bnmetamodel 2.0

T. Griffiths[1], Z. Xuereb Conti[2], and M. Bluck[1]

[1]Department of Mechanical Engineering, Imperial College London, UK
[2]Data-Centric Engineering / TRIC:DT, The Alan Turing Institute, UK

***Abstract*—**

***Index Terms*—Fusion power, metamodels, surrogate modelling, fusion commercialisation, machine learning, fusion economics, energy, Bayesian Networks**

I. Introduction

Commercial-scale fusion power holds the promise of delivering reliable baseload electricity, reducing carbon emissions, and enhancing energy security. However, techno-economic modeling of future fusion power plants faces challenges due to uncertain and imprecise costing models. To address this uncertainty, probabilistic methods have been demonstrated to estimate fusion power economics effectively [1]

Despite the challenges, it's vital for the fusion community to persist in researching both technical and economic performance metrics. The advancement of nuclear fusion research is propelled by a multitude of scientific pursuits in modeling and analysis, drawing upon expertise in theory, experimentation, computation, and inference. Fusion reactions involve intricate processes at extreme conditions within the reactor. Deterministic, theoretical and numerical models must accurately capture these extreme conditions through plasma dynamics, energy transfer, and reaction kinetics. Yet, they often struggle to handle uncertainties inherent in plasma behavior fluctuations and in external factors such as magnetic field perturbations. These uncertainties can significantly impact predicted reactor performance, making consistent achievement of desired output parameters challenging, thus hindering reactor design in roadmapping and slowing R&D. While significant advancements have been made in fusion research, our understanding of plasma physics, fusion reactions, and reactor engineering continues to evolve. Deterministic models may lack the adaptability to incorporate emerging insights or experimental data, potentially leading to discrepancies between predicted and observed reactor behavior. Failing to address these challenges could hinder investment attraction, roadmap target achievement, and the promotion of additional investment opportunities. Implementing statistical methods, like sensitivity analyses, can help assess key modeling variables, identify areas for improvement, and enhance prediction reliability.

Machine Learning (ML) models offer a valuable tool for developing understanding in areas lacking experimental data. They provide the flexibility to simulate various scenarios and conditions rapidly, enabling quick iteration and parameter modification. This facilitates swift exploration and optimisation of designs without the need for physical modifications or repeated experiments. Given the complexity of fusion engineering systems, ML models can effectively handle numerous variables and interactions, facilitating the analysis of large-scale systems with intricate behaviours.

This study extends the surrogate modelling technique applied in [1] to a new case study, introducing enhancements and modifications to the methodology. The objective is to expand the application of Bayesian Networks (BNs) for a deeper understanding of fusion power design spaces. It introduces an innovative, probabilistic approach to managing uncertainty in fusion research, diverging from traditional techniques. The objective of surrogate modelling is to develop a streamlined representation that mirrors the outcomes of a complex deterministic numerical or theoretical model, accounting for its various inputs and parameters. In this scenario, a BN serves as a proxy for a fusion systems code enabling the replication of a spheerical tokamak in its design phase, forecasting the economic and performance viability amidst data uncertainties.

The use of probabilistic Bayesian inference with a Bayesian Network in this analysis offers several advantages over other surrogate modeling techniques. Unlike deterministic models, Bayesian inference accounts for uncertainties in input parameters and output responses, providing a probabilistic framework that captures the inherent variability in fusion systems. Furthermore, Bayesian inference facilitates the integration of prior knowledge and observational data, enabling more robust inference and decision-making in complex engineering systems like fusion reactors. Overall, leveraging Bayesian inference allows fusion developers to extract actionable insights from limited experimental data and make informed design decisions that maximise the performance and cost-effectiveness of fusion reactor systems.

The present investigation expands upon the surrogate modeling framework initially proposed by Griffiths et al. [1] through several notable enhancements. Firstly, it introduces an enhanced and refined methodology, detailed in Section III, to address limitations identified in the previous work. Secondly, it incorporates steps for validation and examines the influence of hyperparameters on the proof-of-concept BN, a feature lacking in [1] due to constraints on publication word count. Thirdly, the study shifts its focus to applying the improved methodology to a new case study, leveraging data from a private fusion developer to explore uncertain parameters of fusion power plants. A notable enhancement is the inclusion of multiple output nodes, distinguishing this study and strengthening the model's capability to address various outputs crucial to developers. Unlike in [1], the input and output data in this study represent a fusion reactor still in its design phase,

†Corresponding author: t.griffiths20@imperial.ac.uk

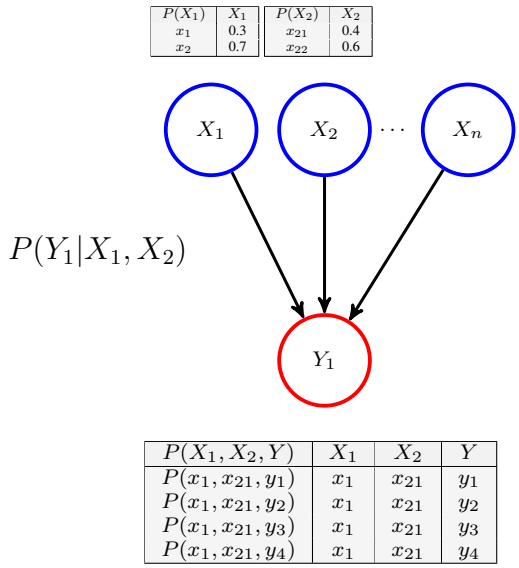


Fig. 1: Graphical representation of BN where nodes X_i represent the input variables and the node Y_1 represents the output variable to the analytical model. The solid lines represent the existing nodes in the model. The double-edged arrow represents additional information.

offering a more practical example of how the BN can support ongoing engineering decisions, enabling real-time optimisation and feedback for fusion developers.

The paper's structure is as follows: Section II provides an overview of BNs and their suitability in aid in fusion research in the methodological framework of surrogate modelling. Section II-A provides a comprehensive literature review, Section III outlines the refined methodology with illustrative examples from the proof-of-concept presented in [1], Section IV details the implementation of the BN in a new case study, Section V discusses the implications of the BN's results on the fusion developer's design approach, and Section VI concludes the paper, along with suggestions for future research directions.

II. Bayesian Networks

Under the umbrella of ML lies a multitude of computational algorithms, each with their own unique characteristics. Despite their diversity, they all share the fundamental trait that they are not programmed to solve specific tasks. Instead, they undergo iterative learning, gradually refining their performance through experience gained from previous iterations. Whilst learning, an ML algorithm utilises a statistical model with adjustable parameters that minimises a loss function through adjustment of internal parameters. For instance, in regression tasks, where the algorithm aims to align its output with a target value, common metrics like mean square error are often employed to evaluate its performance. The training process can be likened to a fitting routine, wherein the model, typically a highly flexible one such as a deep neural network (DNN), learns to capture intricate patterns within the data. The primary objective of training ML algorithms is to enable accurate predictions for new input data, requiring the them to learn solutions applicable beyond the training dataset. Generalisation, the ability to perform well on unseen data, serves as the primary criterion for evaluating an ML model's effectiveness. The main challenge lies in avoiding overfitting and enhancing generalisation. This challenge has led to the recent surge in popularity of deep learning (DL), which excels at approximating complex functions for high-dimensional input and output data, thereby offering solutions to real-world complex problems.

Within the realm of ML, Bayesian Networks (BNs) represent a notable approach, serving as a type of probabilistic graphical model that encapsulates a set of variables and their conditional dependencies through a directed acyclic graph (DAG) [2]. As ML algorithms aim to learn generalized solutions applicable beyond the training dataset, BNs offer a structured framework for capturing and representing probabilistic dependencies among variables. Unlike traditional ML approaches, BNs provide a more explicit representation of probabilistic relationships between variables, facilitating a deeper understanding of the underlying data generating process. This transition from ML to BNs underscores the shift from optimising predictive performance to gaining insight into the underlying mechanisms governing the data. Consequently, understanding BNs becomes increasingly valuable in leveraging the full potential of ML algorithms for real-world applications.

A BN consists of nodes representing variables and edges representing dependencies between variables. The parameters of a BN specify the conditional probabilities associated with each node given its parents in the network. These conditional probabilities determine the relationships and dependencies between variables in the BN. When multiple probability distributions are combined, a Joint Probability Distribution (JPD) is formed. The structure of the BN captures the dependencies between variables, and the conditional probabilities are derived from data or expert knowledge. It is not necessary to explicitly specify all the parameters of the JPD. This allows for efficient representation of the JPD without explicitly specifying all the individual parameters, making BNs a powerful tool for probabilistic modelling. This provides a compact representation by combining the *local* conditional distributions for each node, with respect to its connected parent nodes [3], see Figure 1. The joint probability distribution can be factorised into a product of conditional probability distributions, one for each variable given its parents in the graph. This factorisation is known as the chain rule of probability. Also known as the general product rule, this allows the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities.

Given a set of random variables, say $X_1, X_2, \ldots, X_n$, the chain rule of probability states that the joint probability of these variables is the product of the conditional probabilities of each variable given all the variables that precede it. Mathematically, this can be expressed as:

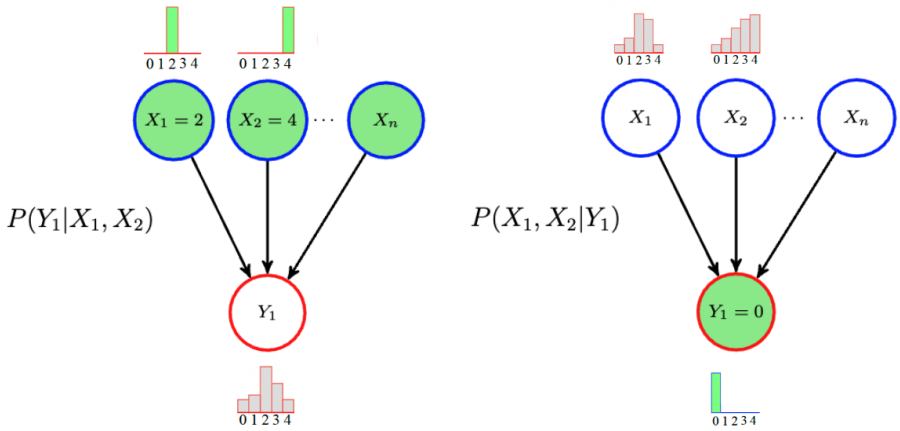


Fig. 2: A visual representation of Bayesian inference in the forward (left) and reverse (right) directions.

$$
\begin{aligned}
P(X_1, X_2, \ldots, X_n) = & P(X_1) \\
& * P(X_2|X_1) \\
& * P(X_3|X_1, X_2) \\
& * \ldots \\
& * P(X_n|X_1, X_2, \ldots, X_n - 1) \quad (1)
\end{aligned}
$$

In Equation 1: $P(X1, X2, \ldots, Xn)$ is the joint probability of $X1$ through Xn. $P(X_i|X1, \ldots, X_i - 1)$ is the conditional probability of X_i given all the preceding variables.

In depth, the term *prior* refers to the initial probability distribution assigned to a variable. On the other hand, the *posterior* probability refers to the updated probability distribution after *prior* beliefs are updated via *inference*, based on available evidence. The primary benefit of a probabilistic representation lies in its ability to perform *probabilistic inference*, which enables bi-directional reasoning with uncertain information using probabilistic methods. Thus, adopting a BN as a surrogate model facilitates exploration of causal relationships between fusion parameters and cost in both forward and reverse directions, allowing for an uncertain fusion parameter design space to be mapped from a desired target output range.

By defining inference as the process of making deductions based on evidence, reasoning and, prior knowledge, it is possible to further characterise Bayesian inference as the act of making alterations to existing probability distributions to discover how, based on their causal relationships, the variable's remaining probability distributions are altered. This acts as a useful tool to measure and understand relationships between inputs and responses in engineering design spaces [3].

The interpretation of simulation outputs from analytical models can be challenging for leadin fusion developers without a deep understanding of the associated engineering or physics principles. This can result in reliance on uncertain numerical data, without the ability to leverage the causal structure of the numerical output. In this scenario, a BN can serve as a two-way translation medium between the overall fusion research and the engineering domains. The use of bi-directional inference can assist engineers in providing feedback in the form of design parameters, based on the interconnections between the two domains. In this manner, fusion researchers can impose constraints on the output distributions of the BN, leveraging their expertise and experience, while engineers and physicists handle the input distributions to interpret engineering constraints in terms of design parameters. The 'translation' process facilitated by a BN could lead to more comprehensive decisions, as probabilistic inference considers the entire network of relationships between inputs and outputs during computation. This is in contrast to traditional deterministic models, which only consider the direct relationship between inputs and outputs.

The ability to perform bi-directional inference is a key feature of BNs, as it allows for the prediction of outputs based on input data, *forward inference*, and inputs based on output data, *reverse inference*. This capability is particularly useful in the context of engineering design, as it enables the exploration of causal relationships between inputs and outputs, and the prediction of inputs based on desired outputs. This can provide valuable insights for decision-making and design optimisation, as it allows for the identification of input values that are most likely to result in a desired output, and the prediction of the likely output given a set of input values. Refer to Figure 2 for a visual representation of Bayesian inferene in the forward (left) and reverse (right) directions.

In addition to its flexibility and robustness, probabilistic Bayesian inference can complement current methods in fusion research by providing a holistic understanding of reactor behaviour and performance. Different ways this technique can be used in conjunction with current methods to enhance fusion research include: integration with experimental data, uncertainty quantification, optimisation of experimental design, decision support for design iterations, and risk assessment and management.

A. Literature review

At present, besides the antecedent proof-of-concept study, no known instances exist where BNs have specifically been employed as surrogates in fusion research [1]. Consequently, this section will delve into the application of ML in fusion research, with a particular focus on the utilisation of other Bayesian modeling techniques in surrogates and uncertainty quantification.

In a review article by Pavonne et al. a comprehensive overview of ML and Bayesian inference in fusion R&D is provided [4]. Bayesian inference facilitates the efficient utilization of shared information among heterogeneous experimental data sources concerning a common underlying description or model of a phenomenon. The study outlines fundamental Bayesian modeling for fusion experiments in physics models via inference of plasma equilibria [5, 6], Gaussian process (GP) tomography [7], and uncertainty propagation [8, 9]. The application of ML in fusion research has encompassed various types, demonstrating the potential to address complex challenges in fusion reactor design and operation.

Pavone et al. details significant studies employing trending ML techniques in fusion, particularly within the realm of deep learning (DL) methods in neural networks. Specifically, in the theory of artificial neural network (ANN) algorithms and convolutional neural networks (CNNs), the latter of which was utilised in [10] for the Wendelstein 7-X Stellarator experiment to reconstruct ion and electron temperature profiles from 2D x-ray spectral images. The article underscores the significance of Bayesian inference for neural networks within the fusion community. Reinforcement learning (RL) has been employed in seminal studies for controlling tokamak plasmas [11] and predicting disruptions [12], showcasing applicability to practical tasks in real-time. In [11], RL enhanced flexibility in plasma control by generating non-linear feedback controllers, thereby simplifying the control system. The RL system learned the control policy through interaction with a simulated environment modeling the dynamic state of the plasma shape and current, magnetic diagnostics sensor data, coil currents, and controller dynamics. In [12], a disruption mitigation system utilizing DL predicts potential disruptions, aiming to reduce thermal loads. The algorithm combines convolutional and recurrent neural networks (CNN and RNN, respectively), trained on data from the JET and DIII-D experiments. The CNN learns the temporal dynamics of the plasma and predicts the likelihood of disruptive states. If the system output exceeds a certain threshold, an alarm triggers the activation of a mitigation action.

Surrogate models aim to expedite computationally intensive models based on numerical approximations and theoretical first principles, a particularly crucial endeavor in fusion research, especially concerning turbulent transport computations. An illustrative case provided in [14] demonstrates the efficacy of surrogate models: leveraging a dataset from [15], a neural network was trained. Integrated into modeling simulations, this neural network efficiently calculated the evolution of ion and electron temperature profiles as well as electron density profiles. The network delivered crucial values such as the main ion and electron heat flux and electron particle flux at a remarkable speed—864 times faster than a numerical model requiring four times the number of cores. Similar instances of neural networks serving as surrogates can be found in [16], focusing on turbulent transport calculations, and [17], exploring physics-informed plasma behavior in a particle-in-cell gyrokinetic code.

Pavone et al. also shed light on the computational complexity inherent in Bayesian inference, particularly in sampling from the posterior distribution. They elucidate that while Markov Chain Monte Carlo (MCMC) methods offer a means to sample from the posterior, their practicality diminishes in many real-world scenarios due to the substantial number of iterations required. The primary bottleneck arises from the need to compute the likelihood function, which entails running a forward modelz or simulation code—a process often prohibitively slow in complex systems. Pavone and colleagues argue that deep learning (DL) models offer a promising alternative, demonstrating their ability to efficiently reconstruct the posterior and infer plasma parameters from diagnostic data within short time frames, as evidenced in studies such as those by Pavone et al. [18, 10, 19, 20]

In navigating the landscape of fusion research, the interplay between machine learning (ML) techniques and Bayesian inference emerges as a pivotal force, offering novel avenues for addressing complex challenges. While BNs have yet to find widespread application as surrogates in fusion studies, the realm of ML, bolstered by Bayesian modeling techniques, presents a rich tapestry of possibilities. From the nuanced inference of plasma equilibria to the predictive power of deep learning methods like convolutional neural networks (CNNs), researchers navigate a diverse array of methodologies to elucidate the intricacies of fusion phenomena. As surrogate models expedite computations and DL models offer efficient alternatives to Bayesian inference, the fusion community stands at the precipice of transformative discovery, guided by the symbiotic relationship between traditional methodologies and cutting-edge computational techniques.

III. Methodology

In this section, aa set of general procedures for creating and applying a BN as a surrogate model is presented, drawing inspiration from [13] and leveraging insights from the proof-of-concept study [1]. Illustrated in Figure 3, a stepwise algorithm delineates the creation, cross-validation, and application of the BN surrogate model, applicable to various input-output problems. This algorithm has been refined and optimised from previous methodologies. In this study, steps are explored methodologically, focusing on identifying optimal deployment strategies such as specific hyperparameters and validation techniques. Throughout the explanation of these steps, specific adaptations and enhancements are highlighted to the previous approach, illustrating how they were implemented in the proof-of-concept study.

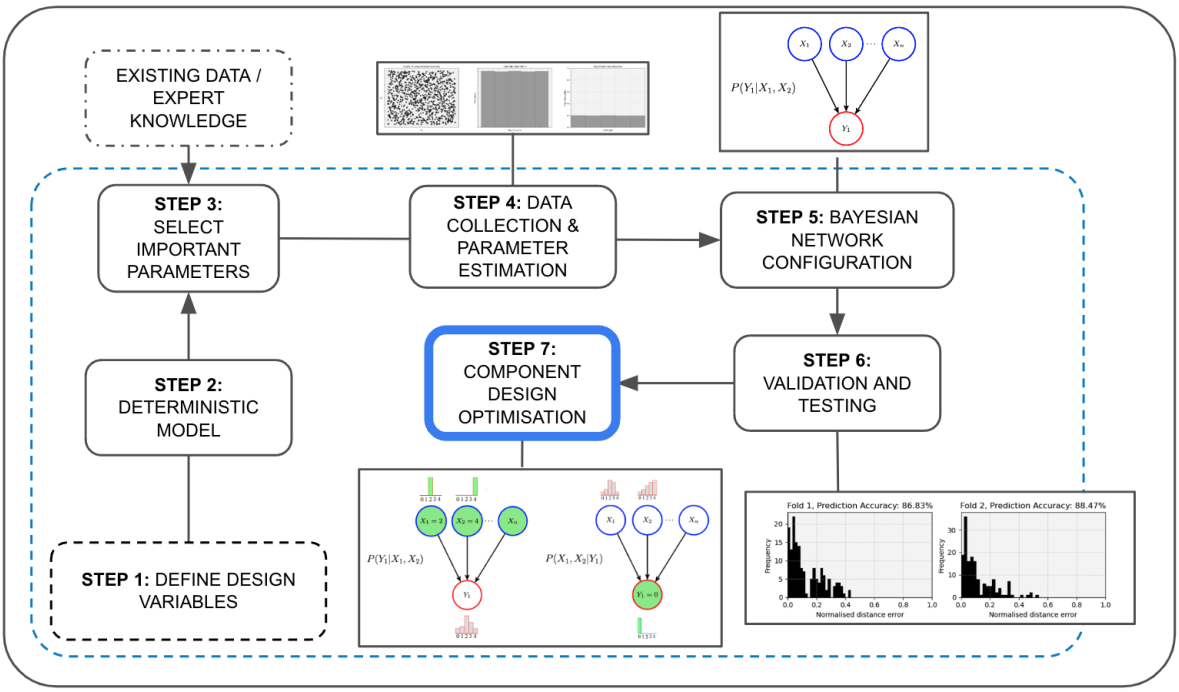


Fig. 3: Workflow of the surrogate modelling process, adapted from [13], and refined from the methodology presented in [1]

A. *Step 1: Define design variables*

In this step, a strategic decision is made regarding the focal point of the model, whether it targets a whole reactor system or a critical component such as the breeder blanket or the divertor. Following this decision, the key parameters and characteristics of the chosen component are then identified and defined as input variables, labeled as X_1, X_2, …, X_n. As exemplified in the proof-of-concept study conducted by Griffiths et al. [1], this process involved the conceptualization of a future-type spherical tokamak reactor system to conduct a comprehensive investigation into its economic viability. In this study, the inputs and outputs of the model were framed as the design variables, allowing for a holistic evaluation of the power plant's economics.

Ahead of executing the selected analytical model, it requires configuration. This process includes choosing the inputs and outputs, along with defining the value ranges for each variable. The selection of inputs and outputs is guided by the objectives of the analysis and the data at hand. The value ranges for each variable are set considering the system's physical limitations and the required granularity of the analysis. Once configured, the analytical model is run to produce the dataset, which is then utilised to set up the BN. Typically, BNs perform better at making predictions when the input distribution is uniform. This is for several reasons: model assumptions: normal distributions assume a bell-shaped curve with values concentrated around a mean, while uniform distributions make no assumptions about data shape and better capture the true underlying distribution. Nonlinearity and Outliers: normal distributions are sensitive to outliers, as they can significantly impact the mean and standard deviation, which may not accurately represent the underlying relationship between variables. Uniform distributions, being less influenced by extreme values, can provide a more robust representation of the data. Flexibility and Non-parametric Modelling: Uniform distributions offer flexibility in modelling unknown or non-normally distributed data without relying on specific distribution assumptions, allowing the BN to adapt and capture complex relationships. Simplification of model complexity: normal distributions require additional parameters (mean and standard deviation) that necessitate estimation, increasing model complexity and the number of parameters to learn. In contrast, using uniform distributions simplifies the model by eliminating the need for extra parameter estimation, resulting in easier training and interpretation [21, 22, 3].

B. *Step 2: Deterministic model*

The deterministic model serves as a foundation for the BN, incorporating a range of models rather than being restricted to a single one. Its primary function is to generate output variables, denoted as Y_1, Y_2, …, Y_m, which serve as the foundation for data collection and parameter estimation in developing the surrogate model. Drawing from the proof-of-concept study by Griffiths et al. [1], an illustrative example of this step involves the utilization of the UKAEA systems code, PROCESS [23, 24]. This model examines the interconnection between the engineering and physics subsystems comprising a fusion power plant while accommodating user-defined constraints.

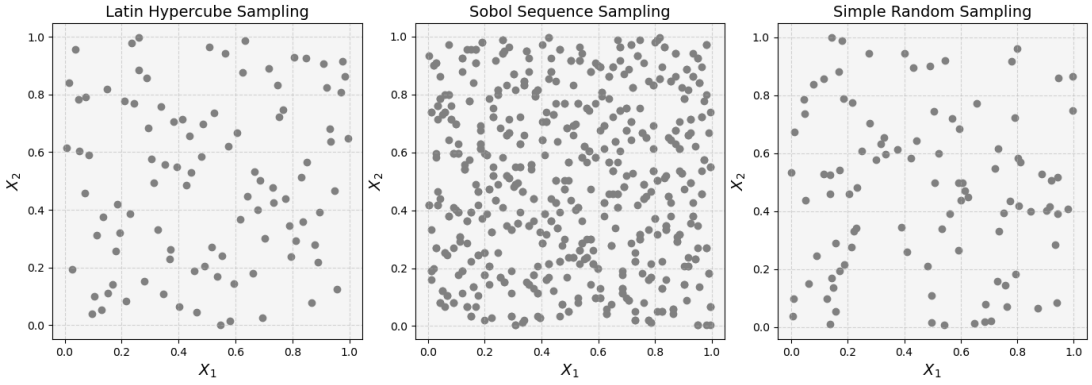


Fig. 4: A comparison of the Sobol and Latin Hypercube sampling methods.

C. ***Step 3***: *Parameter Selection*

In this stage, the focus is on identifying and selecting input parameters that wield substantial influence over the output variables modeled by the deterministic model. This task demands careful deliberation and expertise to discern the most pertinent parameters governing the system's behavior and performance.

In the study by Griffiths et al. [1], an example illustrates this process. They utilised an ST design basis from a previous work by the same authors, presenting a comprehensive techno-economic analysis of ST fusion power plants for hybrid hydrogen-electricity production [25]. To lay the groundwork for the techno-economic investigation, which encompassed evaluations under optimistic, moderate, and conservative scenarios, a rigorous sensitivity analysis was conducted. The objective was to streamline the problem's complexity by screening other potentially economically significant parameters.

Inputs were delineated as the reactor's design parameters, including the Greenwald fraction W impurity fraction Current drive efficiency factor Aspect ratio Toroidal field on axis [T] Total plasma beta NBI plug efficiency Turbine inlet temperature [ºC]. Outputs, on the other hand, were characterized as the economic parameters of the reactor, such as the capital cost [millions USD$]. The ranges for each variable were established considering the system's physical constraints, the required level of analysis granularity, and existing values derived from ST design bases documented in the literature.

D. ***Step 4***: *Data Collection and Parameter Estimation*

Here, the deterministic model is executed to gather data required to configure the surrogate model. It's imperative to collect comprehensive data representing the entire design space to ensure the surrogate model's accuracy. Specific sampling techniques such as Latin Hypercube or Sobol sampling are employed to ensure the collected data adequately covers the design space. Figure 4 illustrates the sampling of the input space. The collected data is then utilized to configure the surrogate model.

For example, the final phase of [25] focused on employing a Sobol analysis, which generated the dataset for [1]. The input space was sampled between the parameter limits using Saltelli's extension of the Sobol sequence [26, 27]; a quasi-random low-discrepancy sequence used to generate uniform samples of parameter space [28]. This analysis extensively investigated the influence of input parameters on the variance of the output, capital cost, by executing PROCESS $>$10,000 times. The analysis considered first-order (S_1), second-order (S_2), and complete interaction effects (S_{tot}), thus identifying key economic parameters and their interrelationships, conceptualising the design space of the nodes for the BN.

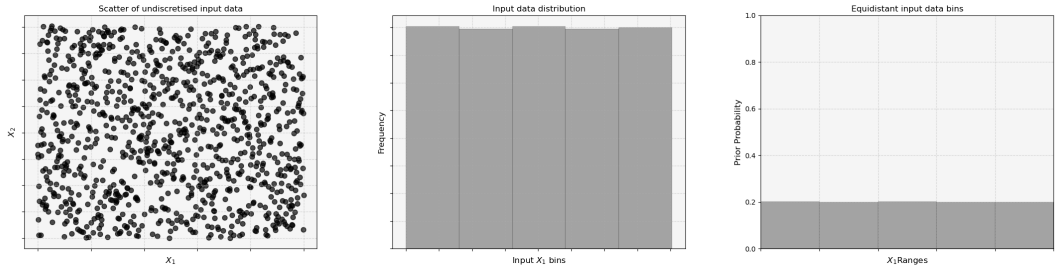
E. ***Step 5***: *Bayesian Network Configuration*

The BN is configured at this stage to mirror the deterministic model, using the inputs selected in Step III-C and the data collected in Step III-D. The data is discretised either by equidistant binning or by percentiles, for inputs and outputs respectively.

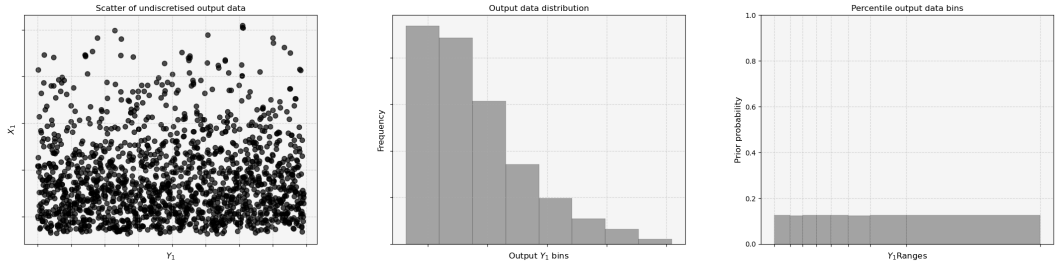
BNs are designed to handle discrete distributions, which means that datasets need to be discretised into bins before they can be used. This discretisation process is necessary to convert continuous data into discrete categories or intervals that the BN can handle effectively. By discretising the dataset into appropriate bins, the BN can leverage the discrete nature of the variables and make probabilistic inferences. Discretisation is a method used in machine learning to transform continuous variables into discrete categories or bins.

For example, in [1], two distinct methods of discretisation were employed: *equal distance* and *percentile binning*. For the inputs, the *equal distance* approach was applied, while the *percentile binning* method is implemented for the target. This differentiation allows for an effective and tailored discretisation process that accommodates the specific characteristics and requirements of the inputs and outputs.

Equal distance binning divides the range of a variable into a fixed number of equal-sized intervals or bins. This method assigns values to bins based on their proximity to the bin boundaries. Equal distance binning is commonly used for discretising input features in machine learning because it preserves the linear relationship between the variable values and allows for easy interpretation of the results. It can be particularly useful when the variable exhibits a linear trend or



(a) inputs



(b) outputs

Fig. 5: An example of the input distribution of a variable demonstrating the need for equidistant binning. An example of the output distribution of a variable demonstrating the need for percentile binning.

when the absolute values of the variable are important for the prediction task, making it the most appropriate method during the initial construction of the BN. Referring to Figure 5a, the input distribution of a variable demonstrates the need for equidistant binning.

On the other hand, percentile binning divides the data based on the distribution of the variable values. Each bin contains an equal number or percentage of data points, ensuring that the bins capture an approximately equal amount of information. Percentile binning is often preferred for discretising target or output variables in ML. This is because output data distributions are typically less uniform. It can thus help handle class imbalance issues, i.e., bias, and create more balanced categories for classification tasks. By grouping data points based on their relative positions in the distribution, percentile binning can ensure that each bin represents a similar portion of the target variable, reducing the impact of outliers and enhancing model performance. Referring to Figure 5b, the input distribution of a variable demonstrates the need for equidistant binning. Overall, the selection of the appropriate discretisation method should be based on an understanding of the data distribution, the specific machine learning task, and the goals of the analysis.

Once discretised into probability distributions, both inputs and outputs from training data configure the BN. Here the BN learns the parameters and develops the causal relationships between them, i.e., populates the conditional probability tables with the prior distributions of the variables. In this context, the pre-configured and validated BN can then be employed to perform bi-directional inference in Step III-G.

*F. **Step 6**: Model Validation and Testing*

Validation is the term used to measure model accuracy. As part of the validation procedure and in order to avoid bias, a partitioning technique was used to partition the dataset into training and testing in k distinct manners, known as k-fold cross-validation. This means the BN was built, and then validated k times, with k unique combinations of training and k testing sets. The ratio of the split between training and testing sets in each fold is determined by the number of folds used to split the dataset by $1/k$, e.g., 3 folds will have a test split size of $1/3 = 33\%$.

In order to quantify how well the model makes predictions compared to the analytical response, the predicted values were compared with the actual values in the testing data. In traditional machine learning evaluation methods like Root Mean Square Error (RMSE), the predictions and actual values are typically scalar values, such as numerical measurements or continuous variables. However, in the context of BNs, the outputs are probabilistic, distributions, or categorical variables rather than single numerical values. As a result, standard evaluation metrics like RMSE are not directly applicable, and alternative or custom approaches must be used to assess the performance of BN models. This can done either by calculating the difference between the mean of the predicted bin and the simulated value in the testing data, or by calculating the difference between the mean of the predicted bin and the mean of the bin in which the actual simulated value lies.

The NDE is a measure used to assess the relative magnitude of the distance error in relation to the range of values in a given bin. It helps to put the distance error into perspective and make it comparable across different bin ranges. By dividing the distance error by the difference between the maximum and minimum bin range values, a normalised distance error is

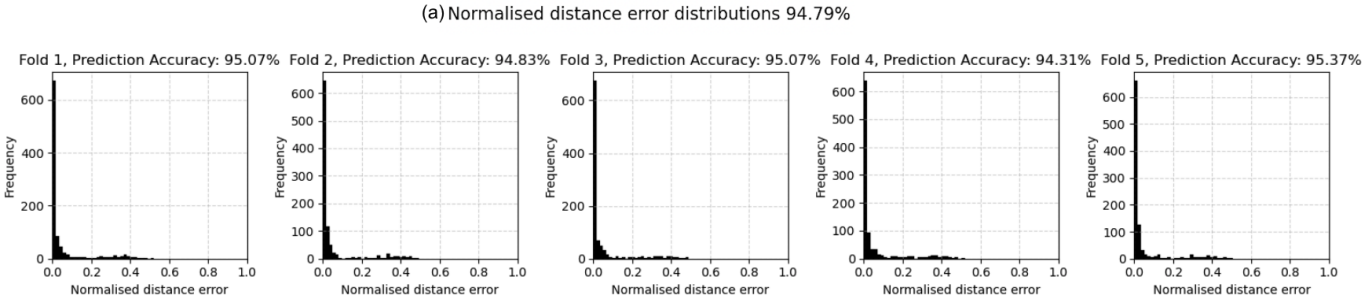


Fig. 6: Normalised probability histograms with prediction accuracy values for the first 5 folds is shown for d_1 bin resolution = 7, dataset size = 10240.

obtained that takes into account the scale and range of the data. This normalisation allows for a fair comparison of the distance error across different bin ranges, providing a standardised measure of the error [13].

1) Validation of the proof-of-concept surrogate

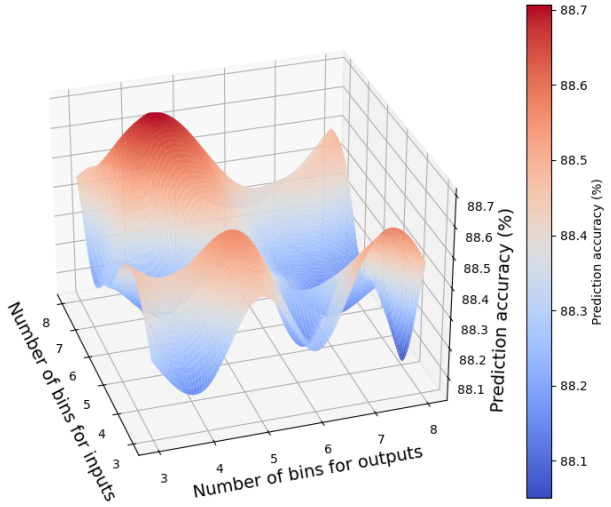


Fig. 7: A 3D surface plot displaying the average accuracy of predictions for d_1 errors while investigating variations in bin resolution.

This section details the implementation of Step III-F using the dataset from the proof-of-concept study [1]. Specifically, this section provides an overview of the outcomes obtained through k-fold cross-validation, and presents the findings from hyperparameter tuning. The validation process was conducted to assess the BN's predictive accuracy and reliability, and to identify the optimal hyperparameters for configuring the model. The results of the validation process provide valuable insights into the performance of the BN model and inform the selection of hyperparameters for future model development.

Using the data, calibration of hyper-parameters (such as bin resolution and dataset sizes) were examined by exploring two factors: (i) utilising three distinct dataset sizes and, (ii) experimenting with different bin resolutions during discretisation. The objective was to determine how modifying these paramters impacted the BN's ability to make predictions that could be usefully interpretted by the user. Consequently, for (i) this meant dataset sizes of 1400, 5120, and 10240 and, for (ii), incrementally increasing the number of bins used in discretisation between inputs and outputs for all combinations between 3 and 12, respectively. For (i), it was found that configuring the BN using the largest dataset resulted in the highest prediction accuracy. For (ii) the findings are presented in 7, illustrating that the model exhibited optimal performance when and input data was discretised into 7 bins and output data into 5 bins, respectively. In general, the average prediction accuracy for d_2 errors tends to be higher than that for d_1 errors, except in the case of using high bin resolutions.

To enable k-fold cross-validation, the dataset was split into 10 folds. 10 fold cross-validation aligns with the recommendation by Marcot et al. [29], who support the prevalent use of k=10 for BNs in literature. Plotting the NDE in a histogram provides a good illustration of model performance, see Figure 6. The resulting validation plot distributions in 6 offers a clear representation of the model's predictive capabilities. Overall, when observing the distribution of the d1 plots for all 10 folds, it can be noted that the BNM built on 10,000 data points, seems to predict the correct response with an average $\sim$95% accuracy.

This high level of accuracy underscores the model's robustness and reliability in making predictions. It also highlights the value of using a substantial dataset for building the model, as in [1]. These findings contribute to the growing body of evidence supporting the use of BNs in making accurate predictions, and they underscore the importance of careful model construction and validation.

The results from the hyper-parameter calibration provide valuable insights into the performance of the BN model. Two key factors were examined: dataset size and bin resolution during discretisation. The investigation found that the size of the dataset used to configure the BN significantly impacts its predictive accuracy. Specifically, the largest dataset (10,240) yielded the highest prediction accuracy. This suggests that the BN model benefits from a larger volume of data, likely because it provides a more comprehensive representation of the underlying distributions and dependencies. Whilst this is not a groundbreaking realisation, it is a useful reminder that the quality of the data used to configure the BN is crucial to its performance. The findings also highlight the importance of collecting a sufficient amount of data to ensure the model's

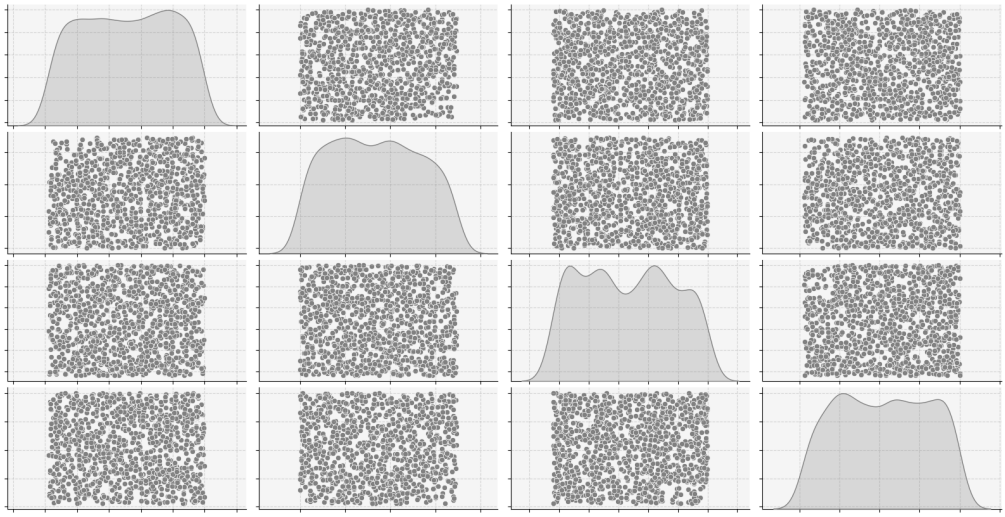


Fig. 8: A 4×4 scatter-grid to illustrate the sampling of an example input space. The scatter-grid is a visual representation where each point represents a unique combination of input values. The scatter-grid is a useful tool for visualising the distribution of the input space and identifying any gaps or biases in the data.

accuracy and reliability, and ensures that these models are not overfitting to the training data. Going forward, this study will inform the collection of data for future BN models, ensuring that the dataset is sufficiently large to support accurate and reliable predictions, especially when the model is used to make critical decisions in engineering design with fusion developers.

The bin resolution during discretisation also played a crucial role in the model's performance. The optimal performance was observed when the input data was discretised into 7 bins and the output data into 5 bins. This indicates a balance is needed in the bin resolution. Too few bins may oversimplify the data, losing important information, while too many bins can lead to overfitting, where the model becomes too tailored to the training data and performs poorly on new data.

Interestingly, the average prediction accuracy for d_2 errors was generally higher than that for d_1 errors, except when using high bin resolutions. This is largely down to how the model handles the different types of errors. The d_1 error measures the difference between the mean of the predicted bin and the simulated value in the testing data, while the d_2 error measures the difference between the mean of the predicted bin and the mean of the bin in which the actual simulated value lies. Thus, the likelihood that the correct bin is predicted is higher for d_2 errors, as it is based on the mean of the bin rather than the actual value. This is an important consideration when interpreting the results and understanding the model's performance. The results suggest that the model may be more sensitive to the bin resolution for d_1 errors, and care should be taken when choosing the bin resolution in this case.

These findings contribute to a deeper understanding of how hyper-parameter choices affect the performance of BN models, and can guide the selection of these parameters in future work by providing a clear indication of the optimal dataset size and bin resolution to benchmark the model. This will help to ensure that the BN model is configured to provide accurate and reliable predictions, and that the results can be interpreted with confidence.

G. ***Step 7**: Design Constraint Decision Support*

In this step application of the BN to provide decision support for engineering design is impleented through Bayesian bi-directional inference. Forward inference involves updating prior assumptions of inputs, utilising available evidence to give the output response. Conversely, reverse inference involves estimating the inputs given the output. The ability to perform reverse inference stems from the fact that once a BN has learned the parameters, it no longer differentiates between inputs and outputs. This characteristic allows the network to infer in both directions, making it possible to predict inputs based on output data., such as determining input ranges for optimal performance based on a desired output using reverse inference. This demonstrates an update from previous work with greater depth and analysis given.

IV. Case Study Appplication

In this section, the BN model is applied to a case study using the methodological framework outline in III to demonstrate its utility in providing decision support for engineering design. The case study focuses on the application of the BN to determine input ranges for optimal performance based on a specified output. This application is particularly relevant in the context of fusion reactor design, where engineers and researchers need to identify the input parameters that will yield the desired output performance. Section IV-A delves into the outcomes of reverse inference with the BN, emphasising the prediction of ranges for economic fusion parameters from uncertain data. Results can provide decision support for components, such as determining input ranges for optimal performance based on a desired output using reverse inference. This demonstrates an update from the proof-of-concept with greater depth and analysis given in Step 7 (III-G).

For **Step 1**: Define Design Variables (III-A), the aim was to model a whole reactor system, determining how important output parameters vary with certain inputs by studying their sensitivity to one another. For **Step 2**: (III-B), the deterministic model used was 'PyTOK', a systems code that finds optimum working point using a series of 0D and 1D approximations that model all subsystems. For **Step 3**: (III-C), the important parameters for the surrogate model were carefully selected based on insights from the PyTOK model, a comprehensive tool for fusion reactor analysis. The chosen input parameters include:

Major Radius (R): This parameter represents the distance from the center of the torus to the center of the plasma. In fusion terminology, it defines the size of the plasma confinement region and influences key plasma stability characteristics. A larger major radius typically allows for increased plasma volume, enhancing overall fusion performance.

Aspect Ratio (A): The aspect ratio signifies the ratio of the major radius (R) to the minor radius (a), which is the radius of the cross-sectional circle of the torus. Higher aspect ratios are characteristic of spherical tokamaks, a specific type of fusion device. Aspect ratio plays a crucial role in shaping the plasma equilibrium and can affect plasma stability and confinement properties.

Effective Ion Charge ($Z_{\mathbf{eff}}$): This parameter serves as a measure of the average charge state of ions within the plasma. It is calculated by summing the square of the charge of each ion species, weighted by their respective densities. A higher effective ion charge may indicate the presence of impurities in the plasma, which can lead to increased energy losses and impact fusion performance.

Toroidal Field on Plasma ($B_{\mathbf{T}}$): The toroidal field refers to the magnetic field that wraps around the torus in the direction of the major radius (R). It plays a crucial role in confining the plasma along the toroidal direction and is typically much stronger than the poloidal field, which confines the plasma in the radial direction. The strength of the toroidal field influences plasma stability, confinement, and overall fusion performance.

The chosen output parameters were:

Engineering Gain ($Q_{\mathbf{eng}}$): This factor is defined as the ratio of the power produced by the fusion reactions to the external heating power. Thus, a Q_{eng}=1 represents breakeven, where the power generated from fusion reactions in a plasma exceeds that of the external power needed to create the plasma state. Achieving a high Q_{eng} is indicative of the reactor's ability to sustain fusion reactions efficiently.

High Grade Waste Heat (H) **(MWt)**: H refers to the thermal energy harnessed in a thermodynamic cooling cycle by the breeder blanket for efficient conversion into usable heat or electricity. The heat can also be utilised for various industrial processes, district heating, or power generation, maximising the reactor's energy output and overall efficiency [30]. Efficient utilisation of waste heat maximizes the overall efficiency of the fusion reactor if non-electric applications can be implemented alongside electricity generation [25].

Net Electrical Output (E) **(MWe)**: This parameter quantifies the amount of electrical power generated by the fusion reactor. Achieving a high E signifies the effectiveness of the fusion reaction in producing usable electrical energy. Thus, an elevated E can contribute significantly to the grid, meeting energy demands to provide baseload power and ultimately reducing reliance on fossil fuels or other non-renewable sources.

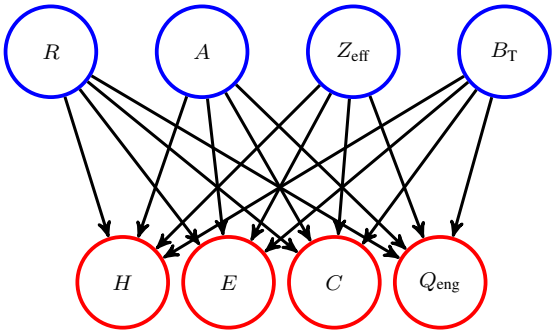


Fig. 9: Graphical representation of Bayesian Network where nodes represent the input variables (Major Radius R, Aspect Ratio A, Effective Ion Charge Z_{eff}, Toroidal Field on Plasma B_{T}) and the output variables (High Grade Wasteheat H, Net Electrical Output E, Capital Cost C, and Engineering Quality Q_{eng}) to the analytical model.

Capital Cost (C) **(million USD)**: This represents the overnight cost of building the fusion reactor. It is a crucial factor in assessing the economic feasibility and viability of implementing fusion energy technology.

For **Step 4**: (III-D), the input space between the parameter limits was sampled using the identical Sobol methods to that in[1]. Data was collected from the PyTOK model to produce a dataset size of 10,000. See Figure 8. For **Step 5**:, III-E, the BN was configured to replicate PyTOK using the inputs outlined in above for III-C, see Figure 9. The data was discretised using the optimal bin resolution of 7 for the inputs and 5 for the outputs, as determined in the hyper-parameter calibration. The BN was then validated using 10(k)-fold cross-validation to assess its predictive accuracy and reliability. The validation process provided valuable insights into the performance of the BN model and informed the selection of hyperparameters for future model development. See the Appendix A figure 12.

A. *Bayesian Reverse Inference*

Corresponding to **Step 7** in III, this section presents the outcomes of reverse inference, employed to predict the input parameters that are most likely to yield a specified output, such as minimising C. In this context, modelling a reactor with elevated E, H, and Q_{eng} within specific ranges is indicative of aiming for a high-performance reactor system.

C, was increased between each result, ranging from \$3.4–3.8 billion to \$4.4–6.5 billion. The aim of this approach was to illustrate the variation in the posterior distributions of the input parameters contingent on C. The outcomes of the reverse inference are depicted in Figures 10 and 11, which illustrate the posterior distributions of the fusion input parameters (highlighted in red) for the selected output parameter bins (in green). The findings suggest that for a reactor with a C of \$3.4–3.8 billion, and output parameters E of 100–180 MWe, H of 800–1100 MWt, and Q_{eng} of 1.70–2.00, the most probable input parameters are a R of 3.02–3.27 m, an Z_{eff} of

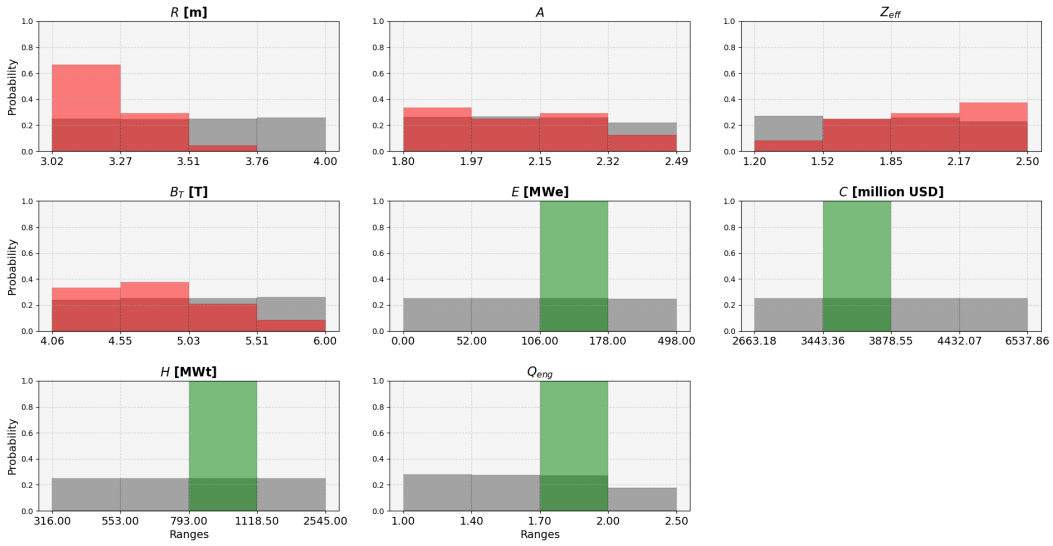


Fig. 10: Resulting posterior distributions depicted in red (variables: Major Radius R, Aspect Ratio A, Effective Ion Charge Z_{eff}, Toroidal Field on Plasma B_{T}) for reverse inference on the selected green bins (High Grade Wasteheat H, Net Electrical Output E, Capital Cost C, and Engineering Quality Q_{eng}) after updating the marginal prior beliefs (grey).

1.85.2.50 and a B_{T} of 4.06–5.03T. The posterior distribution of A is flatter than the other input parameters.

In contrast, for a reactor with a C of \$4.2–4.8 billion, and output parameters E of 80–150 MWe, H of 650–850MWt, and Q_{eng} of 1.70, the most probable input parameters are a R of 3.76–4.00 m, an A of 1.80–1.97, and a B_{T} of 5.51–6.00T. The posterior distributions of Z_{eff} is flatter than the other input parameters, with the largest posterior occuring between 1.20–1.52.

V. Discussion

A. Reverse Inference

The outcomes of the reverse inference analysis provide fusion developers with valuable insights into optimising the design parameters of fusion reactors and their components. The ability to identify the most probable input parameters for a given set of output parameters, can significantly inform the design process. By examining the posterior distributions of fusion input parameters in relation to specified output parameters, such as elevated values of energy output (E), fusion power (H), and Q_{eng}, developers can identify the most probable input parameter ranges that lead to high-performance reactor systems. The results highlight the ability to use the model across different stages of the design process, and it is expected to provide valuable insights for the development of future fusion power plants. In addition, the ambiguity of the model and its ability to handle data from any input-output model highlight the flexibility and adaptability of the BN, making it a valuable tool for fusion developers.

The analysis reveals a significant shift in the most probable input parameters for fusion reactor design, contingent upon the reactor's C. For reactors with a lower C (\$3.2-3.8 billion), optimal parameters tend towards a smaller R, higher Z_{eff}, and lower B_{T}, aligning with specified output parameter ranges. The posteriors of A illustrates a broader distribution compared to other input parameters. This suggests that for lower C machines, there is increased flexibility in selecting the desired value for A. However, as C is increased (\$4.4-6.5 billion), the most probable parameters shift, indicating a need for a larger R, reduced A, reduced s_{eff}, and increased B_{T} to meet the specified output parameters.

The observed decrease in A with increased C can be interpreted through the lens of engineering trade-offs. As C rises, design considerations may prioritise aspects like scalability and efficiency over compactness. A larger R accommodates more plasma volume, potentially enhancing overall reactor performance and output, albeit at the expense of increased material and construction costs. Moreover, a lower A may be favored in larger-scale reactors to optimise plasma stability and confinement, aligning with the operational demands of a higher-budget project.

This model-based insight underscores the intricate interplay between reactor design parameters and capital investment. The shift in optimal parameters with varying Cs highlights the nuanced considerations that developers must navigate to achieve cost-effective and high-performance fusion reactor designs. By interpreting these results, developers can strategically tailor design choices for reactor components, such as toroidal magnets and vacuum vessels, to optimise performance within budgetary constraints, fostering advancements in fusion energy technology.

These insights enable developers to make informed decisions during the design optimisation process. By focusing on the input parameter ranges associated with desired output parameters, developers can refine their reactor designs to enhance performance while considering cost constraints. Additionally, this information can guide the selection of input parameters

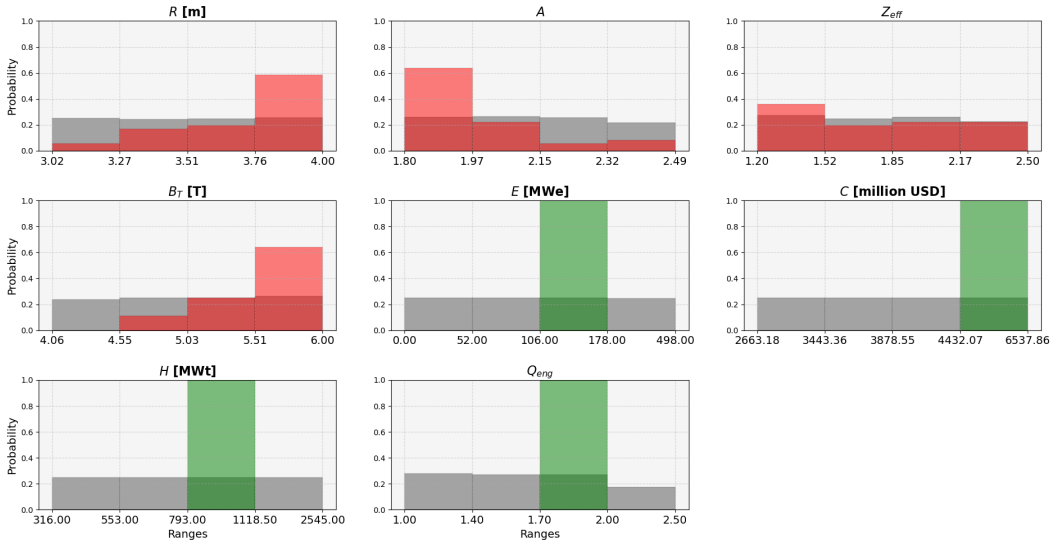


Fig. 11: Resulting posterior distributions depicted in red (variables: Major Radius R, Aspect Ratio A, Effective Ion Charge Z_{eff}, Toroidal Field on Plasma B_{T}) for reverse inference on the selected green bins (High Grade Wasteheat H, Net Electrical Output E, Capital Cost C, and Engineering Quality Q_{eng}) after updating the marginal prior beliefs (grey).

to achieve specific performance targets, such as maximising energy output or Q_{eng}, within budgetary limitations. However, it is important to note that these results are based on the current state of knowledge and technology. As such, it would be beneficial to continually update the Bayesian model as new data becomes available. This could include results from experimental reactors, advancements in plasma physics, or changes in manufacturing costs.

VI. Conclusion and Further Work

In summary, probabilistic Bayesian inference serves as a powerful tool for enhancing fusion research by providing a comprehensive framework for integrating data, quantifying uncertainties, optimising experimental design, supporting decision-making, and managing risks. By leveraging Bayesian methods in conjunction with existing approaches, fusion researchers can gain deeper insights into reactor behaviour, accelerate the pace of innovation, and advance the development of practical fusion energy solutions.

VII. Acknowledgments

This research was supported by the EPSRC (Engineering and Physical Sciences Research Council, UK) Nuclear Energy Futures Centre for Doctoral. Training in Nuclear Energy (NEF CDT). Other research studies under the NEF CDT involving Thomas Griffiths are supported in part by Tokamak Energy Ltd, UK. Views and opinions expressed are however those of the author(s) only a do not nsolidloosely dottedecessarily reflect those of Tokamak Energy Ltd.

References

[1] T. Griffiths, Z. X. Conti, J. Hidalgo-Salaverri, and M. Bluck, "Using bayesian networks as metamodels for predicting uncertain fusion economics in spherical tokamaks," *IEEE Transactions on Plasma Science*, pp. 1–7, 2024.

[2] D. J. Hand and K. Yu, "Idiot's bayes—not so stupid after all?" *International statistical review*, vol. 69, pp. 385–398, 2001.

[3] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[4] A. Pavone, A. Merlo, S. Kwak, and J. Svensson, "Machine learning and bayesian inference in nuclear fusion research: an overview," *Plasma Physics and Controlled Fusion*, 2023.

[5] J. Svensson, A. Dinklage, J. Geiger, and R. Fischer, "An integrated data analysis model for the w7-as stellarator," in *30th EPS Conference on Contr. Fusion and Plasma Phys., St. Petersburg*, vol. 27, 2003.

[6] J. Svensson, A. Dinklage, J. Geiger, A. Werner, and R. Fischer, "Integrating diagnostic data analysis for w 7-as using bayesian graphical models," *Review of Scientific Instruments*, vol. 75, no. 10, pp. 4219–4221, 2004.

[7] J. Svensson, *Non-parametric tomography using Gaussian processes*. EFDA, 2011.

[8] R. Fischer, L. Giannone, J. Illerhaus, P. McCarthy, R. McDermott, A. U. Team *et al.*, "Estimation and uncertainties of profiles and equilibria for fusion modeling codes," *Fusion Science and Technology*, vol. 76, no. 8, pp. 879–893, 2020.

[9] R. Fischer, C. Fuchs, B. Kurzan, W. Suttrop, E. Wolfrum, and A. U. Team, "Integrated data analysis of profile diagnostics at asdex upgrade," *Fusion science and technology*, vol. 58, no. 2, pp. 675–684, 2010.

[10] A. Pavone, J. Svensson, A. Langenberg, U. Höfel, S. Kwak, N. Pablant, R. Wolf *et al.*, "Neural network

approximation of bayesian models for the inference of ion and electron temperature profiles at w7-x," *Plasma Physics and Controlled Fusion*, vol. 61, no. 7, p. 075012, 2019.

[11] J. Degrave, F. Felici, J. Buchli, M. Neunert, B. Tracey, F. Carpanese, T. Ewalds, R. Hafner, A. Abdolmaleki, D. de Las Casas *et al.*, "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature*, vol. 602, no. 7897, pp. 414–419, 2022.

[12] J. Kates-Harbeck, A. Svyatkovskiy, and W. Tang, "Predicting disruptive instabilities in controlled fusion plasmas through deep learning," *Nature*, vol. 568, no. 7753, pp. 526–531, 2019.

[13] Z. X. Conti, T. Advisor, S. Dritsas, S. Lin, and S. Kaijima, "Bayesian network metamodels for inference-driven design space exploration submitted by architecture and sustainable design," 2019.

[14] K. L. van de Plassche, J. Citrin, C. Bourdelle, Y. Camenen, F. J. Casson, V. I. Dagnelie, F. Felici, A. Ho, S. Van Mulders, and J. Contributors, "Fast modeling of turbulent transport in fusion plasmas using neural networks," *Physics of Plasmas*, vol. 27, no. 2, 2020.

[15] J. Citrin, C. Bourdelle, F. J. Casson, C. Angioni, N. Bonanomi, Y. Camenen, X. Garbet, L. Garzotti, T. Görler, O. Gürcan *et al.*, "Tractable flux-driven temperature, density, and rotation profile evolution with the quasilinear gyrokinetic transport model qualikiz," *Plasma Physics and Controlled Fusion*, vol. 59, no. 12, p. 124005, 2017.

[16] O. Meneghini, S. P. Smith, P. B. Snyder, G. M. Staebler, J. Candy, E. Belli, L. Lao, M. Kostuk, T. Luce, T. Luda *et al.*, "Self-consistent core-pedestal transport simulations with neural network accelerated models," *Nuclear Fusion*, vol. 57, no. 8, p. 086034, 2017.

[17] M. A. Miller, R. M. Churchill, A. Dener, C.-S. Chang, T. Munson, and R. Hager, "Encoder–decoder neural network for solving the nonlinear fokker–planck–landau collision operator in xgc," *Journal of Plasma Physics*, vol. 87, no. 2, p. 905870211, 2021.

[18] A. Pavone, J. Svensson, A. Langenberg, N. Pablant, U. Hoefel, S. Kwak, R. C. Wolf *et al.*, "Bayesian uncertainty calculation in neural network inference of ion and electron temperature profiles at w7-x," *Review of Scientific Instruments*, vol. 89, no. 10, 2018.

[19] A. Pavone, J. Svensson, S. Kwak, M. Brix, R. Wolf, and J. Contributors, "Neural network approximated bayesian inference of edge electron density profiles at jet," *Plasma Physics and Controlled Fusion*, vol. 62, no. 4, p. 045019, 2020.

[20] A. Pavone, J. Svensson, M. Krychowiak, U. Hergenhahn, V. Winters, P. Kornejew, S. Kwak, U. Hoefel, R. Koenig, R. Wolf *et al.*, "Neural network surrogates of bayesian diagnostic models for fast inference of plasma parameters," *Review of Scientific Instruments*, vol. 92, no. 3, 2021.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification and scene analysis*. Wiley New York, 1973, vol. 3.

[22] R. E. Neapolitan, *Learning bayesian networks*. Pearson Prentice Hall Upper Saddle River, 2004, vol. 38.

[23] M. Kovari, R. Kemp, H. Lux, P. Knight, J. Morris, and D. J. Ward, "Process: A systems code for fusion power plants - part 1: Physics," vol. 89. Elsevier Ltd, 12 2014, pp. 3054–3069.

[24] M. Kovari, F. Fox, C. Harrington, R. Kembleton, P. Knight, H. Lux, and J. Morris, ""process": A systems code for fusion power plants - part 2: Engineering," *Fusion Engineering and Design*, vol. 104, pp. 9–20, 3 2016.

[25] J. Hidalgo-Salaverri, T. Griffiths, Z. X. Conti, P. Cano-Megias, R. Chacartegui, M. Bluck, J. Ayllon-Guerola, A. Mancini, M. Garcia-Munoz, and E. Viezzer, "Hybrid hydrogen-electricity production using spherical tokamaks: a cost driver sensitivity study and techno economic analysis," *Nuclear Fusion*, 9 2023.

[26] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, pp. 271–280, 2 2001.

[27] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications*, vol. 145, pp. 280–297, 5 2002.

[28] J. Herman and W. Usher, "Salib documentation release 1.4.7," 2023.

[29] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis?" *Computational Statistics*, vol. 36, pp. 2009–2031, 2021. [Online]. Available: https://doi.org/10.1007/s00180-020-00999-9

[30] T. Griffiths, R. Pearson, M. Bluck, and S. Takeda, "The commercialisation of fusion for the energy market: a review of socio-economic studies," *Progress in Energy*, vol. 4, p. 042008, 2022. [Online]. Available: https://dx.doi.org/10.1088/2516-1083/ac84bf

APPENDIX A
APPENDIX: PREDICTION ACCURACY

		Prediction Accuracy %					
		d_1			d_2		
Dataset size		1400	5120	10240	1400	5120	10240
Bin resolution	5	92.73	92.84	93.13	96.36	96.76	96.73
	7	93.74	94.08	94.79	96.98	96.79	97.00
	10	89.37	91.20	90.86	88.25	89.75	89.40

TABLE I: A table displaying the average accuracy of predictions for d_1 and d_2 errors while investigating variations in both (i) the size of the dataset and (ii) the resolution of bins.

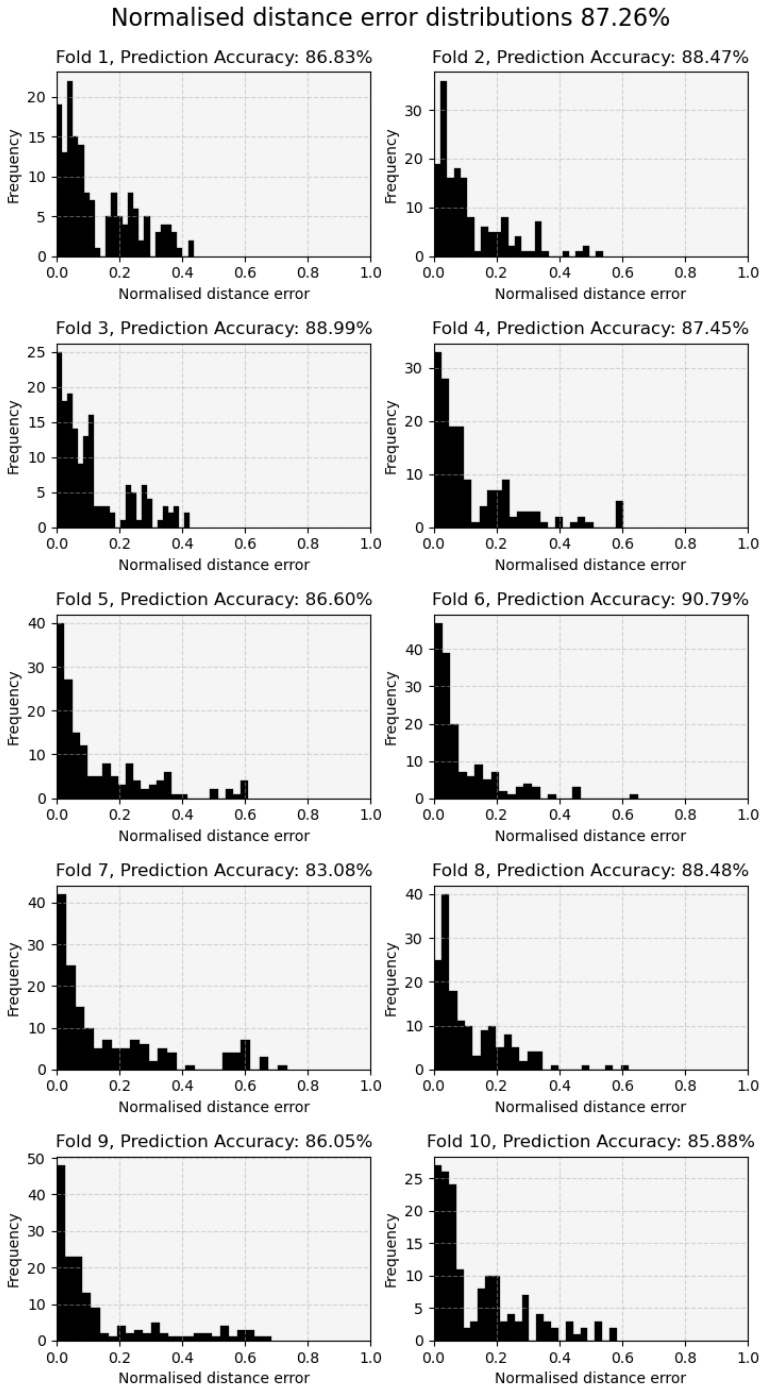


Fig. 12: Normalised probability histograms with prediction accuracy values for each fold is shown for d_1 bin configuration: inputs (7) outputs (5), dataset size = 10240.