Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 53 (2021) 760–772

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

49th SME North American Manufacturing Research Conference, NAMRC 49, Ohio, USA

A Bayesian Framework for Milling Stability Prediction and Reverse Parameter Identification

Aaron Cornelius[a], Jaydeep Karandikar[b], Michael Gomez[a,b], Tony Schmitz[a,b,*]

[a] *Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville, Knoxville, TN 37996, USA*
[b] *Energy and Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA*

* Corresponding author. Tel.: +1-865-974-6141; fax: +1-865-974-5274. *E-mail address:* tony.schmitz@utk.edu

Abstract

This paper describes a physics-guided Bayesian framework for identifying the milling stability boundary and system parameters through iterative testing. Prior uncertainties for the parameters are identified through physical simulation and literature reviews, without physical testing of the actual milling system. Those uncertainties are then propagated to the stability map using a physics-based stability model, which is used to suggest a test point. The uncertainties are updated based on the new information acquired from the cutting test to form a new probability distribution, called the posterior. Finally, the posterior are compared to measured values for the stability boundary and system parameters to evaluate the approach. Based on experimental observations, the advantages and disadvantages of using a physics-guided model are discussed.

Keywords: Milling; modeling; stability; machine learning

Nomenclature

b	Axial depth of cut
b_{lim}	Predicted limiting depth of cut
c	Flexure viscous damping coefficient
k	Flexure stiffness
k_{nc}	Normal cutting force coefficient
k_{tc}	Tangential cutting force coefficient
m	Flexure mass
M	Once-per-tooth stability metric
$\boldsymbol{\theta}$	Probability distribution of input parameters
ω	Results of a specific cutting test
f_n	Flexure natural frequency
ζ	Flexure viscous damping ratio

1. Introduction

The importance of considering the system's vibration response when selecting machining parameters has been well-established in the literature (see [1], for example). If the tool tip and workpiece dynamics are not taken into account, then chatter may result, where regenerative vibrations are caused by varying chip thickness. Stability maps are a powerful tool for selecting the highest productivity milling parameters, where the rotating tool's tooth passing frequency is an integer fraction of the natural frequency that corresponds to the most flexible vibration mode. Frequency domain methods for calculating the stability maps using the tool tip frequency response function (FRF) and cutting force coefficients have been presented [1-2]. However, finding the FRF and cutting force coefficients can be challenging, requiring specialized tools and an experienced operator to perform the measurements. These requirements have prevented industry-wide adoption of the stability map

2351-9789
10.1016/j.promfg.2021.06.073

method for parameter selection.

Several papers have studied stability map identification through iterative testing without the need for direct measurements. Karandikar *et al.* described Bayesian updating methods based on Brownian motion random walks or Gaussian processes to update the probability of stability using test results, as well as algorithms for selecting test parameters based on a probabilistic stability map [3]-[4]. The random walk Bayesian updating method gives fast convergence and is computationally efficient, but does not provide information about the underlying system parameters, including the FRF and cutting force coefficients. Furthermore, the method is not physics-based and incorporates only minimal domain knowledge in its updating. Li *et al.* developed another Bayesian method based on propagating and updating input uncertainties using an ensemble Markov Chain Monte Carlo technique [5]. However, their technique requires a large number of samples which imposes significant calculation time. Additionally, Suzuki *et al.* used stability testing in an inverse analysis to identify the tool tip FRF [6].

This paper describes a unified framework for both identifying the optimal milling parameters and improving the milling model using a physics-based Bayesian updating approach. The framework is described in Figure 1. The major steps are summarized here.

- Establish the prior uncertainties for the FRF modal parameters and cutting force coefficients using simulations and literature reviews.
- Propagate the input uncertainties through the physics-based model to create a probabilistic stability map.
- Use the stability map to select a milling parameters to test.
- Perform the cutting test and classify it as stable or unstable.
- Define a new posterior distribution based on the results of the cutting test.
- Draw samples from the posterior distribution using a Markov Chain Monte Carlo method.
- Update the model input distributions and calculate the posterior stability map by propagating those samples through the physics-based model.

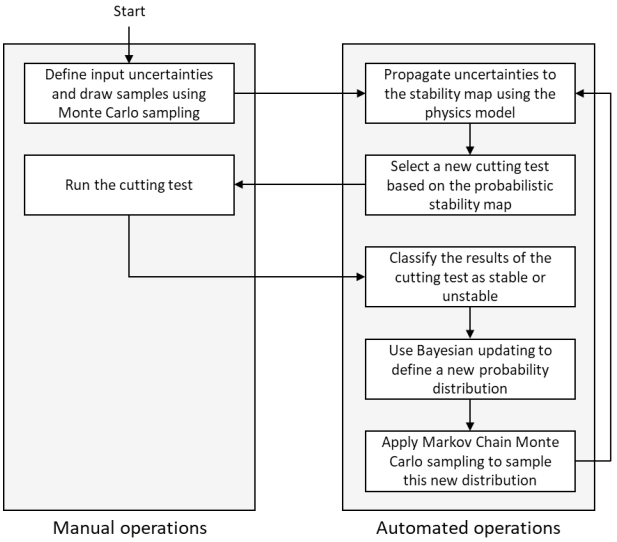


Figure 1: Overview of the Bayesian stability framework.

This paper provides four novel contributions.

- Prior uncertainties are established based on objective analysis, via uncertainty propagation or literature review. This reduces the need for user guidance or expertise in prior selection.
- A combination of rejection sampling and adaptive proposal Markov Chain Monte Carlo sampling is used to approximate the posterior distribution. This method improves mixing and achieves convergence in a smaller number of samples, increasing computational efficiency.
- Using a physics-guided model to propagate uncertainty and an expected-value algorithm for test selection results in faster convergence to the optimal milling parameters.
- The posterior estimates for the FRF and cutting coefficients are compared to measured values to evaluate whether the posterior distributions of the FRF and cutting force parameters converge towards their measured values.

The new Bayesian updating method is applied to an example machining setup on a Haas VF-4, shown in Figure 2. The cutting tool was a single flute inserted endmill (Kennametal part number M1D062E1401W075L150, 15.88 mm cutting diameter, 19.05 mm shank diameter) with a coated carbide insert (part number EC1402FLDJ). The 6061-T6 aluminum workpiece was mounted on a parallelogram-type flexure, shown in Figure 2. The flexure's displacement during cutting was measured using a capacitance probe and tool rotation was measured with a laser tachometer. All test cuts were performed with a fixed radial stepover of 5 mm and a feed per tooth of 0.1 mm. The spindle speed and depth of cut were varied from 2500 rpm to 7000 rpm and 0 mm to 10 mm, respectively.

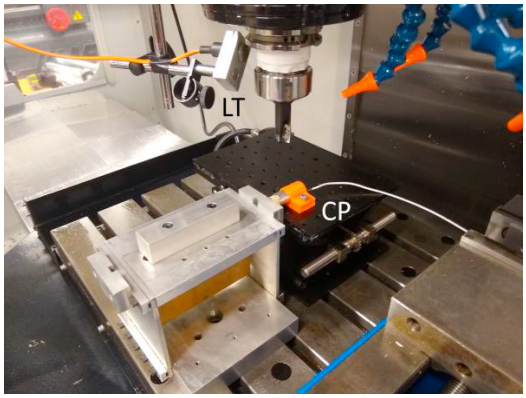


Figure 2: Machining test set up with capacitance probe (CP) and laser tachometer (LT).

The paper is organized as follows:

- In section 2, the prior uncertainties are established and propagated through the physics model to create a prior probabilistic stability map. This prior is compared to the results of an experimental grid test.
- Section 3 describes how Bayesian updating can be used to evaluate the posterior distribution based on cutting test results. The prior distributions are updated based on the experimental grid test using rejection sampling

- Section 4 details the adaptive proposal Markov Chain Monte Carlo technique used to sample the posterior distribution. This is applied to generate samples from the target posterior distribution after grid testing is completed.
- Section 5 describes the expected value method used to select test parameters.
- Section 6 presents the results of the Bayesian optimization framework with the automated test selection.
- Section 7 discusses the results, advantages, and limitations of the physics-guided model.

2. Establishing the prior

The first requirement for establishing a Bayesian model is to determine the *prior*. The prior represents one's current beliefs about the variables using all available information before any new cutting tests have been performed. Five inputs with associated uncertainties are considered in this effort:

- tangential cutting force coefficient k_{tc},
- normal cutting force coefficient k_{nc},
- flexure stiffness k,
- flexure mass m, and
- flexure viscous damping coefficient c.

Uncertainty distributions were established for each of the variables. Each uncertainty was defined as a normal distribution. The parameters and their measured values are listed in Table 1. These were obtained using the following approaches. These methods were selected to demonstrate how a prior can be established without calibration experiments, but other methods can be used depending on what information is known about the system.

- The priors for the cutting force coefficients were found from a literature review of reported force coefficients for milling 6061-T6 aluminum. Twenty different sets of coefficients were taken from five different sources, and the means and standard deviations for k_{tc} and k_{nc} were used as the initial prior [7]-[11]. The "true" values were measured using a cutting force dynamometer by performing a linear regression to the mean cutting force at different feed per tooth values.
- The priors for the flexure mass m and stiffness k were determined through: 1) beam theory predictions for parallelogram, leaf-type flexures [12]; and 2) Monte Carlo sampling to incorporate uncertainties in the analytical model inputs, including flexure dimensions (thickness and length of each of the flexure leaves), elastic modulus, and density. The true values were measured by tap testing with an instrumented hammer and low-mass accelerometer.
- The viscous damping coefficient c distribution was selected based on experience with similar flexures. It was calculated by assuming a mean viscous damping ratio of 1.4% with a standard deviation of 0.14% and propagating that uncertainty to the damping coefficient via Monte Carlo sampling of the mass and stiffness distributions. The true value was measured by tap testing.

Table 1: Input uncertainties

Parameter	Prior distribution ($\mu \pm 1\sigma$)	Measured value	Basis for the prior
$k_{tc}\ \left(\frac{\mathrm{N}}{\mathrm{mm}^2}\right)$	734 ± 83.4	903.9	Literature review
$k_{nc}\ \left(\frac{\mathrm{N}}{\mathrm{mm}^2}\right)$	346 ± 67.3	534.5	Literature review
$k\ \left(\frac{\mathrm{N}}{\mu\mathrm{m}}\right)$	2.17 ± 0.266	1.64	Monte Carlo sampling
m (kg)	2.75 ± 0.057	2.51	Monte Carlo sampling
$c\ \left(\frac{\mathrm{N}\cdot\mathrm{s}}{\mathrm{m}}\right)$	68.3 ± 6.47	114	Author experience, Monte Carlo sampling

The Table 1 entries can be thought of as a series of five univariate normal distributions with standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_5$ and mean values $\mu_1, \mu_2, \ldots, \mu_5$. Sampling from each of these distributions gives a series of values $\theta_1, \theta_2, \ldots, \theta_5$. However, it is more useful to consider the inputs as a single five dimensional multivariate normal distribution $\boldsymbol{\theta} \sim \mathrm{Normal}_5(\boldsymbol{\mu}, \boldsymbol{S})$ with mean vector μ, covariance matrix S, and output vector θ:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_5 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_5 \end{bmatrix}, \boldsymbol{S} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_5^2 \end{bmatrix}, \tag{1}$$

where S is a symmetric matrix which defines the variance and covariance of the output vector θ. The diagonal of the covariance matrix defines the variance (which is the standard deviation squared) for each component in the output vector. The other components give the covariance between the different elements: for example, $\boldsymbol{S}_{12}$ is the covariance between components 1 and 2 (and since the matrix is symmetric, $\boldsymbol{S}_{12} = \boldsymbol{S}_{21}$). This covariance defines the correlation between the components: a positive value indicates that higher values of θ_1 tend to be correlated with higher values of θ_2, while a negative value for $\boldsymbol{S}_{12}$ would indicate that as θ_1 increases, θ_2 tends to decrease. Since all non-diagonal components are 0 in $\boldsymbol{S}$, this indicates that each element of θ is independent; there is no correlation between them.

One thousand (1000) samples were drawn from this input distributions. Figure 3 shows the distributions and correlations of the individual variables listed in Table 1 after 1000 samples θ_1 through θ_{1000} were drawn from the prior. Since the non-diagonal elements of the covariance matrix were zero, the distributions of the variables are independent.

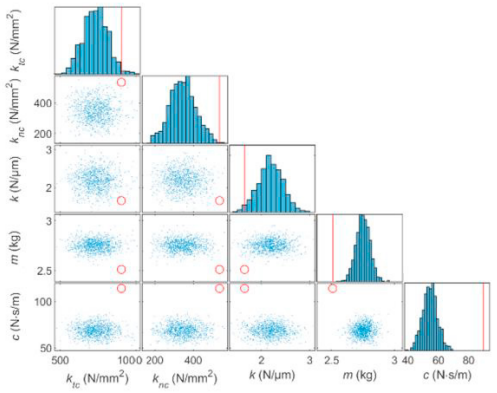


Figure 3: Initial distributions for the input variables. The diagonal plots show the histograms for each individual variable, while each of the non-diagonal subplots plots two of the variables against each other. Red dots and lines show the measured values for each of the variables. The measured values for most of the variables are far away from the prior means.

Once these samples are taken, they can be propagated to the stability map using the Monte Carlo procedure described by Duncan *et al.* [13]. The frequency domain, zero-order approximation stability algorithm described by Altintas and Budak was applied here [2]. While the full algorithm is not presented here since it has already been well-established in literature (consult [2] for a full explanation), a brief summary of the method is presented to establish how the different variables contribute to calculating the stability map.

1. The cutting force coefficients k_{tc}, k_{nc} are used to determine the dynamic cutting force coefficients as a function of cutter rotation. This is approximated as a Fourier series, using only the first (DC) component. Those dynamic cutting force coefficients are then used to calculate the total cutting force vector as a function of relative tool displacement.
2. The frequency response function (FRF) is determined from the modal parameters k, m, c (see [1]). The FRF is used to calculate how much the tool will vibrate during each tooth-passing period as a function of force and chatter frequency.
3. That vibration is plugged into the displacement term of the cutting force function, which is the characteristic equation of the tool motion as a function of chatter frequency.
4. The eigenvalues of that function are used to calculate the limiting axial depth of cut for every chatter frequency. Depths of cut above this limit will chatter, while below will be stable.
5. The spindle speed associated with each chatter frequency is calculated by finding the appropriate tooth-passing period. This calculation is repeated for every lobe of the stability map.

This algorithm calculates the limiting axial depth without chatter, b_{lim}, for some combination of input parameters $\boldsymbol{\theta}$ and spindle speed Ω as $b_{lim}(\Omega, \boldsymbol{\theta})$. Calculating this value for all samples and spindle speeds in the testing range yields the probabilistic stability map shown Figure 4. The grey-scale level of each point $\{\Omega, b\}$ is the probability of stability, which is based on how many of the sampled stability maps predict that a cut at that point will be stable:

$$P_{stable}(b, \Omega) = \frac{1}{n}\sum_{i=1}^{n} b_{lim}(\Omega, \boldsymbol{\theta}_i) < b \qquad (2)$$

where n is the total number of sampled $\boldsymbol{\theta}$ values. If all the stability maps predict that a point $\{\Omega, b\}$ will chatter, then that point is coloured black (zero probability of stability). If all the sampled stability maps predict that point will be stable, then the point is white (100% probability). Intermediate shades of grey represent different expected probabilities of stability. The probabilistic stability map shown in Figure 4 represents the prior.

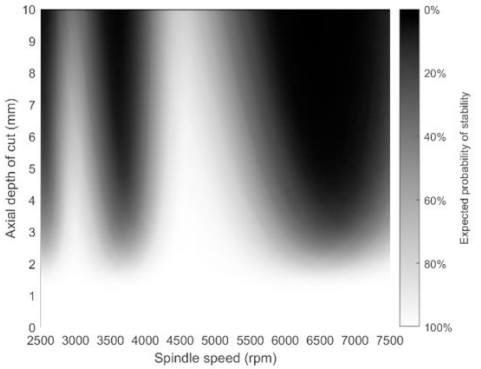


Figure 4: Prior probability of stability map. The more likely it is that a given point is stable, the closer it is to white, while darker points are expected to be more likely to chatter.

Once the probabilistic stability map has been predicted, it can be validated using a grid-type pattern of tests over evenly spaced axial depths of cut and spindle speeds. For this project, tests were performed in increments of 1000 rpm in spindle speed and 2 mm in axial depth, from 2500 rpm to 7500 rpm and 2 mm to 8 mm. Tests at a given spindle speed were discontinued when fully-developed chatter was observed.

The grid-type stability tests were performed and the flexure displacement signal was recorded during cutting. The tachometer pulse was used to sample the displacement once per tooth passage; see Figure 5. Because there was a single tooth, this resulted in once-per-revolution sampling.

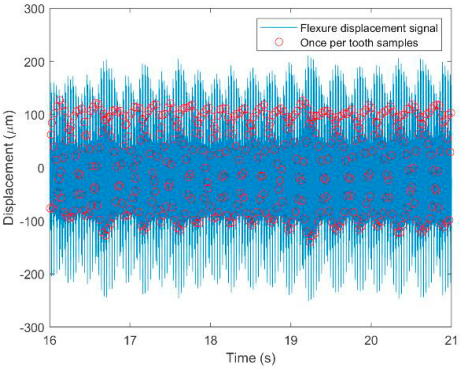


Figure 5: Displacement signal for an unstable test (test 13 at 5500 rpm, 6 mm axial depth in Table 3). The solid blue line represents the displacement signal, while the red circles are the once per tooth samples used to determine stability. For a stable cut, the displacement should repeat every tooth passage.

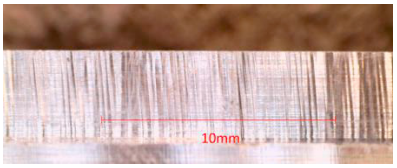
These once-per-tooth samples were classified as either stable or unstable using the sampling metric established by Honeycutt and Schmitz [11]:

$$M = \sum_{i=2}^{N} \frac{|\boldsymbol{x}_s(i) - \boldsymbol{x}_s(i-1)|}{N}, \quad (3)$$

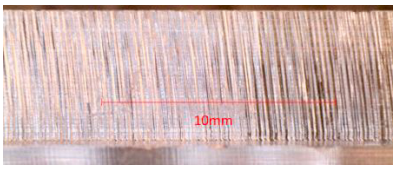
where $\boldsymbol{x}_s$ is a vector containing the once-per-tooth passage samples from the flexure displacement signal, and N is the length of the $\boldsymbol{x}_s$ vector. In a stable cut, the flexure vibration repeats with each tooth passage, meaning that $\boldsymbol{x}_{1:N} = M = 0$ µm ideally. In contrast, the samples vary from tooth to tooth for an unstable cut (chatter), so $M > 0$. A threshold of $M = 10$ µm was selected for automated stability labelling: any cuts with $M \leq 10$ µm were classified as stable, while cuts where $M > 10$ µm were classified as chatter. Figure 6 compares the surfaces from tests 7 ($M = 21$ µm) and 12 ($M = 10$ µm). While both surfaces show noticeable tool marks (due to the low stiffness of the flexure), the tool marks for test 7 are not evenly spaced, which is a key signifier of chatter. In contrast, the tool marks for test 12 are, while large, evenly and regularly spaced, indicating a stable cut (albeit one with large surface dislocation.) While $M = 10$ µm was used for these tests, different cutoffs may be required for different systems or desired surface quality.

Table 2: Grid test points[1]

Test Number	Spindle speed (rpm)	Axial depth (mm)	M (µm)	Stability result
1	2500	2	3	Stable
2	3500	2	104	Unstable
3	4500	2	2	Stable
4	5500	2	2	Stable
5	6500	2	17	Unstable
6	7500	2	3	Stable
7*	2500	4	21	Unstable
8	4500	4	5	Stable
9	5500	4	79	Unstable
10	6500	4	229	Unstable
11	7500	4	1	Stable
12	4500	6	10	Stable
13	5500	6	127	Unstable
14*	4500	8	269	Unstable



(a)



(b)

Figure 6: Cut surfaces for (a) grid test 7 ($M = 21$ µm) and (b) grid test 12 ($M = 10$ µm). Note the uneven spacing of the cutter marks in (a), while in (b) all the tool marks are generally evenly spaced, indicating a stable cut

4. Bayesian updating

Once a cutting test is completed and new information is obtained, the posterior is calculated. The posterior is a new probability distribution which incorporates the new information that's been gained from the outcome of the test cut into the prior. Probability distributions are defined by their probability density function (PDF). This is a function which defines how likely it is for a specific value of θ to be obtained. For example, the PDF of the prior distribution is:

$$P_{prior}(\boldsymbol{\theta}) = \mathrm{PDF}_{\mathrm{Normal}_5}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{S}) = \frac{1}{\left(\sqrt{2\pi}\right)^5 \sqrt{\det(\boldsymbol{S})}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})\right), \quad (4)$$

where $\det(\boldsymbol{S})$ is the determinant of the covariance matrix $\boldsymbol{S}$.The goal is to find a new function which defines the probability of any value of $\boldsymbol{\theta}$ after taking into account the new information acquired from the results of the test cut. Formally, the posterior distribution can be written as $P(\boldsymbol{\theta}|\omega)$, the probability of the input values $\boldsymbol{\theta}$ conditioned on the test result ω. This can be calculated using Bayes' rule[2]:

$$P(\boldsymbol{\theta}|\omega) \propto P(\omega|\boldsymbol{\theta}) \cdot P_{prior}(\boldsymbol{\theta}), \quad (5)$$

where $P(\omega|\boldsymbol{\theta})$ is the probability of observing a test result conditioned on the input values, i.e., how likely it is that a set of input values that say a given stable test will actually be stable or a given unstable test will actually be unstable. The test results ω contain three pieces of information: the test spindle speed Ω_{test}, the test axial depth of cut b_{test}, and whether the cut was stable or not. Using this information, $P(\omega|\boldsymbol{\theta})$ is defined with two piecewise Gaussian dropoff functions, depending on the test cut result. If ω was stable, then:

[1] For cutting tests marked with *, it was observed that the tool shank contacted the part during large vibrations.

[2] $P(\boldsymbol{\theta}|\omega)$ is proportional to $P(\omega|\boldsymbol{\theta}) \cdot P_{prior}(\boldsymbol{\theta})$ since, for it to be a valid probability density function, $\int_\theta P(\boldsymbol{\theta}|\boldsymbol{\omega})\, d\boldsymbol{\theta}$ must be equal to 1. There is therefore some normalization factor that would have to be added to compensate for this to satisfy $P(\boldsymbol{\theta}|\omega) = c_{norm} \cdot P(\omega|\boldsymbol{\theta}) \cdot P_{prior}(\boldsymbol{\theta})$. In practice, this normalization factor is both difficult and, as will be seen, unnecessary to calculate.

$$P(\omega|\boldsymbol{\theta}) \propto \begin{cases} 1 & b_{lim}(\Omega_{\text{test}}, \boldsymbol{\theta}) > b_{test} \\ e^{-0.5\left(\frac{b_{lim}(\Omega_{test},\boldsymbol{\theta}) - b_{test}}{\sigma}\right)^2} & \text{otherwise} \end{cases} \quad (6)$$

If ω was unstable, then:

$$P(\omega|\boldsymbol{\theta}) \propto \begin{cases} 1 & b_{lim}(\Omega_{\text{test}}, \boldsymbol{\theta}) < b_{test} \\ e^{-0.5\left(\frac{b_{lim}(\Omega_{test},\boldsymbol{\theta}) - b_{test}}{\sigma}\right)^2} & \text{otherwise} \end{cases} \quad (7)$$

If the b_{lim} value calculated by $\boldsymbol{\theta}$ at the test spindle speed correctly predicts the results of the test cut, then $P(\omega|\boldsymbol{\theta}) = 1$. This occurs if the cut was stable and b_{lim} was above the test depth of cut, or if the cut was unstable and b_{lim} was below the test depth of cut. If $\boldsymbol{\theta}$ incorrectly predicts the test result, then $P(\omega|\boldsymbol{\theta}) < 1$. The further away $b_{lim}(\omega, \boldsymbol{\theta})$ is from correctly predicting the test results, the lower $P(\omega|\boldsymbol{\theta})$ is, with the rate of drop off set by σ. For this paper, a drop off value of $\sigma = 0.125$ mm was used. This is depicted in Figure 7, which shows the probability density function that results from a stable test cut at a 5 mm axial depth.

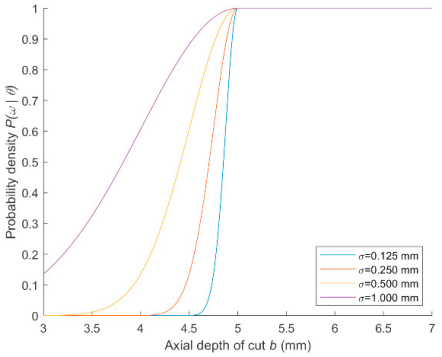


Figure 7: $P(\omega|\boldsymbol{\theta})$ for a stable test result at $b_{test} = 5$ mm with varying values of the dropoff parameter σ. Any value of b_{lim} above 5 mm has a probability of 1 since it predicts that test should be stable. If $b_{lim} < 5$ mm, then the probability density decreases, with the rate of decrease set by σ: the lower σ is, the faster the probability decreases. If the cut was unstable, then the chart would be mirrored horizontally around $b = 5$ mm.

An equation which is proportional to $P(\boldsymbol{\theta}|\omega)$ has now been established. Samples taken from the prior distribution $P_{prior}(\boldsymbol{\theta})$ can then be updated via rejection sampling [14]. The required steps follow. For each sample:

- calculate $P(\omega|\boldsymbol{\theta})$
- select a random value from the distribution Uniform(0,1)
- if the random value is less than $P(\omega|\boldsymbol{\theta})$, then discard $\boldsymbol{\theta}$
- otherwise, accept it.

The remaining samples will approximate the desired distribution $P(\boldsymbol{\theta}|\omega)$. This is called rejection sampling and it's a common method for sampling from complex univariate probability distributions. Figure 8 displays the result of this updating when the first test from the grid at a 2500 rpm spindle speed and 2 mm axial depth is used to update the stability map. Only 75 of the 1000 samples were rejected, since the majority of the prior sampled stability maps predicted that the test would be stable. This indicates that the test had low information value: there was little change in the posterior distribution, and, therefore, little new information was gained from the test.

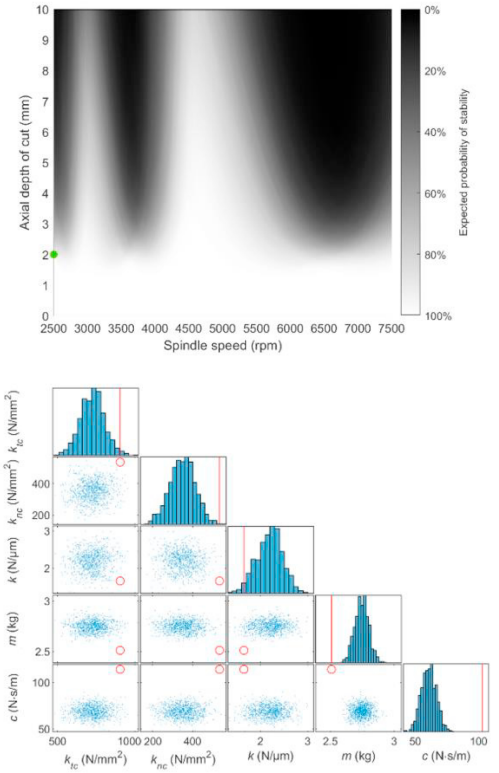


Figure 8: Posterior stability map and histograms after updating based on a stable test run at 2500 rpm, 2mm axial depth of cut. 8 of the samples were rejected because of the results of that test. There are 992 out of the original 1000 samples remaining. This indicates that the test was of low value and demonstrates the inefficiency of the grid test: it was taken at a position which was already expected to be stable with high probability, and there is thus very little change in the stability map since not much has been learned.

This rejection sampling method can be generalized to an arbitrary number of cutting tests by using the posterior of each cutting test $P(\boldsymbol{\theta}|\omega_n)$ as the prior for the next update:

$$P(\boldsymbol{\theta}|\omega_{1:n}) \propto P(\omega_n|\boldsymbol{\theta})P(\boldsymbol{\theta}|\omega_{1:n-1}) = P(\boldsymbol{\theta})\prod_{n=1}^{N} P(\omega_n|\boldsymbol{\theta}) \,. \quad (8)$$

However, executing this strategy over multiple tests will quickly deplete the remaining number of samples. An example is provided in Figure 9, showing the posterior distribution after applying rejection sampling for the entire 14 cut grid test. After 14 cutting tests, 998 of the original 1000 samples have been rejected.[3] The two remaining samples are not sufficient to accurately represent the full posterior distribution.

[3] Note that the exact number and location of samples will vary if the updating process is run again, since both the Monte Carlo sampling and the rejection sampling are stochastic methods.

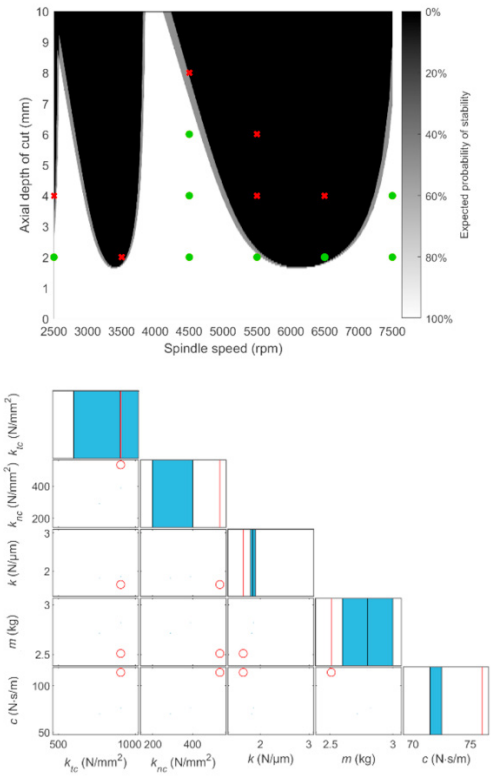


Figure 9: Example of sample depletion after running 14 cutting tests in a grid pattern. There are only two samples left.

This sample depletion issue can be solved by generating more samples from the target distribution via the same rejection sampling method, but this is computationally inefficient. Since only 0.2% of the samples are accepted, approximately 500,000 samples would have to be generated to yield 1000 samples in the posterior.

A better option is to sample directly from the target distribution. However, doing so is non-trivial since there's no closed-form integral for the PDF. Numerically approximating the cumulative density function would require integrating over the entire five-dimensional parameter space. This is computationally infeasible; even a low-resolution integration with 100 samples per dimension would require a total of $100^5 = 10^{10}$ samples (which would take roughly three years to calculate at 100 samples per second). Therefore, an alternative to direct sampling is required.

5. Markov Chain Monte Carlo sampling

One way to efficiently sample the posterior distribution is to use a Markov Chain Monte Carlo (MCMC) method. MCMC methods are a family of techniques used for taking samples from complex high-dimensional distributions and are commonly applied to calculate posteriors for Bayesian inference [15]. Rather than generating each sample independently, they generate samples as a Markov chain where every sample is generated based on the previous sample in the chain. While each sample is correlated with the previous sample, the distribution will eventually converge to the target distribution as sufficient samples are taken.

The specific method used in this paper is a greedy adaptive MCMC method from Haario *et al.* [16]. The adaptive MCMC method is a variation on the classic Metropolis-Hastings (MH) algorithm, which accepts or rejects samples based on the ratio of their probability to the previous value in the Markov Chain [17].[4] This method offers several advantages over simpler sampling methods like rejection sampling or traditional MH.

- Unlike directly sampling the distribution, MH does not require that the probability distribution be normalized; it is sufficient to have a function that is proportional to the true PDF. This is critical, since calculating the appropriate normalization factor to ensure that the PDF integrates to 1 would require numerically integrating over the entire five-dimensional parameter space.
- MCMC methods have higher acceptance rates than simple rejection sampling, especially in high-dimension probability spaces. This is because the Markov chain tends to congregate in areas with high probability and spends less time in areas with lower probability. Thus, the MCMC method will waste fewer samples on areas of the parameter hyper-space where $P(\boldsymbol{\theta}|\omega_{1:N}) \approx 0$.
- Rather than using the same proposal distribution every time like MH, Haario's algorithm updates its proposal distribution after every sample based on the estimated covariance matrix defined by all accepted samples up to that point. Selecting an appropriate proposal distribution for MH is both critical and challenging. If the proposal distribution proposes steps which are very far away from the last accepted sample, then the acceptance rate will be low. However, if the proposal distribution suggests small steps, then the correlation between subsequent samples will be high and more samples will be required to fully explore the parameter space. The adaptive MCMC method ensures that the proposal is always appropriate for the target distribution, and thus provides faster convergence.

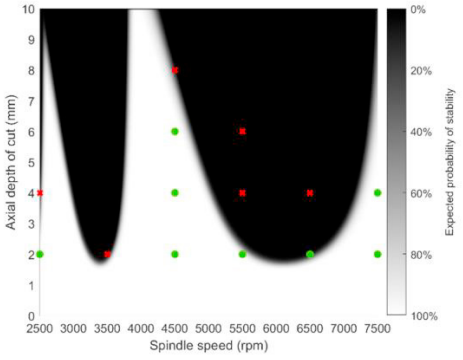
The MCMC process for sampling the target distribution is described here.

1. Collect all samples of $\boldsymbol{\theta}$ that remain after the rejection sampling process described previously. These will be referred to as $\boldsymbol{\theta}^*$ and provide a more accurate initial estimate of the target covariance matrix. This improves convergence since the adaptive algorithm starts with a reasonably good guess for the optimal proposal distribution.
2. Calculate the covariance matrix of all the samples in the union of the starting sample set, $\boldsymbol{S_0} = \mathrm{Cov}(\boldsymbol{\theta}^*)$.

[4] Specifically, it is a variation on the original Metropolis algorithm, a simpler version of Metropolis-Hastings algorithm which requires symmetric proposal distributions. The Haario algorithm is asymptotically symmetric: while the proposal distribution does technically change as each new sample is added, the change is small enough that it can be ignored.

3. Select one of the samples in the starting sample set as the starting point for the Markov chain. Designate this as $\boldsymbol{\theta}_0$.[5]
4. Set the sample tracker $n = 1$.
5. Create a proposal distribution $\boldsymbol{\theta}_{\text{proposal}} \sim (1-\beta)\, N_d(\boldsymbol{\theta}_{n-1}, (2.38)^2 \boldsymbol{S}_{n-1}/d) + \beta\, N(\boldsymbol{\theta}_{d-1}, (0.1)^2 \boldsymbol{I}_d / d)$.[6] The first term is the adaptive normal distribution centered around the last accepted point. The second term is added to help ensure that the parameter space is fully explored even when $\boldsymbol{\theta}^*$ is not a good representation of the actual desired probability distribution [15]. Also, β is a weighting value in the range $0 < \beta \ll 1$. Here, a value of $\beta = 0.1$ was used.
6. Draw a sample $\boldsymbol{\theta}_n$ from $\boldsymbol{\theta}_{\text{proposal}}$.
7. Calculate $P(\boldsymbol{\theta}_n|\omega_{1:N})$ and compare it to a random value x drawn from a uniform distribution between zero and one. If $x > \frac{P(\boldsymbol{\theta}_n|\omega_{1:N})}{P(\boldsymbol{\theta}_{n-1}|\omega_{1:N})}$, then return to step 5 to draw a new candidate $\boldsymbol{\theta}_n$. Otherwise, continue to the next step.
8. Increment n by one. If $n > 1000$, then return $\theta_{1:1000}$. Otherwise, calculate the new covariance matrix $S_n = \text{Cov}(\cup\, \boldsymbol{\theta}^*, \boldsymbol{\theta}_{1:n})$ and return to step 4.[7]

This MCMC method is applied after each new test point is added and the resulting samples are then used as the prior for the next test. Figure 10 shows the posterior distribution after all the grid tests are included. Overall, the results agree with those obtained through rejecting samples, demonstrating that both methods give equivalent results. However, the MCMC method has many more samples and better approximates the target distribution $P(\boldsymbol{\theta}|\omega)$ (note how the larger number of samples in the distributions gives a better idea of the distribution shape).



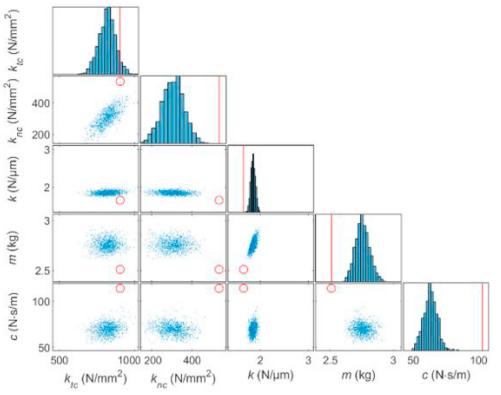


Figure 10: Posterior after the grid tests, using the Markov Chain Monte Carlo method to generate more samples after every test. The red x represents an unstable result, while a green dot represents a stable cut.

There are two interesting things to note about the results in Figure 10. First, the individual values in the posterior distribution are no longer independent. For example, the distributions for k_{nc} and k_{tc} are positively correlated, as shown by the slope of the samples when the two variables are plotted against each other (first column, second row from top.)

Second, the posterior distributions are not necessarily normal. This can be seen by examining the correlation between k and m. The samples from a multivariate normal distribution will take on an elliptical shape. However, the distribution of k and m takes on something more like a wedge, with a tighter correlation at smaller values of k and a wider distribution at higher values of k.

6. Expected value test selection

Taking machines out of production to run cutting tests is expensive. Therefore, the goal of the testing is to identify the best operating parameters in as few tests as possible. As has been demonstrated, grid tests are an inefficient way to select test points; many tests often give little or no information. Other test points are of low value since they have low material removal rate, such as the test at 2500 rpm, 4 mm.

By examining Figure 8, it is clear that some areas are more useful to test than others. Running a test at 6500 rpm, 10 mm , for example, is likely not valuable since it is expected to be unstable with high probability. Similarly, testing at 2500 rpm, 1 mm is also not valuable since it has low material removal rate.

[5] The exact starting position does not matter provided it is of relatively high probability with respect to the number of samples taken. For example, it could be acceptable to start with a point with a probability of 10^{-3} provided that 10^6 samples are taken, since it's expected that 1000 such samples would appear over the entire run. This same starting point would not be acceptable if only 100 samples are taken since such a point should be unlikely to appear. Often, a burn-in cycle is used to help ensure this, running a fixed number of throwaway samples to let the Markov chain move into a higher probability area. For this application, however, this isn't necessary: any of the samples which were retained should be of sufficiently high probability to not bias the Markov chain by starting in a very low probability region [18]. If, however, a test point is added which results in no samples remaining, a burn-in cycle could be used to find an appropriate starting point.

[6] The choice of the constant $\frac{2.38^2}{d}$ has been demonstrated to be close to optimal when both the proposal distribution and target distribution are Gaussian [19]. That assumption does not hold here: the proposal distribution is Gaussian, but the posteriors is not. However, much of the time they are sufficiently close to normal for that constant to give sufficient mixing. Empirically, sufficient mixing is achieved even when the posterior has multiple modes [16].

[7] Note that none of the retained samples are kept. This is to minimize correlation of the samples with the prior: if there were 990 samples retained and MCMC was used to generate 10 new samples to get back to 1000, those 10 samples would be heavily correlated with the starting point since the Markov chain wouldn't be long enough to converge to the target distribution. Therefore, 1000 new samples are always generated.

One effective method to select the new test point is to choose the point which gives the largest expected value, as described by Karandikar *et al.* [3]:

$$V_{test}\{\Omega_{test}, b_{test}\} = \max(P_{stable}(\Omega_{test} \cdot b_{test} - V_{best}), 0) \quad (9)$$

This calculates the expected improvement in material removal rate for some test point $\{\Omega_{test}, b_{test}\}$. The $\Omega_{test} \cdot b_{test}$ term represents a proxy for material removal rate MRR. Since it is assumed the radial depth is constant and that the feedrate is proportional to spindle speed, $\text{MRR}_{test} \propto \Omega_{\text{test}} \cdot b_{test}$. V_{best} is the best identified material remove rate $(\Omega \cdot b)$ that has been identified so far. For this paper, V_{best} is initially set to zero, indicating that no milling parameters have been confirmed to be stable by testing. This method is suitable if there is no prior information on the tool performance available, like when a new model of cutting tool is being used. If there are known stable operating parameters (for example, if a tool was previously used to machine a part successfully), then those parameters can also be used as a starting point.

As shown in Figure 11, this value is calculated for all points in the stability map and the position with the highest expected value is selected as the next test point. For the prior here, the first recommended test point is at $\{\Omega = 4567 \text{ rpm}, b = 9.8 \text{ mm}\}$ at the top center of the map.

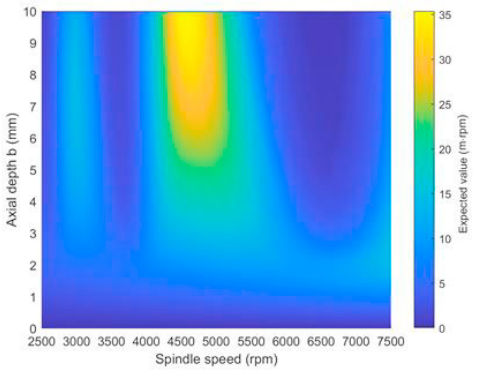


Figure 11: Expected value of the initial (i.e., prior) stability map.

7. Test results

A series of tests were completed using the results from each test to update the probability distribution and form a new posterior. This posterior was approximated using the MCMC sampling technique, yielding a new probabilistic stability map. This stability map was used to select a new test, and so on. This process was iterated, resulting in the series of tests shown in Table 3. The posterior stability map and parameter distributions are shown in Figure 12.

Table 3: Stability tests

Test Number	Spindle speed (rpm)	Axial depth (mm)	M (μm)	Stability result
1	4567	9.8	306	Unstable
2	7500	9.8	10	Stable

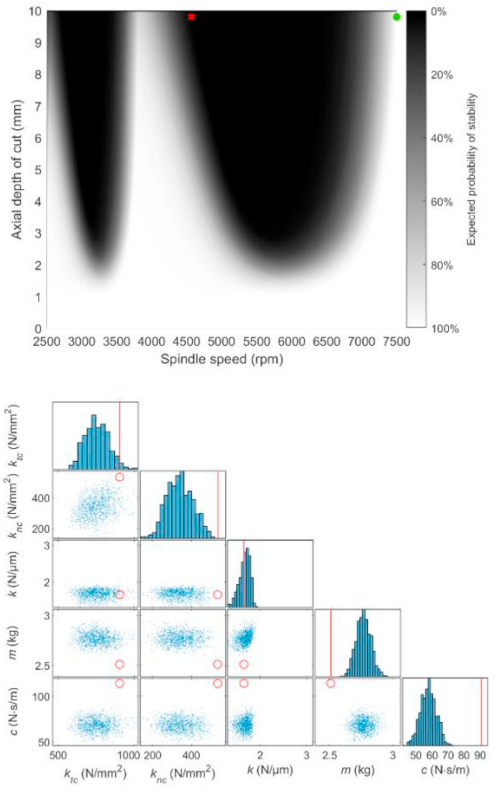


Figure 12: Stability probability map and distributions after two cutting tests.

Testing was stopped after two tests since the expected value algorithm did not find any points which could significantly improve MRR. Figure 13 shows the expected value map used to select the third test point. Since the expected value algorithm will only selected test points which can improve the MRR, nearly the whole map has a value of zero. The maximum expected improvement in MRR is 1.1 m · rpm at 7500 rpm, 10 mm, which represents only a 1.4% improvement to the current best parameters. This is not a significant improvement so testing was stopped.

This example demonstrates how the Bayesian updating framework can quickly converge to the best operating parameters. However, it did not meet the other goal: to reduce uncertainty for the input parameters $k_{tc}, k_{nc}, k, m,$ and c. The distributions of the individual variables, as shown in the histograms and correlation plots, have not converged towards any specific values. There are many possible values for the input variables that could give the test results shown in Figure 12. The ultimate goal for the framework is to identify the same values that can be obtained through direct measurement, which were listed in Table 1.

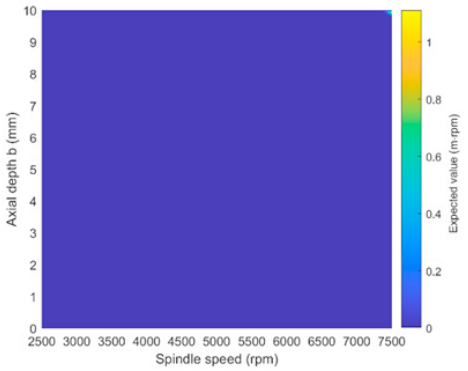


Figure 13: Expected value map after two tests.

Therefore, a second set of tests was completed to evaluate how well the Bayesian updating framework can identify the true value of the parameters. The test setup was identical, except that the maximum spindle speed was reduced from 7500 rpm to 7000 rpm. This limit excluded the stable peak at 7500 rpm, forcing the expected value algorithm to explore the solution space more thoroughly. This resulted in the test set summarized in Table 4.

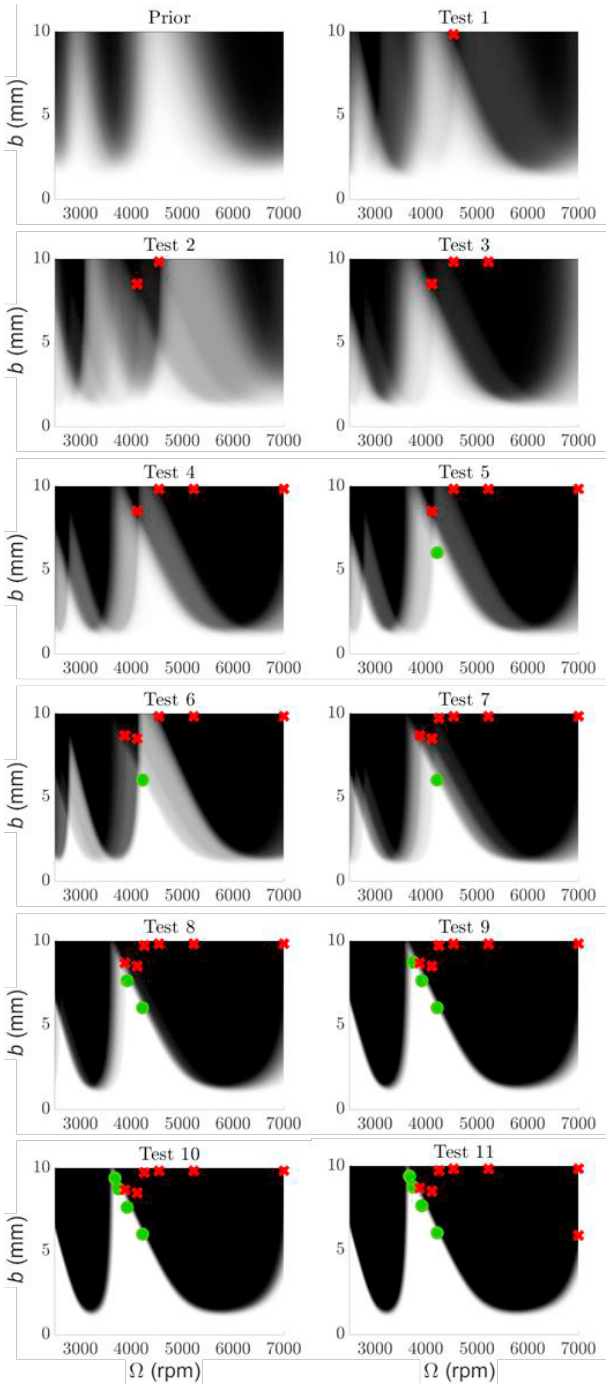


Figure 14 shows the stability map distribution after each cutting test. Figure 15 compares the posterior stability map to the theoretical stability map calculated from the measured values, as well as comparing the posterior distributions to the measured values.

Table 4: Stability tests[8]

[8] For cutting tests marked with *, it was observed that the tool shank contacted the part during large vibrations.

Test Number	Spindle speed (rpm)	Axial depth (mm)	M (μm)	Stability result
1	4567	9.8	306	Unstable
2	4117	8.5	15	Unstable
3	5233	9.8	185	Unstable
4	7000	9.8	156	Unstable
5*	4222	6	7	Stable
6	3873	8.3	15	Unstable
7	4252	9.7	16	Unstable
8*	3920	7.6	8	Stable
9*	3750	8.7	6	Stable
10*	3679	9.4	10	Stable
11	7000	5.9	142	Unstable

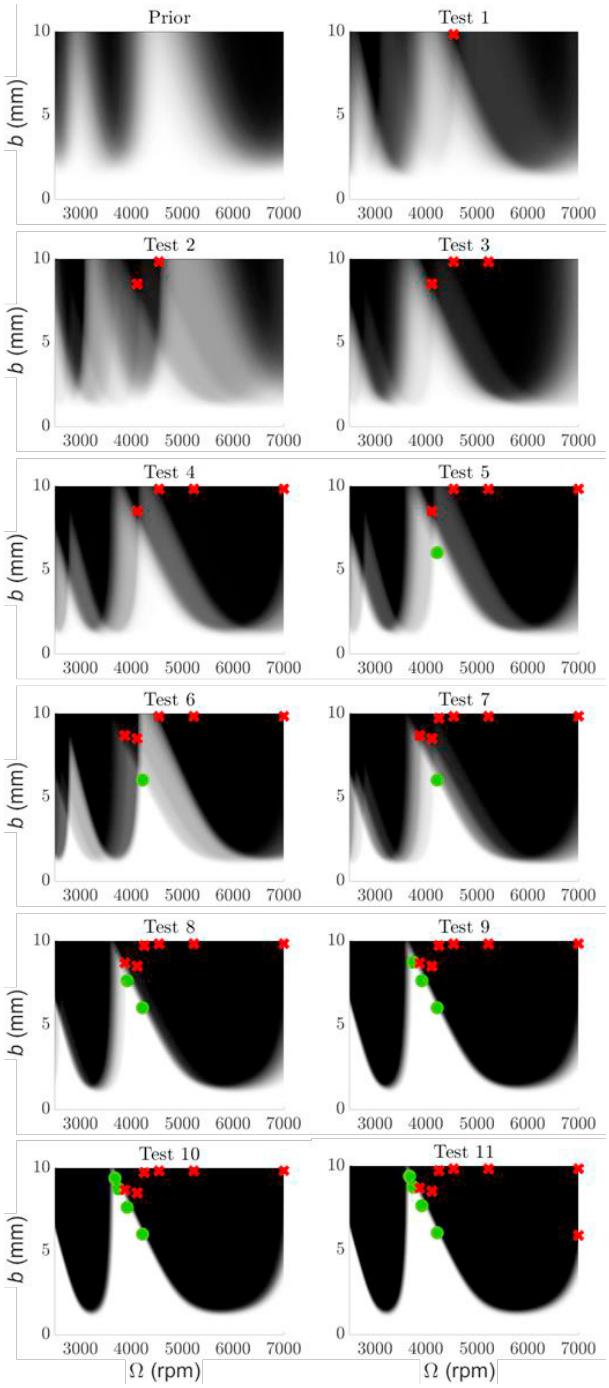


Figure 14: Stability maps after every test in Table 4. The horizontal axis for each plot is the spindle speed in revolutions per minute. The vertical axis for each plot is axial depth of cut. Stable tests are shown in green, while unstable tests are red.

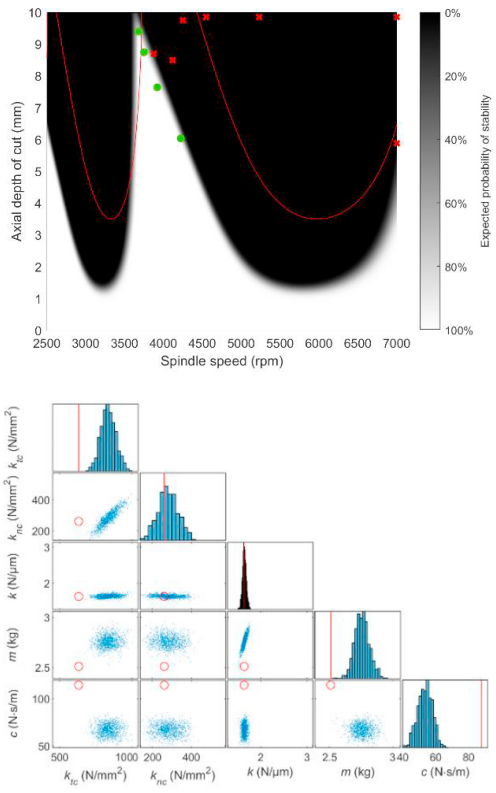


Figure 15: Posterior stability map and distributions after 11 cutting tests. The red solid line in the stability map is the stability map calculated from the measured values using the frequency domain stability limit model [2], while the red dots and lines in the distributions are the measured values.

Table 5 compares the prior, marginal posterior, and measured values for each parameter. This table also includes the derived properties f_n, the flexure natural frequency, and ζ, the flexure damping ratio. The distributions of these derived properties were calculated by Monte Carlo sampling. The prior and posterior uncertainties in the modal parameters was propagated via Monte Carlo simulation to calculate those parameters. Ultimately, the posterior distribution did not converge to the measured values. Additionally, as shown in Figure 15, the posterior stability map did not converge to the stability map calculated from those measured values.

Table 5: Comparison of the prior and posterior uncertainties. The posterior mean/standard deviation do not fully describe the posterior since the distributions of each individual variable are neither normal nor independent. However, they are useful as a point of comparison.

Variable	Prior distribution ($\mu \pm 1\sigma$)	Posterior distribution ($\mu \pm 1\sigma$)	Measured value
$k_{tc}\ \left(\frac{\mathrm{N}}{\mathrm{mm}^2}\right)$	734 ± 83.4	653 ± 53.8	903.9
$k_{nc}\ \left(\frac{\mathrm{N}}{\mathrm{mm}^2}\right)$	346 ± 67.3	387 ± 54.8	534.5
$k\ \left(\frac{\mathrm{N}}{\mu\mathrm{m}}\right)$	2.17 ± 0.266	1.65 ± 0.04	1.64
m (kg)	2.75 ± 0.057	2.75 ± 0.060	2.51
$c\ \left(\frac{\mathrm{N} \cdot \mathrm{s}}{\mathrm{m}}\right)$	68.3 ± 6.47	66.9 ± 6.29	114
f_n (Hz)	142 ± 8.7	123 ± 0.86	129
ζ	$1.41 \pm 0.16\%$	$1.57 \pm 0.15\%$	2.82%

8. Discussion

The framework was able to quickly identify the optimal milling parameters (i.e., the milling parameters with the highest MRR). On the first test set, the Bayesian framework identified the optimal parameters in two test cuts, outperforming other optimization techniques, such as the naïve Bayesian approach proposed by Karandikar *et al.* [3]-[4]. However, as shown in Figure 15, the stability map failed to converge towards the true values of the underlying parameters. More significantly, it did not converge towards the true shape of the stability map.

This occurs due to limitations in the frequency domain, zero-order stability limit approximation. While the approximation is accurate for slotting, lower radial depths of cut yield bifurcations in the stability diagram [11]. These regions display period-n stability, where cuts repeat their displacement profiles every n tooth passes (i.e., they have periods that are integer multiples of the tooth period, rather than repeating every tooth period for forced vibration in stable cuts).

This is illustrated in Figure 16, which compares the analytical stability boundary to time domain simulation results. In total, 400 time domain simulations were completed and the resulting displacement profiles were classified as either stable (period-1), period-n (for n values up to 10), or unstable (secondary Hopf bifurcation) based on subharmonic sampling [20]. This method is an adaptation of the earlier once-per-tooth sampling, except that the metric is calculated over samples taken every nth tooth passage:

$$M(n) = \sum_{i=2}^{N} \frac{|\boldsymbol{x}_{sn}(i) - \boldsymbol{x}_{sn}(i-1)|}{N}, \qquad (10)$$

where $\boldsymbol{x}_{sn}$ is the vector of samples taken from the displacement signal every n tooth passages.

The time domain simulations show a region of period-n and unstable behavior near 4500 rpm which splits the traditional stability lobe. While the analytical solution predicts that cuts in this region should be stable (for example, the test cut at 4567 rpm, 9.8 mm), both the time domain and actual test cuts are unstable. There is also an extra region of stable and points that weren't predicted by the analytical solution at roughly 5200 rpm.

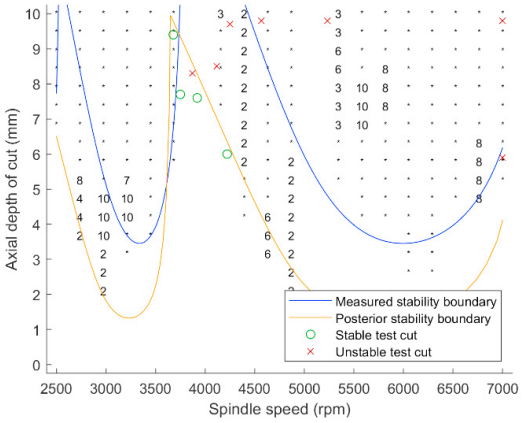


Figure 16: Comparison of analytical and time-domain stability. The numbers in the background provide the M values. Empty regions are predicted to be stable by the time domain simulation. A 2 indicates a period-2 instability, and so on. The dots represent a secondary Hopf bifurcation. The threshold M value was 1 µm. The test cuts generally agree with the time domain simulation and the posterior stability map closely approximates the shape of the left side of the split stability lobe.

Since this intermediate unstable lobe is not predicted by the frequency domain solution, the Bayesian model cannot predict its effects. The end result, as shown in Figure 16, is that the Bayesian method fits a stability lobe to one side of the bifurcated analytical lobe; here, it is observed on the left. This is the best solution given the model limitations.

This illustrates how the choice of model changes the performance of the Bayesian updating and demonstrates the advantages and disadvantages of physics-guided models. A physics-guided model can converge towards the optimal operating parameters more quickly than a more naïve approach. However, it also limits the solutions that the model can identify. Since the frequency domain model does not account for the possibility of bifurcations at low radial engagement, the Bayesian model (which is driven by the frequency domain model) also cannot predict that behavior. In contrast, the naïve Bayesian approach used by Karandikar *et al.* may require more tests to converge, but because it is not physics-guided it does not depend on the assumptions injected from an underlying physics model. It will converge to the true stability map given a sufficient number of test points. Future work can incorporate alternative, more accurate models to reduce the effect of those approximations, such as using time domain simulations. Other areas for future investigation could be different test selection algorithms and expanding the method to other parameters for the stability map algorithm (such as helix angle, variable-pitch flutes, or tool runout.)

9. Conclusion

This paper described a physics-guided framework for optimization and parameter identification of milling setups. Prior uncertainties for the FRF and cutting force coefficients were established through simulation and literature reviews and propagated to the stability map. The Bayesian updating framework was then applied to update those uncertainties based on the results of limited cutting tests. The posterior distributions for the stability map and input parameters were compared to measured values and the advantages and disadvantages of physics-guided frameworks were discussed.

References

[1] Schmitz, T., Smith, S., 2019. Machining Dynamics: Frequency Response to Improved Productivity. Springer.
[2] Altintaş, Y. and Budak, E., 1995. Analytical prediction of stability lobes in milling. CIRP annals, 44(1), pp.357-362.
[3] Karandikar, J., Traverso, M., Abbas, A. and Schmitz, T., 2014. Bayesian inference for milling stability using a random walk approach. Journal of Manufacturing Science and Engineering, 136(3).
[4] Karandikar, J., Honeycutt, A., Smith, S. and Schmitz, T., 2020. Milling stability identification using Bayesian machine learning. Procedia CIRP, 93, pp.1423-1428.
[5] Li, K., He, S., Liu, H., Mao, X., Li, B. and Luo, B., 2020. Bayesian uncertainty quantification and propagation for prediction of milling stability lobe. Mechanical Systems and Signal Processing, 138, p.106532.
[6] Suzuki, N., Kurata, Y., Kato, T., Hino, R. and Shamoto, E., 2012. Identification of transfer function by inverse analysis of self-excited chatter vibration in milling operations. *Precision Engineering*, *36*(4), pp.568-575.
[7] Gomez, M., No, T., Smith, S. and Schmitz, T., 2020. Cutting force and stability prediction for inserted cutters. Procedia Manufacturing, 48, pp.443-451.
[8] Tsai, M.Y., Chang, S.Y., Hung, J.P. and Wang, C.C., 2016. Investigation of milling cutting forces and cutting coefficient for aluminum 6060-T6. Computers & Electrical Engineering, 51, pp.320-330.
[9] Xu, M., Jerard, R.B. and Fussell, B.K., 2007. Energy based cutting force model calibration for milling. Computer-Aided Design and Applications, 4(1-4), pp.341-351.
[10] Rubeo, M.A. and Schmitz, T.L., 2016. Mechanistic force model coefficients: A comparison of linear regression and nonlinear optimization. Precision Engineering, 45, pp.311-321.
[11] Honeycutt, A. and Schmitz, T.L., 2016. A new metric for automated stability identification in time domain milling simulation. Journal of Manufacturing Science and Engineering, 138(7).
[12] Rubeo, M.A., Kiran, K. and Schmitz, T.L., 2016. The design of a multiple degree of freedom flexure stage with tunable dynamics for milling experimentation. Transactions of the NAMRI/SME, 44.
[13] Duncan, G., Schmitz, T.L. and Kurdi, M.H., 2000. Uncertainty propagation for selected analytical milling stability limit analyses. Society of Manufacturing Engineers.
[14] Liu, J.S., 1996. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. Statistics and computing, 6(2), pp.113-119.
[15] Holmes, C., Krzysztof, L. and Pompe, E., 2017. Adaptive MCMC for multimodal distributions. Technical report.
[16] Haario, H., Saksman, E. and Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli, 7(2), pp.223-242.
[17] Andrieu, C., De Freitas, N., Doucet, A. and Jordan, M.I., 2003. An introduction to MCMC for machine learning. Machine learning, 50(1-2), pp.5-43.
[18] Meyn, S.P. and Tweedie, R.L., 2012. Markov chains and stochastic stability. Springer Science & Business Media.
[19] Gelman, A., Roberts, G. and Gilks, W., 1996. Efficient metropolis jumping rules. Bayesian statistics.
[20] Honeycutt, A. and Schmitz, T.L., 2017. Milling stability interrogation by subharmonic sampling. Journal of Manufacturing Science and Engineering, 139(4).