Spanish Journal of Statistics
Vol. **xx** No. xx 2025, Pages 131–139
doi:https://doi.org/10.37830/
SJS.2025.1.07

Invited Article

Address at the 2024 Spanish National Award in Statistics

Concha Bielza

Universidad Politécnica de Madrid, mcbielza@fi.upm.es

Received: November 5, 2025. Returned: November 25, 2025.

Abstract: This article is based on the address given upon receiving the 2024 Spanish national award in statistics, a ceremony made especially meaningful by the attendance of His Majesty the King of Spain. It offers a personal overview of my research trajectory, shaped by the long-standing interplay between statistics and artificial intelligence, and by their applications in neuroscience and industry. After beginning my career in statistical decision theory and probabilistic graphical models, Bayesian networks soon became the central framework of my work, enabling rigorous reasoning under uncertainty across domains.

In neuroscience, my contributions span neuronal classification, spatial analysis of synapses, modeling of dendritic arborizations, biomarker discovery for neurological disorders, and the decoding of brain activity, among others. These efforts, supported by landmark programs such as the Cajal Blue Brain Project and the Human Brain Project, were driven by the need for models capable of capturing complex, high-dimensional, and often unconventional data.

Around 2018, my research turned increasingly toward Industry 4.0, where real-time data streams, dynamic systems, and predictive and prescriptive maintenance posed demanding methodological challenges. This led to advances in dynamic Bayesian networks, latent-variable models, and probabilistic evolutionary algorithms for high-dimensional optimization.

Alongside scientific discovery, I have remained committed to teaching, mentoring, and knowledge transfer through initiatives such as the Machine Learning and Advanced Statistics Summer School at the Universidad Politécnica de Madrid. I conclude with reflections on interdisciplinary work, convergence of statistics and machine learning, ethical use of data and algorithms, and the importance of inspiring new generations of statisticians.

Keywords: Bayesian networks, probabilistic machine learning, Bayesian decision theory, reasoning under uncertainty, heuristic optimization, temporal data, interpretable models, artificial intelligence, neuroscience, industry 4.0

MSC: 62-09, 62P99, 68T37, 90B50

1 Introduction

This article offers a written version of the speech I delivered on April 7, 2025, the day I received the 2024 Spanish national award in statistics. I am grateful to the editor of the *Spanish Journal of Statistics* for the invitation to share it with the journal's readers. The ceremony took place at the Royal Spanish Academy of Exact, Physical, and Natural Sciences in Madrid. The event was made especially memorable by the presence of His Majesty King Felipe VI of Spain.

The video of the ceremony can be seen in https://www.youtube.com/watch?v=3wqZ8h6OrYg.

2 Opening remarks

Good morning. Your Majesty, Secretary of State, President of the Royal Spanish Academy of Exact, Physical, and Natural Sciences, President of the National Statistics Institute, authorities, guests, colleagues, family members, and friends.

I would like to begin by expressing my gratitude to His Majesty for his presence, which elevates this ceremony and lends even greater significance to this event. To the Secretary of State, for the institutional support he represents. And to this Royal Academy, for welcoming us into this magnificent home as host of this occasion.

My special thanks go to the National Statistics Institute for establishing this prestigious award, which highlights the value of statistics as an essential tool for informed decision-making and encourages us to continue developing methodologies that transform data into useful knowledge for society.

It is a great honor to receive this award, which recognizes years of effort and dedication and brings visibility to my work. The stature of previous laureates—three of whom are here today—confirms its relevance (see Figure 1).

Figure 1: His Majesty King Felipe VI (center), together with four of the five laureates—shown from left to right: 2022 (Enrique Castillo), 2021 (Wenzeslao González), 2024 (the author, Concha Bielza), and 2020 (Daniel Peña)—and the Presidents of the National Statistics Institute and of the Royal Academy of Sciences.

My thanks also go to the selection committee for granting me this distinction, to those who supported my candidacy with their letters, and to those who encouraged me to apply. I am likewise grateful to the other candidates, who help keep interest in this award alive. And thank you, Pedro, for your words in the *laudatio* and for being, amid the daily whirlwind, a true friend and my best companion in both hardships and achievements over many years (see Figure 2).

Figure 2: Moments from the ceremony: Pedro Larrañaga delivering the *laudatio* (left), and the applause following the award ceremony, with His Majesty King Felipe VI, the Secretary of State, and the President of the National Statistics Institute (right).

3 Foundations of my research career

My path in statistics began during my undergraduate studies in Mathematics at the Universidad Complutense de Madrid. After finishing my degree, I taught at the beautiful Royal University Center María Cristina in San Lorenzo de El Escorial (see Figure 3, left)—a private university that is part of the famous Monastery—and I also worked in a major Spanish company in the field of computing, where, without realizing it at the time, I was already moving toward artificial intelligence through what were then known as "expert systems."

In 1991, I joined the School of Computer Engineering (then still the Faculty of Computer Science) at the Universidad Politécnica de Madrid (see Figure 3, right), where I began my doctoral studies in the Department of Artificial Intelligence, which included the area of statistics and operations research. This was thanks to Sixto and David Ríos Insua, whose support was fundamental in my early career.

With David—an Academician of this Royal Academy—I carried out my PhD on decision-making under uncertainty within Bayesian decision theory. We represented the decision-making problems using influence diagrams (see Figure 4), a then-recent probabilistic graphical model, and extended them to contexts under partial information.

We applied these ideas in a decision-support system for neonatal jaundice, developed in collaboration with the Gregorio Marañón Hospital, which enabled less invasive treatments and improved clinical understanding. This work gave rise to two PhD dissertations that I supervised: one on modeling and another on explanation generation— a topic we were already addressing in the early 2000s

Figure 3: Royal University Center María Cristina in San Lorenzo de El Escorial (left) and the School of Computer Engineering at the Universidad Politécnica de Madrid (right).

and which is now reflected in European artificial intelligence regulation through the so-called right to a clear and comprehensible explanation, essential for building trust in algorithmic decisions.

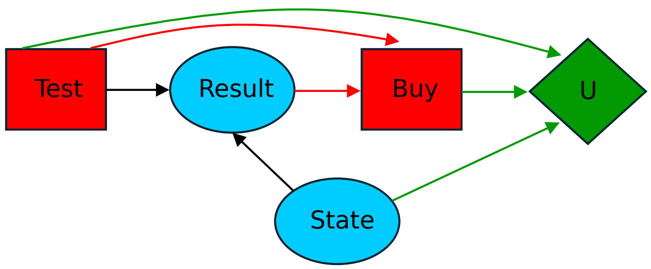

Figure 4: The structure of an influence diagram.

I also wrote several textbooks—on statistics for medical doctors, on decision-support systems, and on operations research (see Figure 5). I worked on decision-making problems across different sectors, and together with the School of Agricultural Engineering at the Universidad Politécnica de Madrid, we developed predictive models in the agri-food domain to estimate the probability of fruit damage as it moved along grading lines, which led to registered software that attracted the interest of cooperatives.

I consolidated my research independence and collaborated with international colleagues, such as Prakash Shenoy. We presented joint works at one of the early editions of AISTATS—then still a workshop—the seed of today's community that brings together statistics and artificial intelligence. I was fascinated to see how both worlds were coming to meet each other: AI was recognizing the need to incorporate uncertainty, and statistics was relying on increasingly powerful computational methods. At this intersection, probabilistic graphical models—particularly Bayesian networks—were beginning to stand out. These networks have been the cornerstone of my research. They represent uncertain knowledge through interpretable graphs and enable probabilistic reasoning of any kind.

Judea Pearl, their principal proponent, has been described as "the most original and influential thinker in statistics today." The influence diagrams that shaped my early research could be seen as their informal precursors, since Bayesian networks offered a more powerful formalism by explicitly incorporating the semantics of conditional independence, thus linking graphs and probabilities.

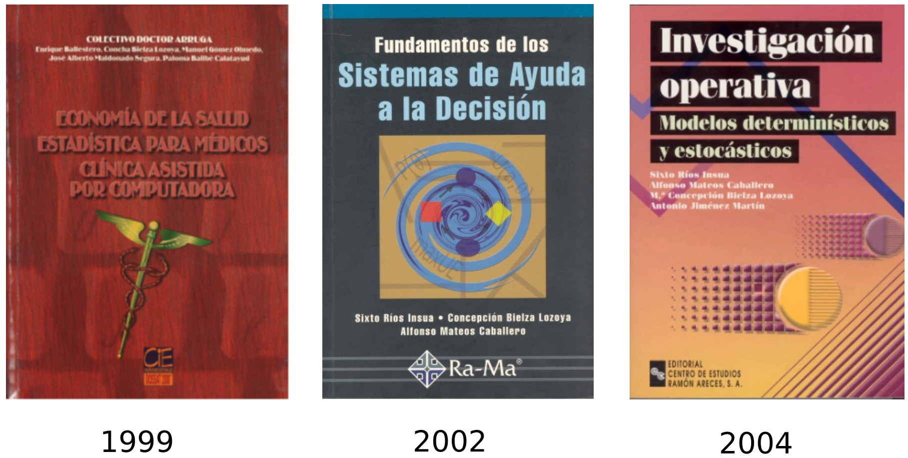

Figure 5: First textbooks authored.

My first applications of Bayesian networks were in predicting surface roughness in high-speed machining, in collaboration with the Institute of Industrial Automation of the Spanish National Research Council (CSIC), a project that led to the third PhD dissertation I supervised. Indeed, some years later, industry would become central to my research.

4 Towards a data-driven neuroscience

The arrival of Pedro Larrañaga at the Universidad Politécnica de Madrid at the end of 2007 marked the beginning of my next stage, which was undoubtedly a fruitful one. Together, we created the Computational Intelligence Group from scratch, with a strong commitment to computational neuroscience, just when our university was joining ambitious international initiatives focused on the study of the brain—one of the great scientific challenges of the 21st century. I will highlight some milestones achieved over more than 15 years of work in neuroscience, mainly within two major projects: the Cajal Blue Brain Project and the Human Brain Project.

In neuroanatomy, we addressed the classification and nomenclature of neurons—a century-old debate—on the basis of their morphology, electrophysiology, and transcriptome, achieving high-impact publications in journals of the *Nature* group. It required a significant effort to establish the value of data-driven models within a community not accustomed to using them.

We also explored the spatial organization of synapses and dendritic spines through spatial statistics—barely used in this field—to investigate whether their arrangement follows structured patterns or is essentially random. Another challenge was to test the hypothesis that neuronal arborizations are designed to maximize connectivity while minimizing wiring length.

It was a real gift to work with the renowned neuroscientists Javier deFelipe (CSIC) and Rafael Yuste (Columbia University), see Figure 6.

In neurological disorders such as Alzheimer's disease, Parkinson's disease, and epilepsy, we worked on the identification of biomarkers, the definition of clinical subtypes, and the prediction of disease progression and quality of life—the latter based on multiple dimensions (see Figure 7). We used highly diverse data—genetic, clinical, neuropsychological, imaging, and questionnaire-based—and collaborated with hospitals and research centers, both national and international. We

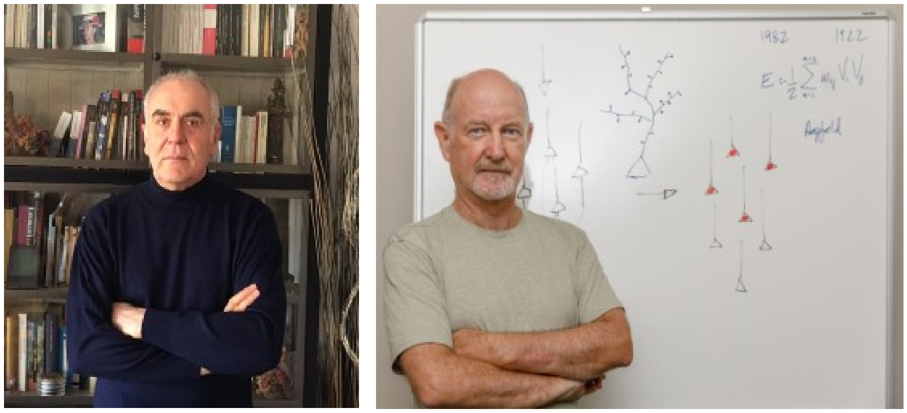

Figure 6: Javier de Felipe (CSIC) (left) and Rafael Yuste (Columbia University) (right).

also worked on classifying mental activity from magnetoencephalography recordings, contributing to the so-called brain decoding, a key line of research in brain–computer interfaces.

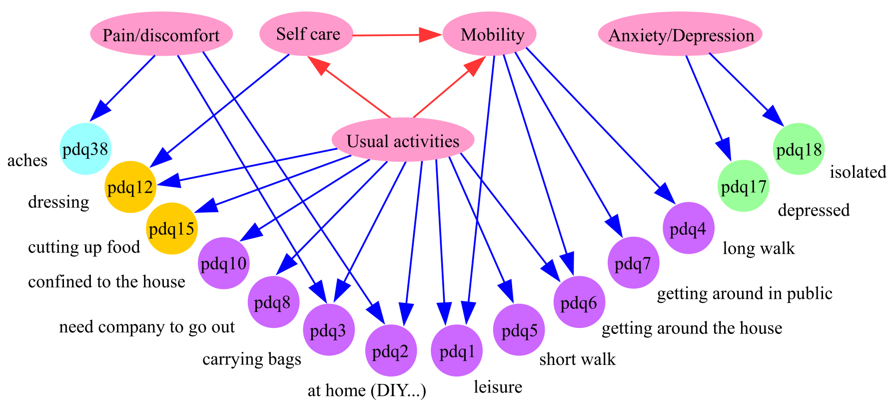

Figure 7: A Bayesian network to predict the European quality of life-5 dimensions (EQ-5D, in pink) from the 39-item Parkinson's disease questionnaire (PDQ-39).

In neuroscience, the lack of data-driven models led me to embark, with Pedro, on the writing of a book published by Cambridge University Press in 2021, which brings together methods from statistics and machine learning and their applications to real data (see Figure 8, left). After more than six years of work and nearly 700 pages, a dear collaborator—who is here today—suggested marking the occasion with steaks matching the weight of the book—almost one and a half kilos. The image has stayed with us ever since… although the meal has yet to be served.

The applications have been highly rewarding because of their real-world impact, but behind each of them there is, more often than not, a methodological gap that needed to be addressed, leading us to develop our own solutions, particularly in the area of Bayesian networks. We have often dealt with atypical types of data: directional data (such as neuronal branching angles); non-Gaussian data (common in real applications); or datasets with far more variables than observations (as in genomics).

We have also faced challenges such as noisy labels (as in the long-standing, inconsistent nomenclature of neurons); multiple interrelated classes (as in the multidimensional assessment of quality of life in Parkinson's disease); model consensus (for example, when integrating the opinions of several experts); computationally efficient learning; the construction of regularized—and even mas-

Figure 8: Recent textbooks authored.

sive—networks; the identification of the most probable or most relevant explanations for a given piece of evidence; and the algebraic characterization of decision boundaries.

5 Probabilistic modeling for Industry 4.0

Around 2018, my research began to shift toward the industrial sector, within the framework of the so-called Industry 4.0, in which sensors, machines, and people collaborate in interconnected systems that require real-time analysis. Time series and data streams arrive at high sampling rates, in the form of electrical, mechanical, or environmental signals, reflecting complex processes that evolve over time.

Working hand in hand with several companies, we have addressed multiple dynamic problems, where the aim is to monitor the operational state of assets, anticipate failures (predictive maintenance), and determine the best course of action (prescriptive maintenance).

Among these applications, I will highlight two. The first concerns industrial furnaces with tubes through which a thermal fluid circulates. The progressive accumulation of impurities on their walls hampers heat transfer and forces an increase in energy consumption to maintain the desired temperature of the fluid. If no action is taken in time, the thermal limit of the material may be reached, causing damage. Using dynamic Bayesian networks, we were able to anticipate several days in advance when a section of the furnace would require cleaning, thereby optimizing maintenance and avoiding inefficiencies or risks.

The second application focuses on estimating the remaining useful life of bearings used in machine tools (Figure 9) and in pumps at desalination plants, whose progressive degradation generates signals that change in complex ways. We developed a model with latent variables—variables that are not directly observed—that automatically captures the different phases of deterioration and adapts to the evolution of the signal.

In all these contexts, we have designed increasingly complex and adaptive dynamic Bayesian networks. It is particularly important to identify different operating modes, detect anomalies, or determine when a model ceases to be valid because the underlying process has changed.

Figure 9: A degraded bearing (image courtesy of Aingura IIoT).

To strengthen this new industrial orientation, I became involved in another book, this one on *Industrial Applications of Machine Learning*, published by CRC Press in 2019 and translated into Chinese in 2023 (see Figure 8, right).

In some of the problems we addressed, it was necessary to optimize complex, multi-objective functions, often in immense spaces where exact methods are not feasible. One example is the optimal design of a biolubricant for MotoGP racing, with millions of possible ingredients for the formulation. Here, we contributed to the development of stochastic evolutionary algorithms based on Bayesian networks—so-called estimation of distribution algorithms—which, although they provide no guarantee of optimality, make it possible to find good solutions and to better understand the optimization process.

6 Reflections and gratitude

Research is not only about discovery; it is also about teaching and sharing. I have maintained that commitment not only in the university classroom, but also in other professional settings, such as with civil servants from the Ministry of Education, the Directorate-General for the Cadastre, the Carlos III Health Institute, or the Cervantes Institute in Berlin. But if I had to highlight one initiative, it would be the Machine Learning and Advanced Statistics Summer School, which I have co-directed for many years at the Universidad Politécnica de Madrid and which is updated regularly to reflect current advances (see Figure 10).

Before closing, I would like to highlight several ideas that have been central to my career. First, the importance of interdisciplinary work in addressing today's major challenges with a richer and more complementary perspective. Second, the value of building bridges between statistics and machine learning—as Bayesian networks do—to approach the analysis of complex modern data with rigor and efficiency. Third, the need to ensure the ethical use of data and algorithms, placing people at the center and promoting statistical literacy; in this spirit, I encourage signing the manifesto promoted by the Spanish Society of Statistics and Operations Research (SEIO). And finally, the importance of a strong commitment to technology transfer, from the university to the industrial ecosystem.

I conclude these words with my deepest gratitude to those who have accompanied me on this journey, for this award belongs to them as well. To the Universidad Politécnica de Madrid, for allowing me to develop my career freely in an environment committed to research and technology transfer. To my collaborators, institutions, and companies with whom I have worked, with the hope of continuing to move forward together. To my doctoral students—past and present—for their talent,

Figure 10: Machine Learning and Advanced Statistics Summer School at the Universidad Politécnica de Madrid.

dedication, and trust, even at a time when job offers outside academia are more tempting than ever. To Laura, our program coordinator, for her constant efficiency and unfailingly positive attitude, even when facing bureaucracy. And to my friends and family, for their endless support, understanding, and love throughout the journey of life.

Allow me to mention three special people from an earlier generation, from whom one always learns. My aunt Concha—Concha Bielza, like me—who is here today and who, at almost 98 years old, remains a model of vitality and affection. And my parents, who always fostered an environment of study and instilled in my three sisters and me a culture of effort. My father would have loved to witness this moment, even if, back in the day, he advised me against studying Mathematics ("What for?"), preferring that I choose an engineering degree as he had done. And my mother, who has been able to accompany me today, making a great effort simply because she loves me.

I hope this award inspires new generations of statisticians—women and men alike—to pursue excellence, to work responsibly, and to contribute with enthusiasm to the progress of our discipline.

THANK YOU VERY MUCH.

Acknowledgments

This work was partially supported by the MINISTRY OF SCIENCE, INNOVATION AND UNIVERSITIES [Project 482 AEI/10.13039/501100011033-PID2022-139977NB-I00, Project PLEC2023-010252/MIG- 483 20232016]. Also, by the AUTONOMOUS REGION OF MADRID [Project 484 ELLIS Unit Madrid and TEC-2024/COM-89].