

DYNABO: DYNAMIC MODEL BAYESIAN OPTIMIZATION FOR TOKAMAK CONTROL

Anonymous authors

Paper under double-blind review

ABSTRACT

Despite recent advances, state-of-the-art machine learning algorithms struggle considerably with control problems where data is scarce relative to model complexity. This problem is further exacerbated if the system changes over time, making past measurements less useful. While tools from reinforcement learning, supervised learning, and Bayesian optimization alleviate some of these issues, they do not address all of them at once. Considering these drawbacks, we present a multi-scale Bayesian optimization for fast and data-efficient decision-making. Our pipeline combines a high-frequency data-driven dynamics model with a low-frequency Gaussian process, resulting in a high-level model with a prior that is specifically tailored to the dynamics model setting. By updating the Gaussian process during Bayesian optimization, our method adapts rapidly to new data points, allowing us to process current high-quality data quickly, which is more representative of the system than past data. We apply our method to avoid tearing instabilities in a tokamak plasma, a control problem where modeling is difficult, and hardware changes potentially between experiments. Our approach is validated through offline testing on historical data and live experiments on the DIII-D tokamak. On the historical data, we show that our method outperforms a naive decision-making algorithm based exclusively on a recurrent neural network and past data. The live experiment corresponds to a high-performance plasma scenario with a high likelihood of instabilities. Despite this base configuration, we achieved a 50% success rate in the live experiment, representing an improvement of over 220% compared to historical data.

1 INTRODUCTION

Controlling real-world systems is generally a difficult task, even when powerful machine-learning tools are employed: nonlinearities are often pronounced, data is scarce, and safety issues impose severe limitations. A prime example of these issues is tokamak control, where good models are unavailable, safety is paramount, and instabilities are notoriously hard to control. These issues are further complicated by the fact that hardware configurations in tokamak change on a regular basis, making a model trained on past data even less reliable. However, despite these challenges, designing good control policies for tokamaks is highly desirable due to their promise to generate abundant clean energy via nuclear fusion.

In many real-world settings, model-free reinforcement learning is a promising solution and has seen successful applications (He et al., 2024; Kumar et al., 2021; Lee et al., 2020). However, most of these methods rely on a prohibitive amount of policy rollouts for training, which is typically only achievable with reliable simulation environments. In complex environments like tokamaks, this is particularly problematic, as operation costs typically only permit a handful of rollouts, and existing simulators do not reflect the true dynamics for many aspects of the plasma (Char et al., 2023a). Offline RL, seeks to overcome these issues by directly learning a policy from offline data which conservatively stays within bounds of observed data (Levine et al., 2020). However, the performance of offline RL methods depends crucially on high-quality expert data that contains advantageous states. If these are not present, then offline RL can suffer from extrapolation errors (Fujimoto et al., 2019). This is a major drawback for tokamak control, where significant exploration and improvement are still required to achieve energy production. Moreover, even offline RL is affected by the sim2real problem which is described in detail below.

054 Alternatively, model-based reinforcement learning offers a solution where dynamics models are
055 trained from historic data and rollouts from the model are then used for policy learning or planning
056 (Deisenroth & Rasmussen, 2011; Chua et al., 2018; Kaiser et al., 2019). In the past, machine learning
057 algorithms have been used to directly model plasma dynamics (Char et al., 2023b; Abbate et al.,
058 2021; Boyer et al., 2021). Reinforcement learning policies have also been trained in models trained
059 solely on fusion data (Char et al., 2023a; Wakatsuki et al., 2023; Degraeve et al., 2022). However, the
060 performance of these approaches crucially hinges on the assumption that the data faithfully captures
061 the model at test time. This is problematic in the case of tokamak dynamics, where time-dependent
062 model changes cannot be neglected. Though this issue can be potentially addressed by updating the
063 model with new data, the scarcity of experiments implies that too little data is typically produced to
064 reliably update the model.

065 In low-dimensional settings, the obstacles posed by conventional RL methods can potentially be
066 addressed by Bayesian optimization (BO). BO is a data-efficient tool for optimizing black box func-
067 tions (Garnett, 2023). By quantifying model uncertainty, BO achieves a tradeoff between exploration
068 and exploitation, leading to fast convergence in many practical settings (Shaloo et al., 2020; Shields
069 et al., 2021). In the case of tokamak control, BO has been used, e.g., to control the rampdown of
070 a real tokamak (Mehta et al., 2024), and to control neutral beams in a tokamak simulator (Char
071 et al., 2019). However, the work of Mehta et al. (2024) does not address critical plasma instabilities,
072 whereas Char et al. (2019) relies on a simulator. Moreover, these methods use a poorly specified
073 prior and require an extensive amount of experiments to perform well.

074 Motivated by the strengths and shortcomings of existing machine learning-based approaches for
075 tokamak control, we design a novel approach that combines a dynamic model predictor and Bayesian
076 Optimization. Our approach employs a multi-scale approach: a recurrent probabilistic neural net-
077 work models the high-frequency model dynamics, while a Gaussian process models the effect of
078 low-frequency marginal statistics on the dynamics. After adequate pre-processing, we use historical
079 data to train both models, where the dynamic model serves as a prior for the Gaussian process. Addi-
080 tionally, by leveraging physics-informed assumptions, we design a low-dimensional state space for
081 the Gaussian process. This naturally leads to a contextual Bayesian optimization algorithm tailored
082 to the task at hand, allowing it to find stabilizing actions in a highly data-efficient manner. Moreover,
083 due to its ability to perform fast updates, it allows us to efficiently leverage small batches of data
collected during experiments to best inform new decisions on the fly.

084 We test our approach on a large dataset from past tokamak experiments, where we can quickly iden-
085 tify stable configurations, outperforming a naive approach based exclusively on the recurrent neural
086 network model. Furthermore, we apply our approach to find stabilizing actions for a high performing
087 plasma scenario in the DIII-D tokamak. High performing plasma scenarios need to maintain high
088 temperature and pressures for increased energy, hence, they are more unstable. Our method was
089 able to find stabilizing ECH actuator values in four of eight experiments despite changes to other
090 actuators, a 133% improvement compared to historical experiments with the same configuration.

091 Our paper is structured as follows: first, we provide some necessary background to nuclear fusion,
092 and define our problem mathematically. Then we discuss our complete pipeline and methodology,
093 followed by the results and analysis on offline historical data and live experiments on a Tokamak
094 reactor. Finally, we provide conclusions and discuss opportunities for future work. Additional
095 details are provided in the Appendix.

096 097 2 BACKGROUND AND PROBLEM STATEMENT

098 In this section, we first provide some background on nuclear fusion and then present the formal
099 problem statement. We also include more details in the Appendix section.

100 101 2.1 NUCLEAR FUSION

102
103 Nuclear fusion is seen as a promising solution for clean, limitless energy, producing no high-level
104 radioactive waste. Among the fusion technologies, tokamaks are the most advanced, using magnetic
105 fields to confine hot plasma to enable fusion conditions. Many countries have invested in tokamak
106 research facilities and currently more 35 nations are collaborating to build ITER, a global project
107

108 aiming to demonstrate the viability of large-scale commercial fusion reactors (Mohamed et al., 2024;
109 Shimada et al., 2007).

110
111 One of the key challenges in tokamak development is plasma disruptions, which can cause se-
112 vere damage to reactor walls and components, particularly in larger reactors like ITER (Schuller,
113 1995; Lehnen et al., 2015). These disruptions often stem from tearing mode instabilities (or tearing
114 modes), where magnetic islands form, leading to energy loss and instability. Prior work proposes
115 avoiding tearing instability with predictive models using real-time control (Fu et al., 2020) and rein-
116 forcement learning (Seo et al., 2024). However, these methods reduce neutral beam power and add
117 torque to stabilize the plasma. This is undesirable, as reducing beam power leads to lower confine-
118 ment energy, decreasing the total energy output of the tokamak. On the other hand, adding torque
119 to large tokamaks is itself a challenging issue. Bardoczi et al. (2024) propose controlling tearing
120 modes utilizing differential rotation of the plasma. However, achieving differential rotation control
121 is itself a challenging problem.

122 **Though still poorly understood, Electron Cyclotron Heating (ECH) has shown promise in counter-**
123 **acting tearing instabilities by driving localized currents at the site of instability (Gantenbein et al.,**
124 **2000; Kolemen et al., 2014). These and other findings have motivated the inclusion of gyrotrons**
125 **capable of delivering ECH in future reactors to potentially control tearing instabilities, e.g., ITER**
126 **will have over 40 gyrotrons. So far, the best results for stabilizing instabilities with ECH have been**
127 **achieved by keeping the ECH constant over time, as this minimizes the chance of plasma disrup-**
128 **tions. However, how to best deploy ECH is still an open question. We provide further details on**
ECH effect on plasma in the appendix.

129 In this work, we aim to control ECH profiles to avoid tearing instability (or modes) in high q_{min}
130 tokamak scenarios. An ECH profile represents the heating achieved by the gyrotrons across the
131 cross section of the plasma. This can be seen in fig 4. High q_{min} is a scenario that supports long
132 duration steady-state plasma operations, making it crucial for future commercial fusion reactors.
133 More details on High q_{min} scenario are provided in the Appendix. We also focus our attention on
134 2-1 tearing instability, a type which is the most common and significantly disruptive.

135 2.2 PROBLEM STATEMENT

136 We treat the tokamak dynamics as an unknown discrete-time stochastic system

$$137 s_{t+1} \sim \Pi_{s_t, a_t}, \quad (1)$$

138 with states $s_t \in \mathcal{S}$ and actions $a_t \in \mathcal{A}$, and the probability of a tearing mode occurring follows a
139 Bernoulli distribution, parameterized by the tokamak states and actions

$$140 T_t \sim \text{Bernoulli}(p(s_t, a_t)). \quad (2)$$

141 Of the state variables describing the plasma, the most important for our approach is the normalized
142 plasma pressure $\beta_{N,t} \in s_t$. A full description of the state space is given in the appendix. The action
143 vector can be decomposed into three different sub-vectors

$$144 a_t := \left[a_t^f, a_t^c, a_t^g \right] \quad (3)$$

145 as follows. The actions a_t^f correspond to feedforward inputs specified before the experiment. These
146 correspond, e.g., to gas flows, plasma density, and shape controls. They are typically picked manu-
147 ally based on the success of previous experiments. The actions a_t^c are part of a feedback control loop
148 that aims to stabilize the normalized plasma pressure $\beta_{N,t} \in s_t$, arguably one of the most impor-
149 tant quantities since it measures the efficiency of plasma confinement relative to the magnetic field
150 strength. The third set of actions a_t^g corresponds to gyrotron angles, operated at constant power,
151 which we use to keep the tearing instability from occurring. The gyrotrons operate on the plasma
152 by generating an ECH profile $a_t^{\text{ech}} = \phi(a_t^g)$. Unlike a_t^f and a_t^c , the number of gyrotrons, i.e., the
153 dimension of a_t^g , potentially changes between each individual experiment. This is due to various
154 reasons, e.g., due to hardware issues or because some gyrotrons might be required for other tasks,
155 such as elm suppression or density control (Hu et al., 2024; Ono et al., 2024).

156 This paper considers the case where the gyrotron angles a_t^g are kept fixed throughout each experi-
157 ment rollout, i.e., $a_0^g = a_1^g = \dots = a_\tau^g =: a^g$, where τ is the length of the rollout horizon. This is

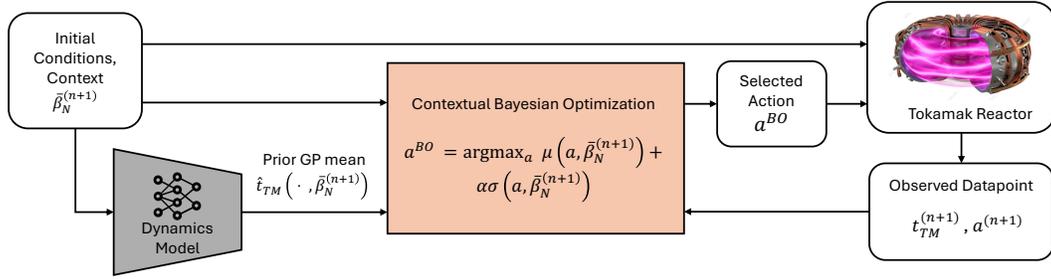


Figure 1: Overall Pipeline to generate trajectory actions. Initial conditions and feedforward actuators are used by RPNN to generate rollouts through which we compute the prior mean of the objective function (time to tearing instability). Our Bayesian optimization algorithm uses this to optimize for actions (ECH). Noisy outputs from the Tokamak are then used to update the Gaussian process model used for Bayesian optimization.

a common operating mode and also a design choice, which we make because we need to search as efficiently as possible within the action space, an impossible task if its dimension is too large. The feedforward actions a_t^f and the target normalized plasma pressure $\bar{\beta}_N$, which defines the set-point for a_t^c , are specified beforehand and can change between rollouts. Our goal is then to select a^g separately for each experiment such that the probability of encountering a tearing mode $T_t = 1$ is minimized over the full rollout horizon.

3 METHODOLOGY

We now introduce our method, which aims to find stationary ECH profiles that mitigate tearing instabilities. Our complete pipeline is illustrated in Fig. 1. On a high level, the process is as follows - We model the system at two different time scales to inform the choice of actuator commands for each experiment. At a smaller, more granular time scale, we use a recurrent probabilistic neural network model (RPNN) to estimate the high-frequency behavior during each experiment. The coarser model corresponds to a Gaussian process model and is trained to predict the behavior of the system based on marginal statistics from experimental observations and RPNN predictions of the objective function, which act as a prior mean. In this case, the objective function is the time-to-tearing instability. Given the target normalized plasma pressure $\bar{\beta}_N$, we leverage the Gaussian process to select actions (ECH profiles) in a low-dimensional space via Bayesian optimization. The desired profile is then converted to gyrotron angles and applied to the tokamak. Finally, we update our model with the resulting time-to-tearing instability and actual ECH profile and repeat the procedure. We update the model with the measured ECH profile because it can diverge significantly from the desired one. In the following sections, we describe the high-frequency RPNN, then the GP, and end with the full Bayesian optimization pipeline.

3.1 RECURRENT PROBABILISTIC NEURAL NETWORKS AND BINARY CLASSIFICATION

We employ a Recurrent Probabilistic Neural Network (RPNN) to model the high-frequency behavior of the tokamak. An RPNN has a Gated Rectifier Unit (GRU) cell, which stores information about past states and actions. The advantage of including a memory unit is that it allows us to model any unobserved variables that influence the state. To bypass the issue that the number of gyrotrons differs for each rollout in the training dataset, we assume that the resulting heating profiles a_t^{ech} can be controlled directly, allowing us to disregard a_t^g both in training and testing. When carrying out experiments, we then project a_t^{ech} onto a_t^g , which can be done for an arbitrary number of gyrotrons, i.e., for an arbitrary dimension of a_t^g .

Given s_t and a_t as inputs, our RPNN outputs a distribution over s_{t+1} as mean η , variance Σ^2 . The mean and variance specify a multivariate normal distribution, which we employ to approximate the system dynamics

$$\mathcal{N}(\eta(s_t, a_t), \Sigma^2(s_t, a_t)) \approx \Pi_{s_t, a_t}. \quad (4)$$

In addition to the RPNN, we train a classifier, which we call the tearing mode predictor h to predict the probability of a tearing mode occurring

$$h(s_t, a_t) \approx \text{Bernoulli}(p(s_t, a_t)). \quad (5)$$

3.2 GAUSSIAN PROCESS MODEL

Exclusively using RPNN for experimental design is challenging for various reasons. Although the RPNN accurately captures some of the tokamak behavior, the resulting predictions often exhibit significant errors, largely due to the sim2real gap caused by time-dependent fluctuations in the environment variables, e.g., due to maintenance or hardware changes provoked by previous experiments. Furthermore, retraining the RPNN between experiments and using it to select gyrotron angles a^g is virtually impossible because the newly collected data is too small and we only have a few minutes between experiments.

We address the above-mentioned issues by employing a Gaussian process (GP) model, a nonparametric model that is very data-efficient, especially in low-dimensional spaces (Deisenroth & Rasmussen, 2011). A GP corresponds to an infinite collection of random variables, of which any finite number is jointly normally distributed. To fully leverage the strengths of GP models, we need to carefully summarize the information collected between experiments before training the GP. This is done as follows.

First, we assume the achieved normalized plasma pressure β_N is independent of the ECH profile a_q . This is a reasonable assumption because β_N is largely determined by neutral beams, which are controlled through the feedback variables a_t^c . We then approximate the feedforward and feedback control actions a_1^f, \dots, a_τ^f and a_1^c, \dots, a_τ^c by assuming that they are uniquely specified by the target normalized plasma pressure, denoted by $\bar{\beta}_N$. This choice is partly justified because the feedforward and feedback control actions are often primarily informed by a target normalized plasma pressure. A further approximation we make is to project the ECH profile a^{ech} to a Gaussian curve, parametrized by the three-dimensional vector a_q containing the center, width, and height of the Gaussian curve. Finally, we employ the GP to predict the time-to-tearing mode t_{TM} , which we use as a proxy for the probability of a tearing mode occurring. The rationale behind this choice is twofold. First, a scenario where tearing instability occur late implies a higher degree of stability than a scenario where they occur earlier. Moreover, it allows us to use the GP in a regression setting, where GPs are strongest and best understood. The GP inputs are thus $\bar{\beta}_N$ and a_q , whereas the output is t_{TM} .

The GP is fully specified by a prior mean function m and a kernel k that specifies the similarity between training inputs. In this work, we employ a squared-exponential kernel k , which is appropriate for approximating most continuous functions. The mean function m corresponds to the average \hat{t}_{TM} predicted by the RPNN and tearing mode predictor,

$$\hat{t}_{\text{TM}}(\bar{\beta}_N, a_q) := \mathbb{E} \left(\arg \min_t \left| T_t \geq 0.5, \quad T_t \sim h(s_t, a_t), \quad s_{t+1} \sim \mathcal{N}(\eta(s_t, a_t), \Sigma^2(s_t, a_t)) \right) \right), \quad (6)$$

where we use the Gaussian curve specified by a_q to choose the ECH component of the actions a_1, \dots, a_τ . The feedforward and control actions a_t^c and a_t^f components of the actions are chosen based on the target $\bar{\beta}_N$ for the experiment. Given training data,

$$\mathcal{D}_n = \{\bar{\beta}_N^{(i)}, a_q^{(i)}, t_{\text{TM}}^{(i)}\}_{i=1, \dots, n},$$

obtained after appropriate pre-processing, we can compute the posterior distribution of t_{TM} for arbitrary test inputs $\bar{\beta}_N^*, a_q^*$, which corresponds to a normal distribution mean and covariance

$$\mu_n(\bar{\beta}_N^*, a_q^*) = \hat{t}_{\text{TM}}(\bar{\beta}_N^*, a_q^*) + k_*^\top (K + \sigma^2 I)^{-1} \Delta_n, \quad (7)$$

$$\sigma_n^2(\bar{\beta}_N^*, a_q^*) = k_{**} - k_*^\top (K + \sigma_{\text{no}}^2 I)^{-1} k_* + \sigma_{\text{no}}^2, \quad (8)$$

where σ_{no}^2 is the noise variance, $[k_*]_i = k(\bar{\beta}_N^*, a_q^*, \bar{\beta}_N^{(i)}, a_q^{(i)})$, $[K]_{ij} = k(\bar{\beta}_N^{(i)}, a_q^{(i)}, \bar{\beta}_N^{(j)}, a_q^{(j)})$, $k_{**} = k(\bar{\beta}_N^*, a_q^*, \bar{\beta}_N^*, a_q^*)$. The vector $[\Delta_n]_i = t_{\text{TM}}^{(i)} - \hat{t}_{\text{TM}}(\bar{\beta}_N^{(i)}, a_q^{(i)})$ contains the difference between the observed and the predicted time-to-tearing mode. In practice, the posterior variance σ_n^2 is

typically small when evaluated in distribution and larger when out of distribution. Hence, intuitively, the posterior GP mean μ_n can be viewed as the predictive model, whereas σ_n^2 quantifies model uncertainty. This distinction for understanding Bayesian optimization, which is introduced in the next section.

3.3 CONTEXTUAL BAYESIAN OPTIMIZATION WITH NOISY INPUTS

Contextual Bayesian optimization is a data-efficient tool that leverages GPs to optimize black-box functions. Given a context that specifies the environment, it optimizes an acquisition function that carefully balances exploration versus exploitation. By recursively updating the acquisition function after every observation, it gradually becomes more confident about its predictions, resulting in convergence. In every experiment, we treat the target normalized plasma pressure $\bar{\beta}_N^{(n+1)}$, specified before the experiment, as the context and choose the ECH profile by optimizing the so-called upper confidence bound (UCB) acquisition function

$$a_q^{\text{BO}} = \arg \max_{a_q} \mu_n(a_q, \bar{\beta}_N^{(n+1)}) + \alpha \sigma_n(a_q, \bar{\beta}_N^{(n+1)}), \quad (9)$$

where α balances exploration and exploitation. In conventional BO methods, the next step consists of setting $a_q^{(n+1)} = a_q^{\text{BO}}$, measuring the time-to-tearing mode $t_{\text{TM}}^{(i+1)}$, and updating the GP accordingly. However, in our setting there is the added challenge that the target plasma $\bar{\beta}_N$ and the desired ECH profile corresponding to a_q^{BO} is not reproduced exactly. This is due to the potentially changing number of available gyrotrons, actuator noise, and unmodeled disturbances. Number of gyrotrons is variable from experiment-to-experiment. To alleviate this, we instead measure the ECH profile obtained during the experiment and use it to determine $a_q^{(n+1)}$ before updating the GP model. Formally, this is equivalent to standard contextual BO where the GP inputs a_q in Equation 9 are perturbed by unknown noise.

4 RESULTS

This section presents results from offline tests using historical data and results from experiments at the General Atomics DIII-D Tokamak Fusion Facility. We use a fixed RPNN in all experiments, trained using 15,000 one-step state transition observations collected between 2010 and 2019 at the DIII-D tokamak.

Through our analysis of offline and online experiments, we aim to answer the following questions:

1. How does DynaBO compare to other baselines? How robust is it in terms of the choice of kernel?
2. Can DynaBO find heating profiles that avoid tearing instabilities altogether using only a handful of trials? If not, can it prolong the stable operation time of the plasma?

We address question 1 by doing running artificial experiments from offline data and comparing performance all methods. We then address question 2 with results from live experiments on the DIII-D Tokamak. As we show in the following, both questions have an affirmative answer.

4.1 OFFLINE DATA ANALYSIS

This section employs historical data from the DIII-D tokamak to compare DynaBO with several baselines. Specifically, we employ data from 281 past experiments carried out at the DIII-D tokamak between 2012 and 2023. We selected the data points based on their similarity with our live experiment, particularly the range of $\bar{\beta}_N$ and the high q_{min} specification. Appendix A.2.2 provides a detailed description of the selection procedure.

We employ the historical data to emulate our live experiment from Section 4.2. This is achieved as follows. At every time step, we sample the target plasma pressure $\bar{\beta}_N$ from a uniform distribution corresponding to the range of the historical data and condition DynaBO on $\bar{\beta}_N$. We then select a

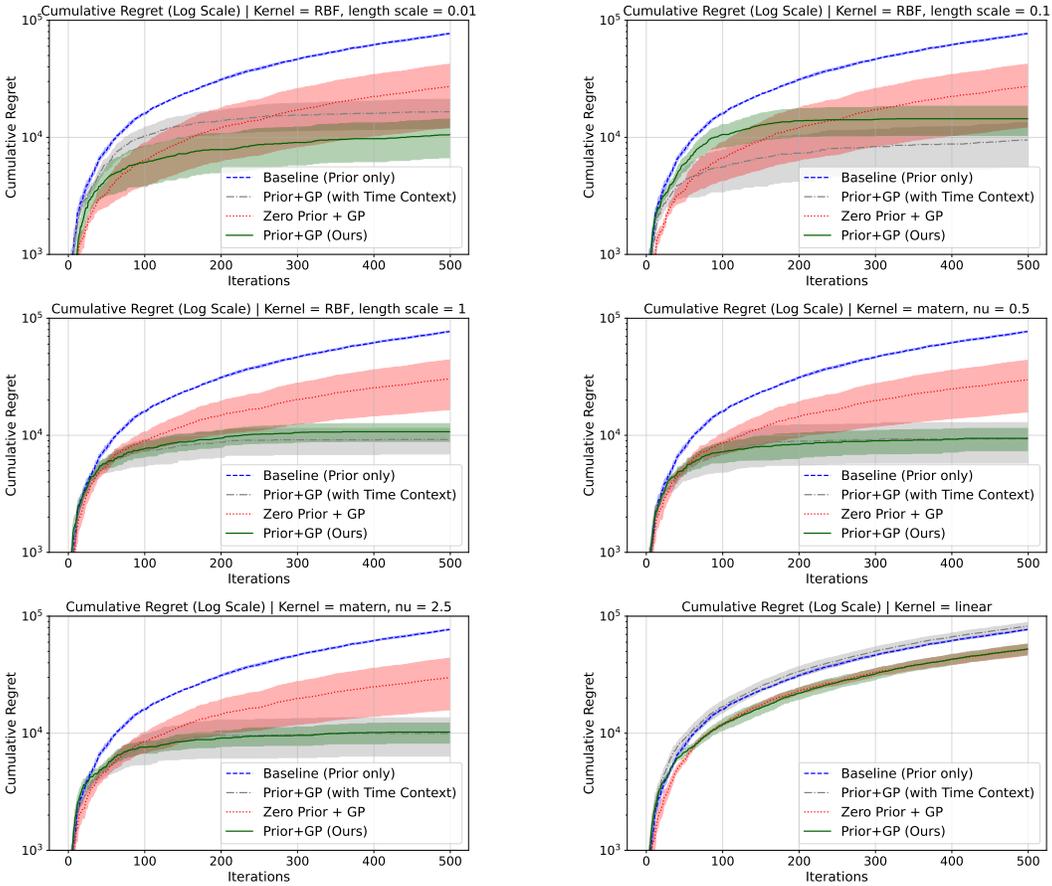


Figure 2: Cumulative Regret (log scale) achieved by DynaBO (green), DynaBO with a time-dependent kernel (gray), the RPNN only (blue), and a vanilla GP with a zero-mean prior (red) using six different kernels.

subset from the historical data with plasma pressure values within the interval $[\bar{\beta}_N - \epsilon, \bar{\beta}_N + \epsilon]$, where $\epsilon = 0.04$, and allow DynaBO to select an ECH value corresponding to an element of that subset. After selecting an ECH value, we treat the historical data point corresponding to that profile as a new observation, which we use to update our GP model.

We compare our approach to three different baselines: the setting where we fully trust the RPNN to predict tearing modes without updating it, a vanilla GP with a zero-mean prior, and our approach using a time-dependent kernel. The motivation for the latter approach is that older data is potentially less trustworthy for present-day experiments due to the number of changes made to the tokamak over time. In addition, we consider a linear kernel, Gaussian kernels, and Matérn kernels with different hyperparameter configurations.

In Fig.2, we depict the cumulative regret

$$\text{Cumulative Regret}(N) = \sum_{i=1}^N (\tau^{\max} - t_{\text{TM}}^{(i)}), \quad (10)$$

where $\tau^{\max} = 10s$ is the maximal shot length. As can be seen, DynaBO and DynaBO with time dependency achieve the highest performance in all settings except the linear kernel setting. By contrast, the RPNN-based method and vanilla BO cannot consistently find good solutions despite performing more steps than the total number of data. This indicates that DynaBO does not become overconfident and is robust to the choice of kernel and hyperparameters except when the kernel is clearly misspecified, e.g., when using a linear kernel. While including time as an input to the GP performs competitively, overall, the improvement seems only marginal. One possible explanation

378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431

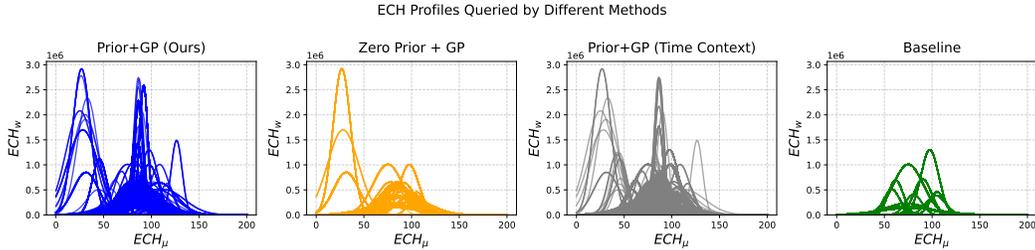


Figure 3: ECH Profiles queried by different methods during simulated offline runs using a Gaussian kernel. We see that DynaBO and DynaBO with a time-dependent GP explore the most, highlighting the importance of our dynamic model prior mean.

is that the reliability of the data depends on multiple factors, many of which cannot be explained exclusively as a continuous function of time, e.g., sensor and actuator upgrades, particle absorption and release by the tokamak wall, and the presence of impurities due to previous experiments.

We note that the vanilla GP does converge after more than 500 steps, i.e., after the plots in Fig. 2 end. However, such a long convergence time is unacceptable for our setting since fusion experiments are very costly, and we only get a handful of experiments to explore.

To compare exploration, in Fig. 3, we display the ECH profiles queried by the different baselines using a Gaussian kernel. DynaBO and DynaBO+time exhibit more variety in the queried ECH profiles than in the vanilla GP and the RPNN baseline. This corresponds to better exploration, resulting in lower regret for our approach. We observed similar trends using all other kernels except the linear one.

4.2 DIII-D TOKAMAK EXPERIMENTS

We tested our algorithm at the General Atomics DIII-D Tokamak during a two-hour time window allocated to us during the FY24 campaign. Each experiment run at DIII-D is known as a ‘shot’. Each shot is then assigned a unique shot number.

To make the most of our time and make significant statements about results, we opted for a pre-specified set of feedforward actuators a_t^f that is highly unstable, having a historical rate of tearing instability occurrences of 77%. We conditioned our GP on 125 historic high q_{min} experiments with similar configurations to the one we opted for on the day of the experiment. Appendix A.2 provides an overview of the data used to train the GP. Our experiment consisted of 8 BO iterations with DynaBO. After each run of DynaBO, the selected heating profiles a_q^{BO} were converted to gyrotron angles and entered into the Plasma Control System, the interface that controls the tokamak. After a few seconds of maintaining the plasma, we ramped down the actuators and terminated the shot.

Experiment ID (Shotnumber)	Target $\bar{\beta}_N$	Tearing Instability Avoided	Stability Time (ms)
199599	3.37	Yes	4566
199601	3.27	Yes	4632
199602	3.27	No	2107
199603	3.27	No	2149
199604	3.27	Yes	4592
199605	3.14	No	1512
199606	3.45	No	3512
199607	3.43	Yes	3654

Table 1: Results from two-hour experiments at DIII-D Tokamak : DynaBO avoids tearing modes in 4/8 runs in a high-performance configuration with a historical rate of occurrence of the tearing instabilities of 77%. The mean time under stability with DynaBO is 3339 ms while the historical time under stability is 2424 ms which represents an improvement of 914 ms in stability time. Generally, for stable experiments at DIII-D, the plasma stability is maintained for 4-5s.

We started our experiments by recreating a high-performing historical high q_{min} experiment with a tearing instability. For this, we recreated the conditions in shot 180636, a high-performance high q_{min} plasma trajectory executed previously at DIII-D. Once the shot was recreated, we ran additional shots where we varied the ECH using DynaBO while keeping the remaining settings identical. The details for each of the 8 shots carried out using DynaBO are shown in Table 1. As can be seen, our algorithm was able to successfully avoid tearing instabilities in 4 out of 8 shots. Moreover, we maintained a stable plasma for 3339 milliseconds on average. Although this number of shots is too low to be statistically significant, we stress that the chosen configuration is highly unstable. For reference, there were 61 historical experiments at DIII-D with very similar settings. Of those experiments, 47 reported tearing instabilities, corresponding to a tearing instability rate of 77%. These experiments include runs from session of 180636 and other sessions where similar high performance high q_{min} conditions were attempted. Moreover, the average time to tearing instability in those experiments is 2424 milliseconds, well below our average of 3339 milliseconds. More details on identification of tearing mode instabilities from raw signals is shown in Section A.3.

5 LIMITATIONS AND FUTURE WORK

While our method is shown to preemptively suppress tearing instabilities, it is mainly data-driven, and potential improvements are possible by incorporating physics knowledge. One possible solution is to develop physics-informed neural network models, such as incorporating elements of the Rutherford equation to improve interpretability. Another shortcoming is that our current method is only applicable to feedforward control scenarios. This means the algorithm cannot adapt to unexpected real-time changes in the plasma, such as MHD activity or impurity changes. In future work, we aim to extend our learning to feedback control systems.

6 CONCLUSION

In this work, motivated by the challenges of tokamak control, we develop a multi-scale modeling approach for making decisions on the fly using a handful of data. Our pipeline leverages a high-frequency neural network model of the system dynamics and a Gaussian process that makes predictions based on marginal statistics. Together, both models form a Bayesian optimization algorithm tailored to the task at hand that can quickly identify stabilizing control actions. This is achieved by making decisions on the fly based on newly collected data. On a historical data set, our method outperforms vanilla BO and a naive baseline that relies exclusively on neural network predictions. This is mainly due to our approach having better exploration capability. Moreover, our method shows promise in live experiments on the DIII-D Fusion reactor. During the experiments, our approach successfully avoided tearing instability in 4/8 runs despite highly unstable conditions, representing an improvement of over 117% percent compared to past experiments.

Our work illustrates the potential of combining complex high-frequency and low-frequency models to improve performance on the fly based on incoming data. In the field of nuclear fusion, the need for similar methods will increase in the future, as new and larger reactors such as ITER become operational, and a significant gap between existing and new models needs to be bridged with very little data. This is the case not only for the stabilization setting considered in this paper but also for settings such as ramp-up design, where a different set of actuators is considered. Moreover, we believe this approach could be of interest to several other applications where the discrepancy between past and present data is considerable, e.g., physical systems that exhibit wear and tear.

REFERENCES

- Joseph Abbate, Rory Conlin, and Egemen Kolemen. Data-driven profile prediction for diii-d. *Nuclear Fusion*, 61(4):046027, 2021.
- L. Bardoczi, N.J. Richner, N.C. Logan, E.J. Strait, C.T. Holcomb, J. Zhu, and C. Rea. The root cause of disruptive ntms and paths to stable operation in diii-d iter baseline scenario plasmas. *Nuclear Fusion*, 64(12):126005, sep 2024. doi: 10.1088/1741-4326/ad7787. URL <https://dx.doi.org/10.1088/1741-4326/ad7787>.

- 486 Mark Boyer, Josiah Wai, Mitchell Clement, Egemen Kolemen, Ian Char, Youngseog Chung, Willie
487 Neiswanger, and Jeff Schneider. Machine learning for tokamak scenario optimization: combining
488 accelerating physics models and empirical models. In *APS Division of Plasma Physics Meeting*
489 *Abstracts*, volume 2021, pp. PP11–164, 2021.
- 490 Ian Char, Youngseog Chung, Willie Neiswanger, Kirthevasan Kandasamy, Andrew Oakleigh Nel-
491 son, Mark Boyer, Egemen Kolemen, and Jeff Schneider. Offline contextual bayesian opti-
492 mization. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and
493 R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran
494 Associates, Inc., 2019. URL [https://proceedings.neurips.cc/paper_files/](https://proceedings.neurips.cc/paper_files/paper/2019/file/7876acb66640bad41f1e1371ef30c180-Paper.pdf)
495 [paper/2019/file/7876acb66640bad41f1e1371ef30c180-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/7876acb66640bad41f1e1371ef30c180-Paper.pdf).
- 496 Ian Char, Joseph Abbate, László Bardóczy, Mark Boyer, Youngseog Chung, Rory Conlin, Keith Er-
497 ickson, Viraj Mehta, Nathan Richner, Egemen Kolemen, et al. Offline model-based reinforcement
498 learning for tokamak control. In *Learning for Dynamics and Control Conference*, pp. 1357–1372.
499 PMLR, 2023a.
- 500 Ian Char, Youngseog Chung, Joseph Abbate, Egemen Kolemen, and Jeff Schneider. Full shot pre-
501 dictions for the diii-d tokamak via deep recurrent networks. In *APS Division of Plasma Physics*
502 *Meeting Abstracts*, volume 2023, pp. UP11–096, 2023b.
- 503 Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learn-
504 ing in a handful of trials using probabilistic dynamics models. *Advances in neural information*
505 *processing systems*, 31, 2018.
- 506 Jonas Degraeve, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco
507 Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de Las Casas, et al. Mag-
508 netic control of tokamak plasmas through deep reinforcement learning. *Nature*, 602(7897):414–
509 419, 2022.
- 510 Marc Deisenroth and Carl E Rasmussen. Pilco: A model-based and data-efficient approach to policy
511 search. In *Proceedings of the 28th International Conference on machine learning (ICML-11)*, pp.
512 465–472, 2011.
- 513 Yichen Fu, David Eldon, Keith Erickson, Kornee Kleijwegt, Leonard Lupin-Jimenez, Mark D Boyer,
514 Nick Eidietis, Nathaniel Barbour, Olivier Izacard, and Egemen Kolemen. Machine learning con-
515 trol for disruption and tearing mode avoidance. *Physics of Plasmas*, 27(2), 2020.
- 516 Scott Fujimoto, David Meger, and Doina Precup. Off-policy deep reinforcement learning without
517 exploration. In *International conference on machine learning*, pp. 2052–2062. PMLR, 2019.
- 518 Gunter Gantenbein, H Zohm, G Giruzzi, S Günter, F Leuterer, M Maraschek, J Meskat, Q Yu,
519 ASDEX Upgrade Team, et al. Complete suppression of neoclassical tearing modes with current
520 drive at the electron-cyclotron-resonance frequency in asdex upgrade tokamak. *Physical Review*
521 *Letters*, 85(6):1242, 2000.
- 522 Roman Garnett. *Bayesian optimization*. Cambridge University Press, 2023.
- 523 Tairan He, Chong Zhang, Wenli Xiao, Guanqi He, Changliu Liu, and Guanya Shi. Agile but safe:
524 Learning collision-free high-speed legged locomotion. *arXiv preprint arXiv:2401.17583*, 2024.
- 525 C.T. Holcomb, J.R. Ferron, T.C. Luce, T.W. Petrie, J.M. Park, F. Turco, M.A. Van Zeeland,
526 M. Okabayashi, C.T. Lasnier, J.M. Hanson, P.A. Politzter, Y. In, A.W. Hyatt, R.J. La Haye,
527 and M.J. Lanctot. Steady state scenario development with elevated minimum safety factor on
528 diii-d. *Nuclear Fusion*, 54(9):093009, aug 2014. doi: 10.1088/0029-5515/54/9/093009. URL
529 <https://dx.doi.org/10.1088/0029-5515/54/9/093009>.
- 530 Qiming M Hu, Nikolas C Logan, Qingquan Yu, and Alessandro Bortolon. Effects of edge-localized
531 electron cyclotron current drive on edge-localized mode suppression by resonant magnetic per-
532 turbations in diii-d. *Nuclear Fusion*, 64(4):046027, 2024.
- 533 Lukasz Kaiser, Mohammad Babaeizadeh, Piotr Milos, Blazej Osinski, Roy H Campbell, Konrad
534 Czechowski, Dumitru Erhan, Chelsea Finn, Piotr Kozakowski, Sergey Levine, et al. Model-based
535 reinforcement learning for atari. *arXiv preprint arXiv:1903.00374*, 2019.

- 540 Egemen Kolemen, AS Welander, RJ La Haye, NW Eidietis, DA Humphreys, J Lohr, V Noraky,
541 BG Penaflor, R Prater, and F Turco. State-of-the-art neoclassical tearing mode control in diii-
542 d using real-time steerable electron cyclotron current drive launchers. *Nuclear Fusion*, 54(7):
543 073020, 2014.
- 544
545 Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rma: Rapid motor adaptation for
546 legged robots. *arXiv preprint arXiv:2107.04034*, 2021.
- 547
548 Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning
549 quadrupedal locomotion over challenging terrain. *Science robotics*, 5(47):eabc5986, 2020.
- 550
551 Michael Lehnen, K Aleynikova, PB Aleynikov, DJ Campbell, P Drewelow, NW Eidietis, Yu Gas-
552 paryan, RS Granetz, Y Gribov, N Hartmann, et al. Disruptions in iter and strategies for their
553 control and mitigation. *Journal of Nuclear materials*, 463:39–48, 2015.
- 554
555 Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tuto-
556 rial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.
- 557
558 Viraj Mehta, Jayson Barr, Joseph Abbate, Mark D Boyer, Ian Char, Willie Neiswanger, Egemen
559 Kolemen, and Jeff Schneider. Automated experimental design of safe rampdowns via probabilistic
560 machine learning. *Nuclear Fusion*, 64(4):046014, 2024.
- 561
562 O. Meneghini, S.P. Smith, L.L. Lao, O. Izacard, Q. Ren, J.M. Park, J. Candy, Z. Wang, C.J.
563 Luna, V.A. Izzo, B.A. Grierson, P.B. Snyder, C. Holland, J. Penna, G. Lu, P. Raum, A. Mc-
564 Cubbin, D.M. Orlov, E.A. Belli, N.M. Ferraro, R. Prater, T.H. Osborne, A.D. Turnbull, and G.M.
565 Staebler. Integrated modeling applications for tokamak experiments with omfit. *Nuclear Fu-
566 sion*, 55(8):083008, 2015. URL [http://iopscience.iop.org/article/10.1088/
567 0029-5515/55/8/083008/meta](http://iopscience.iop.org/article/10.1088/0029-5515/55/8/083008/meta).
- 568
569 Mustakimah Mohamed, Nur Diyana Zakuan, Tengku Nur Adibah Tengku Hassan, Serene Sow Mun
570 Lock, and Azmi Mohd Shariff. Global development and readiness of nuclear fusion technology
571 as the alternative source for clean energy supply. *Sustainability*, 16(10):4089, 2024.
- 572
573 Masayuki Ono, John W Berkery, Nicola Bertelli, Syunichi Shiraiwa, Luis Delgado-Aparicio,
574 Jonathan E Menard, Álvaro Sánchez-Villar, Kajal Shah, Vladimir F Shevchenko, Hiroshi Idei,
575 et al. Efficient eccd non-inductive plasma current start-up, ramp-up, and sustainment for an st
576 fusion reactor. *Nuclear Fusion*, 2024.
- 577
578 FC Schuller. Disruptions in tokamaks. *Plasma Physics and Controlled Fusion*, 37(11A):A135,
579 1995.
- 580
581 Jaemin Seo, SangKyeun Kim, Azarakhsh Jalalvand, Rory Conlin, Andrew Rothstein, Joseph Ab-
582 bate, Keith Erickson, Josiah Wai, Ricardo Shousha, and Egemen Kolemen. Avoiding fusion
583 plasma tearing instability with deep reinforcement learning. *Nature*, 626(8000):746–751, 2024.
- 584
585 RJ Shalloo, SJD Dann, J-N Gruse, CID Underwood, AF Antoine, Christopher Arran, Michael Back-
586 house, CD Baird, MD Balcazar, Nicholas Bourgeois, et al. Automation and control of laser
587 wakefield accelerators using bayesian optimization. *Nature communications*, 11(1):6355, 2020.
- 588
589 Benjamin J Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I Martinez
590 Alvarado, Jacob M Janey, Ryan P Adams, and Abigail G Doyle. Bayesian reaction optimization
591 as a tool for chemical synthesis. *Nature*, 590(7844):89–96, 2021.
- 592
593 M Shimada, DJ Campbell, V Mukhovatov, M Fujiwara, N Kirneva, K Lackner, M Nagami, VD Pus-
594 tovitov, N Uckan, J Wesley, et al. Overview and summary. *Nuclear Fusion*, 47(6):S1, 2007.
- 595
596 T Wakatsuki, M Yoshida, E Narita, T Suzuki, and N Hayashi. Simultaneous control of safety factor
597 profile and normalized beta for jt-60sa using reinforcement learning. *Nuclear Fusion*, 63(7):
598 076017, 2023.

A APPENDIX

A.1 ADDITIONAL FUSION BACKGROUND

Effect of ECH on Plasma : Electron cyclotron heating waves released by gyrotrons interact with the plasma by being absorbed by electrons whose gyrofrequency matches the frequency of the ECH wave. This absorption is highly localized, making electron cyclotron heating (ECH) a powerful tool for precise plasma control. ECH increases the electron temperature at the absorption point and often reduces electron density, a phenomenon known as density pump-out (Wang et al., 2017). As density decreases, plasma rotation tends to speed up due to reduced inertia. By adjusting the toroidal injection angle of the EC wave, different effects can be achieved. When injected perpendicular to the toroidal direction, EC waves primarily heat the plasma. When injected parallel, they drive electron acceleration in the direction of the plasma current, a process called electron cyclotron current drive (ECCD). While the injection angle can be switched between shots, it cannot be actively adjusted during a single shot. In this experiment, ECCD is used for all shots because it effectively suppresses magnetic islands (Kolemen et al., 2014).

Details on high q_{min} plasma scenario : The high q_{min} scenario refers to a group of scenarios related with elevated values of q_{min} , the minimum value of the safety factor profile. Under this umbrella of scenarios, there are three main groups: $q_{min} = 1.4$, $q_{min} = 1.5 - 2$, and $q_{min} > 2$. The lowest of the range with $q_{min} = 1.4$ has shown promise as being stable to TMs, but did not have the greatest confinement, while the highest of $q_{min} > 2$ was stable to TMs but had lower energy confinement. The middle range of $q_{min} = 1.5 - 2$ has very desirable confinement but is very susceptible to TMs (Holcomb et al., 2014). The purpose of this experiment is to work in that middle range of q_{min} , referred to as the elevated q_{min} scenario, to stabilize TMs and achieve higher confinement than either of the other similar scenario options. In the elevated q_{min} scenario, 2/1 TMs are the most prevalent mode because they require the least energy to perform magnetic reconnection and form a magnetic island. Other lower order modes like 3/1 or 5/2 can sometimes occur but are significantly less frequent as they require more energy to form a magnetic island. Stabilizing TMs in the elevated q_{min} scenario would show a path forward for this high-confinement scenario as a possible operating scenario for a fusion power plant.

A.2 DATASET

Plasma trajectories on a Tokamak consists of three phases. The ramp-up phase, where the gases are heated and pressure is increased to generate the plasma state where fusion occurs. During this phase, the normalized plasma pressure β_N rises. Then, we enter the flat-top phase, where the plasma pressure β_N is sustained, allowing fusion to occur. In this phase, β_N is mostly constant and the aim to maintain this state without instabilities. Finally, the actuators are gradually ramped down and the plasma is safely terminated as the shot concludes. In this paper, we stay in the flat top phase and aim to stabilize it. To create our dataset, we hence use only flat top data and only control actuators during this phase of the experiment.

Our complete dataset contains of ~ 15000 plasma trajectories from historical experiments at DIII-D Tokamak. The data contains signals from different diagnostics have different dimensions and spatial resolutions, and the availability and target positions of each channel vary depending on the discharge condition. Therefore, the measured signals are preprocessed into structured data of the same dimension and spatial resolution using the profile reconstruction and equilibrium fitting (EFIT). These shots contain many different signals, some of which are described below. The dataset consists of scalar signals defined at every timestep and profile signals which are defined along 33 or 200 points along the radius of the plasma cross-section. These consists of temperature, ion temperature, pressure, rotation, safety (Q) factor and density. For these signals we first convert them into PCA components and select the top components which are able to explain 99% of the variance in data. The Electron Cyclotron Heating (ECH) profile we choose to control, is also defined at 200 points along the plasma radius. PCA is unable to describe ECH profiles, however they can be described well by a Gaussian curve and are hence parameterized by the center, width and amplitude of the curve. These 3 parameters form our parameterization a_q of the ECH profiles. The model state space s_t is shown in table 2 while the actuator space a_t is shown in table 3.

State Variables	Dimensions
Normalized Plasma Pressure β_N	Scalar
Line averaged density	Scalar
Loop voltage	Scalar
Confinement Energy	Scalar
Temperature Profile	Decomposed to 4 PCA components
Ion Temperature Profile	Decomposed to 4 PCA components
Density Profile	Decomposed to 4 PCA components
Rotation Profile	Decomposed to 4 PCA Components
Pressure Profile	Decomposed to 2 PCA components
q Profile (safety factor)	Decomposed to 2 PCA components

Table 2: Plasma Features used as state space for RPNN model.

Actuator Variables	Dimensions
Power Injected	Scalar
Torque Injected	Scalar
Target Current	Scalar
Target Density	Scalar
Magnetic Field	Scalar
Gas Puffing	Scalar
Shape Controls	6 Scalars
ECH Profile	Decomposed to Gaussian curve with mean, stddev, amplitude (μ, σ, w)

Table 3: Plasma Features used as actuator space of the RPNN model.

A.2.1 DATASET FOR DYNAMICS MODEL

For training the RPNN, we utilize this data set with data points every 20 ms in time intervals with trajectories having an average length of 5 seconds. The RPNN is trained to predict Δs_{t+1} given (s_t, a_t) . We add tearing mode labels to this dataset and train a random forest classifier to predict the probability of tearing modes at every time step. We tried incorporating tearing mode predictions inside the RPNN network however, we did not get good results. This is likely due to the formation of spurious correlations and causality issues formed by introducing tearing modes into the dataset.

A.2.2 DATASET FOR GAUSSIAN PROCESS

To create the dataset for offline testing \mathcal{D}^H , we first limit ourselves to High q_{min} trajectories having high β_N values > 3 . A figure of how the heating profile looks is shown in Fig. 4. These constraints follow our experiment conditions. This leaves us with 281 trajectories. We subsequently convert this data from a time-step scale to a trajectory level scale. We take average β_N of the flat-top phase of the trajectory. For ECH profile a_q , we take a mean of all profiles in the flat-top phase of the experiment. This is the phase where the high-energy plasma state is maintained. We thus get the dataset \mathcal{D}^H where D_i^H consists of triplet $(\beta_N^i, a_q^i, t_{TM}^i)$ i.e. the normalized plasma pressure, parameterized ECH profile and the observed time-to-tearing mode. This dataset is used for offline testing. For online testing, we subset this dataset further by only keeping whose ECH profiles are lie in the achievable parameter space as per experiment configuration, which leaves us with 125 training points. This is used as a training set for the Gaussian Process.

A.3 DETAILS ABOUT ONLINE EXPERIMENT RESULTS

In this section, we analyze the signals from our experiment runs at DIII-D. The results are shown in Fig. 5 and Fig. 6. We show the n1rms signal which measure the n=1 magnetic perturbations. We also show the normalized plasma pressure β_N , a quantity which directly corresponds to energy confinement levels in the plasma. For experiments 199606-199607, it is tricky to detect a tearing instability, hence we also add power injected. A sustained high value in n1rms along with drops in β_N usually denotes tearing modes. However, β_N drops may vary depending on severity.

702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

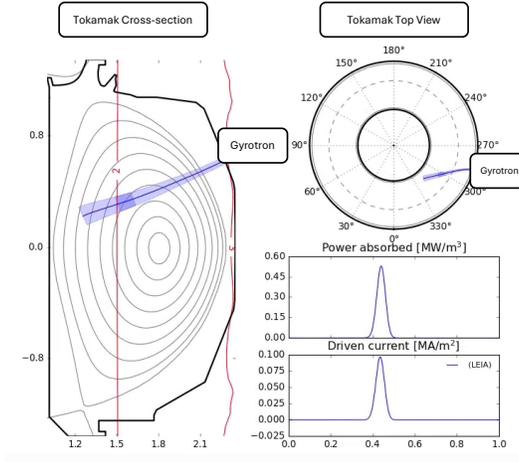


Figure 4: Gyrotron action on the Plasma inside the Tokamak. The bottom 2 curves indicate the power absorbed (heating profile) and current driven in the plasma from the centre to outer region of the plasma.

A.4 APPROXIMATING THE PRIOR

The historical data used to train the RPNN and the GP does not contain the target normalized plasma pressure $\bar{\beta}_N$. Instead, it only contains the actions a_t achieved during the shot. Similarly, the RPNN is trained exclusively on the actions, and not on $\bar{\beta}_N$, hence a direct mapping from $\bar{\beta}_N$ does not take place in the RPNN. In the experiments, we address these issues as follows. In the historical data, we set $\bar{\beta}_N$ to be equal to the average normalized plasma pressure, i.e.,

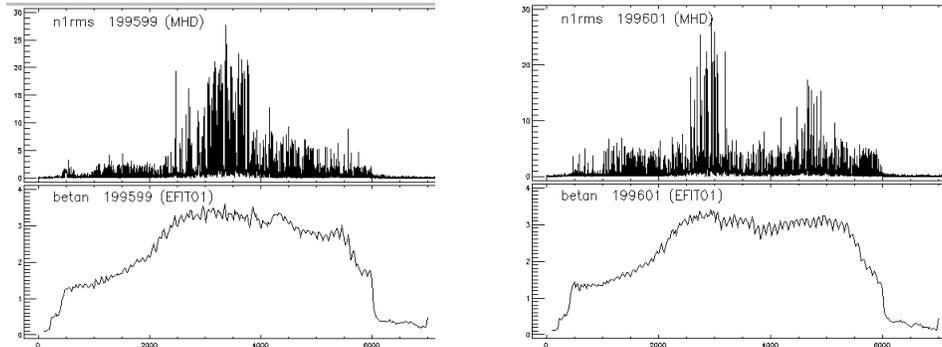
$$\bar{\beta}_N^{(i)} \approx \sum_{t=1}^{\tau} \beta_{N,t}^{(i)}. \quad (11)$$

This is a reasonable assumption since the target $\bar{\beta}_N$ is mostly achieved in practice. We then approximate the time-to-tearing mode $\hat{t}_{TM}(\bar{\beta}_N, a_q)$ predicted by the RPNN given $\bar{\beta}_N$ and a_q as follows. We first use a_q to compute the actions a_t^{ech} . We then compute the remaining actions a_t^c and a_t^f by sampling full rollouts from the historical data and setting a_t^c and a_t^f equal to the corresponding actions. We then look at the resulting average normalized plasma pressure and set it equal to $\bar{\beta}_N^{(i)}$. We do this for all ECH actions a_q within a $10 \times 10 \times 10$ grid within the space of ECH parameters, specified by the historically largest and smallest parameter values in the historical data set. We then separate the results into bins that have the same value of $\bar{\beta}_N^{(i)}$ up to a margin of $\epsilon = 0.04$, and average over all tearing modes within that bin, yielding $\hat{t}_{TM}(\bar{\beta}_N, a_q)$. At test time, we project all points to the closest point on the grid, both when performing queries and before updating the GP model.

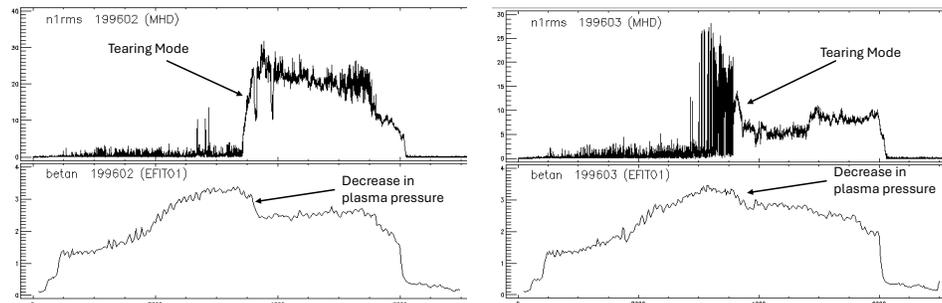
A.5 CONVERSION OF ECH PROFILE TO GYROTRON ANGLES

Even though we selected ECH profiles as our action space, the Plasma Control System (PCS) at DIII-D tokamak expects the output to be Gyrotron angles, which denote locations where they will be aimed. To make this conversion, we used OMFIT software (Meneghini et al., 2015). We selected ECH profiles as our action space instead of gyrotron angles because at experiment time one does not know how many gyrotrons are available. With this choice of action space, we ensure our method is agnostic of number of gyrotrons.

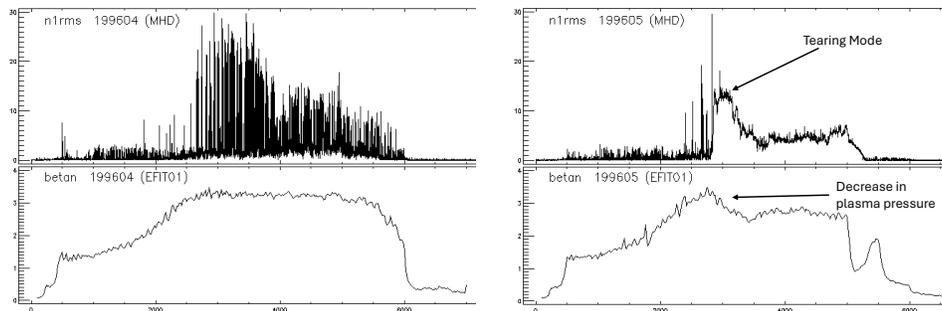
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809



(a) Shots 199599 & 199601 : No tearing modes were observed in these shots



(b) Clear Tearing mode happens in 199602 which leads to loss in normalized plasma pressure β_N . In 199603 we see a tearing mode form however its difficult to spot as the loss in β_N is minor.



(c) No tearing mode happens in 199604. A tearing mode happens in 199605

Figure 5: Identifying Tearing modes from Raw signal data. We use n1rms signal (denotes magnetic perturbations) and normalized plasma pressure β_N to identify tearing modes. A sustained high n1rms signal denotes tearing modes. We label the drop in β_N due to tearing mode formation. Note that in all experiments, β_N drops towards then end as power injected is dropped to safely end the experiment

810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

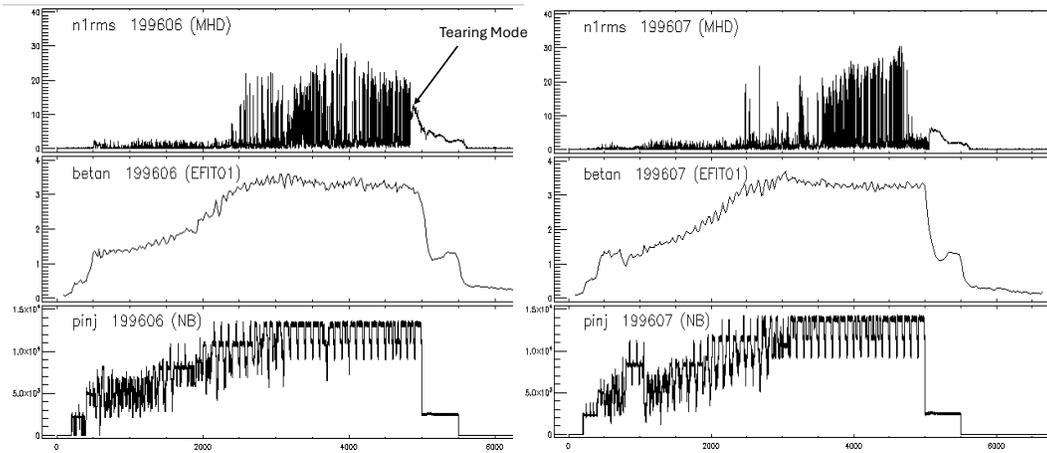


Figure 6: In this figure we also include the power injected ($pinj$) along with $n1rms$ and β_N . in 199606, we see a very late tearing mode which occurs just before power injected is dropped. Very low loss in β_N is seen due to the tearing mode. Finally, in 199607 no tearing modes are seen.