



Review



Cite this article: Kimpton LM, Paun LM, Colebank MJ, Volodina V. 2025 Challenges and opportunities in uncertainty quantification for healthcare and biological systems. *Phil. Trans. R. Soc. A* **383**: 20240232.

<https://doi.org/10.1098/rsta.2024.0232>

Received: 24 August 2024

Accepted: 13 November 2024

One contribution of 13 to a theme issue
'Uncertainty quantification for healthcare and
biological systems (Part 1)'.

Subject Areas:

biomedical engineering, applied mathematics,
statistics, artificial intelligence, computer
modelling and simulation, mathematical
modelling

Keywords:

uncertainty quantification, biology and
healthcare, clinical decision support,
mechanistic models, digital twins

Author for correspondence:

Victoria Volodina

e-mail: v.volodina@exeter.ac.uk

[†]All authors contributed equally to the study.

Challenges and opportunities in uncertainty quantification for healthcare and biological systems

Louise M. Kimpton^{1,†}, L. Mihaela Paun^{2,3,†},
Mitchel J. Colebank^{4,5,†} and Victoria Volodina^{1,†}

¹Department of Mathematics and Statistics, University of Exeter, Exeter, UK

²School of Mathematics and Statistics, University of Glasgow, Glasgow, UK

³School of Mathematical Sciences, University of Southampton, Southampton, UK

⁴Department of Biomedical Engineering, Edwards Lifesciences Foundation
Cardiovascular Innovation and Research Center, University of California, Irvine,
CA, USA

⁵Department of Mathematics, University of South Carolina, Columbia, SC, USA

LMK, 0000-0002-4265-5086; LMP, 0000-0002-8734-8135;
MJC, 0000-0002-2294-9124; VV, 0000-0003-4490-8777

Uncertainty quantification (UQ) is an essential aspect of computational modelling and statistical prediction. Multiple applications, including geophysics, climate science and aerospace engineering, incorporate UQ in the development and translation of new technologies. In contrast, the application of UQ to biological and healthcare models is understudied and suffers from several critical knowledge gaps. In an era of personalized medicine, patient-specific modelling, and *digital twins*, a lack of UQ understanding and appropriate implementation of UQ methodology limits the success of modelling and simulation in a clinical setting. The main contribution of our review article is to emphasize the importance and current deficiencies of UQ in the development of computational frameworks for healthcare and biological systems. As the introduction to the special issue on this topic, we provide an overview of UQ methodologies, their applications in non-biological and biological systems and the current gaps and opportunities for UQ development, as later highlighted by authors publishing in the special issue.

1. Motivation

Computer modelling, simulation and data analysis are recognized tools for understanding our complex physical world. These tools are used in applications such as climate science [1–3] and aerospace [4], though their application in biological and healthcare systems is still in its infancy [5–7]. The COVID-19 pandemic highlighted the benefits, and shortcomings, of predictive modelling in biological and healthcare applications [8]. Numerous computational models quantified and predicted outcomes during the pandemic; however, modelling efforts were not entirely successful, in part owing to the novel strategies for data collection and the unprecedented speed at which the virus spread. In addition, simulation-based decision-making was cross-examined by the general public, as many research articles (in some cases up to 94% of all COVID modelling studies [9]) lacked some sort of confidence or credible interval assessment in the model predictions for the pandemic response. As noted by Jit *et al.* and James *et al.* [8,10], these efforts caused some individuals and groups to cast doubt on the role of pandemic modelling and simulation for planning purposes. This intersection of mathematical modelling, epidemiological expertise, statistical analysis and policy implementation was a spotlight on the topic of UQ within biology and healthcare.

There are fewer applications of UQ in biological and healthcare models for several reasons. First, models describing biological or healthcare-related concepts often lack foundational model infrastructure, such as momentum balance or energy conservation laws commonly used in physics. As a consequence, and as discussed later, models in biology and healthcare are highly susceptible to model misspecification [11,12], and handling of this model uncertainty typically requires sophisticated UQ techniques. A second reason is the heterogeneity of healthcare and biological data [13]. In particular, integrating data of different modalities, e.g. electronic health records, imaging data, as well as clinical data into the UQ framework, is a challenging task, and thus performing holistic UQ to account for as many sources of uncertainty as possible has not been pursued so far. Last, much of the research work in the mechanistic, biological modelling field is typically carried out in the absence of statistical experts (e.g. only in teams of engineers, mathematicians and domain scientists). Although most of these studies are typically focused on developing the models themselves, the lack of statistical input again limits the depth of UQ that may be available for non-statistical (or non-UQ) experts.

Inspired by this, a Lorentz Center workshop on 'Uncertainty Quantification for Healthcare and Biological Systems' was held in April 2023. A group of junior, mid-career and senior scientists across biology, statistics, mathematics, engineering and computer science met to discuss the importance, implementation, gaps and opportunities for UQ in healthcare and biological models. The major outcome of this workshop included a collection of challenges for UQ in biological and healthcare applications, with limitations in communication between modelling and simulation experts, statisticians and domain scientists being at the forefront. To promote UQ in this sector, we have constructed a special issue collection of papers focused on this critical area.

In this introduction to the special issue, we provide an overview of computational modelling and simulation in healthcare and biology, and the key gaps in UQ within this field. We introduce modelling and simulation in established fields (e.g. engineering and physical sciences), as well as newer healthcare applications, and provide definitions and concepts in the field of UQ. Next, we present an extensive summary of gaps and opportunities in UQ in biology and healthcare modelling. We then introduce the special issue, which includes a collection of

articles at the interface of modelling and simulation, healthcare and biology, data integration and machine learning (ML), and UQ.

2. Modelling, Simulation and *Digital Twins*

Mathematical and computational models provide an abstract representation of physical systems of interest, enabling advanced analysis, simulation, optimization and prediction. This leads to significant improvements in performance and efficiency in domain applications. Throughout the main text, we refer to these first-principal-based models as **mechanistic models** [14,15].

With the increase in computer power and the availability of data, terms such as mathematical modelling, simulation and digital twins are becoming increasingly common. Although in many applications they are used interchangeably, there are still communities that have their own precise definitions (e.g. [16]). While modelling creates mathematical representations of a system, simulation refers to applying this model to conduct numerical experiments and analyse results. Simulation uses computational methods to solve model equations and visualize results. We define a **digital twin** to be a dynamic, virtual replica of a physical system that can be continuously updated with new data, allowing up-to-date monitoring, analysis and optimization of the system. A digital twin can run any number of useful simulations to study multiple processes and is linked to physical data sources so that updates reflect the current state of the physical object [17]. Simulations usually do not require real-time data, but digital twins are ideally designed around a two-way flow of information. This is the main drive for the increased use of digital twins, particularly in healthcare applications [18].

(a) Application and value in engineering and physical sciences

In engineering and physical sciences, mathematical models are typically formulated using systems of equations that define the relationships between variables and dynamics based on physical laws. They are powerful analytical tools that help us better understand physical phenomena and simulate complex systems. For instance, the Lotka–Volterra (predator–prey) model [19] in ecology is used to study the interactions between predator and prey species over time. In addition, researchers employ mathematical models to explore different scenarios and their potential effects. For example, four forcing scenarios, representative concentration pathways [20], are used in climate models to study their effects on global climate systems.

Mathematical modelling and simulation are increasingly used as decision support tools. Modern engineering uses computer modelling to develop designs in a cost-effective way [21]. For instance, the Langley grid-back booster for rockets at NASA was developed through the use of computer simulators [22]. In geology, computational models are employed for hazard control with geophysical flow models to assess hazard threats owing to volcanic avalanches [23]. In climate research, the Coupled Model Intercomparison Project (CMIP6) brings together climate modelling centres to produce multi-model climate projections to explore a number of science and policy questions [24] that provide ‘evidence base’ for policy decisions globally [25].

Decision-making can be further enhanced by utilizing digital twins. The concept of a digital twin [1] originated in the early 2000s in manufacturing and was formally introduced by NASA in aerospace engineering [26] where they were used to mirror the behaviour of aircraft engines and other critical system components in space in real time. Today, digital twins can predict engine failures in commercial aircrafts before they occur while also enhancing flight path updates through continuous system monitoring [27]. Other examples include engineering and industrial manufacturing, and smart city simulations [28]. Despite the success of digital twins in industrial sectors, it is still a relatively new concept in environmental sciences with ambitious projects such as Destination Earth (DestinE) starting to appear to support climate change

adaptation [29]. Dale *et al.* [1] outlined the main challenges in adopting digital twins, which include modelling complex environmental systems and environmental information integration.

Advances in computational power have introduced new challenges in modelling and simulation. While they enable more accurate and complex simulators, these models are often computationally demanding, thus increasing computational costs. Multi-fidelity computer models [30–32] address this by combining computationally inexpensive models with fewer runs of expensive high-fidelity models. Multi-scale models are also used to integrate simulations at different scales. For example, aerospace models are built to test the safety of aircraft materials [27]. To reduce costs, a variety of length scales ranging from small sections of carbon fibre-reinforced plastic, to aeroplane wings, to the full aircraft are tested.

(b) Application to biology and healthcare

Biological systems are distinct from physical systems in numerous ways, one of which is the relatively fewer number of ‘laws’ that govern biology. Laws of physics, motion and momentum can be applied in biological systems, but these laws only cover a few applications (e.g. biomechanics, biofluids). Thus, biological and healthcare models are derived in a way that is specific to the population, organism or biological process of interest.

Biological and healthcare simulations have recently made it into the public limelight. A food and drug administration-approved tool for insulin delivery is based, in part, on mechanistic modelling (using ordinary differential equations (ODEs)) [33]. Simulations of heart biomechanics, heart valve function and blood flow have all contributed to new technologies in the healthcare industry [15]. HeartFlow® is an image-based simulation platform that combines medical imaging, computational haemodynamics and ML to predict coronary artery disease risk factors [34]. More resources are now dedicated towards modelling for medical device development [35], ‘patient-specific’ models [36] and *in silico* clinical trials [37].

The use, and criticism, of biological modelling was highlighted during the COVID-19 pandemic [38–40]. The European Union (EU) used the ‘CovidSim’ software, developed in part by Neil Ferguson at Imperial College, London [41], for pandemic response. Though the model fared well in helping plan effective mitigation strategies, a study by Edeling *et al.* [42] quantified the uncertainty and biases in the CovidSim model. CovidSim includes over 900 model parameters, of which 60 were included in the UQ analysis by Edeling *et al.* The authors found that 19 of the 60 parameters were influential on the model outputs and provided nearly 300% variability in the model outputs, yet uncertainty accredited to model misspecification and variability in the simulation settings were not investigated.

One of the most exciting, cutting-edge areas of computational biology and healthcare is in the development of human digital twins. This is part of the broader goal of the Virtual Physiological Human’s [43] EDITH project in the EU. Funded by the European Commission, EDITH aims to develop a virtual human twin of human physiology in an attempt to better address pathological conditions [44]. Within this road map is the idea of model ‘credibility’, which includes aspects of UQ. A workshop hosted by the National Academies of Science, Engineering and Medicine in the USA identified the gaps and future directions in the development of digital twins [17]. A key development of this workshop was updated definitions for digital twin technologies, including (i) the ability of a digital twin to have predictive capabilities and (ii) the need for bidirectional interactions, such that the digital representation of the physical system can be updated with new data and information. In documenting the current gaps and unmet needs in digital twin technologies, the committee noted that the area of verification, validation, and UQ (VV&UQ) was especially important in constructing physiologically digital twin technologies [17,18].

The report [17] also made an important distinction between what are now new ‘digital twins’ and the previous ideas surrounding patient-specific models. The bidirectional nature of the digital twin implies that the model and data are not single instances, but rather

continuously updated. A notable example is the recent study by Chadhuri *et al.* [45], who provided a computational oncology digital twin that was calibrated over multiple time points and provided output uncertainty over an extended period of time. As digital twins become reality, a next step will be in clinical decision support using these twins. This idea is echoed by Huberts *et al.* [46] in the cardiovascular modelling space, whereas the article by Sesen *et al.* [47] considered a purely data-driven Bayesian network approach to decision support in lung cancer. Quantifying uncertainties will be paramount as these techniques and strategies are implemented in clinical decision-making.

There are several examples and outlines of digital twins in specific domains, including cardiology [15,18], pulmonary hypertension [48], oncology [45] and immunology [49–51]. In addition, some initial pipelines for UQ in biological and healthcare systems have been proposed [14,52]. Though UQ is also being applied in the context of ‘model-free’ artificial intelligence (AI) and ML methodologies, this is outside the scope of the present review (see Seoni *et al.* [53] for details).

3. Uncertainty Quantification (UQ)

Mathematical models and digital twins are not perfect replicates of the real-world systems they represent. To support high-stake decisions using computer-based simulations, we need to ensure that predictions produced by these models are credible. The study in [54] identified three key principles to assess model credibility: verification, validation and uncertainty quantification. Verification ensures that the numerical implementation and solution of the model are correct. Validation tests whether the model itself sufficiently captures the dynamics specific to the quantity of interest (sometimes called the ‘context of use’). The last ingredient, UQ, consists of formally identifying and quantifying uncertainties within the model and associated with the data when performing model-based inference. Although UQ is a well-established and widely recognized discipline within the engineering and physical sciences communities, UQ for healthcare and biological systems is under-developed due in part to the complexity of biological and healthcare data, as well as the greater amount of ‘unknowns’ in the development of biological models.

One of the main aims of this special issue is to improve modellers’ and healthcare and biological researchers’ understanding and comprehension of UQ, which can significantly encourage the adoption of UQ as part of the credibility assessment. Many UQ ideas have been echoed in certain healthcare domains, more specifically in the cardiovascular modelling community and the systems biology domain. These ideas were presented in the ‘fickle heart’ meetings [55], as well as the ‘Special Issue on Uncertainty Quantification, Machine Learning, and Data-Driven Modeling of Biological Systems’ [56]. We also note that the field of *in silico* clinical trials has adopted multiple ideas from the UQ field [57], including the Comprehensive *in vitro* Proarrhythmia Assay (CiPA) initiative [58]. The goal of CiPA is to reduce the chance of cardiac toxicity in clinical trials, and has provided a roadmap for how UQ can be used in guiding medical and pharmaceutical directions.

Figure 1 demonstrates a UQ pipeline, including some essential components and their connections. In particular, we consider a computer model, a mathematical (coded) representation of a physical process of interest, expressed through a mathematical function f that takes varying input parameter values \mathbf{x} and returns output $f(\mathbf{x})$. Observations \mathbf{z} are collected from the physical system to aid in the design and calibration of the computer model. We recognize that the computer model is not a perfect representation of the physical system and therefore include the model discrepancy, denoted as δ , in our analysis. As part of the UQ pipeline, we consider both forward and inverse UQ. Forward UQ, typically performed first, involves propagating uncertainties in model inputs through the mathematical model to determine how they affect the outputs. Inverse UQ, typically performed second, works in reverse; it quantifies

the uncertainties in the inputs based on observed data, \mathbf{z} , or desired outcomes. Performing forward and inverse UQ analyses typically require numerous computer model runs. However, limitations arise when simulators have an extensive infrastructure, leading to high computational costs and thus restricting the number of available simulations. To avoid this computational burden, it is common to construct a surrogate model (an emulator) that provides a fast approximation of the computer model output. In the remainder of this section, we discuss UQ methods shown in figure 1 in detail.

(a) Computer simulations and emulators

When dealing with expensive simulation models, a solution is to adopt an emulator, which is a computationally fast approximation to the model. An emulator requires an input training design $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, and the corresponding model realisations $f(\mathbf{X}) = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$. It is common to adopt ‘black-box’ approaches when constructing emulators, where the focus is on modelling the input–output relationships without needing to fully understand the complex components of the simulation model. Extensive research exists in the field of emulators and surrogate modelling with methods including Gaussian processes (GPs) [59–61], polynomial chaos [62,63] and neural networks [64,65]. There are also innovations in *physics-informed* emulation [64,66,67], where underlying mechanisms of the system (e.g. the model equations) are integrated with observations to improve the emulator’s interpretability and ensure that physical constraints of the system are preserved.

As shown in figure 1, after constructing an emulator, it is crucial to validate it to ensure an accurate representation of the model’s behaviour across the input space. An ill-fitting emulator can lead to misleading uncertainty estimates or predictions. The accuracy of an emulator is also greatly influenced by the design of the training points [68,69]. It is important to balance between a space-filling design and one that captures key features of the physical system (exploration versus exploitation) [70]. Optimal designs can be either single-shot, using the entire budget at once, or adaptive, selecting new parameter locations sequentially based on predefined criteria (see [71–75] for details). As shown in the pipeline, an initial design is generated (e.g. a Latin hypercube, Sobol’ sequence or sparse-grid design [75]), and an emulator is built. An adaptive design (e.g. mean-squared-prediction error or mutual information) then uses information from the emulator to optimally select new input locations.

As discussed in §2a, physical systems are often represented by complex mathematical models that can involve high-dimensional input and output spaces. To simplify the analysis of these models within the UQ framework, it is common to adopt projection-based reduced-order modelling (ROM) methods [76–78]. After performing ROM, it is then possible to construct surrogate models and perform an inverse problem using the reduced space [79,80]. In particular, studies in [78,80,81] used principal component analysis before emulation since it allows to compress the large output space into a smaller, more manageable set of principal components that capture the majority of the variation in the data. In applied mathematics, a similar approach is carried out using proper orthogonal decomposition [75]. In addition, as mentioned in §2a, highly complex computer models can be run at different levels of complexity or resolution. Emulators have been developed for both multi-level [82] and multi-scale models [83].

The output of, for example, engineering models is typically deterministic, where the output is identical if the model is run multiple times at the same parameter settings. Alternatively, there are many cases in healthcare and biology, such as epidemiology, where a random component causes the output to be stochastic. The additional variability must be accounted for, thus requiring different emulation strategies, such as heteroscedastic GPs [84].

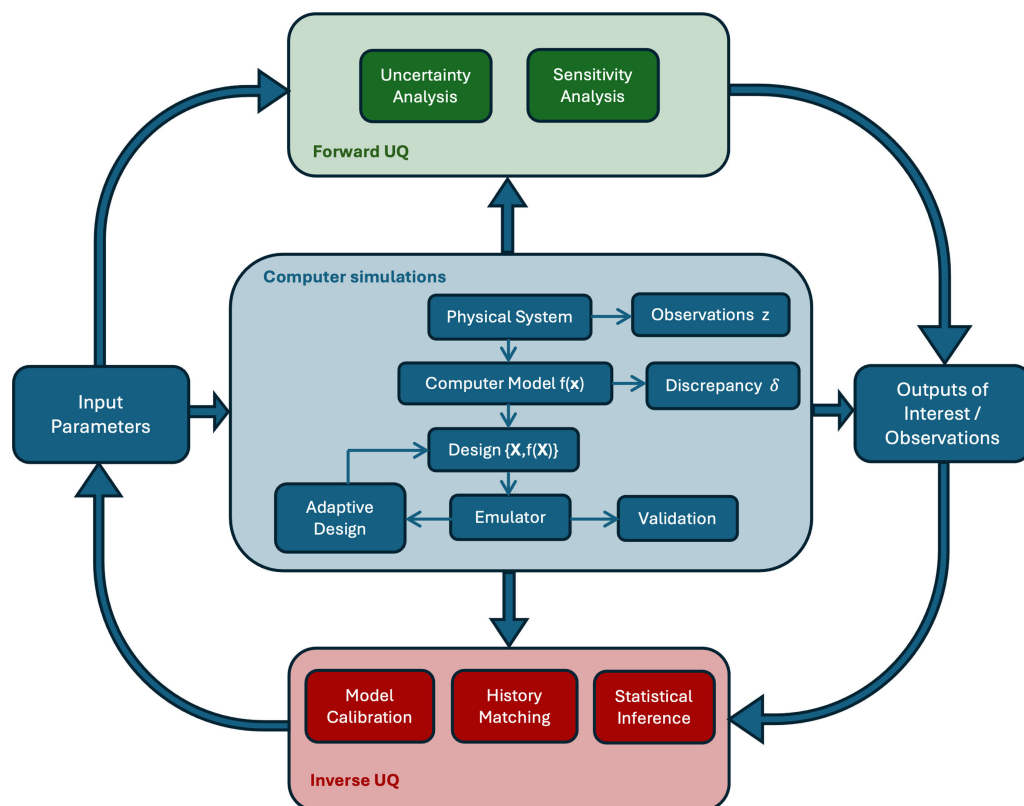


Figure 1. UQ pipeline. At the center of the UQ pipeline is the computer model f that provides a virtual representation of the physical system. Each coloured box corresponds to a distinct branch of UQ methods described in §3. The flowchart demonstrates the iterative nature of UQ, where input parameters \mathbf{x} drive simulations that are continuously refined and validated, linking model outputs to real-world observations (\mathbf{z}) from the physical system.

(b) Forward UQ

Forward UQ methods propagate *prior* input uncertainties through the computer model to assess their effect on the model outputs. Predictions paired with uncertainties give meaningful predictive capabilities and help with robust decision-making. Emulators provide a fast and efficient way to explore model responses under different input scenarios, enabling forward UQ at a lower computational cost. The two main methods of forward UQ (as shown in figure 1) are uncertainty analysis and sensitivity analysis (SA) [85]. The former propagates uncertainty from various sources through the model to compute the model's output uncertainty distribution. This output uncertainty, often measured by its variance, can then be decomposed to evaluate the influence of each input on the model's output, which is SA [75,86].

The main goal of uncertainty analysis is to quantify the information lost owing to uncertainties in model inputs, parameters and assumptions[.]. This is typically done using Monte Carlo techniques, where input parameters are treated to be random variables with known distributions. Large samples of input parameters are drawn, and the corresponding outputs are evaluated to produce the uncertainty distribution. When using emulation, the added uncertainty from using the emulator must be included in the analysis.

SA quantifies how the model inputs contribute to the output uncertainty, allowing input parameters to be ranked according to their influence [87,88]. There are two main approaches: global SA, which examines input–output relationships across the entire input space, and local SA, which focuses on small perturbations around a specific point [14]. Local SA is limited to

small regions, typically using either automatic differentiation or derivative-based sensitivities [14,89], making it less suitable for nonlinear models. Global SA methods consider the effects of all inputs and their interactions, offering a comprehensive overview of their influence on the model output. However, this approach can be computationally expensive, requiring sufficient sampling of the input space to capture variability and interaction effects. Common global SA methods include variance-based approaches such as Sobol' indices [90], Fourier amplitude sensitivity tests [91], Morris screening [14] and Shapley indices [92], which aim to calculate contributions from each input and their interactions on the output uncertainty. Each methodology has its own assumptions (e.g. parameter independence), which subsequently affect the reliability of the results. Emulators can also be used to perform SA more efficiently [93].

(c) Inverse UQ

Inverse UQ focuses on quantifying the *posterior* uncertainty in model inputs, and the associated posterior uncertainty in model outputs. Part of inverse UQ is parameter estimation, which connects models to reality using observational data and predicts future states and outcomes of the system. Note that output uncertainty propagation based on prior input uncertainties can provide substantially different results than output uncertainty propagation based on posterior input uncertainties (i.e. after model calibration from data) [14].

In some cases, modellers adjust parameter values inside the model manually [94], though this can be very time-consuming and unreproducible. Alternatively, a distance function between simulations and observations can be specified, such as least squares, and minimized, providing a deterministic, single best set of parameter values [14,95]. These optimization problems can be ill-posed [96], i.e. with multiple local optima. In addition, optimization over all observations can lead to overfitting (overtuning) [95], where the model performs well for those outputs but exhibits unphysical behaviour in others.

To address these concerns as well as to perform a realistic assessment of parameter uncertainty in mathematical models, inverse UQ methods can be employed. Optimization methods can be combined with an appropriate observation error model (defined in equation (3.1)) to reformulate the least-squares problem into a frequentist statistical inverse problem [14]. This enables the calculation of parameter confidence intervals after estimation, similar to standard methods used in linear and nonlinear regression. In contrast [96], developed the Bayesian approach to inverse problems, which represents the observation z_i as

$$z_i = f(\mathbf{x}^*) + e_i, \quad i = 1, \dots, m, \quad (3.1)$$

where e_i is a zero-mean random variable that corresponds to the observational noise. A prior distribution for model parameters $\pi(\mathbf{x})$ is specified, and observations z_i are used to derive the posterior distribution. This method provides parameter posterior uncertainty and considers the observation error (measurement uncertainty) as part of the analysis. However, failing to recognize that the mathematical model is not a perfect representation of the true process of interest can lead to biased and over-confident parameter estimates and predictions [97]. Similar to the study in [96], Kennedy & O'Hagan [98] assumed the existence of the 'best' input parameter setting \mathbf{x}^* and proposed to define a model discrepancy as the difference between the physical process and the mathematical model representation, and include it in the observation equation, i.e.

$$z_i = f(\mathbf{x}^*) + \delta_i + e_i, \quad i = 1, \dots, m, \quad (3.2)$$

where the model discrepancy term δ is a zero-mean GP. This turns it into a problem where we need to infer parameters and model discrepancy jointly. However, this approach may suffer from non-identifiability issues that can be resolved by informative priors specification on the parameters and model discrepancy [97,99]. A frequentist approach to identifying model

discrepancy is presented in [100]. We consider observation error and model discrepancy as part of our proposed UQ pipeline in figure 1.

Obtaining improved prior specifications can be challenging in most applications. In addition, solving a joint parameter and model discrepancy inference problem can be computationally infeasible in a high-dimensional space [80,101]. Instead, history matching (HM) can be used to obtain the regions of parameter space corresponding to a good match between observations and model representation [102]. Bayesian HM uses a distance function, commonly known as the implausibility measure, to iteratively rule out the implausible regions of parameter space. This implausibility measure only requires expectations and covariance structures specifications—namely, expectation and variance of the model output, variance of the observation error and model discrepancy variance—aligning with a Bayes Linear philosophy [103]. In practical applications, it is common to treat the model discrepancy as the modellers' tolerance to model error (see [3] for more details). We also note that HM can be performed as a pre-calibration step [104].

4. Current trends, gaps, and opportunities for UQ in biological and healthcare domains

The application of UQ in biological and healthcare models is understudied and suffers from several critical knowledge gaps. To successfully adopt computational models in biological or clinical decision support systems, several challenges need to be overcome, as discussed below.

(a) Incomplete model credibility assessment

While VV&UQ is prominent in engineering and biology, UQ is typically lacking. The clinical community shows skepticism in using mechanistic models as assistance tools in patient-specific evidence-based diagnosis, risk assessment and treatment decision support owing to models being simplifications of the complex human physiology. Biological and physiological mechanistic models have relatively fewer concrete 'laws' describing processes than traditional physical science domains [14]. For instance, the Navier–Stokes equation is a principal-based equation for fluid flow, yet its application to blood flow is non-trivial [105]. Standard epidemiological and ecological models, e.g. susceptible-recovered (SIR)-type models, come with their own limitations in mechanistically describing disease transmission [57].

Opportunity: To generate trust between researchers, scientists, clinicians and patients, a thorough assessment of model credibility (VV&UQ) is necessary [46,106–108]. Model credibility has been addressed in detail by several authors, including Erdemir *et al.* [109]. Clearly defining the context of use for a model, performing VV&UQ using appropriate output quantities and listing limitations of the model explicitly are a few of these important rules for credibility.

(b) Lack of UQ understanding and incomplete UQ

We require a deeper understanding of UQ, which is the least understood component in the model credibility process. More specifically, the application of UQ to healthcare and biological systems is under-developed [46,54,107]. There is an urgent need for a synergy between experts in the UQ field and mathematical modellers, who are often underexposed to UQ tools.

In particular, *model discrepancy* between the mechanistic model and the real system is an often neglected source of uncertainty in healthcare and biological modelling studies, consequently leading to distorted predictions and biased uncertainty [12,97,110]. Although some studies have included model discrepancy in the UQ analysis [12,110–112], it remains a difficult problem to tackle owing to confounding interactions between model discrepancy and error

parameters (see §3c) stemming from insufficient data or knowledge about the system [97,99]. This is a prevalent issue in biological problems, owing to missing physics or simplifying assumptions in the modelling framework, or numerical approximations in the numerical scheme used to solve the mechanistic model equations [99,113].

In addition, a complete picture of UQ is usually missing in biological and healthcare modelling research, leading to misestimated uncertainty. More specifically, failing to account for both types of UQ (i.e. forward and inverse) [114,115] can cause uncertainty misestimation. Furthermore, failing to jointly incorporate all inverse UQ sources, e.g. uncertainty in model parameters, measurements, numerical errors from discretization (e.g. partial differential equations (PDEs)) or model error, will cause incomplete UQ. While several studies in the literature [7,42,116] include uncertainty in model parameters, the effects of parameter fixing in the presence of a large number of parameters are an ongoing challenge.

Opportunity: A holistic UQ approach will increase modelling rigour and reliability and generate confidence in them to guide healthcare decisions. Though forward UQ is useful for understanding predictive variability, inverse UQ provides information about posterior uncertainty and is crucial in determining confidence in model predictions in light of data. Coupling forward and inverse UQ can help address the issue of parameter identifiability, where mathematicians can work jointly with statisticians to integrate physiological prior knowledge into the statistical analysis [117]. Identifiability will be key as digital twins and their parameters are used as new disease biomarkers [14,15]. Team-based science that includes experts in mathematical modelling, computational statistics, computer science and biology (or physiology) is necessary for synergistic activities in digital twin development [17]. When addressing model discrepancy, modellers should collaborate with statisticians to identify specific components of the model that lack fidelity and provide prior evidence for discrepancy terms [11]. In addition, more investments should be made in educational literature [75,118] and workshops. We also emphasize, in contrast to prior perspective pieces (and reflected in figure 2), that **statisticians should be consulted and included in the team-science approach within biological and healthcare modelling.**

(c) Adaptation of existing UQ methods

Biological and physiological models have complex, multi-scale mechanisms with limited, noisy observations obtained from experimental measurements [119,120]. Many biological inverse problems are ill-posed owing to non-identifiability between initial conditions, boundary conditions (BCs) and biological parameters, leading to multiple solutions of the system equations [119,120]. Thus, existing UQ methods require adaptation for healthcare and biological systems. As an example, SA for deterministic systems is commonly employed [114,121], yet biological agent-based modelling (e.g. cancer modelling) [5,122] requires innovative SA methods to accommodate the stochastic output. Biological and healthcare models typically have multiple internal states, yet typical SA quantifies parameter effects on a single quantity of interest.

Opportunity: Biological and healthcare models are distinct in their complexity and underlying mechanisms. New mathematical and statistical innovations to handle multi-output quantities of interest and quantify correlations/covariance of the model outputs are warranted. In addition, there is a critical need for determining parameter identifiability [123,124] in complex, biological systems. UQ in systems described by stochastic differential equations [125] requires innovations to deal with the problem of intractability of the data likelihood. For instance, the development of accurate and efficient likelihood-free approaches (e.g. approximate Bayesian computation, ABC[126]) based on data features lends itself to the opportunity of devising an algorithm to extract informative and low-dimensional features from the data, and to appropriately weight and combine them.

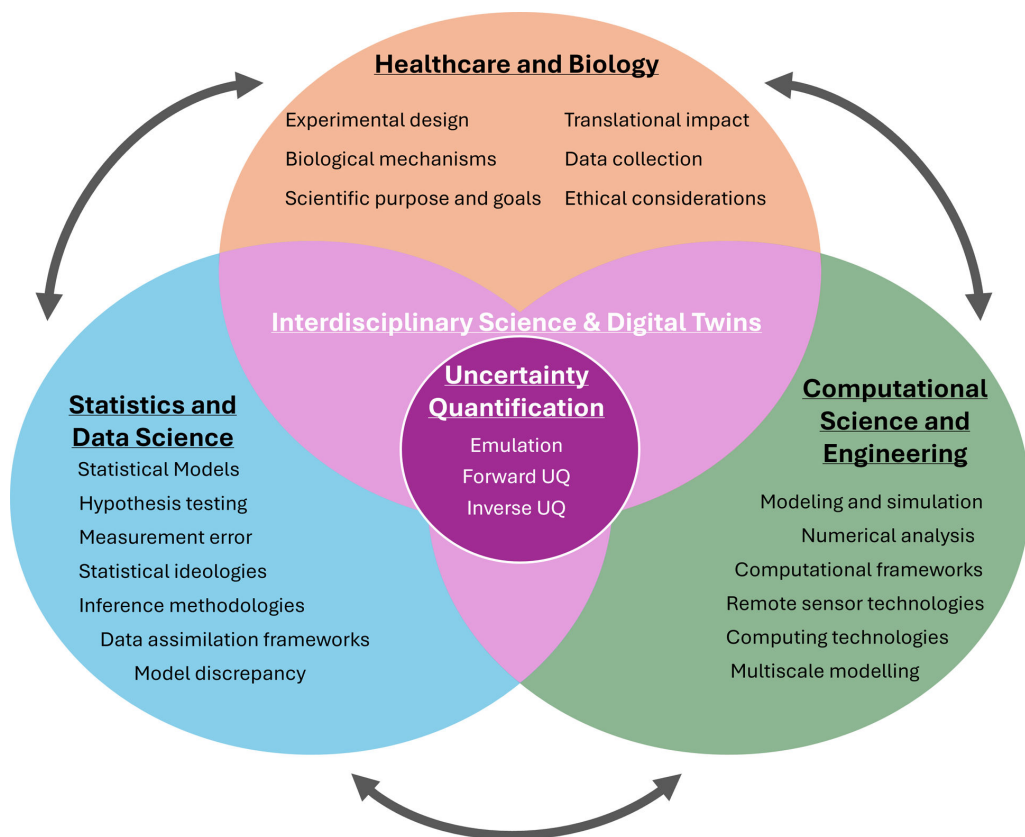


Figure 2. Overarching flow of expertise for UQ. Each subdomain has its own specific skill sets which, when applied together in an interdisciplinary setting, can be used to better explore uncertainty in biological and healthcare problems.

(d) Inconsistent UQ language

Although UQ is adopted in multiple disciplines, the language used differs. For interdisciplinary research, it is important that experts in mathematics, statistics, engineering, healthcare and biology develop a common language. For example, an approximate mathematical model is referred to as a ‘metamodel’, ‘surrogate model’ or ‘emulator’ in different disciplines [110]. In a similar way, ‘calibration’ may imply hand-tuning or manual parameter adjustments, whereas formal model calibration in statistical (and now also engineering) applications requires numerical procedures, measurement error models and transparent assumptions about the model and data [12,14]. The term ‘model’ in engineering typically implies a mechanistic model, while in statistics, it implies a statistical model that may have different (free or fixed) parameters.

Opportunity: Scientists will benefit from an understanding of domain-specific UQ language. Developing collaborations across fields (see figure 2) helps alleviate potential misunderstandings in notation and language. We recommend that scientists read UQ articles from other domains, consider UQ texts that clearly communicate different definitions and languages [75,118], and take time to interact with different scientific domains on UQ. This will drive new research directions where UQ is at the intersection of multiple fields.

(e) Ambiguity about the right model complexity

Determining the correct model complexity for a specific problem is non-trivial and, in UQ, applies to both the mechanistic model and statistical model complexity. The former includes choosing model fidelity (e.g. the inclusion or exclusion of certain biological processes) [8,127], or the choice of ODE versus PDE models [14]. This choice also affects the degree and form of model discrepancy [12]. Keeling and Eames [128] illustrate how epidemiological model fidelity affects simulations of disease infection. In addition, the parameter complexity (i.e., number of inferred parameters) of the model can lead to differences in the overall conclusions of a study. Colunga *et al.* [89] showed that differences in the inferred parameter subset led to different conclusions regarding heart-failure outcomes. The choice of statistical error model and its complexity (e.g., independent and identically distributed versus correlated errors) may affect the estimated uncertainty [110].

Opportunity: An alternative to choosing one fidelity of the mechanistic model is combining models of different fidelities or scales (e.g. from molecule to cell, tissue and last organ level) in a multi-fidelity, multi-scale modelling framework [14,109,119]. Systems with poorly understood physics or limited data using multi-fidelity modelling require new innovations in UQ methods as well [119]. While multiple papers use SA for parameter fixing *a priori* [12,89,107], relatively few papers use statistics-driven criteria, such as information criteria [12,129], for identifying model complexity. Again, we emphasize that computational modellers and statisticians should work together to identify model complexity in both the infrastructure of the mechanistic and statistical model when conducting inverse UQ.

(f) Difficulties with data

Limited, incomplete and noisy data are a common issue in modelling healthcare and biological systems [119]. Many biological and healthcare research use averaged or non-dynamic (e.g. blood pressure) data [14]. However, mechanistic models may predict non-averaged or dynamic outputs [12] and would benefit from higher-fidelity data sources. Other factors include inadequate experimental designs for model calibration, incomplete health history or noisy measurement devices [14]. These factors result in model parameters that are practically non-identifiable when solving statistical inference problems [14,124]. Data bias is another issue. Under-representation of individuals in the data for model training will lead to biased, suboptimal healthcare for the particular group [15]. The guidelines introduced by [14] emphasize data bias in some cardiovascular modelling studies owing to the absence of factors such as sex from the experimental study. This has resulted in biased cardiovascular models that are male specific, limiting their application to female patients and biasing future UQ on these models.

Dealing with high-dimensional and multi-modality data is another challenge in biomedical research. Large and complex ‘omics’ datasets as well as electronic health records and medical imaging data are usually heterogeneous, making a proper UQ study a challenging task [130,131]. Analysing medical images, such as segmenting and meshing the heart for simulation [132], relies on manual interventions by trained experts, rendering clinical digital twins currently infeasible.

Opportunity: Numerous initiatives aim to collect an abundance of data for ML and AI. However, **a major question is whether these data are accurate or informative for modelling and simulation.** Putting UQ at the forefront will enable clear communication about which data are appropriate for a given problem. Data and measurement devices are imperfect, and accounting for their limitations in a statistical framework will ensure that results and conclusions from the data are reliable and credible. Modellers and statisticians should work in concert to identify beneficial data for research projects and collaborate with domain experts and experimentalists to build new experimental designs (see figure 2). Efforts to ensure that computational modelling enhances healthcare for all patients, regardless of their demographics

and especially in the cases when data are limited, are needed. Integrating physical constraints such as conservation, invariance or symmetry into the AI- and ML-based analysis may be beneficial.

(g) Emulators and UQ

Emulation methods, while powerful, have shortcomings. Their approximation of the simulator introduces an extra layer of uncertainty into the analysis. Conducting UQ analysis within an emulation framework poses additional challenges owing to accuracy loss, and the development of emulation UQ methods that are able to achieve accurate results is an active area of research [133,134]. Recently, there has been a surge in studies adopting physics-informed emulation in place of a data-driven approach based on numerical simulation data [64,66,67]. For instance, the study by Grandits *et al.* [64] uses a graph neural network for emulating a cardiac mechanics model. Physics-informed emulation has more often been employed from a purely predictive perspective [64,66,67] rather than from an inference perspective. An example from the latter is the study by Ahmadi Daryakenari *et al.* [135] that employs a physics-informed neural network for parameter estimation and missing physics identification in systems biology.

Opportunity: Mathematicians, engineers and statisticians all employ different emulation strategies (e.g. GPs, polynomial chaos expansions, neural networks, etc.) [64,119]. Clear communication about emulation performance metrics and their ability to solve forward and inverse problems is warranted. In particular, understanding which emulation strategies return the most accurate predictions and investigating how predictive accuracy correlates with inference accuracy is needed. In biological mechanistic modelling, substantial efforts are required to integrate physics-based emulation within a multi-fidelity, multi-scale and multi-physics framework [136]. We emphasize that integrating physics with data provides a possible solution to the challenge of contemporary AI by making ‘black-box’ (data-driven) ML methods ‘explainable’ and advocate for educating and encouraging non-computational scientists to use mechanistic modelling in a synergistic fashion with AI and ML.

(h) Incorporating patient-specific healthcare into a population framework

Extending predictive models to a general patient cohort is non-trivial [15]. First, mechanistic models must be extended to reflect physiological and demographic heterogeneity of the population (e.g. biologically female and male patients) to avoid model bias in clinical decision-making. Studies so far have only dealt with relatively small patient groups (approx. 100 patients) [137], and hence substantial efforts will be required for larger scalability. Second, AI and ML method extensions are lacking, more specifically building emulators that account for anatomical variability for a population, or conducting the UQ analysis in a population framework using a Bayesian hierarchical model. Integrating population data in a dual mechanistic-statistical framework will require significant innovations in inverse UQ, as measurement noise and parameter influence may be patient-specific, making it difficult to expand to a population.

Opportunity: Statisticians have experience in handling both individual and population-level data (e.g. Bayesian hierarchical models) and should be consulted in the development of mechanistic models at the population level. Population data may have distinct characteristics to patient-specific data owing to issues with data collection strategies at large scales, possibly resulting in missing or incomplete data, and even outliers. To tackle this, the available data may be supplemented with synthetic cases obtained through computational simulations. In addition, reliable data-sharing platforms will play an important role in providing additional datasets [15]. Epidemiology modellers routinely simulate large-scale dynamics using population data

and will be a valuable contributor to statisticians and healthcare modellers to better translate patient-specific to population simulation.

(i) Ethical and legal considerations

Before the incorporation of digital twins and AI in a clinical practice, several critical ethical considerations need addressing, such as (i) patient consent for their medical data to be used, (ii) data privacy and protection, and (iii) transparency and accountability [138,139]. In the context of extending the UQ framework from a patient to a population level, where access to data-sharing platforms is important, consent and data sharing are not straightforward to implement in a clinical pipeline. For instance, it has been argued that access to shared patient data should be granted to national and international bodies under certain conditions, such as securing and de-identifying data (i.e. removing personal information, such as name, address and social security numbers) [140]. There is also ambiguity about transparency and accountability when using black-box algorithms for UQ. For example, how transparent should a clinical professional be about their limited understanding of how these algorithms work to guide the medical decision support, how will this relate to the *legal* ‘right to explanation’ under the EU General Data Protection Regulation [106,139], and who will be held accountable if anything goes wrong? These concerns could be partially alleviated by designing explainable AI methodologies that enable UQ.

Opportunity: Science will only have its effect on everyday life when it is clearly communicated and disseminated through appropriate policy experts. In light of the COVID-19 pandemic and in preparation of the era of digital twins, modellers, statisticians and healthcare experts must communicate ideas in a digestible manner. Moreover, UQ must be incorporated in the future to emphasize the possible outcomes and precision of predictions or conclusions from computational frameworks. The use of modelling without proper UQ can over-promise the potential of computational science. As argued throughout this text, UQ and the concept of uncertainty must be (i) integrated within every problem that uses computational modelling, and (ii) clearly explained for multiple audiences to ensure that credibility and trust are at the forefront of a given application area.

5. Conclusions and articles in the special issue-Part 1

Computational modelling and simulation of healthcare and biological systems is a rapidly evolving area, with numerous opportunities to affect policy decisions, healthcare practices and new technologies. Like physics, climate science and aerospace, the results in healthcare and biological applications should be scrutinized via a proper quantification of uncertainty. In this review article, we emphasize the need for synergy between experts in statistics, modelling and simulation, and domain science. This volume of the two-part special issue contains articles that address multiple challenges discussed in §4.

We begin with the opinion article by Tsaneva-Atanasova *et al.* [141], which examines in detail the role that UQ plays in informing clinical decision-making processes. The authors discuss the crucial role of understanding and managing the uncertainties present in clinical data (such as measurement error), diagnostic tools and treatment outcomes. The authors demonstrate how effectively addressing and decoding uncertainties could significantly enhance the accuracy and robustness of clinical decisions, ultimately leading to better patient outcomes and more informed healthcare practices.

Scientific ML holds significant value, especially in expediting simulation and UQ. For instance, the study by Böttcher *et al.* [142] aims to improve existing control and optimization techniques specifically for human biological models where typical physics-based approaches are not suitable. They use dynamics-informed artificial neural network (ANN) controllers as

an alternative approach to using medical digital twins, focusing on the control of agent-based models in biomedical applications. The authors show that the ANN approach is robust and computationally efficient, allowing for the calculation of parametric uncertainty. In addition, the study by De Florio *et al.* [143] provides a novel strategy for handling ‘total uncertainty’, including epistemic, aleatoric and model-form uncertainty within the framework of physics-informed ML. Their methodology, applied to a reduced-order computational haemodynamics model, showcases the importance of model misspecification in the emulator training process.

In the field of emulation, the study by Paun *et al.* [144] compares GPs and polynomial chaos expansions as emulators for forward and inverse problems in the context of pulmonary blood flow modelling. Their results indicate that GPs outperform slightly, but consistently polynomial chaos expansions across every comparison, and that a similar performance is obtained for the emulators of multivariate output and reduced output from the fluid-dynamics models. The study by Salter *et al.* [145] provides a novel emulation approach to overcome issues in population dynamics via a Poisson log-normal principal component analysis emulation strategy applied to a COVID-19 model. The authors show that the emulator provides predictions that capture complex correlations between the true model outputs.

In the study by Bajaj *et al.* [146], the authors develop a new Bayesian approach to account for the dynamics of the reproduction number, R_t , in epidemiological models. Their approach highlights the importance of R_t history and provides a method for quantifying uncertainty in the time-dependent values of R_t . In addition, the use of novel Bayesian approaches using computationally cheaper surrogate models is presented in the study by Richter *et al.* [147], which demonstrates how to efficiently quantify uncertainties in three-dimensional cardiovascular fluid-dynamics simulations using clinical measurements. The authors only evaluate the three-dimensional model once for an initial choice of boundary conditions and use the result to create a highly accurate zero-dimensional (0D) surrogate. The authors then perform sequential Monte Carlo using the optimized 0D model to derive the high-dimensional (Windkessel) boundary conditions posterior distribution, thus significantly reducing the computational complexity of the problem.

The study by Dadashova *et al.* [148] develops a method for determining identifiable parameters in the physiologically based pharmacokinetic brain model that comprises coupled systems of ODEs and a large number of parameters. These models are crucial for understanding the movement of drugs and other chemical compounds within the body. The authors propose to combine the parameter subset selection method with Bayesian inference to obtain a subset of identifiable parameters with quantified uncertainties.

In the field of cardiac electrophysiology, Del Core & Mirams [149] consider the stochastic nature of ion-channel behaviour. The study presents a state-space model based on stochastic reaction networks, which retain statistical properties that can be used to identify parameters in the system. The authors show that the proposed method outperforms standard inference in ion-channel models and is able to detect distinct components of measurement error that are typically missed in deterministic-based inference. The study by Shuttleworth *et al.* [150] calibrates a model with a voltage-sensitive potassium ion-channel current to newly collected data from patch-clamp experiments. The authors assess how each model predicts the dynamics of the patch-clamp protocol and report variability in parameter estimates with each protocol. The authors show that heterogeneity across experimental designs and cell types accounts for a missing latent effect within the modelling framework, and that incorporating this missing effect may benefit the applicability of these models to a wider set of experiments. Furthermore, Patten-Elliott *et al.* [151] focus on defining the mechanisms by which drug compounds bind to the human Ether-a-go-go Related Gene to reduce uncertainty when assessing the risk of cardiac arrhythmia. They present automated optimal experimental design techniques to design voltage-clamp protocols that can be used in voltage-clamp experiments to better distinguish between different models of the drug-binding mechanism.

We end the special issue with a study by Rabbani *et al.* [152] that develops an adaptive three-dimensional image synthesis technique to reconstruct the inhomogeneous, porous microstructure of the meniscus, a fibrous knee tissue. These reconstructions are used to model hydraulic permeation and mechanical testing. By analysing 1500 synthesized geometries, they introduce a feature extraction technique, allowing for the systematic generation of perturbed three-dimensional biomaterial models. This framework helps identify structure–function relationships, offering insights into the connection between macroscopic biophysical properties and microscale tissue characteristics.

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. L.M.K.: conceptualization, writing—original draft, writing—review and editing; L.M.P.: conceptualization, writing—original draft, writing—review and editing; M.J.C.: conceptualization, writing—original draft, writing—review and editing; V.V.: conceptualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. This theme issue was put together by the Guest Editor team under supervision from the journal's Editorial staff, following the Royal Society's ethical codes and best-practice guidelines. The Guest Editor team invited contributions and handled the review process. Individual Guest Editors were not involved in assessing papers where they had a personal, professional or financial conflict of interest with the authors or the research described. Independent reviewers assessed all papers. Invitation to contribute did not guarantee inclusion.

Funding. This special issue originated from discussions held during the 'Uncertainty Quantification for Healthcare and Biological Systems' workshop at the Lorentz Center in the Netherlands. LMP was supported by the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom, grant reference number EP/T017899/1. MJC was supported in part through TL1 TR001415 through the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health (NIH). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

References

1. Dale KI, Pope ECD, Hopkinson AR, McCaie T, Lowe JA. 2023 Environment-Aware Digital Twins: Incorporating Weather and Climate Information to Support Risk-Based Decision-Making. *Artif. Intell. Earth Syst.* **2**, e230023. (doi:10.1175/aies-d-23-0023.1)
2. Williamson D, Blaker AT, Hampton C, Salter J. 2015 Identifying and removing structural biases in climate models with history matching. *Clim. Dyn.* **45**, 1299–1324. (doi:10.1007/s00382-014-2378-z)
3. Rougier J. 2007 Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations. *Clim. Chang.* **81**, 247–264. (doi:10.1007/s10584-006-9156-9)
4. Stuckner J, Piekenbrock M, Arnold SM, Ricks TM. 2021 Optimal experimental design with fast neural network surrogate models. *Comput. Mater. Sci.* **200**, 110747. (doi:10.1016/j.commatsci.2021.110747)
5. Paun LM *et al.* 2024 SECRET: Statistical Emulation for Computational Reverse Engineering and Translation with applications in healthcare. *Comput. Methods Appl. Mech. Eng.* **430**, 117193. (doi:10.1016/j.cma.2024.117193)
6. Madireddy S, Sista B, Vemaganti K. 2015 A Bayesian approach to selecting hyperelastic constitutive models of soft tissue. *Comput. Methods Appl. Mech. Eng.* **291**, 102–122. (doi:10.1016/j.cma.2015.03.012)
7. Lima EABF, Faghihi D, Philley R, Yang J, Virostko J, Phillips CM, Yankeelov TE. 2021 Bayesian calibration of a stochastic, multiscale agent-based model for predicting in vitro tumor growth. *PLoS Comput. Biol.* **17**, e1008845. (doi:10.1371/journal.pcbi.1008845)

8. James LP, Salomon JA, Buckee CO, Menzies NA. 2021 The Use and Misuse of Mathematical Modeling for Infectious Disease Policymaking: Lessons for the COVID-19 Pandemic. *Med. Decis. Mak.* **41**, 379–385. (doi:[10.1177/0272989x21990391](https://doi.org/10.1177/0272989x21990391))
9. Gnanvi JE, Salako KV, Kotanmi GB, Kakai RG. 2021 On the reliability of predictions on Covid-19 dynamics: A systematic and critical review of modelling techniques. *Infect. Dis. Model.* **6**, 258–272. (doi:[10.1016/j.idm.2020.12.008](https://doi.org/10.1016/j.idm.2020.12.008))
10. Jit M *et al.* 2023 Reflections On Epidemiological Modeling To Inform Policy During The COVID-19 Pandemic In Western Europe, 2020–23. *Health Aff.* **42**, 1630–1636. (doi:[10.1377/hlthaff.2023.00688](https://doi.org/10.1377/hlthaff.2023.00688))
11. Paun LM, Colebank MJ, Olufsen MS, Hill NA, Husmeier D. 2020 Assessing model mismatch and model selection in a Bayesian uncertainty quantification analysis of a fluid-dynamics model of pulmonary blood circulation. *J. R. Soc. Interface* **17**, 20200886. (doi:[10.1098/rsif.2020.0886](https://doi.org/10.1098/rsif.2020.0886))
12. Goldstein M, Huntley N. 2017 Bayes linear emulation, history matching, and forecasting for complex computer simulators. In *Handbook of uncertainty quantification*, pp. 9–32. Cham, Switzerland: Springer International Publishing. (doi:[10.1007/978-3-319-11259-6_14-1](https://doi.org/10.1007/978-3-319-11259-6_14-1))
13. Rossi RL, Grifantini RM. 2018 Big Data: Challenge and Opportunity for Translational and Industrial Research in Healthcare. *Front. Digit. Humanit.* **5**. (doi:[10.3389/fdigh.2018.00013](https://doi.org/10.3389/fdigh.2018.00013))
14. Colebank MJ, Oomen PA, Witzenburg CM, Grosberg A, Beard DA, Husmeier D, Olufsen MS, Chesler NC. 2024 Guidelines for mechanistic modeling and analysis in cardiovascular research. *Am. J. Physiol. Heart Circ. Physiol.* **327**, H473–H503. (doi:[10.1152/ajpheart.00766.2023](https://doi.org/10.1152/ajpheart.00766.2023))
15. Corral-Acero J *et al.* 2020 The ‘Digital Twin’ to enable the vision of precision cardiology. *Eur. Heart J.* **41**, 4556–4564. (doi:[10.1093/eurheartj/ehaa159](https://doi.org/10.1093/eurheartj/ehaa159))
16. Negri E, Fumagalli L, Macchi M. 2017 A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manuf.* **11**, 1. (doi:[10.1016/j.promfg.2017.07.198](https://doi.org/10.1016/j.promfg.2017.07.198))
17. Casola L, National Academy of Engineering, National Academies of Sciences, Engineering, and Medicine; Division on Engineering and Physical Sciences, Division on Earth and Life Studies, Board on Mathematical Sciences and Analytics, Computer Science and Telecommunications Board, Board on Atmospheric Sciences and Climate, Board on Life Sciences. 2023 *Opportunities and Challenges for Digital Twins in Biomedical Research*. Technical report, National Academies of Science.
18. Niederer SA, Sacks MS, Girolami M, Willcox K. 2021 Scaling digital twins from the artisanal to the industrial. *Nat. Comput. Sci.* **1**, 313–320. (doi:[10.1038/s43588-021-00072-5](https://doi.org/10.1038/s43588-021-00072-5))
19. Wangersky PJ. 1978 Lotka-Volterra Population Models. *Annu. Rev. Ecol. Syst.* **9**, 189–218. (doi:[10.1146/annurev.es.09.110178.001201](https://doi.org/10.1146/annurev.es.09.110178.001201))
20. van Vuuren DP *et al.* 2011 The representative concentration pathways: an overview. *Clim. Chang.* **109**, 5–31. (doi:[10.1007/s10584-011-0148-z](https://doi.org/10.1007/s10584-011-0148-z))
21. Gramacy RB, Lee HKH. 2008 Bayesian Treed Gaussian Process Models With an Application to Computer Modeling. *J. Am. Stat. Assoc.* **103**, 1119–1130. (doi:[10.1198/016214508000000689](https://doi.org/10.1198/016214508000000689))
22. Pamadi B, Covell P, Tartabini P, Murphy K. 2004 Aerodynamic Characteristics and Glide-Back Performance of Langley Glide-Back Booster. In *22nd Applied Aerodynamics Conference and Exhibit*, Providence, Rhode Island. Reston, Virginia. (doi:[10.2514/6.2004-5382](https://doi.org/10.2514/6.2004-5382)). <https://arc.aiaa.org/doi/book/10.2514/MAAC04>.
23. Rutarindwa R, Spiller ET, Bevilacqua A, Bursik MI, Patra AK. 2019 Dynamic Probabilistic Hazard Mapping in the Long Valley Volcanic Region CA: Integrating Vent Opening Maps and Statistical Surrogates of Physical Models of Pyroclastic Density Currents. *J. Geophys. Res.* **124**, 9600–9621. (doi:[10.1029/2019jb017352](https://doi.org/10.1029/2019jb017352))
24. O’Neill BC *et al.* 2016 The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.* **9**, 3461–3482. (doi:[10.5194/gmd-9-3461-2016](https://doi.org/10.5194/gmd-9-3461-2016))
25. Schleussner CF *et al.* 2016 Science and policy characteristics of the Paris Agreement temperature goal. *Nat. Clim. Chang.* **6**, 827–835. (doi:[10.1038/nclimate3096](https://doi.org/10.1038/nclimate3096))
26. Rosen R, von Wichert G, Lo G, Bettenhausen KD. 2015 About The Importance of Autonomy and Digital Twins for the Future of Manufacturing. *IFAC Pap.* **48**, 567–572. (doi:[10.1016/j.ifacol.2015.06.141](https://doi.org/10.1016/j.ifacol.2015.06.141))

27. Dodwell TJ, Kynaston S, Butler R, Haftka RT, Kim NH, Scheichl R. 2021 Multilevel Monte Carlo simulations of composite structures with uncertain manufacturing defects. *Probabilistic Eng. Mech.* **63**, 103116. (doi:10.1016/j.probengmech.2020.103116)
28. Grieves M, Vickers J. 2017 Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary perspectives on complex systems: new findings and approaches* (ed. FJ Kahlen), pp. 85–113. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-319-38756-7_4)
29. Hoffmann J, Bauer P, Sandu I, Wedi N, Geenen T, Thiemert D. 2023 Destination Earth – A digital twin in support of climate services. *Clim. Serv.* **30**, 100394. (doi:10.1016/j.cliser.2023.100394)
30. Fernández-Godino MG. 2023 Review of multi-fidelity models. *Adv. Comput. Sci. Eng.* **1**, 351–400. (doi:10.3934/acse.2023015)
31. Huang E, Xu J, Zhang S, Chen CH. 2015 Multi-fidelity Model Integration for Engineering Design. *Procedia Comput. Sci.* **44**, 336–344. (doi:10.1016/j.procs.2015.03.002)
32. Peherstorfer B, Willcox K, Gunzburger M. 2018 Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. *SIAM Rev.* **60**, 550–591. (doi:10.1137/16m1082469)
33. Cobelli C, Kovatchev B. 2023 Developing the UVA/Padova Type 1 Diabetes Simulator: Modeling, Validation, Refinements, and Utility. *J. Diabetes Sci. Technol.* **17**, 1493–1505. (doi:10.1177/19322968231195081)
34. Yang S *et al.* 2024 Long-term prognostic implications of CT angiography-derived fractional flow reserve: Results from the discover-flow study. *J. Cardiovasc. Comput. Tomogr.* **18**, 251–258. (doi:10.1016/j.jcct.2024.01.016)
35. Morrison TM, Pathmanathan P, Adwan M, Margerrison E. 2018 Advancing Regulatory Science With Computational Modeling for Medical Devices at the FDA's Office of Science and Engineering Laboratories. *Front. Med.* **5**. (doi:10.3389/fmed.2018.00241)
36. Galappaththige S, Gray RA, Costa CM, Niederer S, Pathmanathan P. 2022 Credibility assessment of patient-specific computational modeling using patient-specific cardiac modeling as an exemplar. *PLoS Comput. Biol.* **18**, e1010541. (doi:10.1371/journal.pcbi.1010541)
37. Pathmanathan P, Aycock K, Badal A, Bighamian R, Bodner J, Craven BA, Niederer S. 2024 Credibility assessment of in silico clinical trials for medical devices. *PLoS Comput. Biol.* **20**, e1012289. (doi:10.1371/journal.pcbi.1012289)
38. Linka K, Peirlinck M, Sahli Costabal F, Kuhl E. 2020 Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Comput. Methods Biomech. Biomed. Eng.* **23**, 710–717. (doi:10.1080/10255842.2020.1759560)
39. Anastassopoulou C, Russo L, Tsakris A, Siettos C. 2020 Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One* **15**, e0230405. (doi:10.1371/journal.pone.0230405)
40. Chen TM, Rui J, Wang QP, Zhao ZY, Cui JA, Yin L. 2020 A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infect. Dis. Poverty* **9**, 24. (doi:10.1186/s40249-020-00640-3)
41. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S. 2020 *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. Imperial College London.
42. Edeling W *et al.* 2021 The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nat. Comput. Sci.* **1**, 128–135. (doi:10.1038/s43588-021-00028-9)
43. Brook BS, Kohl P, King JR. 2011 Towards the virtual physiological human: mathematical and computational case studies. *Phil. Trans. R. Soc. A* **369**, 4145–4148. (doi:10.1098/rsta.2011.0278)
44. Viceconti M, De Vos M, Mellone S, Geris L. 2024 Position Paper From the Digital Twins in Healthcare to the Virtual Human Twin: A Moon-Shot Project for Digital Health Research. *IEEE J. Biomed. Health Informatics* **28**, 491–501. (doi:10.1109/jbhi.2023.3323688)
45. Chaudhuri A, Pash G, Hormuth DA II, Lorenzo G, Kapteyn M, Wu C, Lima EABF, Yankeelov TE, Willcox K. 2023 Predictive digital twin for optimizing patient-specific radiotherapy regimens under uncertainty in high-grade gliomas. *Front. Artif. Intell.* **6**, 1222612. (doi:10.3389/frai.2023.1222612)

46. Huberts W, Heinen SGH, Zonnebeld N, van den Heuvel DAF, de Vries JPPM, Tordoir JHM, Hose DR, Delhaas T, van de Vosse FN. 2018 What is needed to make cardiovascular models suitable for clinical decision support? A viewpoint paper. *J. Comput. Sci.* **24**, 68–84. (doi:10.1016/j.jocs.2017.07.006)
47. Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. 2013 Bayesian Networks for Clinical Decision Support in Lung Cancer Care. *PLoS One* **8**, e82349. (doi:10.1371/journal.pone.0082349)
48. Walker M, Moore H, Ataya A, Pham A, Corris PA, Laubenbacher R, Bryant AJ. 2024 A perfectly imperfect engine: Utilizing the digital twin paradigm in pulmonary hypertension. *Pulm. Circ.* **14**, e12392. (doi:10.1002/pul2.12392)
49. Laubenbacher R, Sluka JP, Glazier JA. 2021 Using digital twins in viral infection. *Science* **371**, 1105–1106. (doi:10.1126/science.abf3370)
50. Laubenbacher R *et al.* 2022 Building digital twins of the human immune system: toward a roadmap. *NPJ Digit. Med.* **5**, 64. (doi:10.1038/s41746-022-00610-z)
51. Laubenbacher R *et al.* 2024 Forum on immune digital twins: a meeting report. *NPJ Syst. Biol. Appl.* **10**, 19. (doi:10.1038/s41540-024-00345-5)
52. Linden NJ, Kramer B, Rangamani P. 2022 Bayesian parameter estimation for dynamical models in systems biology. *PLoS Comput. Biol.* **18**, e1010651. (doi:10.1371/journal.pcbi.1010651)
53. Seoni S, Jahmunah V, Salvi M, Barua PD, Molinari F, Acharya UR. 2023 Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023). *Comput. Biol. Med.* **165**, 107441. (doi:10.1016/j.combiomed.2023.107441)
54. Coveney PV, Groen D, Hoekstra AG. 2021 Reliability and reproducibility in computational science: implementing validation, verification and uncertainty quantification *in silico*. *Phil. Trans. R. Soc. A* **379**, 20200409. (doi:10.1098/rsta.2020.0409)
55. Mirams GR, Niederer SA, Clayton RH. 2020 The fickle heart: uncertainty quantification in cardiac and cardiovascular modelling and simulation. *Phil. Trans. R. Soc. A* **378**, 20200119. (doi:10.1098/rsta.2020.0119)
56. Tepole AB, Nordsletten D, Garikipati K, Kuhl E. 2020 Special Issue on Uncertainty Quantification, Machine Learning, and Data-Driven Modeling of Biological Systems. *Comput. Methods Appl. Mech. Eng.* **362**, 112832. (doi:10.1016/j.cma.2020.112832)
57. Morrison RE, Cunha A. 2020 Embedded model discrepancy: A case study of Zika modeling. *Chaos* **30**, 051103. (doi:10.1063/5.0005204)
58. Li Z *et al.* 2020 General Principles for the Validation of Proarrhythmia Risk Prediction Models: An Extension of the CiPA In Silico Strategy. *Clin. Pharmacol. Ther.* **107**, 102–111. (doi:10.1002/cpt.1647)
59. Haylock RG, O'Hagan A. 1996 On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. In *Bayesian statistics 5*, pp. 629–638. Oxford, UK: Oxford University Press. (doi:10.1093/oso/9780198523567.003.0041)
60. O'Hagan A. 2006 Bayesian analysis of computer code outputs: A tutorial. *Reliab. Eng. Syst. Saf.* **91**, 1290–1300. (doi:10.1016/j.res.2005.11.025)
61. Rasmussen CE, Williams CKI. 2006 *Gaussian processes for machine learning*, 1st edn. Cambridge: MIT Press. (doi:10.7551/mitpress/3206.001.0001)
62. Owen NE, Challenor P, Menon PP, Bennani S. 2017 Comparison of Surrogate-Based Uncertainty Quantification Methods for Computationally Expensive Simulators. *SIAM/ASA J. Uncertain. Quantif.* **5**, 403–435. (doi:10.1137/15m1046812)
63. Najm HN. 2009 Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annu. Rev. Fluid Mech.* **41**, 35–52. (doi:10.1146/annurev.fluid.010908.165248)
64. Grandits T *et al.* 2024 Neural network emulation of the human ventricular cardiomyocyte action potential for more efficient computations in pharmacological studies. *eLife* **12**, P91911. (doi:10.7554/elife.91911)
65. Dalton D, Husmeier D, Gao H. 2023 Physics-informed graph neural network emulation of soft-tissue mechanics. *Comput. Methods Appl. Mech. Eng.* **417**, 116351. (doi:10.1016/j.cma.2023.116351)
66. Pang G, Karniadakis GE. 2020 Physics-Informed Learning Machines for Partial Differential Equations: Gaussian Processes Versus Neural Networks. In *Emerging frontiers in nonlinear*

- science nonlinear systems and complexity*, pp. 323–343. Cham, Switzerland: Springer International Publishing. (doi:[10.1007/978-3-030-44992-6_14](https://doi.org/10.1007/978-3-030-44992-6_14))
67. Mann J, Meshkin H, Zirkle J, Han X, Thrasher B, Chaturbedi A, Arabidarrehdor G, Li Z. 2024 Mechanism-based organization of neural networks to emulate systems biology and pharmacology models. *Sci. Rep.* **14**, 12082. (doi:[10.1038/s41598-024-59378-9](https://doi.org/10.1038/s41598-024-59378-9))
 68. Currin C, Mitchell T, Morris M, Ylvisaker D. 1991 Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. *J. Am. Stat. Assoc.* **86**, 95363. (doi:[10.2307/2290511](https://doi.org/10.2307/2290511))
 69. Santner TJ, Williams BJ, Notz WI. 2018 *The design and analysis of computer experiments*, 1st edn. New York: Springer New York. (doi:[10.1007/978-1-4939-8847-1_1](https://doi.org/10.1007/978-1-4939-8847-1_1))
 70. Joseph VR. 2016 Space-filling designs for computer experiments: A review. *Qual. Eng.* **28**, 28–35. (doi:[10.1080/08982112.2015.1100447](https://doi.org/10.1080/08982112.2015.1100447))
 71. Novák L, Vořechovský M, Sadílek V, Shields MD. 2021 Variance-based adaptive sequential sampling for Polynomial Chaos Expansion. *Comput. Methods Appl. Mech. Eng.* **386**, 114105. (doi:[10.1016/j.cma.2021.114105](https://doi.org/10.1016/j.cma.2021.114105))
 72. McKay MD, Beckman RJ, Conover WJ. 1979 Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21**, 239–245. (doi:[10.1080/00401706.1979.10489755](https://doi.org/10.1080/00401706.1979.10489755))
 73. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. 1989 Design and Analysis of Computer Experiments. *Stat. Sci.* **4**, 409–423. (doi:[10.1126/science.103.2667.174-c](https://doi.org/10.1126/science.103.2667.174-c))
 74. Fajraoui N, Marelli S, Sudret B. 2017 Sequential Design of Experiment for Sparse Polynomial Chaos Expansions. *SIAM/ASA J. Uncertain. Quantif.* **5**, 1061–1085. (doi:[10.1137/16m1103488](https://doi.org/10.1137/16m1103488))
 75. Smith RC. 2013 *Uncertainty quantification: theory, implementation, and applications*. Philadelphia, PA: SIAM. (doi:[10.1137/1.9781611973228](https://doi.org/10.1137/1.9781611973228))
 76. Higdon D, Gattiker J, Williams B, Rightley M. 2008 Computer Model Calibration Using High-Dimensional Output. *J. Am. Stat. Assoc.* **103**, 570–583. (doi:[10.1198/016214507000000888](https://doi.org/10.1198/016214507000000888))
 77. Chen P, Quarteroni A, Rozza G. 2017 Reduced Basis Methods for Uncertainty Quantification. *SIAM/ASA J. Uncertain. Quantif.* **5**, 813–869. (doi:[10.1137/151004550](https://doi.org/10.1137/151004550))
 78. Constantine PG, Dow E, Wang Q. 2014 Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM J. Sci. Comput.* **36**, A1500–A1524. (doi:[10.1137/130916138](https://doi.org/10.1137/130916138))
 79. Salter JM, Williamson DB, Scinocca J, Kharin V. 2019 Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases. *J. Am. Stat. Assoc.* **114**, 1800–1814. (doi:[10.1080/01621459.2018.1514306](https://doi.org/10.1080/01621459.2018.1514306))
 80. Hijazi S, Freitag M, Landwehr N. 2023 POD-Galerkin reduced order models and physics-informed neural networks for solving inverse problems for the Navier–Stokes equations. *Adv. Model. Simul. Eng. Sci.* **10**, 5. (doi:[10.1186/s40323-023-00242-2](https://doi.org/10.1186/s40323-023-00242-2))
 81. Sexton DMH, Murphy JM, Collins M, Webb MJ. 2012 Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Clim. Dyn.* **38**, 2513–2542. (doi:[10.1007/s00382-011-1208-9](https://doi.org/10.1007/s00382-011-1208-9))
 82. Kennedy MC, O'Hagan A. 2000 Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13. (doi:[10.1093/biomet/87.1.1](https://doi.org/10.1093/biomet/87.1.1))
 83. Ye D, Veen L, Nikishova A, Lakhilil J, Edeling W, Luk OO, Krzhizhanovskaya VV, Hoekstra AG. 2021 Uncertainty quantification patterns for multiscale models. *Phil. Trans. R. Soc. A* **379**, 20200072. (doi:[10.1098/rsta.2020.0072](https://doi.org/10.1098/rsta.2020.0072))
 84. Binois M, Gramacy RB, Ludkovski M. 2018 Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments. *J. Comput. Graph. Stat.* **27**, 808–821. (doi:[10.1080/10618600.2018.1458625](https://doi.org/10.1080/10618600.2018.1458625))
 85. Hofer E. 1999 Sensitivity analysis in the context of uncertainty analysis for computationally intensive models. *Comput. Phys. Commun.* **117**, 21–34. (doi:[10.1016/s0010-4655\(98\)00153-2](https://doi.org/10.1016/s0010-4655(98)00153-2))
 86. Saltelli A, Aleksankina K, Becker W, Fennell P, Ferretti F, Holst N, Li S, Wu Q. 2019 Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environ. Model. Softw.* **114**, 29–39. (doi:[10.1016/j.envsoft.2019.01.012](https://doi.org/10.1016/j.envsoft.2019.01.012))
 87. Saltelli A. 2008 *Global Sensitivity Analysis: the primer*, 1 edn. Hoboken, NJ: John Wiley, Sons Inc. (doi:[10.1002/9780470725184](https://doi.org/10.1002/9780470725184))

88. Saltelli A, Tarantola S, Campolongo F, Ratto M. 2002 *Sensitivity analysis in practice: a guide to assessing scientific models*, 1st edn. Hoboken, NJ: John Wiley and Sons Inc. (doi:10.1002/0470870958)
89. Colunga AL, Kim KG, Woodall NP, Dardas TF, Gennari JH, Olufsen MS, Carlson BE. 2020 Deep phenotyping of cardiac function in heart transplant patients using cardiovascular system models. *J. Physiol.* **598**, 3203–3222. (doi:10.1113/jp279393)
90. Sobol IM. 1993 Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414.
91. Xu C, Gertner G. 2011 Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). *Comput. Stat. Data Anal.* **55**, 20. (doi:10.1016/j.csda.2010.06.028)
92. Owen AB. 2014 Sobol' Indices and Shapley Value. *SIAM/ASA J. Uncertain. Quantif.* **2**, 245–251. (doi:10.1137/130936233)
93. Oakley JE, O'Hagan A. 2004 Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach. *J. R. Stat. Soc. Ser. B* **66**, 751–769. (doi:10.1111/j.1467-9868.2004.05304.x)
94. Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003 A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **39**. (doi:10.1029/2002wr001642)
95. Hourdin F *et al.* 2017 The Art and Science of Climate Model Tuning. *Bull. Am. Meteorol. Soc.* **98**, 589–602. (doi:10.1175/bams-d-15-00135.1)
96. Stuart AM. 2010 Inverse problems: A Bayesian perspective. *Acta Numer.* **19**, 451–559. (doi:10.1017/s0962492910000061)
97. Brynjarsdóttir J, O'Hagan A. 2014 Learning about physical parameters: the importance of model discrepancy. *Inverse Probl.* **30**, 114007. (doi:10.1088/0266-5611/30/11/114007)
98. Kennedy MC, O'Hagan A. 2001 Bayesian Calibration of Computer Models. *J. R. Stat. Soc.* **63**, 425–464.
99. Gardner P, Rogers TJ, Lord C, Barthorpe RJ. 2021 Learning model discrepancy: A Gaussian process and sampling-based approach. *Mech. Syst. Signal Process.* **152**, 107381. (doi:10.1016/j.ymssp.2020.107381)
100. Wong RKW, Storlie CB, Lee TCM. 2017 A Frequentist Approach to Computer Model Calibration. *J. R. Stat. Soc. Ser. B.* **79**, 635–648. (doi:10.1111/rssb.12182)
101. Andrianakis I, Vernon IR, McCreesh N, McKinley TJ, Oakley JE, Nsubuga RN, Goldstein M, White RG. 2015 Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda. *PLoS Comput. Biol.* **11**, e1003968. (doi:10.1371/journal.pcbi.1003968)
102. Craig PS, Goldstein M, Seheult AH, Smith JA. 1997 Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments. In *Case Studies in Bayesian Statistics*, pp. 37–93, vol. III. New York, NY: Springer. (doi:10.1007/978-1-4612-2290-3_2)
103. Goldstein M, Rougier J. 2006 Bayes Linear Calibrated Prediction for Complex Systems. *J. Am. Stat. Assoc.* **101**, 1132–1143. (doi:10.1198/016214506000000203)
104. Salter JM. 2017 Uncertainty Quantification for Spatial Field Data Using Expensive Computer Models: Refocused Bayesian Calibration with Optimal Projection. University of Exeter (United Kingdom).
105. Nolte D, Bertoglio C. 2022 Inverse problems in blood flow modeling: A review. *Int. J. Numer. Methods Biomed. Eng.* **38**, e3613. (doi:10.1002/cnm.3613)
106. Viceconti M, Pappalardo F, Rodriguez B, Horner M, Bischoff J, Musuamba Tshinanu F. 2021 In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. *Methods* **185**, 120–127. (doi:10.1016/j.ymeth.2020.01.011)
107. Hose DR, Lawford PV, Huberts W, Hellevik LR, Omholt SW, van de Vosse FN. 2019 Cardiovascular models for personalised medicine: Where now and where next? *Med. Eng. Phys.* **72**, 38–48. (doi:10.1016/j.medengphy.2019.08.007)
108. Collin CB *et al.* 2022 Computational Models for Clinical Applications in Personalized Medicine—Guidelines and Recommendations for Data Integration and Model Validation. *J. Pers. Med.* **12**, 166. (doi:10.3390/jpm12020166)

109. Erdemir A *et al.* 2020 Credible practice of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective. *J. Transl. Med.* **18**, 369. (doi:10.1186/s12967-020-02540-4)
110. Lei CL *et al.* 2020 Considering discrepancy when calibrating a mechanistic electrophysiology model. *Phil. Trans. R. Soc. A* **378**, 20190349. (doi:10.1098/rsta.2019.0349)
111. Shuttleworth JG, Lei CL, Whittaker DG, Windley MJ, Hill AP, Preston SP, Mirams GR. 2024 Empirical Quantification of Predictive Uncertainty Due to Model Discrepancy by Training with an Ensemble of Experimental Designs: An Application to Ion Channel Kinetics. *Bull. Math. Biol.* **86**, 2. (doi:10.1007/s11538-023-01224-6)
112. Zou Z, Meng X, Karniadakis GE. 2024 Correcting model misspecification in physics-informed neural networks (PINNs). *J. Comput. Phys.* **505**, 112918. (doi:10.1016/j.jcp.2024.112918)
113. Lei CL, Mirams GR. 2021 Neural Network Differential Equations For Ion Channel Modelling. *Front. Physiol.* **12**, 708944. (doi:10.3389/fphys.2021.708944)
114. Renardy M, Hult C, Evans S, Linderman JJ, Kirschner DE. 2019 Global sensitivity analysis of biological multiscale models. *Curr. Opin. Biomed. Eng.* **11**, 109–116. (doi:10.1016/j.cobme.2019.09.012)
115. Schiavazzi DE, Arbia G, Baker C, Hlavacek AM, Hsia TY, Marsden AL, Vignon - Clementel IE. 2016 Uncertainty quantification in virtual surgery hemodynamics predictions for single ventricle palliation. *Int. J. Numer. Method. Biomed. Eng.* **32**, e02737. (doi:10.1002/cnm.2737)
116. Longobardi S, Sher A, Niederer SA. 2021 In silico identification of potential calcium dynamics and sarcomere targets for recovering left ventricular function in rat heart failure with preserved ejection fraction. *PLOS Comput. Biol.* **17**, e1009646. (doi:10.1371/journal.pcbi.1009646)
117. Lazarus A, Gao H, Luo X, Husmeier D. 2022 Improving Cardio-Mechanic Inference by Combining in Vivo Strain Data with Ex Vivo Volume–Pressure Data. *J. R. Stat. Soc. Ser. C* **71**, 906–931. (doi:10.1111/rssc.12560)
118. McClarren RG, McClarren P, Penrose R. 2018 *Uncertainty quantification and predictive computational science*. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-319-99525-0_1)
119. Alber M *et al.* 2019 Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digit. Med.* **2**, 115. (doi:10.1038/s41746-019-0193-y)
120. Yazdani A, Lu L, Raissi M, Karniadakis GE. 2020 Systems biology informed deep learning for inferring parameters and hidden dynamics. *PLOS Comput. Biol.* **16**, e1007575. (doi:10.1371/journal.pcbi.1007575)
121. Coveney S, Clayton RH. 2018 Fitting two human atrial cell models to experimental data using Bayesian history matching. *Prog. Biophys. Mol. Biol.* **139**, 43–58. (doi:10.1016/j.pbiomolbio.2018.08.001)
122. Mester R, Landeros A, Rackauckas C, Lange K. 2022 Differential methods for assessing sensitivity in biological models. *PLoS Comput. Biol.* **18**, e1009598. (doi:10.1371/journal.pcbi.1009598)
123. Wieland FG, Hauber AL, Rosenblatt M, Tönsing C, Timmer J. 2021 On structural and practical identifiability. *Curr. Opin. Syst. Biol.* **25**, 60–69. (doi:10.1016/j.coisb.2021.03.005)
124. Chis OT, Banga JR, Balsa-Canto E. 2011 Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. *PLoS ONE* **6**, e27755. (doi:10.1371/journal.pone.0027755)
125. Borowska A, Giurghita D, Husmeier D. 2021 Gaussian process enhanced semi-automatic approximate Bayesian computation: parameter inference in a stochastic differential equation system for chemotaxis. *J. Comput. Phys.* **429**, 109999. (doi:10.1016/j.jcp.2020.109999)
126. Price LF, Drovandi CC, Lee A, Nott DJ. 2018 Bayesian Synthetic Likelihood. *J. Comput. Graph. Stat.* **27**, 1–11. (doi:10.1080/10618600.2017.1302882)
127. Fleeter CM, Geraci G, Schiavazzi DE, Kahn AM, Marsden AL. 2020 Multilevel and multifidelity uncertainty quantification for cardiovascular hemodynamics. *Comput. Methods Appl. Mech. Eng.* **365**, 113030. (doi:10.1016/j.cma.2020.113030)
128. Keeling MJ, Eames KTD. 2005 Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307. (doi:10.1098/rsif.2005.0051)

129. Randall EB, Randolph NZ, Alexanderian A, Olufsen MS. 2021 Global sensitivity analysis informed model reduction and selection applied to a Valsalva maneuver model. *J. Theor. Biol.* **526**, 110759. (doi:[10.1016/j.jtbi.2021.110759](https://doi.org/10.1016/j.jtbi.2021.110759))
130. Aldoseri A, Al-Khalifa KN, Hamouda AM. 2023 Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Appl. Sci.* **13**, 7082. (doi:[10.3390/app13127082](https://doi.org/10.3390/app13127082))
131. Awrahman BJ, Aziz Fatah C, Hamaamin MY. 2022 A Review of the Role and Challenges of Big Data in Healthcare Informatics and Analytics. *Comput. Intell. Neurosci.* **2022**, 1–10. (doi:[10.1155/2022/5317760](https://doi.org/10.1155/2022/5317760))
132. Feng L, Gao H, Luo X. 2024 Whole-heart modelling with valves in a fluid–structure interaction framework. *Comput. Methods Appl. Mech. Eng.* **420**, 116724. (doi:[10.1016/j.cma.2023.116724](https://doi.org/10.1016/j.cma.2023.116724))
133. Paun LM, Husmeier D. 2022 Emulation-accelerated Hamiltonian Monte Carlo algorithms for parameter estimation and uncertainty quantification in differential equation models. *Stat. Comput.* **32**, 25. (doi:[10.1007/s11222-021-10060-4](https://doi.org/10.1007/s11222-021-10060-4))
134. Vernon I *et al.* 2022 Bayesian emulation and history matching of JUNE. *Phil. Trans. R. Soc. A* **380**, 20220039. (doi:[10.1098/rsta.2022.0039](https://doi.org/10.1098/rsta.2022.0039))
135. Ahmadi Daryakenari N, De Florio M, Shukla K, Karniadakis GE. 2024 AI-Aristotle: A physics-informed framework for systems biology gray-box identification. *PLoS Comput. Biol.* **20**, e1011916. (doi:[10.1371/journal.pcbi.1011916](https://doi.org/10.1371/journal.pcbi.1011916))
136. Penwarden M, Zhe S, Narayan A, Kirby RM. 2022 Multifidelity modeling for Physics-Informed Neural Networks (PINNs). *J. Comput. Phys.* **451**, 110844. (doi:[10.1016/j.jcp.2021.110844](https://doi.org/10.1016/j.jcp.2021.110844))
137. Qian S *et al.* 2024 Developing Cardiac Digital Twins at Scale: Insights from Personalised Myocardial Conduction Velocity. *medRxiv*.
138. Bruynseels K, Santoni de Sio F, van den Hoven J. 2018 Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Front. Genet.* **9**. (doi:[10.3389/fgene.2018.00031](https://doi.org/10.3389/fgene.2018.00031))
139. Freitas AT. 2024 Data-Driven Approaches in Healthcare: Challenges and Emerging Trends. In *Multidisciplinary perspectives on artificial intelligence and the law*, pp. 65–80. Cham: Springer International Publishing. (doi:[10.1007/978-3-031-41264-6_4](https://doi.org/10.1007/978-3-031-41264-6_4))
140. Johnson SB, Slade I, Giubilini A, Graham M. 2020 Rethinking the ethical principles of genomic medicine services. *Eur. J. Hum. Genet.* **28**, 147–154. (doi:[10.1038/s41431-019-0507-1](https://doi.org/10.1038/s41431-019-0507-1))
141. Tsaneva-Atanasova K, Pederzanil G, Laviola M. 2025 Decoding uncertainty for clinical decision-making. *Phil. Trans. R. Soc. A* **383**, 20240207. (doi:[10.1098/rsta.2024.0207](https://doi.org/10.1098/rsta.2024.0207))
142. Böttcher L, Fonseca LL, Laubenbacher RC. 2024 Control of Medical Digital Twins with Artificial Neural Networks. *bioRxiv* 2024.03.18.585589. (doi:[10.1101/2024.03.18.585589](https://doi.org/10.1101/2024.03.18.585589))
143. De Florio M, Zou Z, Schiavazzi DE, Karniadakis GE. 2025 Quantification of total uncertainty in the physics-informed reconstruction of CVSim-6 physiology. *Phil. Trans. R. Soc. A* **383**, 20240221. (doi:[10.1098/rsta.2024.0221](https://doi.org/10.1098/rsta.2024.0221))
144. Paun LM, Colebank MJ, Husmeier D. 2025 A comparison of Gaussian processes and polynomial chaos emulators in the context of haemodynamic pulse-wave propagation modelling. *Phil. Trans. R. Soc. A* **383**, 20240222. (doi:[10.1098/rsta.2024.0222](https://doi.org/10.1098/rsta.2024.0222))
145. Salter JM, McKinley TJ, Xiong X, Williamson D. 2025 Emulating computer models with high-dimensional count output. *Phil. Trans. R. Soc. A* **383**, 20240216. (doi:[10.1098/rsta.2024.0216](https://doi.org/10.1098/rsta.2024.0216))
146. Bajaj S, Thompson R, Lambert B. 2025 A renewal-equation approach to estimating R_t and infectious disease case counts in the presence of reporting delays. *Phil. Trans. R. Soc. A* **383**, 20240357. (doi:[10.1098/rsta.2024.0357](https://doi.org/10.1098/rsta.2024.0357))
147. Richter J, Nitzler J, Pegolotti L, Menon K, Biehler J, Wall WA, Schiavazzi DE, Marsden AL, Pfaller MR. 2025 Bayesian Windkessel calibration using optimized zero-dimensional surrogate models. *Phil. Trans. R. Soc. A* **383**, 20240223. (doi:[10.1098/rsta.2024.0223](https://doi.org/10.1098/rsta.2024.0223))
148. Dadashova K, Smith R, Haider M, Reich B. 2025 Bayesian inference informed by parameter subset selection for a minimal PBPB brain model. *Phil. Trans. R. Soc. A* **383**, 20240219. (doi:[10.1098/rsta.2024.0219](https://doi.org/10.1098/rsta.2024.0219))

149. Del Core L, Mirams GR. 2025 Parameter inference for stochastic reaction models of ion channel gating from whole-cell voltage-clamp data. *Phil. Trans. R. Soc. A* **383**, 20240224. (doi:[10.1098/rsta.2024.0224](https://doi.org/10.1098/rsta.2024.0224))
150. Shuttleworth JG, Lei CL, Windley MJ, Hill AP, Preston SP, Mirams GR. 2025 Evaluating the predictive accuracy of ion channel models using data from multiple experimental designs. *Biophysics*. (doi:[10.1101/2024.08.16.608289](https://doi.org/10.1101/2024.08.16.608289))
151. Patten-Elliott F, Lei CL, Preston SP, Wilkinson RD, Mirams GR. 2025 Optimizing experimental designs for model selection of ion channel drug-binding mechanisms. *Phil. Trans. R. Soc. A* **383**, 20240227. (doi:[10.1098/rsta.2024.0227](https://doi.org/10.1098/rsta.2024.0227))
152. Rabbani A, Sadeghkhanian A, Holland A, Besharat M, Fang H, Babaei M, Barrera O. 2025 Structure–property relationships in fibrous meniscal tissue through image-based augmentation. *Phil. Trans. R. Soc. A* **383**, 20240225. (doi:[10.1098/rsta.2024.0225](https://doi.org/10.1098/rsta.2024.0225))