



This is a repository copy of *Review of material modeling and digitalization in industry: barriers and perspectives*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/207079/>

Version: Published Version

---

**Article:**

Scotti, L. [orcid.org/0000-0002-3807-1694](https://orcid.org/0000-0002-3807-1694), Basoalto, H., Moffat, J. et al. (1 more author) (2023) Review of material modeling and digitalization in industry: barriers and perspectives. *Integrating Materials and Manufacturing Innovation*, 12 (4). pp. 397-420. ISSN 2193-9764

<https://doi.org/10.1007/s40192-023-00318-2>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



# Review of Material Modeling and Digitalization in Industry: Barriers and Perspectives

Lucia Scotti<sup>1</sup> · Hector Basoalto<sup>1</sup> · James Moffat<sup>2</sup> · Daniel Cogswell<sup>1</sup>

Received: 11 November 2022 / Accepted: 4 October 2023 / Published online: 10 November 2023  
© The Author(s) 2023

## Abstract

Materials modeling technologies are fundamental to explore, understand, and ultimately predict materials behavior. They are essential to solve challenges posed by the need to reduce human impact on the environment. Modeling and simulation of materials behavior have been recognized over the years as fundamental as an asset in industrial R & D, guiding the decision-making process regarding the design or optimization of new products and manufacturing processes. At the same time, it reduces product cost and development time. However, highlighting the revenue brought by using such tools is not trivial, especially because they mainly affect the complex activities such as the innovation process, whose return only becomes available in the long run and it is difficult to measure. This means that the materials modeling field is often overlooked in an industry setting, where it is not integrated in the company workflow. In some cases, modeling provides the potential to capture tacit knowledge preventing the loss of capability in an aging specialist community, that why its industrial integration is important. This paper explores the reason behind this dichotomy, presenting first what it is intended for the modeling process, and the main types used in materials application. The current industrial adoption is reviewed by outlining success stories, economic impact, business uptake, and barriers. Past and current approaches and strategies are also presented and discussed. In prospective, materials modeling plays a key role in developing material-centric industry for sustainable economy, providing physical understating (physics-based models) and fast approaches (data-driven solutions). Digitalization is the mean for the green economy and it needs to push for a more integration at the core of the business of materials modeling.

**Keywords** Materials modeling · Digitization · Industrial adoption · Modeling process

## Introduction

Over the last few decades, Modeling and Simulations (M & S) have become familiar assets used in different sectors from weather forecasting to medicine. From an engineering

and science point of view, these technologies are nowadays ubiquitous tools, including the understanding and prediction of materials behavior. Modeling techniques have in fact accelerated materials development, leading to discovery of new materials and applications, reducing expensive and time-consuming physical tests [1]. The Integrated Computational Materials Engineering (ICME) [2] paradigm has demonstrated to be a great success in promoting progress in the aerospace industry for example. Materials models and simulations are today exploited to effectively and efficiently design new products and/or optimize processes by reducing the need and time of prototyping and testing [3]. Materials modeling has driven the acceleration of the better-performing product design, with a corresponding reduction in the production costs and time-to-market. For sectors that are typically characterized by highly volatile and competitive markets, materials modeling plays a key role for the innovation and competitiveness.

---

✉ Lucia Scotti  
luciascottis@gmail.com  
Hector Basoalto  
h.basoalto@sheffield.ac.uk  
James Moffat  
james.moffat@manchester.ac.uk  
Daniel Cogswell  
d.cogswell@sheffield.ac.uk

<sup>1</sup> Department of Materials Science and Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

<sup>2</sup> Department of Materials - Henry Royce Institute, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

However, technological and cultural barriers still remain, and the full potential of these methods has been not achieved [4]. Materials modeling is often not considered as an essential component of commercial and business processes, and is not well integrated with other tools such as Business Decision System, stopping it to become a decisive rather than supportive tool in the material design and selection [5]. Modeling tools are often regarded as too complicated to use, inaccurate, and/or time consuming. Their development and maintenance require specific expertise (mathematics, physics, materials science, computer science, and engineering), infrastructure (high performance computing (HPC)), and trusted databases. There are different stakeholders (scientists, modelers, and end-users) involved in the development and application of modeling tools. They often do not share common language, backgrounds, goals or methods, therefore making communication and project delivery difficult.

Today, sustainability policies and increasing market competition have put materials in the center of the product and process design, with materials modeling as a key part of the design process. Actions are needed to facilitate the use and development of mature models and simulation tools across various industries and application areas.

There are national and international projects that try to assess these challenges, including the Materials Genome Initiative (MGI) [1], NASA Vision 2040 [4], European Materials Modeling Council (EMMC) [6], and in both academia and industry materials science communities. The aim of the MGI project is to facilitate materials development by exploiting computational techniques, use of standards, and data management by all stakeholders in the materials development community: academic institutions, small businesses, large industrial enterprises, professional societies, and government. The NASA roadmap identifies the technical and cultural challenges, deficiencies, and the technical work areas required to build a collaborative digital environment for the aerospace industry. Similarly, the EMMC mission is to improve collaboration and integration between all the stakeholders involved in modeling and digitalization of materials, by increasing awareness of the materials modeling industry, supporting the software development industry, and creating collaborative partnerships between academia and industry.

The purpose of this paper aligns with aims of these activities to create awareness of the materials modeling sector outside the community, to review the challenges that obstruct industrial adoption, and to analyze the promotion methods that have been pursued over the years in the UK and internationally. For this reason, the paper is divided into three main sections. The first presents the modeling background, introducing the terminology behind the model design process, from real world problems to the verification and validation process, including a general categorization

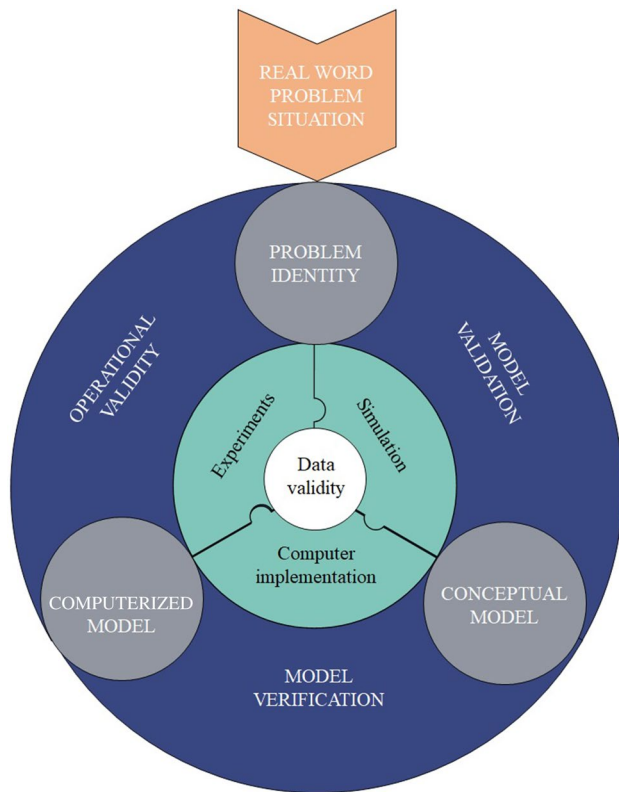
of models. Multiscale materials modeling approaches are then illustrated. The uncertainties affecting the model output and their quantification is also discussed. The current state of industrial adoption is then examined in the second part of the paper. The methodology used to assess the maturity level of the technology is presented, followed by a review of application and economic revenue in industry. This section closes with discussion on the current industrial integration of materials modeling and the barriers and challenges that impede industrial adoption. The final parts regards a review of initiatives adopted by the community to overcome these barriers and the needs for materials modeling integration via digitalization to promote a more sustainable business.

## Background

### Modeling Process

The aim of any scientific research is to investigate, understand, and ultimately predict the evolution of the system under study. Given a real problem situation, the first step in research study is defining a so called problem entity which identifies the system and process to be investigated [7] (see Fig. 1). The conceptual model is then defined to represent the physical system that is appropriate to solve the problem entity [7]. Due to the physical world complexity, conceptual models are hardly an exhaustive description of the system, and need to be regarded as an accurate approximation of the reality, where only the most-effecting physics phenomena are considered. For this reason, conceptual models need to be validated by executing a series of investigations, either experimentally and computationally, to establish the accuracy and precision of the conceptual model.

In the computational modeling route, the conceptual model is translated into a computerized model. The computerized model consists in a numerical formulation of the conceptual model and its implementation in computational form through an appropriate algorithm using a chosen computer language. It is important to ensure that the code solves correctly the mathematical model. This step is called model verification. It consists of two parts: the code verification and the solution verification. Code verification compares the simulation results with a known solution in order to determine if the computerized model is a good representation of the conceptual model. The known solution used to test the code can be an analytical approach or assumed *a priori* (Method of Manufactured Solutions [8]). Solution verification is applied to minimize the numerical errors without excessively increasing the computational expenses. Similarly, to the code verification, the numerical error is evaluated by comparing the results with a known solution, either analytically or arbitrarily. Once the verification process is



**Fig. 1** Modeling process from the definition of a problem entity based on real word situation problem, to formulation of conceptual model and its implementation into a computerized model

concluded, the validation process can begin. The validation process requires the comparison of simulation results with real physical data. The set of simulations and experiments to be performed need to be carefully designed accounting for the accessibility of methodologies, facilities, and data (e.g., characterization techniques, computational time on HPC), operability of selected conditions (e.g., high temperature, boundary conditions, input parameter availability), ability to perform measurement with an acceptable degree of confidence (e.g., 95% confidence interval is usually accepted in industrial settings). This process is called experiment design, and it requires the formulation of a physical model to reproduce the conceptual model. During the experiment design, a set of variable values that can be obtained from both experiments and simulations for comparison. These variables are called Quantities of Interest (QoI).

Both modeling and experiment QoIs are affected by uncertainty due to the variability of physical parameters within the system itself, and simplification within the model. Model uncertainties can be only partially reduced, so it is important to estimate how much they affect the results. Similarly, the blocking process of arranging experimental units into homogenous groups aims to reduce the variability of experiment results. The experimental and simulation

outcomes are then compared during validation to understand how accurately the simulations reproduce the experiments. If the agreement is acceptable the model is validated, otherwise the model and experiments are revised. This process is called Verification and Validation (V & V) of the model. V & V assesses the accuracy of the computational and conceptual model to assure researchers that they well represent the problem entity to solve. This process is not trivial and requires expertise in model development, software development, simulation analysis, and experimental methodology. Figure 2 summarizes the flowchart of the V & V process as presented in the American Society of Mechanical Engineers (ASME) standard for verification and validation of modeling [9].

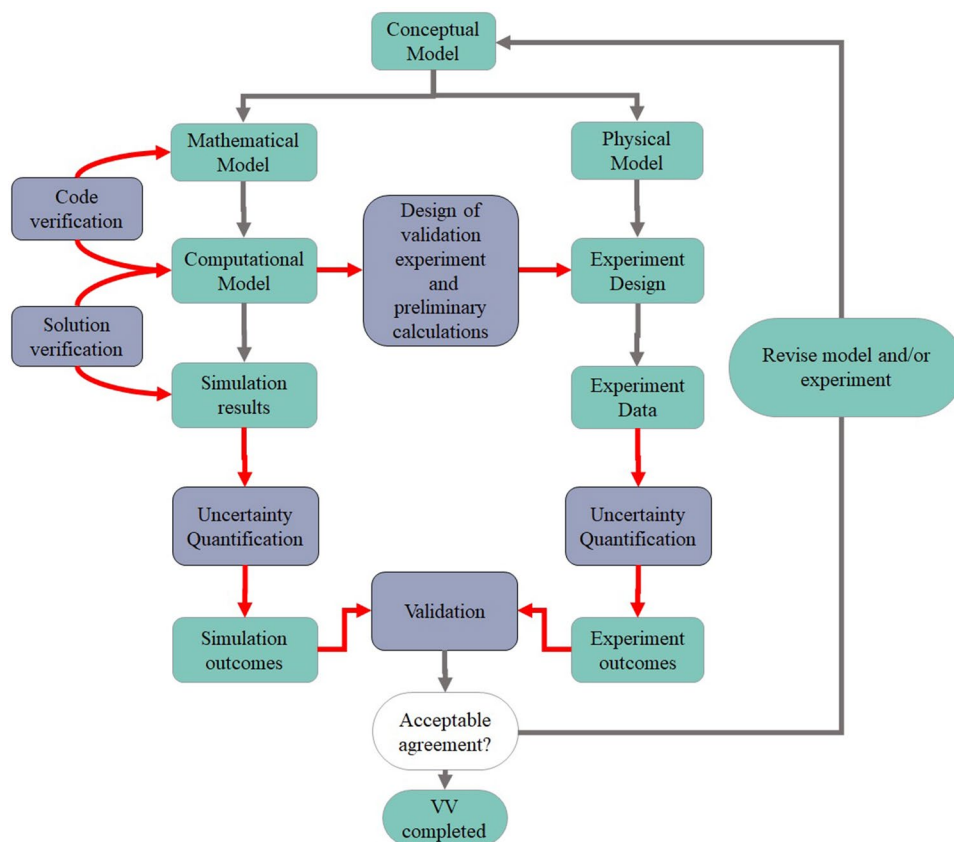
### Model Types and Definition

Computational models can be divided in two main categories: physics-based and empirical models, which can include data-driven. Physics-based modeling translates our understanding of the physical system into a mathematical form [10]. These models are less biased with respect to unverified assumptions or spurious correlations, because they follow the underlying principles governing the behavior of the material. Biases can be still introduced by the choice of the laws, but it remains a transparent choice. These models can be applied to different systems with the same underlying physics, and not restricted to the application they were created for. However, spatial and temporal resolution required by this models to simulate physical phenomena can impose high computational times, limiting their use to the design stage. Outputs are affected by errors and uncertainties in parameter values as well as lack of understanding of the system.

While physics-based models are extensively used in research and academia, they struggle to be widely accepted in industry. The main objection is that they are slow compared to the empirical methods and experimental trials [11]. However, they provide a wide range of data that would take a considerable amount of time and money to generate experimentally. For example, in addition to simulation of processes, simulation datasets can include microstructure and defect distributions as well as variations in property scatter. The underline problem is that this is not always well perceived.

For empirical models, the opposite is true, with the input/output relationship extrapolated from observed data. This may be done by the researchers directly, using experimental data to find correlations, or by using Data-driven modeling like Artificial Intelligence (AI) and Machine Learning (ML) approaches. Data-driven modeling is becoming popular especially in the “digital twin” context, due to the availability of vast quantities of digital data and the increasing

**Fig. 2** Verification and Validation framework of modeling technology showing the fundamental passages in blue boxes: code verification, design of validation experiments and preliminary calculations, calculation/solution verification, uncertainty quantification on the model and on the experiments, and validation. The green boxes define the fundamental entities in V & V process: conceptual model, mathematical and physical model, computational model, experiment design, simulation and experimental results, and experimental and model results [9]



availability of low-cost, high-performance computer hardware. For example, in the image-recognition field, digital capture of images is the norm, effectively automating the data capture process and readily creating libraries of images to process. This results in a rise of open-source platform for AI/ML such as TensorFlow [12] and scikit-learn [13].

Compared to physical-based models, empirical-based models suffer less of instability issues. They take into account long-term historical data and experiences. They are significantly cheaper in terms of time and computational resource in comparison to a physical-based model. However, they cannot be easily generalized and the choice of the training data is subject to bias, which can influence the results. The validity of their outputs generates debate, especially in the materials field. The output behavior is controlled by a large number of variables and obtaining a reliable dataset is a non-trivial process, due to the complexity associated with the effect of composition, microstructures, and process parameters. However, some argue that since each manifestation of the system is a result of both known and unknown physics, data-driven models can account for the full physics and potentially help to uncover the unknown phenomena [10]. However, for any set of data an arbitrary relationship can be always drawn. Traditionally in industry, empirical models are preferred over physics-based models, because in many cases, the computational resources and skills

necessary are not available [11]. The empirical models used/developed rely on data availability, and their applicability is restricted to the particular condition modeled. They require the calibration of a number of parameters, often obtained through extensive experimental trials [11, 14]. They represent are computationally fast and once calibrated can provide readily solutions for a specific system, but the extrapolation, such as to another alloy, can result in severe drop in accuracy of the prediction [11, 15]. Accounting for some physical knowledge in the building of the data sample and in design of the model relations can increase the chance these models can capture the full physics. For example, physical-based models can be used to inform data-driven methods, providing the relevant physics to guide training of these the algorithms, by adding physics-based constraints in the optimization process [16], or by a physics-based architecture of neural network [17]. These types of hybrid methods are known as Reduced Order Models (ROMs), Surrogate Models, or Metamodels, and they aim to reduce the complexity of a fully physics-based model and maintain the fundamental physics and features. They provide a balance between fidelity and required computational power. In materials science, the Hall–Petch relationship [18] is an example of a ROM, quantitatively describes the grain size strengthening effect due to dislocation pile-up mechanisms.

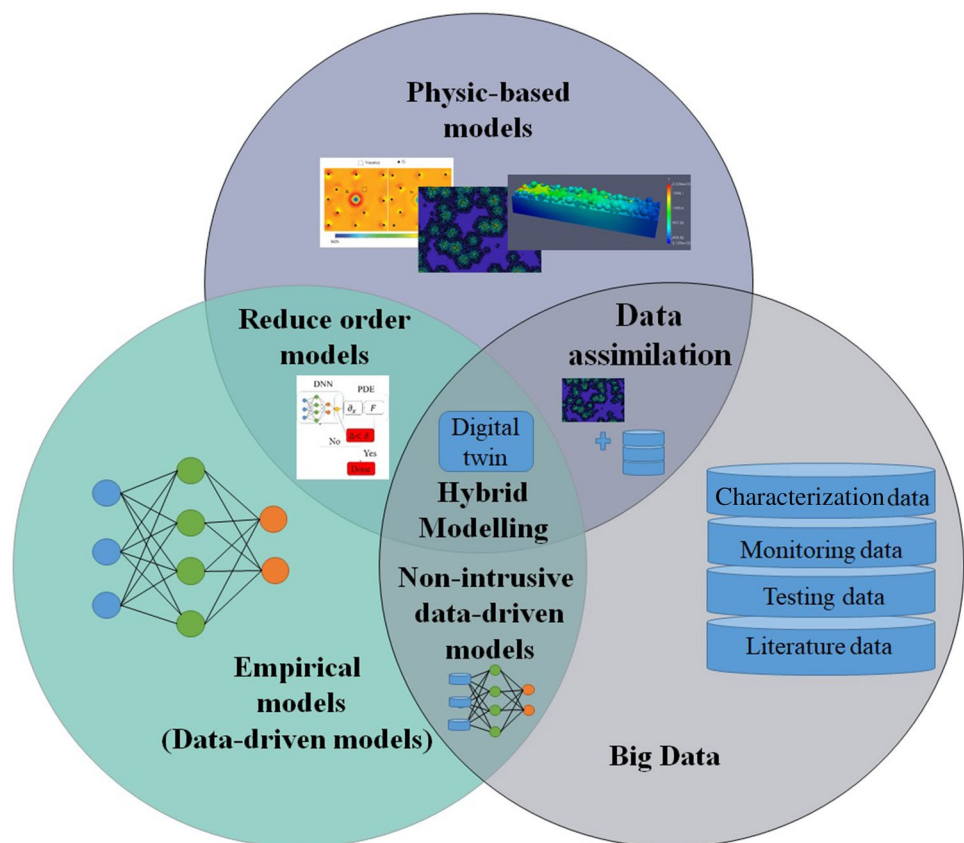
Big Data are also gaining attention as tools to estimate the QoIs. Big data are very large structured and unstructured sets of information that can not be handled in a traditional way. Data Assimilation is a type of hybrid technique, where the model predictions of a physics-based model are improved by using Big Data. This approach is mainly used in weather forecasting [19]. The major challenges are the quality, standardization and processing the data [20]. A part from the Materials Genome Initiative in the US [21], which has substantially funded Open Science, Big Data, and AI initiative, Big Data applications in materials has been met with resistance especially from industry, due to the lack of specialized skills, knowledge, and resources in this field to transform industries into data-driven organizations [22]. Calibration of the Master Curve transition toughness model through Bayesian analysis is an example of this approach in materials science [23]. Data-driven models are also applied in the context of Big Data to establish trends. This type of modeling capability is referred to as Non-intrusive data-driven model. They are defined as Non-intrusive, because they are based solely on data, and there is no intervention or modification of the algorithm. A new field called Big Data Cybernetics or Hybrid Analysis and Modeling is arising, where the combination of data-driven and physical-based models with Big Data allows the inclusion of more complex physics to solve the problem under study. Digital

twins are an example of hybrid analysis, where the architecture combines different types of modeling to inform a decision-making process. Figure 3 summarized all the different types of modeling.

### Multi-scale Materials Modeling

Materials are highly complex systems, whose performances are governed by convoluted physical and chemical mechanisms occurring at different spatial and time scales. For example, atomic-scale defects like dislocations and grain boundaries are well known to have an impact on the macro-scale mechanical behavior [24]. Macro-segregation during casting is micro-scale segregation during solidification [25]. Experimental characterization of these phenomena are difficult or impossible to perform as real-time observations cannot be readily made. Nevertheless, understanding these mechanisms is fundamental to predict materials behavior during each phase of the product life, from design to service. Application of modeling tools plays a significant role to accelerate discoveries in materials science. However, the complexity of the material system is such that one monolithic approach to model all aspects of the physical system across different length and time scales is not presently feasible.

**Fig. 3** Diagram of the three main modeling areas, physic-based, data-driven and Big Data, and their hybrid methods [10]



Multi-scale approaches, instead, use the “divide and conquer” concept: the problem is solved by using a framework of simulations, each one describing a mechanism happening at the nanoscale, microscale or macroscale, which are then connected together like in a jigsaw puzzle. While models at each scale have their own limitations and difficulties, the greater, more central challenge in multi-scale modeling is the parsing and integration of information between scales. This methodology is called Scale bridging, and the approach can be divided into hierarchical and concurrent techniques [26].

The hierarchical method involves separating the different scales, and connecting them in an sequential manner through, for example, volume averaging of a field variable or simple parameter identification from one model to another. Each model in a hierarchical framework runs independently from the others. A homogenization technique using a Representative Volume Element (RVE) is an example of the hierarchical method [27]. A fine-scale model is run on a series of RVEs, each with a given microstructure to calculate the QoIs, usually the mechanical properties [28]. This process generates a distribution of microstructure-informed QoIs which are then input in the higher scale model. Another hierarchical approach consists of developing scaling laws to describe the lower level mechanisms, and incorporating them in a higher-scale model: e.g., a constitutive model [29].

Alternatively, in the concurrent approach, the domain is divided in two or more subdomains that are treated with different scale approaches within the same code and time step. This approach is useful in fracture mechanics, where the crack nucleation and propagation at the macro-scale result from a series of underlying atomistic processes, such as dislocation kinetics and/or nano-void formation, which in turn are affected by the stress field at the continuum level [30]. Quasicontinuum Modeling is another example of the concurrent approach where the atomistic calculation replaces the constitutive law in the continuum framework [31].

The multi-scale approach is an especially powerful technique in the industrial sector, by increasing product development efficiency, developing new materials, and improving materials quality [32]. However, there are still challenges to resolve. Coupling different scales requires the generation of usable and transferable information. In some cases, information transferability is limited by the different assumptions, physics-based mathematical equations, and parameters, which may not be valid at each level [33, 34]. Determining the link between the various scales is important but not trivial to do. In other cases, the assumption of scale separation cannot be applied [33]. Transferability is also limited by the output form of each model, such as deterministic to stochastic, or discrete to continuous descriptions. Defining the output form is necessary to develop bridging techniques to link the multi-scale information together [35, 36]. The set-up

of the architecture to integrate the modeling framework is also challenging. The architecture needs to be appropriate to handle the information transfer output, long simulation time, pre- and post-processing, capturing and managing data [36].

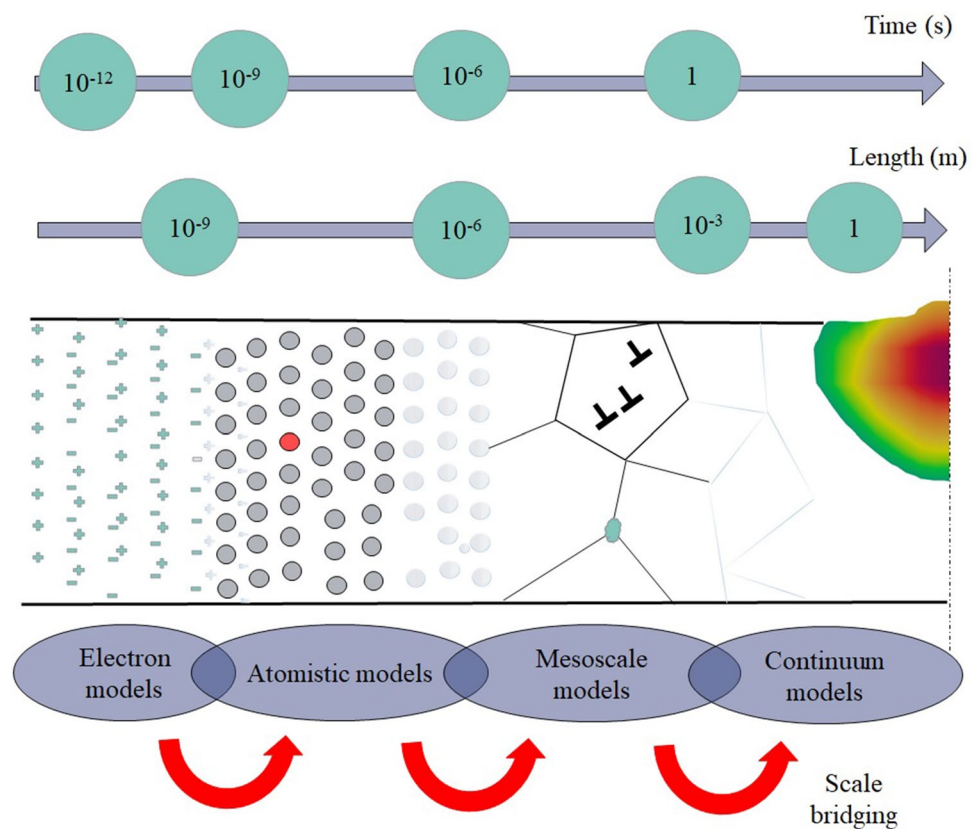
There is also a need for balance between predictability and computational cost of modeling tools. The choice of the model to use at a specific scale is not trivial and it should consider the system under consideration. Smaller-scale models have higher accuracy in describing the physical system but running these models for large domains becomes unfeasible. In order to reduce the computational cost, the appropriate number of physical phenomena should be modeled. From an engineering point of view, this is achieved by inductive goal-mean or top-down approach where the problem is solved from the component level going down in scale. However, materials scientists use a bottom-up approach, aiming to defined a cause-effect relation [37]. The level of details in modeling tools should be chosen in a way to reduce execution time guaranteeing the appropriate accuracy, since a grade of uncertainty is associated with any simulation. Therefore, uncertainty quantification of the model output is important to make appropriate use of them. It is also beneficial to identify the validity range of the model itself. In multiscale modeling frameworks, it is also essential understand how the uncertainty propagates from one scale to another and for it to remain acceptable. The uncertainty quantification process can also identify the most influential model on the overall error, so action can be taken to improve its accuracy.

## Multi-scale Materials Model Classification

There are different modeling techniques used to study materials behavior, many of which overlap different length and time scales. Therefore, it is difficult to identify each modeling technique based only on the spacial and temporal scales. Classification based on four fundamental entities has been suggested to be more appropriate: electrons, atoms, mesoscale entities (e.g., particles, grains and molecules), and continuum (see Fig. 4) [38].

Electron models simulate the evolution of molecules and atomic systems by describing their electronic interactions. They resolve systems at atomic to nanometer length scales and their evolution between pico to nanosecond time frames. The position, momentum, and spin of electrons in the system are described as an electronic wave function, whose Hamiltonian operator represents the total energy. In most of cases, the Schrödinger equation [39] is the physical model, while the Hamiltonian can be approximated by the Kohn-Sham theory [40] or the impurity model [41], to name a few, to solve many-particle problems. The equations are commonly solved using plane-wave basis because it allows an easier implementation of the total energy expressions and

**Fig. 4** Diagram shows the four typologies of modeling used in materials science based on the described entity: electrons, atoms, mesoscale entities such as particles, and continuum [38]



the Hamiltonian, and the use of Fourier transforms [42]. The plane-wave basis allows to compute the electrostatic interaction, however they are limited to small systems with period boundaries that restricted their applications to bulk properties of perfect materials [42, 43]. Orbital-free approaches have recently gained attention thanks to a reduced computational complexity that allows application to systems with larger number of electrons and their implementation in concurrent scheme with finite-element approaches [43, 44]. Green's functions can also be applied to solve the Schrödinger equation [45]. These models are used to provide input parameters for atomistic, mesoscopic and continuum simulations, such as activation energies [46], or giving an insight on phenomena such as thermodynamic stability [47], kinetics of atomic defects [48], and reaction pathways [49]. Their applications include the evaluation of electronic, conductive, and optical properties too, as well as being used to calibrate the inter-atomic potentials or partial atomic charge distributions used for atomistic simulations [50].

Atomistic models describe the behavior of atoms and molecules using molecular or classic mechanics. The atomic interactions (e.g., force fields or inter-atomic potentials) are used instead of the electronic field. This allows simulation of a larger system, spanning from nanometer to micronmeter, and nano to microsecond, with respect to the electronic based methods. The atomic system can be treated

statistically or by deterministic methods. The most common statistical methods are Monte Carlo (MC) [51] based approaches like Markovian Chain or kinetic Monte Carlo [52]. Molecular Dynamic (MD) methods [53] are instead deterministic as they solve the classical dynamic equations to determine the atom motions. Some applications of these model methods include diffusion mechanisms [54], phase transformations (evaporation [55], melting and solidification [56]), evaluation of surface [57] and interface energies [58], characterization of fundamental dislocation properties, such as core energies [59], or grain boundary evolution [60].

Mesoscale models focus on the behavior of nanoparticles, grains or molecules in a length scale that spans from  $10^{-9}$  to  $10^{-3}$  m, from  $10^{-6}$ s to seconds. They do not resolve the atomic motions, instead, these are averaged out or replaced by stochastic terms. They comprise methods to predict microstructural evolution such as Phase Fields (PF) [61] and Cellular Automaton (CA) [62], and material mechanics behavioral methods, such as Discrete Dislocation Dynamics (DDD) methods [63]. These three types of models are all Discrete models, meaning the state variables of the system change only with a countable number of points, while Continuum models treat the system as a continuum, where the space is occupied by the substance of the object in the domain. The continuum volume is divided either into finite volumes, cells or elements. They can be applied to a



large range of processes, from nano to macro level, and they are commonly sorted by their applications. Solid mechanics modeling, for example, studies the behavior in terms of deformation and motion of solid matter under one or more external influences including thermal and/or mechanical loading, or electrochemical reactions. Fluid mechanics modeling regards the motion of fluids, such as liquids, gases or dense plasma. Thermodynamics models are used to study energy conservation. Chemistry reaction models and electromagnetism investigate chemical reactions and electrically charged particles, respectively. These models are based on a set of equations that describe the process under investigation and constitutive laws describing the material behavior. For example, fluid mechanics theory is based on solving the conservation laws of energy, momentum and mass to predict velocity, temperature, pressure and density fields. Linear elasticity is an example of a constitutive law that describes the material strain response as directly proportional to the stress applied.

### Model Uncertainties

Computational models are powerful tools and have accelerated scientific progress in many fields; however, they can be subject to errors. There are four types of error: implementation mistake, numerical error, parameter uncertainty, and model form error.

Implementation mistakes are the errors made during the implementation of the conceptual model. The code verification is the practical process to correct these errors and make sure the model solution is what is expected. Along with techniques presented in previous sections, practices like cross-check and visualization of the output are good methods to find and eliminate this type of error. It is also a good practice to write the code in distinct modules that can be independently reviewed.

Numerical errors arise from the computational approximations that are introduced to practically solve the equations. For example, each number is represented by computer with a finite number of digits, known as finite precision, that round-off its value. In addition, most of the equations cannot be solved analytically, but they are solved in an iterative way. This introduces an error called convergence error, which is the difference between the exact value and the iterative solution. The exact solution is defined as the solution to the equations with zero residual. This error can be reduced by imposing a smaller residual tolerance. Another type of numerical error is the simplification introduced in the equations so that they can be handled by the machine, for example, the truncation of an infinite series expansion. Likewise, continuum variables, and their function, are treated as discrete quantities by computer leading to discretization error. Discretization errors are reduced by increasing the

mesh resolution. The reduction in numerical errors comprises higher computational cost. The solution verification process should provide a balance between accuracy and computational power requirements.

Parameter uncertainty concerns the errors deriving from input parameters. The uncertainty is due to the natural variability of the system such as microstructure variability in a component. It can also arise from the ability to measure the QoI in an accurate way. By nature, this last error can be reduced by advances in characterization techniques, therefore obtaining higher quality data. On the other hand, the natural randomness of the system is not reducible, but it can be quantified usually as a probability distribution.

Finally, model form error is due to the disparity between the real system and the conceptual model. This error depends on a lack of knowledge on the physics governing the system, and model simplifications assumed to reduce the computational expense. It is the most difficult error to quantify, because it requires comparisons with true experimental values that also contain uncertainties, and may not be easy to obtain. The uncertainties in experimental data are due to human errors (e.g., misreading instrument scale), systematic errors (e.g., faulty calibration, poor maintenance), and random errors (e.g., reading fluctuations).

### Uncertainty Quantification

The quantifying of uncertainties is very important during the validation process of a model, which ideally involves comparing the amount of scatter in values from the model and experiment. The scientific approach to measure the uncertainty in modeling and experimental is called Uncertainty Quantification (UQ). UQ methods are classified as forward or inverse analysis (see Fig. 5). The UQ forward problems focus on the analysis on the uncertainty of the model outcomes, and they are conducted to assess the uncertainties of the QoIs (i.e., computed volume mechanical properties) as a function of small scale-features (i.e., microstructure features). Forward studies can be applied to both experimental and computational researches. In materials science, a typical example of a forward problem is the investigation and determination of process-structure–property relationships.

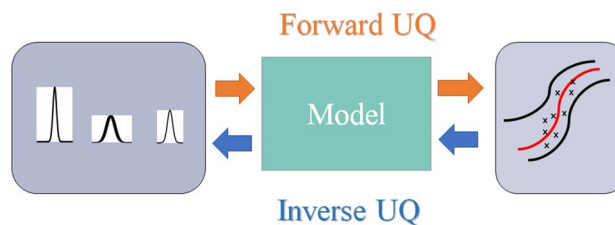


Fig. 5 Forward and inverse uncertainty quantification approaches

Conversely, UQ of an inverse problem aims to optimize the distribution of the stochastic inputs to obtain the target features. Materials design and calibration are examples of inverse problems. The evaluation of uncertainty in an inverse problem is more difficult than a forward analysis, because of the larger number of variables and the possibility of multiple solutions.

Forward and inverse UQ methods are then subdivided in non-intrusive and intrusive methods. Non-intrusive methods treat the model as a black box, as they are implemented along side the conceptual model. MC approaches are an example of non-intrusive methods, largely applied to study the uncertainty in forward problems. They are robust techniques but require a large number of samples for each input, becoming costly quickly both in terms of physical measurements and computation. The efficiency can be improved by using design sampling techniques such as Latin Hypercube Sampling [64, 65] to better cover the input space, or Importance Sampling [66] to reduce the sample size. The input sample can be built by statistical methods such as Karhunen-Loeve expansion [67, 68] or Auto-regressive moving average [69].

Alternative to MC, surrogates models are less expensive. They use outcomes of a physic-based model as training points for the calibration. The selection of these data points needs to be made in a sensible way, balancing accuracy without excessively increasing the number of simulations. Surrogate models speed up the process, however, they are prone to oversimplification, and they can be unreliable in the case of complex systems. Examples of surrogate models are Polynomial Chaos Expansion [70], Singular Value Decomposition [71], which are forward analyses, or the Bayesian method [72], and Gaussian Process regression [73], which can be used for both forward and inverse problems. If the UQ approaches are implemented directly in the original model to take into account the uncertainty internally, they are called intrusive methods. Polynomial Chaos Expansion can be used as a non-intrusive or intrusive surrogate model [74, 75].

The identification of the source of uncertainty reduces the risk of an incorrect outcome in pass-fail decision making. Where possible, these errors can be reduced by improving our knowledge (more accurate models of the system, and/or model parameter) or by increasing the accuracy of the calculation (higher resolution, lower tolerance). When the errors cannot be reduced (usually due to computational constraints and/or to physical measurement limitations), their quantification indicates the level of trust one can put in the predictions. The sources of uncertainty can vary from model to model; here, a summary of the most common in multi-scale materials modeling is presented.

In electronic models, uncertainties are mainly derived from the approximation of the electronic wave functions. In

density functional theory (DFT) [40], the electrons interactions are approximated by fitted exchange-correlation functions introducing a model-form type of error in the outputs. Another model error is the use of pseudopotentials to reduce the computational cost. Pseudopotentials consider only the effect of valence electrons, while the core electrons are considered frozen. This approximation can affect the calculated properties. The choice of the exchange-correlation functionals and pseudopotentials is fundamental and there are a considerable number of publications addressing the fidelity of the results of these simulations. There are also numerical errors due to the threshold on the energy and force convergence calculations, and the size of the simulations domains. In this case, convergence studies are needed to reduce these errors.

In atomistic models like MD, the interatomic potential function introduces model-form type and parameter errors in the results. Other model form errors are the choice of the boundary conditions that can introduce artifacts, and the deviation of the constructed model from the real process. Numerical errors are caused by the picked cut-off distance, round-offs in floating-point, task distribution and sequencing in parallel computation. In kinetic Monte Carlo simulations, the incomplete list of events or wrong event rates are model-form and parameters uncertainties. Another source of model-form uncertainty are the unknown correlations between the events.

In mesoscale modeling, model-form type of errors are introduced by using an empirical model for the stress field and dislocation interactions in DDD simulations or energy functionals in the PF model. Uncertainties also come from the numerical treatment to solve the ordinary or partial differential equations. At the continuum level, sources of model-form errors are wrong or incomplete constitutive laws, and the assumptions made to reduce the computational cost of the simulation. Uncertainties are introduced by the mesh size and numerical calculation to solve the set of differential equations.

## Materials Modeling in Industry

Modeling and simulation are valuable assets for industry. Industrial interests on these methodologies are several, from a better insight of materials behavior to cost efficiency in materials and process designs. These tools are used to find Process-Structure-Property-Performance (PSPP) relationships in a material system [37]; a key link between composition, process parameters, microstructure features, and final material properties. This enables the forecast of the behavior of materials during production and service, optimizing the manufacturing process, and accelerating the design of new alloys. Modeling outcomes can also guide experimental

trials, by scaling down to the elements or features that affect the final results. They are suitable tools to reduce time, cost, and risks associated with the materials design processes, by minimizing expensive or dangerous experiments. A well-documented model additionally formalizes knowledge that could be easily lost over time in order to avoid mistakes and repetition. They are also useful to scale up production and reduce time-to-market through process design.

## Model Quality Levels

As with any other new technology, models lead to improved performance and advance understanding but also introduce a great deal of uncertainty and risk in terms of capabilities, limitations, and development trajectory. It is very important to assess the quality and maturity of any technology, including modeling tools, to ensure they are integrated into engineering projects.

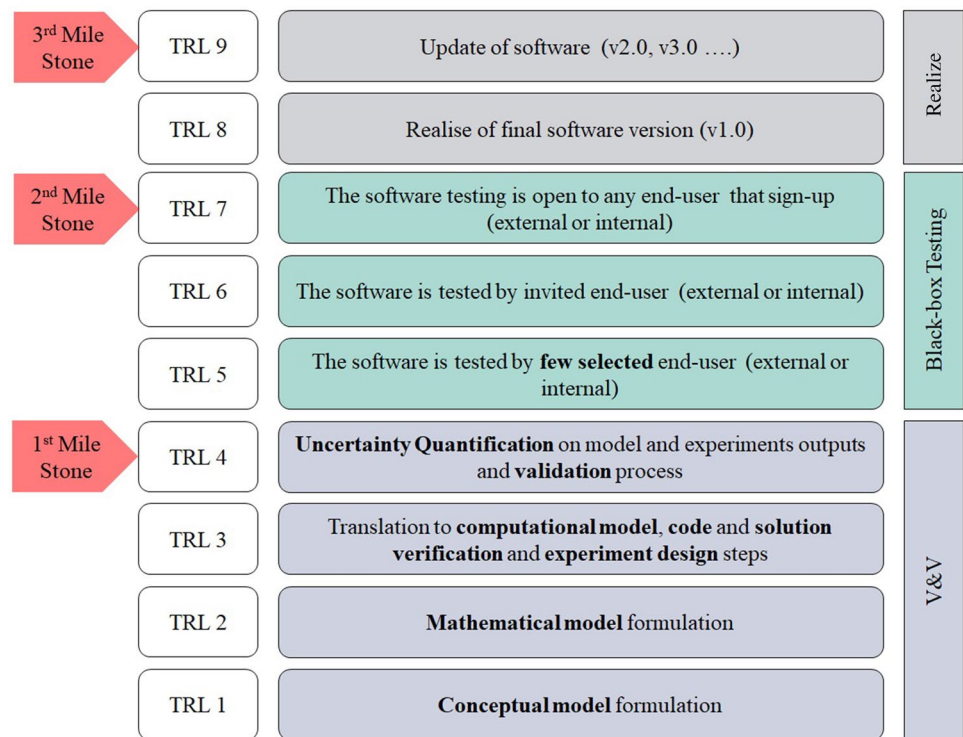
The widely used methodology to assess the maturity level is the Technology Readiness Level (TRL) scale developed by NASA in the 1970s [76] to assure safety, and define emergency strategies before commencing missions. The TRL scale comprises nine levels as summarized in Table 1. There is an initial research phase consisting of levels one to four. In this phase the scientific knowledge (observations) is translated into applied research by the formulation of practical concepts (level 1) into an actual technology or applications (level 2). Level 3 concerns the proof-of concept of the technology through analytical and experimental studies. Low-fidelity laboratory-scale validation of the component assembly is performed at level 4. Following this level, it starts a second phase of fidelity tests conducted in relevant ambient conditions and more representative of service conditions. The component assembly is validated in the relevant environment (TRL 5), then a prototype is created and tested in the relevant ambient conditions (TRL 6) and in an operational system (TRL 7). After level 7, the realize phase starts, where the technology is in its final form and a series of final tests and demonstrations are performed (TRL 8), following by adoption and application of the technology (TRL 9).

In case of modeling, the TRL maturity levels can be translated, as summarized in Fig. 6. There are three milestones defining the three phases earlier described. The research phase corresponds to the V & V process that starts with the formulation of a conceptual model (TRL 1) and its translation to a mathematical form (TRL 2). The computational code is implemented and verified (code and solution verification) at the TRL 3 of the chart. This stage also includes the design of the experiment for the validation. The validation process (TRL 4) concludes this phase together with the UQ on the experimental and model data. After V & V, a so-called black-box testing phase starts, where end-users, who have no knowledge of the implementation, test the code. The code is initially tested by few selected end-users, who can be from outside or inside the company (TRL 5). The testing is then enlarged to invited end-users, both internal and external (TRL 6). At the last stage, the testing is open to any end-users who sign up (TRL 7). The fulfillment of the testing phase corresponds to the second milestone. After the black box testing phase, the first version of the software can be realized in the market (TRL 8), and its updated versions are realized over time until it satisfies the application requirements or it becomes obsolete (TRL 9). In recently years, the Air Force Research Laboratory (AFRL) in the United States sponsored a Technical Interchange Meeting aiming to develop strategies for V & V within integrated computational materials engineering, in order to upraise computational tools to more fundamental role in design and decision-making [77]. The principal outcome was the draft of guidelines for software evaluation called Tool Maturity Levels (TML) [78]. The guidelines consider six categories: model basis and definition, complexity and documentation, supporting data, model verification, rules of applicability, uncertainty quantification, and validation. The maturity of a models is assessed using 5 levels. The first level required definition of the model, inputs and outputs, its application, and its flow diagram. The supporting data are identified. The code, verification and uncertainty quantification plans need to be developed. Level two is reached when the sub-models are also defined with their inputs and outputs, with definition of ranges of model inputs and outputs as

**Table 1** Technology readiness levels developed by NASA [76]

TRL 9	Actual system “flight proven” through successful mission operation
TRL 8	Actual system completed and “flight qualified” through test and demonstration (ground or space)
TRL 7	System prototype demonstration in a space environment
TRL 6	System/subsystem model or prototype demonstration in a relevant environment (ground or space)
TRL 5	Component and/or breadboard validation in relevant ambient
TRL 4	Component and/or breadboard validation in laboratory ambient
TRL 3	Analytical and experimental critical function and/or characteristic proof-of concept
TRL 2	Technology concept and/or application formulated
TRL 1	Basic principles observed and reported

**Fig. 6** The chart shows the nine TRLs used to assess the maturity level of a model. Three Mile stone can be identified corresponding to the completion of the V & V, testing and diploid processes

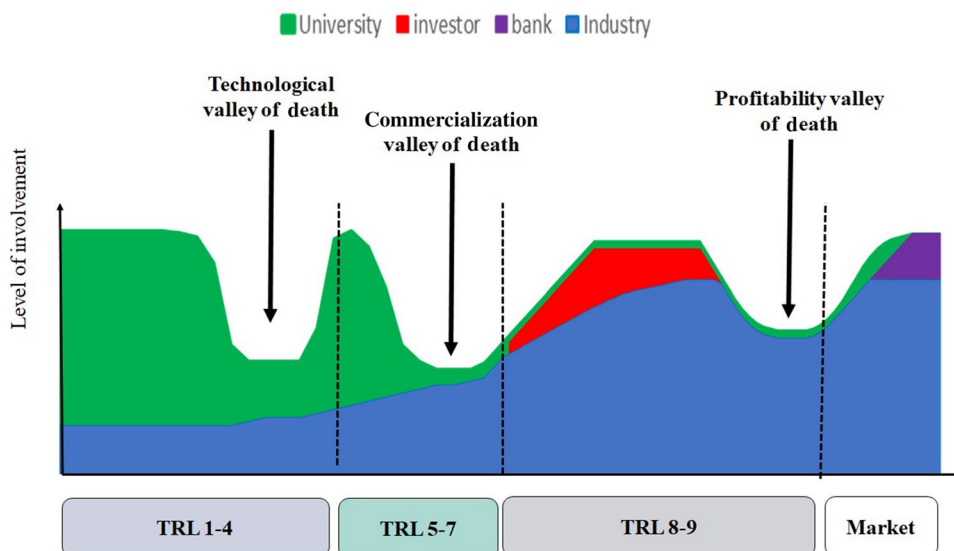


well as the sequential or interdependent computations. The supporting data is documented and archived and used to verified the model computations. The code verification is also completed. A range of applications is defined with the inputs required. The sensitivity analysis is completed and the results trends are consistent with known data. A validation plan is drawn out with risks assessment. Level three consists of finalization of the user guide, definition of dependence of inputs data from other analytical tools, implementation and documentation of version control. The model is verified for range of applications with definition of the supporting data and limitations for each of the applications. Risks assessment is performed for each application. Validation is carried out for sub-models and selected cases. The user guide is updated with reference to the supporting data, uncertainty quantification and validated cases at level four. The model and its accuracy is validated for a range of cases. In the last level, benchmarking cases are carried out and validated with UQ. This assessment was developed for ICME applications and it is thought to be flexible allowing tailoring for the specific case. A model is assessed with gated process meaning it moves to the next level only if it meets all the TML criteria of the current level.

In terms of technology development from R & D to commercialization, these three milestones correspond to failure points, as commonly referred to in the Technology-to-market flow as valley of death as shown in see Fig. 7. The diagram indicates the involvement level of academia

(green area), industry (blue area), investment (red area), and bank loan (purple area), in the Technology-to-market flow. Initially, the maximum involvement is represented by university research with a relatively small amount of support from industry. The first valley of death is the technological one that occurs at TRLs 3 and 4. In the case of computational modeling, it corresponds to the failure of the V & V process. Gated reviews are performed to assess the commercial risk of pursuing further into product development. Evidence needs to be provided that the technology could be successful. At this stage, the modeling tools review processes correspond to proof of concept and V & V. The second block corresponds to the testing procedure. In this case, the university drives this process at the beginning, but industry takes slowly over from TRL 6 onward. During this process, there is the second valley of death: commercialization. Another series of reviews is conducted between level 6 and 7 to assess the economic impact of the technology development. If the review is successful, further investments are made to deploy the commercial software in TRLs 8 and 9. Industry leads over the commercialization of the technology with little-to-no effort from academia. The final failure point, preventing the distribution in the market (e.g., constant improvement, standardization), is the profitability valley of death. This failure is driven by the small revenue at market entry of a new technology. As shown in Fig. 7, loans and/or investments are required to overcome this last valley.

**Fig. 7** This chart shows the development stages of a new technology from initial research (TRLs 1–4) to commercialisation (TRLs 8–9) indicating the effort from University (green area), industry (blue area), investment (red area), and bank loan (purple area)



## Application Cases and Economic Revenue

There are few examples of successful industrial applications of multi-scale materials modeling within industry in the literature. The Virtual Aluminum Castings (VAC) software package [79] was developed, for example, in 2006 by Ford Motor Company to simulate casting and heat treatment. The framework predicts microstructure and residual stress, and it forecasts virtual performance of the service life of the component. It consists of an integrated multi-scale modeling framework including atomic to component level models. It is the result of a deep understanding of physical phenomena involved through theoretical, experimental and computational studies. It has enabled a 15–25% reduction in the development time of a component, such as a cylinder head [79]. More recently, Rolls-Royce has developed precipitation modeling tools in collaboration with the University of Birmingham [80]. This model allows simulation of the precipitation kinetics as a function of their position in the matrix to optimize the heat treatment and component life of the alloys developed [81]. The tool was successfully used to predict a quantitative trend of the manufacturing process, and accelerate the design of new alloys for gas turbine application within five years [81].

In terms of economic revenue, a survey conducted by Goldeck Consulting Ltd in 2015 reports that the adoption of material modeling resulted in improved innovation accomplishment of 80% and revenue increase for 30% of interviewed companies [82]. The average investment in material modeling was £ 1 M, with an average cost saving due to materials modeling equal to £ 12 M [82]. Continuum Materials models are the most ubiquitous in industry with an estimated 75% of the materials software market [83]. The total market size of Computer Aided Engineering (CAE)

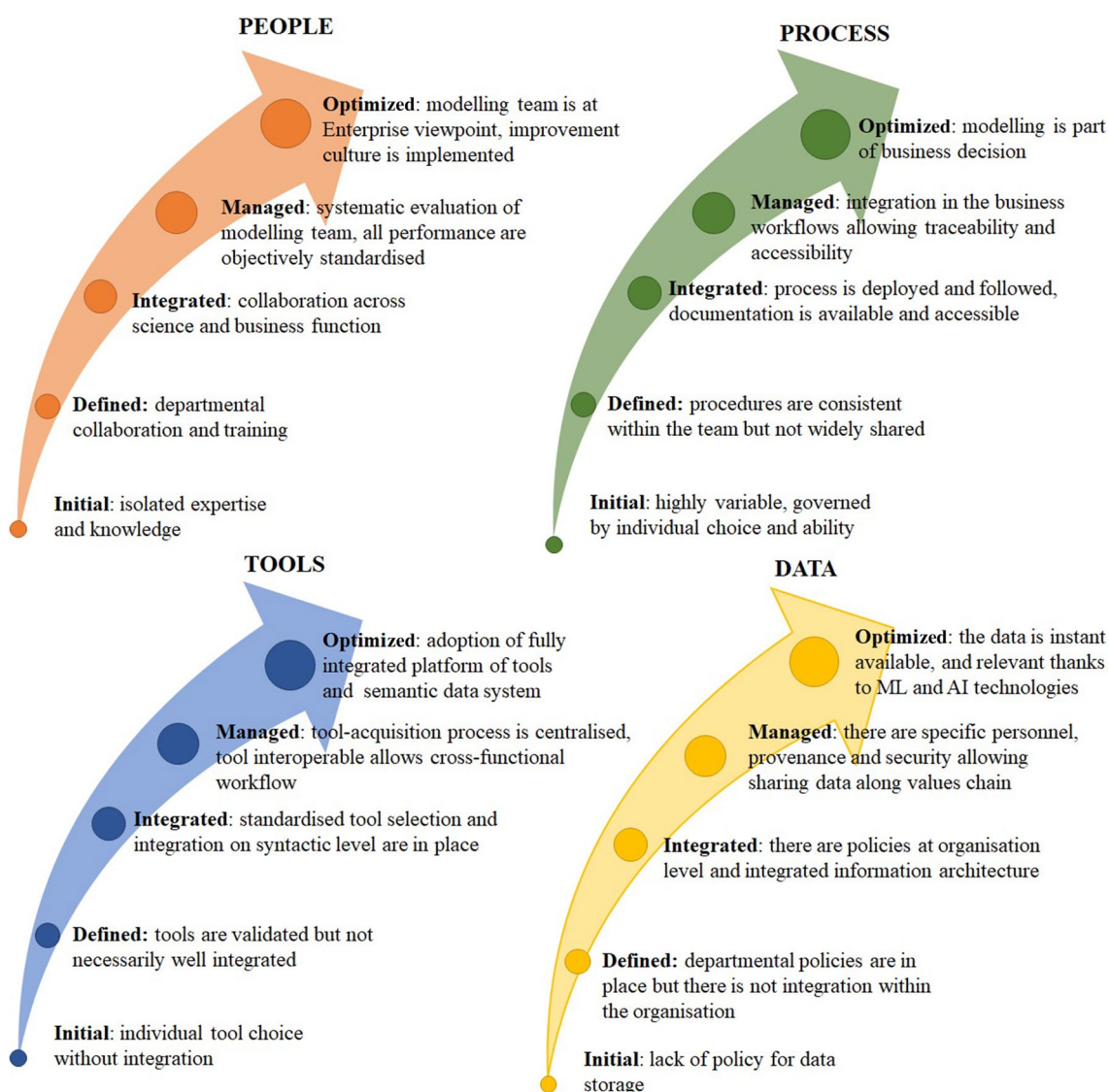
modeling is estimated to be in a range of \$5–8bn [83], where the software development for materials applications is 5–10% of the market [82]. The discrete modeling (electronic, atomistic and mesoscale) softwares started to be developed in the 1970s concomitantly with the spread of HPC in the life science industries. In the 1990s, HPC was adopted by other industrial sectors such as chemistry and materials engineering. Since then, the market of discrete modeling outside the pharmaceutical field has increased, reaching a value of ~ \$50 M in 2012 [84].

## Company Values

These are valid examples of how materials modeling is a credible tool for industry, in terms of cost and time saving, and technological and scientific advancement. However, like other technologies, a consistent advantage through time can only be achieved if materials modeling is valued through the core of the industrial business. A protocol to monitor the level of materials modeling adoption in industry from a business point of view has been suggested by Goldbeck and Simperler [85]. The method is based on Capability Maturity Model Integration (CMMI) [86] and it assesses the impact of material modeling on four factors: People, Process, Tool and Data. The People factor includes all the relevant personnel within the organization to materials modeling, from the expert modelers to occasional end-user. The Process factor regards not only the modeling workflow itself but also the workflow within the whole organization. The Tool factor comprehends the selection and application of software and hardware within an organization IT infrastructure. The Data factor concerns the storage, handling and management of the information obtained within the organization.

For each of these factors, a five-level scale (see Fig. 8) was set to measure the exploitation of modeling techniques in the company. In the initial stage, the knowledge is restricted to individual people, and the tool choice is subjective to the modeler. At this stage, the workflow is not established and inefficient, and there are insufficient data management policies to reliably record parameters and simulation outputs. The next level is called defined, the knowledge is shared within a larger group of interested subject matter experts (a research group or specialized department). The workflows inside this materials modeling group is consistent but not shared outside the group with the extended organization. The individual tools have been validated internally, but they are not yet integrated into an automated system to share

information. Data policies have been established but they are not consistently applied across the organization. At the next step (integrated level), the research strategy is consolidated at the business level, resulting in the integration of policies and objectives for materials modeling. The process workflow is well-documented, assimilated, and straightforward to follow. The selection and application of modeling tools is standardized and integrated on a synthetic level. The data are managed at an organization level through an integrated information architecture defined by uniform data policies. At the managed level, the evaluation performance and objectives of the modeling teams performance is standardized. The modeling process is integrated in the business workflow, and all data and metadata are traceable. The modeling tools



**Fig. 8** Capability maturity model integration protocol measures the integration of modeling tools in an industrial setting in relevant four factors: people, process, tools and data [85, 86]

acquisition is now centrally managed, and the models are inter-operable across domains and cross-functional across the workflow. The company has data-dedicated personnel, provenance, and security, allowing data sharing along the value chain. The highest level of adoption, the optimized level, requires full integration of the modeling process in the business plan, and the highest level data-infrastructure management where all data are available, prompt and relevant across the entire organization. The modeling team is part of the enterprise level viewpoint, and resources are spent toward continuous improvement of modeling tools and data infrastructure. The tools are fully integrated with semantic data flows.

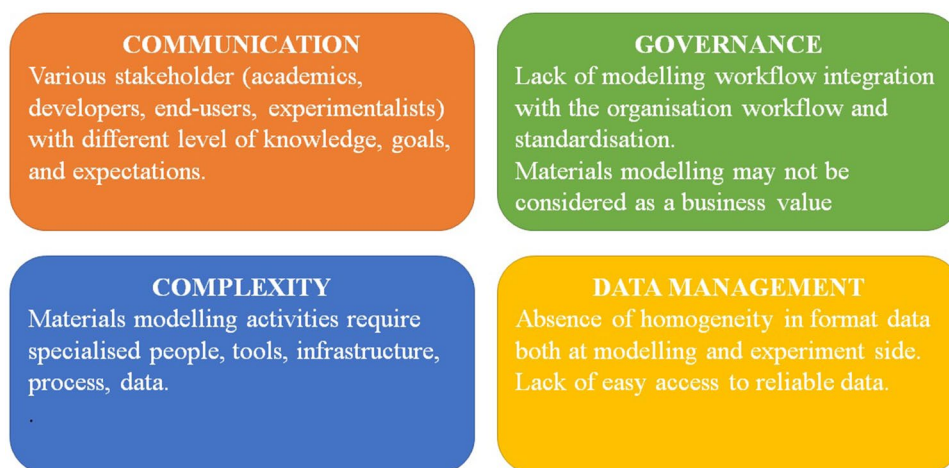
In 2018, this protocol was used by Goldbeck and Simperler to assess the adoption of materials modeling in industry [85]. They interviewed 120 materials modelers across 16 companies, from pharmaceutical through special chemical and materials fields, to the manufacturing industry [85]. Their study showed that the majority of companies reached the integrated level in regards of people and tool factors, while process and data factors lag behind between defined and integrated levels. Overall the average CMMI level was 3.2 in manufacturing industries and 2.5 in chemical and materials industries [85]. These data show that materials modeling technologies are recognized as valuable tools mainly by specialized personnel. Each modeling group reaches an advanced level of knowledge and standardization. However, this is not translated at the management and business levels. The materials modeling development and evaluation are not considered as part of the business values. The know-how acquired is not immediately traceable, readable, transferable, or retainable. The Goldbeck and Simperler study concluded that, although there remain barriers for their complete endorsement, modeling tools have demonstrated themselves to be technologically invaluable and successful.

## Barriers

The report of Goldbeck and Simperler [85] shows that materials modeling is not considered as a core industry value and barriers to full adoption still exist. In general, the main obstacles to the adoption of materials modeling can be summarized in a lack of communication and governance, high complexity, and difficulties with data management (see summary in Fig. 9). The trust in integrated modeling tools is severely effected by these problems, and in some cases the loss of confidence can completely undermine developments in industry and discourage final users from embracing these technologies.

Various actors are involved in the development and use of these technologies: modelers, software developer and end-user who work as researchers in academia, research technology organization, or R & D intensive industry. The variety of understanding of the technology between these groups can be stark, leading to potentially contrasting goals. Communication between these roles can be complicated due to the different backgrounds, motivations, and terminologies, and it is recognized as one of the main barriers to adoption. Academics, for example, are interested in solving very complicated problems, and looking in depth at the physics and chemistry that govern the phenomena involved. They sometimes underestimate the problems that occur in an industrial setting, or find them unengaging [87]. Their research interests are also not static but they evolve moving from one subject to another. They do not concentrate on one aspect, but they like to keep exploring and pushing the boundaries of their knowledge. Moreover, this community tends to be more motivated by exploration and pushing the boundaries of the state of the art, rather than long-term continuous focus on a particular solution or improved user-experience. These aspects may conflict with the industrial point of view. Industry organizations tend to push for simple solutions that are fast and easy to implement. This is clearly

**Fig. 9** Summary of the barriers that stop materials modeling integration



at odds with the university experts who tend to push for more complicated solutions. In the academia-industry relationship, these cultural barriers limit the creation of strong collaborations, and the communication can be difficult in both directions. Industry partners sometimes fail on synthesizing and communicating the problem they want to solve. As discussed earlier, defining the problem entity is an important step for the development of modeling tools. With a vast community like materials modeling, where research groups tend to focus on a specific methodology and/or materials application, a well-defined question assists with identifying the appropriate technologies and expertise. At the same time, academics and modelers may struggle to transfer efficiently the technologies to people that are not highly-trained across an industrial organization. Alternatively, end-users frequently apply modeling tools as black boxes, which can lead to improper, potentially dangerous, results. Academics can also over-estimate the scope of applicability of modeling capabilities in more complex settings, reducing the confidence of the industrial partner in case of large errors and/or failure. Strong communication is also needed between modelers and experimentalists. As already said, experiments are essential to provide data for validation of the modeling tools and to provide model parameters. The experiment design is, in fact, a fundamental step during the V & V process, and it requires a mutual understanding of problem system and conceptual model. Modelers need to explain how the model works, and what inputs and outputs are required. Experimentalists need to make clear what the type of experiment can be conducted to validate the model. It is important to know what can be measured, monitored, the limitations of technologies used, and how to quantify the uncertainties. A mutual understanding from the beginning is important for the success of the V & V process and can require the dedication of time and resources to ensure it is done correctly. Like the academic experts, the experimentalists can also overestimate their in-house capabilities, undermining the V & V process. Modelers may struggle to provide a complete description of their work, impairing the communication of the advantages of these technologies. Often, their work is not recognized, either because it requires specific knowledge, or because of skepticism in the accuracy/ability of the model [85]. The skepticism with regard to modeling tools arises as a result of a lack of awareness or complete description of the quantitative accuracy of a solution for a specific problem. Models can provide, instead, direction and/or reasonable range (output that includes uncertainties) to decision-making direction. Providing realistic and transparent information from the start of the process would assist the integration of materials modeling in industry. In addition, there is an evident gap between modelers and other important figures such as design engineers and business managers who are not able to recognize the potential benefits gleaned from R

& D, product development, process optimizations, and cost effectiveness of materials modeling. In general, outside of the modeling community, modeling is not well understood, defined, or well-established within industrial organizations [85]. This impedes the integration of materials modeling in the vertical value chain of the company process [85]. The materials modeling position is not consolidated or, worse, not accounted for in the business processes, policies, and standardization of the organization. In many cases, materials modelers are only involved in a problem as a last resort, while to fully exploit their potential, modelers should be involved as early as is practical for a given project.

Related to poor communication links, the lack of workflows and standardization in the organization can significantly affect the governance of materials data and information [22]. The first barrier in the workflow is between the university and industry during the TRLs 1–4 as shown in the technology development chart (see Fig. 7). The V & V process and software development are demanding and require skills across multiple disciplines. Academic research groups, for example, alone do not have the capability to deliver user-friendly software. The adequate standard to produce documentation and software, and perform a V & V process that can be easily translated to industry may not be followed by university researchers. In particular, the ability to create a user-friendly code that has been verified across several use-cases is out of scope for most of the academics [88]. Even when modeling tools are successfully developed, there are barriers inside the industrial organization that stop modeling tools from being effectively used. Modeling tools available are not well integrated with other tools in multidisciplinary design and optimization architecture [5]. This makes it difficult to access these technologies, especially for end-users. Developed codes can be archived even if they are successful as they are created to solve a specific problem. The absence of standards also leads to bias in the choice of which tool to be used. The materials modeling field is populated by a vast heterogeneity of codes and software to choose from. As already discussed, the choice of modeling tool to be used is a difficult one, and it needs to be addressed in relation to the problem to be solved. The lack of standardization and policies can lead to an arbitrary choice of the modeler who may favor one technique over another depending on their background. The lack of governance is frequently due to the fact that materials modeling is not well enough established as an asset to business, which is often connected to insufficient ability of modelers of shining a light on the business advantages in terms of spending and time. This affects the recruiting process, making it difficult to attract and retain a valuable workforce.

Complexity is also a major barrier to the adoption of these techniques and it affects mainly Small and Medium Enterprises (SMEs), which do not undertake a dedicated



research activity or have access to HPC facilities [85]. Establishing a hardware ecosystem from scratch is neither trivial nor cheap, and it requires a costly process that involves putting in place specialized tools, processes, data management, and a trained workforce. Making materials modeling technologies more accessible to all industries including SMEs is a requirement for their integration across the supply chain. Nonetheless, problems can arise from over-simplification. The naive use of desktop software can, in fact, lead to misleading results if used with the wrong set-up for the simulation (e.g., boundary conditions, mesh size), effecting the reliability of the output. The expertise of end-users required for model development (mathematical methods, physics, materials science, and computing) may not be at an appropriate level to prevent escapes of inaccurate outputs. However, these challenges existed equally in other modeling domains, such as finite element analysis (FEA) and computational fluid dynamics (CFD), and have been overcome and matured for wide-scale industrial adoption and standardization.

Another obstacle to modeling is the absence of extensive and reliable materials data. In fact, modeling activity needs data as input parameters, boundary conditions, and output for validations. Data availability needs to be considered in the building of a model and their capital investment should be evaluated early on [87]. The difficulties to gather data both for inputs and outputs leads to technical issues and cost effecting even more accepted methods such FEA and CFD. Another problem that affects data is the presence of different types of format, which makes it difficult in the handling and storage [85]. The variety of data formats is driven by the prevalence of proprietary data formats used in scientific instrumentation and modeling technologies. This variation makes it difficult to extract, establish, and support metadata

for the the creation and implementation of a Materials Data Infrastructure (MDI) [89].

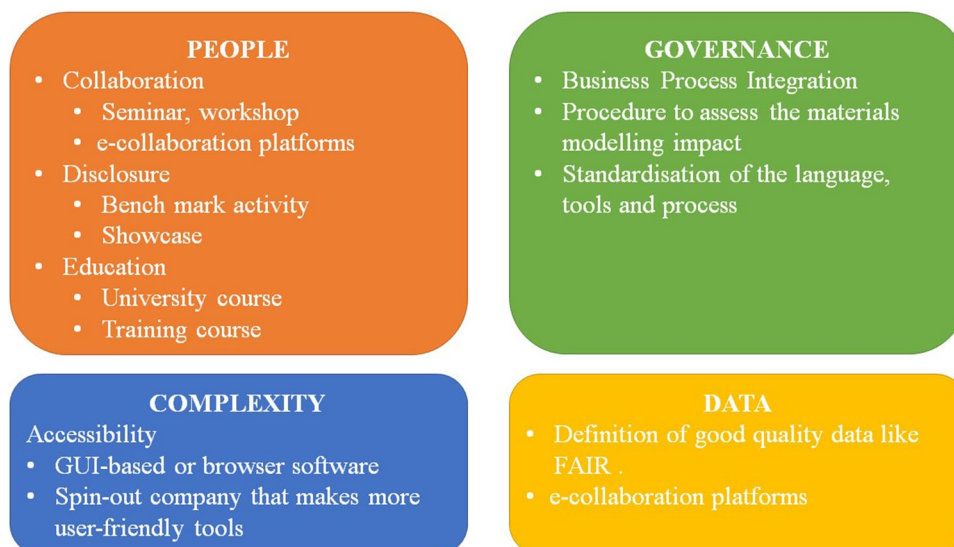
## Initiatives

Materials modeling is recognized as fundamental for the discovery and development of new materials and processes [90]. Many initiatives have been proposed over the years to tackle the barriers discussed in the previous section [3, 4, 21, 83, 87]. The main areas recognized to be essential to advocate for materials modeling in industry are collaboration, disclosure, education, accessibility, integration and data [3, 4, 21, 85, 87]. Collaboration, disclosure and education directly effect the people factor. The governance is affected by the integration, and complexity by accessibility (see Fig. 10).

## Collaboration

In term of collaboration, it has been recognized that cooperative work between academic model developers, industrial stakeholders, and across industrial departments and research modeling groups, is not only beneficial but essential to the development of tools and frameworks ready for industrial applications [87]. Materials modeling does require different skills, involving relatively large research groups (up to 20 or more people). Alternatively, collaboration between smaller groups can achieve the same outcomes. Promotion of collaborative work between modeling groups can help the creation of a framework that can span across length scales and purposes. In order to promote it, cooperation between universities is now a prerequisite to apply for funding. Other activities to support partnerships between universities are the

**Fig. 10** The main area of initiative are summarized in relation of the four factors: people (collaboration, disclosure, and education), governance (integration), complexity (accessibility), and data



organization of workshops and seminars, possibly extended to a wider group of researchers and industry. These can also help to find new applications for existing techniques, especially for the physics-based models, in other fields [87]. Research centers can have an important role on encouraging collaboration. It has been proposed that a one-day meeting between modeling and non-modeling experts (experimentalists and industrial users) organized by the Engineering and Physical Sciences Research Council (EPSRC) or a center of excellence would be beneficial to bridge this gap [87]. They can play a key role by creating international relationships with universities and organizations outside the UK. Collaboration is promoted by communities such as nanoHUB [91] which aims to gather together students, researchers and instructors working and teaching in the nanotechnology field around the world. The sharing of tools and teaching materials are encouraged in exchange of benefits including wider visibility, automatic data usage, possibility to discover new applications in other fields, the usage of HPC resources, citation tracking, and control of reuse and assets via licensing.

### Conflict of Interest

The modeling community needs to put major effort on disclosure and reaching out. Activities such as training, showcases, software benchmarking and consultancy can speed up the adoption of these technologies. Centers of excellence are points of aggregation for academics and industry, therefore, ideally placed to run these activities. At a European level, the European Materials Modelling Council [6] was created in 2019 aiming to overcome the barriers to industrial integration of materials modeling. The EMMC focus is divided into six areas: Model development, interoperability, digitalization, software, impact of industry, and policy. They produce survey studies, whitepapers, roadmaps, and run workshops to identify the barriers and suggest strategies. Another example is the Interdisciplinary Centre for Advanced Materials Simulation (ICAMS), in Germany, which is a research center at the Ruhr-Universität Bochum [92] focusing on the development and application of simulation tools for multi-scale materials modeling [92]. They actively contribute to simulation software development that can be found in their GitHub repositories [93]. The knowledge and know-how was transferred to two new enterprises, one of which offers support for microstructure modeling. They are active in industrial research, and they organize showcases of three of the six project groups every two years, to show the last development and get input from the industrial stakeholders. In France, the French National Centre for Scientific Research (Centre national de la Recherche Scientifique, CNRS) [94] is a state research organization, which coordinates different institutes. Each institute promotes cooperation between disciplines and oversees the research activities of laboratories

in the country such as the Institut Jean Lamour (IJL) [95], which focuses on materials science research including the use and development of in-house software. Five companies have been established as spin-outs from this research laboratory, two of which are modeling and simulation focused. The activities of benchmark exercises are also a good way to showcase the state-of-the-art technology and disclose the latest achievement in the community. The Additive Manufacturing Benchmark Test Series (AM-Bench) activities promoted by the American National Institute of Standards and Technology (NIST) [96] is a good example of benchmark exercises for additive manufacturing applications. A first benchmark exercise was run in 2018, the results of which were disclosed in a series of articles and at the conference organized by The Minerals, Metals & Materials Society (TMS), which saw the participation of academics; software companies such as Dassault Systèmes SIMULIA [97] and QuesTek Innovations LLC [98]; research centers such as the Advanced Manufacturing Research Center (AMRC) [99]; and industry such as Rolls-Royce [100] and Siemens [101]. The French national research agency (Agence Nationale de la Recherche, ANR) [102] also promotes benchmark exercises, for example on macrosegregation simulations during solidification run in 2010 [103], which involved different research centers in France including IJL. The aim of the project was to refine the processes of macrosegregation in the final product. At the time of publishing this review, there is no existing research center that specializes in materials modeling, because the decision was to create research centers focusing on specific processes [87]. Nowadays, most of these centers have materials modeling departments as they conduct projects involving modeling and simulations. In terms of SME engagement, there are projects founded by the European Regional Development Fund (EDRF) that aim to facilitate the access to materials modeling technologies for SMEs [104]. Other fundamental actors for materials modeling disclosure are communities such as nanoHUB [91] or the International Association for the Engineering Modeling, Analysis and Simulation Community (NAFEMS) [105]. They offer a vast selection of resources including software that can be run in the cloud, and learning materials such as videos of lectures and courses.

### Education

It is clear that industrial adoption of materials modeling requires the formation of a new generation of modelers, engineers, and materials scientists in this discipline. The importance of education in materials modeling was already recognized by Sheriff [87]. He recommended to develop appropriate degrees both at undergraduate and postgraduate levels, which foster an open-minded approach, ability to work in a team, and the clear idea that models are not the

final goal but useful tools to accomplish the aims of research, design or monitoring activities. In his final remarks, he indicated the EPSRC as the main actor in promoting and developing materials modeling education. Most undergraduate studies in materials science in the UK university offer some materials modeling course in their curricula. Most of these courses focus on continuum techniques, but with very few on discrete modeling tools. In terms of higher education, two Centers of Doctoral Training (CDT) have been founded by EPSRC, Theory and Simulation of Materials (CDT-TSM) at Oxford [106] and Computational Methods for Materials Science (CDT-CMMS) [107] in Cambridge. Both courses aim to generate materials scientists and researchers in the state-of-the-art of materials modeling, helping them to develop and select the right modeling tools for their research. The state of materials modeling in university teaching in the USA seems similar to the UK, as reported surveys conducted by Thorton et al. [108] and by Enrique et al. [109]. The surveys were conducted nine years apart between 2009 and 2018, proposing the same questions to a select group of American academics [110]. The comparison between answers show an increment of Computational Materials Science and Engineering (CMSE) courses and a clear support for including them in the core curriculum. In nine years, the number of universities offering one or two CMSE undergrad courses almost doubled, and 96% of institutions interviewed has at least one CMSE module in their graduate courses while it was 41% in 2009. The surveys also show that there is no strong preference for standalone courses or the introduction of specific modules in the existing materials science courses. Some interviewees indicated that barriers stopping further development were the lack of teaching resources such as graphics-based operating system interface (GUI) software, and preparation of the students in mathematics, physics, and computer science fundamentals. In Europe, the teaching includes some complete courses in materials modeling for graduate students [111, 112]. In addition, short training courses and seminars for industry are organized by universities [113], software companies [97] and associations [105]. Online resources are also available such as the nanoHUB catalog of courses [91] or other sources [114]. These are fundamental activities for promoting a wider use of materials modeling tools outside the modeling community.

### Accessibility

The complexity characterizing these technologies is also a technical barrier that prevents users from taking advantage continuously and efficiently applying these tools. Most of the people working in industry today do not have a very strong background in modeling, especially in discrete techniques, program languages and operating systems. The EMMC suggested that the community of modeling experts, the software

developers, and the funding agencies should focus on transforming a promising tool into a technology that can be easily adopted throughout the industrial stakeholders [3]. The development of GUI-based or browser software is a fundamental process in the industrial distribution of these methodologies. According to the EMMC, the technology transfer and software development do not need to be performed by university, actually small companies and research institutions are better intermediate entities between university and industry [87]. While user friendly software should not be one of the objectives of academic work, more effort should be dedicated to write clear documentation of the code developed, to make the conversion easier for software developers, and help to set the training of users [3]. Different university spin-out companies have created web-based interfaces that simplify the use of techniques such as CALPHAD [115] and MD [3]. The use of an open source web app framework, such as the Django project [116], Flask [117], and Pyramid [118] that allows running remote applications using a web interface should be also investigated.

### Integration

Even when the technology is accessible to industry, it is still not widely used. As shown in the surveys of Goldbeck and Simperler [85] materials modeling is not well integrated in the industrial processes or workflows. The authors believe that this is a sign of the fact that materials modeling is not *de facto* recognized as value in the organization. Work should be done in promoting these technologies, emphasizing the benefits, efficiency and effectiveness. As already mentioned, benchmark activities, showcases, seminars and workshops organized by research centers can help show evidence of the benefits coming from the use of materials modeling. Internally to the organization, activities such as Business Processes analysis and Quantitative Benchmark for Time to Market framework (QBTM) [119] should be performed to show the impact of materials modeling [85]. Materials modeling and simulation need to be part of Business Process Integration (BPI) Paradigm. BPI aim is to connect systems and information efficiently allowing the automation of business processes, integration of systems and services, and the secure sharing of data across numerous applications internally and externally [120]. Belouettar and co-workers [5] have proposed an integrated multi-disciplinary, multi-model and multi-field approach. The tool developed integrates materials modeling with business tools and database in a single open platform, connecting a rich materials modeling layer with industry standard business process models. In terms of modeling governance, the non-government organization, Modelica Association, aims to develop and promote a standard language for modeling, simulation and programming of physical and technical systems and

processes [121]. The standards include an object-oriented, equation based language (Modelica Language) to describe cyber-physical systems, zip-file (Functional Mock-Up Interface) container for the description, exchange and storage of simulation input and outputs, a workflow (System Structure and Parameterization) that describes the logical connections between the different models and how data are stored and exchanged, and a communication protocol (Distributed Co-Simulation Protocol) for the integration of models and real-time system into simulation environments. This organization includes members from around the world, from academia, industry, and software developer companies. This project started in 1996, and it has promoted the developing of a series of open source and commercial tools such as simulation environments, developer, and data visualization, managing and storage. In industry, the Modelica language and libraries are used in automotive industries and by plant power providers.

## Data

The availability and sharing of data are crucial to accelerate the development of new materials, optimize manufacturing processes, and validate theory, but also assists to identify gaps and avoid redundancy [21]. The creation of a MDI is explicitly indicated as a strategic challenge in a Materials Genome Initiative white paper [122]. The MDI should be easy to access and use, with high-quality standardized datasets [4, 21]. Quick and easy access should be guaranteed by online accessibility so the materials science and engineering community can take advantage of these data. Standards need to be developed for format, metadata, data types, data quality criteria, data management, data stewardship, and protocols necessary for interoperability and seamless data transfer. The goal is to extract the maximum benefit from a set of data by reusing them for different applications and research. This, however, requires a data quality metric to be defined. Four principles have been recognized as fundamental to guide toward a good-quality and open data: Findability, Accessibility, Interoperability, and Reusability (FAIR) [123]. Findability principally concerns the fact that data should be associated with a globally unique and persistent identifier, contain accurate and relevant attributes (metadata), and be registered in a searchable resource. Accessibility concerns an open, free, and universal communication protocol for identification authentication and authorization procedures enabling access to metadata. Interoperability regards the language for knowledge representation of the data and metadata. Reusability aims to assure that the relevant attributes are described in the metadata in order to specify the data usage license, detailed data provenance, and that they meet the domain-relevant community standards. These four principles were recognized and indicated by a G20 statement,

"G20 2016 Innovation Action Plan," as a guideline to promote open science and facilitate appropriate access to publicly funded research results [124]. The NASA Vision 2040 also indicates FAIR principles as important emerging paradigm for the management of technical data [4] and in the creation of MDIs [89].

Three fundamental digital components are recognized at the core of MDI: repositories, tools, and e-collaboration platforms [125]. Repositories are not simple databases but they need to provide an Application Programming Interface (API) to facilitate query and data extraction [89]. The materials science and engineering community should *in primis* try to define policies for capturing data, and how to incorporate them into existing workflows, or develop a new one [21]. It has been identified that national agencies such as NASA in the USA should coordinate these efforts, and promote the production, maintenance and dissemination of high-quality datasets by establishing clear incentives for the community [4]. The Thermodynamics Research Center by NIST [126] and MatNAVI [127] by the Japanese National Institute of Materials Science (NIMS) [128] are good examples of materials databases created and maintained by government entities. Other examples outside the government institutions are the Cambridge Crystallographic Data Centre (CCDC) database for organic and metal-organic compounds studied by X-ray and neutron diffraction [129], and ASMI Online Databases [130]. These databases were available for decades and they collect a small volume of high-quality data in each submission, which can be hand-curated. Although these have been a valuable source of information, in order to respond to FAIR principles, new repository infrastructures are required to be able to manage a higher volume of data, be more accessible to both expert and non-expert users, have APIs, searchable tools, and allow easy tracking of data [89]. There are different initiatives supported by government institutions such as Materials Commons [131], and independent ones such as Zenodo [132] that offer these capabilities. In terms of tools, NIST has developed the Materials Resource Registry (MRR) [133] that allows an organization to make their resource findable by registering metadata of data collections, application programming interfaces, informational sites, and materials software, like the Virtual Observatory (VO) does for astronomy data. However, it is better to underline that materials are characterized by variability and their data are, therefore, very specific, while astronomic information are obtained pointing the telescope at the same object in space. Therefore, a VO solution in materials may not be achievable without considering the intrinsic uncertain nature of the materials. In terms of accessibility, interoperability, and reusability, tools may be needed to be developed for specific sub-fields of materials science such as the Crystallographic Information File (CIF) format [134] used by the crystallographic community. The Materials Data Curation

System (MDCS) [135] by NIST aims to provide a more general tool to manage and share data that uses one structured format: Extensible Markup Language (XML). The main issue with materials data remains the formulation of standards for sharing data. To mitigate this, the Office for Data and Informatics was created by NIST in 2014 to provide leadership and expertise from other communities to meet the challenge in modern data-driven research [136]. Ultimately, e-collaboration platforms serve as generation of data. These platforms use web-based technologies to manage materials data by implementing the FAIR principles. The nanoHUB [91] is an example of such platforms. These, however, do not solve the problem of the prevalence of proprietary data formats used in scientific instrumentation that makes extracting experimental data and the corresponding metadata challenging [89].

### Outlook: Material Informatics Approach for More Sustainable Manufacturing

The today task is to achieve a green economy, putting materials in the center of manufacturing processes and life product cycle. Recent introduction of new regulation to tackle climate change [137], depletion of critical natural resources [138], and social pressure are pushing industry to reduce their environment impact focusing on reducing CO<sub>2</sub> emissions while increasing energy efficiency of manufacturing process, producing sustainable products with responsible sourcing of raw materials [139]. Identification of innovative materials for clean energy generation, optimization of processes to improve efficiency, and establishing recycling route are the key strategies for the transition to green economy. The demand of more environmentally-friendly solutions is not new [140], but the adoption of digital solutions for materials selection, process optimization and control can be critical to accelerate the transition. With the classical approaches, material discovery and optimization can take three to four times the time-to-market of a new product [141]. New technologies such as Internet of Things (IoT), artificial intelligence (AI), Big Data, blockchain, digital twins, and digital threads are going to change the product life within supply chain, production, service and end-of life to satisfy the need for more sustainability not just in terms of environment but also from economic and social perspectives. The need for a material-centric industry requires these tools to be supported by material models. In particular, the combination of the three typologies - physic-based, data driven and Big Data - offer the creation of digital instruments to rapidly explore materials and manufacturing technologies with an understating of the structure-process–property relationship [142]. Digital twins with integrated physic-based models will enable exploration of what-if scenarios with

behavior prediction. Data-driven methods trained using enriched datasets that comprises experimental data and modeling predictions can be used in digital threading tools that connect sensor data from the machine and digital data to perform machine control during production to assure quality regardless the uncertainty of the process conditions. Materials informatics technologies will also allow a better integration of material modeling with other technology such as business modeling tools and engineering design. However, it is fundamental that these tools are robust, necessitating the application of rigorous processes, such as V & V [9] and TML [78].

### Conclusions

In summary, modeling techniques and digitalization still struggle to be accepted as valuable resources for industry applications by all stakeholders and not just by modelers and theoretical scientists. The barriers are technology, cultural and infrastructural. The complexity of these methods and the absence of processes prevents overcoming the first technological gap between developer (usually academic) and industrial users, usually only interacting reactively to an urgent issue. Reluctance and cultural barriers have a significant impact on embracing modeling and digitalization as a business value. This results in difficulty to find and retain skilled labor, and stops the integration of this technology in the workflow and standardization in the industrial setting. The absence of infrastructure such as access to HPC and appropriate databases also stops a full and systematic acquisition of models for SMEs as well as for large companies. However, the drive to a green economy requires industry to put materials at the center of their process and product designs to assure clean energy, recycling, efficiency and responsible sourcing. Digitalization and materials informatics are going to play a key role in the transition, and they will heavily depend on material modeling technologies that comprise physics knowledge and fast computing. This requires using physics-based simulation to train data-driven modeling. In addition, digital threading will enable integrated material modeling in Model based Engineering frameworks, in order to share data from and to the physical world, but also from/to other model technology, i.e., business modeling tools.

In terms of perspectives, six actions areas were identified and discussed having the major impact on enhancing industrial awareness: collaboration, disclosure, education, accessibility, integration, and data.

- Collaboration is vital for the development of modeling tools but also to increase their visibility by sharing tools and materials.

- Conflict of interest activities such as show cases, benchmark and consultancy also are important to expand modeling awareness.
- Education plays a key role in changing the cultural approach of the next generation of scientists and engineers.
- On the technical side, the adoption of modeling requires an increase in the accessibility by reducing the complexity of these methods working on the development of GUI-based or browser software, and the integration in the industrial process.
- Data availability is important not only for modeling exploitation but to accelerate materials developing, and assure industrial competitiveness.

Materials modeling is a well-developed research area that has shown great benefit when applied reactively to industrial challenges. By introducing modeling proactively into the design process, real benefit in the quality, cost, and time of manufacturing can be realized.

**Acknowledgements** This work was supported by UKRI Future Leaders Fellowship [grant number MR/T02058X/1]. The authors are also grateful for the comments from Dr. Fabian Sorce (Imperial College London), and Dr. Chaitanya Paramatmuni (University of Sheffield).

**Author contributions** L.S., D. C. conceived and coordinated the review. L.S. did the literature review and collected data. All authors contributed to the manuscript.

## Declarations

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Materials Genome Initiative for Global Competitiveness (2011) Executive office of the president. National Science and Technology Council
2. National Research Council (2008) Integrated computational materials engineering: a transformational discipline for improved competitiveness and national security. National Academies Press, Washington
3. EMMC (2018) The EMMC roadmap for materials modelling and informatics. <https://emmc.eu/emmc-roadmaps/>
4. Liu X, Furrer D, Kusters J, Holmes J (2018) Vision 2040: a roadmap for integrated, multiscale modeling and simulation of materials and systems NASA report NASA/CR—2018–219771
5. Belouettar S, Kavka C, Patzak B, Koelman H, Rauchs G, Giunta G, Madeo A, Pricl S, Daouadji A (2018) Integration of material and process modelling in a business decision support system: case of COMPOSELECTOR H2020 project. *Compos Struct* 204:778–790
6. European Materials Modelling Council (EMMC). <https://emmc.eu/>
7. Sargent RG (1981) An assessment procedure and a set of criteria for use in the evaluation of 'computerized models and computer-based modelling tools' Syracuse University Department of Engineering and Operations Research report ADA098785
8. Roache PJ (2002) Code verification by the method of manufactured solutions. *J Fluids Eng* 124(1):4–10
9. ASME (2020) Standard for verification and validation in computational solid mechanics. US Patent 3,120,606
10. Rasheed A, San O, Kvamsdal T (2020) Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access* 8:21980–22012
11. Grong O, Shercliff HR (2002) Microstructural modelling in metals processing. *Prog Mater Sci* 47(2):163–282
12. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow: large-scale machine learning on heterogeneous systems. Software available from <https://www.tensorflow.org/>
13. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
14. Ashoor BB, Giwa A, Hasan SW (2019) Full-scale membrane distillation systems and performance improvement through modeling: a review. *Curr Trends Future Dev Bio Membr* 2019:105–140
15. Cugnet M, Dubarry M, Liaw BY (2009) Secondary batteries—lead–acid systems—modeling. In: Garche J (ed) *Encyclopedia of Electrochemical Power Sources*. Elsevier, Amsterdam, pp 816–828
16. Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707
17. Sun L, Gao H, Pan S, Wang J-X (2020) Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Comput Methods Appl Mech Eng* 361:112732
18. Smith WF, Hashemi J (2006) *Foundations of materials science and engineering*. McGraw-Hill Publishing, Sydney
19. Lahoz BKW, Menard R (2010) *Data assimilation*. Springer, Berlin
20. White AA (2013) Big data are shaping the future of materials science. *Mater Res Soc Bull* 38(8):594–595
21. Materials Genome Initiative. <https://www.mgi.gov/>
22. Himanen L, Geurts A, Foster AS, Rinke P (2019) Data-driven materials science: status, challenges, and perspectives. *Adv Sci* 6(21):1900808
23. Takamizawa H, Nishiyama Y, Hirano T (2020) Bayesian uncertainty evaluation of Charpy ductile-to-brittle transition

- temperature for reactor pressure vessel steels. In: Pressure vessels and piping conference, vol 83815. American Society of Mechanical Engineers, pp 001–01065
24. Dyson BF (1988) Creep and fracture of metals: mechanisms and mechanics. *Revue Phys Appl* 23(4):605–613
  25. Beckermann C (2002) Modeling of solidification. *Purdue Heat Transfer Celebration*. April, 3–5
  26. Van Der Giessen E, Schultz PA, Bertin N, Bulatov VV, Cai W, Csányi G, Foiles SM, Geers MG, González C, Hütter M et al (2020) Roadmap on multiscale materials modeling. *Modell Simul Mater Sci Eng* 28(4):043001
  27. Geers MGD, Kouznetsova VG, Matouš K, Yvonnet J (2017) Homogenization methods and multiscale modeling: nonlinear problems. In: *Encyclopedia of computational mechanics*, 2nd edn, (eds E. Stein, R. Borst and T.J.R. Hughes). <https://doi.org/10.1002/9781119176817.ecm2107>
  28. Mori T, Tanaka K (1973) Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metall* 21(5):571–574
  29. Dyson BF (2009) Microstructure based creep constitutive model for precipitation strengthened alloys: theory and application. *Mater Sci Technol* 25(2):213–220
  30. Tong Q, Li S (2020) A concurrent multiscale study of dynamic fracture. *Comput Methods Appl Mech Eng* 366:113075
  31. Tadmor EB, Ortiz M, Phillips R (1996) Quasicontinuum analysis of defects in solids. *Philos Mag A* 73(6):1529–1563
  32. Horstemeyer MF (2009) Multiscale modeling: a review. In: Leszczynski, J., Shukla, M. (eds) *Practical aspects of computational chemistry*. Springer, Dordrecht. [https://doi.org/10.1007/978-90-481-2687-3\\_4](https://doi.org/10.1007/978-90-481-2687-3_4)
  33. Vassaux M, Sinclair RC, Richardson RA, Suter JL, Coveney PV (2020) Toward high fidelity materials property prediction from multiscale modeling and simulation. *Adv Theory Simul* 3(1):1900122
  34. Panchal JH (2005) A framework for simulation-based integrated design of multiscale products and design processes. Georgia Institute of Technology
  35. Weinan E, Engquist B (2003) Multiscale modeling and computation. *AMS* 50(9):1062–1070
  36. Khaleel M (2004) Multiscale mathematics initiative: a roadmap. US Department of Energy
  37. Olson GB (1997) Computational design of hierarchically structured materials. *Science* 277(5330):1237–1242
  38. de Baas AF (2017) What makes a material function? Let me compute the ways: modelling in H2020 LEIT-NMBP programme materials and nanotechnology projects
  39. Schrödinger E (1926) An undulatory theory of the mechanics of atoms and molecules. *Phys Rev* 28(6):1049
  40. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev* 140:1133–1138
  41. Anderson PW (1961) Localized magnetic states in metals. *Phys Rev* 124(1):41
  42. Gavini V, Knap J, Bhattacharya K, Ortiz M (2007) Non-periodic finite-element formulation of orbital-free density functional theory. *J Mech Phys Solids* 55(4):669–696
  43. Radhakrishnan BG, Gavini V (2013) Electronic structure calculations at macroscopic scales using orbital-free DFT. World Scientific, Singapore
  44. Svejstrup W, Maiani A, Van Hoogdalem K, Flensburg K (2023) Orbital-free approach for large-scale electrostatic simulations of quantum nanoelectronics devices. *Semicond Sci Technol* 38(4):045004
  45. Balian R, Bloch C (1974) Solution of the Schrödinger equation in terms of classical paths. *Ann Phys* 85(2):514–545
  46. Kristinsdóttir L, Skúlason E (2012) A systematic DFT study of hydrogen diffusion on transition metal surfaces. *Surf Sci* 606(17–18):1400–1404
  47. Zhao S, Liu X-W, Huo C-F, Li Y-W, Wang J, Jiao H (2015) Determining surface structure and stability of  $\epsilon$ -Fe<sub>2</sub>C,  $\chi$ -Fe<sub>5</sub>C<sub>2</sub>,  $\theta$ -Fe<sub>3</sub>C and Fe<sub>4</sub>C phases under carburization environment from combined DFT and atomistic thermodynamic studies. *Catal Struct React* 1(1):44–60
  48. Wu HH, Wisesa P, Trinkle DR (2016) Oxygen diffusion in hcp metals from first principles. *Phys Rev B* 94(1):014307
  49. Fomine S, Vargas SM, Tlenkopatchev MA (2003) Molecular modeling of ruthenium alkylidene mediated olefin metathesis reactions. DFT study of reaction pathways. *Organometallics* 22(1):93–99
  50. Ghazisaeidi M, Hector LG Jr, Curtin W (2014) First-principles core structures of  $\langle c + a \rangle$  edge and screw dislocations in Mg. *Scr Mater* 75:42–45
  51. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092
  52. Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, Uberuaga BP (eds) *Radiation effects in solids*. Springer, Dordrecht, pp 1–23
  53. Rahman A (1964) Correlations in the motion of atoms in liquid argon. *Phys Rev* 136(2A):405
  54. Herschberg R, Fu C-C, Nastar M, Soisson F (2019) Atomistic modelling of the diffusion of C in FeCr alloys. *Acta Mater* 165:638–653
  55. Ma B, Guye K, Dogruoz B, Agonafer D (2021) Molecular dynamics simulations of thin-film evaporation: the influence of interfacial thermal resistance on a graphene-coated heated silicon substrate. *Appl Therm Eng* 195:117142
  56. Hoyt JJ, Asta M, Karma A (2003) Atomistic and continuum modeling of dendritic solidification. *Mater Sci Eng R Rep* 41(6):121–163
  57. Becker C, Olmsted D, Asta M, Hoyt J, Foiles S (2007) Atomistic underpinnings for orientation selection in alloy dendritic growth. *Phys Rev Lett* 98(12):125701
  58. Sangid MD, Ezaz T, Sehitoglu H, Robertson IM (2011) Energy of slip transmission and nucleation at grain boundaries. *Acta Mater* 59(1):283–296
  59. Zhou XW, Sills RB, Ward DK, Karnesky RA (2017) Atomistic calculations of dislocation core energy in aluminium. *Phys Rev B* 95(5):054112
  60. Frolov T, Asta M, Mishin Y (2016) Phase transformations at interfaces: observations from atomistic modeling. *Curr Opin Solid State Mater Sci* 20(5):308–315
  61. Fix GJ (1982) Phase field methods for free boundary problems. Carnegie Mellon University, Pittsburgh
  62. Neumann J, Burks AW (1966) *Theory of self-reproducing automata*. University of Illinois Press, Urbana
  63. Sills RB, Kuykendall WP, Aghaei A, Cai W (2016) Fundamentals of dislocation dynamics simulations, pp 53–87
  64. McKay MD, Beckman RJ, Conover WJ (2000) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42(1):55–61
  65. Eglajs V, Audze P (1977) New approach to the design of multifactor experiments. *Probl. Dyn. Strengths* 35(1):104–107
  66. Hammersley JM, Handscomb DC (1964) *General principles of the Monte Carlo method*. Springer, Dordrecht, pp 50–75
  67. Karhunen K (1947) Ueber lineare methoden in der wahrscheinlichkeitsrechnung. *Soumalainen Tiedeakatemia*
  68. Loève M (1955) *MM, probability theory*. Van Nostrand Company, Princeton
  69. Whittle P (1951) *Hypothesis testing in time series analysis*. Almqvist & Wiksells Boktr, Stockholm

70. Tapia G, King W, Johnson L, Arroyave R, Karaman I, Elwany A (2018) Uncertainty propagation analysis of computational models in laser powder bed fusion additive manufacturing using polynomial chaos expansions. *J Manuf Sci Eng* 140(12):121006
71. Golub GH, Reinsch C (1971) Singular value decomposition and least squares solutions. Springer, Berlin, Heidelberg, pp 134–151
72. Xie Z, Jiang W, Wang C, Wu X (2022) Bayesian inverse uncertainty quantification of a MOOSE-based melt pool model for additive manufacturing using experimental data. *Ann Nucl Energy* 165:108782
73. Cressie N (1990) The origins of kriging. *Math Geol* 22(3):239–252
74. Ghanem R, Spanos PD (1990) Polynomial chaos in stochastic finite elements. *J Appl Mech* 57(1):197–202
75. Xiu D, Karniadakis GE (2002) The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 24(2):619–644
76. Tzinis I Technology readiness level. [https://www.nasa.gov/directorates/heo/scan/engineering/technology/technology\\_readiness\\_level](https://www.nasa.gov/directorates/heo/scan/engineering/technology/technology_readiness_level)
77. McLelland AN (2013) Simulation and experimental predictions for the structural response of satellites. PhD thesis
78. Cowles B, Backman D, Dutton R (2012) Verification and validation of ICME methods and models for aerospace applications. *Integr Mater Manuf Innov* 1:3–18
79. Allison J, Li M, Wolverton C, Su X (2006) Virtual aluminum castings: an industrial application of ICME. *JOM* 58(11):28–35
80. Anderson MJ, Panwisawas C, Sovani Y, Turner RP, Brooks JW, Basoalto HC (2018) Mean-field modelling of the intermetallic precipitate phases during heat treatment and additive manufacture of Inconel 718. *Acta Mater* 156:432–445
81. Simperler A, Goldbeck G EMMC case study: optimisation of manufacturing conditions for gas turbine rotating components. [https://emmc.eu/wp-content/uploads/2022/02/EMMC\\_RR\\_CASESTUDY.pdf](https://emmc.eu/wp-content/uploads/2022/02/EMMC_RR_CASESTUDY.pdf)
82. Goldbeck G, Court C (2016) The economic impact of materials modelling. Zenodo
83. Goldbeck G, Simperler A (2020) Materials modelling software market. Zenodo
84. Goldbeck G (2012) The economic impact of molecular modelling. Zenodo
85. Goldbeck G, Simperler A (2019) Strategies for industry to engage in materials modelling. Zenodo
86. Paulk MC, Curtis B, Chrissis MB, Weber CV (1993) Capability maturity model, version 1.1. *IEEE Softw* 10(4):18–27
87. Shercliff HR (1997) Modelling of materials and processes. Department of engineering, University of Cambridge, technical report CUED/CMATS/TR 243
88. Goldbeck G, Simperler A Business models and sustainability for materials modelling software
89. Warren J.A, Ward C.H (2018) Evolution of a materials data infrastructure. *JOM* 70(9):1652–1658
90. Materials engineering—metals and alloys. <https://epsrc.ukri.org/research/ourportfolio/researchareas/matengmetal/>
91. nanoHUB. <https://nanohub.org/>
92. Interdisciplinary centre for advanced materials simulation (ICAMS). <http://www.icams.de/content/>
93. github-interdisciplinary centre for advanced materials simulation (ICAMS). <https://github.com/ICAMS>
94. Centre national de la recherche scientifique (CNRS). <https://www.cnrs.fr/en>
95. Institut jean lamour (IJL). <https://ijl.univ-lorraine.fr/>
96. NIST, National institute of standards and technology. <https://www.nist.gov/>
97. Dassault systèmes. <https://www.3ds.com/>
98. QUESTEK. <https://www.questek.com/>
99. The university of Sheffield. Additive Manufacturing Research Centre. <https://www.amrc.co.uk/>
100. Rolls-Royce plc. <https://www.rolls-royce.com/>
101. SIEMENS. <https://www.siemens.com/global/en.html>
102. Agence nationale de la recherche (ANR). <https://anr.fr/en/>
103. Collaborative Workshop on Experimental and Numerical Simulation of MACroseggregation and Structure—SMACS (2010) University of Lorraine <https://benchmark-solidification.ijl.univ-lorraine.fr/>
104. AMCASH. <https://www.amcash.co.uk/>
105. NAFEMS. <https://www.nafems.org/>
106. Imperial College London—centre for doctoral training on theory and simulation of materials. London, UK. <https://www.imperial.ac.uk/theory-and-simulation-of-materials/>
107. University of Cambridge—Centre for doctoral training in computational methods for materials science, Cambridge, UK. <https://www.csc.cam.ac.uk/academic/cdtcompmat>
108. Thornton K, Nola S, Garcia R.E, Asta M, Olson G.B (2009) Computational materials science and engineering education: a survey of trends and needs. *JOM* 61(10):12–17
109. Enrique R.A, Asta M, Thornton K (2018) Computational materials science and engineering education: an updated survey of trends and needs. *JOM* 70(9):1644–1651
110. Thornton K, Asta M (2005) Current status and outlook of computational materials science education in the US. *Modell Simul Mater Sci Eng* 13(2):53
111. Malmö University—computational materials science, Master’s programme (two-year), Malmö, Sweden. <https://mau.se/en/study-education/programme/tacma>
112. Università di Genova—continuous and discrete modelling and simulation, Genoa, Italy. <http://www.itim.unige.it/cs/strategos/edu/continuousdiscrete/>
113. Cranfield University—finite element analysis and materials modelling, Cranfield, UK. <https://www.cranfield.ac.uk/courses/short/manufacturing/finite-element-analysis-and-materials-modelling>
114. ANSYS/GRANTA education HUB. <http://www.grantadesign.com/education/>
115. Kaufman L, Bernstein H (1970) Calculation of regular solution phase diagrams for titanium base binary systems. In: Jaffee RI, Promisel NE (eds) *The science, technology and application of titanium*. Pergamon, London, pp 361–372
116. django project. <https://www.djangoproject.com/>
117. Flask. <https://flask.palletsprojects.com/en/2.0.x/>
118. Pyramid. <https://trypyramid.com/>
119. Quantitative Benchmark for Time to Market (QBTM) for new materials innovation: an analytical framework (2016) National Institute of Standards and Technology, Gaithersburg, MD. [https://www.nist.gov/system/files/documents/2018/06/26/qbtm\\_final\\_analytical\\_framework\\_011216.pdf](https://www.nist.gov/system/files/documents/2018/06/26/qbtm_final_analytical_framework_011216.pdf)
120. MuleSoft: business process integration: creating connectivity (2023) <https://www.mulesoft.com/resources/esb/business-process-integration>
121. The Modelica association. <https://modelica.org/index.html>
122. Materials genome initiative strategic plan, The White House, Washington, DC (2014). [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/mgi\\_strategic\\_plan\\_-\\_dec\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/mgi_strategic_plan_-_dec_2014.pdf)
123. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3(1):1–9
124. G20 Leaders’ Communiqué Hangzhou Summit (2016). Hangzhou, China. [https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT\\_16\\_2967](https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967)



125. Metals TM (2017) Building a materials data infrastructure: opening new pathways to discovery and innovation in science and engineering. TMS, Pittsburgh
126. Thermodynamics research center (TRC). <https://trc.nist.gov/>
127. NIMS materials database (MatNavi). <https://www.nims.go.jp/eng/index.html>
128. National institute of materials science (NIMS). <https://www.nims.go.jp/eng/index.html>
129. The Cambridge crystallographic data centre (CCDC). <https://www.ccdc.cam.ac.uk/>
130. Online databases—ASM International. <https://www.asminternational.org/materials-resources/online-databases>
131. Materials commons 2.0. <https://materialscommons.org/>
132. Zenodo. <https://zenodo.org/>
133. Materials resource registry. <https://materials.registry.nist.gov/>
134. Hall SR, Allen FH, Brown ID (1991) The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallogr A* 47(6):655–685
135. NIST-materials data curation system. [https://www.nist.gov/progr\\_ams-projects/materials-data-curation-system](https://www.nist.gov/progr_ams-projects/materials-data-curation-system)
136. NIST-office for data and informatics. <https://www.nist.gov/mml/odi>
137. Commission E (2020) Proposal for a regulation of the European parliament and of the council establishing the framework for achieving climate neutrality and amending regulation (EU) 2018/1999 (European climate law). European Commission (EC), Brussels
138. Northey SA, Mudd GM, Werner TT (2018) Unresolved complexity in assessments of mineral resource depletion and availability. *Nat Resour Res* 27:241–255
139. Melia HR, Muckley ES, Saal JE (2021) Materials informatics and sustainability: the case for urgency. *Data Centric Eng* 2:19
140. United Nations: UN (1992) Agenda 21. United Nations Conference on Environment & Development. United Nations, Rio de Janeiro
141. Ramakrishna S, Zhang T-Y, Lu W-C, Qian Q, Low JSC, Yune JHR, Tan DZL, Bressan S, Sanvito S, Kalidindi SR (2019) Materials informatics. *J Intell Manuf* 30:2307–2326
142. Hara K, Yamada S, Kurotani A, Chikayama E, Kikuchi J (2022) Materials informatics approach using domain modelling for exploring structure-property relationships of polymers. *Sci Rep* 12(1):10558