018). Schwer et al. (2018), on the other hand, use a 7-step comparison method that makes it possible to identify 147 'variables of digitisation' and classify them according to 6 dimensions. In both cases, the qualitative approaches make it possible to identify the scopes of the selected maturity models.

Earlier research (Westerman et al., 2014) presumed that the development of a specific set of the above types of digital capabilities leads to higher digital maturity, and moreover, that a higher degree of digital maturity can lead to superior corporate performance. Maturity models also provide the basis for guiding a digital transformation (Schumacher et al., 2016), but the development of a roadmap is necessary to ensure the actions will be performed in the right order (De Carolis et al., 2018). However, such maturity models vary in terms of their structure, scope and industry focus (Schwer et al., 2018). Furthermore, while Rossman's (2018) recent study has been among the first to present a more unified conceptualisations of the topic, the current research still reflects conceptual unclarity and fragmented views about the concept and the measurement frameworks for digital maturity. In general, the development of a maturity model, digital or not, requires a literature review to identify the existing models, along with a comparison of these models, facilitated by a group of maturity and digital experts (Becker et al., 2009).

Currently, there is no clear definition of what digital transformation really is. Many points of view can be observed in the literature (Moeuf et al., 2017; Oztemel and Gursev, 2020; Pereira and Romero, 2017). This diversity is also reflected in digital maturity assessment models (Gökalp et al., 2017; Rossmann, 2018). An analysis of the digital maturity assessment models available in the literature shows that they do not all assess the same aspects (Gökalp et al., 2017; Mittal et al., 2018). Instead, they only focus on an aspect of the modifications induced by the desire for digital transformation. Their evaluation may focus on the human aspect of the fourth industrial revolution, while others have a more technical dimension focused on the evaluation of installed technologies. However, the aggregation of all these digital maturity models should provide a more complete assessment of a company's digital maturity.

In addition, each model proposes its own system of dividing into dimensions and sub-dimensions (Gökalp et al., 2017; Mittal et al., 2018). The issues of the different models partially overlap. Thus, in order to establish a complete list of KPIs, it is important to identify the similarities between the different questions rather than trying keep individual statements from the existing maturity models. The development of a maturity model requires a methodology to be followed. Comprehensive literature reviews show that this can be a lengthy process and suggest that the choices made at each stage should be justified (Becker et al., 2009; Wendler, 2012). A review of the literature makes it possible to identify existing maturity models, and a comparison of the models selected can then serve as a basis for the new model.

Thus, assessing the similarities of different maturity models is not straightforward, as it requires a deep understanding of and expertise in the widely varying domains and dimensions covered by the available multiple models. Cognet et al. (2019) developed a manual similarity assessment framework that can manually identify the level of match between the KPIs extracted from two digital maturity models used as the inputs in their framework. The matches are identified with the help of several experts. This framework is time consuming, as the matches must be evaluated one by one.

This paper intends to overcome these limits and extend this work, using a more efficient semi-automatic framework that performs similarity assessment by comparing keywords associated manually to the KPIs. The literature shows that the assessment of similarity between two sentences can be performed and automated using a word analysis (Liu and Wang, 2013). The approach developed and detailed in this paper is a simplified version of what is observed in the literature. These methods also make it possible to quantify the similarity between two sentences and to limit subjective bias in the sentence analysis. Comparing a known number of words to a database that enables the classification of these same words according to their meaning or the desired grouping can give a numerical indicator for maturity indicator comparison.

The first results obtained with the semi-automatic process were slightly different than those obtained following the manual process, thus requiring a further fine-tuning step. This paper introduces a new methodology that compares the manual and semi-automatic frameworks and improves the semi-automatic framework accordingly.

The large variety of available digital maturity models justifies the need to develop a framework to assess their similarities in an efficient and accurate manner. The literature review shows that existing comparison approaches are qualitative. Our semi-automatic framework, on the other hand, is quantitative. This section introduced the background literature of maturity models, and the next section focuses on presenting the overall methodology utilised to compare two digital maturity model similarity assessment frameworks.

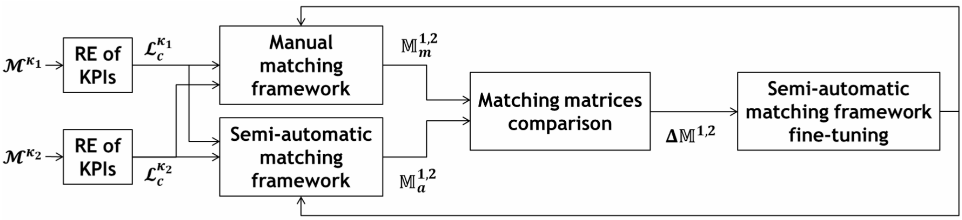
3 Overall comparison methodology

The literature review revealed that there is a large diversity between digital maturity models. The manual and semi-automatic similarity assessment frameworks were developed to compare maturity models for this study. In this paper, a methodology is developed to compare those two frameworks. This methodology is composed of several steps, illustrated in Figure 1 and detailed in the next sections:

- reverse engineering (RE) of the lists of KPIs (or criteria) $L_c^{K_1}$ and $L_c^{K_2}$ of the two maturity models M^{K_1} and M^{K_2}, respectively, to be compared
- similarity assessment using both the manual and the semi-automatic frameworks. The output of the manual framework is a matching matrix $M_m^{1,2}$ characterising how much the two maturity models match. The output of the semi-automatic framework is the matching matrix $M_a^{1,2}$
- computation of the deviation matrix $\Delta M^{1,2}$ characterising the deviations between the results obtained by two similarity assessment frameworks

- fine-tuning of the semi-automatic framework.

Figure 1 Overall methodology for the comparison of manual and semi-automatic similarity assessment frameworks



The reverse engineering step is required for all maturity models, even if they provide their KPIs and information about models' KPIs. This step is designed to extract and formalise the list of KPIs that best characterises the criteria adopted by a given maturity model in order to assess the maturity levels, accomplished by using all the available resources describing the considered maturity model (e.g., online self-assessment tools, questionnaires, benchmarking reports, articles). How this list is determined has been detailed in Cognet et al. (2019). In short, the output list of KPIs results from consensual exchanges involving a pool of experts in the domain. During the evaluation process, experts are requested to focus on the explicitly available information rather than on more implicit data whose interpretation could be questionable. Following this process, the risk of bias due to reinterpretations is reduced, but cannot be fully disregarded.

The reverse engineering is a pre-processing step required to obtain the KPIs used as the inputs of the manual and semi-automatic similarity assessment frameworks. This is a prerequisite common to both frameworks and detailed in Cognet et al. (2019). Maturity models use dimensions and sub-dimensions to group the questions/answers by broad categories. The reverse engineering of KPIs homogenises the maturity models' questions/answers. For instance, IMPULS has broad questions for which multiple answers are expected, whereas PwC has more precise questions and only two sliders that need to be positioned to answer. Therefore, the KPIs are also grouped along these dimensions and sub-dimensions, since they are directly derived from the questions/answers. The other steps of the proposed methodology are detailed in sections 4 and 5.

4 Manual similarity assessment framework

The manual framework evaluates the level of matching between the KPIs of the two digital maturity models whose similarity is being assessed (Cognet et al., 2019).

A maturity model M^K (with $K \in$ {IMPULS,PwC,ADN,…}) makes use of N_c^K criteria denoted as C_i^K (with $i \in [1..N_c^K]$) and grouped in N_d^K dimensions denoted as D_j^K (with $j \in [1..N_d^K]$). The j^{th} dimension D_j^K contains $N_{c,j}^K$ criteria, which start at index s_j^K and end at index e_j^K. The criteria can be gathered together in a single list

$L_c^K = \{C_i^K, i \in [1..N_c^K]\}$, or in separate lists $L_{d,j}^K = \{C_i^K, i \in [s_j^K..e_j^K]\}$ associated to their respective dimensions and with $j \in [1..N_d^K]$. The following rules apply:

$$N_c^K = \sum_{j=1}^{N_d^K} N_{c,j}^K \tag{1}$$

$$s_1^K = 1, and \forall j \in \left[2..N_d^K\right], s_j^K = s_{j-1}^K + N_{c,j-1}^K \tag{2}$$

$$\forall j \in [1..N_d^K], e_j^K = \sum_{k=1}^{j} N_{c,k}^K \tag{3}$$

The criteria C_i^K (with $i \in [1..N_c^K]$) of a maturity model M^K are the KPIs used to evaluate the maturity.

The similarity assessment of the maturity models is performed at the level of the KPIs. To characterise how much the KPIs match, three levels are introduced: Strong match, Partial match, and No match. Two KPIs are considered a Strong match if the experts involved in this process identify a strong similarity between the two. Conversely, if the two KPIs do not share any similar features, a No match is considered. In between these two situations, when the KPIs share some similar features, but also have dissimilarities, a Partial match is assigned. Such a three-level matching analysis presents a good trade-off between an under-segmentation, which would lead to a coarse analysis, and an over-segmentation that would complexify the comparison making it cumbersome and unworkable.

Therefore, the matching function f_m, with m referring to the manual framework, evaluates the matching level of two KPIs, $C_{i_1}^{K_1}$ and $C_{i_2}^{K_2}$, of two maturity models, M^{K_1} and M^{K_2}, respectively. It is defined as follows:

$$f_m\left(C_{i_1}^{K_1}, C_{i_2}^{K_2}\right) = \begin{cases} \text{Strong} & \text{if } C_{i_1}^{K_1} \text{ and } C_{i_2}^{K_2} \text{ strongly match} \\ \text{Partial} & \text{if } C_{i_1}^{K_1} \text{ and } C_{i_2}^{K_2} \text{ partially match} \\ \text{No} & \text{otherwise} \end{cases} \tag{4}$$

This function is then called to fill in the matching matrix $M_m^{1,2}$ containing the $\left(N_c^{K_1} \times N_c^{K_2}\right)$ values returned when applied with the KPIs of M^{K_1} and M^{K_2}. Clearly, due to this procedure, it is possible to observe that the matching function is symmetric, i.e., it returns the same matching level no matter the order of the arguments.

Here again, the assessment of the matching levels results from consensual exchanges involving a pool of experts. In a first individual phase, experts are asked to suggest a matching level for each KPI pair. Then, during a consensus phase, experts exchange information about their classifications and discuss the matching levels for which there are discrepancies. When the discussion fails to reach an adequate consensus, a simple majority rule can be used, giving greater weight to the choice(s) of the most experience experts. Ultimately, an additional expert may be considered to solve any residual conflicts. Thus, the matching process results strongly rely on the exchanges between the experts, and consequently on their knowledge and experience in the domain. Clearly, similar results could not really be obtained using simple text-based similarity analysis tools.

5 Semi-automatic similarity assessment framework

During the manual comparison of the IMPULS and PwC models, the experts compared the KPIs directly with one another. After this first comparison, two observations were noted:

1 working at the KPI level is time-consuming

2 the experts were concentrating on the important words of the KPIs.

Working at the KPI level is time-consuming because the experts must evaluate the matching level of each pair of KPIs, one-by-one. For instance, considering two maturity models based on an average of $N_c^{K_1} = N_c^{K_2} =$ 30 KPIs each, the number of evaluations increases rapidly to 900. After testing with the manual framework, only 49 matches were identified out of the 825 KPI pairs formed by the two selected models.

The manual comparison also revealed that the experts were focusing on the important words of the KPIs. This observation led to the creation of the semi-automatic framework. The switch from KPI comparison (manual comparison) to keyword comparison (semi-automatic comparison) is also motivated by the number of KPIs to be compared (in this paper, 20×33 and then about 400×400).

The keywords work at a different level of abstraction and overcome the formulation. In other words, two KPIs can express the same idea but be written in different ways. The semi-automatic framework is composed of three steps: the definition of keywords for each KPI, the identification of matches between keywords, and an automatic scoring to evaluate the matches between KPIs. Thus, extracting keywords from each KPI and creating a matrix to classify the keywords then makes it possible to identify similar KPIs without focusing on the formulation.

In this framework, the similarity assessment is not based on the KPIs but on the keywords, which are defined to characterise the KPIs. Thus, for a maturity model M^K, experts define $N_q^{K,i}$ keywords $Q_h^{K,i}$ (with $h \in [1..N_q^{K,i}]$) in $L_{q,i}^K$ which best characterise each KPI C_i^K (with $i \in [1..N_c^K]$). Each maturity model leads to a list of keywords $L_q^K = \{L_{q,i}^K, i \in [1..N_c^K]\}$. These lists are then grouped into an overall list of keywords L_q.

To perform an automatic correspondence between KPIs, the keywords of L_q must be classified in a matching matrix Q which characterises the matches at the level of the keywords. Thus, each row of this matrix gathers keywords with a similar meaning. Thus, each row contains a distinct concept which could help in the development and structuring of a new maturity model in a later step.

This last step of the semi-automatic similarity assessment framework aims at automatically computing the scores characterising the matching levels of all the KPI pairs between two maturity models. This step takes as its input the lists of keywords of two maturity models M^{K_1} and M^{K_2}, as well as the matrix Q. The output is a matching matrix $M_a^{1,2}$ containing $(N_c^{K_1} \times N_c^{K_2})$ real values ranging from 0 (no match) to 1 (perfect match). This computation relies on a low-level counting function f_a (where a stands for automatic) that can verify if a keyword Q matches at least one of the elements of a list of keywords L according to the matching matrix Q. This function is defined as follows:

$$f_a(Q,L) = \begin{cases} 1 & \text{if } \exists\, Q_h \in L/Q \text{ and } Q_h \text{ are on the same row in } Q \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

Thus, for each KPI pair $\left(C_i^{K_1}, C_j^{K_2}\right) \in \left(L_c^{K_1}, L_c^{K_2}\right)$, with $(i,j) \in \left[1..N_c^{K_1}\right] \times \left[1..N_c^{K_2}\right]$, the two related lists of keywords $L_{q,i}^{K_1}$ and $L_{q,j}^{K_2}$ are automatically analysed so as to compute their level of match and fill the term (i,j) of the matrix $M_a^{1,2}$ as follows:

$$\left(M_a^{1,2}\right)_{i,j} = \frac{1}{N_q^{K_1,i} + N_q^{K_2,j}} \left[\sum_{h=1}^{N_q^{K_1,i}} f_a\left(Q_h^{K_1,i}, L_{q,j}^{K_2}\right) + \sum_{k=1}^{N_q^{K_2,j}} f_a\left(Q_h^{K_2,j}, L_{q,i}^{K_1}\right) \right] \tag{6}$$

The terms of this matrix are always between 0 and 1. Compared to the manual similarity assessment framework, the semi-automatic one helps to better characterise the so-called Partial matches. However, in the next section, the comparison of the two matrices $M_m^{1,2}$ and $M_a^{1,2}$ will not consider those real values, and all values in between 0 (no match) and 1 (perfect match) will be associated with a Partial match.

It is worth emphasising that, at its core, the manual framework acts at the KPI level rather than at the question/answer or dimension/sub-dimension level, whereas the semi-automatic framework acts at the keyword level to overcome the formulation of the KPIs. Focusing on KPIs and keywords reduces the subjectivity when comparing different models, since the keywords characterise the essence of the KPIs. This is the strategy followed in this paper. The alternative, i.e., working at the level of the dimensions would make the comparison significantly more difficult. Indeed, the dimensions are very abstract and are highly dependent on the design of each maturity model. Each model groups the questions into dimensions (and sometimes sub-dimensions) depending on its own vision of maturity assessment. Using keywords to compare KPIs is simpler than comparing KPIs directly, as the keywords convey the essence of the KPIs. The final list of keywords is the lowest common denominator and establishes the 'dictionary' that defines the common space for comparing maturity models.

6 Fine-tuning of the semi-automatic framework

This section details the two last steps of the overall comparison methodology, i.e., the comparison of the matching matrices $M_m^{1,2}$ and $M_a^{1,2}$ and its use to fine-tune the semi-automatic similarity assessment framework, as described in Figure 1.

The matrix $M_m^{1,2}$ is the result of quite a long analysis (more than two days) by several experts involving complex discussions and consensus phases. As a result, it is assumed in this work to be the reference matrix that embeds the experts' knowledge and expertise, to be reproduced by the semi-automatic matrix during the fine-tuning process.

6.1 Comparison of the matching matrices

As briefly introduced at the end of section 4, the two matrices to be compared do not own comparable values directly. The elements of $M_m^{1,2}$ can be of three types (Strong, Partial or No match), whereas $M_a^{1,2}$ is filled with real values in between 0 and 1. Thus, to be able to compare the two matrices, and to observe the possible differences, a deviation matrix $\Delta M^{1,2}$ is defined as follows:

$$\left(\Delta M^{1,2}\right)_{i,j}=\begin{cases}0 & \text{if}\left(\left(M_m^{1,2}\right)_{i,j}=\text{No AND}\left(M_a^{1,2}\right)_{i,j}=0.0\right)\\-1 & \text{if}\left(\left(M_m^{1,2}\right)_{i,j}\neq\text{No AND}\left(M_a^{1,2}\right)_{i,j}=0.0\right)\\-2 & \text{if}\left(\left(M_m^{1,2}\right)_{i,j}=\text{No AND}\left(M_a^{1,2}\right)_{i,j}>0.0\right)\\0 & \text{if}\left(\left(M_m^{1,2}\right)_{i,j}\neq\text{No AND}\left(M_a^{1,2}\right)_{i,j}>0.0\right)\end{cases} \quad (7)$$

In other words, when an element of $\Delta M^{1,2}$ is equal to 0, it means that the related elements in $M_m^{1,2}$ and $M_a^{1,2}$ match, and consequently that the manual and semi-automatic frameworks result in the same conclusion at the level of the underlying KPIs. When an element of $\Delta M^{1,2}$ is equal to –2, it means that the two frameworks differ in their conclusion, and that the semi-automatic framework identifies a false-positive match, i.e., a match not found by the experts involved in the manual framework. Thus, in this case the semi-automatic framework over-estimates the match level of the underlying KPIs. Finally, when an element of $\Delta M^{1,2}$ is equal to –1, it means that the semi-automatic framework does not identify a match that has been found by the experts following the manual framework. In this case, the value returned by the semi-automatic framework is a false-negative, which is not desirable as it means that some matches are not properly captured.

It is important to understand that the deviation matrix is used only to fine-tune the semi-automatic framework and not used anymore afterwards.

6.2 Three types of change for fine-tuning of the semi-automatic framework

Given that the ultimate objective is to replace the manual similarity assessment framework by a semi-automatic one, it is important to fine-tune the latter so that it identifies at least the matches found manually by the experts. Following such a fine-tuning process, the false-positive matches would remain, while the false-negative ones would be removed. In other words, at the end of this process, the deviation matrix $\Delta M^{1,2}$ should only contain the digits 0 or –2.

During the fine-tuning process, the KPI pairs whose match level is equal to –1 in the deviation matrix $\Delta M^{1,2}$ are identified and treated one by one. Three types of changes can therefore be considered and are listed below in order of preference:

- First, the lists of keywords used to characterise the KPI pairs whose match level can be considered as a false-negative can be changed. Keywords can either be removed if found to be irrelevant, or added if those already used do not allow for a complete characterisation of the KPIs. New keywords can either be selected from the complete list L_q, or added to it if none of the available keywords conveys the related meaning;
- Second, the matching matrix Q can be changed by either reconsidering the position of some keywords on the various rows, or by splitting or merging some rows; and

- Lastly, the matrix $M_m^{1,2}$, obtained by the experts at the end of the manual framework, may be changed and some of its values reconsidered during a new consensus discussion. This last solution is to be followed only if the first two changes did not produce a convergence.

Of course, the two first types of changes may affect other matches that were not initially considered as false-negative, i.e., the new value of –1 could appear in the deviation matrix. Thus, as soon as a list of keywords or the matching matrix is modified, all the values of matrices that may have changed must be revaluated. At the end of this process, the semi-automatic framework is fine-tuned, and it produces results similar to those of the manual framework, with no false-negatives.

7 Results and discussion

Even though there is a wide variety of maturity models available in literature, the proposed comparison frameworks have been tested and validated with two of the earliest maturity models: IMPULS and PwC. Those two maturity models were selected because they are both digital maturity self-assessment tools easily available online, and because their number of questions and number of dimensions are quite similar and reasonably low, making it simpler to validate the proposed comparison methodology. The resulting fine-tuned semi-automatic similarity assessment framework can be used on any other maturity model.

7.1 IMPULS and PwC maturity models

Considering the formalisation introduced above, the two maturity models can be characterised by the values presented in Table 1.

Table 1 Numerical characteristics of the two maturity models being compared

K	K_1=IMPULS	K_2=PwC
N_d^K	6	6
N_c^K	25	33
$\{N_{c,j}^K, j \in [1..N_d^K]\}$	{6,5,7,1,4,2}	{6,6,5,6,6,4}

This table shows that both maturity models have the same number of dimensions, whose details are given in Table 2. Clearly, this is a unique case, as there is no obvious reason to have $N_d^{IMPULS} = N_d^{PwC}$. Table 2 indicates that there is no correspondence between the dimensions of the two selected models, which is why the comparison is based on smaller components, the KPIs. The table also reveals that each dimension is not evaluated with the same number of KPIs. For instance, the first dimension of IMPULS contains six KPIs, whereas its last dimension has only two. This might be a good indicator of the importance of a given dimension in the overall maturity model.

Table 2 Dimensions of the two maturity models being compared

j_1	Dimension $D_{j_1}^{IMPULS}$	j_2	Dimension $D_{j_2}^{PwC}$
1	Strategy and organisation	1	Business models, product and service portfolio
2	Smart factory	2	Market and customer access
3	Smart operations	3	Value chains and processes
4	Smart products	4	IT architecture
5	Data-driven services	5	Compliance, legal, risk, security and tax
6	Employees	6	Organisation and culture

Table 3 KPIs reverse engineered from the IMPULS maturity model

i_1	Reverse engineered KPI $C_{i_1}^{IMPULS}$
1	Implementation status of Industry 4.0 strategy
2	Operationalisation and review of Industry 4.0 strategy through a system of indicators
3	IT and digital technologies used in the company
4	Level of financial investment in the implementation of Industry 4.0 in various company sectors in the next 5 years
5	Level of financial investment in the implementation of Industry 4.0 in various company sectors in the past 2 years
6	Company sectors of systematic innovation management
7	Availability of communication, control and interoperability functionalities of the equipment infrastructure
8	Upgradability of communication, control and interoperability functionalities of the equipment infrastructure
9	Level of digital modelling of the factory through the collection, storage and processing of data during production
10	Digital tools used in the company's sectors
11	Interface of the digital tools to the leading system
12	Internal cross-sectors' level of information sharing
13	External information sharing between sectors, and with customers and/or suppliers
14	Level of deployment of workpiece self-guiding capacities through production
15	Level of deployment of autonomous real-time response capacities to changes in production conditions
16	Organisation/distribution of IT expertise across the company's departments
17	Level of implementation of IT security solutions for internal/external data storage and data communication
18	Use of cloud-based software, and of cloud services for data analysis and storage
19	Availability of add-on functionalities (e.g., memorisation, localisation, self-reporting) in the company's products in order to make them smart
20	Availability of, and customer integration with data-driven services that use data gathered during the production and usage phases
21	Capacity to collect and analyse data from the usage phase
22	Share of revenues derived from data-driven services
23	External share of data collected all along the product lifecycle
24	Levels of the employees' skills with respect to the future requirements (e.g., IT infrastructure, automation technology, data analytics) of Industry 4.0
25	Level of the company's effort to acquire new skills and train its employees

Table 4 KPIs reverse engineered from the PwC maturity model

i_2	*Reverse engineered KPI* $C_{i_2}^{PwC}$
1	Contribution of digital features, products and services to the overall value creation of the organisation's portfolio
2	Degree of digitisation of the organisation's products and/or services
3	Possibilities for customer customisation of products
4	Degree of digitisation of the product lifecycle phases
5	Importance of data usage and analysis for the organisation's business model
6	Intensity of the collaboration with external partners and clients for the development of products and services
7	Level of integration of sales channels used to sell the organisation's products
8	Level of integration of communication channels for customer interaction
9	Availability of digital tools and technologies to support the organisation's sales force
10	Degree of dynamic customisation of the prices based on customers' willingness to pay
11	Degree of customer data analysis to increase customer insight
12	Level of collaboration with partners regarding customers' access approach
13	Degree of digitisation of activities from product development to production
14	Ability to monitor production and to dynamically respond to changes in demand
15	Degree of integration of the end-to-end IT-enabled planning and steering process over the entire value chain
16	Degree of digitisation of the production equipment up to a virtual representation of the factory
17	Degree of digitisation of activities from customer's order to service
18	Degree of consideration of the digitisation and Industry 4.0 requirements in IT architecture
19	Level of use of a manufacturing execution system (MES) or of a similar system control the manufacturing process
20	Level of maturity of the IT and data architecture to gather, aggregate and interpret real-time manufacturing, product and client data
21	Importance of new technologies (social media, mobility, analytics and cloud computing) to enable business operations
22	Ability of the IT organisation to fulfil business requirements within the requested time, quality and cost
23	Level of IT integration with customers and partners
24	Degree of sophistication of the digital compliance policy
25	Levels of the organisation's IP protection and of the external IP consideration
26	Level of consideration of the digital product portfolio and production factory in the risk management
27	Level of management of the digital components of the organisation's value chain with respect to tax-related topics (IP location...)
28	Level of consideration of production in the organisation's IT security concept
29	Level of consideration of the service-partners or customers in the organisation's compliance and risk management
30	Ability to create value from data so as to optimise operations and foster new business models
31	Level of the organisation's capabilities and resources related to Industry 4.0
32	Level of involvement, support and expertise of the organisation's managers with regards to Industry 4.0
33	Level of collaboration of the organisation with external partners (e.g., academia, industry, suppliers, customers) on Industry 4.0 topics

The first step of the proposed methodology is designed to reverse engineer the KPIs of the maturity models to be compared (Figure 1). In Cognet et al. (2019), it was decided to keep the KPIs already formalised in the models' documentation, even though they were sometimes quite generic, without considering the underlying dimensions and corresponding questions. However, at the first attempt with the manual framework, the experts encountered problems in understanding some of the indicators. In a second attempt, the experts decided to reverse engineer the KPIs and to disregard those that were formalised too synthetically. Table 3 contains the new list L_c^{IMPULS} of the N_c^{IMPULS} =25 KPIs of IMPULS obtained through a consensus workshop involving four experts. Similarly, the list L_c^{PwC} of the N_c^{PwC} =33 KPIs of PwC have been reverse engineered and are shown in Table 4. Starting from the available online self-assessment tools of IMPULS and PwC, each question and possible answers have been carefully analysed and discussed to come out with a consensual formalisation of the KPIs. This step is not straightforward and required several in-depth discussions to achieve a consensus. The main difficulty was to avoid over-interpretation of the online questionnaire and to remain as objective and factual as possible.

7.2 Initial matching matrices and deviation matrix

Once the lists of the IMPULS and PwC KPIs have been obtained, the two similarity assessment frameworks can be run separately.

Following the manual framework, the matching matrix $M_m^{1,2}$ can be filled out while evaluating the matching levels between all the KPI pairs of the two maturity models. As mentioned in Cognet et al. (2019), the experts felt the need to develop self-understandable KPIs that directly embed the context within their formulation. This was done during the reverse engineering step, and the identified KPIs of Table 3 and 4 follow this rule. The obtained matching matrix $M_m^{1,2}$ is presented in Table 5. Green colours correspond to Strong matches between two KPIs, and yellow colours to Partial matches, whereas no colour indicates No match. For instance, five KPIs from PwC have a strong match with five different KPIs from IMPULS. It is also clear that KPI 13 from IMPULS partly matches four KPIs from PwC, and so forth. To better understand the strong matches, the matching matrix of IMPULS compared to itself would be a square matrix with only green cells on its diagonal and with 100% of the KPIs strongly matched.

The semi-automatic framework was utilised next, and the matching matrix $M_m^{1,2}$ completed and presented in Table 6. Experts used 133 keywords overall (i.e., card $\left(L_p\right)$ =133), with an average of 3 keywords per KPI. These keywords are sorted in the matching matrix Q, defined by 62 rows, which means that the adopted keywords cover 62 different concepts. This value is directly comparable to the number of KPIs in both maturity models (25 for IMPULS and 33 for PwC), which suggests that the KPIs do not overlap very much. In Table 6, each non-zero value is coloured orange, independently of whether it has a high or low matching level. As for $M_m^{1,2}$ this matrix clearly highlights the KPIs that have no correspondence in the other maturity model. For instance, KPI 9 of PwC can be considered as specific to PwC as it does not match any of the IMPULS KPIs.

Table 5 Initial matching matrix $M_m^{1,2}$ (see online version for colours)

KPIs PwC		KPIs IMPULS																								
		1						*2*					*3*							*4*	*5*				*6*	
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*
1	1																			P	P		S			
	2																			S	P					
	3																			P						
	4																				P	P		P		
	5									P				P								S		P		
	6																									P
2	7																									
	8																									
	9																									
	10																									
	11																				P	P		P		
	12									P																
3	13										P		P													
	14									P						S										
	15											P	P	P												
	16							P		S																
	17										P			P								P				
4	18							P																		
	19										P															
	20									P											P	P				
	21			P															P			P				
	22																									
	23													P								P				
5	24																									
	25						P																			
	26						P																			
	27																									
	28																	P								
	29																									
6	30									P												P		P		
	31																								P	P
	32																									
	33																									P

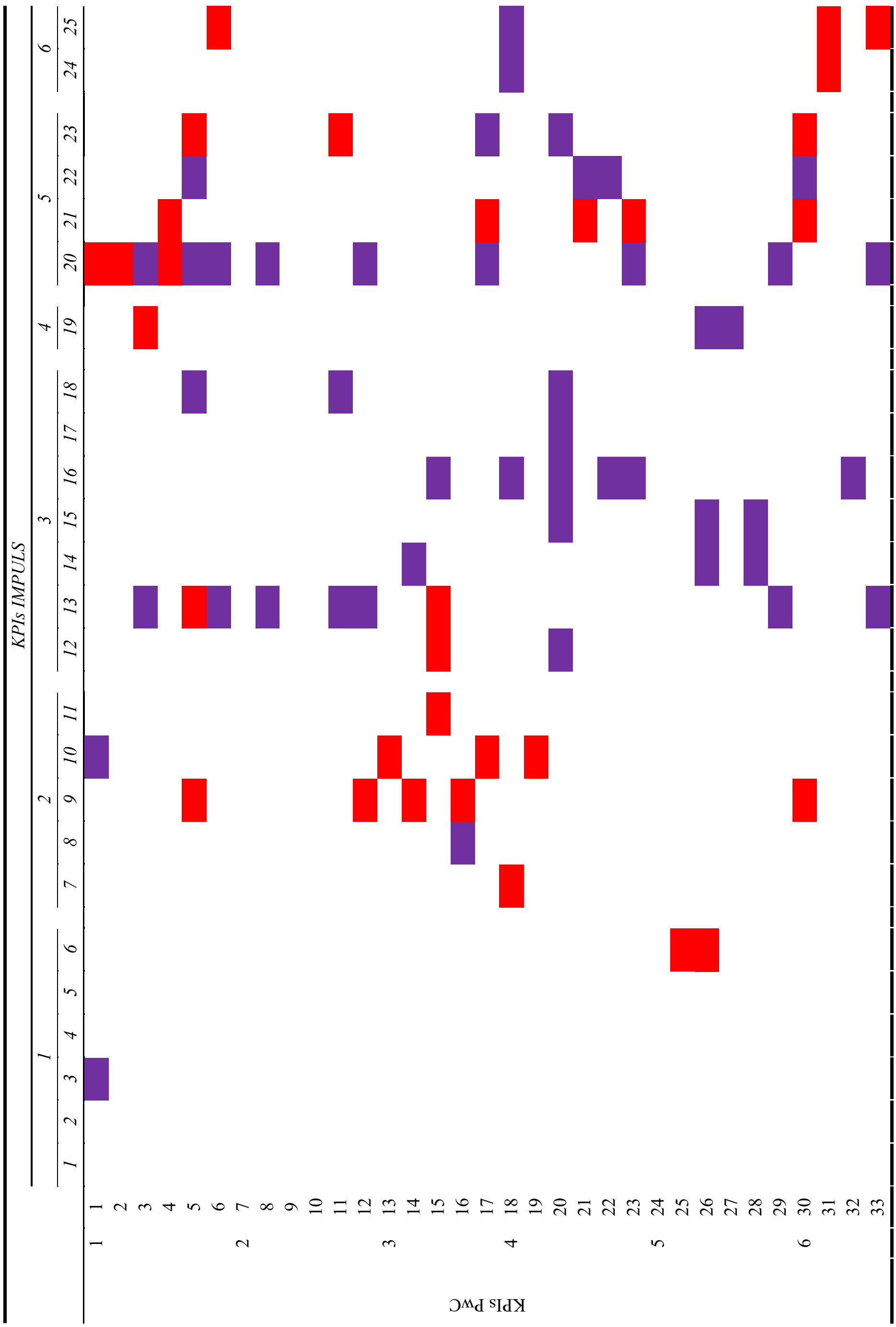
Notes: Strong match; Partial match.

Table 6 Initial matching matrix $M_a^{1,2}$ (see online version for colours)

		KPIs IMPULS																								
		1						2					3							4	5				6	
KPIs PwC		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1			■							■									■			■			
	2																			■						
	3													■							■					
	4																							■		
	5																		■		■	■	■			
	6													■							■					
2	7																									
	8													■							■					
	9																									
	10																									
	11													■					■		■	■				
	12													■							■					
3	13												■													
	14														■	■										
	15																■									
	16							■	■																	
	17													■							■			■		
4	18																■								■	■
	19																									
	20									■			■			■	■	■	■		■	■		■		
	21			■															■				■			
	22																■						■			
	23													■			■				■					
5	24																									
	25																									
	26														■	■				■						
	27																			■						
	28														■	■		■								
	29													■							■					
6	30																						■			
	31																									
	32																■									
	33													■							■					

Notes: ■ Automatic match.

Table 7 Initial deviation matrix $\Delta m^{1,2}$ (see online version for colours)



Notes: ■ Automatic false-positive match. ■ Automatic false-negative match.

Table 8 Final matching matrix $M_m^{1,2}$ (see online version for colours)

		KPIs IMPULS																								
		1						*2*					*3*							*4*	*5*				*6*	
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*
KPIs PwC	1 1																			S	P		S			
	2																			S	P					
	3																									
	4																				P	P		P		
	5									P												S				
	6																									
	2 7																									
	8																									
	9																									
	10																									
	11																				P	P		P		
	12																									
	3 13										P		P													
	14															S										
	15											P	P	P												
	16									S																
	17													P								P				
	4 18							P																		
	19										P															
	20									P											P					
	21			P															P							
	22																									
	23													P								P				
	5 24																									
	25																									
	26																									
	27																									
	28																	P								
	29																									
	6 30									P												P		P		
	31																								P	P
	32																									
	33																									P

Notes: Strong match. Partial match.

The deviation matrix $\Delta M^{1,2}$ is then computed and displayed in Table 7. It characterises the deviations between the two similarity assessment frameworks. A red colour corresponds to a value of –1, which means that a match has been identified with the manual framework but not with the semi-automatic one (i.e., a false-negative with respect to the semi-automatic framework). In another example, a purple colour represents a value of –2, which means that a match was revealed by the semi-automatic framework but was not been detected by the manual one (i.e., a false-positive). Of course, if the two frameworks both identify a match, or they identify no match between two KPIs, the colour remains white.

During the first test of the semi-automatic framework, some manual matches were not found (red cells or false-negatives). Indeed, Table 7 directly highlights the 31 potential conflicts (i.e., red cells) resulting from a first try of the methodology defined to compare the manual and the semi-automatic frameworks (Figure 1). This corresponds to only 3.8% of the complete list of 25×33 = 825 cells, which can be considered reasonably good for a first attempt. However, this result is not satisfactory, as the semi-automatic framework should detect all matches that would be detected with a manual assessment process. Thus, some fine-tuning is therefore needed, as presented next.

7.3 *Comparison of the two frameworks and fine-tuning of the semi-automatic framework*

As mentioned before, the last step of the proposed methodology aims at removing the undesired deviations revealed by the deviation matrix. Thus, particular attention must be paid to the fine-tuning of this semi-automatic framework. As discussed in section 6.2, three types of actions can be taken. Thus, some KPIs were not described by the right keywords (some keywords have thus been removed, added, or modified), some have been moved in the matrix Q, and as a last resort, the matching matrix $M_m^{1,2}$ itself has been slightly modified. Some examples of the three types of actions are detailed in the following paragraphs to better understand these adjustments to the semi-automatic framework.

The experts' first task was to review all the KPIs and ensure they were characterised by the right keywords. This review made it possible to modify, add or delete keywords that had a broad meaning. For example, the KPI C_{32}^{PWC} was characterised by the following three keywords: 'expertise', 'organisation's managers' and 'i4.0'. However, 'i4.0' is a broad keyword that could characterise many KPIs and be placed on several lines of the keyword matrix, so it was removed. The keyword 'expertise' evolved into 'expertise wrt i4.0' to be more precise, and 'organisation's managers' was replaced with 'managers support'. For the KPI C_{12}^{PWC}, the different iterations made it possible to modify the keyword 'collaboration' (to 'external collaboration') to specify its meaning, keep the keyword 'partners' and add the keyword 'customers'.

At the end of the first test of the semi-automatic framework, the experts noticed that there were a lot of lines with only one keyword in the keyword matrix and decided to classify the keywords in a different way. For example, the first keyword matrix had 'i4.0 strategy', 'indicators' and 'strategy review' on three separate lines. These three keywords were grouped together on the same line in the matrix, as they are all related to the digital strategy. Another way is to split a line across multiple existing lines. In the first attempt, the experts grouped 'MES' (manufacturing execution system) and 'manufacturing process' on one line. When modifying the keyword matrix, 'MES' was placed on a line grouping digital tools, and 'manufacturing process', on a line dedicated to production digitalisation. Each time a keyword was modified, the experts made sure the keywords were still related to the KPIs they characterised.

The last way for the experts to tune the semi-automatic framework was to check on the manual matches. Manual matches can reveal a misunderstanding on the part of the experts or that the modification of statements for IMPULS KPIs has not been taken into account. For example, the KPIs C_9^{IMPULS} and C_{12}^{PWC} evolved between the time the manual framework was implemented and the time the semi-automatic framework was used. Therefore, these KPIs are not similar. The first KPI evaluates the organisation's capacity to create a digital model of the factory from production data, while the second KPI focuses on evaluating the collaboration between the organisation and partners. Thus, the experts decided to remove this match from the manual matching matrix.

Ultimately, 139 keywords have been used, with an average of 3 keywords per KPI, and distributed on 57 rows of the matrix Q. These modifications took place in a series of consensual discussions. An intriguing observation -- when checking the matches in $M_m^{1,2}$ the experts noticed that some of the manually identified matches were occurring because of some over-interpretations of the core meaning of the different KPIs -- and this had to be corrected. At the end, the experts had modified 11 partial matches out of the 49 matches in the $M_m^{1,2}$.

While there is a definite increase in the false-positive matches, as there are more purple cells in Table 10 than in Table 7 (57 vs. 46), it should be noted that the semi-automatic framework saves a significant amount of time. Indeed, the experts no longer need to evaluate all possible KPI couples (825), but only those recorded by the semi-automatic framework (95 matches, Table 9).

It should be emphasised that the number of false-positives increases between Tables 7 and 10 because the fine-tuning stage results in:

1 the aggregation of several lines in the keyword matrix, which increases the number of false-positives

2 the attribution of more keywords to each KPI, which increases the number of false-positives as well.

However, the increase in false-positives remains acceptable since the semi-automatic framework automatically eliminates about 80% of the KPI pairs that have no link (true-negatives). This is particularly valuable when comparing more than two maturity models.

The new matrices are displayed in Tables 8, 9 and 10. Having reached the end of the process, the semi-automatic framework has been properly fine-tuned and the deviation matrix $\Delta M^{1,2}$ no longer reveals any false-negative matches (the red cells have been eliminated). The semi-automatic framework has been validated, since the results clearly demonstrate that it retrieves all matches identified by the manual framework. It therefore yields the same results as the manual framework, but in just a second, which was the primary objective.

Table 9 Final matching matrix $M_a^{1,2}$ (see online version for colours)

KPIs PwC		KPIs IMPULS 1						2					3							4	5				6	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	1																			■	■		■			
	2																			■	■		■			
	3													■								■				
	4																				■	■		■		
	5									■									■			■	■			
	6													■								■				
2	7																									
	8													■							■					
	9																									
	10													■								■				
	11									■				■					■		■	■		■		
	12													■								■				
3	13										■		■								■	■		■		
	14															■										
	15							■	■		■	■	■	■							■					
	16									■																
	17												■	■							■	■		■		
4	18							■	■			■													■	
	19										■	■			■											
	20							■	■	■		■						■	■		■	■		■		
	21			■															■				■			
	22							■	■			■											■			
	23							■	■			■		■								■				
5	24																									
	25																									
	26														■											
	27												■													
	28																	■								
	29													■								■				
6	30									■								■	■			■	■	■		
	31																								■	■
	32																■									
	33													■			■					■			■	■

Notes: ■ Automatic match.

Table 10 Final deviation matrix $\Delta m^{1,2}$ (see online version for colours)

Notes: ■ Automatic false-positive match.

8 Conclusions and future works

Digital maturity models help to characterise the maturity level of SMEs with respect to specific KPIs and dimensions, and consequently they provide important inputs with which to design improved digital transformation plans. Many countries and consulting firms have been engaged in developing their own models, and so it is important to understand the positioning of each model with respect to the others. It is possible to envision different strategies to identify similarities and differences between the available models within the Industry 4.0 paradigm. This paper has compared two similarity assessment frameworks: a manual framework and a semi-automatic framework. The manual framework relies on consensual discussions between experts who directly identify how much the KPIs used in one maturity model match the KPIs adopted by another. The semi-automatic framework also compares the KPI pairs, but through set of keywords associated to each KPI. Keywords are sorted in a matching matrix, which helps to characterise the level of matching of the KPI pairs to which they are associated. The manual and semi-automatic frameworks are compared in a common space, the space of the keywords. The KPIs are quantitatively compared based on matching scores obtained using Equation (7). This methodology can be reused to limit the subjective bias of the study when the input data are statements.

Maturity models often do not explicitly propose roadmaps, and it is not the aim of this paper to do so either. However, from the reverse engineered KPIs, one can directly identify the important aspects required for a successful digital transformation. Thus, in a sense, the KPI grid can also give good hints as to what should be improved, even if it does not explain the roadmap steps to be put in place.

The comparison methodology was tested and validated with two maturity models, namely IMPULS and PwC. The results show that the proposed methodology was successful in fine-tuning the semi-automatic framework using results from the manual framework. As an outcome of this fine-tuning, the semi-automatic framework accurately captured the similarities and differences between the KPIs of two maturity models. A pool of four experts played a key role in the reverse engineering of KPIs and of keywords' definition steps. Once this fine-tuning was accomplished, the semi-automatic framework definitely accelerated the comparison process.

The analysis process revealed four valuable take-away lessons. First, mitigation measures had to be set up to avoid over-interpreting the KPIs. In this sense, experts were asked to focus on explicit and tangible information rather than on implicit information that could be open to interpretation. Second, KPIs should be as self-explanatory as possible in order to avoid having to go back to the dimensions or questions to clearly understand the context. Third, the definition of keywords for each KPI is particularly important. It is necessary to characterise all the ideas of the KPIs in order not to miss any possible matches. Fourth, the classification of keywords in the matching matrix should not be neglected, as it is that classification which makes it possible to define a match between two KPIs.

This validated semi-automatic framework now needs to be exploited with additional maturity models. Comparing more maturity models will help to define a common kernel of KPIs shared by many models. In turn, this common kernel will likely help to specify a new maturity model, together with its KPIs and dimensions.

References

Akdil, K.Y., Ustundag, A. and Cevikcan, E. (2018) 'Maturity and readiness model for Industry 4.0 strategy', in Ustundag, A. and Cevikcan, E. (Eds.): *Industry 4.0: Managing The Digital Transformation, Springer Series in Advanced Manufacturing*, pp.61–94, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-57870-5_4.

Becker, J., Knackstedt, R. and Pöppelbuß, J. (2009) 'Developing maturity models for IT management', *Bus. Inf. Syst. Eng.*, Vol. 1, pp.213–222, https://doi.org/10.1007/s12599-009-0044-5.

Cognet, B., Pernot, J-P., Rivest, L., Danjou, C., Wuest, T., Kärkkäinen, H. and Lafleur, M. (2019) 'Towards a novel comparison framework of digital maturity assessment models', in Fortin, C., Rivest, Louis, Bernard, A. and Bouras, A. (Eds.): *Product Lifecycle Management in the Digital Twin Era, IFIP Advances in Information and Communication Technology*, pp.58–71, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-42250-9_6.

De Carolis, A., Macchi, M., Negri, E., Terzi, S. (2018) 'Guiding manufacturing companies towards digitalization – a methodology for supporting manufacturing companies in defining their digitalization roadmap', Presented at the *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 – Proceedings*, pp.487–495, https://doi.org/10.1109/ICE.2017.8279925.

Felch, V., Asdecker, B. and Sucky, E. (2019) 'Maturity models in the age of industry 4.0 – do the available models correspond to the needs of business practice? Presented at the *52nd Hawaii International Conference on System Sciences*, pp.5165–5174, https://doi.org/10.24251/HICSS. 2019.620.

Geissbauer, R., Vedso, J. and Schrauf, S. (2016) 'Industry 4.0: Building the digital enterprise', *2016 Global Industry 4.0 Survey 36*.

Gökalp, E., Sener, U. and Eren, P. (2017) 'Development of an assessment model for Industry 4.0: Industry 4.0-MM. pp.128–142, https://doi.org/10.1007/978-3-319-67383-7_10.

Lichtblau, K., Stich, V., Bertenrath, R., Blum, M., Bleider, M., Millack, A., Schmitt, K., Schmitz, E. and Schröter, M. (2015) *IMPULS - Industrie 4.0 Readiness*, Impuls –Stiftung des VDMA, Aacghen – Köln.

Liu, H. and Wang, P. (2013) 'Assessing sentence similarity using wordNet based word similarity', *JSW*, Vol. 8, pp.1451–1458, https://doi.org/10.4304/jsw.8.6.1451-1458.

Mittal, S., Khan, M.A., Romero, D. and Wuest, T. (2018) 'A critical review of smart manufacturing & Industry 4.0 maturity models: implications for small and medium-sized enterprises (SMEs)', *Journal of Manufacturing Systems*, Vol. 49, pp.194–214, https://doi.org/10.1016/j. jmsy.2018.10.005.

Moeuf, A., Pellerin, R., Lamouri, S., Tamayo Giraldo, S. and Barbaray, R. (2017) 'The industrial management of SMEs in the era of Industry 4.0', *International Journal of Production Research*, pp.1–19, https://doi.org/10.1080/00207543.2017.1372647.

Oztemel, E. and Gursev, S. (2020) 'Literature review of Industry 4.0 and related technologies', *J. Intell. Manuf.*, Vol. 31, pp.127–182, https://doi.org/10.1007/s10845-018-1433-8.

Pereira, A. and Romero, F. (2017) 'A review of the meanings and the implications of the Industry 4.0 concept', *Procedia Manufacturing*, Vol. 13, pp.1206–1214, https://doi.org/10.1016/j. promfg.2017.09.032.

Rossmann, A. (2018) 'Digital maturity: conceptualization and measurement model', Presented at the *Thirty Ninth International Conference on Information Systems*, San Francisco, p.9.

Salkin, C., Oner, M., Ustundag, A. and Cevikcan, E. (2018) 'A conceptual framework for Industry 4.0', in Ustundag, A. and Cevikcan, E. (Eds.): *Industry 4.0: Managing The Digital Transformation, Springer Series in Advanced Manufacturing*, Springer International Publishing, Cham, pp.3–23, https://doi.org/10.1007/978-3-319-57870-5_1.

Samaranayake, P., Ramanathan, K. and Laosirihongthong, T. (2017) 'Implementing industry 4.0 – a technological readiness perspective', in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp.529–533, https://doi.org/10.1109/IEEM.2017.8289947.

Schumacher, A., Erol, S. and Sihn, W. (2016) 'A maturity model for assessing industry 4.0 readiness and maturity of manufacturing enterprises', *Procedia CIRP, The Sixth International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2016)*, Vol. 52, pp.161–166, https://doi.org/10.1016/j.procir.2016.07.040.

Schwer, K., Hitz, C., Wyss, R., Wirz, D. and Minonne, C. (2018) 'Digital maturity variables and their impact on the enterprise architecture layers', *Problems and Perspectives in Management*, Vol. 16, pp.141–154, https://doi.org/10.21256/zhaw-3303.

Scremin, L., Armellini, F., Brun, A., Solar-Pelletier, L. and Beaudry, C. (2018) 'Towards a Framework for Assessing the maturity of manufacturing companies in Industry 4.0 adoption', *Analyzing the Impacts of Industry 4.0 in Modern Business Environments*, pp.224–254, https://doi.org/10.4018/978-1-5225-3468-6.ch012.

Simpson, J.A. and Weiner, E.S.C. (1989) *The Oxford English Dictionary*, Oxford University Press, Clarendon Press.

Taisch, M., Arena, D.N., Gorobtcova, P., Kiritsis, D., Luglietti, R., May, G., Morin, T.R. and Wuest, T. (2018) *The 2018 World Manufacturing Forum Report, "Recommendations for the Future of Manufacturing," World Manufacturing Forum.*

Wendler, R. (2012) 'The maturity of maturity model research: a systematic mapping study', *Information and Software Technology, Special Section on Software Reliability and Security*, Vol. 54, pp.1317–1339, https://doi.org/10.1016/j.infsof.2012.07.007.

Westerman, G., Bonnet, D. and McAfee, A. (2014) *Leading Digital: Turning Technology Into Business Transformation*, Harvard Business Press, Boston.