

Objective Model Selection in Physics: Exploring the Finite Information Quantity Approach

Boris Menin

Independent Researcher, Beer-Sheva, Israel

Email: meninbm@gmail.com

How to cite this paper: Menin, B. (2024) Objective Model Selection in Physics: Exploring the Finite Information Quantity Approach. *Journal of Applied Mathematics and Physics*, 12, 1848-1889. <https://doi.org/10.4236/jamp.2024.125115>

Received: April 26, 2024

Accepted: May 27, 2024

Published: May 30, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Traditional methods for selecting models in experimental data analysis are susceptible to researcher bias, hindering exploration of alternative explanations and potentially leading to overfitting. The Finite Information Quantity (FIQ) approach offers a novel solution by acknowledging the inherent limitations in information processing capacity of physical systems. This framework facilitates the development of objective criteria for model selection (comparative uncertainty) and paves the way for a more comprehensive understanding of phenomena through exploring diverse explanations. This work presents a detailed comparison of the FIQ approach with ten established model selection methods, highlighting the advantages and limitations of each. We demonstrate the potential of FIQ to enhance the objectivity and robustness of scientific inquiry through three practical examples: selecting appropriate models for measuring fundamental constants, sound velocity, and underwater electrical discharges. Further research is warranted to explore the full applicability of FIQ across various scientific disciplines.

Keywords

Comparative Uncertainty, Finite Information Quantity, Formulating a Model, Measurement Accuracy Limit, Objective Model Selection

1. Introduction

Accurate measurements of physical processes are fundamental to scientific research and technological progress. These processes cover a wide range of phenomena: from the behavior of particles at the quantum level to the dynamics of celestial bodies in space. The accuracy and reliability of these measurements are essential to improving our understanding of the universe and promoting innovation in various fields.

In addition to its theoretical importance, accurate measurements of physical processes play a crucial role in practical applications. They provide the basis for the development and optimization of technologies that shape our daily lives, from medical devices to renewable energy systems. The reliability of these measurements directly impacts the efficiency and effectiveness of these technologies, impacting everything from healthcare outcomes to environmental sustainability.

Ensuring measurement accuracy requires not only precise technical equipment, but also the selection of appropriate mathematical models. Models serve as tools for interpreting experimental data and predicting the behavior of physical systems. Choosing the right model is important to minimize errors and uncertainties and ensure that the data accurately reflects the underlying phenomena.

However, selecting the best model can be challenging due to the complexities inherent in real-world phenomena. Physical processes often exhibit nonlinear behavior, unpredictable interactions, and emergent properties that defy easy explanation. As a result, researchers must carefully evaluate different modeling approaches and their suitability to capture the nuances of the phenomena being studied.

Moreover, traditionally, model selection—the process of selecting the most appropriate mathematical structure to represent observed phenomena—has relied heavily on the knowledge, intuition, and experience of the researcher. While these factors are valuable, they may lead to certain limitations:

- 1) Subjectivity. Individual experiences and biases of researchers can influence their choice of model, potentially leading to inconsistencies and preventing the search for truly objective solutions.

- 2) Limited scope. Often the focus is on identifying one “best” model based on specific criteria, ignoring the potential value of exploring alternative explanations and understanding the entire landscape of possibilities surrounding a phenomenon.

- 3) Risk of overtraining. Complex models, although offering a seemingly close fit to existing data, can be prone to overfitting, which affects their generalizability and accuracy in predicting new observations.

These limitations highlight the need for innovative approaches to model selection that minimize subjective bias, enhance our understanding of physical phenomena, and reduce the risk of overfitting.

We will look at various methods for selecting the best model when measuring physical processes, their advantages and disadvantages, considering their applicability to different types of data and scientific purposes. By gaining a better understanding of model selection techniques, researchers and engineers can improve the accuracy and reliability of their measurements, advancing both scientific knowledge and technological innovation.

This paper introduces the Finite Information Quantity (FIQ) approach, a new methodology emerging from physics that addresses these problems by incorporating the fundamental principle of finite information processing into physical

systems.

The proposed approach involves analyzing the amount of information contained within a model, which is constructed based on the observer's knowledge and experience. The study argues that this information can play a critical role in assessing the attainable accuracy of representing a modeled phenomenon. The method also addresses the concept of *comparative uncertainty*, which is a fundamental aspect for evaluating the accuracy limit of a model [1].

The main idea of the FIQ approach is that information, unlike classical physics where it is considered infinitely divisible, is a finite and limited resource in any physical system. This limitation may be due to the fundamental laws of thermodynamics or quantum mechanics. Recognizing these limitations, the FIQ approach seeks to establish *objective criteria (comparative uncertainty)* for model selection based on the information processing capabilities of the system being modeled.

This article will detail the details of the FIQ approach, exploring its theoretical underpinnings and potential benefits in overcoming the limitations of traditional model selection methods. We compare the FIQ approach with existing methodologies, highlighting its unique contribution to achieving greater objectivity and reliability in our understanding of physical phenomena.

In addition, the discussion will explore possible applications of the FIQ approach in various fields of physics, exploring its potential for discovering new ideas and revolutionizing the way we interpret experimental data and formulate accurate physical models.

2. Short Review of the Applied Methods to Verify the Most Accurate Model

The field of scientific inquiry relies on a rich toolbox of methods for processing experimental data and formulating models of physical phenomena. These methods are particularly valuable when they excel at modeling uncertainty and enabling probabilistic inferences. This paper introduces ten prominent methods that have gained widespread application in this domain.

2.1. List Squares Method

This method minimizes the sum of squared differences between observed and predicted values. It is widely used when fitting linear models to experimental data, helping identify the coefficients that best describe the relationship. It is a venerable and widely employed statistical technique that serves as a linchpin in the realms of regression analysis and parameter estimation. It has found ubiquitous application across various scientific disciplines, owing to its simplicity and versatility. In this detailed exploration, we will meticulously examine the advantages and disadvantages of the LSM, elucidating its strengths and limitations in the context of precision and reliability. The LSM's longevity and widespread adoption underscore its status as a well-established tool in statistical

modeling. Originating in the early 19th century, it has evolved to become a cornerstone in data analysis. Its enduring popularity is attributed to its simplicity, ease of implementation, and effectiveness in various applications. Researchers and practitioners appreciate the method's familiarity and reliability, making it a default choice in situations where a quick, robust solution is required.

Advantages:

1) Simplicity: LSM is straightforward and easy to implement. Its simplicity makes it accessible to a wide range of users, especially in cases where a quick and simple model is sufficient.

2) Analytical Solutions: For linear models, there are analytical solutions available for finding the optimal coefficients. This allows for a direct calculation of the best-fit line, providing insights into the relationship between variables.

3) Interpretability: The coefficients obtained from the regression analysis have clear interpretations. In a linear model, each coefficient represents the change in the dependent variable for a one-unit change in the corresponding independent variable, facilitating the interpretation of results [2].

4) Widespread Applicability: LSM is widely used in various fields, including physics, economics, biology, and engineering. Its versatility makes it a go-to method for initial exploratory data analysis [3].

5) Efficient with Large Datasets: When dealing with large datasets, the computational efficiency of the LSM is advantageous. It can handle a considerable amount of data without significant computational burden. In [4] the authors highlight recent advancements in efficient optimization algorithms for Least Squares Regression with massive datasets.

Disadvantages:

1) Sensitive to outliers: despite its robustness to noise, the Least Squares Method is sensitive to outliers, which are data points significantly deviating from the general trend. Outliers can unduly influence the regression coefficients, leading to biased estimates. The method's reliance on minimizing squared differences amplifies the impact of extreme values, potentially distorting the overall fit of the model. Careful consideration and outlier detection techniques are necessary to mitigate the influence of outliers on the results. Outliers can significantly impact the results of least squares regression. Since the method minimizes the sum of squared errors, a single outlier can disproportionately influence the estimated coefficients and distort the model. This sensitivity is highlighted by Shi *et al.* in their review of robust regression techniques, emphasizing the need for outlier detection and mitigation strategies [5].

2) Assumes a linear relationship between variables: a fundamental assumption of the Least Squares Method is the linearity of the relationship between variables. While this assumption holds in many practical scenarios, it imposes a limitation on the types of relationships the method can effectively model. Nonlinear relationships may go undetected, leading to inaccurate predictions. In cases where the true relationship is nonlinear, alternative modeling techniques, such as po-

ynomial regression or nonlinear least squares, may be more appropriate. If the underlying relationship is nonlinear, as suggested by Rolnick *et al.* in their work on deep learning for climate data, LSR can lead to inaccurate predictions compared to more flexible approaches [6].

3) Does not account for uncertainties in independent variables: another notable limitation of the LSM is its failure to account for uncertainties in the independent variables. The method assumes that the predictor variables are measured precisely, neglecting any potential errors associated with their measurements. In situations where the independent variables have inherent uncertainties, this assumption can lead to underestimated standard errors and, consequently, inaccurate confidence intervals for the regression coefficients. Researchers must exercise caution when applying the method in situations where uncertainty in predictor variables is a critical consideration.

4) Assumption of homoscedasticity: least squares regression assumes that the variance of the errors is constant across all levels of the independent variable. The authors of [7] propose non-parametric approaches for cases where heteroscedasticity violates Least Squares Regression assumptions, offering alternatives like conformal prediction for more reliable uncertainty estimation.

5) Multicollinearity issues: highly correlated independent variables (multicollinearity), as discussed by Sirimongkolkasem *et al.* (2019) in their work on sparse regularization, can lead to unstable coefficient estimates [8]. Identifying the individual contribution of each variable becomes challenging, and the precision of the estimates may be compromised.

6) Doesn't handle missing data well: the LSM relies on complete datasets. If there are missing values in the data, traditional least squares regression may not be applicable, and imputation or other methods would be needed. Missing data presents a challenge for Least Squares Regression, requiring imputation techniques like those reviewed by Liu *et al.* with deep learning approaches, or alternative model structures that can accommodate missingness [9].

It must be mentioned that LSM is not suitable for modeling complex nonlinear relationships. In such cases, alternative methods like nonlinear regression or machine learning algorithms may be more appropriate. While LSM offers simplicity, interpretability, and efficiency, its effectiveness is contingent on meeting certain assumptions. Care should be taken to assess the linearity of relationships, handle outliers, and consider alternative methods when faced with nonlinearities or other violations of assumptions.

In conclusion, the LSM stands as a venerable and widely used tool in statistical modeling, appreciated for its simplicity and effectiveness. Its advantages, including its robustness to noise and straightforward interpretation of results, make it a valuable asset in various scientific and engineering applications. However, its sensitivity to outliers, assumption of linearity, and neglect of uncertainties in independent variables highlight the importance of careful consideration when applying the method. Researchers should be aware of these limitations

and, when necessary, explore alternative modeling approaches to better capture the complexities of real-world data.

2.2. Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a potent statistical method employed for parameter estimation, widely recognized for its versatility and applicability across diverse fields, including but not limited to finance, biology, and engineering. This comprehensive exploration aims to provide a detailed understanding of MLE, delving into its advantages, disadvantages, and the intricacies that shape its use in scientific research and data analysis.

Conceptual Framework:

At its core, Maximum Likelihood Estimation is a method for finding the parameter values of a statistical model that maximize the likelihood function, which represents the probability of observing the given data under the specified model. In other words, MLE seeks the parameter values that make the observed data most probable under the assumed model.

Advantages of the MLE:

- 1) Asymptotic Consistency: Under certain regularity conditions, MLEs are asymptotically consistent, meaning they converge to the true parameter value as the sample size grows towards infinity. This robustness means even if the distribution of data is not perfectly known, MLEs tend to get closer to the true value with more data [10].
- 2) Large-Sample Normality: Under specific conditions, MLEs become asymptotically normal, allowing for the construction of confidence intervals and hypothesis tests. This allows for rigorous statistical inference despite potential deviations from normality in the data [11].
- 3) Efficient Estimator under Certain Conditions: when specific regularity conditions are met, MLEs achieve the Cramer-Rao lower bound, meaning they have the smallest possible variance among all unbiased estimators. This translates to efficient estimates that extract the most information from the available data [12].
- 4) Asymptotic Efficiency: As the sample size increases, MLEs become asymptotically efficient compared to other estimators, meaning they approach the minimum achievable variance. This makes them particularly attractive for large datasets [10].
- 5) Allows for the incorporation of prior knowledge through likelihood functions: MLE can be integrated within the Bayesian framework by using prior information to construct the likelihood function. This allows for incorporating expert knowledge or domain-specific constraints, potentially leading to more accurate estimates [13].
- 6) Flexible Likelihood Design: The likelihood function can be customized to incorporate specific knowledge about the data-generating process or the nature of the parameters, leading to more informative estimates [14].

It's important to remember that MLE has limitations. Here's a closer look at its disadvantages:

1) Requires knowledge of the distribution of errors:

Assumption Dependence: MLE relies on the assumption that the errors follow a specific probability distribution (e.g., normal, Poisson). If this assumption is incorrect, the estimates can be biased and unreliable [15].

2) Complex computation, especially for nonlinear models:

Optimization Challenges: Finding the maximum likelihood estimate often involves optimization algorithms, which can be computationally expensive and prone to finding local maxima instead of the true global maximum [16]. Intractability with Non-linearity: for complex, non-linear models, analytical solutions might not be available, making optimization even more challenging and requiring specialized algorithms [17].

3) Sensitivity to model misspecification: Assumption Dependence: MLE assumes a specific model for the relationship between variables. If this model is incorrect, the estimates can be biased and misleading, even if the error distribution is correctly specified [18].

4) Outlier Impact: MLE can be sensitive to outliers, which can significantly influence the estimates and distort the results [19].

While these are notable disadvantages, MLE remains a widely used and versatile method. Understanding its limitations is crucial for interpreting results and selecting appropriate estimation techniques. MLE stands as a cornerstone in statistical modeling, offering a principled and versatile approach to parameter estimation. Its advantages, including statistical robustness, efficiency under certain conditions, and flexibility in incorporating prior knowledge, make it a go-to method in various scientific disciplines. However, the method is not without challenges, with dependencies on correct distributional assumptions, computational complexity, and sensitivity to model misspecification.

The real-world application of MLE involves thoughtful model specification, the transformation of likelihood functions, optimization procedures, and subsequent inference. Through examples like modeling exam scores with a normal distribution, the practicality of MLE becomes evident. However, researchers must remain vigilant about potential pitfalls, conduct thorough model diagnostics, and be mindful of the assumptions underpinning the chosen statistical model.

As statistical techniques continue to evolve, Maximum Likelihood Estimation retains its significance, providing a robust and widely applicable framework for extracting meaningful insights from data. Its continued integration into diverse fields underscores its enduring impact on scientific research and data-driven decision-making.

2.3. Nonlinear Regression

Nonlinear regression stands as a powerful tool in unveiling the intricacies of ex-

perimental data and crafting accurate models of physical processes. Unlike its linear counterpart, it delves beyond straight-line relationships, venturing into the realm of complex, curvilinear dynamics that govern diverse phenomena. Understanding both the advantages and limitations of this technique is crucial for researchers navigating the intricate pathways of data analysis and model building.

Advantages: Capturing the Nuances of Reality

1) Flexibility for Complicated Relationships: One of the most compelling advantages of nonlinear regression lies in its inherent flexibility. Unlike linear models, it doesn't impose a restrictive, straight-line relationship between variables. Instead, it can accommodate a wide range of functional forms, such as exponential, logarithmic, sigmoidal, and power functions, allowing it to capture the nuances of intricate relationships often observed in real-world data. This flexibility is particularly valuable when dealing with phenomena like population growth, enzyme kinetics, or chemical reactions, where linear models would prove inadequate [20].

2) Improved Model Accuracy: By venturing beyond the limitations of linearity, nonlinear regression often leads to more accurate models that better fit the observed data. This enhanced accuracy translates to more reliable predictions and a deeper understanding of the underlying processes at play. For instance, in studies of biological systems, where feedback loops and complex interactions are commonplace, nonlinear models can provide superior insights compared to linear approaches [21].

3) Insights into Underlying Mechanisms: The functional forms employed in nonlinear regression models can sometimes offer valuable insights into the mechanisms driving the observed phenomenon. By analyzing the parameters estimated by the model, researchers can gain clues about the nature of the interactions between variables, potentially leading to a more comprehensive understanding of the system under study [22].

Disadvantages: Navigating the Challenges

1) Increased Complexity: The very flexibility that makes nonlinear regression powerful also presents a challenge. With a wider array of possible functional forms comes the burden of choosing the most appropriate model for the given data. This selection process can be complex, requiring careful consideration of theoretical knowledge, data characteristics, and statistical criteria [23].

2) Overfitting and Interpretability: Overfitting, where the model closely fits the training data but fails to generalize to unseen data, is a significant concern in nonlinear regression. The abundance of parameters in complex models can make them susceptible to overfitting, leading to unreliable predictions. Therefore, careful evaluation and techniques like regularization are crucial to ensure model generalizability [24].

3) Sensitivity to Data Quality and Outliers: Nonlinear regression models can be more sensitive to noise and outliers in the data compared to linear models.

The presence of outliers can significantly impact parameter estimates and model fit, necessitating careful data cleaning and outlier analysis before proceeding with the analysis [25].

4) Computational Demands: Finding the optimal parameters for nonlinear models often requires iterative optimization algorithms, which can be computationally expensive, especially for large datasets. This can limit the applicability of nonlinear regression in certain scenarios where computational resources are constrained [26].

Nonlinear regression offers a powerful tool for researchers analyzing complex data and building accurate models of physical processes. By understanding its advantages, such as flexibility, improved accuracy, and potential mechanistic insights, and being aware of its limitations, including increased complexity, overfitting, and sensitivity to data quality, researchers can leverage this technique effectively to extract valuable knowledge from their data. The ever-evolving field of nonlinear regression, with its advancements and resources, promises to continue empowering researchers in their pursuit of understanding the intricate relationships that govern the world around us.

2.4. Bayesian Inference

Bayesian methods are powerful for incorporating prior knowledge and uncertainty into statistical models, making them highly valuable for scientific research. They allow researchers to update their beliefs about parameters or hypotheses based on observed data, providing a principled framework for making probabilistic inferences. This statistical method updates the probability for a hypothesis as more evidence or data becomes available. It is useful when dealing with uncertainty and incorporating prior knowledge into the model.

Bayesian methods are versatile and can be applied to a wide range of scientific endeavors, including the measurement of physical processes. While the application of Bayesian methods to physical processes might not be as common as in some other fields, there are instances where these methods can provide valuable insights.

Detailed explanations of the advantages and disadvantages of Bayesian Inference, along with links to relevant resources:

Advantages:

1) Incorporation of prior knowledge: Bayesian methods allow researchers to incorporate prior knowledge or beliefs about parameters into their statistical models. This prior information can help constrain parameter estimates and improve the accuracy of inference [27].

2) Flexible handling of Uncertainty: Bayesian inference provides a flexible framework for quantifying and handling uncertainty. By representing uncertainty using probability distributions.

3) Bayesian models can capture complex sources of uncertainty and provide probabilistic estimates of model parameters [28].

4) Sequential updating of beliefs: Bayesian inference allows for sequential updating of beliefs as new data becomes available. This sequential updating process, known as Bayesian updating, enables researchers to iteratively refine their estimates and incorporate new evidence into their models [29].

5) Model comparison and selection: Bayesian methods facilitate model comparison and selection by quantifying the evidence in favor of different models. Techniques such as Bayes Factors and Deviance Information Criterion allow researchers to compare the fit of competing models and identify the most plausible model given the data [30].

Disadvantages:

1) Computational complexity: Bayesian inference can be computationally demanding, particularly for complex models or large datasets. Markov Chain Monte Carlo (MCMC) methods, commonly used for Bayesian inference, may require extensive computational resources and time [31].

2) Subjectivity in prior specification: The choice of prior distributions in Bayesian analysis can influence the resulting posterior estimates. Subjective or poorly specified priors may lead to biased inference or misleading results, highlighting the importance of careful prior elicitation [32].

3) Interpretability of results: Bayesian models can sometimes be more complex and challenging to interpret compared to frequentist models. The interpretation of Bayesian posterior distributions and uncertainty intervals may require specialized knowledge and expertise [33].

4) Sensitivity to model assumptions: Like any statistical method, Bayesian inference relies on certain assumptions about the underlying data-generating process. Violations of these assumptions or misspecification of the model can lead to biased or unreliable inference [34].

Bayesian inference remains a powerful and versatile tool for statistical modeling and inference, offering unique advantages in handling uncertainty and incorporating prior knowledge. However, researchers should be aware of the computational challenges, subjective nature of priors, and potential pitfalls associated with Bayesian analysis.

2.5. Monte Carlo Simulation

Monte Carlo methods involve generating random samples from probability distributions to estimate unknown quantities or simulate complex systems. They are particularly useful for modeling uncertainty and conducting sensitivity analyses in scientific research. Involves generating random samples to simulate various outcomes of a model. This method is particularly useful when dealing with systems with inherent randomness or when uncertainty in parameters needs to be considered.

Detailed explanations of the advantages and disadvantages of Monte Carlo Simulation, along with links to relevant resources:

Advantages:

1) Flexibility in model complexity: Monte Carlo simulation allows for the

modeling of complex systems with multiple interacting components and non-linear relationships. It can handle models with arbitrary complexity, making it suitable for a wide range of scientific applications [35].

2) Incorporation of Uncertainty: Monte Carlo simulation is well-suited for incorporating uncertainty into models by sampling from probability distributions of uncertain parameters. This allows researchers to quantify and propagate uncertainty through the model, providing probabilistic estimates of model outputs [36].

3) Sensitivity Analysis: Monte Carlo simulation enables sensitivity analysis by systematically varying input parameters and observing the resulting changes in model outputs. This helps identify critical parameters that have the greatest impact on model predictions and assess the robustness of the model [37].

4) Estimation of Rare Events: Monte Carlo simulation can accurately estimate the probabilities of rare or extreme events by generating a large number of samples from the distribution of interest. This makes it valuable for risk assessment and reliability analysis in engineering, finance, and other fields [38].

Disadvantages:

1) Computational intensity: Monte Carlo simulation can be computationally intensive, particularly for models with a large number of parameters or complex simulation algorithms. Generating a sufficient number of samples to achieve reliable results may require significant computational resources and time [39].

2) Sampling Errors: Monte Carlo simulation results are subject to sampling errors, especially when using a finite number of samples to estimate model outputs. As a result, the accuracy of Monte Carlo estimates depends on the number of samples generated and the convergence properties of the simulation algorithm [40].

3) Difficulties in Convergence: Convergence can be a challenge in Monte Carlo simulation, particularly for models with complex or high-dimensional parameter spaces. Assessing convergence and determining when to stop the simulation may require careful monitoring and diagnostics [41].

4) Difficulty in Model Specification: Monte Carlo simulation requires specifying a probabilistic model for the system of interest, including probability distributions for uncertain parameters and relationships between variables. Model specification errors or misspecification can lead to biased or unreliable simulation results [42].

Monte Carlo simulation remains a powerful and versatile tool for modeling uncertainty and conducting sensitivity analyses in scientific research. However, researchers should be aware of the computational challenges, sampling errors, convergence issues, and difficulties in model specification associated with Monte Carlo simulation.

2.6. Optimization Techniques

Optimization methods such as gradient descent, genetic algorithms, and simu-

lated annealing are essential for finding optimal model parameters and minimizing errors when fitting models to experimental data. They are widely used in parameter estimation and model calibration tasks.

Detailed explanations of the advantages and disadvantages of Optimization Techniques, along with links to relevant resources:

Advantages:

1) Efficient parameter estimation: optimization techniques allow for efficient estimation of model parameters by searching for the values that minimize or maximize a given objective function. This enables researchers to find optimal model fits to experimental data and improve the accuracy of their models [43].

2) Global search capability: certain optimization methods, such as genetic algorithms, are capable of performing global search across a large parameter space. This helps avoid local optima and ensures that the optimization process converges to a globally optimal solution [44].

3) Robustness to noise: Optimization techniques are often robust to noise in the objective function or parameter estimates. They can handle noisy or imperfect data and still converge to reasonable solutions, making them suitable for real-world applications where data quality may be less than ideal [45].

4) Versatility and adaptability: Optimization methods can be adapted to a wide range of optimization problems and objective functions. They are versatile tools that can handle various types of constraints and optimization objectives, making them suitable for diverse applications [46].

Disadvantages:

1) Sensitivity to initial conditions: some optimization techniques, such as gradient descent, are sensitive to the choice of initial conditions. Poor initial guesses may lead to convergence to suboptimal solutions or even divergence from the optimal solution [47].

2) Computational complexity: certain optimization methods, especially those that involve evaluating the objective function multiple times, can be computationally intensive. This may pose challenges when dealing with large-scale optimization problems or when real-time performance is required [48].

3) Local optima traps: optimization techniques that rely on local search, such as gradient descent, may get trapped in local optima and fail to find the global optimum. This can limit the effectiveness of the optimization process, particularly for non-convex and multimodal objective functions [49].

4) Difficulty in tuning parameters: some optimization methods require careful tuning of hyperparameters or algorithmic parameters to achieve optimal performance. Selecting appropriate parameter values may require expertise and experimentation, adding complexity to the optimization process [50].

Optimization techniques remain indispensable tools for parameter estimation and model calibration tasks in scientific research. However, researchers should be mindful of their sensitivity to initial conditions, computational complexity, susceptibility to local optima traps, and the challenges associated with tuning

algorithmic parameters.

2.7. Machine Learning Algorithms

Machine learning techniques like neural networks, support vector machines, and decision trees are increasingly used for data modeling and analysis. They offer powerful tools for capturing complex patterns in experimental data and making predictions based on probabilistic reasoning. Techniques like neural networks, support vector machines, or decision trees can be applied for complex data patterns. Machine learning is valuable when the relationship between variables is intricate and not easily captured by traditional methods.

By combining these mathematical methods judiciously, researchers can effectively analyze experimental data, extract meaningful insights, and formulate accurate models of physical processes. The choice of method depends on the nature of the data and the characteristics of the underlying physical phenomenon.

Detailed explanations of the advantages and disadvantages of Machine Learning Algorithms, along with links to relevant resources:

Advantages:

1) Ability to capture complex patterns: machine learning algorithms excel at capturing complex patterns in experimental data, including nonlinear relationships and interactions between variables. Techniques like neural networks, support vector machines, and decision trees can learn intricate patterns from data, making them valuable for modeling complex phenomena [51].

2) Flexibility and adaptability: machine learning algorithms are highly flexible and adaptable to various types of data and modeling tasks. They can handle diverse data formats, including structured and unstructured data, and are suitable for regression, classification, and clustering tasks [52].

3) Scalability to large datasets: many machine learning algorithms are scalable to large datasets, allowing researchers to analyze massive amounts of experimental data efficiently. Techniques such as deep learning, in particular, have been shown to perform well on large-scale data analysis tasks [53].

4) Automated feature engineering: machine learning algorithms can automatically extract relevant features from raw data, eliminating the need for manual feature engineering. This can save time and effort in the modeling process and may lead to more robust models [54].

Disadvantages:

1) Black-box nature: many machine learning algorithms, particularly deep learning models, are often viewed as black boxes due to their complex internal workings. This lack of interpretability can make it challenging to understand how predictions are made and may hinder the adoption of machine learning models in some domains [55].

2) Data requirements: machine learning algorithms typically require large amounts of labeled data for training, which may be costly or time-consuming to collect. Insufficient or biased training data can lead to poor model performance and generalization errors [56].

3) Overfitting: overfitting occurs when a machine learning model learns to capture noise or irrelevant patterns in the training data, leading to poor generalization to new, unseen data. Regularization techniques and careful model selection are needed to mitigate the risk of overfitting [57].

4) Hyperparameter tuning: machine learning algorithms often require tuning of hyperparameters to achieve optimal performance. Selecting the right combination of hyperparameters can be challenging and may require extensive experimentation [58].

Machine learning algorithms offer powerful tools for analyzing experimental data and formulating models of physical processes. However, researchers should be aware of the challenges associated with their black-box nature, data requirements, risk of overfitting, and the need for hyperparameter tuning.

2.8. Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique that identifies the most important features or patterns in high-dimensional data. It is valuable for simplifying complex datasets and identifying underlying trends or relationships in experimental data. PCA reduces the dimensionality of the data by identifying the principal components that capture the maximum variance. This technique is beneficial for simplifying complex datasets and focusing on essential information.

Detailed explanations of the advantages and disadvantages of Principal Component Analysis (PCA), along with links to relevant resources:

Advantages:

1) Dimensionality reduction: PCA effectively reduces the dimensionality of high-dimensional data by identifying a smaller number of principal components that capture the most variance in the dataset. This simplification makes it easier to visualize and interpret the data while retaining most of the important information [59].

2) Feature extraction: PCA extracts meaningful features or patterns from the original data, allowing researchers to focus on the most relevant information. By representing data in terms of principal components, PCA can reveal underlying trends, relationships, or clusters in the data that may not be apparent in the original high-dimensional space [60].

3) Noise reduction: PCA can help mitigate the effects of noise in the data by filtering out components with low variance. By focusing on the principal components that capture the most variance, PCA enhances signal-to-noise ratio and improves the quality of data analysis and interpretation [61].

4) Visualization: PCA facilitates data visualization by projecting high-dimensional data onto a lower-dimensional subspace spanned by the principal components. This allows researchers to visualize the structure of the data in a more interpretable and insightful manner, aiding in data exploration and understanding [62].

Disadvantages:

1) Linearity assumption: PCA assumes that the underlying relationships in the

data are linear, which may not always hold true for complex datasets with non-linear relationships. In such cases, PCA may not effectively capture the underlying structure of the data [63].

2) Loss of interpretability: While PCA simplifies the data by reducing dimensionality, the resulting principal components may not always be directly interpretable in terms of the original features. This loss of interpretability can make it challenging to relate the principal components back to the original variables [64].

3) Variance bias: PCA prioritizes components that capture the most variance in the data, which may not always align with the most meaningful or informative features. This variance bias can lead to suboptimal representation of the data and may overlook important but low-variance features [65].

4) Data scaling sensitivity: PCA is sensitive to the scale of the original variables, and features with larger scales may dominate the principal components. Proper scaling of the data is necessary to ensure that PCA captures the true underlying structure of the data [66].

Principal Component Analysis (PCA) offers valuable insights into high-dimensional data by reducing dimensionality and extracting meaningful features. However, researchers should be mindful of its assumptions, loss of interpretability, variance bias, and sensitivity to data scaling.

2.9. Gaussian Processes

Gaussian processes are a Bayesian non-parametric approach for modeling complex data distributions. They are particularly useful for modeling uncertainty in regression tasks and making probabilistic predictions with uncertainty estimates.

Detailed explanations of the advantages and disadvantages of Gaussian Processes (GPs), along with links to relevant resources:

Advantages:

1) Flexibility: Gaussian Processes offer flexibility in modeling complex data distributions without assuming a specific parametric form. They can capture intricate patterns and nonlinear relationships in data, making them suitable for a wide range of regression and classification tasks [67].

2) Uncertainty estimation: GPs provide probabilistic predictions with uncertainty estimates, allowing researchers to quantify the uncertainty associated with predictions. This is particularly valuable in decision-making scenarios where understanding prediction uncertainty is crucial [68].

3) Robustness to noise: Gaussian Processes are robust to noise in the data and can effectively model noisy observations. By capturing the underlying trends in the data, GPs can filter out noise and provide more accurate predictions compared to deterministic models [69].

4) Adaptability to small datasets: GPs perform well even with small datasets, making them suitable for scenarios where data availability is limited. They can provide reliable predictions and uncertainty estimates even when trained on a limited amount of data [70].

Disadvantages:

1) Computational Complexity: Gaussian Processes can be computationally expensive, especially for large datasets or high-dimensional input spaces. Inference and prediction with GPs involve matrix computations that scale cubically with the number of data points, limiting their scalability [71].

2) Limited scalability: GPs may struggle to scale to large datasets due to their computational complexity. Approximation techniques such as sparse GPs or stochastic variational inference can help mitigate this limitation but may sacrifice some accuracy [72].

3) Choice of kernel function: The performance of Gaussian Processes is highly dependent on the choice of kernel function, which determines the characteristics of the covariance structure. Selecting an appropriate kernel function requires domain knowledge and experimentation [73].

4) Interpretability: Gaussian Processes are often viewed as black-box models, making them less interpretable compared to simpler regression techniques. Understanding the relationship between input variables and predictions may be challenging, especially for complex kernel functions [74].

Gaussian Processes offer powerful capabilities for modeling uncertainty and making probabilistic predictions, but researchers should be mindful of their computational complexity, scalability limitations, kernel function selection, and interpretability challenges.

2.10. Information Criteria (AIC, BIC)

Information criteria, such as the Akaike Information Criterion (AIC) is statistical measures used for model selection and comparison. It provides a quantitative framework for evaluating the trade-off between model complexity and goodness of fit to the data.

AIC is based on information theory and is derived from the Kullback-Leibler divergence, which measures the discrepancy between the true underlying model and the model being evaluated.

AIC balances the goodness of fit of the model (how well it explains the observed data) with the complexity of the model (the number of parameters). It penalizes complex models more heavily to prevent overfitting. A lower AIC value indicates a better balance between model fit and complexity, so models with lower AIC values are preferred. AIC provides quantitative measures for comparing models by considering their fit to the data and their complexity. It helps researchers select the most appropriate model that explains the data well while avoiding overly complex models that may overfit the data.

Detailed explanations of the advantages and disadvantages of the AIC Information Criterion, as well as links to related resources:

Advantages:

1) Model selection: Information criteria, such as AIC and BIC, provide a systematic and quantitative approach to model selection. They help researchers

compare competing models and select the one that strikes the best balance between goodness of fit and model complexity [75].

2) Balancing complexity and fit: AIC and BIC penalize models for complexity, encouraging the selection of simpler models that explain the data well. This helps prevent overfitting and ensures that the selected model is not overly complex given the available data [76].

3) General applicability: Information criteria are applicable across various statistical models and techniques, making them widely applicable in different domains of research. They can be used for model selection in regression, time series analysis, machine learning, and other fields [77].

4) Incorporation of uncertainty: AIC and BIC implicitly account for uncertainty in model selection by considering both the goodness of fit and the number of parameters. This allows researchers to make informed decisions while acknowledging the inherent uncertainty in modeling [78].

Disadvantages:

1) Assumption of large sample sizes: AIC and BIC are derived under the assumption of large sample sizes, which may not always hold true in practice. For small sample sizes, the performance of information criteria may be suboptimal, leading to unreliable model selection [79].

2) Sensitivity to model misspecification: Information criteria are sensitive to model misspecification, and their performance may degrade if the underlying assumptions of the models are violated. Researchers should carefully assess model adequacy before relying solely on information criteria for model selection [80].

3) Inability to capture complexity beyond parameters: AIC and BIC penalize models based solely on the number of parameters, which may not fully capture the complexity of the model. Models with complex structures or interactions may be penalized less than warranted by their true complexity [81].

4) Subjectivity in penalty functions: the penalty functions used in AIC and BIC are subjective and may not always reflect the true trade-off between model complexity and goodness of fit. Different penalty functions may lead to different model selection outcomes [82].

Information criteria, such as AIC, offers valuable tools for model selection and comparison by balancing model complexity and goodness of fit. However, researchers should be aware of its limitations, including sensitivity to sample size, model misspecification, inability to capture complexity beyond parameters, and subjectivity in penalty functions.

2.11. Limitations of Existing Model Selection Methods

Current methodologies for selecting models of physical phenomena and analyzing experimental data face several key limitations. These limitations stem from their reliance on researcher subjectivity:

1) Inherent subjectivity: Traditional methods depend heavily on researchers'

knowledge, intuition, and experience. This can introduce biases and inconsistencies in model selection, hindering the pursuit of truly objective solutions.

2) Limited scope: Many existing approaches prioritize identifying the “best” model based on specific criteria. However, a broader understanding of the model landscape is crucial. This includes exploring alternative explanations and potential shortcomings inherent in any chosen model.

3) Furthermore, these methods often focus on minimizing overfitting. Overfitting occurs when complex models closely mimic training data but fail to generalize accurately to new data. Techniques like regularization and cross-validation help address this limitation.

These limitations are intrinsic to current model selection methods and can influence the chosen model’s reliability and validity. Careful consideration of these limitations and implementation of appropriate mitigation strategies are crucial for robust scientific inference.

Modern scientific literature generally overlooks an additional, fundamental limitation of existing methods. These methods analyze experimental data using models designed to minimize known uncertainties. However, they neglect a critical source of uncertainty: the model itself. This uncertainty arises from the qualitative and quantitative set of variables chosen to construct the model. Existing methods fail to account for this inherent and primary uncertainty, which ultimately influences the choice of the most suitable model for a specific phenomenon.

Subsequent chapters introduce a novel method based on the concept of “finite information quantity.” This approach allows us to define a “comparative uncertainty” criterion, facilitating the selection of the optimal model for studying a given physical phenomenon or technological process.

3. Finite Information Quantity (FIQ) Approach

We need to take a closer look at what it means to formulate a mathematical model of the physical or technological process being studied.

The model does not exist in empty space. Its components (variables) are selected by scientists and engineers from any system of units, for example, the International System called SI, the Gaussian System, etc. Any system of units is a finite Abelian group (group theory) [83], the number of elements of which can be calculated.

It can be assumed that there must be some objective criteria by which it is possible to judge the preferred system of units and decide which one to choose. However, it can be shown that the subsequent results presented are realized for any known system of units [84].

The “finite information quantity approach” (FIQ approach) is a relatively new approach in physics that explores the link between information and physical systems. While the concept can be applied in various contexts, its core idea revolves around two key points:

1) Information is not infinite: Unlike classical physics where information is treated as an infinitely divisible quantity, the FIQ approach acknowledges limitations on the information that can be stored or processed within a physical system. This limitation might arise from fundamental aspects of nature like the laws of thermodynamics or quantum mechanics.

2) This finiteness has physical implications: By imposing a limit on the available information, the FIQ approach seeks to understand how this constraint affects various physical phenomena. For example, it can be used to:

a) Estimate the maximum information storage capacity of black holes. This challenges the traditional Bekenstein bound, suggesting that black holes might have a finite information limit even though they can seemingly store infinite entropy.

b) Predict intrinsic uncertainties in physical measurements: by considering the finite information available during measurement, the FIQ approach can predict an inherent uncertainty associated with the measured value, independent of statistical errors.

Traditional model selection methods often rely on criteria like goodness-of-fit, which might not fully capture the limitations of information processing in the physical world. The FIQ approach proposes alternative criteria based on the finite information available to the model.

While the Finite Information Quantity (FIQ) approach remains under development, its potential is actively explored by leading physicists. Pioneered by Del Santo and Gisin [85]. Prof. Wojciech H. Zurek has explored the limitations of complex systems theory in the face of finite information [86], while Prof. Rafael Sorkin has investigated the information loss paradox in black holes from the FIQ perspective [87]. Dr. Donald Marolf has proposed a novel information-theoretic approach to black hole entropy informed by FIQ [88].

FIQ's broader implications for the nature of reality are being investigated by Prof. Max Tegmark, who explores the possibility of a "computational universe" with finite information [89]. Similarly, Dr. Antony Valentini has proposed interpretations of quantum mechanics that incorporate finite information [90]. Dr. Fay Dowker has delved into the connection between information and the geometry of spacetime using insights from FIQ [91].

Despite being under development, FIQ offers a groundbreaking perspective by bridging information theory and physical laws, potentially leading to groundbreaking discoveries across various physics subfields.

3.1. Initial Uncertainties in Model Building

This chapter addresses the concept of initial uncertainties in model building, a topic lacking sufficient exploration in current scientific literature. Traditionally, validation and verification methods focus on uncertainties arising from chosen variables, model structure, testing procedures, and data scatter. However, the crucial role of the initial qualitative and quantitative set of variables and the un-

derlying system of units in introducing inherent uncertainties remains largely unacknowledged.

This work highlights the finite information quantity (FIQ) method ([92]), which posits that uncertainty of perception is inherent to the observer's mind, leading to a "blurring" of the object being modeled. This philosophical stance, often overlooked in the scientific community, contrasts with established uncertainties stemming from limitations in:

- Measurement accuracy;
- Observability of the object;
- Measurement-induced perturbations;
- Quantum mechanical interpretations.

The author proposes the inclusion of perceptual uncertainty as a fifth fundamental uncertainty in model building. The FIQ method rests upon five key axioms [92]:

1) Choice of system of units: the observer selects from standardized systems like SI, CGS, or Planck units, influencing the model's group of phenomena (GoP). GoP defines the specific physical processes described by the model's variables and characterizes relevant features of the material object. For instance, an electric arc model typically utilizes variables with dimensions involving length, mass, time, current, and temperature, belonging to the $\text{GoP}_{\text{SI}} = \text{LMTI}$ class.

2) Observer bias and variable selection: each observer, informed by their unique perspective, selects a set of qualitative and quantitative variables to represent the observed phenomenon. This selection process aims to minimize distortions and subjective biases inherent in their individual viewpoint.

3) Finite information quantities (FIQ): FIQs are variables within the FIQ framework which include time, universal constants, one-dimensional components of position or momentum, and dimensionless numbers. Their values are drawn from the set of real numbers, \mathbb{R} [85].

4) Finiteness of information: the model contains a finite amount of information due to the limited number of variables and the inherent information limitation within each variable [85] [93].

5) Equiprobable variable selection: given a chosen system of units, if no prior information about the phenomenon exists, all variables possess an equal probability of inclusion in the model. Any variable is chosen by a conscious observer based on their background knowledge and research goals. This selection process inherently introduces a level of subjectivity. If a system of units is chosen without prior knowledge of the phenomenon under study, we can estimate the probability of including a relevant variable in the model to be equiprobable (equal for all possibilities). This essentially means that, in the absence of any guiding information, any variable within the chosen system of units has an equal chance of being included in the model. To illustrate this, consider the well-known case of an electron. Depending on the experimental setup, an electron can exhibit wave-like or particle-like behavior. If we don't know beforehand whether

we're studying the wave nature or the particle nature of the electron, selecting variables like wavelength or momentum becomes a matter of educated guesswork within the chosen unit system. Therefore, both versions have a right to exist before further experiments reveal the dominant characteristic.

The first three axioms align with common scientific practices. The fourth axiom resonates with the growing application of information theory across various disciplines. However, the fifth axiom, concerning equiprobable variable selection, is likely to spark discussion. As an illustration, the historical debate surrounding the wave-particle duality of the electron exemplifies how researchers, guided by intuition and existing knowledge, can propose radically different models for the same phenomenon, both potentially valid and experimentally supported.

3.2. Model as an Information Channel

The thesis that a model serves as a channel of information between the object or phenomenon being studied and the observer represents a deep and insightful view of the nature of models and their role in scientific research.

In [84] [94] [95] [96] the nature of models and their relationship to the objects or phenomena they represent is explored. Models are not simply passive representations of reality, but rather active tools that shape and mediate our understanding of the world. Models are not just descriptive but also generative because they allow us to make predictions, test hypotheses, and explore the consequences of different scenarios. In other words, models are not simply tools for representing or manipulating data, but also for communicating and translating information between the world and the mind.

One way to think about this is in terms of the concept of “mediation.” According to this view, models serve as intermediaries between the object or phenomenon being studied and the observer, facilitating the flow of information and understanding in both directions. On the one hand, models help the observer understand the world by providing a structured and simplified representation of complex phenomena. On the other hand, models also allow the observer to communicate his ideas and conclusions to the world by providing a common language and framework for describing and explaining his observations.

Another way to understand this thesis is to use the concept of “abstraction.” According to this view, models are abstract representations of reality that capture the essential features of a phenomenon while leaving out unnecessary details. By abstracting away the complexity and noise of the world, models allow us to focus on the underlying patterns and structures that govern a particular domain. In this sense, models serve as a kind of “filter” or “lens” that allows us to see the world in a new and more honest way.

Regardless of how one interprets this thesis, it is clear that it has important implications for the way we think about models and their role in scientific research. By emphasizing the active and generative nature of models, this ap-

proach encourages us to think of models not simply as passive representations of reality, but as active tools for shaping and mediating our understanding of the world. Whether we are studying natural phenomena, social systems, or technological artifacts, the ideas and perspectives offered can help us develop more powerful, more efficient, and more insightful models that can deepen our understanding and improve our ability to predict and explain. And control the world around us.

The unique aspect here is that any system of units, like the International System of Units (SI), relies on variables. These variables can include scalar parameters like time, universal constants, one-dimensional components of position or momentum, dimensionless numbers. These variables take values from the set of real numbers [85]. Each variable (q) carries a finite amount of information, and this information has an upper bound [85] [93]. Consequently, these variables are termed “finite information quantities” (FIQs) [85]. The number of dimensionless FIQs based on the SI can be calculated as $\mu_{\text{SI}} = 38,265$ [95]. The subsequent reasoning and formulas hold true for models incorporating any FIQs, dimensional or dimensionless. These formulas are independent of the specific system of units employed in the model [94] [95] [96] [97].

When constructing a model, the observer must make a deliberate decision to select only a few quantities, thereby defining one or more groups of phenomena (GoPs) for the model. GoP is a set of physical phenomena and processes that can be described by a finite number of basic quantities and derived variables from any system of units, for example, SI. These characteristics help determine the characteristics of a material object, both qualitatively and quantitatively [98].

For example, when modeling an electric arc, variables commonly used include the basic SI quantities of length (L), mass (M), time (T), current (I), and thermodynamic temperature (Θ). This means that the model falls into the category of phenomena $\text{GoP}_{\text{SI}} \equiv \text{LMT}\Theta\text{I}$. At this stage (GoP selection), the number of variables is significantly reduced compared to μ_{SI} . However, due to practical constraints such as limited time, financial resources and computational capabilities, the researcher ends up selecting only a very small number of variables for the final model compared to μ_{SI} .

In this context, we consider two separate sets of random variables with equal probability: $X \in \{x_1, \dots, x_j\}$ represents the total number of FIQs in the observed physical system, excluding any hidden variables. $Y \in \{y_1, \dots, y_p\}$ represents the number of output FIQs reflected in the model, chosen by the observer. Set Y is essentially a “noisy” version of set X , where the observed phenomenon is compressed. This means the number of variables is significantly reduced, but without any energy expenditure. The observer simply focuses on the model without disturbing the actual process.

Given that μ_{SI} is constant, each FIQ has a limited information content, and the number of FIQs in a model is always finite, we can infer that the total amount of information contained within both the SI and the model is inherently limited.

The key idea of the principle of choosing the best model to describe the phenomenon under study is that when constructing a model of a physical phenomenon, we (observers) determine a set of variables. These include Ψ - dimensional quantities: they represent measurable properties with units of measurement (for example, length, time); ξ base quantities: these are fundamental units included in one or another system of units from which others are formed (for example, meter, second).

The modeler/thinker/observer then selects a specific dimensionless quantity of interest (u) whose values fall within a certain range (S). Importantly, the modeling approach is non-invasive and introduces no perturbations to the system under study. In addition, the researcher indicates the type of phenomenon being modeled, characterized by the total number of FIQs in the selected GoP (z'), β is the number of base quantities in the selected GoP, z'' is the number of FIQs recorded in a model, and β' is the number of independent base quantities recorded in a model. In other words, the absolute uncertainty (Δ_Σ) when measuring a selected quantity can be determined using the following Equation (1) [94] [99] [100] [101] [102] [103]:

$$\Delta_\Sigma = S \cdot \left[(z' - \beta') / \mu_{SI} + (z'' - \beta'') / (z' - \beta') \right] \quad (1)$$

In the Equation (1), Δ_Σ represents the a priori total model uncertainty, which arises from the selection of the GoP and the number of recorded FIQs. At its heart, Δ_Σ represents an inherent, fundamental uncertainty in any physical-mathematical model. This uncertainty exists before any measurements are made and is independent of the measurement process itself. It arises solely from the number of variables chosen and the selected GoP. Consequently, the total uncertainty of the model, which includes additional uncertainties from the model's structure and computerization, will be significantly larger than Δ_{pmm} . In essence, Equation (1) can be seen as the uncertainty principle for model development. It implies that any change in the level of detail used to describe the observed object ($z'' - \beta''$; $z' - \beta$) will cause a shift in both Δ_{pmm} and the accuracy of the key variables representing the object's internal properties.

The term ε is the comparative uncertainty of the model, defined as $\varepsilon = \Delta_\Sigma / S$. Despite its significance in information theory [104], the value of ε has often been overlooked by researchers.

Table 1 summarizes the optimal values of ε_{opt} for different GoPs and recommends the corresponding number of FIQs required to achieve them including the optimal number of FIQs inherent in a model, $\gamma_{mod} = z'' - \beta'$.

The concept of “amount of information” as a physical quantity introduces a unique perspective on model uncertainty. This perspective reveals an inherent uncertainty arising from the researcher's worldview, which cannot be captured by traditional statistical methods, weighting factors, or consistency criteria. These tools are limited because they operate on the results of experiments and simulations for existing models. This distinction highlights the difference between treating information as a physical entity and the well-established theory of measurements [94].

Table 1. Comparison of measurement uncertainties and optimal dimensionless parameters.

GoP _{SI}	Comparative uncertainty, ε_{opt}	Number of FIQs inherent in GoP _{SI} , $\gamma_{\text{GoP}} = z' - \beta'$	Optimal number of FIQs inherent in a model, $\gamma_{\text{mod}} = z'' - \beta''$
<i>LMT</i>	0.0048	91	$\approx 0.2 < 1$
<i>LMTF</i>	0.0146	279	≈ 2
<i>LMTI</i>	0.0245	468	≈ 6
<i>LMTθ</i>	0.0442	846	≈ 19
<i>LMTIF</i>	0.0738	1412	≈ 52
<i>LMTθF</i>	0.1331	2546	≈ 169
<i>LMTθI</i>	0.2220	4247	≈ 471
<i>LMTθFI</i>	0.6665	12 751	≈ 4249

While, traditionally, models with more variables are favored, our approach prioritizes informativeness. Here, models with an FIQ count closer to γ_{GoP} are considered more informative. This information-centric approach allows us to identify the most suitable model for the object under study, and consequently, the optimal method for calculating its relevant researched variable.

4. FIQs for Model Selection: A New Approach to Understanding Uncertainty

The FIQ-based approach recommends analyzing scientific research results by comparing the achieved model uncertainty (ε_{mod}) with the theoretically optimal uncertainty (ε_{opt}) as shown in **Table 1**. The ratio $\varepsilon_{\text{mod}}/\varepsilon_{\text{opt}}$ serves as an objective criterion for assessing a model's acceptability, measurement method effectiveness, and accuracy when comparing models for a specific physical phenomenon or technological process. A ratio close to 1 ($1 > \varepsilon_{\text{mod}}/\varepsilon_{\text{opt}} \approx 1$) indicates the model's suitability for describing the studied process. Conversely, a large difference suggests the model's limitations. It's important to note that reaching the theoretical limit (ε_{opt}) might not be achievable (ε_{mod} is always less than ε_{opt}) due to inherent constraints. The following analysis will explore the challenges that need to be addressed to optimize model development.

4.1. Measuring Physical Constants with Improved Uncertainty Analysis

This chapter explores the challenges associated with achieving optimal accuracy in measuring fundamental physical constants. The recent adoption of the International System of Units (SI) by CODATA (Committee on Data for Science and Technology) signifies significant progress in this domain [105]. However, limitations remain in the current methodology for analyzing measurement uncertainties.

Limitations of CODATA Methodology: the established CODATA methodology relies on Bayesian Linear Regression with least squares adjustments (LSA) to harmonize data from various research centers [106]. This approach, while ensuring consistency, can introduce subjective bias, particularly when reconciling conflicting results. Additionally, concerns exist regarding the potential influence of personal opinions on statistical analysis [107].

This chapter proposes the FIQ-based method as an alternative for uncertainty analysis. This method avoids subjectivity by focusing on the concept of “optimal uncertainty” (ϵ_{opt}) and “experimental comparative uncertainty” (ϵ_{exp}) [108]. ϵ_{opt} defines the inherent limitations of a measurement model, considering the chosen variables and phenomena. ϵ_{exp} , on the other hand, reflects the actual uncertainty achieved in a specific experiment.

The FIQ-based method offers several advantages. Firstly, it avoids subjective adjustments inherent in the CODATA approach. Secondly, it focuses on ϵ_{opt} , a fundamental limit on measurement accuracy. Finally, it emphasizes the importance of including a sufficient number of variables in the measurement model to minimize ϵ_{exp} .

In [109] a detailed analysis of various physical constant measurements through the FIQ-based method is presented. The analysis, based on data from 2000-2019, revealed a clear trend: models incorporating a larger number of base quantities (LMT θ , LMTI, etc.) and FIQs (γ_{mod}) generally achieved lower $\epsilon_{\text{exp}}/\epsilon_{\text{opt}}$ ratios (Table 2). This suggests that considering a broader range of variables leads to a more accurate understanding of the underlying phenomenon and reduces the discrepancy between the optimal and achieved uncertainty.

Table 2. Summary of $\epsilon_{\text{M}}/\epsilon_{\text{opt}}$ values.

Ratio $\epsilon_{\text{exp}}/\epsilon_{\text{opt}}$	Physical Constant	Measurement method	GoP
2.3	Boltzmann constant	DCGT ¹	LMT θ
3.6	Planck constant	AGT ²	LMT θ F
4.1	Hubble constant	CMB ³	LMT θ
7.9	Gravitational constant	Electro-mechanical methods	LMTI
15.9	Planck constant	KB ⁴	LMTI
32.6	Planck constant	XRCD ⁵	LMT θ F
100	Gravitational constant	Mechanical methods	LMT
104	Hubble constant	BAO ⁶	LMT
710	Boltzmann constant	BDL ⁷	LMT

¹DCGT—dielectric constant gas thermometer, ²AGT—acoustic gas, thermometer, ³CMB—cosmic microwave background, ⁴KB—Kibble balance, ⁵XRCD—X-ray crystal density, ⁶BAO—baryonic acoustic oscillations, ⁷BDL—brightness of distance ladder.

The analysis also highlights the limitations of certain measurement techniques for specific constants (**Table 2**). For instance, the data suggests that methods like BDL (brightness of distance ladder) for the Boltzmann constant, BAO (baryonic acoustic oscillations) for the Hubble constant, and mechanical methods for the gravitational constant, are less promising in terms of achieving optimal accuracy.

The FIQ-based method offers a valuable framework for analyzing measurement uncertainties with greater objectivity and accuracy. However, further efforts are needed to promote its adoption by the scientific community. This includes reformulating the concept of “comparative uncertainty” in terms of relative uncertainty ($r_{\text{exp}}/r_{\text{CoP}}$) which is more readily understood by a wider range of scientists (**Table 3**).

Several key trends emerge from analyzing the data in **Table 3**. First, the ratio of $r_{\text{exp}}/r_{\text{CoP}}$ increases significantly when using the GoP with a limited number of base quantities and a low γ_{mod} (complexity factor) like LMTF or LMT. This suggests that models incorporating a broader range of variables (higher γ_{mod}) and fundamental quantities (more than LMT) can potentially achieve lower comparative uncertainty (ε_{exp}) compared to their optimal uncertainty (ε_{opt}). Second, all $r_{\text{exp}}/r_{\text{CoP}}$ values are greater than 1. This supports the core tenet of the FIQ-based method [109]: the inherent limit on accuracy for any model (ε_{opt} , or r_{CoP}) is theoretically unattainable in practice. Third, comparing the DBT and JNT methods for measuring the Boltzmann constant, the JNT method offers potential for improved accuracy. This can be achieved by refining the experimental setup and incorporating additional relevant variables. Fourth, the data suggests that electro-mechanical methods for measuring the gravitational constant hold promise for achieving higher accuracy with greater confidence. Finally, within

Table 3. Comparison of achieved vs. optimal relative uncertainty.

Ratio $r_{\text{exp}}/r_{\text{CoP}}$	Physical constant	Measurement method	GoP
1.1	Boltzmann constant	DBT ¹	LMT θ F
1.9	Boltzmann constant	JNT ²	LMT θ I
1.9	Gravitational constant	Electro-mechanical methods	LMTI
2.4	Hubble constant	CMB	LMT θ
2.6	Planck constant	AGT	LMT θ F
2.9	Planck constant	KB	LMTI
9.1	Planck constant	XRCD	LMTF
12.7	Gravitational constant	Mechanical methods	LMT
44	Boltzmann constant	BDL	LMT
56	Hubble constant	BAO	LMT

¹DBT—Doppler broadening thermometer, ²JNT—Johnson noise thermometer.

the FIQ-based framework, the AGT method appears most promising for improving the accuracy of Planck constant measurements compared to KB and XRCD methods.

The FIQ-based method presents a promising alternative for uncertainty analysis in measuring fundamental physical constants. By emphasizing ϵ_{opt} and the importance of a comprehensive measurement model, this methodology has the potential to enhance the accuracy and objectivity of scientific inquiry in this critical domain.

4.2. Unveiling Accuracy in Underwater Electrical Discharges

This chapter critically analyzes research on underwater electrical discharge (UED) published between 2011 and 2021. A comprehensive search across various databases (IEEE Xplore, ScienceDirect, etc.) yielded 800 articles [110]. To ensure rigor, four selection criteria were applied:

- 1) Solid Theoretical Foundation: Articles required a well-defined mathematical model with theoretical UED calculations, providing a strong theoretical framework.
- 2) Experimental Validation: Inclusion of experiments and their results was essential for validating theoretical models with empirical data.
- 3) Theory-Experiment Comparison: Articles explicitly comparing theoretical calculations with experimental findings were prioritized to assess model accuracy.
- 4) Uncertainty Quantification: Studies calculating the total absolute or relative uncertainty in experiments were preferred. Ideally, the uncertainty (EU) should be lower than the discrepancy between theory (TD) and experiment (ED) to validate the model's practical applicability.

These criteria aimed to identify high-quality studies encompassing theoretical models, experimental data, theory-experiment comparisons, and uncertainty considerations.

The review identified valuable insights. Authors presented diverse experimental setups and emphasized the scientific significance of their findings. However, some concerning trends emerged:

- 1) Limited Theoretical-Experimental Comparison: While many acknowledged the importance of comparing results with other studies, some lacked theoretical data for comparison with their own experiments.
- 2) Incomplete Uncertainty Analysis: Although some highlighted the relevance of uncertainty for optimizing designs, most studies didn't thoroughly explain the calculation of relative uncertainty in their experiments. Notably, even the study by W. Yao *et al.* (2019) that addressed uncertainty did not detail the individual contributions of uncertainty sources.

These shortcomings suggest a potential gap in UED research. Many studies focus on theoretical models, experimental data, and achieving good agreement between them, overlooking the need for rigorous testing through comprehensive

theory-experiment comparisons and detailed uncertainty analysis. Minor discrepancies are sometimes acknowledged, but their importance is often underappreciated.

This critical review highlights the need for a more balanced approach in UED research, emphasizing the importance of rigorous comparisons between theory and experiment alongside detailed uncertainty quantification. Such a shift could lead to more robust and reliable UED models with greater practical applicability.

Despite the wealth of available publications, a strategic selection process was necessary. Following the spirit of the proverb “make do with what you have,” six key articles were meticulously chosen for in-depth analysis [111]-[116]. Results are introduced in **Table 4**.

Several key trends emerge from analyzing the data in **Table 4**. First, the ratios of $\varepsilon_i/\varepsilon_{\text{opt}i}$ (achieved vs. optimal uncertainty) suggest a potential bias towards models with fewer variables (GoP with low $\gamma_{\text{mod}i}$ and base quantities). Examples include $\varepsilon_1/\varepsilon_{\text{opt}1}$ (9.67), $\varepsilon_2/\varepsilon_{\text{opt}2}$ (1.37), and $\varepsilon_3/\varepsilon_{\text{opt}3}$ (1.2) all exceeding 1 in studies [113] [114]. This contradicts the core principle of the information method [109]: the inherent accuracy limit (ε_i) of any model should be lower than the optimal limit ($\varepsilon_{\text{opt}i}$). As a result, these models [113] [114] [115] [116] may require significant reformulation to achieve optimal accuracy.

Conversely, the ratios $\varepsilon_5/\varepsilon_{\text{opt}5}$ and $\varepsilon_6/\varepsilon_{\text{opt}6}$ support the models proposed in [111] [112]. These models appear more promising in representing underwater electrical discharges with a higher degree of accuracy.

Remarkable achievement in UDM [112]: the research presented in [112] stands out for its exceptional results, comparable to the achievements of NASA engineers [117]. This work utilized the GoP_{SI} framework (LMT θ F), which incorporates variables expressed as combinations of five fundamental quantities

Table 4. Prioritization of UED models based on uncertainty and interpretability.

Variable/ Reference	Chosen GoP _{SI} of the model	Number of FIQs inherent in GoP _{SI} , $\gamma_{\text{GoP}} = z' - \beta'$	Optimal number of dimensionless FIQs inherent in a model, $\gamma_{\text{mod}i} = z'' - \beta''$, $i = 1, 2, 3$	Number of dimensionless FIQs inherent in a formulated model*, $\gamma_{\text{exp}i} = z'' - \beta''$, $i = 1, 2, 3$	The achieved experimental comparative uncertainty of the model**, ε_i	The comparative uncertainty of the model, theoretically justified for the selected GoP, $\varepsilon_{\text{opt}i}$	Ratio of $\varepsilon_i/\varepsilon_{\text{opt}i}$
[113]	LMT	91	$\gamma_{\text{mod}1} \approx 0.2 < 1$	$\gamma_{\text{exp}1} \approx 4$	$\varepsilon_1 = 0.0464$	$\varepsilon_{\text{opt}1} = 0.0048$	≈ 9.67
[114]	LMTI	468	$\gamma_{\text{mod}2} \approx 6$	$\gamma_{\text{exp}2} \approx 10$	$\varepsilon_2 = 0.0336$	$\varepsilon_{\text{opt}2} = 0.0245$	≈ 1.37
[115]	LMTI	468	$\gamma_{\text{mod}3} \approx 6$	$\gamma_{\text{exp}3} \approx 8$	$\varepsilon_3 = 0.0293$	$\varepsilon_{\text{opt}3} = 0.0245$	≈ 1.20
[116]	LMTI θ	4247	$\gamma_{\text{mod}4} \approx 471$	$\gamma_{\text{exp}4} \approx 19$	$\varepsilon_4 = 0.0484$	$\varepsilon_{\text{opt}4} = 0.2220$	≈ 0.22
[111]	LMTI	468	$\gamma_{\text{mod}5} \approx 6$	$\gamma_{\text{exp}5} \approx 3$	$\varepsilon_5 = 0.0186$	$\varepsilon_{\text{opt}5} = 0.0245$	≈ 0.76
[112]	LMTI	468	$\gamma_{\text{mod}6} \approx 6$	$\gamma_{\text{exp}6} \approx 5$	$\varepsilon_6 = 0.0229$	$\varepsilon_{\text{opt}6} = 0.0245$	≈ 0.93

* While crucial for calculating comparative uncertainty (2), explicitly stating the number of variables considered in a model is not standard scientific practice. Additionally, some researchers neglect to define the variables used within their formulas. This necessitates independent calculation of the number of variables in the reviewed articles, potentially introducing inaccuracies in representing this critical aspect of the model. ** ε_i is calculated according to Equation (2).

(length, mass, time, temperature, and force) at varying degrees [98]. Notably, the model in [112] employs a significant number of variables (130) and achieves a $\varepsilon_{\text{mod}}/\varepsilon_{\text{opt}}$ ratio close to 0.9, demonstrating a close alignment between achieved and optimal uncertainty. This success highlights the importance of considering a broader range of variables, even if the researchers were unaware of the specific information method. The numerous successful Mars rover landings by NASA [117] provide a compelling real-world validation of such a comprehensive modeling approach.

Building on the progress made in prior studies [111]-[117], the information method emphasizes the critical role of incorporating a specific, well-defined number of variables in models. In the context of underwater electrical discharges, the model proposed in [112] appears most promising due to its inclusion of a variable count closer to the optimal values suggested by the information method. This approach encourages researchers to move beyond models with limited variables and strive for a more comprehensive representation of the phenomenon by incorporating a wider range of relevant factors.

4.3. Optimizing Speed of Sound Measurements with the FIQ-Based Approach

While numerous studies have explored sound speed measurement, this analysis focuses on three specific works investigating sound propagation in hydrogen chloride [118], various solids [119], and $\text{N}_2\text{-H}_2$ mixtures [120]. The sound speed data generated (Table 5) is evaluated using the FIQ-based information method [84].

Quantifying measurement quality through relative uncertainty (r) poses limitations in directly comparing the presented models' accuracy. This is because the studies assume reliable measurements based on agreement between model calculations, experimental data, and achieved relative uncertainty (EU). However, a

Table 5. Comparison of research results.

Variable/Reference	[116]	[118]	[117]
Chosen GoP_{SI} of the model	$\text{LMT}\theta$	$\text{LMT}\theta$	LMTIF
FIQs amount contained in a given GoP_{SI} , $\gamma_{\text{GoP}} = z' - \beta'$	846	846	1412
Amount of FIQs recommended for the model, $\gamma_{\text{mod}} = z'' - \beta''$, $\gamma_{\text{mod}i} = z'' - \beta''$, $i = 1, 2, 3$	$\gamma_{\text{mod}1} \approx 19$	$\gamma_{\text{mod}2} \approx 19$	$\gamma_{\text{mod}3} \approx 52$
Achieved experimental comparative uncertainty of the model, ε_i	$\varepsilon_1 = 0.0233$	$\varepsilon_2 = 0.0305$	$\varepsilon_3 = 0.0596$
Comparative uncertainty of the model, theoretically justified for the selected CoP, $\varepsilon_{\text{opt}i}$	$\varepsilon_{\text{opt}1} = 0.0442$	$\varepsilon_{\text{opt}2} = 0.0442$	$\varepsilon_{\text{opt}3} = 0.0738$
Ratio of $\varepsilon_i/\varepsilon_{\text{opt}i}$	≈ 0.53	≈ 0.69	≈ 0.8

crucial comparison is often missing contrasting the achieved EU with the discrepancy between theoretical calculations (TC) and experimental results (ER) [121] [122]. When $|TC - ER|$ falls within the margin of EU, the model's validity and applicability become questionable [123].

To address these limitations and guide model selection, we employ the concept of optimal uncertainty (ε) as outlined in Axiom 5 [92]. This approach assumes equiprobable consideration of model variables. However, researchers often rely on intuition and experience to select variables, potentially neglecting important factors influencing sound propagation.

The findings in [119] highlight the importance of considering a broader range of variables in the model. This approach not only deepens our understanding of the true sound speed value but also opens doors for further exploration of seemingly well-understood phenomena.

Analysis [124] reveals that the model in [119] exhibits the closest agreement between achieved uncertainty (ε) and optimal uncertainty (ε_{opt}) compared to the other two models [118] [120]: $\varepsilon_1/\varepsilon_{opt1} = 0.53 < \varepsilon_2/\varepsilon_{opt2} = 0.69 < \varepsilon_3/\varepsilon_{opt3} = 0.8$. This preference is further supported by the ratio of model complexity factors (γ) to their optimal values (γ_{mod}). The model in [119] incorporates a higher number of variables ($\gamma_3 = 18$) closer to the optimal value ($\gamma_{mod3} = 52$) compared to the models in [118] ($\gamma_1 = 1$, $\gamma_{mod1} = 19$) and [120] ($\gamma_2 = 4$, $\gamma_{mod2} = 19$).

5. Discussion

The burgeoning scientific literature, fueled by escalating research and development costs and a growing number of researchers, has raised concerns about the quality and reliability of published findings. Replication problems, fraudulent practices, and a lack of expertise in measurement theory and uncertainty analysis threaten the very foundation of scientific progress.

Our findings unveil a fundamental challenge in the realm of experimental data processing: the information-theoretic bottleneck imposed by model complexity. Traditional methods often lack the sophistication to capture the intricate details encoded within a model during its construction. This limitation stems from the absence of tools to quantify and address model complexity itself.

Complexity, a well-established concept rooted in both physics and mathematics, reflects the inherent difficulty associated with a task [125]. In the context of modeling an observed object with high accuracy, complexity demonstrably relates to the chosen "frame" - the system of units and variables selected by the researcher. The intricacy and information content of this frame directly influence the reconstruction difficulty. Existing data processing methods typically overlook the impact of frame selection on model complexity.

To address these concerns, a universally applicable criterion for assessing model-phenomena discrepancies is urgently needed. This criterion, termed "comparative certainty," aims to evaluate these mismatches and provide a theoretically robust framework applicable across scientific disciplines adhering to the

International System of Units (SI). By establishing this criterion, scientific investigations can achieve greater reproducibility and reliability, bolstering confidence in published results.

Our work highlights the Frame of Finite Information Quantity (FIQ) method as a potential solution. Unlike traditional approaches, FIQ empowers researchers to select the most plausible model for the object under study by explicitly considering the information content within the chosen system of units. This system acts as a metaphorical “shell” encapsulating the essence of the investigated physical phenomenon.

The prevailing paradigm suggests that a model’s accuracy hinges solely on the quality of the experimental data processing method. However, our findings necessitate a paradigm shift. The accuracy of a model constructed using the FIQ method is demonstrably proportional to the information content embedded within the chosen system of units. This challenges the long-held belief within the scientific community.

The informational approach offers a fresh perspective on quantifying model uncertainty. Traditionally, statistical methods dominated this field. However, the informational approach focuses on the information transmission, accumulation, and transformation processes inherent in model construction. It captures the irreducible uncertainty associated with the model’s qualitative and quantitative variables, providing a holistic measure of overall uncertainty.

The significance of the SI in scientific research cannot be overstated. By providing a standardized framework for measurements, the SI ensures consistency, traceability, and comparability, enabling accurate and replicable experiments. This promotes interdisciplinary collaboration, quality control, and error analysis. Using SI units fosters global communication, enhances research impact, and upholds scientific integrity.

The selection of a specific unit system, such as the SI, plays a crucial role in model formulation. This system comprises a finite set of physical dimensional variables that characterize the world’s physical properties. It serves as the foundation for all scientific knowledge and establishes a framework for modeling phenomena. By conceptualizing a model as an information channel bridging the phenomenon and the observer, information theory’s concepts and mathematical tools can be applied to assess the model’s accuracy and determine its permissible discrepancy.

The comparative certainty criterion has wide-ranging implications for diverse experimental data. It provides a universal metric (ϵ) for quantifying the model’s proximity to the studied object. This metric transcends statistical methods and offers insights into the fundamental nature of reality. By analyzing experimental data using relative uncertainty and considering the conditions and requirements of the informational approach, researchers can detect subtle deviations from established principles in modeling physical phenomena, potentially revealing new

discoveries.

However, applying comparative uncertainty analysis to diverse experimental data presents challenges and considerations. The informational approach necessitates careful consideration of the unit system, variable selection, and potential information distortion during the model-building process. Researchers need to account for various uncertainty sources and potential limitations to enhance the accuracy and reliability of their predictions.

The comparative uncertainty criterion, grounded in the informational approach, holds significant promise for advancing scientific rigor and addressing concerns about research reliability and credibility. By quantifying model-phenomena mismatches and providing a theoretically sound framework, this criterion can enhance reproducibility and instill greater confidence in published findings. Nevertheless, challenges in applying this approach to diverse experimental data necessitate careful consideration and further research. Overall, establishing the comparative certainty criterion represents a substantial step towards ensuring the robustness and credibility of scientific research across all disciplines.

6. Conclusions

The paper aims to enhance the understanding of the modeling process, measurement accuracy, and the role of information in representing physical phenomena. It emphasizes the need to consider the philosophical perspectives and subjective judgments of researchers in constructing accurate models.

The paper provides valuable insights into the complex process of modeling physical phenomena, incorporating information theory principles to evaluate and enhance the accuracy of scientific models.

Traditional data processing methods struggle to capture the intricate details within complex models due to limitations in addressing model complexity itself.

The frame of finite information quantity (FIQ) offers a novel approach by explicitly considering the information content within the chosen system of units during model selection. This methodological shift acknowledges the critical role of frame selection in influencing model complexity and accuracy.

Our findings challenge the prevailing paradigm that solely focuses on data processing methods for achieving model accuracy. The FIQ method highlights that the information content embedded within the chosen units themselves demonstrably impacts model accuracy.

The FIQ method goes beyond traditional approaches by providing a framework to address the inherent uncertainty associated with models. By enabling the selection of optimal variables, FIQ demonstrably reduces this uncertainty within the constructed model.

The FIQ method presents a promising avenue for future scientific exploration. While its potential is vast, further research is necessary to:

- 1) Explore applicability across disciplines: investigate the effectiveness of the FIQ method in various scientific fields beyond the domain in which it was in-

initially developed.

2) Address potential limitations: identify and address potential limitations of the FIQ method, such as computational complexity or challenges arising in specific experimental scenarios.

By emphasizing the importance of information content within the modeling frame, our work paves the way for overcoming the information-theoretic bottleneck associated with model complexity. The FIQ method offers researchers a powerful tool to construct more accurate and reliable models from experimental data, ultimately leading to a deeper understanding of the scientific phenomena under investigation. The integration of the FIQ method holds the potential to transform various scientific disciplines by facilitating the development of more robust and informative models.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Menin, B. (2017) Information Approach for Calculating the Resolutions of Energy, Length and Information. *Journal of Multidisciplinary Engineering Science and Technology*, **4**, 6859-6862.
<https://www.jmest.org/wp-content/uploads/JMESTN42352095.pdf>
- [2] Li, Y., Yang, B., Yang, N. and Wang, T. (2019) Application of Interpretable Machine Learning Models for the Intelligent Decision. *Neurocomputing*, **333**, 273-283.
<https://doi.org/10.1016/j.neucom.2018.12.012>
- [3] Humphries, G., Magness, D.R. and Huettmann, F. (2018) Machine Learning for Ecology and Sustainable Natural Resource Management. Springer, Cham.
<https://doi.org/10.1007/978-3-319-96978-7>
- [4] Fehrman, B., Gess, B. and Jentzen, A. (2020) Convergence Rates for the Stochastic Gradient Descent Method for Non-Convex Objective Functions. *Journal of Machine Learning Research*, **21**, 1-48.
<https://www.jmlr.org/papers/volume21/19-636/19-636.pdf>
- [5] Shi, H., Zhang, X., Gao, Y., Wang, S. and Ning, Y. (2023) Robust Total Least Squares Estimation Method for Uncertain Linear Regression Model. *Mathematics*, **11**, Article 4354. <https://doi.org/10.3390/math11204354>
- [6] Rolnick, D., *et al.* (2022) Tackling Climate Change with Machine Learning. *ACM Computing Surveys (CSUR)*, **55**, Article 42.
<https://dl.acm.org/doi/pdf/10.1145/3485128>
- [7] De Baets, N.B. and Waegeman, W. (2023) Conditional Validity of Heteroskedastic Conformal Regression. 1-31. <https://arxiv.org/pdf/2309.08313>
- [8] Sirimongkolkasem, T. and Drikvandi, R. (2019) On Regularization Methods for Analysis of High Dimensional Data. *Annals of Data Science*, **6**, 737-763.
<https://doi.org/10.1007/s40745-019-00209-4>
- [9] Liu, M., *et al.* (2022) Handling Missing Values in Healthcare Data: A Systematic Review of Deep Learning-Based Imputation Techniques. 1-34.
<https://arxiv.org/ftp/arxiv/papers/2210/2210.08258.pdf>

- [10] Van der Vaart, A.W. (2000) Asymptotic Statistics. Cambridge University Press. https://assets.cambridge.org/97805214/96032/frontmatter/9780521496032_frontmatter.pdf
- [11] Lehmann, E.L. and Casella, G. (2015) Theory of Point Estimation. Springer, Berlin.
- [12] Casella, G. and Berger, R.L. (2002) Statistical Inference. Duxbury Press. https://www.academia.edu/34751941/Casella_berger_statistical_inference
- [13] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2014) Bayesian Data Analysis. Chapman and Hall/CRC, New York. <https://doi.org/10.1201/b16018>
- [14] Murphy, K.P. (2012) Machine Learning: A Probabilistic Perspective. MIT Press. https://www.academia.edu/35856835/Machine_Learning_A_Probabilistic_Perspective
- [15] Varin, C., Reid, N. and Firth, D. (2011) An Overview of Composite Likelihood Methods. *Statistica Sinica*, **21**, 5-42. <https://www3.stat.sinica.edu.tw/statistica/oldpdf/A21n11.pdf>
- [16] Nocedal, J. and Wright, S.J. (2006) Numerical Optimization, 2nd Edition, Springer. <https://www.math.uci.edu/~qnie/Publications/NumericalOptimization.pdf>
- [17] Betancourt, M. (2017) A Conceptual Introduction to Stochastic Gradient Methods. <https://arxiv.org/pdf/1701.02434>
- [18] Hjort, K., *et al.* (2018) Open Problems in Likelihood and Bayesian Inference. *International Statistical Review*, **86**, 219-252. <https://link.springer.com/book/10.1007/978-3-662-60792-3>
- [19] Li, J., Wang, Z., Li, R. and Wu, R. (2015) Bayesian Group Lasso for Nonparametric Varying-Coefficient Models with Application to Functional Genome-Wide Association Studies. *The Annals of Applied Statistics*, **9**, 640-664. <https://www.jstor.org/stable/24522596>
- [20] Chowdhury, S., Uddin, G., Hemmati, H. and Holmes, R. (2024) Method-Level Bug Prediction: Problems and Promises. *ACM Transactions on Software Engineering and Methodology*, **33**, Article No. 98. <https://doi.org/10.1145/3640331>
- [21] Villaverde, A.F., Bongard, S., Mauch, K., Müller, D., Balsa-Canto, E., Schmid, J. and Banga, J.R. (2015) A Consensus Approach for Estimating the Predictive Accuracy of Dynamic Models in Biology. *Computer Methods and Programs in Biomedicine*, **119**, 17-28. <https://doi.org/10.1016/j.cmpb.2015.02.001>
- [22] Wolf, B.J., Jiang, Y., Wilson, S.H. and Oates, J.C. (2021) Variable Selection Methods for Identifying Predictor Interactions in Data with Repeatedly Measured Binary Outcomes. *Journal of Clinical and Translational Science*, **5**, e59. <https://doi.org/10.1017/cts.2020.556>
- [23] Fay, D.S. and Gerow, K. (2013) A Biologist's Guide to Statistical Thinking and Analysis. WormBook. <http://www.wormbook.org> <https://doi.org/10.1895/wormbook.1.159.1>
- [24] Barbierato, E. and Gatti, A. (2024) The Challenges of Machine Learning: A Critical Review. *Electronics*, **13**, Article 416. <https://doi.org/10.3390/electronics13020416>
- [25] Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012) Introduction to Linear Regression Analysis. 5th Edition. <https://ocd.lcwu.edu.pk/cfiles/Statistics/Stat-503/IntroductiontoLinearRegressionAnalysisbyDouglasC.MontgomeryElizabethA.PeckG.GeoffreyViningz-lib.org.pdf>
- [26] Jarantow, S.W., Pisors, E.D. and Chiu, M.L. (2023) Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays. *Current Protocols*, **3**, e801. <https://doi.org/10.1002/cpz1.801>

- [27] Mikkola, P., Martin, O.A., Chandramouli, S., *et al.* (2023) Prior Knowledge Elicitation: The Past, Present, and Future. *Bayesian Analysis*, Advance Publication, 1-33. <https://doi.org/10.1214/23-BA1381>
- [28] Ferienc, M., Maji, P., Mattina, M. and Rodrigues, M. (2021) On the Effects of Quantisation on Model Uncertainty in Bayesian Neural Networks. *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI 2021), Proceedings of Machine Learning Research*, **161**, 929-938. <https://proceedings.mlr.press/v161/ferienc21a/ferienc21a.pdf>
- [29] Glasauer, S. (2019) Chapter 1—Sequential Bayesian Updating as a Model for Human Perception. *Progress in Brain Research*, **249**, 3-18. <https://sci-hub.se/https://doi.org/10.1016/bs.pbr.2019.04.025>
<https://doi.org/10.1016/bs.pbr.2019.04.025>
- [30] Alstona, C., *et al.* (2005) Bayesian Model Comparison: Review and Discussion. International Statistical Institute. https://www.researchgate.net/publication/239442576_Bayesian_Model_Comparison_Review_and_Discussion
- [31] Craiu, R.V., Gustafson, P. and Rosenthal, J.S. (2022) Reflections on Bayesian Inference and Markov Chain Monte Carlo. *The Canadian Journal of Statistics*, **50**, 1213-1227. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjs.11707>
<https://doi.org/10.1002/cjs.11707>
- [32] Goldstein, M. (2006) Subjective Bayesian Analysis: Principles and Practice. *Bayesian Analysis*, **1**, 403-420. <https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-3/Subjective-Bayesian-Analysis-Principles-and-Practice/10.1214/06-BA116.pdf>
<https://doi.org/10.1214/06-BA116>
- [33] Taka, E., Stein, S. and Williamson, J.H. (2020) Increasing Interpretability of Bayesian Probabilistic Programming Models through Interactive Representations. *Human-Media Interaction-Frontiers in Computer Science*, **2**, Article 567344. <https://sci-hub.se/https://doi.org/10.3389/fcomp.2020.567344>
<https://doi.org/10.3389/fcomp.2020.567344>
- [34] Akanbi, O.B., Olubusoye, O.E. and Odeyemi, O.O. (2020) Sensitivity of the Posterior Mean on the Prior Assumptions: An Application of the Ellipsoid Bound Theorem. *Journal of Scientific Research & Reports*, **26**, 134-149. <https://doi.org/10.9734/jsrr/2020/v26i730291>
- [35] Dąbrowska, E. (2020) Monte Carlo Simulation Approach to Reliability Analysis of Complex Systems. *Journal of KONBiN*, **50**, 155-170. <https://doi.org/10.2478/jok-2020-0010>
- [36] Albert, D.R. (2020) Monte Carlo Uncertainty Propagation with the NIST Uncertainty Machine. *Journal of Chemical Education*, **97**, 1491-1494. <https://doi.org/10.1021/acs.jchemed.0c00096>
- [37] Bond, S.D., Franke, B.C., Lehoucq, R.B., Smith, J.D. and McKinley, S.A. (2022) Sensitivity Analyses for Monte Carlo Sampling-Based Particle Simulations. Sandia National Laboratories, 1-50. <https://www.sandia.gov/app/uploads/sites/205/2022/09/SAND2022-12721.pdf>
<https://doi.org/10.2172/1889334>
- [38] Krystul, J. and Blom, H.A.P. (2006) Sequential Monte Carlo Simulation of Rare Event Probability in Stochastic Hybrid Systems. National Aerospace Laboratory NLR, 1-7. <https://core.ac.uk/download/pdf/53034179.pdf>
<https://doi.org/10.3182/20050703-6-CZ-1902.00382>

- [39] Muraro, S., Battistoni, G. and Kraan, A.C. (2020) Challenges in Monte Carlo Simulations as Clinical and Research Tool in Particle Therapy: A Review. *Frontiers in Physics*, **8**, Article 567800. <https://sci-hub.se/10.3389/fphy.2020.567800>
<https://doi.org/10.3389/fphy.2020.567800>
- [40] Hickey, J.M., Veerkamp, R.F., Calus, M.P.L., Mulder, H.A. and Thompson, R. (2009) Estimation of Prediction Error Variances via Monte Carlo Sampling Methods Using Different Formulations of the Prediction Error Variance. *Genetics Selection Evolution*, **41**, Article 23. <https://sci-hub.se/10.1186/1297-9686-41-23>
<https://doi.org/10.1186/1297-9686-41-23>
- [41] Roslan, N.R., Fauzi, N. and Ridzuan, M. (2022) Monte Carlo Simulation Convergences' Percentage and Position in Future Reliability Evaluation. *International Journal of Electrical and Computer Engineering*, **12**, 6218-6227. <https://doi.org/10.11591/ijece.v12i6.pp6218-6227>
- [42] Qin, N., *et al.* (2018) Full Monte Carlo-Based Biologic Treatment Plan Optimization System for Intensity Modulated Carbon Ion Therapy on Graphics Processing Unit. *International Journal of Radiation Oncology Biology Physics*, **100**, 235-243. <https://sci-hub.se/https://doi.org/10.1016/j.ijrobp.2017.09.002>
- [43] Kern, S., McGuinn, M.E., Smith, K.M., *et al.* (2023) Computationally Efficient Parameter Estimation for High-Dimensional Ocean Biogeochemical Models. *Geoscientific Model Development*, **17**, 1-34. <https://doi.org/10.5194/gmd-2023-107>
- [44] Gornov, A., Sorokovikov, P. and Zarodnyuk, T. (2019) Computational Technology for Global Search Based on the Modified Algorithm of the Univariate Nonlocal Optimization. *Advances in Intelligent Systems Research*, **169**, 189-193. <https://www.atlantis-press.com/article/125917325.pdf>
<https://doi.org/10.2991/iwci-19.2019.33>
- [45] Iwasaki, Y., *et al.* (2022) Evaluation of Optimization Algorithms and Noise Robustness of DMDsp. *IEEE Access*, **10**, 80748-80763. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9837072>
<https://doi.org/10.1109/ACCESS.2022.3193157>
- [46] Koppen, S., Langelaar, M. and van Keulen, F. (2022) A Simple and Versatile Topology Optimization Formulation for Flexure Synthesis. *Mechanism and Machine Theory*, **172**, Article 104743. <https://doi.org/10.1016/j.mechmachtheory.2022.104743>
- [47] Esward, T., Matthews, C., Wright, L. and Yang, X.-S. (2010) Sensitivity Analysis, Optimization, and Sampling Methods Applied to Continuous Models. NPL Report MS 2, 1-44. <https://eprintspublications.npl.co.uk/4783/1/MS2.pdf>
- [48] Arora, S. and Barak, B. (2007) Computational Complexity: A Modern Approach. <https://theory.cs.princeton.edu/complexity/book.pdf>
- [49] Arora, S. and Singh, S. (2016) An Effective Hybrid Butterfly Optimization Algorithm with Artificial Bee Colony for Numerical Optimization. *International Journal of Interactive Multimedia and Artificial Intelligence*, **4**, 14-21. https://reunir.unir.net/bitstream/handle/123456789/11743/ijimai20174_4_2_pdf_16914.pdf?sequence=1&isAllowed=y
<https://doi.org/10.9781/ijimai.2017.442>
- [50] Sebastjan, P. and Kuś, W. (2023) Method for Parameter Tuning of Hybrid Optimization Algorithms for Problems with High Computational Costs of Objective Function Evaluations. *Applied Sciences*, **13**, Article 6307. <https://doi.org/10.3390/app13106307>
- [51] Singh, C. (2023) Machine Learning in Pattern Recognition. *European Journal of*

- Engineering and Technology Research*, **8**, 63-68.
<https://doi.org/10.24018/ejeng.2023.8.2.3025>
- [52] Caputo, C. and Cardin, M.-A. (2021) The Role of Machine Learning for Flexibility and Real Options Analysis in Engineering Systems Design. *Proceedings of the International Conference on Engineering Design (ICED21)*, Gothenburg, 16-20 August 2021, 3121-3130. <https://doi.org/10.1017/pds.2021.573>
 - [53] Sharm, V. (2022) A Study on Data Scaling Methods for Machine Learning. *International Journal for Global Academic & Scientific Research*, **1**, 31-42.
<https://doi.org/10.55938/ijgasr.v1i1.4>
 - [54] Diaz, R.S., Neutatz, F. and Abedjan, Z. (2021) Automated Feature Engineering for Algorithmic Fairness. *Proceedings of the VLDB Endowment*, **14**, 1694-1702.
<https://doi.org/10.14778/3461535.3463474>
 - [55] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L. and Zhong, C. (2022) Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *Statistics Surveys*, **16**, 1-85. <https://doi.org/10.1214/21-SS133>
 - [56] Pei, Z., Liu, L., Wang, C. and Wang, J. (2021) Requirements Engineering for Machine Learning: A Review and Reflection, 1-10.
https://aire-ws.github.io/aire22/papers/AIRE_05.pdf
 - [57] Ying, X. (2019) An Overview of Overfitting and Its Solutions. *Journal of Physics: Conference Series*, **1168**, Article 022022.
<https://doi.org/10.1088/1742-6596/1168/2/022022>
 - [58] Elgeldawi, E., Sayed, A., Galal, A.R. and Zaki, A.M. (2021) Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. *Informatics*, **8**, Article 79. <https://doi.org/10.3390/informatics8040079>
 - [59] Salema, N. and Hussein, S. (2019) Data Dimensional Reduction and Principal Components Analysis. *Procedia Computer Science*, **163**, 292-299.
<https://doi.org/10.1016/j.procs.2019.12.111>
 - [60] Ali, E., Hossain, A. and Islam, R. (2019) Analysis of PCA Based Feature Extraction Methods for Classification of Hyperspectral Image. 2019 *2nd International Conference on Innovation in Engineering and Technology (ICIET)*, Dhaka, 23-24 December 2019, 1-6. <https://sci-hub.se/10.1109/ICIET48527.2019.9290629>
<https://doi.org/10.1109/ICIET48527.2019.9290629>
 - [61] Colom, M. and Buades, A. (2016) Analysis and Extension of the PCA Method, Estimating a Noise Curve from a Single Image. *Image Processing On Line*, **6**, 365-390.
<https://doi.org/10.5201/ipol.2016.124>
 - [62] Misue, K., Sugiyama, K. and Tanaka, J. (2006) Asia-Pacific Symposium on Information Visualization. *Conferences in Research and Practice in Information Technology (CRPIT)*, **60**, 1-10.
<https://crpit.scem.westernsydney.edu.au/Vol60.html>
 - [63] Weeraratne, N., Hunt, L. and Kurz, J. (2024) Challenges of Principal Component Analysis in High-Dimensional Settings when $n < p$. Preprint, 1-32.
<https://www.researchsquare.com/article/rs-4033858/v1>
<https://doi.org/10.21203/rs.3.rs-4033858/v1>
 - [64] Briscik, M., Dillies, M.-A. and Déjean, S. (2023) Improvement of Variables Interpretability in Kernel PCA. *BMC Bioinformatics*, **24**, Article No. 282.
<https://doi.org/10.1186/s12859-023-05404-y>
 - [65] Hong, J. and Kent, E.M. (2000) Bias in Principal Components Analysis Due to Correlated Observations. *Conference on Applied Statistics in Agriculture*, Manhattan, 30April-2May, 2000, 148-160.

- [66] Yang, X., Chen, J., Gu, X., He, R. and Wang, J. (2023) Sensitivity Analysis of Scalable Data on Three PCA Related Fault Detection Methods Considering Data Window and Thermal Load Matching Strategies. *Expert Systems with Applications*, **234**, Article 121024.
<https://www.sciencedirect.com/science/article/abs/pii/S0957417423015269>
<https://doi.org/10.1016/j.eswa.2023.121024>
- [67] Lin, Z., Yin, F. and Maronas, J. (2023) Towards Flexibility and Interpretability of Gaussian Process State-Space Model. 1-22. <https://arxiv.org/pdf/2301.08843.pdf>
- [68] Li, Y., Rao, S., Hassaine, A., *et al.* (2021) Deep Bayesian Gaussian Processes for Uncertainty Estimation in Electronic Health Records. *Scientific Reports*, **11**, Article No. 20685. <https://doi.org/10.1038/s41598-021-00144-6>
- [69] Patan, A., *et al.* (2022) Adversarial Robustness Guarantees for Gaussian Processes. *Journal of Machine Learning Research*, **23**, 1-55.
<https://jmlr.org/papers/volume23/21-0382/21-0382.pdf>
- [70] Galy-Fajou, T. and Opper, M. (2021) Adaptive Inducing Points Selection for Gaussian Processes. 1-9. <https://arxiv.org/pdf/2107.10066.pdf>
- [71] Belyaev, M., Burnaev, E. and Kapushev, Y. (2016) Computationally Efficient Algorithm for Gaussian Process Regression in Case of Structured Samples. *Computational Mathematics and Mathematical Physics*, **56**, 499-513.
<https://doi.org/10.1134/S0965542516040163>
- [72] Liu, H., Cai, J., Ong, Y.-S. and Wang, Y. (2019) Understanding and Comparing Scalable Gaussian Process Regression for Big Data. *Knowledge-Based Systems*, **164**, 324-335. <https://sci-hub.se/https://doi.org/10.1016/j.knosys.2018.11.002>
<https://doi.org/10.1016/j.knosys.2018.11.002>
- [73] Abdessalem, A.B., Dervilis, N., Wagg, D.J. and Worden, K. (2017) Automatic Kernel Selection for Gaussian Processes Regression with Approximate Bayesian Computation and Sequential Monte Carlo. *Frontiers in Built Environment*, **3**, Article 52.
<https://doi.org/10.3389/fbuil.2017.00052>
- [74] Yoshikawa, Y. and Iwata, T. (2023) Gaussian Process Regression with Interpretable Sample-Wise Feature Weights. *IEEE Transactions on Neural Networks and Learning Systems*, **34**, 5789-5803. <https://doi.org/10.1109/TNNLS.2021.3131234>
- [75] Zhang, J., Yang, Y. and Ding, J. (2023) Information Criteria for Model Selection. *WIREs Computational Statistics*, **15**, e1607.
<https://wires.onlinelibrary.wiley.com/doi/epdf/10.1002/wics.1607>
<https://doi.org/10.1002/wics.1607>
- [76] Kuha, J. (2004) AIC and BIC: Comparisons of Assumptions and Performance. *Sociological Methods & Research*, **33**, 188-229.
<https://sci-hub.se/10.1177/0049124103262065>
<https://doi.org/10.1177/0049124103262065>
- [77] Emiliano, P.C., Vivanco, M.J.F. and de Menezes, F.S. (2013) Information Criteria: How Do They Behave in Different Models? *Computational Statistics & Data Analysis*, **69**, 141-153. <https://sci-hub.se/10.1016/j.csda.2013.07.032>
<https://doi.org/10.1016/j.csda.2013.07.032>
- [78] Preacher, K.J. and Merkle, E.C. (2012) The Problem of Model Selection Uncertainty in Structural Equation Modeling. *Psychological Methods*, **17**, 1-14.
https://quantpsy.org/pubs/preacher_merkle_2012.pdf
<https://doi.org/10.1037/a0026804>
- [79] Brewer, M.J., Butler, A. and Cooksley, S.L. (2016) The Relative Performance of AIC, AICC and BIC in the Presence of Unobserved Heterogeneity. *Methods in Ecology*

- and Evolution*, **7**, 679-692.
<https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12541>
<https://doi.org/10.1111/2041-210X.12541>
- [80] Harbecke, J., Grunau, J. and Samanek, P. (2024) Are the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) Applicable in Determining the Optimal Fit and Simplicity of Mechanistic Models? *International Studies in the Philosophy of Science*, 1-20. <https://doi.org/10.1080/02698595.2024.2304487>
 - [81] Chakrabarti, A. and Ghosh, J.K. (2011) AIC, BIC and Recent Advances in Model Selection. *Philosophy of Statistics*, **7**, 583-605.
<https://sci-hub.se/https://doi.org/10.1016/B978-0-444-51862-0.50018-6>
<https://doi.org/10.1016/B978-0-444-51862-0.50018-6>
 - [82] Dziak, J.J., Coffman, D.L., Lanza, S.T. and Li, R. (2012) Sensitivity and Specificity of Information Criteria. The Pennsylvania State University, Technical Report Series, 1-31.
<https://www.latentclassanalysis.com/wp-content/uploads/2021/04/12-119-1.pdf>
 - [83] Menin, B. (2019) Progress in Reducing the Uncertainty of Measurement of Planck's Constant in Terms of the Information Approach. *Physical Science International Journal*, **21**, 1-11. <https://journalspsij.com/index.php/PSIJ/article/view/531>
 - [84] Menin, B. (2018) h , k , N_A : Evaluating the Relative Uncertainty of Measurement. *American Journal of Computational and Applied Mathematics*, **8**, 93-102.
<http://article.sapub.org/10.5923.j.ajcam.20180805.02.html>
 - [85] Del Santo, F. and Gisin, N. (2019) Physics without Determinism: Alternative Interpretations of Classical Physics. *Physical Review A*, **100**, Article 062107.
<https://doi.org/10.1103/PhysRevA.100.062107>
 - [86] Zurek, W.H. (2022) Quantum Theory of the Classical: Einselection, Envariance, Quantum Darwinism and Extantons. *Entropy*, **24**, Article 1520.
<https://doi.org/10.3390/e24111520>
 - [87] Bombelli, L., Koul, R.K., Lee, J. and Sorkin, R.D. (1986) Quantum Source of Entropy for Black Holes. *Physical Review D*, **34**, 373-383.
<https://doi.org/10.1103/PhysRevD.34.373>
 - [88] Marolf, D. (2017) The Black Hole Information Problem: Past, Present, and Future. *Reports on Progress in Physics*, **80**, Article 092001.
<https://sci-hub.se/10.1088/1361-6633/aa77cc>
<https://doi.org/10.1088/1361-6633/aa77cc>
 - [89] Tegmark, M. (2014) Our Mathematical Universe: My Quest for the Ultimate Nature of Reality. Vintage Books, New York.
 - [90] Valentini, A. (2002) Subquantum Information and Computation. *Pramana Journal of Physics*, **59**, 269-277. <https://sci-hub.se/10.1007/s12043-002-0117-1>
<https://doi.org/10.1007/s12043-002-0117-1>
 - [91] Dowker, F. and Zalel S. (2017) Evolution of Universes in Causal Set Cosmology. *Comptes Rendus Physique*, **18**, 246-253. <https://doi.org/10.1016/j.crhy.2017.03.002>
 - [92] Menin, B. (2021) Construction of a Model as an Information Channel between the Physical Phenomenon and Observer. *Journal of the Association for Information Science and Technology*, **72**, 1198-1210. <https://doi.org/10.1002/asi.24473>
 - [93] Burgin, M. (2003) Information Theory: A Multifaceted Model of Information. *Entropy*, **5**, 146-160. <https://doi.org/10.3390/e5020146>
<https://doi.org/10.3390/e5020146>
 - [94] Menin, B. (2018) Applying Measurement Theory and Information-Based Measure

- in Modelling Physical Phenomena and Technological Processes. *European Journal of Engineering Research and Science*, **3**, 28-34.
<https://ej-eng.org/index.php/ejeng/article/view/594>
<https://doi.org/10.24018/ejers.2018.3.1.594>
- [95] Menin, B. (2019) A Look at the Uncertainty of Measuring the Fundamental Constants and the Maxwell Demon from the Perspective of the Information Approach. *Global Journal of Researchers in Engineering: A mechanical and mechanics engineering*, **19**, 1-17.
https://globaljournals.org/GJRE_Volume19/1-A-Look-at-the-Uncertainty.pdf
- [96] Menin, B. (2019) Precise Measurements of the Gravitational Constant: Revaluation by the Information Approach. *Journal of Applied Mathematics and Physics*, **7**, 1272-1288. http://file.scirp.org/pdf/JAMP_2019062614403787.pdf
<https://doi.org/10.4236/jamp.2019.76087>
- [97] Menin, B. (2019) Hubble Constant Tension in Terms of Information Approach. *Physical Science International Journal*, **23**, 1-15.
<https://doi.org/10.9734/psij/2019/v23i430165>
- [98] Sedov, L.I. (1993) Similarity and Dimensional Methods in Mechanic. CRC Press, Florida.
- [99] Menin, B. (2017) Simplest Method for Calculating the Lowest Achievable Uncertainty of Model at Measurements of Fundamental Physical Constants. *Journal of Applied Mathematics and Physics*, **5**, 2162-2171.
<https://www.scirp.org/journal/paperinformation?paperid=80237>
<https://doi.org/10.4236/jamp.2017.511176>
- [100] Menin, B. (2017) Novel Approach: Information Quantity for Calculating Uncertainty of Mathematical Model. *Proceedings*, **1**, Article 214.
<https://www.mdpi.com/2504-3900/1/3/214>
<https://doi.org/10.3390/IS4SI-2017-04034>
- [101] Menin, B. (2017) Universal Metric for the Assessing the Magnitude of the Uncertainty in the Measurement of Fundamental Physical Constants. *Journal of Applied Mathematics and Physics*, **5**, 365-385.
<https://www.scirp.org/journal/paperinformation?paperid=74189>
<https://doi.org/10.4236/jamp.2017.52033>
- [102] Menin, B.M. (2018) Optimal Mathematical Model for Description of Physical Phenomena and Technological Processes. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, **18**, 322-330 (In Russian).
<https://doi.org/10.17586/2226-1494-2018-18-2-322-330>
- [103] Menin, B. (2017) Information Measure Approach for Calculating Model Uncertainty of Physical Phenomena. *American Journal of Computational and Applied Mathematics*, **7**, 11-24.
- [104] Brillouin, L. (1953) *Science and Information Theory*. New York, NY, USA: Academic Press. <https://doi.org/10.1063/1.3057866>
- [105] Newell, D.B. and Tiesinga, E. (2019) The International System of Units (SI). NIST Special Publication 330, 1-138. <https://doi.org/10.6028/NIST.SP.330-2019>
- [106] Wübbeler, G., Bodnar, O. and Elster, C. (2017) Robust Bayesian Linear Regression with Application to an Analysis of the CODATA Values for the Planck Constant. *Metrologia*, **55**, 20-28. <https://sci-hub.se/10.1088/1681-7575/aa98aa>
<https://doi.org/10.1088/1681-7575/aa98aa>
- [107] Dodson, D. (2013) Quantum Physics and the Nature of Reality (QPNR) Survey: 2011. <https://www.scirp.org/reference/referencespapers?referenceid=2739480>

- [108] Henrion, M. and Fischhoff, B. (1986) Assessing Uncertainty in Physical Constants. *American Journal of Physics*, **54**, 791-798. <https://sci-hub.se/10.1119/1.14447>
<https://doi.org/10.1119/1.14447>
- [109] Menin, B. (2020) High Accuracy When Measuring Physical Constants: From the Perspective of the Information-Theoretic Approach. *Journal of Applied Mathematics and Physics*, **8**, 861-867. <https://doi.org/10.4236/jamp.2020.85067>
- [110] Menin, B. (2023) Advancing Scientific Rigor: Towards a Universal Informational Criterion for Assessing Model-Phenomenon Mismatch. *Journal of Applied Mathematics and Physics*, **11**, 1817-1836.
https://www.scirp.org/pdf/jamp_2023071113462472.pdf
<https://doi.org/10.4236/jamp.2023.117117>
- [111] Yao, W., *et al.* (2019) An Empirical Approach for Parameters Estimation of Underwater Electrical Wire Explosion. *Physics of Plasmas*, **26**, Article 093502.
<https://doi.org/10.1063/1.5111518>
- [112] Han, Z., Zhang, X., Yan, B., Qiao, L. and Li, Z. (2022) Methods on the Determination of the Circuit Parameters in an Underwater Spark Discharge. *Mathematical Problems in Engineering*, **2022**, Article ID: 7168375.
<https://doi.org/10.1155/2022/7168375>
- [113] Shafer, D., *et al.* (2015) Generation of Ultra-Fast Cumulative Water Jets by Sub-Microsecond Underwater Electrical Explosion of Conical Wire Arrays. *Physics of Plasmas*, **22**, Article 122703. <https://doi.org/10.1063/1.4937370>
- [114] Henzan, R., Higa, Y., Higa, O., Shimojima, K. and Itoh, S. (2018) Numerical Simulation of Electrical Discharge Characteristics Induced by Underwater Wire Explosion. *Materials Science Forum*, **910**, 72-77.
<https://sci-hub.se/10.4028/www.scientific.net/MSF.910.72>
<https://doi.org/10.4028/www.scientific.net/MSF.910.72>
- [115] Tuholukov, A. and Stelmashuk, V. (2020) Comparison of Underwater Spark Simulation Using Elliptical and Cylindrical Models. *WDS 20 Proceedings of Contributed Papers, Physics*, Prague, 22-24 September 2020, 111-117.
https://www.mff.cuni.cz/veda/konference/wds/proc/pdf20/WDS20_17_f2_Tuholukov.pdf
- [116] Wojtowicz, J., Wojtowicz, H. and Wajs, W. (2015) Simulation of Electrohydrodynamic Phenomenon Using Computational Intelligence Methods. *Procedia Computer Science*, **60**, 188-196. <https://sci-hub.se/10.1016/j.procs.2015.08.118>
<https://doi.org/10.1016/j.procs.2015.08.118>
- [117] Bose, D., Palmer, G.E. and Wright, M.J. (2006) Uncertainty Analysis of Laminar Aeroheating Predictions for Mars Entries. *Journal of Thermophysics and Heat Transfer*, **20**, 652-662. <https://doi.org/10.2514/1.20993>
- [118] Thol, M., Dubberke, F.H., Baumhögger, E., Span, R. and Vrabec, J. (2018) Speed of Sound Measurements and a Fundamental Equation of State for Hydrogen Chloride. *Journal of Chemical & Engineering Data*, **63**, 2533-2547.
<https://doi.org/10.1021/acs.jced.7b01031>
- [119] Trachenko, K., Monserrat, B., Pickard, C.J. and Brazhkin, V.V. (2020) Speed of Sound from Fundamental Physical Constants. *Science Advances*, **6**, eabc8662.
<https://sci-hub.se/10.1126/sciadv.abc8662>
<https://doi.org/10.1126/sciadv.abc8662>
- [120] Segovia, J.J., Lozano-Martin, D., Tuma, D., Moreau, A., Carmen Martín, M. and Vega-Maza, D. (2022) Speed of Sound Data and Acoustic Virial Coefficients of Two Binary (N_2+H_2) Mixtures at Temperatures between (260 and 350) K and at Pres-

- tures between (0.5 and 20) MPa. *The Journal of Chemical Thermodynamics*, **171**, Article 106791. <https://doi.org/10.1016/j.jct.2022.106791>
- [121] Gourgoulas, K., Katsoulakis, M.A., Rey-Bellet, L. and Wang, J. (2020) How Biased Is Your Model? Concentration Inequalities. *Information and Model Bias. IEEE Transactions on Information Theory*, **66**, 3079-3097. <https://arxiv.org/abs/1706.10260>
<https://doi.org/10.1109/TIT.2020.2977067>
- [122] Patra, L.K., Kayal, S. and Kumar, S. (2020) Measuring Uncertainty Under Prior Information. *IEEE Transactions on Information Theory*, **66**, 2570-2580. <https://sci-hub.se/10.1109/TIT.2020.2970408>
<https://doi.org/10.1109/TIT.2020.2970408>
- [123] Cunha Jr., A. (2017) Modeling and Quantification of Physical Systems Uncertainties in a Probabilistic Framework. Probabilistic Prognostics and Health Management of Energy Systems, Springer International Publishing, New York, 1-34. <https://hal.science/hal-01516295/document>
https://doi.org/10.1007/978-3-319-55852-3_8
- [124] Menin, B. (2022) The Role of Thinker Consciousness in Measurement Accuracy: An Informational Approach. *International Journal Information Theories & Applications*, **29**, 203-229. <https://doi.org/10.54521/ijita29-03-p01>
- [125] Golan, A. and Harte, J. (2022) Information Theory: A Foundation for Complexity Science. *Proceedings of the National Academy of Sciences of the United States of America*, **119**, e2119089119. <https://doi.org/10.1073/pnas.2119089119>