Irving Fisher Committee on
Central Bank Statistics

BIS

12th biennial IFC Conference: "Statistics and beyond: new data for decision making in central banks"

22-23 August 2024

Is there a future for stochastic modeling in business and industry in the era of machine learning and artificial intelligence?[1]

Fabrizio Ruggeri,
President-elect, International Statistical Institute &
Istituto di Matematica Applicata e Tecnologie Informatiche

[1] This contribution was prepared for the conference. The views expressed in this publication are those of the authors and do not necessarily represent the official views of the Committee, its members, or the BIS.

Is there a future for traditional stochastic models (in business and industry) in the AI and ML era?

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Alfonso Corti 12, I-20133, Milano, Italy, European Union
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/fabrizio/

AN ASMBI COLLABORATIVE PAPERS

- The talk is based on a forthcoming collaborative paper by many researchers in the field who discuss, and illustrate through examples, when the ”traditional” stochastic models should be preferred to AI and ML methods and vice versa

- Actually, sentences from the paper are presented and commented

- The paper showcases examples from business and industry, since it will be published in *Applied Stochastic Models in Business and Industry*

- So far the contributors are:

  David Banks, Marcos Escobar, Nicholas Fisher/William Cleveland, Paolo Giudici, Roger Hoerl/Dennis Lin, Ron Kenett/Nalini Ravishanker/Marco Reis, Wai Keung Li/Philip Yu, Jean-Michel Poggi, Marco Reis, Gilbert Saporta, Piercesare Secchi, Rituparna Sen, Ansgar Steland, and Zhanpan Zhang

FEW QUOTES

- *We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot*.
  (Christopher Anderson, computer scientist)

- *All models are wrong, but some are useful.*
  (George P. Box)

- *All models are wrong, and increasingly you can succeed without them.*
  (Peter Norvig, computer scientist and former director of research at Google)

- *The goal of science is not to make predictions. It is also offering an image of reality, a conceptual framework for thinking about things. This ambition has made scientific thinking effective. If the goal of science were only predictions, Copernicus would not have discovered anything compared to Ptolemy: his astronomical predictions were no better than Ptolemy's. But Copernicus found a key to rethink everything and understand better.*
  (Carlo Rovelli, physicist, text translated from Italian by Piercesare Secchi)

SOME QUESTIONS

- Is there a future for traditional stochastic models (in business and industry) in the AI and ML era?
- Would stochastic modelling still be useful with the impressive performance of ML methods?
- Is it still justified to push research and teaching in stochastic modelling?
- What can stochastic modelling contribute to ML and vice versa?
- Are there areas where stochastic modelling performs better and others where ML does?
- How does AI affect (and is affected by) stochastic modelling and ML?
- What are ”traditional” and ”new” stochastic models? Regression vs. deep neural network?

SOME PRELIMINARY COMMENTS

- The advances in analytics under headings of AI, ML or deep learning have changed the approach to data analysis. Part of this change followed the big data, sensor technology and computational capabilities advances.

- A proper methodology for assessing the right balance between purely data driven empirical methods, physics-based models and probabilistic stochastic models is requiring further development. This is necessary for the effective and robust achievement of data driven knowledge.

- This requires substantial revisions in educational curricula where statistics and data science methods need to have stronger ties with physics and engineering, and rely on software such as `R, Python, JMP`.

- Papacharalampous et al (2019) compare forecast performances of the two approaches and found that ML methods do not differ dramatically from the stochastic, while none of the methods under comparison dominates the other. From a scientific point of view both tools would be valuable to an investigator who would like to seek a better understanding of his/her problem.

SOME PRELIMINARY COMMENTS

- In contrast to deterministic modelling, which many ML methods employ, a key advantage of statistical modelling is its ability to quantify uncertainty. This quantification is crucial for enhancing model reliability and supporting decision-making in real-world applications.

- In addition, statistical models generally offer better explainability compared to black-box ML methods. A thorough understanding of model behaviour enhances transparency and builds trust in high-risk domains such as healthcare and finance.

- Often, sample size is another important factor in determining the selection of appropriate modelling approaches. It is well known that ML, particularly deep learning methods, require a large amount of data to achieve satisfactory performance.

- When physical simulation is computationally slow and/or experimental data collection is costly, practitioners will need to explore alternative approaches. For example, Bayesian networks can be an effective method for addressing inverse design problems and time-dependent dynamic issues.

SOME PRELIMINARY COMMENTS

- If the question is predicting the occurrence of a future event, e.g., future conditional mean of a stochastic process, and there is a large dataset with many variables, then ML methods may be a first choice.

- However, in many investigations, predictions may not be the only objective and one would like to have more understanding of the underlying data generating process.

- It is true that all models may be only approximations at best but a good one should be able to give us insight into why the data are behaving in a certain way.

- In such cases, classical stochastic modelling may provide more insights into the data generating mechanism, especially if the number of observations is larger than the number of variables.

- The principles of randomization, replication and blocking need to be taken into account for any designed experiment. For sample surveys, it is extremely important to ensure that common sources of bias like selection, response/non-response etc. do not creep in. Given the particular scenario, stochastic modelling provides a guiding principle on data collection for the most efficient utilisation of resources geared towards answering the business and industry question at hand.

A WORLD IN EVOLUTION

- A simple example motivated by finance, but easily extrapolated to other areas, is the evolution of stock prices in the last 60 years.
- The price of a share is not only dictated by rational economic forces, but it is also a by-product of human individual and collective behaviour, desires, passion, wishes, and mistakes; all this summarised on a single number at any given time.
- The modelling of stocks had to adapt/evolve from a quite robust and simple Gaussian model in the 60's to, e.g., the inclusion of jumps in the 70's (Levy models), random local volatility as detected in the 80's (i.e., GARCH, CEV* models), stochastic volatility in the 90's motivated by the need to trade volatilities and the emergence of the volatility index (VIX), and more recently the presence of stochastic volatility of volatility and stochastic skewness on prices as captured by the new indexes VVIX and SKEW respectively. The history for multi-asset modelling is even richer.
- The increase in complexity will only continue, requiring more advanced models (based on ML and AI?) to properly explain the dynamics of single and multiple stocks.
- Would this mean the end of stochastic modelling?

*Constant elasticity of variance

STOCHASTIC MODELLING: EXAMPLES

- Many consider logistic regression a stochastic model since it produces individual or collective probabilities of default, but this is an abuse of the term, as it is very difficult to define a generative model of default, and it is actually an empirical approach.

- The method is the industry standard for credit scoring and fulfills regulatory requirements (at least in EU), thanks to the simplicity of its formulation and interpretation (e.g. on macroeconomic variables like GDP and interest rates affecting the corporate probability of default).

- Numerous attempts to use more complex models have not led to significant progress as quoted in, e.g., Bazzana et al (2024): *This paper uses a large sample of small Italian companies to compare the performance of various ML classifiers and a more traditional logistic regression approach. In particular, we perform feature selection, use the algorithms for default prediction, evaluate their accuracy, and find a more suitable threshold as a function of sensitivity and specificity. Our outcomes suggest that ML is slightly better than logistic regression. However, the relatively small performance gain is insufficient to conclude that classical statistical classifiers should be abandoned, as they are characterized by more straightforward interpretation and implementation.*

STOCHASTIC MODELLING: EXAMPLES

- Stochastic modelling is valuable when considering massive amount of data on individual electricity consumption provided by new metering technologies and smart grids, for load profiling and modelling at different scales of the electricity network.

- A methodology based on a mixture of high-dimensional regression models is used to perform clustering of individual customers. It accounts for forecasting, combined with the partitioning of the electrical signal into successive curves to consider it as functional data.

- The method extracts nice features from individual consumption data with little information (2 days) and no other prior, focusing on the discovery of clusters corresponding to different regression models, which could then be used directly for profiling, but can also be useful for forecasting purposes,

- The statistical approach allows a deeper analysis of the use of the internal objects of the method from a practical perspective, focusing not only on the results of the method but also on the by-products of the method, providing visualisation tools to understand the estimates and facilitate interpretation.

STOCHASTIC MODELLING: EXAMPLES

- In the context of business, there is great potential for Statistics in computational advertising.

- One way that stochastic processes arise is in contract fulfilment for showing online advertisements.

- When a user visits a website (e.g., cnn.com), it triggers a complex sequence of activity lasting less than ten milliseconds.

- There is a virtual auction among demand-side platforms who must decide

  1. whether to bid on showing an ad to the user
  2. which ad to show
  3. how much to bid

- Typically, the demand-side platform knows quite a lot about the user: gender, approximate age, approximate income, marital status, location, and previous purchase history

STOCHASTIC MODELLING: EXAMPLES

- The whole topic of survival analysis has been developed to handle censored and truncated observations efficiently. Without stochastic modelling such topics will remain completely out of reach.

- There are works about the use of deep learning (Wiegrebe et al, 2024) especially exploiting unstructured or high-dimensional data such as images, text or omics data, but not much exists about censored and truncated observations as confirmed by the regularly updated website `https://survival-org.github.io/DL4Survival/`

- The corporate probability of default is an important quantity for regulatory purposes as it measures how much risk a bank is taking overall. A ML algorithm will be useful in identifying which customers are more liable to default. But when the interest is in the overall probability of default, which is a population parameter, concepts of population and stochastic modelling naturally need to be taken into account.

STOCHASTIC MODELLING: EXAMPLES

- Even for point predictions of future time series the knowledge of the underlying probability model can provide many insights about interpretation and quality of the prediction.

- Wong and Li (2000) considered a three-component mixture autoregressive model for the first difference y_t of the IBM stock price data from 1961 to 1962 with 369 observations and easily obtained one-step ahead predictive distributions.

- It was observed that the predictive distributions for different time points would exhibit distinct bimodality when the volatility of y_t was high. In contrast unimodality was observed when the market was less volatile.

- In other words, the market would have higher chances of a sharp increase or decrease when volatility was high.

- In such a case a point forecast of the future time series would not be informative but knowledge of bimodality about the predictive distribution and the fitted mixture model would be useful to the investigator and risk manager.

MACHINE LEARNING

- The main advantage of ML methods is in prediction tasks and this is most common in regression, clustering, classification and time series or spatial forecasting settings.
- ML models are boosting AI applications in many domains, such as finance, healthcare, and automotive.
- This is mainly due to their advantage, in terms of predictive accuracy, with respect to "classic" statistical learning models.
- However, although complex ML models may reach high predictive performance, their predictions are not explainable, and have an intrinsic black-box nature.
- Furthermore, these models may not be robust, and may use data that are not representative, thereby generating biases and discriminations
- The success of advanced ML methods in areas such as image and face recognition is spectacular, although there are a few shortcomings in terms of robustness, where the modification of a single pixel can dramatically change the prediction, but the intelligibility of the models was hardly questioned until recently.

MACHINE LEARNING

- There are attempts to make the algorithms explainable: this is the XAI, with methods that seek to open the black box, with new measures of variable importance or the use of surrogate models to explain a decision using trees or local linear models.

- Some researchers are even calling for black boxes to be abandoned in favour of simple models.

- It should be noted that works on feature importance in ML, with Shapley measures for example, have renewed the classic problems which already showed that even simple models are not so easy to interpret.

- As well as being transparent, algorithms need to be fair and non-discriminatory when applied to human groups.

- Algorithmic fairness is a major area of development in computer science, but one in which statisticians are not yet very involved.

- Transparency or explicability are not enough: if a model is not causal, which is in line with a definition of stochastic modelling, and is based solely on correlations, it can lead to erroneous conclusions.

MACHINE LEARNING: EXAMPLES

- We shall someday have all vehicles on the road being networked and autonomous, and then there will be an ocean of travel data.

- We will not need differential equation models for traffic flow since we shall have the empirical process in great detail and the need for process modelling will be much less, although the stochastic process itself will remain important.

- Considering data on flows of networked vehicles, they are actually the superposition of observations from multiple and perhaps simpler data generation mechanisms, like daily commuters, long-haul trucking, school buses, holiday and weekend travel.

- An analyst might use simple stochastic models to describe each component, and then attempt to decompose the complex empirical process into its constituent parts.

- We have a superposition of results from many distinct data generation mechanisms, some of which are well understood, some which are partially understood, and some of which may represent new ones.

MACHINE LEARNING: EXAMPLES

- The inverse design problem, aimed to identify input values producing desired outputs, has long been a challenge in natural science and engineering areas, due to the many-to-one relationships between inputs and outputs that are commonly embedded in the data.

- Recently, several deep learning-based methods have been developed which tackle the problem in a more direct way: one of them is conditional invertible neural network (cINN), which benefits from both mathematical and statistical principles.

- First, cINN is a generative probabilistic model which stochastically generates posterior samples of inputs X given specified outputs Y, i.e. $P(X|Y = y)$.

- Second, the network architecture of cINN is carefully designed to enable the computation of a Jacobian determinant.

- Furthermore, it is critical to effectively select a subset of posterior samples of inputs for the follow-up validation study, which may incorporate user constraints and potentially involve an optimisation process.

MACHINE LEARNING: EXAMPLES

- Sequential aggregation of individual predictions aims to predict y_t, given past values up to t: $y_1, \ldots, y_{t-1}$, using K experts whose only known information consists of their immediate predictions of y_t and the history of their predictions.

- The idea is to optimally combine the predictions by adjusting the weights (assigned to each expert) at each step based on instantaneous losses (e.g. quadratic), with no stochastic modelling.

- Auder et al (2016) considered daily average concentration of PM10 in a location in Normandy, as well as the corresponding forecasts given by 10 different statistical models

- The sequential prediction strategy significantly improves the performance of the best expert, both in terms of errors and alerts.

STOCHASTIC MODELS AND MACHINE LEARNING

- Fernandez-Delgado et al (2014) compared 179 classifiers over 121 (non-large-scale) data sets from the UCI ML classification database, and found that parallel random forest (mostly due to a statistician, Breiman), performs best, followed by support vector machines. However, when no training data is available at the design phase or when relatively small amounts of data need to be analysed, stochastic modelling still shines.

- There is also an increasing number of applications, such as embedded systems (HW+SW) and implanted medical devices, which can only access very limited computational power. To provide them with ML capabilities, one may employ randomised neural networks, which can be trained with extremely low computational costs, combined with computationally efficient statistical methods to assess uncertainty, such as fixed-length confidence intervals to determine required sample sizes.

- Emulators are (usually) Gaussian process approximations to the agent-based model and offer fast and practical solutions that are sufficiently faithful to it to provide useful guidance. Remarkably, emulators can give posterior distributions over the discrepancy function, which indicates for which regions of the input space the emulator does a poor job of matching the agent-based model (computer simulations used to study the interactions between people, things, places, and time).

STOCHASTIC MODELS AND MACHINE LEARNING

- Although various state-of-the-art AI/ML approaches do not rely on explicit stochastic modelling, this is not the rule.

- Indeed, many ML problems can greatly benefit from stochastic modelling expertise and require results from Probability and Statistics for their improvement.

- For example, in industrial quality control, when setting up inspection processes, manually selecting key characteristics requires substantial resources and efforts.

- An auto-encoding neural network was used to learn such features from training data, with correlated features identified and evaluated by means of a risk analysis.

- The overall analysis includes pre-processing, design of net parameters, specification of the dependence measure, and combines ML as well as human expertise from Statistics and Engineering.

STOCHASTIC MODELS AND MACHINE LEARNING

- Active learning is a highly relevant learning problem in industry, dealing with the automatised sequential exploration of a sample space to identify the safe region where a system can optimally operate.

- It aims at reducing the number of labelled examples needed to achieve a certain accuracy by selecting the most informative safe examples from a large unlabelled dataset and determining their labels from a costly additional experiment.

- When it comes to dynamic systems, for example when a robot explores an environment, exploration takes place along trajectories instead of single points, and then the whole trajectory needs to be safe.

- Gaussian processes are an attractive approach to model the sequential data collection mechanism as well as the prediction uncertainty at a new point. For this purpose the Borel-TIS inequality is used instead of Monte Carlo simulations.

- It turns out that the combination of stochastic modelling and results from statistical theory and probability theory allows to produce a state-of-the-art ML method.

ARTIFICIAL INTELLIGENCE

- AI is susceptible to how a question is posed
- AI is limited in its ability to separate the good from the bad, in terms of the resources it uses to carry out its 'reasoning'. These resources may include resources generated by AI and derived from unsound resources or algorithms.
- The recent European regulation on AI (*AI Act*) aims to regulate the use of AI with a set of requirements of trustworthiness for AI applications, to be embedded in a risk management model. The requirements established for high-risk applications in the AI Act can be classified in four main variables to measure: Security, Accuracy, Fairness and Explainability. All of them need a set of consistent metrics that can establish not only if, but also how much, the requirements are satisfied over time.
- Regarding the role stochastic modelling can play in the AI era, one can distinguish between specialised ML approaches competing with stochastic methods, e.g.,
  - prediction using deep neural networks vs. (non)parametric regression
  - AI systems which automatically plan and model a statistical problem and then analyse real data vs. traditional computer-aided modelling and analysis done by a trained human analyst

ARTIFICIAL INTELLIGENCE

- Any answer to the initial question needs to predict to some extent the future development of AI, and thus may fail, but one can make an educated guess.

- The rapid progress of LLNs (Liquid Neural Networks) and their fine-tuning to specific domains and tasks allows to set up interacting AI agents to generate, check and curate output.

- It seems that software development is the field where this approach is most advanced and already provides convincing results.

- Here one agent outputs source code based on a user prompt, and a further specialised agent interacting with the user performs code checking and outputs directives for source code revisions, in order to eliminate errors and improve the program.

- It is clear that the concept of several AI agents, which interact among each other and with the user to solve certain problems, will become widespread and has a substantial potential for better and less error-prone AI.

ARTIFICIAL INTELLIGENCE

- LLNs are trained from massive internet data which contains huge amounts of source code in many programming languages and this is certainly the reason for their capabilities in generating computer programs.

- Probability, Theoretical Statistics and large parts of Applied Statistics and stochastic modelling are exact sciences and follow relatively simple grammars or, at least, they can be put into a formal language following a simple grammar.

- Therefore, one can expect that future AI systems have capabilities in these areas which are comparable to their skills in software development, thus going beyond the already remarkable functionality of Open AI's Code Interpreter.

- It is likely that future AI systems substantially simplify the development and application of stochastic models and statistical analyses and provide access to a large amount of knowledge.

- This might have devastating effects on the job market for graduates, since then only a few experts are needed to supervise the AI.

ARTIFICIAL INTELLIGENCE

- However, that development will be probably relatively slow.
  - Compared to software, there are less data for training of a Statistics agent.
  - Whereas software is written in a formal language, this does not strictly apply to scientific papers and books which form the training corpus.
  - Although not completely transparent, the state-of-the-art AI systems use lots of manual input from human experts.
- When it comes to Statistics, this needs well trained experts which are not available on a large scale, and it is questionable whether providers will invest here.
- Thus, as long as human input is needed, the capabilities in such fields will substantially lag behind areas such as programming.
- AI systems capable of producing valid stochastic models, methods for their analysis and derivations of their properties might result as a side product of efforts to create AI systems which can solve scientific research questions addressing hot topics such as finding cancer treatments, understanding the human brain or finding the grand unified theory of physics.

ARTIFICIAL INTELLIGENCE

- Nevertheless, it is an open question whether future AI systems will be smart enough to produce valid novel scientific results going beyond correct and nice sounding science prose, and whether they will be superior to humans in identifying and resolving challenging problems arising in modelling and analysing real-world data.

- Examples for the latter are causality (versus association), confounders and colliders, bias and multiple testing.

- For example, in order to identify causality, controlled experiments and randomization are the methods of choice, and available data and those obtained from observational studies often suffer from confounders (variables affecting response and regressor) and colliders (variables affected by response and regressor), which lead to distorted estimates of causal effects.

- Generally, blind analysis of available data collections may lead to severe biases which quickly result in discrimination and harm for people.

MY OWN INTERESTS

- Bayesian networks to split complex problems in many simpler ones, exploiting information and keeping interpretability
- Adversarial classification as an example of adversarial ML based on a proper Bayesian statistical approach
- AI-driven expert elicitation for prior choice in high dimensional problems

CONCLUSIONS

- It is worth noting that both the statistical modelling and ML communities are rapidly evolving. Therefore, a good practice for problem solving is to leverage multiple approaches and assess their performance and usability. Integrating insights from multiple approaches addresses the problem from different perspectives, thereby guiding users toward a clear path for continuous improvement.

- There is no any antagonism between statistical modelling and ML, which in some respects is its 21st century version, just as the emergence of Computer Science led to the development of Multivariate Statistics in the second half of the 20th century. ML has enriched the statistician's toolbox and, above all, has provided him/her with the fundamental concept of generalisation and the need to go beyond simply fitting a model on the basis of training data alone.

- For their part, statisticians can provide ML practitioners with their knowledge of biases, their sense of data (missing, aberrant, etc.) and their culture to avoid reinventing techniques such as categorical data encoding or principal component analysis! But statisticians must not be afraid to enter this new field, otherwise they will be marginalised.

CONCLUSIONS

- Statistical modelling remains a powerful tool for problem formulation and solution development, and ML can yield more insightful outcomes when it is integrated into the problem-solving process.

- Statistics contributed and contributes a lot to the development of ML and AI, since recognising the importance to formulate the problem of learning from data as an inference on the underlying data generating process (e.g. random sample or time series) is fundamental for a deeper understanding of ML methods.

- Statistics provides a more advanced and complete general framework for data analysis and is highly relevant for any AI development since AI applications often neglect issues such as evaluation of uncertainty, interpretability, model stability and reproducibility, or issues such as confounding, especially in areas where ML is regarded as state-of-the-art.

- This certainly applies to problems dealing with large-scale data such as image classification, which require enormous computational resources and are almost exclusively researched by computer scientists. However, for other task where small to moderately large samples are sufficient, the situation can be different.

Applied Stochastic Models in Business and Industry

WILEY

RESEARCH ARTICLE OPEN ACCESS

Is There a Future for Stochastic Modeling in Business and Industry in the Era of Machine Learning and Artificial Intelligence?

Fabrizio Ruggeri[1] | David Banks[2] | William S. Cleveland[3] | Nicholas I. Fisher[4] | Marcos Escobar-Anel[5] | Paolo Giudici[6] | Emanuela Raffinetti[6] | Roger W. Hoerl[7] | Dennis K. J. Lin[3] | Ron S. Kenett[8] | Wai Keung Li[9] | Philip L. H. Yu[9] | Jean-Michel Poggi[10] | Marco S. Reis[11] | Gilbert Saporta[12] | Piercesare Secchi[13] | Rituparna Sen[14] | Ansgar Steland[15] | Zhanpan Zhang[16]

Correspondence: Fabrizio Ruggeri (fabrizio@mi.imati.cnr.it)

Received: 15 November 2024 | **Accepted:** 13 February 2025

Funding: This work was supported by Fundação para a Ciência e a Tecnologia and European Commission.

Keywords: artificial intelligence | machine learning | stochastic models

ABSTRACT

The paper arises from the experience of *Applied Stochastic Models in Business and Industry* which has seen, over the years, more and more contributions related to Machine Learning rather than to what was intended as a stochastic model. The very notion of a stochastic model (e.g., a Gaussian process or a Dynamic Linear Model) can be subject to change: What is a Deep Neural Network if not a stochastic model? The paper presents the views, supported by examples, of distinguished researchers in the field of business and industrial statistics. They are discussing not only whether there is a future for traditional stochastic models in the era of Machine Learning and Artificial Intelligence, but also how these fields can interact and gain new life for their development.

1 | Introduction

The paper comes from the idea of one of the authors who has been Editor-in-Chief of *Applied Stochastic Models in Business and Industry* for 17 years. Based on his experience in the journal and his participation in conferences, he observed what is nowadays evident to everyone: Machine Learning methods have proved to be very effective in many fields (like prediction and classification) which were traditionally covered by Statistics. This fact triggered the question which gives the title to this contribution: Is there a future for stochastic modeling in business and industry in the era of Machine Learning and Artificial Intelligence? This is a question that does not imply a simple yes/no answer, but requires deep thinking and discussion, pointing out differences, common aspects, interactions, and new opportunities, as well as the pros and cons of those approaches. The goal of this paper is to promote a common reflection about all those aspects, in particular when applied to business and industrial problems. Many leading researchers in the field were contacted and most of them enthusiastically agreed to contribute. What you can read now is the result of a collective work involving more than 20 researchers, with different expertise. It is not like a traditional paper where a new model is presented, with a discussion about its property and a possible illustration of its utility in addressing a real problem. Here we have contributions that address the initial question from different points of view. Grouping them under some common themes was not possible, so the paper is organized into sections corresponding to each contributor (sometimes two of them jointly), presented in alphabetical order.

For affiliations, refer to page 22.

I feel like a chatbot. The editor's prompt was *Will stochastic processes continue to be relevant to business and industry?* And now I get to make up an answer, or possibly even hallucinate.

As Yogi Berra is said to have said, *Prediction is difficult, especially about the future*. My guess is that stochastic processes will remain important, but that mathematical stochastic process modeling will become less so. Although my response emphasizes business and industry, some of my examples and discussion speak to more general applications.

2.1 | Why?

Currently, we regularly use Gaussian processes [1], treed Gaussian processes [2], and Gaussian process regression [1] to describe all sorts of phenomena. An early application was for estimating the amount of gold ore in South African mines, a methodology now known as kriging [3]. We use conditional autoregressive process models for syndromic surveillance [4], Poisson flow models for Internet traffic [5], and heavy-tailed random fields for mapping regions of neural activation [6]. Time series models are ubiquitous and have been built out in many directions, as have spatial process models and spatio-temporal processes [7, 8]. Self-exciting Hawkes processes are used in finance [9], neural connections [10], and in tracking political blog activity [11]. There are many other mathematically defined processes with corresponding application areas.

But the world is changing. In some fields, the amount of data that is available has exploded. *Applied Stochastic Models in Business and Industry* (ASMBI) recently had a special issue on autonomous vehicles. If, as several of those ASMBI articles suggested, we shall someday have all vehicles on the road being networked and autonomous, then there will be an ocean of travel data. We will not need differential equation models for traffic flow [12, 13]; we shall have the empirical process in great detail. In that regime, the need for process modeling is much less, although the stochastic process itself remains important.

Given large quantities of data, the nature of process modeling itself will probably change. For example, if we have data on flows of networked vehicles, what we are observing is actually the superposition of observations from multiple and perhaps simpler data generation mechanisms [14, 15]. Specifically, some of the traffic corresponds to daily commutes, with peak volumes in the early morning and late afternoon. But another component of the traffic corresponds to long-haul trucking and is well described by flows on a distribution tree, perhaps with travel timed to avoid rush hour congestion. A third component is produced by school buses, a fourth by weekend travel, a fifth by holiday travel, and so forth. An analyst might use simple stochastic models to describe each component, and then attempt to decompose the complex empirical process into its constituent parts.

Similar situations arise in many other applications. In high-energy particle physics, the search for the Higgs boson entailed analysis of five petabytes of time-stamped data [16]. Those data are an accumulation of collision events and decays, many of which correspond to known laws of physics, and the goal is the discovery of anomalous behavior. As before, one has a superposition of results from many distinct data generation mechanisms, some of which are well understood, some which are partially understood, and some of which may represent new physics.

In the context of business, I see great potential for statistics in computational advertising, and I hope that more statisticians will get involved. Computational advertising is a complex economic ecology and a transformational driver in modern e-commerce. It includes click-through prediction [17], recommender systems [18], causal inference [19], and experimental design [20], among many other statistical subfields. The stochastic process perspective informs several aspects of its development.

One way that stochastic processes arise is in contract fulfillment for showing online advertisements. When a user visits a website (e.g., cnn.com), it triggers a complex sequence of activity lasting less than ten milliseconds. There is a virtual auction among demand-side platforms that must decide (1) whether to bid on showing an ad to the user, (2) which ad to show, and (3) how much to bid. Typically, the demand-side platform knows quite a lot about the user: Gender, approximate age, approximate income, marital status, location, and previous purchase history.

A demand-side platform contracts with clients to show the clients' ads to a certain number of people with specified characteristics before a certain date. For example, McDonald's might contract to have 200,000 hamburger ads shown to males between the ages of 16 and 35 in southern California before January 1, 2025. Similarly, Wendy's might contract to have its ad shown to 400,000 people in California between the ages of 15 and 40 before November 30, 2024. Obviously, sometimes the same person could be used to fulfill either of the contracts and then the demand-side platform must decide which ad should be shown.

If a demand-side platform does not completely fulfill the contract, then it must return some portion of the client's money, and the amount depends upon how near to the target the platform has come. Third parties are routinely used to ensure that ads are being displayed to people that match the contract specifications (of course, there is noise—sometimes my wife uses my laptop). This leads to an interesting optimization problem, and one aspect of that problem is the uncertain future. The demand-side platform does not know what contracts it will write next month, nor does it know who is going to log in to one of the websites upon which it can bid for eyeballs.

I do not know if any of the demand-side platforms are using dynamic programming for stochastic optimization [21]. It is unlikely—calculating a Gittins index would be hard, and the underlying stochastic process is surely non-stationary. But, the stochastic process perspective may inform the heuristics that managers of demand-side platforms need to use to ensure economic viability.

Agent-based models are used in many different disciplines, and although I do not have direct knowledge of their degree of penetration in industry, there is much academic discussion about

such applications [22]. They can be tools for logistics and supply chain management [23], they can inform various aspects of insurance [24], and they can, in principle, describe manufacturing processes [25, 26]. However, in many applications, they can be computationally cumbersome, requiring a lot of memory and/or a lot of time.

In such cases, one standard workaround is to use an emulator. Emulators are (usually) Gaussian process approximations to the agent-based model. Sometimes they offer fast and practical solutions that are sufficiently faithful to the agent-based model to provide useful guidance [27]. Remarkably, emulators can give posterior distributions over the discrepancy function, which indicates for which regions of the input space the emulator does a poor job of matching the agent-based model.

2.2 | Conclusion

My guess is that stochastic processes will continue to be important in business and industry, but that attention will shift from closed-form expressions and theory towards computation. In commercial competition, better predictive accuracy and better fit are comparative advantages. I think companies will be happy to give up mathematical elegance in exchange for slightly better performance.

Nonetheless, the theory may have persistent value as the starting point for building computation-based processes. That theoretical insight can help the company statistician to flag which features of the new processes are most important to the firm's financial future. Statisticians and probabilists have built an extraordinarily broad and flexible intellectual toolkit, but the real world will generally still be too complex to be fully described by such models.

I must end with an apology. Usually, I deplore people who relentlessly cite themselves. This paper sins grievously in that regard. My excuse is that this exercise seemed more like an op-ed piece than the usual ASMBI article, and so I responded to the prompt using ideas that have been in my head for a while. I beg your indulgence.

3 | William S. Cleveland and Nicholas I. Fisher

Why will Human Intelligence continue to prevail in Data Science?

The spark that creates Life has yet to be purposefully created in a laboratory. This spark is intrinsic to being able to ascend all of the following steps:

1. Create or gather **data**.
2. Turn data into **information**.
3. Transform information into **knowledge**.
4. Apply knowledge with **wisdom**[1].

Artificial intelligence (hereafter AI) is totally dependent on the quality of what it can find on the Web relating to Items (1–3). For many problems, this brings enormous advantages compared with a group of data scientists applying Human Intelligence (hereafter HI) to the same problems. However, there are, and will continue to be, problems where HI will prevail, not least because of the spark that provides us access to Item (4).

HI brings to stochastic modeling some qualities that are essential to good work: Scepticism, Doubt, Suspicion, ... all of which add up to a form of Statistical Wisdom in both solving a problem and assessing whether a good job has actually been done. We illustrate this with two incredibly contrasting examples, one relating to massive data sets and development at the leading edge of research where little methodology is available, and the other involving very small data sets and very localized inference where, again, there is little specific or at least inferable information available.

We preface these examples by making two general assertions about AI that are a consequence of the qualities that it lacks.

- AI is limited in its ability to separate the good from the bad, in terms of the resources it uses to carry out its "reasoning". These resources may include resources generated by AI and derived from unsound resources or algorithms.
- AI is susceptible to how a question is posed.

Of course, data scientists also face the same challenges. However, Statistical Wisdom provides a reasonable measure of protection in this regard.

Example 3.1. The data of the Internet are the traffic transmitted. Deep data analysis requires deep knowledge of Internet technology and stochastic process modeling. The Internet is a complex system of links and routers. An example process is the transport of a "file" from one "host", a computer, to another host, establishing a "connection". The file is broken up into "packets", no more than 1500 bytes each. This is the "payload". Each packet also carries 40 bytes of transmission information, the "headers", such as the IP addresses of the source and the destination hosts of the connection. Traffic Modeling (TM) is critical for Traffic Engineering (TE). TM enables TE simulations to find optimal pathways, the "Edges" of the Internet consist of connections by people at home, businesses, schools, etc., with connections to Internet service providers who enable entry to the Internet. The user technology is a part of the Internet of Things (Iot) with host hardware devices, and interface software for building, sending, and receiving connections. The many entering connections occur randomly, payload sizes are random, and the packets of connections merge randomly. This is why Internet transport is a "stochastic process". Interestingly, traffic engineers refer to the merging of packets as "statistical multiplexing". Further along the pathways, merged streams merge with one another, and the traffic becomes the Internet "Core" with rates substantially larger. Routers are much more powerful. The stochastic properties of the Core differ from those of the Edges. Technically, that is, mathematically, Traffic overall is Long-Range-Dependent (LRT). Edge traffic is still bursting. This makes Edge router design much more complex – and therefore more costly – than if traffic were smooth. In the core, technically (mathematically) the traffic is still LRD, but the ratio of the traffic rate standard deviation over the mean declines, so the traffic gets smoother in the Core. Burstiness declines. Core

Routers there are very costly since they have much more to manage, but with much more smoothness, traffic engineering is much easier, and cost goes down. When this was discovered, many traffic "experts" disagreed very strongly, but as time went on, statistical traffic diagnostics showed the smoothness was truth[2]. See [28] for more details.

Model validation was critical to confirm the accuracy of the model. It was carried out by analyzing live packet traces in both directions from gateway links in Auckland in New Zealand, Leipzig in Germany, and Bell Labs in Murray Hill, New Jersey, USA. Arrival time traces of 715,665,213 packets were collected. The collected segments were large enough to be informative, but not so large that the rate was non-stationary.

The processes to study the internet and carry out model building for traffic engineering require knowledge of the internet and of stochastic model building and very importantly, experience in using this knowledge. AI technology consists of algorithms, but there is no apparent need for foundational algorithms either for stochastic processes or for the Internet, which has protocols that run the system. There are tools for detecting protocol violations but calling them algorithms is probably an overstatement. For Internet random processes there is far too much Internet subject matter involved in applying Statistics for model building to the point of being able to carry out simulations for Internet processes, and for discovery of ways to improve Internet performance, for AI processes to analyze and model and improve the Internet.

Example 3.2. Processes for measuring and managing relationships are ubiquitous throughout business and industry, in order to maintain market share, ensure staff retention, improve safety performance, and so on. These surveys are critically dependent on being able to capture high-quality satisfaction data from survey-satiated groups, so the survey processes need to be able to work well with short, carefully constructed survey instruments that furnish actionable data from small but representative surveys (e.g., [29]). Such instruments tend to capture both qualitative and quantitative data, the numbers providing the basis for improvement priorities, and the comments insight into what, specifically, needs to be fixed. It is this latter aspect that calls for HI, rather than AI. Many workplaces have workforces from a wide variety of cultural and linguistic backgrounds let alone levels of literacy, so constructing a bespoke survey instrument from focus groups is itself a challenge as is looking for root and systemic causes in the small numbers of (often semi-literate) comments. And, of course, enterprises keep this sort of material confidential, so there are no vast resources of comparable studies elsewhere to feed into the AI maw and churn out actions.

4 | Marcos Escobar-Anel

Interactions among humans, between humans and the environment, together with changes in human behavior due to evolving cultures and technological advances, will continue to increase the complexity of phenomena in the field of social sciences.

A simple example motivated by finance but easily extrapolable to other areas is the evolution of stock prices in the last 60 years. The price of a share is not only dictated by rational economic forces but is also a by-product of human individual and collective behavior, our desires, passion, wishes, and mistakes; all this is summarized on a single number at any given time. This means the modeling of stocks has had to adapt/evolve from a quite robust and straightforward Gaussian-based model in the '60s too, for example, the inclusion of jumps in the '70s with Levy models (e.g., [30]), random local volatility as detected in the '70 and '80s (i.e., GARCH, and constant elasticity of volatility (CEV) models, e.g., [31, 32]), stochastic volatility in the '90s (e.g., [33]), motivating the emergence of the volatility index (VIX) by the Chicago Board Options Exchange (CBOE). Advanced models, combining these features and more, continue to emerge, for example, [34–37]. Nonetheless, new features, like stochastic volatility of volatility and stochastic skewness on prices, as captured by the relatively new indexes VVIX and SKEW respectively (CBOE), are yet to be tamed. This is just for single assets; multi-asset modeling is even richer. It, therefore, stands to reason that the increase in complexity will only continue, requiring more advanced models to explain the dynamics of single and multiple stocks adequately.

It is in this ever-increasing complex environment where the new field of Machine Learning (ML) and the immense potential of Artificial Intelligence (AI) have emerged, ideally as prospective saviors to help humans tame reality. Would this mean the end of stochastic modeling?

At first sight, we have gone through this before. Interestingly, the development of new mathematical/scientific disciplines, for example, algebra, calculus, probability, differential equations, time series, and stochastic calculus, has not only created topics arguably more complex and abstract than their predecessors but has never made predecessors obsolete. On the contrary, it has breathed new life into its predecessors, widening interpretations and connections and enriching our toolbox to explain reality as a consequence. Even more encouraging is that, despite the increasing complexity of every new discipline, more humans than ever dare to study and understand it, as evidenced by the steady increase in PhD graduates worldwide. This means we continue to show capacity as a species to meet the challenges.

On the other hand, at a deeper level and depending on our risk-aversion levels, the emergence of ML/AI is so revolutionary that it might look like a change point in the dynamics of our scientific evolution. This resonates with a common pitfall in careless modeling, that is, we should not blindly use the past to explain the future, assumptions should be checked periodically, and models should be updated. Could AI be so disruptive that it would render models obsolete?

In my opinion, the answer is no for two reasons. First, the increasing complexity described above is mainly human-made, irrational to some extent, and fed by the widening spectrum of our individual needs and diversity. It is in the study of humans where the answers are found, and fortunately, we have an obvious advantage. Secondly, a model, a formula, like a painting, is the human way to understand and explain ourselves and the world. We have a genetic need to satisfy our curiosity and to understand reality; we do not like imposed solutions. So, we will always need to translate phenomena (either created by the universe, by us, or by a machine) to a level we can understand and share, and it is in this context where stochastic models will always appear.

5 | Paolo Giudici and Emanuela Raffinetti

Machine Learning models are boosting Artificial Intelligence (AI) applications in all human activities, particularly in domains such as finance, health care, and automotive. AI applications have improved in general products and services, in terms of costs, user experience, and inclusion. Moreover, their use in finance gave rise to the diffusion of financial technologies (fintech), resulting in an improvement of the payment processes, asset management, lending, and insurance services.

This is mainly due to their advantage, in terms of predictive accuracy, with respect to "classic" statistical learning models. However, although complex Machine Learning models may reach high predictive performance, their predictions are not explainable, and have an intrinsic black-box nature: Input data are transformed through complex processes without effective control and monitoring of the risks arising from potential biases in forecasting. This is a problem in regulated industries, as authorities aimed at monitoring the risks arising from the application of AI methods may not validate them. For example, the application of AI to finance may lead to automated decisions that can classify a company at risk of default, without explaining the underlying rationale and, therefore, limiting possible intervention actions in case of wrong predictions.

Furthermore, they may not be robust, they may use private data, or data that are not representative, thereby generating bias and unfairness.

Indeed, differently from ordinary computer software and its applications, AI not only converts inputs into outputs, but can also change the surrounding environment, with the risk of creating harm to individuals, organizations, and countries. This is the reason why authorities, regulators, and standard bodies around the world have begun to monitor the risks arising from the adoption of AI methods.

For example, the European Union has introduced the AI Act, which puts forward a number of key compliance requirements to AI in terms of security, accuracy, fairness, and explainability, compulsory for high-risk applications [38].

The above developments require, to be practically implemented, the availability of a set of statistical metrics that can actually measure whether AI applications are compliant or the probability that they are not compliant, along with the expected harm if they are not so.

This requires research, particularly in the field of stochastic modeling, aimed at developing a consistent set of metrics that can estimate the probability distribution of AI harms, to be employed in AI risk management models.

The main compliance requirements for AI established by the international regulations and standards [38–41] can be summarized in four main principles, which are measurable, and not only auditable.

The first principle is related to the performance of AI systems: Accuracy, especially predictive, but also in terms of the "authenticity" of the AI-generated context.

The second principle is sustainability, related to the robustness and resilience of AI systems to extreme events, for example environmental, or to cyberattacks that may violate their security and integrity.

The third principle is fairness, related to the impact of AI systems on the external environment and particularly on human rights.

The fourth principle is explainability, which refers to human-AI interaction, and requires the output from AI systems to be transparent, accountable, and interpretable.

Together, the four principles make the acronym S.A.F.E. An example of a set of consistent statistical metrics to assess the S.A.F.E.ty of Machine Learning models is contained in [42], which extends [43]. The paper [42] provides a model agnostic approach to assess S.A.F.E. machine learning, valid for all AI applications, independently on the underlying field domain, data, and models, along with a Python software implementation: The `safeaipackage`, which allows full reproducibility of the proposed model.

The metrics proposed in the paper are consistent with each other, according to a common mathematical framework: The Lorenz curve (see, e.g., [44]). The Lorenz curve is a well-known robust statistical tool, which has been employed, along with the related Gini index (see [45]) to measure income and wealth inequalities. It thus appears as a natural methodology on which to build an integrated set of trustworthy AI measurement metrics, which allows their integration into a unified decision-theoretic framework.

An important advantage of the S.A.F.E. model contained in [42] is that all four proposed metrics are based on the same notion of variability, derived from the Lorenz curve. They can therefore be similarly normalized to [0, 1] and integrated into a single measure that can assess the trustworthiness of any AI application.

On the other hand, a possible weakness of the model is that it does not yet fully take into account model uncertainty. It does so within the testing procedure which allows, by means of the developed jackknife procedure, to decide whether a certain value of a metric is significantly higher than a certain threshold. This may be useful to estimate whether an AI application is compliant with regulations and standards, but may not be sufficient to assess its risks, in terms of probability and expected impact.

Further research should be pursued, to improve the model and/or provide alternative modelizations that can help assess the uncertainty surrounding the S.A.F.E. metrics and, more generally, quantify the uncertainty of Machine Learning models.

6 | Roger W. Hoerl and Dennis K.J. Lin

6.1 | Introduction

Data science, especially its most recent manifestation emphasizing artificial intelligence (AI), is clearly one of the hottest, if not *the* hottest, of technical fields today. This has led many computer scientists to view statistics as an "old technology", no

longer useful in a big data era. Perhaps more concerning, many within the statistics community appear to be focusing their teaching and research more on machine learning, coding, and cloud computing than on traditional statistical strengths, such as randomized experiments, applied probability, process control, and uncertainty quantification. AI clearly provides exciting new capabilities, as anyone who has used ChaptGPT can testify. We argue, however, that in a big data era, statistical foundations are not only still relevant, but in fact even more critical than they used to be. This is because AI's weaknesses tend to be statistics' strengths, and without incorporating statistical principles AI can be quite dangerous. One obvious example is the Boeing 737 Max tragedy, in which faulty sensor data fed into an automated flight control system led to the deaths of 346 people [46].

To be clear, we are "bullish" on AI, and feel that it has significantly contributed to society already. Our main point, rather, is that AI will be most successful when it integrates core statistical principles, as well as domain knowledge. This implies that data science and AI development teams need to be technically diverse, and include people well-trained in traditional statistics. Similarly, while it behooves statisticians to stay up with the times, and know as much about machine learning and AI as possible, we suggest that this should be accomplished without sacrificing expertise in core statistical principles, which only those well-trained in statistics can provide.

6.2 | Data (AI) Scientists and Statisticians

Much of the dialogue between AI researchers, data scientists, and statisticians seems to be competitive. That is, "our group is better than your group". From roughly 2015 until 2020, data scientist was the "hottest" job title. Now, it seems to have been surpassed by AI researchers or AI scientists. This competition over who is "hottest" overlooks the obvious fact that different skills are needed to succeed in any data activity in practice. As an analogy, a basketball team of all point guards is not likely to perform well, nor is a football (soccer) team made up of all goalies. Succeeding in medicine, sports analytics, science and engineering, or political polling, for example, requires diverse teams, rather than teams that are "an inch wide and a mile deep." Technical diversity seems to be a grossly overlooked and undervalued form of diversity.

Those working in AI and machine learning (which we will abbreviate as "AI scientists" for simplicity) and more traditional statisticians actually need each other. We suggest that statisticians have much to learn from AI scientists, including the following:

- AI scientists have been much more successful in communicating their value to senior management. For example, senior leaders are much more likely to value "deep learning" than a "hierarchical generalized linear model." Statisticians might write this off as "marketing," but it is undeniable that without leadership support there is little data analysts can accomplish.
- Skills in data acquisition, transmission, storage, and retrieval are critical to extracting information from massive data sets. While statisticians are studying these topics more now, they are generally part of the core training of computer scientists.
- Similarly, computer scientists as a group will always be more proficient in computation and coding, especially in optimizing code for computational speed.
- AI scientists understand that there are problems for which causal reasoning and interpretable models are not required. Many problems require interpretable models, but in some cases, being able to accurately predict is sufficient. Empiricism has its place.

We also feel that statisticians have much to offer to AI scientists:

- From the literature, it is clear that many AI scientists are not well grounded on core statistical principles, perhaps viewing them as outdated or no longer relevant. We refer to principles such as the distinction between a sample and a population, the concept of inference—which is not the same as the AI concept of generalizability, potential sources of bias in data and models, the superiority of data from randomized experiments versus observational data, the need for holistic evaluation of models beyond prediction accuracy out of sample, and the criticality of incorporating sound domain knowledge in models. All of these concepts are even more relevant in a big data era.
- As pointed out by Li [47], many statisticians have extensive consulting experience interacting with clients to discuss the problem in question, data that would be needed to solve it, what would constitute a successful solution, and so on. Most AI scientists work in organizations separated from the operations where the model might actually be deployed, hence they often do not understand the real problem in need of solution [48]. Statisticians learned long ago that the stated problem often turns out not to be the real problem in need of a solution.
- Statisticians generally have experience with problem "triage," that is determining what type of problem one is facing, so that an appropriate approach can be tailored. Is it a predicting problem, an estimation problem, a problem requiring validation of scientific hypotheses, or something else? Too often, AI scientists have been trained to view every problem as a prediction problem, a limitation only exacerbated by the proliferation of data competitions. While this has "simplified" the task of model evaluation, we argue that it has resulted in over-simplification of this critical task.
- A rich underlying theory is needed to provide a basis for inference or generalizability for AI [49]. That is, in order to have confidence that a model will work well on new data, there must be some scientific basis for confidence. If we do not understand why a model works, we obviously will not understand its limitations, and when it will not work. Statisticians understand this and have developed a deep underlying theory for sampling, inference, experimental design, and so on. This theory provides a much richer basis for inference than performing out-of-sample predictions on a single test set, which is typically taken from the same original data as the training set.

6.3 | Types of Intelligence

In order to dig deeper into the relationship between statistics and AI, we see the need to consider three types of intelligence (AI, BI, and SI). Typically, AI employs a large number of inputs (training data), super-efficient computer power/memory, and smart algorithms to perform various tasks "intelligently," such as driving a car or evaluating a loan application. Of course, AI is by definition "artificial," implying that there is "natural" intelligence. We use the term Biological Intelligence (BI) to refer to the natural intelligence innate in humans, and other living creatures. BI combines instinct with proper Soulware [50], and little—even no—input data to achieve its performance. By "soulware" Kuo refers to cultural, cognitive, and behavioral patterns that may differ significantly between people groups. We name a third type of intelligence Statistical Intelligence (SI), by which we refer to employing sample data, statistical inference/models with solid theoretical foundations, and BI to solve problems more complicated than can be solved with BI alone. We view SI as being on a continuum between AI and BI, in that ideally it combines domain knowledge with empiricism. We argue that SI can serve as a bridge between AI, which is objective and powerful, but also lacking in domain knowledge and "common sense," and BI, which provides what AI lacks, but is of course subjective and highly variable from person to person.

These types of intelligence relay on different types of data. AI requires digitized data, for example when converting images to pixels. More and more subjects are being digitized, hence we anticipate that this will result in AI becoming more extensively applied, and more powerful. SI, on the other hand, requires that this digitized data be structured to some degree. Analyzing unstructured data remains a challenge to most statisticians. Much of the "data" utilized by BI is unknown to us, such as something we feel but can't articulate ("I just have a bad feeling about that person."). We refer to this as "soft" data in the sense that it cannot be measured, at least not as of yet.

We further argue that "learning" is broader than "learning from data." There are many important components of learning. In BI, one often learns from direct observation, without any quantification. For example, lion cubs learn to hunt by watching mature adults in the pride hunt. Humans frequently learn from subtle signals, such as body language, tone of voice, or even lack of conversation. While one can define "data" quite broadly, minimizing such differences, it is still true that AI, BI, and SI require fundamentally different types of data. AI requires digitalized data, not necessarily structured, SI requires structured data, while BI can effectively utilize "soft data," including "data" that we do not know how to quantify or digitize. For example, most of us have experienced someone "saying the right thing" to us, but clearly sensing that the speaker is not being sincere. Historically, detection of such subtle signals was often referred to as "intuition," because people were not able to precisely explain exactly how they detected the lack of sincerity.

6.4 | Four Critical Considerations for Intelligence

We want to consider four specific issues that we feel are particularly relevant in the integration of AI, SI, and BI: Samples versus populations; statistical Inference; interpolation versus extrapolation; and domain knowledge.

In our experience, AI researchers too often assume that an extremely large sample can be considered the population. However, the goal for almost all real problems is to consider the entire population, for example, to evaluate future loan applications not yet received. SI can add significant value here, showing how a thorough study of the sample can reveal the properties of the population. The key idea behind such an approach, to be discussed below, is the assumption that the population of interest is well represented by the sample. Sample size matters but is not critical; data quality is more important than data quantity.

Consider two simplistic examples—soup tasting and cake tasting. For soup tasting, one does not need to drink the entire pot of the soup to learn if the soup tastes good or not. In fact, if one stirs the entire pot and takes one spoon to taste, one can quickly evaluate the quality of the soup. In this case, the sample size is one and it is sufficient. On the other hand, for tasting a multi-layer cake, one spoon from the top layer is clearly not sufficient. However, even if you take thousands of samples at the top layer this will not be sufficient either. What is needed from the sample is "representation," a vertical slice of cake in this case. As we discuss below, the "right data" are what we need, not necessarily "big data." Ronald Fisher discovered over a century ago that a small amount of the right data is usually more useful than a large amount of the wrong data. This principle is still true today.

Statistical inference, part of SI, involves uncertainty, typically in at least three forms—point estimates, interval estimates (confidence intervals), and hypothesis testing. In other words, SI provides more than a point estimate, but also its variation or uncertainty. This uncertainty quantification helps people know how far off from the "truth" the point estimate might be. AI is very powerful in providing point estimates, but struggles to provide the uncertainty of its estimates. In other words, AI does not know what it does not know, and thus AI will always provide an answer "confidently," whether the answer is accurate or not. Application of the concepts of statistical significance and especially equivalence testing are somewhat new to AI, and in our view, lacking such important concepts is unfortunate.

When two objects are to be compared, SI will test the hypothesis $H_0 : \delta = 0$, for example, where δ is the "true" difference in the population, while AI will typically perform a direct comparison. For example, a tourist visiting China saw a beautiful piece of ancient art at Xi'An. He asked the owner how old the artwork was. The owner replied that it was 5,003 years old. The tourist was puzzled by the exactness of the answer, and asked "How do you know it is exactly 5,003 years old?" The owner replied "Well, three years ago, a university professor dated it at 5,000 years old,

so it must be 5,003 years old now". It is unlikely that a statistician would accept this argument, but an AI system might.

While this example may seem trivial, similar situations occur in the determination of "winners" of data competitions. The winning participant may, hypothetically, produce a mean absolute error (MAE) in predicting the hold-out set of 0.3245638, while several other participants produce MAEs of 0.3245639. Is the winner's model really "best"? Would an equivalence test suggest that the winner's solution should be considered different than those of the runner-ups? We argue that AI needs to implement such SI and BI considerations into future systems.

When publishing research, statisticians, like mathematicians, aim to show that their conclusions are correct in all cases, or at least under a wide set of reasonable assumptions. That is, they will typically make an argument that their results can be applied beyond any actual data they have analyzed in the paper. We consider this a form of extrapolation – looking beyond the current data. On the hand, it is much more common in AI publications to demonstrate that one's model or prediction system works very well for a given set of data, that is, under very special conditions. In our view this is a form of interpolation – looking within the current data. Note that the lack of distinction between the sample and the population of interest naturally leads to such a viewpoint. In other words, answers from AI tend to be very good, but they provide no guarantee, or even confidence, that they will function well when "outside the box." This relates to our previous discussion of needing a "basis for inference." As one example, an AI system to transcribe handwritten English script into digital form may work quite well, but this provides no evidence that the same system will work well with handwritten Mandarin script.

By "domain knowledge" we mean everything that is currently known about a given phenomenon under study. Such knowledge may or may not be properly documented, and even if documented, it may or may not be digitized. For example, experienced operators may be able to detect a "weird" noise coming from a noisy manufacturing process that they have never heard before. They may, however, be unable to easily train a new operator as to which noises are "weird," and need to be investigated, and which are not. Domain knowledge is obviously associated with BI.

It has been recognized for some time that AI lacks "common sense," that is, basic domain knowledge, frequently resulting in "dumb" errors. We recall the embarrassing but humorous case of Amazon's automated pricing algorithm that suggested an introductory biology textbook should sell for $24 million [51]. Obviously, any child with even a modicum of BI would have known that this was a ridiculous price. BI can augment AI to help ensure that such blunders are avoided, hopefully when they involve more serious issues, such as airplane flight management systems or medical treatment. Even if not originally trained to do so, practicing statisticians learned long ago that integration of BI is critical to proper data analysis.

6.5 | Data Quality: Data Are Right Versus Right Data

Data quality has long been recognized as critical to the validity of any empirical analysis, whether it be in traditional statistics, machine learning, or an AI system. In our view, however, most data quality evaluations have focused almost exclusively on the question of whether the data are right or not. That is, are there errors, blunders, outliers, missing value codes of −999 or perhaps 0, and so on?

Ensuring that the data are right is of course an important concern, and bad data are unfortunately ubiquitous. This is why most depictions of the model-building process in AI incorporate a "data cleaning" step. However, we ask if it makes sense to tolerate sloppy data collection, assuming that the bad data will later be filtered out? How do we know if the unusual data are actually "bad" data, versus unusual but accurate? That is, how do we avoid "throwing the baby out with the bath water?" Would it not make more sense to focus efforts on improving our data collection processes to avoid bad data in the first place? This is basic process control, a historical part of SI. Modern technologies, such as sensors, RFID, or auto-camera, allow efficient means of accurate data collection. Of course, technology does not guarantee data accuracy. Therefore, any sensors utilized must have evaluation and calibration systems to ensure accuracy. No one wants another 737 Max catastrophe.

Admitting that the "data are right" question is important, we argue that the question of whether these are the *right data* to address the current problem is even more critical, but often overlooked. The growth of data competitions has exacerbated this problem, since in such competitions the data are generally provided as a "given," and the sole objective is to fit the provided data as closely as possible. Rarely is the original practical problem clearly defined, nor is there an opportunity to discuss what data might be most appropriate for that specific problem. In other words, the "right data" question is not even considered.

We argue that many of the well-publicized AI failures, such as the Amazon facial recognition system that matched pictures of twenty-eight known criminals to members of the US Congress, are the result of not using the right data to train the system [52]. While this particular blunder might seem funny at first, almost all of the mismatches involved African-American and Hispanic members of Congress. In this case, the training data did not include sufficient people of color, an oversight some have attributed to the lack of diversity at Amazon.

As noted by Li [53], the data that are readily available are rarely the most appropriate data, that is, the right data, to address a specific question or problem. This realization led Ronald Fisher, among others, to the development of the field of experimental design, as noted previously. Since Fisher published *The Design of Experiments* [54], statisticians have been studying, researching, and practicing the principles of experimental design to obtain the right data for a well-defined problem. This thinking is a critical advantage of SI and one that we feel statisticians have not emphasized sufficiently. Even in cases where designed experiments are not feasible, the principles of obtaining the right data for a particular problem are still relevant.

There are, of course, many aspects of the right data, all of which should be documented in a complete data pedigree [55], to be evaluated relative to the problem at hand. We highlight just a few criteria here:

- Specific relevancy: The data should be the most relevant data possible to the specific problem at hand. For example, data on the spread of COVID in children in Australia are relevant to addressing COVID, but would not be the right data to address the spread of COVID among the elderly in the US.
- Completeness: The data set should include all relevant variables and observations, minimizing any "dark data" [56] or "ghost data" [57], that is, lurking variables or relevant data that are hidden from the analyst. To consider this criterion, the problem in need of a solution, the population of interest, and the potential future use of any models generated must be defined and clarified.
- Freedom from bias: Many types of biases can be hidden in data; some intentional, some unintentional.
- Timeliness: As an obvious example, financial and economic modeling are dependent on the timeframe of the data collection, relative to the potential use of models in the future.

Interestingly, when obtaining data from `data.gov`, `Kaggle.com`, the UC Irvine machine learning depository, or similar online sites, a full data pedigree that would allow evaluation versus these criteria is rarely provided. It is virtually impossible in most cases to determine if the data provided are, in fact, the right data.

We feel strongly that for both the "data are right" and the "right data" questions, SI has significant value to add.

6.6 | Summary

While AI has much to contribute to society, it is clear that to date, the hype has exceeded the tangible successes. Further, the discussions between AI researchers, data scientists, and statisticians have too often been pejorative and competitive. We argue that with greater cooperation across disciplines, the future of AI can greatly exceed its current state. An important step is to recognize that AI becomes much "smarter" when it incorporates other types of intelligence, such as SI and BI. Along the same lines, statisticians need to make a clearer case as to why they deserve a "seat at the table" in AI development. Key points to emphasize include the importance of focusing on the right data rather than big data, and the benefits of a more holistic evaluation of models, beyond train/test splits. While it is in the best interests of statisticians to learn as much about machine learning and AI as possible, we repeat that this should be accomplished without sacrificing expertise in core statistical principles, which only those well-trained in statistics can provide.

7 | Ron S. Kenett

Before presenting comments on the topic under discussion, I would like to recognize the contribution of *Applied Stochastic Models in Business and Industry* (ASMBI) to the scientific body of knowledge on applied stochastic models in business and industry. In their seminal book, Efron and Hastie discuss the development of statistics since the 19th century using a simplex with nodes on Application, Mathematics and Computations [58], page 448. The path started on the Application corner in the 1900s, moved to the Mathematics corner in the 1950s, and then shifted to the Computation side with a crossroad in 1995 splitting into two paths, one leading back to Applications. ASMBI helped the movement towards that fork where stochastics are considered in the context of business and industrial applications.

The advances in analytics under headings of artificial intelligence (AI), machine learning (ML), or deep learning (DL) have changed the approach to data analysis. Part of this change followed the big data, sensor technology, and computational capabilities advances. Kenett and Francq [59] review this evolution and propose checklists to assess applied statistics studies. This range of approaches puts varying roles to model-based probability methods and empirical data-driven model assessment. In general, the complementary role of statistics, stochastics, and AI deserves much discussion.[3] In assessing model performance, one often uses cross-validation. Kenett et al. [60] discuss the importance of accounting for the data generation process. They propose an approach labeled befitting cross-validation (BCV) to ensure that the validation approach is appropriate and permits to generalize of the model findings.

BCV is particularly important in time series [61]. In the industrial context, the data generation process is complex and describable [62]. For examples of cybermanufacturing and digital twins see [63, 64]. This presents a contrast between big data from contexts such as social media to examples with physics-based models and engineering considerations. It circumvents the role of stochastics.

We see a future in stochastic modeling in business and industry, in combination with AI, ML, and DL methods. This brings up the need to achieve an appropriate balance in specific situations. A methodology for assessing the right balance between purely data-driven empirical methods, physics-based models, and probabilistic stochastic models requires further development. This is necessary for the effective and robust achievement of data-driven knowledge.

Another challenge associated with this forking path is to provide the skills and experience necessary for professionals to meet this balance. This requires substantial revisions in educational curricula where statistics and data science method need to be merged with physics and engineering. Part of this needs an emphasis on software delivery platforms such as R, Python, JMP, and MINITAB; see [64–66].

Finally, a topic of growing interest in industry and engineering is the role of digital twins. This digital platform that accompanies a physical system provides enhanced monitoring, diagnostic, prognostic, and optimization capabilities. Such digital twins provide new opportunities for applied stochastics. Some references are [67–69].

8 | Wai Keung Li and Philip L.H. Yu

Would stochastic modeling (SM) still be useful with the impressive performance of machine learning (ML) methods? We think that the answer lies in what questions you want to answer and what sort of datasets you have. If your question is predicting the

occurrence of a future event, which may be the future conditional mean of a stochastic process and you have a large dataset with many variables then ML methods may be a first choice. However, in many investigations predictions may not be the only objective and one would like to have more understanding of the underlying data-generating process. It is true that all models may be only approximations at best but a good one should be able to give us insight into why the data are behaving in a certain way. In such cases, classical stochastic modeling may provide more insights into the data-generating mechanism, especially if the number of observations is larger than the number of variables [70, 71]. Even in the context of doing point predictions of future time series the knowledge of the underlying probability model can provide many insights into the interpretation cum quality of the prediction [72]. Considered a three-component mixture autoregressive model for the first difference y_t of the IBM stock price data from 1961 to 1962 with 369 observations [73]. From the modeling result (example 1 in [72]), one-step ahead predictive distributions for the IBM data can be obtained easily. It was observed that the predictive distributions for different time points in the dataset would exhibit distinct bimodality when the volatility of y_t is high. In contrast, unimodality was observed when the market was less volatile. In other words, the market would have higher chances of a sharp increase or decrease when volatility was high. In such a case a point forecast of the future time series would not be informative but knowledge of bimodality about the predictive distribution and the fitted mixture model would be useful to the investigator and risk manager.

Furthermore, in many applications, interest may not be in the future estimate but in a set of estimates. In such cases, a stochastic model would easily lend itself to provide the set of estimates which may be in the form of an interval. Indeed, if a Bayesian approach is adopted various predictive or tolerance intervals may be obtained under different coverage criteria. Please refer to the work by [74] for a classical and thorough discussion on statistical prediction.

There is no doubt about the usefulness and potential of ML methods. However, traditional stochastic modeling would provide an alternative even complementary view of the data that may be beneficial to the investigators. In this connection [75], compares forecast performances of the two approaches and finds that ML methods do not differ dramatically from the stochastic, while none of the methods under comparison dominates the other. From a scientific point of view, both tools would be valuable to an investigator who would like to seek a better understanding of his/her problem.

9 | Jean-Michel Poggi

As an applied statistician in academia, I would like to highlight, within the specific domain of time series forecasting in industry, two common challenges where machine learning (ML) methods are particularly promising, and one where stochastic modeling remains valuable, especially from an applied point of view.

9.1 | Sequential Learning

Sequential aggregation of individual predictions explores the principles of aggregating a group of experts and methods for weighting and integrating these "experts" (see [76] and [77]). The challenge is to predict y_t, given past values up to the moment t: $y_1, \ldots, y_{t-1}$, using K experts whose only known information consists of their immediate predictions of y_t and the history of their predictions. The idea is to optimally combine the predictions by adjusting the weights at each step based on instantaneous losses (e.g., quadratic). Observations and forecasts are made sequentially, without any stochastic modeling. The theoretical goal is to forecast almost as accurately as the best expert, but this expert is an oracle, since it remains unidentified in real-time and can only be determined at the end of the period (at time T).

Theoretical guarantees exist for several methods. Approaches such as EWA (Exponential Weighted Average) result in a convex combination of experts, ideal for mixing unbiased experts. Techniques that optimize a global criterion based on the history of measurements and expert predictions at each step, such as RR (Ridge Regression type criterion), become particularly relevant when the set of experts includes biased ones by relaxing the convexity constraint.

As a result, sequential prediction allows multiple models constructed under very different assumptions to be mixed in a unified and agnostic manner, as it does not require prior knowledge of how each expert internally generates predictions.

9.1.1 | Agnostic Combination of Heterogeneous Forecasts

The first example deals with heterogeneous experts to predict the concentration for the next day. In [78] a dataset from April 2013 to March 2014, is considered, giving the measures in the monitoring station of Normandy of the daily average concentration of PM10 and the corresponding forecasts of the day for the day after coming from 10 different prediction models. These models comprise 6 statistical models developed regionally by national air quality associations, 3 numerical models at varying spatial resolutions from meteorological agencies, and 1 classical baseline model known as the persistence model. Even if we had some information on the design ingredients for some of these models, we could not take this into account to combine the forecasts properly.

Sequential prediction allows several models based on very different assumptions to be mixed in a unified approach, and secondly, it does not require any prior knowledge of the internal way in which each expert generates predictions. It is therefore particularly relevant for combining the outputs of models of different types (statistical models and physical-chemical deterministic models). The sequential prediction strategy significantly improves the performance of the best expert, both in terms of errors and alerts, and, for the non-convex weighting

strategy, achieves the "unbiasedness" of the observed-predicted scatterplot, which is extremely difficult to obtain with classical methods.

In this example, we have a given set of experts corresponding to the outputs of different models and sequential aggregation provides an agnostic way to optimally combine them.

9.1.2 | Multiscale Selection for Disaggregated Forecasting

The second example is different and illustrates another aspect related to the flexibility of ML models. Indeed, the strategy starts with the generation of a large number of models, the experts, and sequential aggregation is used to select the appropriate ones and automatically reject the others. In the context of bottom-up forecasting of electricity demand (see [79]), starting from the consumption of individual customers, the problem is to forecast the aggregate demand for the next day. The challenge is to disaggregate the global signal in order to improve the forecasts. The idea is to find a hierarchical clustering that generates groups of customers at different scales, predicts the demand of the different groups, and combines them to get the best aggregate forecast.

A solution involving the sequential aggregation of multiscale random forest-based experts is provided in [80]. It considers N individual customers and the problem is to disaggregate the global signal to improve the forecast of global demand. The bottom-up forecasting strategy consists of grouping individuals into clusters, fitting forecasting models to each cluster, and aggregating the forecasts to predict the total. The intuition is that the population could be divided into sub-populations with different consumption habits, requiring different models.

The approach is to build experts using random forests trained on some subsets of customers, then normalize their predictions and aggregate them using a convex expert aggregation algorithm to predict system load. This leads to the automatic generation of a large number of models from clusters at different scales using random forests, a flexible non-parametric method introduced by Breiman [81] that allows exogenous variables to be easily considered with only a small number of hyper-parameters to be tuned. The final step is to combine these different model outputs to produce a forecast of aggregate demand at time t.

Applying these ideas to the well-known Irish public dataset, the new aggregation method is compared with two strategies for building subsets of customers: Hierarchical clustering based on survey data and/or load characteristics, and the baseline: Random clustering strategy. We find that disaggregation leads to a large gain (but no more than random) of about 25% in terms of error.

Random forests provide useful predictors for all aggregation scales, but also many irrelevant ones and crucial additional gains are obtained thanks to the sequential combination of group predictors.

9.1.3 | Finite Mixture of Regression Models for Forecasting

On the contrary, I would like to highlight a situation where stochastic modeling is still valuable, also from an applied perspective. The context of [82] deals with the massive amount of data on individual electricity consumption provided by new metering technologies and smart grids, for load profiling and load modeling at different scales of the electricity network.

A methodology based on a mixture of high-dimensional regression models is used to perform clustering of individual customers. The theoretical framework is a finite mixture of regression models to account for forecasting (the model selection step is theoretically justified in [83]) combined with the partitioning of the electrical signal into successive curves to consider it as functional data. Focusing on the discovery of clusters corresponding to different regression models, which could then be used directly for profiling, but can also be useful for forecasting purposes, the method is able to extract nice features from individual consumption data with little information (2 days of individual consumption) and no other prior.

The statistical approach allows a deeper analysis of the use of the internal objects of the method from a practical perspective, focusing not only on the results of the method but also on the by-products of the method, providing visualization tools to understand the estimates and facilitating interpretation.

10 | Marco Seabra Reis

10.1 | Introduction

The recent advances in Artificial Intelligence & Machine Learning (AI/ML) technology in the fields of image & video analysis and natural language processing (NLP) have generated a plethora of methods and tools, spiking the interest of the research community to explore their application outside these domains, namely in the process industry, which includes the chemical, food, biotechnological, semiconductor, and pharmaceutical sectors, among others. This boost in analytical diversity has also increased the difficulty of finding the best approach to apply in each situation, as well as fully understanding the multiple underlying rationales for applying each one of them. Moreover, there is a certain subliminal message currently being passed that, sooner or later, modern/future AI methods will make classic statistical/ML methods outdated. Therefore, a shadow of doubt is cast about the focus given in the future to statistical and probabilistic foundations of data analysis during the initial training of future professionals, which may be reduced to accommodate the new AI methods.

Science evolves, and so does the way it is taught and applied. Data analysis, in its broadest scope, is not an exception. The diversity of areas interested in knowledge induction from all forms of data is growing tremendously. More room should be given to learn about the analytical methodologies available, but also when each one of them is likely to bring value to the analysis and generate information with higher quality [84, 85]. Therefore, in this short

contribution, I will address different application scenarios from the Process Industry and present the analytical solutions that generate more information or meet more effectively their goals.

10.2 | Application Scenarios From the Process Industry

10.2.1 | Quality by Design (Qbd)

The capability of designing new products or processes with a high degree of confidence that they meet the target specifications is a valuable resource for companies. Its importance and supporting methods have been developed by Juran and others [86] and gained renewed attention and interest in the period of 2004–2008 with the publication of the ICH-Q8 guidelines and the FDA's PAT initiative in the scope of the highly regulated pharmaceutical sector. Such guidelines put forward the concept of Design Space, as *the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality* [87]. The Design Space is obtained after a risk assessment process where the potential aspects with impact on Critical Quality Attributes are identified and put to test, usually through a suitable Design of Experiments (DOE) methodology. The data available in these studies is scarce, but highly valuable, as well as all the pharmaceutical and engineering knowledge. Classical DOE methods, regression modeling, and ANOVA analysis are among the analytical approaches most often adopted in QbD. The scarcity of data and the interpretative limitations of Deep AI methods make them not so suited for this activity, even though applications have been found for fast screening of drugs and molecular discovery [88, 89]. The use of historical data, especially for legacy products, is also a possibility usually referred to as *retrospective Quality by Design, rQbD* [90]. Furthermore, DOE methodologies have been developed to handle new scenarios of QbD, such as copying with multivariate responses [91, 92] and profiles/functional responses [93], under a frequentist or Bayesian setting [94, 95]. Some AI/ML-driven approaches have also been suggested to be applied for designing experiments, such as Bayesian Optimization, Derivative Free Optimization, or Reinforcement Learning. However, they tend to be less efficient when the number of experiments to run is low, finding applications in fields where budget restrictions are not so strong and assumptions about the convexity of the surface response are harder to make (or it is likely that the response surface has a complex shape), such as in the analysis of computer experiments or robotics [96–99].

10.2.2 | Process Improvement by Variance Reduction

Process improvement by variance reduction is a common challenge across the industry. It involves exploring all potential sources for injecting variance into the system, assessing their impact, and developing mitigating solutions. This process typically has several stages, including analyzing measurement systems, collecting data from different sources, suggesting new sensors and data collecting systems, discussing intermediate results with plant personnel, etc. Solutions may address the removal of special causes (using statistical process monitoring methods), a reduction of common causes (e.g., using feedback process control), or both. In any case, the solution is not obtained by submitting data to a modern AI method and, in a single pass, obtaining the final outcome. Rather, it results from an interactive data collection and analysis process, using mostly classical methods (rational sampling, variance components, regression analysis, control charts, etc.), where the next step is decided depending on the results obtained in the previous, until a suitable solution is found. For example, the author's team developed a variance reduction solution for a major cork producer, following a systematic analysis process with multiple data analysis/experimentation cycles. In the end, EWMA control charts were developed for monitoring the innovations (incorporating the IMA behavior for the disturbances), as well as to know the current estimate of the average level of the quality variable (cork stopper density) and the current machine variation. The average level was then used to adjust the process by feedback control, using a discrete control with integral action designed for optimally compensating a process with an IMA disturbance. The pilot implementation of this monitoring & feedback control strategy in the process for one month led to an estimated decrease in the process variation (standard deviation) from 45.4% to 67.7% in one type of raw material and from 47.3% to 54.3% in another type of raw material. Noting that this industrial unit produces roughly 4,5 million cork stoppers per day, it is possible to have a sense of the impact this reduction can have on the quality of the process and business bottom-line results.

10.2.3 | Monitoring Large-Scale Assembling Processes

Even when data abounds, variance management and reduction may not be possible via modern AI tools. One reason is that, despite all the data collected, common cause variation may not be fully represented there, especially if some variation components only change over larger time horizons (e.g., lot-to-lot variation). One such case is reported in [100], regarding a Surface Mount Technology (SMT) production line of Bosch Car Multimedia, where more than 17 thousand product variables are simultaneously monitored. The information available from the earlier dozens of lots is not sufficient to develop control limits for future operation, given the underrepresentation of certain "normal" operational zones. However, the project team has deep knowledge of the process and the factors inducing such long-term variations. Such knowledge, including the plausible statistical distribution of certain variation sources, was translated into a Digital Twin of the process, which was used to realistically emulate long-term variation, which finally allowed to establish multivariate control limits that remain valid across future lots. This solution relied on deep knowledge, rather than deep learning from a high volume of data available, to make up for the underrepresented process variation. For more on this case study, see [101, 102]. In summary, it is often the case in Process Industry that the amount of data is not paralleled by the information needed to address the analytical goal. Here, the information was not sufficient to delimit the normal operation conditions, for which expert knowledge was used. Another aspect where AI frameworks face difficulties in process monitoring applications regards *fault diagnosis* – the process of delimiting or even finding the root cause of the abnormality, once a detection takes place. This requires cause-effect reasoning, which is not embedded in current AI models. In this regard, classical approaches can be

integrated with process knowledge (as described above, or in [103]) or even causal discovery and inference, providing more effective solutions to this stage of analysis [104–108].

10.2.4 | Inferential Models From Process Data (Process Soft Sensors)

The development of predictive models for relevant end-product quality properties from data contained in large process historians, databases or data lakes, is an area where the number of deep AI/ML applications found in the literature is increasing. Such quality properties tend to be available less frequently due to the nature of the associated measurement systems, which are usually obtained offline, with significant delays, involving complex, expensive, and time-consuming experimental protocols [109–111]. Given their importance, such variables are often the target for process supervision, monitoring, control, and optimization, for which more frequent estimates are required. *Soft sensors* are inferential models developed to achieve this goal, bringing other associated benefits, such as a reduction of the inspection overhead, improvements in the consistency of the final product quality, and in process efficiency, among others [109, 112–115]. As they are based on process data, they are also referred to *process soft sensors*. Many such models have been developed for continuous [116–118] and batch processes [119–124]. Applications include the prediction of compositions from the outgoing streams in distillation columns [125], the prediction of the Research Octane Number (RON) in industrial catalytic reforming units [116, 126], the estimation of the product's quality in batch polymerization processes [127], the estimation of cement properties [128]; and the prediction of NO_x and CO_2 emissions in industrial boilers [129] and commercial ships [130]. Despite the large number of applications, certain features of industrial data are, however, sometimes overlooked and should be kept in mind when developing process soft sensors, as they limit the range of applications of these models and provide clues on the type of predictive frameworks to use. The following list summarizes some relevant ones, according to our experience:

- Data collected span narrow operational regions. Processes are designed to operate around set points (or trajectories in the case of batch processes), with minimum variation. Therefore, the dynamic excitation is small, and the information available for effectively learning the relevant relationships is scarce. This limits the quality of the models estimated. Additionally, some variables remain "silent", that is, not frequently manipulated, and their influence cannot be assessed or reliably inferred.
- The narrow operational regions have another consequence. According to the Taylor series expansion, even when the process is non-linear, the linear part dominates in the neighborhood of the reference point, namely of the process set point. Therefore, it is often the case that linear methods perform as well, or even better than the non-linear ones (especially those methods that handle well features like collinearity, sparsity, and uncertainty [111, 131–135]), and is not expected that deep AI models bring much added-value in such low-excitation, information-poor settings.
- As data is collected in a passive way (*observational* data), process soft sensors are, in general, acausal, and therefore should not be used for process control and optimization, but only for prediction of the target variables under normal operating conditions. This misuse of process soft sensors is sometimes seen or implied in applications and should be avoided.
- Certain process features, such as long-term degradation patterns, are known to exist and must be addressed during the modeling stage. One may argue that with enough data, deep learning methods can apply to their feature extraction capabilities and be able to figure out the long-term degradation trend and include it tacitly in a deep model. For instance, in a different context, the deep convolutional neural network VGG-16 was able to successfully classify objects from 1,000 categories using a training dataset composed of 1.3 million images (it won the ImageNet Large-Scale Visual Recognition Challenge, ILSVRC). This approximately 1000:1 ratio for a number of instances, class, could be translated, with the due adaptation, to roughly 1,000 degradation patterns being necessary for achieving a similar performance. However, if each degradation cycle runs for 2 years, as happens with catalyst deactivation in certain industrial chemical processes, this would require approximately 2,000 years of uninterrupted operation data. This is obviously impossible, but using expert knowledge and engineering principles, together with stochastic modeling, a solution can still be found.

10.2.5 | Image-Based Classification

Among the areas of application of modern AI where major achievements have been reported, one can find those related to image analysis and classification. These areas are also among the first successful applications of deep learning methods, just before the emergence of the generative large natural language processing models that have been attracting much attention recently. Deep learning models are composed of millions of adjustable parameters, requiring equally large databases for their learning with diverse instances spanning all relevant aspects to the prediction task. This richness of the training dataset is crucial for estimating the huge number of parameters of the network (a stable learning method is also very important and has motivated the development of a variety of stochastic gradient descent methods). From what was stated in the previous paragraphs, it is not likely that process data, including images, can be collected in such large amounts and variety, for adopting a deep learning neural network in a classification problem. However, image classification problems may benefit from information extracted elsewhere if it can be assumed that it contains relevant information. We call this property, *knowledge transferability*. Using a human analogy, humans learn to identify persons, animals, surrounding objects, etc., and are able to transfer such knowledge to identify shapes, classes, and even functions of objects they have never seen or worked with. In the same way, it often happens that artificial neural network models developed to perform certain image classification tasks can be useful to address others, if the knowledge gathered earlier is updated and relevant for the new activity. One example from the process industry can be found in [136], where a pre-trained deep learning model, with more than 130M of parameters was adapted (through transfer learning) to a different task,

using only 2,961 images for training and 1777 for validation, and improved the prediction performance under independence testing conditions (the test set was composed of 1,185 images). This was only possible because of the plausibility of *transferability* conditions, which makes the effective training sample size much larger than the available image set when a pre-trained network can be used.

10.3 | Discussion and Conclusions

From the situations described above, spanning different application scenarios in the process industry, some guidelines can be extracted (whose validity is not claimed to be universal) on the use of classical data-driven methods versus modern AI/ML approaches, which can be summarized as follows:

Use stochastic modeling and classical data-driven approaches when/for:

- *Quality by Design* activities;
- Complex process improvement problems (with multiple stages with conditional decisions);
- An intense and iterative interaction with the process is necessary to develop the solution;
- Unstructured variation is an important part of the problem definition, analysis, and solution strategy;
- Exploiting tacit/expert knowledge is required or recommended;
- Data sets available are small and regard different aspects of the problem;
- Data have low operational variation;
- Presence of long-term degradation behavior;
- Transferability conditions are not met.

Use deep AI approaches when/for:

- Simple, well-defined, repetitive problems;
- Closed solution spaces (even though they can be large);
- Unstructured variation is not an important part of the problem definition, analysis, and solution strategy;
- Analysis tasks can be defined a priori and programmed;
- No interaction with the process is required. All the relevant information lies in the collected data;
- Data sets available are large and rich (representative of the different aspects of the problem);
- There are no major non-stationary components.
- Transferability conditions are met.

11 | Fabrizio Ruggeri

I grew up in a scientific world where I was used to thinking about the properties of the system I was interested in and then looking for an appropriate model among the many developed in literature and using it, with possible adaptations. Years ago, I was interested in modeling gas escapes in a city network with the aim of suggesting a replacement policy starting from the pipelines most subject to gas escapes. These pipelines differ in many ways, such as materials and installation conditions. We considered the gas escapes as rare (and countable) events, and we thought of using a Poisson process to model them. The first issue was about the aging of the pipelines: The experts told us that the cast iron pipelines were not subject to corrosion, while steel ones were. We translated such information into homogeneous (HPP) and non-homogeneous (NHPP) Poisson processes, respectively. (see [137] and [138] respectively). We faced many practical problems, typical in a statistical analysis. We had to deal with bad quality data since six gas escapes, out of 30 in 6 years, occurred in different parts of the city and were recorded in a 24 h period in which nothing significant, such as an earthquake, had occurred. We considered whether there were differences between steel pipelines installed in different years, since we had to split them by the installation time (we chose years), and their age at escape matters more than calendar time. We had to model the gas escapes considering different cases, thinking if they had the same, completely different, or similar behavior. We translated that, in a Bayesian framework, in performing a unique analysis with a common parameter for all the years, a separate analysis with different parameters for all years, or a hierarchical model with different parameters coming from the same distribution, respectively. To get a better model, we had even to resort to Bayesian non-parametric, namely Gamma processes as conjugate prior for the mean value function of a Poisson process [139]. We also had to propose a new intensity function for an NHPP [140] when forecasting the reliability of train doors before warrant expiration to help the transportation company decide if the doors were reliable as stipulated in the contract with the manufacturer. We had to deal with seasonality and different time scales (calendar and kilometers traveled). In both examples, gas escapes and the failure of train doors, we built models that were "explaining" what was going on. I am confident that, without "traditional" stochastic models like those in the cited papers, someone could train a Deep Neural Network which will make, for example, predictions on future failures of the train doors. But what about the explainability of the results and our conclusions, which led us to identify the worst combination of installation conditions citecagno and the effect of different causes on the failure of train doors [141]? Here, we have a limited number of data and parameters, so it might be difficult to think of reliable training and testing data (but this might be the biased opinion of a statistician!)

These are the thoughts, combined with my experience as Editor-in-Chief of ASMBI for 17 years and participant in many conferences in business and industrial statistics, which led me to ask many researchers the question which is the title of this paper.

I am a mathematical statistician and as such I love working on "traditional" stochastic processes [142] as well as more fundamental aspects, such as Bayesian robustness, but I think it is appropriate for statisticians to take up the challenge and see if cooperation between our field and Machine Learning and Artificial Intelligence can be fruitful. This is already happening, not only because many people work at the interface of those fields, but also because it is not always clear when one begins and the

other ends! I am making some comments on possible and actual collaboration, which are part of my recent or planned research. The first one concerns the choice of priors in a Bayesian framework. Everyone who has tried to get a prior from experts is aware of the shortcomings of the process [143], sometimes related to the use of questionable methods, such as the Analytic Hierarchy Process [144] that uses qualitative judgments to obtain probabilities, as in [137].

Problems in the Big Data era involve so many parameters that it is impossible for a human to derive appropriate priors on all of them. However, a Bayesian approach is still used, with non-informative priors, for several reasons, such as the relative ease of making predictions. My research question here is whether it is possible (positive opinion, at least after proper training of the algorithms) and useful (very uncertain about it!) to try to obtain prior distributions using, for example, ChatGPT. The move away from "traditional" stochastic models is sometimes motivated by the greatly increased complexity of the problems being addressed which involve a huge number of parameters. Computational methods have been developed, and deep neural networks are just one example among many. I wonder whether it is not possible to consider, in some situations, (dynamic) Bayesian networks (BNs) that can split a complex problem into a simpler, but not simple, one, preserving the interpretability of the results. There are many computational challenges, also due to the need to propagate the acquired evidence, but there are some works that try to reduce this burden. Clustering has been proposed as a method to improve computational efficiency and/or improve model learning. One of the main motivating contexts for BN cluster mapping is complex systems, characterized by multiple interacting components and by emergent and complex behavior under uncertainty [145]. In a recent paper [146], we propose a new algorithm, DCMAP, that also exploits the notion of layers, that is, nodes that have the same maximum distance from the BN leaves. The algorithm, given an arbitrarily defined positive cost function (based on the number and type of operations), iteratively and rapidly finds near-optimal, hence optimal, cluster mappings. DCMP is applied in a case study based on a DBN of a seagrass ecosystem [147]. Seagrass ecosystems are a critical primary habitat for fish and many endangered species such as the dugong and green turtle, impacted by human activities such as dredging and interactions between biological, ecological, and environmental factors. The complexity of the system is also due to its inhomogeneity, caused by the switch between loss and recovery regimes. As a result, the DBN itself is complex and computationally intensive and inference is unfeasible within an MCMC framework. DCMAP was applied at 25 nodes, when considering only one time epoch and 50 nodes when considering two epochs, showing a clear improvement over previously used methods. Open questions concern the scalability of methods like this and the possible range of applications.

The final comments concern Adversarial Risk Analysis (ARA), a relatively new field that overcomes the practical limitations of game-theoretic approaches that require shared knowledge of preferences and beliefs (utilities and probabilities, respectively) by the agents involved (see [148] for more details on ARA). Here I focus my discussion on classifiers and their protection from attackers (STS and IJAR). Recently, an increasing number of processes are being automated using classification algorithms [149]. They need to be robust and protected from possible attacks in order to trust key operations based on their output. Security can be an issue in classification due to the presence of adversaries ready to modify the data to gain an advantage, influencing training and operations. These are aspects considered within the emerging field of Adversarial Machine Learning (AML) [150]. I refer to [151] and [152] for an in-depth explanation of the approach, but here I would like to mention that this is an example of integrating ML and statistical approaches, namely Bayesian. The Bayesian decision-theoretic approach is used to model beliefs and preferences, considering both the defender and the attacker as maximizers of expected utility. The defender knows his/her utility and probability, but makes assumptions about those of the attacker as random utilities and random probabilities. By finding the optimal decision of the attacker for each realization of the random utilities and probabilities, the defender can find the distribution of optimal attacks and use it to find his/her optimal decision.

I apologize for the many self-citations, but my contribution is mainly based on my own research and how the new "era" is influencing my work.

12 | Gilbert Saporta

If by stochastic modeling we are referring to generative models as described by L. Breiman [153], I am more from the "data-driven" culture where models are based on data. But which models are we talking about? As George Box (1987) wrote: *Essentially, all models are wrong, but some are useful*; there is no *true* model. For a long time, we were content with explicit models that could be represented by equations or rules, within the reach of anyone with a sufficiently high academic level. The models became more complex, and statistical learning theory validated the idea that we could predict without understanding [154]. According to L. Bottou [155], V. Vapnik [156] stated that *Better models are sometimes obtained by deliberately avoiding to reproduce the true mechanisms*. After all, we can use a car or a television without knowing its inner workings. There have been numerous successes in many areas of business and industrial statistics, such as fraud detection and purchasing recommendations. The famous economist H. Varian (also chief economist at Google) encouraged his fellow econometricians to take an interest in Machine Learning methods [157].

Let's take a classic example: Credit scoring. Logistic regression is considered by some to be a stochastic model because it produces individual or collective probabilities of default. But this is an abuse of the term, as it is very difficult to define a generative model of default. It is therefore an empirical approach. If this method has become the industry standard, it is thanks to the simplicity of its formulation, which also makes it possible to comply with regulatory requirements, in Europe at least. Numerous attempts to use more complex models have not led to significant progress [158].

The last two decades have seen both the deluge of data (Big Data) and the development of black-box models using thousands, if not millions, of parameters and highly sophisticated non-linear

architectures (Deep Learning). We even forget that Vapnik's theory of learning was a modern theory of parsimony!

The success of advanced Machine Learning methods in areas such as image and face recognition is spectacular, although there are a few shortcomings in terms of robustness, where the modification of a single pixel can dramatically change the prediction [159], but the intelligibility of the models was hardly questioned until recently.

The emergence of whistleblowing literature [160] motivated by unethical applications in automatic decisions concerning individuals, such as recruitment assistance, predictive justice and facial recognition, changed the situation.

On the one hand, there are attempts to make the algorithms explainable: This is the XAI, with methods that seek to open the black box, as it were, with new measures of variable importance, or the use of surrogate models to explain a decision using trees or local linear models. On the other hand, some researchers like Cynthia Rudin [161] are calling for black boxes to be abandoned in favor of simple models. It should be noted that works on features importance in ML, with Shapley measures for example, have renewed the classic problems [162] which already showed that even simple models are not so easy to interpret.

As well as being transparent, algorithms need to be fair and non-discriminatory when applied to human groups. Algorithmic fairness is a major area of development in computer science [163], but one in which statisticians are not yet very involved.

Transparency or explicability is not enough: If a model is not causal, which is in line with a definition of stochastic modeling, and is based solely on correlations, it can lead to erroneous conclusions. Big names in Machine Learning such as L. Bottou already quoted and B. Schölkopf [164] have established links with J.Pearl's theory of causality.

To conclude, I do not see any antagonism between statistical modeling and Machine Learning, which in some respects is its 21st-century version, just as the emergence of computer science led to the development of multivariate statistics in the second half of the 20th century. Machine Learning has enriched the statistician's toolbox and, above all, has provided him with the fundamental concept of generalization and the need to go beyond simply fitting a model on the basis of training data alone. As early as 1941, Paul Horst et al. [165] wrote that *the usefulness of a prediction procedure is not established when it is found to predict adequately on the original sample; the necessary next step must be its application to at least a second group. Only if it predicts adequately on subsequent samples can the value of the procedure be regarded as established*, but his lesson had been lost for a long time!

For their part, statisticians can provide Machine Learning practitioners with their knowledge of biases, their sense of data (missing, aberrant, etc.), and their culture to avoid reinventing techniques such as categorical data encoding [166] or principal component analysis! But statisticians must not be afraid to enter this new field, otherwise, they will be marginalized.

13 | Piercesare Secchi

Let me begin my contribution with two quotations. The first one is from Chris Anderson [167]:

> We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

In the same Wired's article, Anderson attributes to Peter Norvig, a well-known American computer scientist who was director of research at Google, the revision of George Box's famous aphorism — "all models are wrong, but some are useful" — into: "All models are wrong, and increasingly you can succeed without them". Peter Norvig [168] disavowed it, but the quote remains indicative of a mainstream belief.

In fact, Anderson focuses on prediction and finds his captivating conclusions on the observation that good predictions could be generated by algorithmic black boxes trained on massive amounts of data, those black boxes that today we consider the products of Machine Learning and Artificial Intelligence. My second quotation is from Carlo Rovelli [169], translated from Italian by me; it reminds us that a scientific model does not have as sole purpose that of making predictions:

> The goal of science is not to make predictions. It is also offering an image of reality, a conceptual framework for thinking about things. This ambition has made scientific thinking effective. If the goal of science were only predictions, Copernicus would not have discovered anything compared to Ptolemy: His astronomical predictions were no better than Ptolemy's. But Copernicus found a key to rethink everything and understand better.[4]

In [170] we pointed out that a model is not a magic box; rather, it serves as a powerful tool that enhances the capabilities of human knowledge. Its true value lies in its capacity to increase and refine our understanding of complex systems. This expansion occurs when the model represents the interactions and dependencies among the variable entities identified by stakeholders and data scientists as essential for describing the system under investigation. By effectively capturing these relationships, a model becomes a conduit for informed decision-making. It empowers stakeholders to make sense of intricate data landscapes by allowing them to manipulate independent input variables and observe their effects on the system's behavior. Furthermore, a model offers a platform for experimentation and scenario analysis, enabling stakeholders to simulate various conditions and explore potential outcomes. A model, that conforms to the data for representing a phenomenon in which variability is a constitutive element, must be open to quantifying uncertainty. This uncertainty arises from the variability of the observed phenomenon, from the imprecision of the formal and computational representation, from the error in predictions, and from the effects of actions taken based on plausible scenarios. In essence, a model

serves as a tool for understanding complex systems, facilitating decision-making, and exploring potential outcomes under different conditions. It embodies a balance, driven by data but ruled by the data scientist, between simplification and fidelity to reality, while acknowledging and quantifying the inherent uncertainty in the represented phenomenon.

Traditionally in Statistics, the model represents in an idealized mathematical form the data generating process. However, I will here argue that when dealing with complex data we should expand the notion of the model to include also the problem of *data representation*. Hence I will take the side with the data modeling culture [153] when it comes to representing the atoms of the statistical analysis, although sympathizing with the algorithmic modeling culture when it comes to their analysis, especially when the analysis is conducted in mathematical spaces distant from the standard finite-dimensional Euclidean ones, the realm of classical multivariate Statistics. To uphold my position, I'll leverage different personal experiences in real-world data analysis, where I contend that statistical data modeling *lato sensu* holds greater relevance than machine learning algorithms aimed solely at producing accurate predictions.

The emergence of new digital devices and systems capable of gathering data with fine temporal and spatial granularity—for example, medical imaging, mobile devices generating every few time instants data tracking individual positions, remote sensing data for environmental monitoring, automatic people counting data recovered by systems installed on vehicles of public transportation networks, 3D meshes obtained by x-ray computed tomography in additive manufacturing, data from sensors recording displacements and accelerations of large civil infrastructures, digital administrative data recorded by public administrations, . . . —has created a demand for innovative data analytics models and algorithms beyond those traditionally developed in multivariate Statistics. Object Oriented Data Analysis (OODA), introduced by Wang and Marron [171] (see also the recent book by Marron and Dryden [172]), is a branch of Statistics specializing in the interdisciplinary analysis of complex data. It is my belief that at the core of OODA is the modeling problem of *representation* wherein each raw datum—an object like a curve, an image, a network—is transformed through a *reduction by sufficiency* into an atom embedded in a mathematical space. This space must be suitable for the representation of the information carried by each datum and deemed relevant for the application, and must have enough mathematical structure to allow for the statistical analysis aimed at the understanding of the data's relevant variability. I am here referring to a notion of sufficiency that does not immediately fit the traditional Fisherian definition derived from a preceding assumption on the statistical model generating the data. Indeed, this reduction by sufficiency is not an agnostic process, but it is strongly driven by a prior knowledge of the phenomenon under study, explicitly or implicitly informed by a model founded on science (physics, chemistry, economics, medicine . . .) or on experience, which is however impractical to capture as a formal probabilistic model generating the data.

The Aneurisk65[5] data set [173] stands for a paradigmatic example. The AneuRisk Project, supported by Fondazione Politecnico di Milano and Siemens Medical Solutions Italia, was devoted to the study of cerebral aneurysms, investigating the role of vessel morphology on aneurysms pathogenesis, via the effects that the morphology has on the blood fluid dynamics within the vessel. The AneuRisk65 data set collects 65 bivariate smooth functions providing, as a function of the arc length of the vessel centerline, the local radius and the curvature of the internal carotid artery of 65 subjects; these are the atoms providing the sufficient representation for the statistical analysis. In fact, the original raw data were 65 collections of 100 B/W Xray-scans taken from 100 different angular perspectives of each subject head. For each subject, a B/W 3D-array was obtained; in this representation, the gray level of each voxel was related to the amount of flowing blood in that part of the head. Then, vessel surfaces were identified together with their centerlines [174]; finally smooth radius and curvature functions were estimated [175, 176] for all vessels, generating the final representation of the original raw data. This final representation was considered sufficient for the statistical analysis, because the theory of fluid dynamics in curved vessels posits that they are governed locally by a parameter known as the Dean number. This number is calculated based on factors related to the physical properties of the blood, such as viscosity, density, and velocity, as well as factors concerning the vessel's geometry, and in particular the local radius and curvature. This example illustrates how a Physics-based model drives each and every step of the complex pipeline which generates the representation of the atoms of the statistical analysis. Moreover, when exploring the variability of these functional data one is immediately confronted with the problem of their alignment, that is decoupling their phase variability—which, for this specific study, is ancillary—from their amplitude variability. A further reduction by sufficiency could indeed be reached with the tools of algebraic topology by representing these functions as merge trees [177]. A merge tree representation is in fact invariant under homeomorphic re-parametrizations of the arc length argument of the radius and curvature functions, thus allowing for a statistical analysis that is indifferent to their misalignment. Merge trees can be embedded in a metric space; non-parametric algorithms for supervised classification, even those developed by Machine Learning, can then be used for the construction of an accurate classifier of patients.

Analogous problems of representation emerge in additive manufacturing (AM), a new industrial production method that makes available novel types of complex shapes that go beyond traditionally manufactured geometries and 2.5D free-form surfaces. New challenges must be faced to characterize, model, and monitor the variability of such complex shapes. In [178] we analyzed the deviations between a set of measured shapes generated by polymer fused deposition modeling (FDM) and their nominal model. Traditionally the reconstructed geometry of an AM object is obtained via an x-ray computed tomography which generates a 3D mesh. This is compared with the nominal geometry of the object represented by the 3D mesh of the originating model. Since there is no one-to-one correspondence between the points in the two meshes, the deviations of points in the reconstructed geometry from the nominal do not coincide with the deviations of points in the nominal geometry from the reconstructed one. Hence, as a first step in reduction by sufficiency, we introduced two directional deviation maps, defined consistently with the Hausdorff distance between the reconstructed mesh and the nominal one. As a second step of reduction by sufficiency, we summarized these two deviation maps by means of the probability density

functions (PDFs) of their logarithms. Indeed, we were driven to this reduction by the prior experiential knowledge that anomalies in FDM are often generated by a local excess of material or by a local lack of it; hence the right tail of the PDF of the deviation map of the reconstructed geometry from the nominal might identify large anomalous deviations likely generated by a localized excess of material, while the right tail of the PDF of the deviation map of the nominal geometry from the reconstructed might highlight anomalies generated by a lack of material, for example, the absence of a strut in a trabecular geometry. Therefore, each raw datum—the 3D reconstructed mesh capturing the shape of the manufactured object—is now represented by a couple of PDFs summarizing the distribution of its local deviations from its nominal model. These PDFs are then embedded in a Bayes Hilbert space [179] where a further dimensional reduction can be carried out by means of Simplicial Functional Principal Component Analysis [180]. Profile monitoring of geometrical discrepancies can then be carried out based on Hotelling T^2 and Q statistics of the principal components scores [181].

When spatial dependence is the significant issue, Object Oriented Spatial Statistics (O2S2), as outlined in [182], offers a conceptual framework to address the novel challenges brought about by the geo-data revolution, leveraging a potent combination of geometric and topological approaches for the analysis. The main problem is here that of understanding spatial dependence through a model that allows for interpretability and also for prediction, for instance in sites where the stochastic field generating the data has not been observed. For example, in [183] we used O2S2 ideas to spatially predict the probability density function of dissolved oxygen in water—treated as an atomic datum embedded in a Bayes Hilbert space [179]—in each site of the Chesapeake Bay, the largest, most productive and biologically diverse estuary in North America, handling both the data and the domain complexity. In this case study, stationarity of the stochastic spatial field generating the distributional data is not a viable assumption, and yet through the localization of the Kriging model, suitably extended for the prediction of constrained functional data, we can estimate in each location of the Bay the probability that dissolved oxygen is below any given threshold, thus allowing for the identification of Dead Zones. Moreover, by extending the Kriging model to treat covariance data—which belong to the Riemannian manifold of positive definite matrices, a metric space without a vectorial structure—and by localizing it, in [184] we described the spatial dependence and variability of the covariance structure between surface temperature and dissolved oxygen, thus offering a first interpretable insight on the spatial variability of the joint distribution of these two important environmental descriptors. Spatially localizing the Kriging model takes care of the complexity of the spatial domain, characterized by the presence of holes and non-convexities, and also allows for the linearization of the manifold data. When the target variable is a real number, a different approach for dealing with the complexities of textured domains is that offered by Spatial Regression with Partial Differential Equation (SR-PDE), whose roots are in [185, 186]. This approach considers semiparametric regression models for spatial smoothing which implicitly capture the spatial dependence of the field generating the data by controlling the regularity of a deterministic term in the model which accounts for the spatial effects. The regression model is estimated by minimizing a penalized sum-of-squares-error where the penalization embodies prior knowledge of the phenomenon; for instance, it might penalize the misfit with respect to the solution of a set of differential equations which are believed to govern the system generating the data, at least in an ideal situation [187]. Furthermore, SR-PDE has the capability to adhere to particular conditions set at the boundaries of the spatial domain, a crucial aspect in numerous applications for acquiring significant estimates.

The personal experiences mentioned earlier were driven by real-world data analysis challenges relevant to industry or science. These challenges were addressed using a model-driven approach, although the model may not always capture the stochastic process behind the data generation. Nonetheless, these analyses provide a depiction of the phenomenon being studied that allows for interpretation and informed decision-making.

14 | Rituparna Sen

At this time, when machine learning(ML) and artificial intelligence have become very popular, it is necessary to introspect and answer this question. There is an obvious need to justify the continuation of pushing research in stochastic modeling. As a side effect, we expect to identify areas where machine learning methods need to be developed, if possible.

The main advantage of ML methods is in prediction tasks. This is most common in regression, clustering, classification, and time series or spatial forecasting settings. The set-up is that we have several examples to learn from and then we need to predict for other test cases. While this is a very important and common problem, stochastic analysis has a much broader scope. We discuss this in what follows.

Stochastic modeling starts with data collection. According to Ronald Fisher, the founder of modern statistics, "To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: He may be able to say what the experiment died of." [188] This statement remains true even today. The principles of randomization, replication, and blocking need to be taken into account for any designed experiment. For sample surveys, it is extremely important to ensure that common sources of bias, like selection, response, non-response, etc, do not creep in. Given the particular scenario, stochastic modeling provides us with a guiding principle on data collection for the most efficient utilization of resources geared towards answering the business and industry question at hand.

The set-up of statistical inference is somewhat different from prediction, see [189] for elaboration on this. For practical purposes in business and industry, the goal is often to predict. Even in those cases several limitations and discrepancies exist, as explained in [190]. But in a lot of situations, where randomness appears naturally, stochastic modeling is essential to understand a process and its dynamics.

Statistical inference consists of estimation and hypothesis testing. Some estimation problems remain challenging even today, in areas like high-dimensional or functional data, particularly when data are observed with noise and with missingness. In

high-frequency finance, estimation of even variances and covariances remain challenging due to microstructure noise and asynchronicity. A recent work in this area is [191]. It is necessary to estimate these quantities for risk management and optimization of investment portfolios.

Similarly, in the hypothesis testing situation, there is a need for stochastic modeling. A natural question to ask is whether there has been a shift in the returns and volatilities of stock prices due to covid lockdown. One has data on returns pre and post-lock-down. The stochastic modeling way of comparing these will be to assume that these returns are random samples from two populations, and subsequently to compare the population means or variances. Without such a structure, it is unclear how the two groups of numbers can be compared and how to conclude if the difference between them is large enough to indicate a shift or not. An interesting recent paper on hypothesis testing is [192]. Here we are interested in testing if there is a trend in the distribution of recurrence times and not predicting individual recurrences, so stochastic modeling is unavoidable.

Even in the regression set-up, in several areas, it remains important to estimate the parameters in the population and not predict individual observations. The interpretation of the regression slope β is essential in many applications. For example, how macroeconomic variables like GDP and interest rates affect the corporate probability of default. This is an important quantity for regulatory purposes as it measures how much risk a bank is taking overall. A machine learning algorithm will be useful in identifying which customers are more liable to default. But when the interest is in the overall probability of default, which is a population parameter, concepts of population and stochastic modeling naturally need to be taken into account.

Next, we focus on the area of estimation of a probability distribution. This could be density estimation, survival function estimation, or estimation of quantiles. Each of these has important practical applications. Quantile estimation is an important problem in financial risk management, see [193] for an interesting recent application. Stochastic analysis is inherent in these areas as the basic premise is built on the existence of a probability distribution. The whole topic of survival analysis has been developed to handle censored and truncated observations efficiently. Without stochastic modeling, such topics will remain completely out of reach.

15 | Ansgar Steland

The recent years have seen an unprecedented uprise of machine learning and artificial intelligence not even been foreseen by leading computer scientists like Geoffrey Hinton, who laid important foundations for the current approach behind large language models.

The question arises as to what role stochastic modeling and statistical analysis can play in an era of AI tools and AI agents. One can distinguish between specialized machine learning approaches competing with stochastic methods, for example, prediction using deep neural networks versus parametric or non-parametric regression models employing statistical methods, and AI systems, which automatically plan, design, and model a statistical problem, and then analyze real data on the fly, versus traditional computer-aided and AI-supported modeling and analysis done by a trained human analyst.

15.1 | The Role of Statistics, Probability, and Modeling in AI

In [194] a large author collective provided an extensive discussion of the history of machine learning and its interplay with statistics, of the notions of weak and strong artificial intelligence, and of the role of statistics. Although written in the pre-GPT3 era and thus almost exclusively dealing with the first question, the main findings and conclusions still hold. Essentially, it is argued that statistics has contributed a lot to the development of ML and AI and is still doing so. Indeed, recognizing the importance of formulating the problem of learning from data as the problem to infer the underlying data-generating process from a random sample or time series was and is fundamental for a deeper understanding of machine learning methods. In [194] it is further argued that statistics provides a more advanced and complete general framework for data analysis and is highly relevant for any AI development. AI applications often neglect issues such as data quality, evaluation of uncertainty, interpretability, causality, model stability, and reproducibility, or issues such as confounding. Although these issues are also discussed in the ML community, sometimes from a different perspective and using other notions, there is a well-established expertise in statistics about theoretical questions and practical solutions. This is so despite the widely accepted fact that there are various areas where machine learning methods are state of the art. This certainly applies to problems genuinely dealing with large-scale data such as image classification, which require enormous computational resources and are therefore almost exclusively researched by computer scientists. However, for other tasks where small to moderately large samples are sufficient, the situation can be different. For example, in an extensive study [195], compared 179 classifiers over 121 (non-large-scale) data sets from the UCI machine learning classification database. It was found that parallel random forest, a method mainly developed by the statistician Leo Breiman, see [81], performs best, followed by support vector machines, both outperforming extreme learning networks (i.e., networks with random weights of the hidden layers, thus optimizing only the output layer by ridge regression) and multi-layer perceptrons. However, when it comes to problems where no training data is available at the design phase or where relatively small amounts of data need to be analyzed, stochastic modeling and statistics still shine. Contrary, typically, ML methods are non-superior in such settings, as they are designed for large-scale data. There is also an increasing number of applications, such as embedded systems, small devices for the Internet of things, wearables, or implanted medical devices, which can only access very limited computational power. To provide them with AI capabilities that learn from the device's input obtained by its sensors, they either need to communicate with AI servers that carry the computational load, which requires reliable and fast wireless networking, or they need to rely on AI methods with extremely low computational costs. For the latter approach, one may employ randomized neural networks (such as extreme learning networks or echo-state neural networks [196]). Extreme learning networks do not train coefficients of hidden layers but

select them randomly. They are strongly related to regularized regression such as ridge regression, and therefore can be trained with extremely low computational costs. Echo-state networks are more complex and belong to the class of recursive networks, but using random weights for hidden layers they also substantially reduce the computational load for training. A third option is classical approaches using stochastic models and statistical estimation techniques. Here, one should avoid numerical algorithms to compute maximum likelihood estimators and instead rely on Le Cam's one-step estimation procedure [197], for instance, or use computationally less demanding methods such as moment estimators. Required distributions can be approximated by simulation and bootstrap approaches which are often appealing from a computational viewpoint, if applied to estimators given in closed form. To limit the training time, that is, estimating the model, all these approaches can be combined with computationally efficient statistical methods to assess uncertainty, such as fixed-length confidence intervals to determine required sample sizes, see the discussion in [198] and the references given there.

15.2 | Stochastic Modeling in ML/AI

Although various state-of-the-art AI/ML approaches do not rely on explicit stochastic modeling, this is not the rule. Indeed, many ML problems can greatly benefit from stochastic modeling expertise and require results from probability and statistics for their improvement. For example, in industrial quality control, when setting up inspection processes, manually selecting key characteristics requires substantial resources and effort. In [199], an auto-encoding neural network was used to learn such features from training data. In order to select a small number of features, including the case of grouped features, a tailor-made algorithm was developed with specific regularization terms. Correlated features are identified and evaluated by means of a risk analysis. Here, auto-encoders with one hidden layer are trained for each combination of two features as input. Then, the similarity of weights can be used to define how dependent the features are. The overall analysis includes pre-processing, design of net parameters, and specification of the dependence measure. It combines machine learning as well as human expertise from statistics and engineering, see [199] for details and references to related literature. A second insightful example is a recent improvement of active learning [200]. Active learning is a highly relevant learning problem in industry. It deals with the automatized sequential exploration of a sample space to identify the safe region where a system can optimally operate. It aims at reducing the number of labeled examples needed to achieve a certain accuracy by selecting the most informative safe examples from a large unlabeled dataset and determining their labels from a costly additional experiment. When it comes to dynamic systems, for example, when a robot explores an environment, exploration takes place along trajectories instead of single points, and then the whole trajectory needs to be safe. Gaussian processes (GPs) are an attractive approach to model the sequential data collection mechanism $x_1, x_2, \ldots$ as well as the prediction uncertainty at a new point x^*. It is natural to select x^* as a minimizer of the predictive variance under the constraint that x^* is unsafe with probability at most $\alpha \in (0, 1)$. The latter problem boils down to a tail event of the supremum of a GP. Instead of relying on brute-force Monte Carlo simulations of the whole GP, one can draw on the Borel-TIS inequality, which only requires estimating median and supremal variance. For these two crucial quantities, one may draw on sequential statistics to construct an adaptive Monte-Carlo procedure that outperforms pure Monte Carlo. It turns out that the combination of stochastic modeling and results from statistical theory and probability theory allows to produce a state of a art ML method. Best results are achieved, if machine learning, statistics, and probability go hand in hand.

For applications in business and industry, it seems to be crucial to have access to models, methods, and tools from stochastic modeling and statistical inference as well as access to ML/AI methods, in order to be able to experiment with all of them and select the approach which works best for the problem at hand. Personal communication with an AI startup developing quality control software addressing classical questions studied in Statistical Quality Control and Statistical Process Control revealed that under the hood a whole bunch of ML methods is compared and the best performing one is used to generate signals. ML methods are usually defined in terms of an ultra-high-dimensional parametric model, which is fitted to the training data by explicit (drop-out, penalties) or implicit (early stopping) regularization. Since they scale well with large numbers of input variables and massive training data, they often have an edge compared to stochastic modeling for large data sets. However, as discussed above, this requires good data quality and often careful preprocessing of data. Data quality and clarifying what the data is representative of all is crucial to relating findings and conclusions drawn from the data to the real world. Manually defining the right input variables and the right data transformations is more important than one might think even for ML methods. Indeed, even the best training algorithms are far from being able to learn this automatically in a satisfactory manner. This becomes clear from the fact that neural networks with convolutional layers for image classification are formally special cases of fully-connected feedforward neural networks, but it is pointless to hope that a learning algorithm will output a fitted network with input layers coming close to convolutional layers. But it also applies to more or less low-dimensional problems such as regression problems with a handful of regressors. Although (shallow or deep) neural networks implicitly compute non-linear transformations of the input variables, practical experience shows that often the performance can be substantially improved, when feeding the net with the specific transformations (such as logs, polynomials, differences, or percentages instead of levels) suggested by domain knowledge and/or previous analyses, instead of trying to learn them implicitly.

15.3 | Llns and AI Agents

Concerning the second question posed at the beginning, the subsequent discussion will ignore the fundamental question of whether or not AI systems can exhibit human-like intelligence and are on par with the human mind, or whether AGI and superintelligence are possible. These questions are beyond the scope of this contribution. Further, any answer needs to predict to some extent the future development of AI, and thus may fail. But let us try to make an educated guess. The rapid progress of LLNs and their fine-tuning to specific domains and tasks allows to set up of interacting AI agents to generate, check, and curate output. It

seems that software development is the field where this approach is most advanced and already provides convincing results at the time of writing. Here one agent outputs source code based on a user prompt, and a further specialized agent interacting with the user performs code checking and outputs directives for source code revisions, in order to eliminate errors and improve the program. It is clear that the concept of several AI agents, that interact with each other and with the user to solve certain problems, will become widespread and has a substantial potential for better and less error-prone AI. LLNs are trained from massive internet data which contains huge amounts of source code in many programming languages. This is certainly the reason for their capabilities in generating computer programs. Probability, theoretical statistics, and large parts of applied statistics and stochastic modeling are exact sciences and follow relatively simple grammar. At least, they can be put into a formal language following simple grammar. Therefore, one can expect that future AI systems have capabilities in these areas that are comparable to their skills in software development, thus going beyond the already remarkable functionality of Open AI's Code Interpreter. It is likely that future AI systems substantially simplify the development and application of stochastic models and statistical analyses and provide access to a large amount of knowledge. This might have devastating effects on the job market for graduates, since then only a few experts are needed to supervise the AI. However, that development will be probably relatively slow. Firstly, compared to software, there is less data available for the training of a statistics agent. Second, whereas software is written in a formal language, this does not strictly apply to scientific papers and books that form the training corpus. Third, although not completely transparent, state-of-the-art AI systems use lots of manual input from human experts. When it comes to a domain such as statistics, this needs well-trained experts which are not available on a large scale, and it is questionable whether providers will invest here. Thus, as long as human input is needed, the capabilities in such fields will substantially lag behind areas such as programming.[6] Perhaps, AI systems capable of producing valid stochastic models, methods for their analysis, and derivations of their properties will result as a side product of efforts to create AI systems that can solve scientific research questions addressing hot topics such as finding cancer treatments, understanding the human brain or finding the grand unified theory of physics.

Nevertheless, it is an open question, whether future AI systems will be smart enough to produce valid novel scientific results going beyond correct and nice sounding science prosa, and whether they will be superior to humans in identifying and resolving challenging problems arising in modeling and analyzing real-world data. Examples of the latter are causality (vs. association), confounders and colliders, bias, and multiple testing. For example, in order to identify causality, controlled experiments, and randomization are the methods of choice, and available data and data obtained from observational studies often suffer from confounders (variables affecting response and regressor) and colliders (variables affected by response and regressor), which lead to distorted estimates of causal effects. Generally, blind analysis of available data collections may lead to severe biases which quickly result in discrimination and harm to people. Last but not least, extensive testing requires appropriate multiple-testing corrections to avoid irreproducible large-scale false-positive results and pseudo-significant patterns in data. Ubiquitous issues are Simpson's paradox, the fact that a given pattern appears in a random sequence with probability one, or the problem that important notions such as fairness in automated decision systems cannot be uniquely defined and lead to different, inconsistent decisions. For example, should one look at the true and false positive rates or at the positive and negative predictive values? And should one balance the former or the latter between groups? Or should one equalize the odds? Questions such as these do not disappear by using an ML method regarded as state of the art in terms of a criterion such as predictive power. In practice, one needs to consider these issues and decide what to do and one has to take responsibility for it. Statistics still provides a more complete and accessible framework, and combining statistical thinking and expertise with machine learning might be the way to go. And if future AI systems produce human-like solutions and make choices on their own, it is important that their generated solutions and choices are transparent and verifiable by humans, and that humans remain control over the AI system.

Stochastic modeling, and more broadly statistical modeling, will continue to play an important role in various aspects of problem-solving. Simply put, problem-solving is a process rather than a single step, involving multiple components structured in a specific way. In the era of big data, the contribution of machine learning techniques has become increasingly significant, especially when dealing with complex systems. However, statistical modeling remains a powerful tool for problem formulation and solution development, and machine learning can yield more insightful outcomes when it is integrated into the problem-solving process alongside other critical components such as stochastic modeling.

One such application is the inverse design problem, which aims to identify input values that produce desired outputs. Due to the many-to-one relationships between inputs and outputs that are commonly embedded in the data, the inverse design problem has long been a challenge in natural science and engineering areas. Recently, several deep learning-based methods [201] have been developed which tackle the problem in a more direct way. One of these methods is a conditional invertible neural network (cINN) [202, 203]. Essentially, the effectiveness of cINN benefits from both mathematical and statistical principles. First, cINN is a generative probabilistic model that stochastically generates posterior samples of inputs given specified outputs. Second, the network architecture of cINN is carefully designed to enable the computation of Jacobian determinants. Furthermore, it is critical to effectively select a subset of posterior samples of inputs for the follow-up validation study, which may incorporate user constraints and potentially involve an optimization process.

There are certainly applications where stochastic and statistical modeling is preferred and more suitable. In contrast to deterministic modeling, which many machine learning methods employ, a key advantage of statistical modeling is its ability to quantify uncertainty. This quantification is crucial for enhancing model reliability and supporting decision-making in real-world applications. In addition, statistical models generally offer better explainability compared to black-box machine learning methods.

A thorough understanding of model behavior enhances transparency and builds trust in high-risk domains such as healthcare and finance. Often, sample size is another important factor in determining the selection of appropriate modeling approaches. It is well known that machine learning, particularly deep learning methods, requires a large amount of data to achieve satisfactory performance. When physical simulation is computationally slow and/or experimental data collection is costly, practitioners will need to explore alternative approaches. For example, Bayesian networks [204] can be an effective method for addressing inverse design problems and time-dependent dynamic issues.

It is worth noting that both the statistical modeling and machine learning communities are rapidly evolving. Therefore, a good practice for problem-solving is to leverage multiple approaches and assess their performance and usability. Integrating insights from multiple approaches addresses the problem from different perspectives, thereby guiding users toward a clear path for continuous improvement.

17 | Final Discussion

Actually, these are not conclusions since the hope of the authors is to stimulate a debate with sound contributions using many possible channels, especially those offered by the International Society for Business and Industrial Statistics (ISBIS), of which Applied Stochastic Models in Business and Industry (ASMBI) is the official journal. The former Editor-in-Chief of ASMBI would like to thank all those who contributed to this collective work.

Affiliations

[1]Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, Milano, Italy | [2]Department of Statistical Science, Duke University, Durham, North Carolina, USA | [3]Department of Statistics, Purdue University, West Lafayette, Indiana, USA | [4]School of Mathematics and Statistics, University of Sydney & ValueMetrics, Sydney, New South Wales, Australia | [5]Department of Statistical and Actuarial Sciences, Western University, London, Ontario, Canada | [6]Dipartimento di Scienze Economiche e Aziendali, Università di Pavia, Pavia, Italy | [7]Department of Mathematics, Union College, Schenectady, New York, USA | [8]KPA Ltd & Samuel Neaman Institute, Haifa, Israel | [9]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, Hong Kong | [10]Laboratoire de Mathématiques d'Orsay, University Paris-Saclay, Orsay, France | [11]University of Coimbra, CERES, Department of Chemical Engineering, Coimbra, Portugal | [12]Cédric — Conservatoire National des Arts et Métiers, Paris, France | [13]MOX — Dipartimento di Matematica, Politecnico di Milano, Milano, Italy | [14]Indian Statistical Institute, Bangalore, India | [15]Institute of Statistics, RWTH Aachen University, Aachen, Germany | [16]GE Vernova Advanced Research Center, Niskayuna, New York, USA

Acknowledgments

The contribution of Paolo Giudici and Emanuela Raffinetti was funded by the European Union — NextGenerationEU, in the framework of GRINS — Growing Resilient, Inclusive and Sustainable (GRINS PE00000018). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Marco Reis acknowledges support from Fundação para a Ciência e Tecnologia, I.P., through the projects with DOI:10.54499/UIDB/00102/2020 and DOI:10.54499/UIDP/00102/2020. Open access publishing facilitated by Consiglio Nazionale delle Ricerche, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Endnotes

[1] We interpret Wisdom as skill in applying Knowledge in the presence of uncertainty.

[2] A very successful traffic modeling has facets: A Multifractal Fractional Sum-Difference model (MFSD) is a monotone transformation of a Gaussian Fractional Sum-Difference model GFSD. The GFSD is the sum of two independent components: A moving sum of length 2 of discrete fractional Gaussian noise (fGn), and white noise. Internet traffic interarrival times are very well modeled by an MFSD in which the marginal distribution is Weibull.

[3] For a blog and report on this see https://errorstatistics.com/2024/07/11/guest-post-ron-kenett-whats-happening-in-statistical-practice-since-the-abandon-statistical-significance-call-5-years-ago/ and https://www.neaman.org.il/en/a-tripartite-view-on-the-role-of-ai-in-modern-analytics.

[4] L'obiettivo della scienza non è fare predizioni. E' anche offrire un'immagine della realtà, un quadro concettuale per pensare le cose. Questa ambizione ha reso efficace il pensiero scientifico. Se l'obiettivo della scienza fossero solo le predizioni, Copernico non avrebbe scoperto nulla rispetto a Tolomeo: Le sue previsioni astronomiche non erano migliori di quelle di Tolomeo. Ma Copernico ha trovato una chiave per ripensare tutto e comprendere meglio.

[5] Available at https://statistics.mox.polimi.it/aneurisk/.

[6] Although this could be a historical remark at the time of reading, at the time of writing, Microsoft's Bing copilot does not even correctly reproduce the definition of a probability measure.

References

1. R. B. Gramacy, *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences* (Chapman and Hall/CRC, 2020).

2. R. B. Gramacy, H. K. H. Lee, and W. G. Macready, "Parameter Space Exploration With Gaussian Process Trees. Proceedings of the 21st International Conference on Machine Learning," (2004), 45.

3. D. G. Krige and D. G. Krige, *Lognormal-de Wijsian Geostatistics for Ore Evaluation* (South African Institute of Mining and Metallurgy, 1981).

4. D. L. Banks, G. Datta, A. Karr, J. Lynch, J. Niemi, and F. Vera, "Bayesian CAR Models for Syndromic Surveillance on Multiple Data Streams: Theory and Practice," *Information Fusion* 13, no. 2 (2012): 105–116.

5. X. Chen, K. Irie, D. L. Banks, R. Haslinger, J. Thomas, and M. West, "Scalable Bayesian Modeling, Monitoring, and Analysis of Dynamic Network Flow Data," *Journal of the American Statistical Association* 113, no. 522 (2018): 519–533.

6. K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans, "A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation," *Human Brain Mapping* 4, no. 1 (1996): 58–73.

7. S. Sahu, *Bayesian Modeling of Spatio-Temporal Data With R* (Chapman and Hall/CRC, 2022).

8. N. Cressie and C. K. Wikle, *Statistics for Spatio-Temporal Data* (Wiley, 2011).

9. A. G. Hawkes, "Hawkes Processes and Their Applications to Finance: A Review," *Quantitative Finance* 18, no. 2 (2018): 193–198.

10. C. Blundell, J. Beck, and K. A. Heller, "Modelling Reciprocating Relationships With Hawkes Processes," *Advances in Neural Information Processing Systems* 25 (2012).

11. T. R. Henry, D. L. Banks, D. Owens-Oas, and C. Chai, "Modeling Community Structure and Topics in Dynamic Text Networks," *Journal of Classification* 36 (2019): 322–349.

12. M. J. Lighthill and G. B. Whitham, "On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads," *Proc R Soc. Lond. Ser. A* 229, no. 1178 (1955): 317–345.

13. M. R. Flynn, A. R. Kasimov, J. C. Nave, R. R. Rosales, and B. Seibold, "Self-Sustained Nonlinear Waves in Traffic Flow," *Physical Review E* 79, no. 5 (2009): 056113.

14. P. Zilber and B. Nadler, "Imbalanced Mixed Linear Regression," *Advances in Neural Information Processing Systems* 36 (2023): 54540–54553.

15. D. L. Banks, L. House, and K. Killourhy, "Cherry-Picking for Complex Data: Robust Structure Discovery," *Philosophical Transactions of the Royal Society A* 2009, no. 367 (1906): 4339–4359.

16. CMS Collaboration, *CMS Releases Higgs Boson Discovery Data to the Public* (CERN, 2024), https://home.web.cern.ch/news/news/experiments/cms-releases-higgs-boson-discovery-data-public.

17. Y. Yang and P. Zhai, "Click-Through Rate Prediction in Online Advertising: A Literature Review," *Information Processing & Management* 59, no. 2 (2022): 102853.

18. P. M. LeBlanc, D. L. Banks, L. Fu, M. Li, Z. Tang, and Q. Wu, "Recommender Systems: A Review," *Journal of the American Statistical Association* 119, no. 545 (2024): 773–785.

19. C. Waisman, H. S. Nair, and C. Carrion, "Online Causal Inference for Advertising in Real-Time Bidding Auctions. To Appear in *Mark Sci*," (2024).

20. N. T. Stevens and L. Hagar, "Comparative Probability Metrics: Using Posterior Probabilities to Account for Practical Equivalence in A/B Tests," *American Statistician* 76, no. 3 (2022): 224–237.

21. P. Whittle, *Optimization Over Time* (Wiley, 1982).

22. D. L. Banks and M. B. Hooten, "Statistical Challenges in Agent-Based Modeling," *American Statistician* 75, no. 3 (2021): 235–242.

23. J. Li, E. Rombaut, and L. Vanhaverbeke, "A Systematic Review of Agent-Based Models for Autonomous Vehicles in Urban Mobility and Logistics: Possibilities for Integrated Simulation Models," *Computers, Environment and Urban Systems* 89 (2021): 101686.

24. J. Dubbelboer, I. Nikolic, K. Jenkins, and J. Hall, "An Agent-Based Model of Flood Risk and Insurance," *JASSS-Journal of Artificial Societies and Social Simulation* 20, no. 1 (2017): 6.

25. M. Rolón and E. Martínez, "Agent-Based Modeling and Simulation of an Autonomic Manufacturing Execution System," *Computers in Industry* 63, no. 1 (2012): 53–78.

26. D. Z. Zhang, A. I. Anosike, M. K. Lim, and O. M. Akanle, "An Agent-Based Approach for e-Manufacturing and Supply Chain Integration," *Computers and Industrial Engineering* 51, no. 2 (2006): 343–360.

27. M. Kieu, H. Nguyen, J. A. Ward, and N. Malleson, "Towards Real-Time Predictions Using Emulators of Agent-Based Models," *Journal of Simulation* 18, no. 1 (2024): 29–46.

28. D. Anderson, W. S. Cleveland, and B. Xi, "Multifractal and Gaussian Fractional Sum–Difference Models for Internet Traffic," *Performance Evaluation* 107 (2017): 1–33.

29. N. I. Fisher, P. D. Lunn, and S. M. Sasse, "Enhancing Value by Continuously Improving Enterprise Culture," *Journal of Creating Value* 7, no. 2 (2021): 232–254.

30. R. C. Merton, "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics* 3, no. 1–2 (1976): 125–144.

31. R. F. Engle, "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50, no. 4 (1982): 987–1007.

32. J. C. Cox and S. A. Ross, "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics* 3, no. 1–2 (1976): 145–166.

33. S. L. Heston, "A Closed-Form Solution for Options With Stochastic Volatility With Applications to Bond and Currency Options," *Review of Financial Studies* 6, no. 2 (1993): 327–343.

34. M. Grasselli, "The 4/2 Stochastic Volatility Model: A Unified Approach for the Heston and the 3/2 Model," *Mathematical Finance* 27, no. 4 (2017): 1013–1034.

35. J. Gatheral, T. Jaisson, and M. Rosenbaum, "Volatility Is Rough," *Quantitative Finance* 18, no. 6 (2018): 933–949.

36. M. Escobar-Anel and W. L. Fan, "A New Type of CEV Model: Properties, Comparison, and Application to Portfolio Optimization," *Stoch Models* 41, no. 1 (2025): 85–119.

37. M. Escobar-Anel, L. Stentoft, and X. Ye, "Setting the VIX Free: A Generalized Affine GARCH Model. SSRN 4664927," 2023.

38. European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* (European Commission, 2021).

39. ISO-IEC 23894: 2023 AI Risk Management, "International Organisation for Standardisation and International Electrotechnical Commission," 2023.

40. *AI Risk Management Framework* (National Institute for Standards and Technology, 2023).

41. "Defining AI Incidents and Related Terms. Organisation for Economic Cooperation and Development (OECD) Artificial Intelligence Papers No.16, May," 2024.

42. G. Babaei, P. Giudici, and E. Raffinetti, "A Rank Graduation Box for SAFE Artificial Intelligence," *Expert Systems with Applications* 259 (2025): 125239.

43. P. Giudici and E. Raffinetti, "SAFE Artificial Intelligence in Finance," *Finance Research Letters* 56 (2023): 104088.

44. M. O. Lorenz, "Methods of Measuring the Concentration of Wealth," *Publications of the American Statistical Association* 9, no. 70 (1905): 209–219.

45. C. Gini, "Measurement of Inequality of Incomes," *Econometrics Journal* 31, no. 121 (1921): 124–125.

46. W. George, "Why Boeing's Problems With the 737 Max Began More Thank 25 Years Ago," *Harvard Business School Working Knowledge* (2024), https://www.library.hbs.edu/working-knowledge/why-boeings-problems-with-737-max-began-more-than-25-years-ago.

47. M. Li, *Personal Communication* (2023).

48. J. A. Cazier and D. T. Rawls, *Succeeding in Analytics by Leading in Analytics* (Informs, 2023), https://doi.org/10.1287/LYTX.2024.01.01/full/.

49. T. C. Redman and R. W. Hoerl, "AI and Statistics: Perfect Together," *MIT Sloan Management Review* (2024), https://sloanreview.mit.edu/article/ai-and-statistics-perfect-together/.

50. W. Kuo, *Soulware: The American Way in China's Higher Education* (Wiley, 2019).

51. O. Solon, "How a Book About Flies Came to Be Priced $24 Million on Amazon," *Wired* (2011), https://www.wired.com/2011/04/amazon-flies-24-million/.

52. N. Singer, "Amazon's Facial Recognition Wrongly Identifies 28 Lawmakers A.C.L.U. Says. New York Times," 2018, https://www.nytimes.

com/2018/07/26/technology/amazon-aclu-facial-recognition-congress.html.

53. R. W. Hoerl and T. C. Redman, "What Managers Should Ask About Ai Models and Data Sets," *MIT Sloan Management Review* 65, no. 2 (2024): 1–5.

54. R. A. Fisher, *The Design of Experiments* (Oliver and Boyd, 1935).

55. R. W. Hoerl and R. D. Snee, *Statistical Thinking: Improving Business Performance*, 3rd ed. (Wiley, 2020).

56. D. J. Hand, *Dark Data: Why What You Don't Know Matters* (Princeton University Press, 2020).

57. D. K. J. Lin, *Ghost Data*. Unpublished manuscript (, 2017).

58. B. Efron and T. Hastie, *Computer Age Statistical Inference: Algorithms, Evidence and Data Science* (Cambridge University Press, 2016).

59. R. S. Kenett and B. G. Francq, "Helping Reviewers Assess Statistical Analysis: A Case Study From Analytic Methods," *Analytical Science Advances* 3, no. 5–6 (2022): 212–222.

60. R. S. Kenett, C. Gotwalt, L. Freeman, and X. Deng, "Self-Supervised Cross Validation Using Data Generation Structure," *Applied Stochastic Models in Business and Industry* 38, no. 5 (2022): 750–765.

61. N. Ravishanker, V. Serhiyenko, and M. R. Willig, "Hierarchical Dynamic Models for Multivariate Times Series of Counts," *Stat Its Interface* 7, no. 4 (2014): 559–570.

62. M. S. Reis, "Multiscale and Multi-Granularity Process Analytics: A Review," *Processes* 7, no. 2 (2019): 61.

63. R. S. Kenett and J. Bortman, "The Digital Twin in Industry 4.0: A Wide-Angle Perspective," *Quality and Reliability Engineering International* 38, no. 3 (2022): 1357–1366.

64. R. S. Kenett, S. Zacks, and P. Gedeck, *Industrial Statistics: A Computer-Based Approach With Python* (Birkhäuser, 2023).

65. R. S. Kenett and S. Zacks, *Modern Industrial Statistics: With Applications in R, MINITAB, and JMP* (Wiley, 2021).

66. R. S. Kenett, S. Zacks, and P. Gedeck, *Modern Statistics: A Computer-Based Approach With Python* (Birkhäuser, 2022).

67. G. Davidyan, J. Bortman, and R. S. Kenett, "Development of an Operational Digital Twin of a Locomotive Parking Brake for Fault Diagnosis," *Scientific Reports* 13, no. 1 (2023): 17959.

68. G. Davidyan, J. Bortman, and R. S. Kenett, "Development of an Operational Digital Twin of a Freight car Braking System for Fault Diagnosis," *Advanced Theory and Simulations* 7, no. 6 (2024): 2301257.

69. R. S. Kenett, "Engineering, Emulators, Digital Twins, and Performance Engineering," *Electronics* 13, no. 10 (2024): 1829.

70. D. Bzdok, N. Altman, and M. Krzywinski, "Statistics Versus Machine Learning," *Nature Methods* 15, no. 4 (2018): 233–234.

71. H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, and V. Fanos, "Comparison of Conventional Statistical Methods With Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment," *Medicina* 56, no. 9 (2020): 455.

72. C. S. Wong and W. K. Li, "On a Mixture Autoregressive Model," *Journal of the Royal Statistical Society B* 62, no. 1 (2000): 95–115.

73. G. E. P. Box, G. M. Jenkins, G. C. Reinsell, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. (Wiley, 2016).

74. J. Aitchison and I. R. Dunsmore, *Statistical Prediction Analysis* (Cambridge University Press, 1975).

75. G. A. Papacharalampous, H. Tyralis, and D. Koutsoyiannis, "Comparison of Stochastic and Machine Learning Methods for Multi-Step Ahead Forecasting of Hydrological Processes," *Stochastic Environmental Research and Risk Assessment* 33, no. 2 (2019): 481–514.

76. R. T. Clemen, "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting* 5, no. 4 (1989): 559–583.

77. N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games* (Cambridge University Press, 2006).

78. B. Auder, M. Bobbia, J. M. Poggi, and B. Portier, "Sequential Aggregation of Heterogeneous Experts for PM10 Forecasting," *Atmospheric Pollution Research* 7, no. 6 (2016): 1101–1109.

79. J. Cugliari, Y. Goude, and J. M. Poggi, "Disaggregated Electricity Forecasting Using Wavelet-Based Clustering of Individual Consumers. Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON). New York: IEEE; 2016:1–6,".

80. B. Goehry, Y. Goude, P. Massart, and J. M. Poggi, "Aggregation of Multi-Scale Experts for Bottom-Up Load Forecasting," *IEEE Transactions on Smart Grid* 11, no. 3 (2019): 1895–1904.

81. L. Breiman, "Random Forests," *Machine Learning* 45 (2001): 5–32.

82. E. Devijver, Y. Goude, and J. M. Poggi, "Clustering Electricity Consumers Using High-Dimensional Regression Mixture Models," *Applied Stochastic Models in Business and Industry* 36, no. 1 (2020): 159–177.

83. E. Devijver, "Finite Mixture Regression: A Sparse Variable Selection by Model Selection for Clustering," *Electronic Journal of Statistics* 9, no. 2 (2015): 2642–2674.

84. M. S. Reis and R. S. Kenett, "Assessing the Value of Information of Data-Centric Activities in the Chemical Processing Industry 4.0," *AICHE Journal* 64, no. 11 (2018): 3868–3881.

85. R. S. Kenett and G. Shmueli, "On Information Quality," *J Roy Statist Soc Ser A* 177, no. 1 (2014): 3–38.

86. J. M. Juran, *Juran on Quality by Design* (Free Press, 1992).

87. H. B. Grangeia, C. Silva, S. P. Simões, and M. S. Reis, "Quality by Design in Pharmaceutical Manufacturing: A Systematic Review of Current Status, Challenges and Future Perspectives," *European Journal of Pharmaceutics and Biopharmaceutics* 147 (2020): 19–37.

88. C. W. Coley, D. A. Thomas, J. A. M. Lummiss, et al., "A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning," *Science* 365, no. 6453 (2019): eaax1566.

89. F. Grisoni, "Chemical Language Models for de Novo Drug Design: Challenges and Opportunities," *Current Opinion in Structural Biology* 79 (2023): 102527.

90. B. M. A. Silva, S. Vicente, S. Cunha, et al., "Retrospective Quality by Design (rQbD) Applied to the Optimization of Orodispersible Films," *International Journal of Pharmaceutics* 528, no. 1–2 (2017): 655–663.

91. A. C. Pereira, M. S. Reis, J. M. Leça, P. M. Rodrigues, and J. C. Marques, "Definitive Screening Designs and Latent Variable Modelling for the Optimization of Solid Phase Microextraction (SPME): Case Study - Quantification of Volatile Fatty Acids in Wines," *Chemometrics and Intelligent Laboratory Systems* 179 (2018): 73–81.

92. M. S. Reis, A. C. Pereira, J. M. Leça, P. M. Rodrigues, and J. C. Marques, "Multiresponse and Multiobjective Latent Variable Optimization of Modern Analytical Instrumentation for the Quantification of Chemically Related Families of Compounds: Case Study-Solid-Phase Microextraction (SPME) Applied to the Quantification of Analytes With Impact on Wine Aroma," *Journal of Chemometrics* 33, no. 3 (2018): e3103.

93. M. S. Reis and P. M. Saraiva, "Prediction of Profiles in the Process Industries," *Industrial and Engineering Chemistry Research* 51, no. 11 (2012): 4254–4266.

94. E. Del Castillo and M. S. Reis, "Bayesian Predictive Optimization of Multiple and Profile Response Systems in the Process Industry: A Review and Extensions," *Chemometrics and Intelligent Laboratory Systems* 206 (2020): 104121.

95. G. Miró-Quesada, E. Del Castillo, and J. J. Peterson, "A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables," *Journal of Applied Statistics* 31, no. 3 (2004): 251–270.

96. K. Eggensperger, M. Feurer, F. Hutter, et al., "Towards an Empirical Foundation for Assessing Bayesian Optimization of Hyperparameters," *NIPS Workshop on Bayesian Optimization in Theory and Practice* 10, no. 3 (2013): 1–5.

97. J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," *Advances in Neural Information Processing Systems* (2012): 2951–2959.

98. V. Mnih, K. Kavukcuoglu, D. Silver, et al., "Human-Level Control Through Deep Reinforcement Learning," *Nature* 518, no. 7540 (2015): 529–533.

99. R. Nian, J. Liu, and B. Huang, "A Review on Reinforcement Learning: Introduction and Applications in Industrial Process Control," *Computers and Chemical Engineering* 139 (2020): 106886.

100. T. J. Rato, P. Delgado, C. Martins, and M. S. Reis, "First Principles Statistical Process Monitoring of High-Dimensional Industrial Microelectronics Assembly Processes," *PRO* 8, no. 11 (2020): 1520.

101. M. S. Reis and P. Delgado, "A Large-Scale Statistical Process Control Approach for the Monitoring of Electronic Devices Assemblage," *Computers and Chemical Engineering* 39 (2012): 163–169.

102. M. S. Reis, R. Rendall, T. J. Rato, C. Martins, and P. Delgado, "Improving the Sensitivity of Statistical Process Monitoring of Manifolds Embedded in High-Dimensional Spaces: The Truncated-Q Statistic," *Chemometrics and Intelligent Laboratory Systems* 215 (2021): 104369.

103. M. S. Reis, G. Gins, and T. J. Rato, "Incorporation of Process-Specific Structure in Statistical Process Monitoring: A Review," *Journal of Quality Technology* 51, no. 4 (2019): 407–421.

104. L. H. Chiang and R. D. Braatz, "Process Monitoring Using Causal Map and Multivariate Statistics: Fault Detection and Identification," *Chemometrics and Intelligent Laboratory Systems* 65, no. 2 (2003): 159–178.

105. R. Paredes, T. J. Rato, and M. S. Reis, "Causal Network Inference and Functional Decomposition for Decentralized Statistical Process Monitoring: Detection and Diagnosis," *Chemical Engineering Science* 267 (2023): 118338.

106. W. T. Yang, J. Blue, A. Roussy, J. Pinaton, and M. S. Reis, "A Physics-Informed Run-To-Run Control Framework for Semiconductor Manufacturing," *Expert Systems with Applications* 155 (2020): 113424.

107. T. J. Rato and M. S. Reis, "Markovian and Non-Markovian Sensitivity Enhancing Transformations for Process Monitoring," *Chemical Engineering Science* 163 (2017): 223–233.

108. T. J. Rato and M. S. Reis, "Sensitivity Enhancing Transformations for Monitoring the Process Correlation Structure," *Journal of Process Control* 24, no. 6 (2014): 905–915.

109. F. A. A. Souza, R. Araújo, and J. Mendes, "Review of Soft Sensor Methods for Regression Applications," *Chemometrics and Intelligent Laboratory Systems* 152 (2016): 69–79.

110. B. Lin, B. Recke, T. M. Schmidt, J. K. H. Knudsen, and S. B. Jørgensen, "Data-Driven Soft Sensor Design With Multiple-Rate Sampled Data: A Comparative Study," *Industrial and Engineering Chemistry Research* 48, no. 11 (2009): 5379–5387.

111. M. S. Reis, R. Rendall, S. T. Chin, and L. Chiang, "Challenges in the Specification and Integration of Measurement Uncertainty in the Development of Data-Driven Models for the Chemical Processing Industry," *Industrial and Engineering Chemistry Research* 54, no. 37 (2015): 9159–9177.

112. C. Shang, X. Huang, J. A. K. Suykens, and D. Huang, "Enhancing Dynamic Soft Sensors Based on DPLS: A Temporal Smoothness Regularization Approach," *Journal of Process Control* 28 (2015): 17–26.

113. P. Kadlec, B. Gabrys, and S. Strandt, "Data-Driven Soft Sensors in the Process Industry," *Computers and Chemical Engineering* 33, no. 4 (2009): 795–814.

114. P. Kadlec, R. Grbić, and B. Gabrys, "Review of Adaptation Mechanisms for Data-Driven Soft Sensors," *Computers and Chemical Engineering* 35, no. 1 (2011): 1–24.

115. K. Fujiwara, M. Kano, S. Hasebe, and A. Takinami, "Soft-Sensor Development Using Correlation-Based Just-In-Time Modeling," *AICHE Journal* 55, no. 7 (2009): 1754–1765.

116. T. Dias, R. Oliveira, P. M. Saraiva, and M. S. Reis, "Predictive Analytics in the Petrochemical Industry: Research Octane Number (RON) Forecasting and Analysis in an Industrial Catalytic Reforming Unit," *Computers and Chemical Engineering* 139 (2020): 106912.

117. T. J. Rato and M. S. Reis, "Multiresolution Soft Sensors (MR-SS): A New Class of Model Structures for Handling Multiresolution Data," *Industrial and Engineering Chemistry Research* 56, no. 13 (2017): 3640–3654.

118. T. J. Rato and M. S. Reis, "Building Optimal Multiresolution Soft Sensors for Continuous Processes," *Industrial and Engineering Chemistry Research* 57, no. 30 (2018): 9750–9765.

119. J. Flores-Cerrillo and J. F. MacGregor, "Control of Batch Product Quality by Trajectory Manipulation Using Latent Variable Models," *Journal of Process Control* 14, no. 5 (2004): 539–553.

120. J. M. González-Martínez, A. Ferrer, and J. A. Westerhuis, "Real-Time Synchronization of Batch Trajectories for On-Line Multivariate Statistical Process Control Using Dynamic Time Warping," *Chemometrics and Intelligent Laboratory Systems* 105, no. 2 (2011): 195–206.

121. P. Nomikos and J. F. MacGregor, "Multi-Way Partial Least Squares in Monitoring Batch Processes," *Chemometrics and Intelligent Laboratory Systems* 30, no. 1 (1995): 97–108.

122. T. J. Rato, J. Blue, J. Pinaton, and M. S. Reis, "Translation-Invariant Multiscale Energy-Based PCA for Monitoring Batch Processes in Semiconductor Manufacturing," *IEEE Transactions on Automation Science and Engineering* 14, no. 2 (2017): 894–904.

123. T. J. Rato and M. S. Reis, "Optimal Selection of Time Resolution for Batch Data Analysis. Part I: Predictive Modeling," *AIChE Journal* 64, no. 11 (2018): 3923–3933.

124. R. Rendall, L. H. Chiang, and M. S. Reis, "Data-Driven Methods for Batch Data Analysis – A Critical Overview and Mapping on the Complexity Scale," *Computers and Chemical Engineering* 124 (2019): 1–13.

125. M. Kano, K. Miyazaki, S. Hasebe, and I. Hashimoto, "Inferential Control System of Distillation Compositions Using Dynamic Partial Least Squares Regression," *Journal of Process Control* 10, no. 2–3 (2000): 157–166.

126. T. Dias, R. Oliveira, P. M. Saraiva, and M. S. Reis, "Linear and Non-Linear Soft Sensors for Predicting the Research Octane Number (RON) Through Integrated Synchronization, Resolution Selection and Modelling," *Sensors* 22, no. 10 (2022): 3734.

127. P. Facco, F. Doplicher, F. Bezzo, and M. Barolo, "Moving Average PLS Soft Sensor for Online Product Quality Estimation in an Industrial Batch Polymerization Process," *Journal of Process Control* 19, no. 3 (2009): 520–529.

128. D. Stanišić, N. Jorgovanović, N. Popov, and V. Čongradac, "Soft Sensor for Real-Time Cement Fineness Estimation," *ISA Transactions* 55 (2015): 250–259.

129. M. Shakil, M. Elshafei, M. A. Habib, and F. A. Maleki, "Soft Sensor for NO_x and O_2 Using Dynamic Neural Networks," *Computers and Electrical Engineering* 35, no. 4 (2009): 578–586.

130. A. Lepore, M. S. Reis, B. Palumbo, R. Rendall, and C. Capezza, "A Comparison of Advanced Regression Techniques for Predicting Ship CO_2

Emissions," *Quality and Reliability Engineering International* 33, no. 6 (2017): 1281–1292.

131. M. S. Reis and P. M. Saraiva, "Integration of Data Uncertainty in Linear Regression and Process Optimization," *AICHE Journal* 51, no. 11 (2005): 3007–3019.

132. R. Rendall and M. S. Reis, "Which Regression Method to Use? Making Informed Decisions in Data-Rich/Knowledge Poor Scenarios – The Predictive Analytics Comparison Framework (PAC)," *Chemometrics and Intelligent Laboratory Systems* 181 (2018): 52–63.

133. W. Sun and R. D. Braatz, "Smart Process Analytics for Predictive Modeling," *Computers and Chemical Engineering* 144 (2021): 107134.

134. L. E. Frank and J. H. Friedman, "A Statistical View of Some Chemometrics Regression Tools," *Technometrics* 35, no. 2 (1993): 109–135.

135. M. S. Reis, "Network-Induced Supervised Learning: Network-Induced Classification (NI-C) and Network-Induced Regression (NI-R)," *AICHE Journal* 59, no. 5 (2013): 1570–1587.

136. R. Rendall, I. Castillo, B. Lu, et al., "Image-Based Manufacturing Analytics: Improving the Accuracy of an Industrial Pellet Classification System Using Deep Neural Networks," *Chemometrics and Intelligent Laboratory Systems* 180 (2018): 26–35.

137. E. Cagno, F. Caron, M. Mancini, and F. Ruggeri, "Using AHP in Determining the Prior Distributions on Gas Pipeline Failures in a Robust Bayesian Approach," *Reliability Engineering and System Safety* 67 (2000): 275–284.

138. A. Pievatolo and F. Ruggeri, "Bayesian Reliability Analysis of Complex Repairable Systems," *Applied Stochastic Models in Business and Industry* 20 (2004): 253–264.

139. D. Cavallo and F. Ruggeri, "Bayesian Models for Failures in a Gas Network," in *Safety and Reliability*, ed. E. Zio, M. Demichela, and N. Piccinini (Politecnico di Torino Editore, 2001), 1963–1970.

140. A. Pievatolo, F. Ruggeri, and R. Argiento, "Bayesian Analysis and Prediction of Failures in Underground Trains," *Quality and Reliability Engineering International* 19 (2003): 327–336.

141. A. Pievatolo and F. Ruggeri, "Bayesian Modelling of Train Doors' Reliability," in *Handbook of Applied Bayesian Analysis*, ed. A. O'Hagan and M. West (Oxford Press, 2010).

142. D. Rios Insua, F. Ruggeri, and M. P. Wiper, *Bayesian Analysis of Stochastic Processes Models* (Wiley, 2012).

143. A. O'Hagan, C. E. Buck, A. Daneshkhah, et al., *Uncertain Judgements: Eliciting Experts' Probabilities* (Wiley, 2006).

144. T. L. Saaty, *The Analytic Hierarchy Process* (McGraw-Hill, 1980).

145. S. A. Levin and J. Lubchenco, "Resilience, Robustness, and Marine Ecosystem-Based Management," *BioScience* 58 (2008): 27–32.

146. P. P. Y. Wu, F. Ruggeri, and K. Mengersen, "Optimal Clustering With Dependent Costs in Bayesian Networks," 2024, Submitted.

147. P. P. Y. Wu, M. J. Caley, G. A. Kendrick, K. McMahon, and K. Mengersen, "Dynamic Bayesian Network Inferencing for Non-Homogeneous Complex Systems," *Journal of the Royal Statistical Society. Series C, Applied Statistics* 67 (2018): 417–434.

148. D. Banks, J. Rios, and I. D. Rios, *Adversarial Risk Analysis* (Chapman and Hall/CRC, 2015).

149. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).

150. Y. Vorobeichyk and M. Kantarcioglu, *Adversarial Machine Learning* (Morgan & Claypool, 2019).

151. V. Gallego, R. Naveiro, A. Redondo, D. Rios Insua, and F. Ruggeri, "Protecting Classifiers From Attacks," *Statistical Science* 39 (2024): 449–468.

152. R. Naveiro, A. Redondo, D. Rios Insua, and F. Ruggeri, "Adversarial Classification: An Adversarial Risk Analysis Approach," *International Journal of Approximate Reasoning* 113 (2019): 133–148.

153. L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science* 16, no. 3 (2001): 199–231.

154. G. Saporta, "Models for Understanding Versus Models for Prediction," in *COMPSTAT 2008: Proceedings in Computational Statistics* (Physica-Verlag HD, 2008), 315–322.

155. L. Bottou, J. Peters, J. Quiñonero-Candela, et al., "Counterfactual Reasoning and Learning Systems: The Example of Computational Advertising," *Journal of Machine Learning Research* 14, no. 11 (2013): 3207–3260.

156. V. Vapnik, *Estimation of Dependences Based on Empirical Data* (Springer, 1982).

157. H. R. Varian, "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives* 28, no. 2 (2014): 3–28.

158. D. J. Hand, "Classifier Technology and the Illusion of Progress," *Statistical Science* 21, no. 1 (2006): 1–14.

159. J. Su, D. V. Vargas, and K. Sakurai, "One Pixel Attack for Fooling Deep Neural Networks," *IEEE Transactions on Evolutionary Computation* 23, no. 5 (2019): 828–841.

160. C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, 2017).

161. C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence* 1, no. 5 (2019): 206–215.

162. U. Grömping, "Variable Importance in Regression Models," *Wiley Interdisciplinary Reviews: Computational Statistics* 7, no. 2 (2015): 137–152.

163. S. Mitchell, E. Potash, S. Barocas, A. D'Amour, and K. Lum, "Algorithmic Fairness: Choices, Assumptions, and Definitions," *Annual Review of Statistics and Its Application* 8, no. 1 (2021): 141–163.

164. B. Schölkopf and J. von Kügelgen, "From Statistical to Causal Learning," in *Proceedings of the International Congress of Mathematicians*, vol. 7 (EMS Press, 2022), 5540–5593.

165. P. Horst, P. Wallin, L. Guttman, et al., "Testing of Prediction Procedure and Revision of Hypotheses," in *(Collaborators), the Prediction of Personal Adjustment: A Survey of Logical Problems and Research Techniques, With Illustrative Application to Problems of Vocational Selection, School Success, Marriage, and Crime*, ed. P. Horst, P. Wallin, L. Guttman, et al. (Social Science Research Council, 1941), 115–118.

166. H. Abdi, A. Di Ciaccio, and G. Saporta, "Old and New Perspectives on Optimal Scaling," in *Analysis of Categorical Data From Historical Perspectives: Essays in Honour of Shizuhiko Nishisato*, ed. E. J. Beh, R. Lombardo, and J. G. Clavel (Springer Nature Singapore, 2024), 131–154.

167. C. Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired* (2008), https://www.wired.com/2008/06/pb-theory/.

168. P. Norvig, "Fact-Checking,", https://norvig.com/fact-check.html.

169. C. Rovelli, *Helgoland* (Adelphi Editore, 2020).

170. B. Iooss, R. Kenett, and P. Secchi, "Different Views of Interpretability," in *Interpretability for Industry 4.0: Statistical and Machine Learning Approaches*, ed. A. Lepore, B. Palumbo, and J. M. Poggi (Springer, 2022).

171. H. Wang and J. S. Marron, "Object Oriented Data Analysis: Sets of Trees," *Annals of Statistics* 35, no. 5 (2007): 1849–1873.

172. J. S. Marron and I. L. Dryden, *Object Oriented Data Analysis* (Chapman and Hall/CRC, 2022).

173. L. M. Sangalli, P. Secchi, and S. Vantini, "AneuRisk65: A Dataset of Three-Dimensional Cerebral Vascular Geometries," *Electronic Journal of Statistics* 8 (2014): 1879–1890.

174. M. Piccinelli, A. Veneziani, D. Steinman, A. Remuzzi, and L. Antiga, "A Framework for Geometric Analysis of Vascular Structures: Application to Cerebral Aneurysms," *IEEE Transactions on Medical Imaging* 28, no. 8 (2009): 1141–1155.

175. L. M. Sangalli, P. Secchi, S. Vantini, and A. Veneziani, "A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery," *Journal of the American Statistical Association* 104, no. 485 (2009): 37–48.

176. L. M. Sangalli, P. Secchi, S. Vantini, and A. Veneziani, "Efficient Estimation of Three-Dimensional Curves and Their Derivatives by Free-Knot Regression Splines, Applied to the Analysis of Inner Carotid Artery Centrelines," *Journal of the Royal Statistical Society. Series C, Applied Statistics* 58, no. 3 (2009): 285–306.

177. M. Pegoraro and P. Secchi, "Functional Data Representation With Merge Trees," 2021, *arXiv:2108.13147*.

178. R. Scimone, T. Taormina, B. M. Colosimo, M. Grasso, A. Menafoglio, and P. Secchi, "Statistical Modeling and Monitoring of Geometrical Deviations in Complex Shapes With Application to Additive Manufacturing," *Technometrics* 64, no. 4 (2022): 437–456.

179. K. G. van den Boogaart, J. J. Egozcue, and V. Pawlowsky-Glahn, "Bayes Hilbert Spaces," *Australian & New Zealand Journal of Statistics* 56, no. 2 (2014): 171–194.

180. K. Hron, A. Menafoglio, M. Templ, K. Hrůzová, and P. Filzmoser, "Simplicial Principal Component Analysis for Density Functions in Bayes Spaces," *Computational Statistics & Data Analysis* 94 (2016): 330–350.

181. B. M. Colosimo and M. Pacella, "A Comparison Study of Control Charts for Statistical Monitoring of Functional Data," *International Journal of Production Research* 48, no. 6 (2010): 1575–1601.

182. A. Menafoglio and P. Secchi, "Statistical Analysis of Complex and Spatially Dependent Data: A Review of Object Oriented Spatial Statistics," *European Journal of Operational Research* 258, no. 2 (2017): 401–410.

183. A. Menafoglio, G. Gaetani, and P. Secchi, "Random Domain Decompositions for Object-Oriented Kriging Over Complex Domains," *Stochastic Environmental Research and Risk Assessment* 32, no. 12 (2018): 3421–3437.

184. A. Menafoglio, D. Pigoli, and P. Secchi, "Kriging Riemannian Data via Random Domain Decompositions," *Journal of Computational and Graphical Statistics* 30, no. 3 (2021): 709–727.

185. L. M. Sangalli, J. O. Ramsay, and T. O. Ramsay, "Spatial Spline Regression Models," *Journal of the Royal Statistical Society, Series B (Methodology)* 75, no. 4 (2013): 681–703.

186. S. N. Wood, M. V. Bravington, and S. L. Hedley, "Soap Film Smoothing," *Journal of the Royal Statistical Society, Series B (Methodology)* 70, no. 5 (2008): 931–955.

187. L. Azzimonti, L. M. Sangalli, P. Secchi, M. Domanin, and F. Nobile, "Blood Flow Velocity Field Estimation via Spatial Regression With PDE Penalization," *Journal of the American Statistical Association* 110, no. 511 (2015): 1057–1071.

188. A. L. Mackay, *A Dictionary of Scientific Quotations* (Routledge, 2019).

189. G. Shmueli, "To Explain or to Predict?," *Statistical Science* 25, no. 3 (2010): 289–310.

190. B. Efron, "Prediction, Estimation, and Attribution," *International Statistical Review* 88 (2020): S28–S59.

191. A. Chakrabarti and R. Sen, "Copula Estimation for Nonsynchronous Financial Data," *Sankhya B* 85, no. Suppl 1 (2023): 116–149.

192. B. H. Lindqvist and J. T. Kvaløy, "New Tests for Trend in Time Censored Recurrent Event Data," *Applied Stochastic Models in Business and Industry* 40 (2024): 1275–1290.

193. S. Biswas and R. Sen, "Kernel Based Estimation of Spectral Risk Measures," *Journal of Risk* 26, no. 5 (2024): 17–47.

194. S. Friedrich, G. Antes, S. Behr, et al., "Is There a Role for Statistics in Artificial Intelligence?," *Advances in Data Analysis and Classification* 16, no. 4 (2020): 823–846.

195. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?," *Journal of Machine Learning Research* 15, no. 1 (2014): 3133–3181.

196. H. Jaeger, *The Echo State Approach to Analysing and Training Recurrent Neural Networks-With an Erratum Note. GMD Technical Report*, vol. 148 (German National Research Center for Information Technology, 2001), 13.

197. L. Le Cam, "On the Asymptotic Theory of Estimation and Testing Hypotheses," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, vol. 3 (University of California Press, 1956), 129–157.

198. A. Steland and B. E. Pieters, "One-Round Cross-Validation and Uncertainty Determination for Randomized Neural Networks With Applications to Mobile Sensors," in *Artificial Intelligence, Big Data and Data Science in Statistics: Challenges and Solutions in Environmetrics, the Natural Sciences and Technology*, ed. A. Steland and K.-L. Tsui (Springer, 2022), 3–24.

199. J. S. Greipel, R. M. Frank, M. Huber, A. Steland, and R. H. Schmitt, "Auto-Encoder-Based Algorithm for the Selection of Key Characteristics for Products to Reduce Inspection Efforts," *International Journal of Quality & Reliability Management* 40, no. 7 (2023): 1597–1620.

200. J. Tebbe, C. Zimmer, A. Steland, M. Lange-Hegermann, and F. Mies, "Efficiently Computable Safety Bounds for Gaussian Processes in Active Learning," in *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research*, vol. 238 (IEEE, 2024), 1333–1341.

201. A. Khaireh-Walieh, D. Langevin, P. Bennet, O. Teytaud, A. Moreau, and P. Wiecha, "A Newcomer's Guide to Deep Learning for Inverse Design in Nano-Photonics," *Nano* 12, no. 24 (2023): 4387–4414.

202. L. Ardizzone, J. Kruse, S. Wirkert, et al., "Analyzing Inverse Problems With Invertible Neural Networks, *arXiv:1808.04730*," 2018.

203. L. Ardizzone, C. Lüth, J. Kruse, C. Rother, and U. Köthe, "Guided Image Generation With Conditional Invertible Neural Networks, *arXiv:1907.02392*," 2019.

204. M. Scutari and J. B. Denis, *Bayesian Networks With Examples in R*, 2nd ed. (Chapman and Hall/CRC, 2021).