TU Dortmund

Case Studies, LSF, 053014

Project I: BTA Deep Hole Drilling

Lecturers:

Dr. Uwe Ligges

M. Sc. Leonie Schuermeyer

Author: Nancy Bou Kamel

Group number: 2

Group members: Nancy Bou Kamel, Nourah Buhamra, Mariia Hrechyn, Veronika Tsishetska

July 26, 2023

Contents

1. Introduction

The BTA stands for Boring and Trepanning Association, a vital cutting process in the metal processing industry. It can drill deep slim holes up to 400:1 depth-to-diameter ratio. Using the BTA makes it possible to produce exact and smooth holes, for example, axial bores in turbines and compressor shafts. Two main types of dynamic process disturbances can occur during drilling, namely, chattering and spiralling. In a chattering operation, self-excited rotational vibration leads to tool wear, noisy working conditions, and poor quality work. In a spiralling process, bending vibrations cause holes with multiple lobes, adversely affecting the borehole. The main emphasis of this project revolves around the presentation and analysis of data, involving various stages such as data importing, examination of different parameters in the header files, analysis of moment sensors during distinct phases, and time discretization to extract frequencies associated with each phase.

Chapter 2 provides an introductory overview of the available data and the specific objectives we are aiming to achieve through our analysis. In Chapter 3, fundamental concepts of Time series analysis, autocorrelation, Fourier transform, and the use of Hanning window are thoroughly explained, setting the groundwork for the subsequent statistical models and tests.

Chapter 4 delves into the implementation of statistical models and tests, which help us understand the relationships between parameters and moment behavior. This phase of the analysis enables us to draw meaningful conclusions based on the data's statistical characteristics and correlations.

Finally, Chapter 5 presents a concise summary of the results obtained from our analysis. This chapter provides an organized and comprehensive overview of the key findings, allowing us to draw meaningful insights and conclusions from the data exploration and statistical tests performed in the previous chapters.

2. Problem statement

This data is collected by the Institute for Machining, Faculty of Mechanical Engineering, TU Dortmund, using an A4-sized GX-1 data recorder with a signal change from question to statement. The GX-1 integrates a wide range of functions, such as connecting sensors, recording data, viewing recorded data, and transferring data to a high-level analysis and processing system. Data recording, monitoring, and data transfer to PC are all controlled by the GX Navi software.Whenever recording pauses or stops, the GX-1 creates a 16-bit binary-format data file and an ASCII-format header file. The data file is binary-format data converted through A/D (.dat extension).Data converted from analogue to

digital is stored as a two-byte integer ranging from -32,768 to +32,767. Two's complement numbers represent negative values.The header file is an ASCII text file in which information such as recording condition is written (.hdr extension). Besides a header and a data file, a wave file is created when a voice memo is recorded in the before or after mode. Files in wave format can be viewed in Windows media player. The data file and header file have the same name. In total, five alphanumeric characters and three-digit numbers beginning with 001 represents the ID number. When recording is stopped or paused, the ID number is automatically incremented if the file number auto-incremented is selected.Otherwise, the file is overwritten. In this project, the GXI generated 11 data files and 11 header files, all with the same ID number, 001. However, they had different alphanumeric numbers. Only the torque sensor measurement from V2_0001 is contained in the data file and header file of V2_00001_1. Each data file contains seven sensors and their parameters are stored in their header files. Appendix table 8 explains the parameters of the data file. Since the sensors automatically record their values, there are no null values in the data files. All variables in the dataset are described in Appendix table 9.The data and header files for V6_0001, V10_0001, V2_00001, V17_0001, V24_0001, V25a_001 and V20_0001 contain recorders for CH1_Moment, CH2_Force, CH3_SyncSig, CH4_Acoustic, CH5_a1_WSAS, CH6_a2_WSAF and CH7_a3_BOZA sensors. They were recorded in 2001. Data and header files of D0400001, D0600001, and D0800001 have CH1_Moment, CH2_Force, CH3_Bending, CH4_SyncSig, CH5_a3_BOZA, CH6_Acoustic and CH7_a4_Drilling. They were recorded in 2002. In order to detect drilling torque, a torque sensor is located on the drilling bar above the borehole of the drilling tool. It is excited by forces such as chipping, friction, and deformation at the guide rails. V2_00001 data file contains the largest number of observations (5,629,936), whereas D0400001, D0600001, and D0800001 data files contain the fewest number of observations (2,453). Both data files, V20_0001 and V17_0001 data files contain the same number of observations (5,689,691). There are 5,389,292 observations in the V24_0001 data file and 5,290,814 observations in the V10_001 data file. The primary objective of our analysis is to investigate the impact of various parameters present in the data files on the moment sensor readings. By examining these parameters, we aim to understand their influence on the dynamic behavior of the moment data during different phases of the machining process. Furthermore, our goal is to determine the frequencies associated with each phase. Through spectral analysis and techniques like Fast Fourier Transform (FFT) on the moment data, we can extract the frequency components present in the signal during distinct phases. Identifying these frequencies is crucial in understanding the periodicity and vibration characteristics of the machining process. By analyzing the frequencies and their correlation with specific parameters, we aim to gain valuable insights into the relationship between cutting conditions, workpiece properties, and the dynamic response of the moment sensor. This knowledge can be instrumental in optimizing cutting parameters

and workpiece entry timing to enhance machining performance, workpiece quality, and productivity.

3. Statistical methods

3.1. Time series and stationary processes:

A time series [Haben et al., 2023, pp.55-57] consists of data points gathered over successive time intervals, with each data point associated with a specific timestamp. The order of time series data is chronological. Univariate time series data consist of measuring a single variable over time, whereas multivariate data consist of recording multiple variables simultaneously. In time series analysis, data points are equally spaced. It is used in finance, economics, physics, environmental sciences, engineering, and more. Time series data is denoted as Y_k with time steps $k = 1, 2, ..., N$, where N is the total number of observed time steps in the time series. If the time series continues in the future, then $N = \infty$, and K becomes the set of natural numbers such that $k \in \mathbb{N}$. In time series analysis, stationary data plays an important role in understanding and modeling the data. During a stationary period, the mean, variance, and autocorrelation structure of a time series remain constant. Stationary time series have a constant mean over time. This means that the average value of the series does not change as we move through time. Stationary time series also exhibit a consistent level of variability. In other words, the variance or standard deviation must remain constant over time. Stationary time series have an autocorrelation function (ACF) independent of time. Consequently, the relationship between past and current observations is constant with time. A stationary time series is an effective tool for easier interpretation, reliable forecasting, time series modeling, including ARIMA, as well as statistical analysis. When a time series is non-stationary, both the mean and the variance do not remain constant over time.

3.2. Autocorrelation function

An autocovariance is a measure that quantifies the difference between a time series and its lagged version. It measures the correlation between the values of a time series at different points in time. Given a stationary process X_t with constant mean μ and constant variance σ^2 at time points $t \in \mathbb{N}$, the autocovariance between X_t and X_{t+s} is defined as

$$\gamma(s) = Cov(X_t, X_{t+s}) = \mathbb{E}((X_t - \mu)(X_{t+s} - \mu)) \quad [Broersen,\ 2006, pp.31]$$

where X_t is the observation at time t, X_{t+s} is the observation at time t+s. For all different values of s, it measures the covariance between pairs at a distance or lag s. The lag s is the time distance between the two data points being considered. Positive autocovariance indicates a positive correlation between the time series values at a distance of s time units. This suggests that when the time series value at t is high (or low), the value at t+s is also more likely to be high (or low). Negative autocovariance indicates a negative correlation between the time series values at a distance of s time units. This means that when the time series value at t is high (or low), the value at t+s is more likely to be low (or high). If the autocovariance $\gamma(s)$ is close to zero, it suggests no significant correlation between the time series values at a distance of s time units. In time series analysis, autocovariance provides information about the temporal relationship and patterns within the data. An autocorrelation function (ACF) is often computed using autocovariance, which is a normalized version of autocovariance. Autocorrelation describes the degree of correlation between a time series and a lagged version of it at different time lags. The autocorrelation between X_t and X_{t+s} is

$$\rho(s) = Cov(X_t, X_{t+s}) = \frac{Cov(X_t, X_{t+s})}{\sigma(X_t)\sigma(X_{t+s})} = \frac{\gamma(s)}{\gamma(0)} \quad [Broersen,\ 2006, pp.32]$$

where $\rho(s)$ is the autocorrelation at lag s, $Cov(X_t, X_{t+s})$ is the autocovariance at lag s, which measures the covariance between $X(t)$ and $X(t+s)$, and $\sigma(X_t)\,\sigma(X_{t+s})$ are the standard deviations of the time series values $X(t)$ and $X(t+s)$, respectively. Using the autocorrelation function (ACF), you can plot the autocorrelation value at different lags. Using it, we can gain valuable insight into how the time series are correlated and patterned over time.

For a stationary process, the autocovariance function $\gamma(s)$ and the autocorrelation function $\rho(s)$ have the following properties [Broersen, 2006, pp.32]:

- The autocorrelation at lag 0 is $\rho(0) = 1$, as it represents the correlation of a data point with itself.
- The autocorrelation $\rho(s)$ at any lag s is between -1 and 1 i.e. $-1 \leq \rho(s) \leq +1$. A positive autocorrelation $\rho(s) > 0$ indicates a positive relationship between a data point and its lagged counterpart, whereas a negative autocorrelation $\rho(s) < 0$ indicates a negative relationship. A zero autocorrelation $\rho(s) = 0$ means there is no correlation between a data point and its lagged version.
- The autocorrelation is symmetric around lag s=0 ($\rho(s) = \rho(-s)$). The autocorrelaion function is plotted for the non negative lags.

In time series data with seasonality or periodic patterns, autocorrelation may exhibit periodicity at specific lags. The autocorrelation of a stationary series is determined only

by the time lag, not by the specific time points. Thus, autocorrelation remains constant over time. The pxp dimensional autocovariance matrix R_p is defined as

$$R_p = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ \vdots \\ x_{n-p+1} \end{bmatrix} \begin{bmatrix} x_n & x_{n-1} & \cdots & \cdots & x_{n-p+1} \end{bmatrix} \quad [Broersen,\ 2006, pp.33]$$

The sample mean of a stationary process is defined as

$$\hat{\mu}_x = \frac{1}{n}\sum_{t=1}^{n} x_t \quad [Broersen,\ 2006, pp.38]$$

where $\bar{x}$ is the sample mean, n is the number of observations and x_t is the set of values at time t. The sample mean $\bar{x}$ is unbiased because $\mathbb{E}(\bar{x}) = \mu$. The expectation of the sample mean is

$$\mathbf{E}[\hat{\mu}_x] = \mu \quad [Broersen,\ 2006, pp.39]$$

The sample variance is

$$\sigma_x^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \hat{\mu}_x)^2 \quad [Broersen,\ 2006, pp.38]$$

The sample variance in terms of autocorrelation and autocovariance. can be defined as

$$Var(\hat{\mu}_x) = \frac{\gamma(0)}{n}\sum_{s=-(n-1)}^{n-1}(1 - \frac{|s|}{n})\rho(s) \quad [Broersen,\ 2006, pp.40]$$

where n is the number of observations, $\gamma(0)$ is the autocovariance at lag 0, $\rho(s)$ is the autocorrelation at lag s .When $n \to \infty$ then $\lim_{n\to\infty} Var(\frac{1}{n}\sum_{t=1}^{n} x_t) = 0$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} x_t = \mu$ thus, its consistent estimator. Given a single realization x_t of a stochastic process $\{X_t\}$ at time t, the sample autocovariance function $\hat{\gamma}(s)$ at lag s is given by

$$\hat{\gamma}(s) = \frac{1}{n-s}\sum_{t=1}^{n-s}(x_t - \hat{\mu}_x)(x_{t+s} - \hat{\mu}_x) \quad [Broersen,\ 2006, pp.42]$$

where n is the length of time series, and s is a real number such that s<n. The mean of the sample autocovariance $\hat{\gamma}(s)$ at lag s is

$$\mathbb{E}(\hat{\gamma}(s)) \approx \gamma(s) - \frac{s}{n}\gamma(s) - (\frac{n-s}{n})\sigma_x^2 \quad [Broersen,\ 2006, pp.44]$$

.

where n is the number of observations, s is the lag, $\gamma(s)$ is the autocovariance at lag s, and $Var(\bar{x})$ is the variance of sample mean $\bar{x}$. When the stochatsic process $\{X_t\}$ is gaussian i.e. every finite collection of those random variables has a multivariate normal distribution, then covariance between sample autocovariance $\hat{\gamma}(s)$ at lag s and sample autocovariance $\hat{\gamma}(s+j)$ at lag s+j is defined as

$$Cov(\hat{\gamma}(s), \hat{\gamma}(s+v)) \approx \frac{\sigma_x^4}{N} \sum_{m=-\infty}^{\infty} (\gamma(m)\gamma(m+v) + \gamma(m+s+v)\gamma(m-s) \quad [Broersen,\ 2006, pp.45]$$

where $\gamma(m)$, $\gamma(m+v)$, $\gamma(m+s+v)\ and\ \gamma(m-s)$ are the autocovariance at lags m, m+v, m+s+v, and m-s respectively such that $m,\ v, and\ s\ \in \mathbb{R}$. The variance of sample autocovariance $\hat{\gamma}(s)$ at lag s is

$$Var(\hat{\gamma}(s)) \approx \frac{\sigma_x^4}{n} \sum_{m=-\infty}^{\infty} (\gamma^2(m) + \gamma(m+s)\gamma(m-s)) \quad [Broersen,\ 2006, pp.44]$$

.

where n is the number of observations, $\gamma(m)$, $\gamma(m+s)$, $and\ \gamma(m-s)$ are the autocopvariance at lags m, m+s, m-s respectively such that $m, s \in \mathbb{R}$. For a given time series x_t, the sample autocorrelation $\hat{\rho}(s)$ at lag s is given by

$$\hat{\rho}(s) = \frac{\sum_{t=1}^{n-s}(x_t - \hat{\mu}_x)(x_{t+s} - \hat{\mu}_x)}{\sum_{t=1}^{n}(x_t - \hat{\mu}_x)^2} = \frac{\hat{\gamma}(s)}{\hat{\gamma}(0)} \quad for\ s = 0,\ ,1\ ,\ 2,\ ... \quad [Broersen,\ 2006, pp.44]$$

where $\hat{\gamma}(0)$ is the sample autocovariance at lag 0, and $\hat{\gamma}(k)$ is the sample autocovariance at lag k. The mean of the sample autocorrelation can be defined as

$$\mathbf{E}[\hat{\rho}(s)] = (1 - \frac{s}{N})\rho(s) + \frac{\gamma(s)}{\gamma(0)^3} Var(\hat{\gamma}(0)) - \frac{cov[\hat{\gamma}(s), \hat{\gamma}(0)}{\gamma(0)^2} \quad [Broersen,\ 2006, pp.44]$$

where

$$Var(\hat{\rho}(0)) = \frac{2\sigma_x^4}{n} \sum_{m=-\infty}^{\infty} \rho^2(m) \quad [Broersen,\ 2006, pp.45]$$

The variance of sample autocorrelation is defined as [Broersen, 2006, pp.45]

$$Var(\hat{\rho}(s)) \approx \frac{1}{n} \sum_{m=-\infty}^{\infty} [\rho^2(m)(1+\rho^2(s)) + \rho(m+s)\rho(m-s) - 4\rho(s)\rho(m)\rho(m-s)]$$

3.3. Fourier transform, window function and periodograms:

Fourier Transform is a mathematical transformation that transforms signals between the time (or spatial) and frequency domains. It is a tool that breaks a waveform (a functional or signal) into an alternate representation characterized by sine and cosines. It shows that any waveform can be rewritten as the weighted sum of sinusoidal functions. In the Fourier transformation, the time domain, which is continuous-time and infinite duration, is transformed into the frequency domain, which is discrete, whereas in the Fourier inversion, the frequency domain into the time domain.

3.3.1. Discrete Fourier transform:

The discrete Fourier transform (DFT) is a specific implementation of the Fourier transform that uses a finite set of discrete time points. In DFT the discrete time domain are converted into a discrete frequency domain whereas, the inverse DFT transforms the discrete frequency domain into time domain. The computational complexity of DFT requires $O(N^2)$. This means that DFT requires N^2 complex multiplies and $N(N-1)$ complex additions. A DFT is very useful in many fields, including signal processing, telecommunications, audio processing, and image processing, among others. The DFT can be can be defined as

$$X(k) = \sum_{n=0}^{N-1} x(t) e^{\frac{-j2\pi kt}{N}} \qquad k = 0,\ 1,\ 2,\ ...,\ N-1 \quad [Walker,\ 2017, pp.39]$$

The inverse DFT is

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{\frac{j2\pi kt}{N}} \qquad n = 0,\ 1,\ 2,\ ...,\ N-1 \quad [Walker,\ 2017, pp.39]$$

where j is the imaginary unit, N is the length of the input sequence, $X(k)$ is the k^{th} element in the frequency domain sequence and $x(t)$ is the t^{th} element in the time domain sequence and $\pi = 3.14$. To simplify the notation define

$$W_N = e^{-j2\pi/N}$$

.

We can rewrite the DFT as

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kt} \qquad k = 0, \ , 1, \ , ..., N-1 \quad [Walker, \ 2017, pp.38]$$

and the inverse DFT as

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W_N^{-kt} \qquad n = 0, \ , 1, \ , ..., N-1 \quad [Walker, \ 2017, pp.38]$$

3.3.2. Properties of discrete fourier transform:

Let $X(k)$ be the k^{th} element in frequency domain and $X(t)$ be the t^{th} element in frequency domain. The properties of discrete fourier transform can be listed as following [Walker, 2017, pp.38]

- **Linearity:**
  Let $(x_1(t), X_1(k))$ and $(x_2(t), X_2(k))$ be two DFT pairs with the same duration of N such that $x_1(t) \overset{\text{DFT}}{\longleftrightarrow} X_1(k)$ and $x_2(t) \overset{\textbf{DFT}}{\longleftrightarrow} X_2(k)$. we have

  $$ax_1(t) + bx_2(t) \overset{\textbf{DFT}}{\longleftrightarrow} aX_1(k) + bX_2(k)$$

  where a and b are constants.

- **Periodicity:**
  Let $(x_1(t), X_1(k))$ be a DFT pair with the same duration N such that $x_1(t) \overset{\textbf{DFT}}{\longleftrightarrow} X_1(k)$ then

  $$x(t+N) = x(t) \ for \ all \ t$$

  and

  $$X(k+N) = x(k) \ for \ all \ k$$

- **Shift theorem:**
  The DFT of time-shifted sequences is determined by multiplying their original DFT by their phase factor.

  $$x(t-m) \overset{\textbf{DFT}}{\longleftrightarrow} X(k) e^{\frac{-j2\pi km}{N}}$$

  where m is an integer.

- **Convolution theorem:**
  The DFT of the convolution of two sequences is equal to the element-wise product of their individual DFTs. Let $x_1(t) \overset{\text{DFT}}{\longleftrightarrow} X_1(k)$ and $x_2(t) \overset{\textbf{DFT}}{\longleftrightarrow} X_2(k)$

  $$x_1(t) * x_2(t) \overset{\textbf{DFT}}{\longleftrightarrow} X_1(k) * X_2(k)$$

3.3.3. Fast fourier transform:

The fast fourier transform (FFT) [Walker, 2017, pp.53-54] is an algorithm that computes the DFT or its inverse more efficiently. It is used in digital signal processing, spectral analysis and image processing. FFT reduces the computational complexity of DFT from $O(N^2)$ to $O(NlogN)$ which produce same result and significantly fewer multiplication. The FFT is more parallelizable than the DFT, meaning it can be computed on multiple processors simultaneously. The main difference between an FFT and a DFT is that with an FFT, you are transforming the data in both directions at once - turning data in the time domain into data in the frequency domain and converting data in the frequency domain back into data in the time domain. One of the most widely used FFT algoritm is the Decimation in Time (DIT) Radix 2. This algorithm converts the time domain N point sequence x(n) to a frequency domain N-point sequence $X(k)$. In Decimation in Time algorithm the time domain sequence x(n) is decimated and smaller point DFT are performed. The results of smaller point DFTs are combined to get the result of N-point DFT. Let $x(t)$ be N sample sequence, we can decimate $x(t)$ into two sequences of $\frac{N}{2}$ samples. Let the two sequences be $f_1(n)$ and $f_2(n)$.Let $f_1(n)$ consists of even numbered samples of $x(n)$ and $f_2(n)$ consists of odd numbered samples of $x(t)$.

$$f_1(t) = x(2n) \qquad n = 0,\ 1,\ 2,\ ...,\ , \frac{N}{2} - 1$$

.

and

$$f_2(t) = x(2n+1) \qquad n = 0,\ 1,\ 2,\ ...,\ , \frac{N}{2} - 1$$

.

Let $X(k) = N - point$ DFT of x(n), $F_1(k) = \frac{N}{2}$ point DFT of $f_1(t)$ and $F_2(k) = \frac{N}{2}$ point DFT of $f_2(t)$. By definition of DFT the $\frac{N}{2}$ point DFT of f_1 and f_2 are given by

$$F_1(k) = \sum_{t=0}^{(\frac{N}{2})-1} f_1(n) W_{\frac{N}{2}}^{kt} \qquad k = 0,\ 1,\ ...,\ , \frac{N}{2} - 1 \quad [Walker,\ 2017, pp.53]$$

and

$$F_2(k) = \sum_{t=0}^{(\frac{N}{2})-1} f_2(t) W_{\frac{N}{2}}^{kt} \qquad k = 0,\ 1,\ ...,\ , \frac{N}{2} - 1 \quad [Walker,\ 2017, pp.53]$$

where $W_{\frac{N}{2}} = e^{\frac{-j2\pi}{\frac{N}{2}}} = e^{\frac{-j4\pi}{N}}$. Now-point DFT $X(k)$, in terms of $\frac{N}{2}$ point DFTs $F_1(k)$ *and* $F_2(k)$ is given by

$$X_k = F_1(k) + W_N^k F_2(k) \qquad k = 0,\ 1,\ ...,\ ,N-1 \quad [Walker,\ 2017, pp.54]$$

The computation of $F_1(k)$ and $F_2(k)$ each requires $(\frac{N}{2})^2 = \frac{N^2}{4}$ multiplications. The scaling by W_N^k requires $\frac{N}{2}$ additional multiplications. The total computation is $\frac{N^2}{2} + N^2$ multiplications, which represents $\frac{N^2}{2} - N^2$ fewer multiplications than the N^2 multiplications required by the DFT. Repeat the decimation until the resulting sequences are reduced to 2-point sequences.

3.3.4. Window function:

The DFT operation is processing a input signal through a set of filter banks with bandwidth centered on the bin (frequency) of interest. If the signal under study contains several spectral components and broadband noise, the computed amplitude of the DFT output in each bin (filter bank) is significantly affected by the noise contained in the corresponding bandwidth of the filter bank. When the time-limited signal slice is subjected to analysis using FFT algorithm (used for computing DFT), the FFT implicitly assumes that the signal essentially repeats itself after the observed interval. This may lead to discontinuities at the edges of each slice, that causes the energy contained in each frequency bin to spill into other bins. This phenomenon is called spectral leakage. To suppress the spectral leakage, the signals are multiplied with a window function so as to smooth the discontinuity at the edges of the FFT slices. As a result, the choice of window function affects the amount of signal and noise that goes inside each filter bank. In effect, the amplitude estimate in each frequency bin is influenced by the amount of accumulated noise contributed by the window function of choice. In other words, the noise floor displayed at the FFT output, varies according to the window of choice. In each window, the main lobe is centered at each frequency component of the time-domain signal, and the side lobes approach zero. The height of the side lobes indicates the affect the windowing function has on frequencies around main lobes. The Hann window [Prabhu, 2018, pp.148] has a sinusoidal shape with a wide peak but low side lobes. It touches zero at both ends eliminating all discontinuity. In general, the Hanning window is satisfactory in 95 percent of cases. It has good frequency resolution and reduced spectral leakage. Hann window can be defined as

$$a(k) = \frac{1}{2}(1 - cos\frac{2\pi k}{N}) \quad [Prabhu,\ 2018, pp.148]$$

where N is the length of the filter and k=0, 1, ..., N.

3.4. Software tools

The statistical software R [R Core Team, 2023] version 4.2.2 was used for this analysis. In addition to the following packages **tuneR**[Krey et al., 2023] version 3.3.2 to construct wave objects, **seewave** Tidy verse [Aubin et al., 2023] version 2.2.1 to get the dominant frequency of time wave, **fftw**[Krey et al., 2022] version 1.0.7 to calculate (inverse) DFT using the FFT method, and **dplyr** [Fransois et al., 2023] to change the name of variables.

4. Statistical analysis

This section of the report presents the results of the exploratory analysis of the data.

4.1. Data description

The variable names were made more meaningful by renaming the last substrings, specifically Akustik, Kraft, Biegemo, and Bohrst, to Acoustic, Force, Bending, and Drilling, respectively. The data and header files of V2_00001_1 were not analyzed since they duplicate torque sensor records that already exist in V2_00001. In the data files, the data is arranged in interlaced format. The interlaced character is stored in STORAGE_MODE of each header file. The data has an interlaced arrangement where each sampling includes the first channel, followed by the second channel, and so on until the last channel, then this pattern continues with subsequent samplings. In MIX mode, when a voice memo is recorded, the voice data is added to the end of the last channel's data for each sampling. The base sample frequency used during the recording is 20 kHz, and this information is stored in the RATE parameter of each header file. The RATE parameter is a crucial metadata field that indicates the sampling rate at which the data in the corresponding file was recorded. During our analysis, we scanned data recorded at a sampling frequency of 20 kHz for each of the seven channels (channels 1 to 7). The reason for using a sampling frequency of 20 kHz is that it was assumed that all the relevant frequencies to monitor were below 10 kHz. By using a sampling frequency that is at least twice the highest frequency of interest (in this case, 10 kHz), the data can be adequately captured and analyzed without aliasing or loss of information. In table 1, the order of data in V6_00001, V10_0001, V2_00001, V17_0001, V24_0001, V25a_001, and V20_0001 is shown. In D0400001, D0600001, and D080001, the order of the data is shown in table 2. The first two chapters of D* data files are the same as those in V* data files.

CH1_Moment
CH2_Force
CH3_SyncSig
CH4_Acoustic
CH5_a1_WSAS
CH6_a2_WSAF
CH7_a3_BOZA

Table 1: Data order in V* header files.

CH1_Moment
CH2_Force
CH3_Bending
CH4_SyncSig
CH5_a3_BOZA
CH6_Acoustic
CH7_a4_Drilling

Table 2: Data order in D* header files.

The parameters of the header files are presented in table 3. In the experiments conducted, only the data files labeled as D0400001, D0800001, and D060001 contain information about damper position before climbing (1240 mm) and oil pressure (12 bar). These data files were used in experiments where dampers were utilized, and as a result, they have the specific parameters related to damper position and oil pressure recorded in them. Other data files may not have this information as they were not part of experiments involving dampers. Data files D0400001, D0800001 and D060001 do not contain oil flow rate. In the data files D0400001, V6_00001, and V10_0001, the cutting speed is recorded as 111 m/min, and the feed speed is recorded as 0.231 mm/sec. However, these data files have different oil flow rates, indicating variations in the amount of oil used in each experiment. The oil flow rate is an essential parameter that affects the lubrication and cooling of the cutting process, and its variation may have implications for the experimental outcomes and results. Specifically, V6_00001 has a higher oil flow rate of 371 l/min, while V10_0001 has a lower oil flow rate of 229 l/min. The variation in oil flow rates between these two experiments may have significant implications for the cutting process, affecting factors such as lubrication, cooling, and overall machining performance. The header files D0600001, V2_00001, V17_0001, V24_0001, V25a_001 have the same cutting speed (120 m/min) and the same feed speed (0.185 mm/sec) but different oil flow rates. Oil flow rate (300 l/min) exists in V17_0001, V24_0001, and V25a_001 but not in V2_00001. D0800001 and V20_0001 have the same cutting speed (90 m/min) and feed speed (0.250 mm/sec). It appears that in the data files V2_00001 and D040001, there are variations in the duration of the recorded data. Specifically, the longest duration is found in V2_00001, where the recording duration is 291.5 seconds, while the shortest duration is found in D040001, with a recording duration of 234.43 seconds.These variations in data recording duration may be due to differences in experimental setups, data collection procedures, or specific conditions in each experiment. By varying these parameters, researchers can conduct a comprehensive analysis to identify the optimal combination of cutting conditions that result in high-quality workpieces. This approach allows for a systematic exploration of the effects of different process parameters on the final product's quality and helps to

determine the key factors that influence concentricity and surface roughness. The data collected from these experiments with varying parameters can provide valuable insights into the machining process and aid in optimizing machining conditions to achieve the desired workpiece quality. Understanding how each parameter affects the final outcome is essential for making informed decisions in manufacturing processes and ensuring the production of high-quality components. Moreover, they are interested in the case that has a higher speed (low duration) and low cost of toolwear.

In order to convert the binary data to physical values, the following formula is used

$$X_{data} * SLOPE + Y\_OFFSET$$

.

The value in data file of each chapter is represented as X_{data}, SLOPE and Y_OFFSET are the value of slope and y offset for each chapter in the header files. Each value in data file is multiplied by the slope and added to the y offset for each chapter. There are n rows and 1 column in the vector X_data, where n is the number of observations for each sensor in each data file. For V* header files, table 10 shows slopes and y offsets for each chapter. All V* data files except V2_00001 have the same slopes, which are shown in column SLOPE 2 in table 10. Table 10 shows the slopes of data file V2_00001 in column SLOPE 1. D* data files have different slopes than V* data files. The slopes of D* data files for each chapter are shown in 11. The value of Y_OFFSET is set to zero for all chapters in V* and D* data files. As the data is only 16 bits, the possible values are rather limited, and the software does some transformations to make the best use of it.

Header file	Cutting speed (m/min)	Feed speed (mm/sec)	Oil flow rate (l/min)	Duration (sec.)	Position of damper (mm)	Oil pressure (bar)
D0400001	111	0.231	-	234.43	1240	12
V6_00001	111	0.231	371	265.07	-	-
V10_0001	111	0.231	229	264.54	-	-
D0600001	120	0.185	-	268.49	1240	12
V2_00001	120	0.185	-	291.5	-	-
V17_0001	120	0.185	300	284.48	-	-
V24_0001	120	0.185	300	269.46	-	-
V25a_001	120	0.185	300	272.88	-	-
D0800001	90	0.250	-	264.46	1240	12
V20_0001	90	0.250	300	284.48	-	-

Table 3: The parameters of header files

4.2. Time series analysis of moment in V24_0001 data file

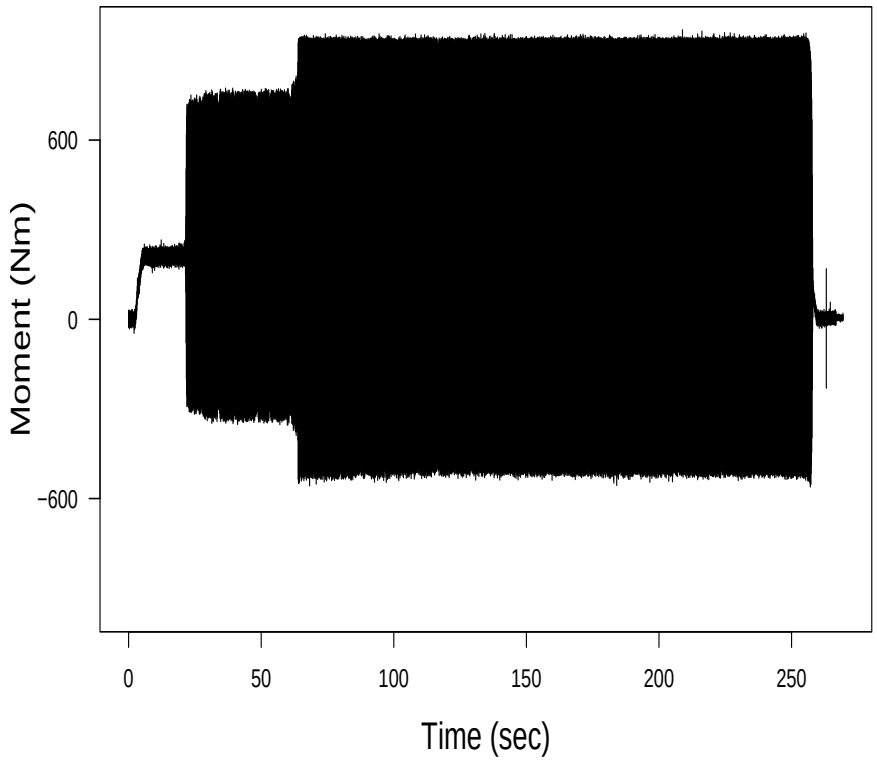

Figure 1: Moment versus time for the V24_0001 data file

Using time series and autocorrelation plots, this analysis compares different phases of the V24_0001 moment. The time series plots provide a visual representation of the moment's amplitude over time, while the autocorrelation plots reveal patterns of cyclic behavior and correlations within the moment data. The comparison allows for a deeper understanding of how the moment changes over time and how different factors, such as cutting parameters or tool wear, may influence the machining performance. In this analysis, we incorporated the concept of time series data and stationary, as described in 3.1.

The moment sensor (Nm) versus time (sec) from 0 sec to 269.46 sec is shown in figure 1. At the beginning of the process, at 2 sec., the torsional moment increase from -46 Nm to approximately 251.44 Nm. The gradual increase in the moment is attributed to the distribution of cuts on three axially offset main cutting edges with different setting angles. Due to the different setting angles, the cutting edges experience non-uniform cutting forces as they interact with the workpiece. This non-uniformity in the cutting forces leads to a gradual increase in the moment (torque) required for machining [Webber, 2006, pp.14-15]. At around 5.5 sec, the torque reaches a constant medium level. The first section from 15 to 15.03 sec, figure 2 (a), shows that the measured values in this process phase with a

small amplitude vary irregularly around an almost constant mean value. Self-excited torsional vibrations occur after approximately 22 seconds.The moment's amplitude increases continuously and rapidly from the range of 182.03 Nm to 238.71 Nm to a much larger range of -301.97 Nm to 732 Nm. After the change of state, the signal exhibits minimal noise and takes on a shape similar to a sine wave, as depicted in Figure 2 (b). Around 64 seconds, another change of state occurs, characterized by a rapid change in amplitude. At this point, the signal becomes periodic but resembles more of a square wave, exhibiting a much lower fundamental frequency compared to the previous sinusoidal signal. This behavior is illustrated in Figure 2 (c).

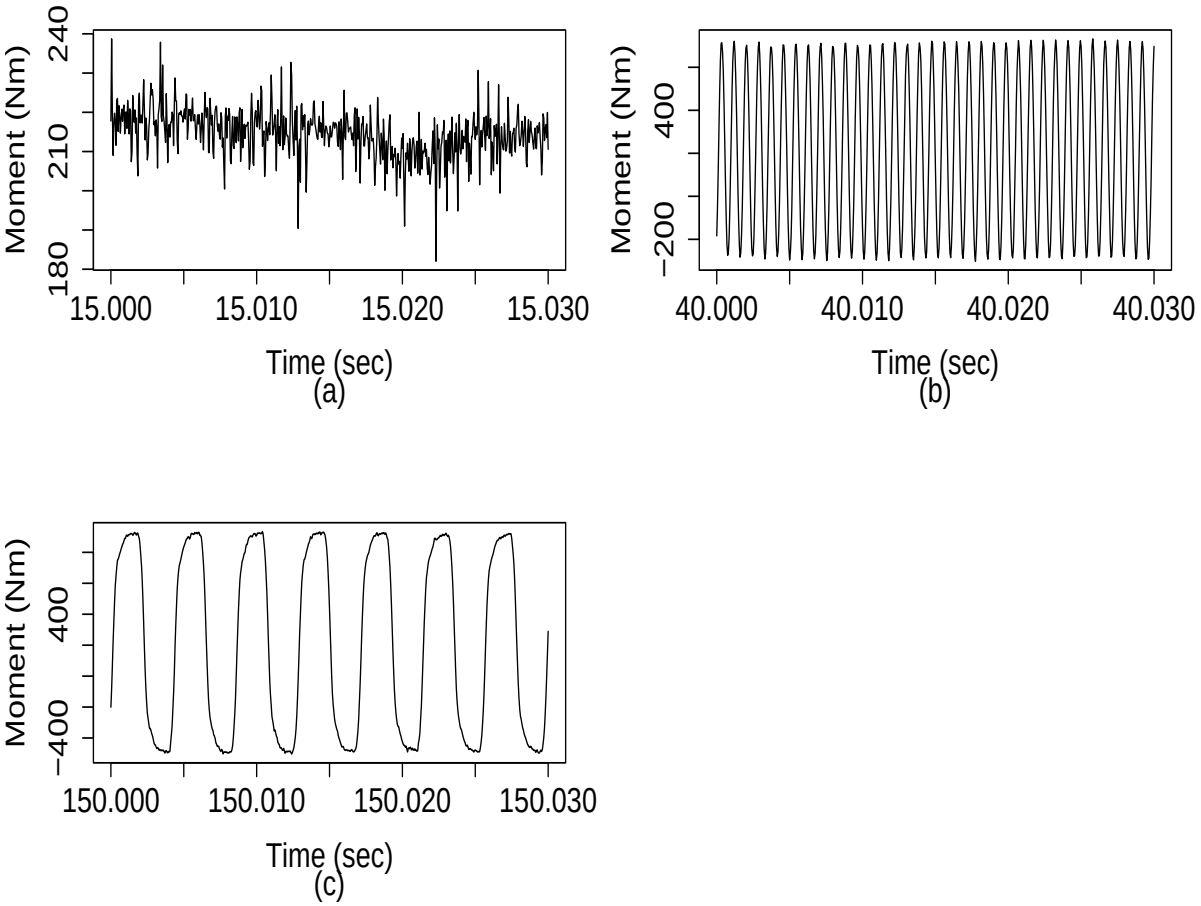

Figure 2: V24_0001 moment versus time with windows of 0.03 sec for each phase (a), (b) and (c)

The autocorrelation plots for the moment of V24_0001 during three different time intervals (15 sec to 15.03 sec, 40 sec to 40.03 sec, and 150 sec to 150.03 sec) are displayed in Figure 3 (a), (b), and (c), respectively. According to Figure 3 (a), the autocorrelation of moment V24_00001 is positive for the first 25 lags and varies within the range of 0.2 to 0.4. The fact that more than 25 lags are needed to complete a cycle indicates a very low frequency of oscillation in the moment data. This observation suggests that the moment

signal exhibits a slow, periodic behavior with a long duration for each cycle. The absence of a clear and distinct repetitive pattern in the autocorrelation plot suggests that there is no regular triggering mechanism or periodic event influencing the moment fluctuations during this specific time interval. Instead, the moment data appears to exhibit slow and random fluctuations without any specific triggering event at this particular phase of the machining process. According to Figure 3 (c), the autocorrelation of moment V24_00001 rapidly increases after 22 seconds and varies between -1 and 1. The autocorrelation plot exhibits a sinusoidal shape, suggesting a periodic behavior in the moment data. The cycle in the autocorrelation plot requires 21 lags, which corresponds to a duration of less than 0.03 seconds. This indicates a high frequency of oscillation in the moment data during this phase of the machining process. The presence of a distinct sinusoidal pattern and the short duration for each cycle suggest the occurrence of a repetitive triggering event or periodic phenomenon influencing the moment fluctuations at this specific time interval.According to Figure 3 (d), the autocorrelation of moment V24_00001 remains periodic after 64 seconds, but it takes more than 21 lags for a cycle to be completed. As a result, the frequency of oscillation has decreased compared to Figure 3 (c). The longer duration for each cycle suggests a lower frequency of repetitive oscillations in the moment data during this phase of the machining process. During chattering or spiraling phases, the autocorrelation values are higher, indicating a stronger correlation between consecutive moments. Different phases of the machining process exhibit distinct autocorrelations, signifying variations in the periodicity and dynamic behavior of the moment data.

4.3. Anomalies detecting for V6_00001 versus V10_0001

The objective of this analysis is to compare moment sensors for V6_00001 and V10_0001 under varying oil pressures, with V6_00001 having an oil pressure of 371 l/min and V10_0001 having an oil pressure of 229 l/min. Both files share fixed cutting and feed speeds of 111 m/min and 0.231 mm/sec, respectively. Additionally, it is noteworthy that V6_00001 has a duration that is 0.53 seconds longer than V10_0001. In this analysis, we employed a Hanning window of size 20,000 observations and applied the Fast Fourier Transform (FFT) method, as described in 3.3.3, to analyze the data. The periodogram and time series of the moment sensor for V6_0001 and V10_0001 are shown in figure 4. In the first phase, the torsional moment increases immediately for both V* files after the workpiece contacts the drill tip and guide rails at 30 sec. At this time, the frequency was 0.001 kHz. As a result, there is no disturbance in the process. When the self excited torsional vibration starts at 52 sec, the moment increase gradually ion both V* files to achieve the highest frequency of 1.184 kHz. At the third phase, the torsional moment increases in amplitude to take a sinusoidal shape with square wave at 91 sec for V10_0001 comparing to 135 sec for V2_00001. This phase takes longer time in V10_0001 and lower

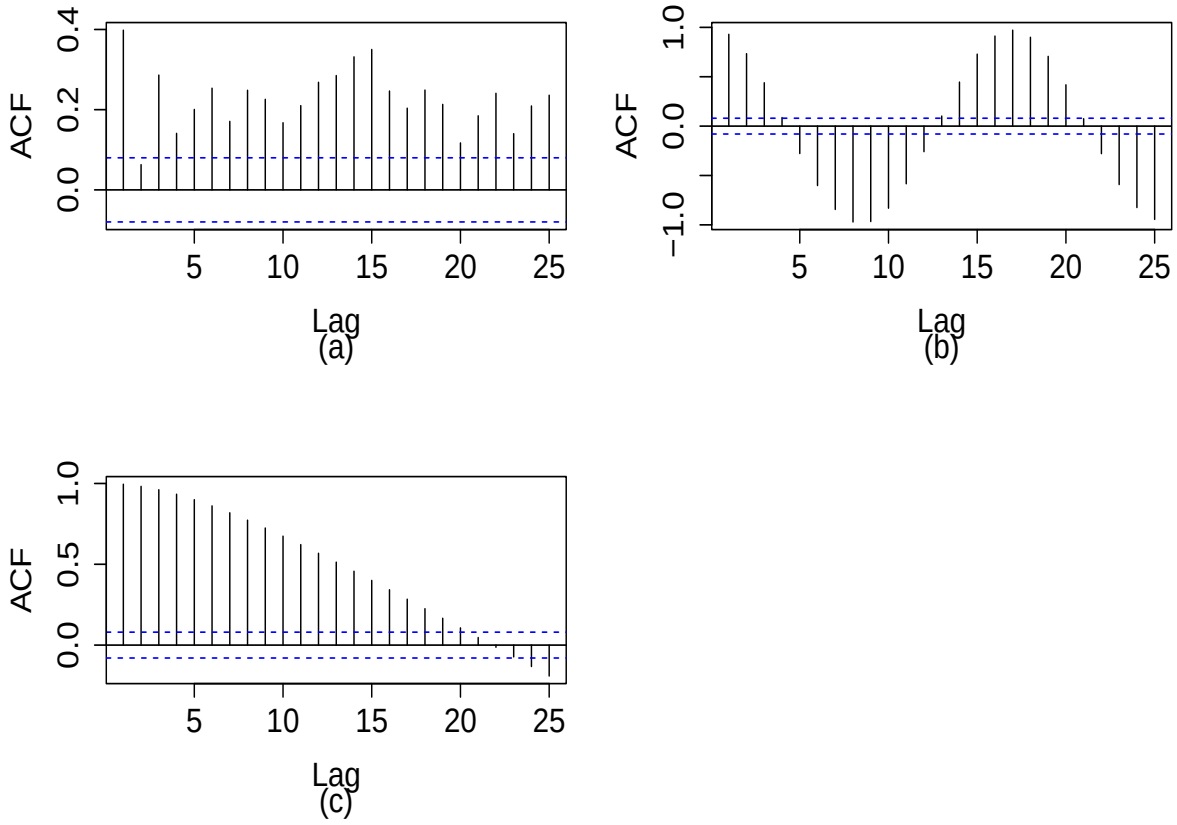

Figure 3: Moment autocorrelation plot for V24_0001 with window of 0.03 sec for each phase (a), (b), (c) and (d)

frequency of 0.236 kHz than the previous phase. As a conclusion, we can say that the lower oil flow in V10_00001 cause the chatter phase to take longer than V6_00001. Our results can't be generalized, however, because we have limited observations and experiments.

4.4. Anomalies detecting for V20_0001 versus V24_0001

This analysis compares moments sensors for V20_0001 and V24_0001 while varying cutting and feed speeds. V24_0001 has a faster cutting speed (120 m/min) than V20_0001. The feed speed of V20_0001 is higher than V24_0001. In both files, the oil pressure rate is fixed at 300 l/min. The duration of V24_0001 is 15.02 seconds longer than that of V20_0001. In this analysis, we employed a Hanning window of size 20,000 observations and applied the Fast Fourier Transform (FFT) method, as described in 3.3.3, to analyze the data. The spectrogram and time series plots of the moment sensor for V20_0001 and V24_0001 are shown in Figure 5. In V24_0001, the workpiece enters the drill tip at zero seconds, whereas in V20_0001, the workpiece enters at 11 seconds. Torsional vibration occurs earlier in V24_0001 at 11 seconds, while in V20_0001, it occurs at 22 seconds. This observation suggests that the machining process in V24_0001 experiences torsional

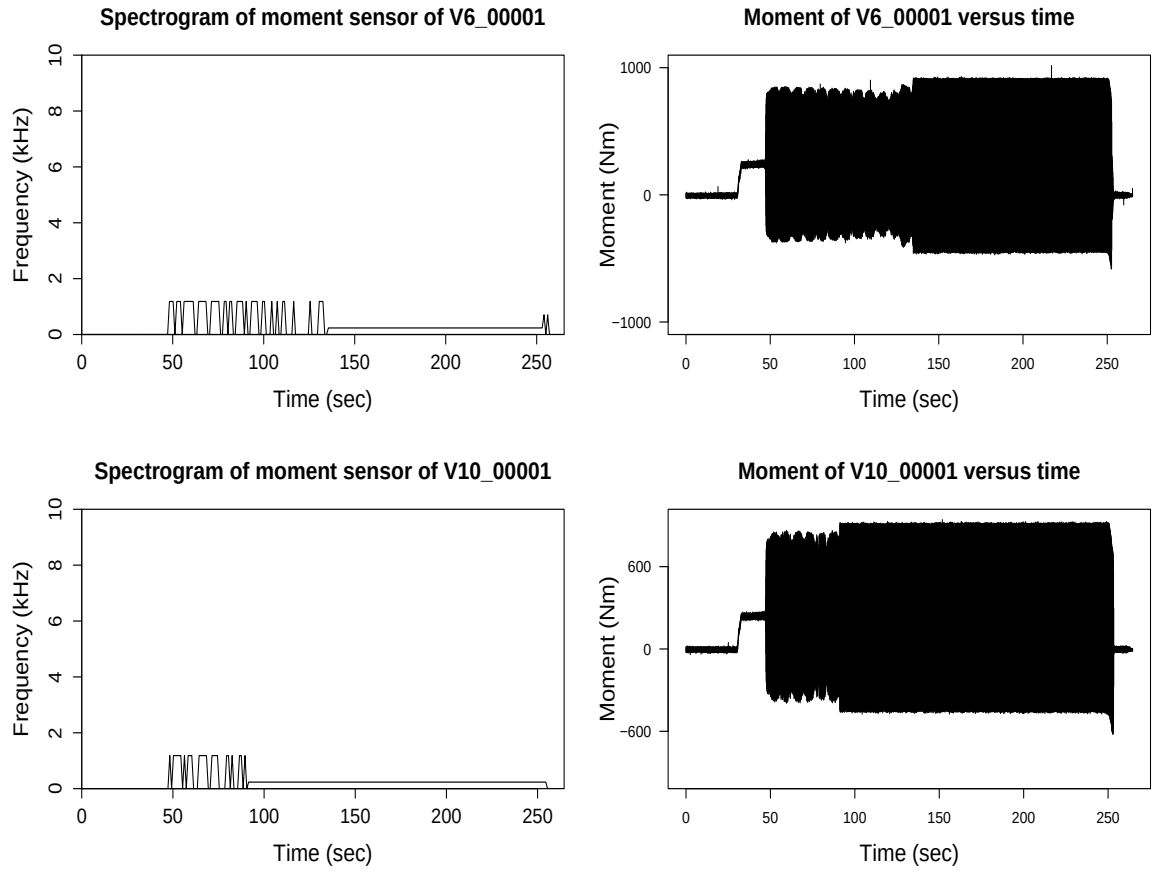

Figure 4: V6_00001 and V10_0001 periodograms and moment time series

vibrations sooner after the workpiece enters the drill tip compared to V20_0001. The early occurrence of torsional vibration in V24_0001 may be attributed to the differences in cutting parameters, workpiece entry timing, or other process-related factors between the two experiments. The second phase in V24_0001 spans from 22 to 61 seconds, while in V20_0001, it is confined to the time interval from 42 to 45 seconds. During the extended second phase of V24_0001, the moment amplitude remains higher, whereas the frequency content is lower compared to the shorter second phase of V20_0001. from During the second phase, V24_0001's moment amplitude was higher than V20_0001. At this particular time, both V20_00001 and V24_0001 had the same frequency at 42 seconds, which was 1.179 kHz. However, it's worth nothing that this frequency occurred at different time points during the machining process for the two experiments.During the machining process, the third phase occurs at 64.349 seconds in V24_0001, whereas in V20_0001, it takes place at 45 seconds. It appears that during this phase, both V20_0001 and V24_0001 exhibit similar characteristics, including a square wave sinusoidal signal and a frequency of 0.235 kHz as can be seen in table 4 and table 5. It is evident that the third phase was longer in V20_0001, lasting from 45 seconds to 272 seconds, compared to V24_0001, which spanned from 64 seconds to 259 seconds. Based on the information

provided, it is possible to draw the conclusion that the low cutting speed and high feed speed may indeed contribute to the longer duration of the third phase in V20_0001 compared to V24_0001. Lower cutting speed typically results in a slower rate of material removal from the workpiece, which could lead to a prolonged machining process. On the other hand, a high feed speed means that the cutting tool advances rapidly into the workpiece, potentially resulting in increased dynamic forces and vibrations during the cutting operation. The combination of these factors could influence the moment behavior and frequency patterns during the third phase, leading to an extended duration in V20_0001. The specific behavior characterized by the square wave sinusoidal signal and 0.235 kHz frequency may persist for a longer time due to the interaction of low cutting speed and high feed speed.

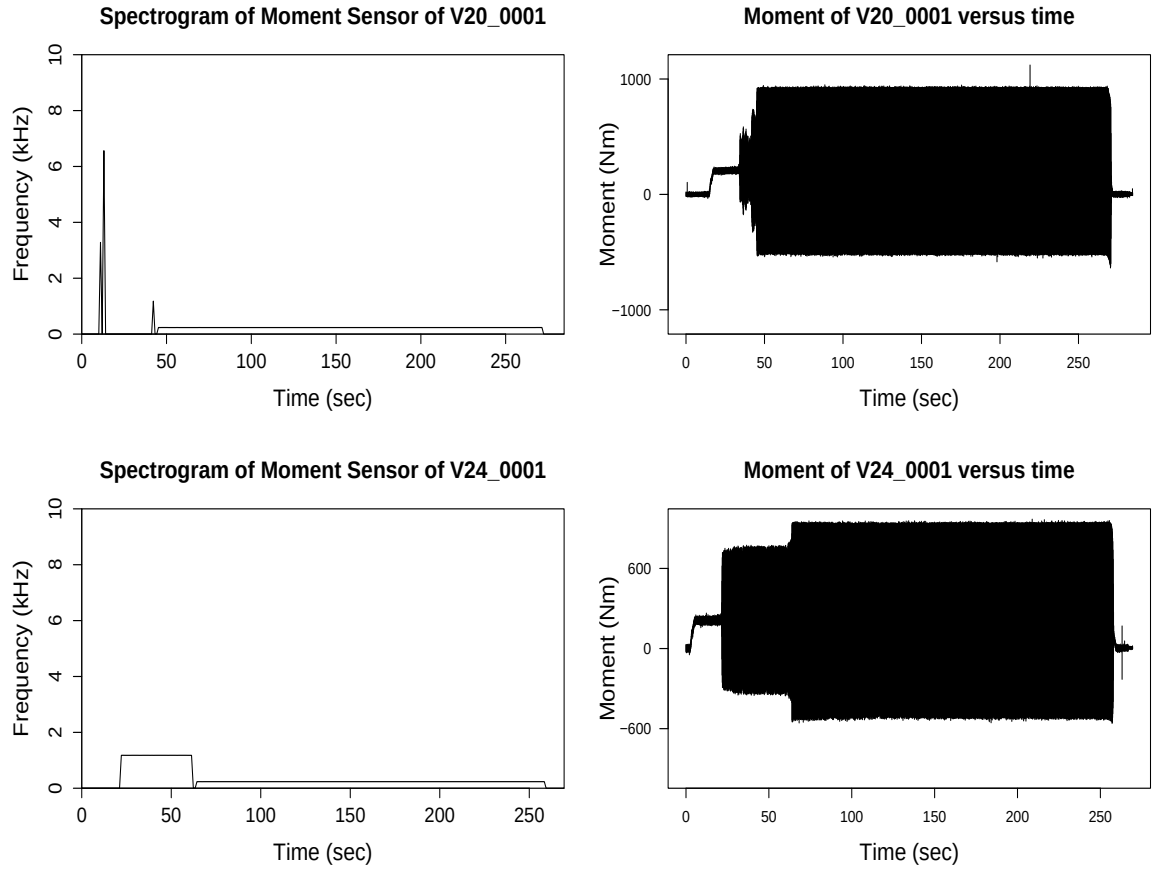

Figure 5: V20_0001 and V24_0001 periodograms and moment time series

Start time (sec.)	End time (sec.)	Frequency (kHz)
0	21	0
22	61	1.178
64	258	0.235
259	269	0

Table 4: Time, and frequency of V24_0001

Start time (sec.)	End time (sec.)	Frequency (kHz)
0	10	0
11	11	3.282
12	12	0
13	13	6.565
14	41	0
42	42	1.180
43	44	0
45	271	0.235
272	284	0

Table 5: Time, and frequency of V20_00001

5. Summary:

In this project, data was generated by the GX-1 device, producing various data files and corresponding header files with different cutting speed, feed speed, oil flow, duration, damper position, and oil flow values. During our analysis section 4.2 **??**, we observed three distinct phases in V24_0001, where the moment gradually increased. The first phase had a small amplitude, fluctuating irregularly around a nearly constant mean value. The second phase exhibited minimal noise and appeared as a sine wave-like shape. In the third phase, the signal became periodic but resembled more of a square wave, with a lower fundamental frequency compared to the previous sinusoidal signal. Autocorrelation values were higher during chattering or spiraling phases, indicating a stronger correlation between consecutive moments. Different phases of the machining process showed distinct autocorrelations, indicating variations in periodicity and dynamic behavior of the moment data.

In section 4.3 5, our analysis of V6_0001 and V10_0001 led us to conclude that the lower oil flow in V10_00001 caused the chatter phase to take longer than in V6_00001. In section 4.4 4, when comparing V20_0001 and V24_0001, we observed that lower cutting

speed and higher feed speed influenced the third phase, leading to an extended duration in V20_0001. The specific behavior characterized by the square wave sinusoidal signal and 0.235 kHz frequency persisted for a longer time due to the interaction of low cutting speed and high feed speed.

However, it's essential to note that the conclusions drawn from limited observations and specific parameter settings should be considered preliminary and not generalized without further validation and experimentation. In scientific and engineering studies, having a sufficient sample size and exploring a broader range of parameter settings is crucial to ensure the reliability and validity of the findings. The behavior of the machining process can be influenced by various factors, and a small dataset or limited parameter range might not fully capture the complexities and variations in the system. To strengthen the conclusions and generalize the results, additional experiments with different parameter combinations, larger datasets, and possibly controlled variations are necessary. This iterative process of investigation and analysis is crucial in scientific research and engineering to ensure the accuracy and robustness of the conclusions.

Bibliography

T. Aubin, C. Simonis, and J. Sueur. Seewave: Sound analysis and synthesis. 2023. URL `https://CRAN.R-project.org/package=seewave`. R package version 2.2.1.

Petrus M.T. Broersen. *Automatic Autocorrelation and Spectral Analysis*. Springer, 2006.

Romain Fransois, Lionel Henry, Kirill Mueller, Davis Vaughan, and Hadley Wickham. *dplyr: A Grammar of Data Manipulation*, 2023. URL `https://CRAN.R-project.org/package=dplyr`. R package version 1.1.2.

Stephen Haben, William Holderbaum, and Marcus Voss. *Core Concepts and Methods in Load Forecasting With Applications in Distribution Networks*. Springer, 2023.

Sebastian Krey, Uwe Ligges, and Olaf Mersmann. *fftw: Fast FFT and DCT Based on the FFTW Library*, 2022. URL `https://CRAN.R-project.org/package=fftw`. R package version 1.0-7.

Sebastian Krey, Uwe Ligges, Olaf Mersmann, and Sarah Schnackenberg. *tuneR: Analysis of Music and Speech*, 2023. URL `https://CRAN.R-project.org/package=tuneR`. R package version 1.4.4.

K. M. M. Prabhu. *Window Functions and Their Applications in Signal Processing*. CRC Press, 2018.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

James S. Walker. *Fast Fourier Transforms*. CRC Press, second edition, 2017.

Oliver Webber. Untersuchungen zur bohrtiefenabh¨angigen prozessdynamik beim bta-tiefbohren. *The Computer Journal*, pages 14–15, 2006.

A. Appendix

Variable	Data type	Description
DATASET	Categorical	The file name
VERSION	Integer	Data version number (fixed at 1)
SERIES	Categorical	Channel name follow the underbars "MEMO" is appended when voice memo is recorded in the MIX mode
DATE	DATE - format: mm-dd-yyyy	Date when recording started
TIME	TIME - format: hh:mm:ss	Time when recording started
VERT_UNITS	Categorical	Units for each channel."V" is appended when voice memo is recorded in the MIX mode
HORZ_UNITS	Categorical	Units for time axis (fixed at sec.)
COMMENT	Categorical	Comment entered in the New submenu
NUM_SERIES	Integer	Number of recording channels one channel is added when voice is recorded in the MIX mode
STORAGE_MODE	Categorical	Order of stored data fixed at INTERLACED because data is stored in the order of sampling
FILE_TYPE	Categorical	Fixed at INTEGER because data is two byte integer
SLOPE	Integer	coefficients for physical-value conversion "0.000040000" is appended when voice memo is recorded in the MIX mode
X_OFFSET	Integer	Position of the first data on the time axis Normally 0.
Y_OFFSET	Categorical	Offset for physical value conversion "0.0" is appended when voice memo is recorded in the MIX mode
NUM_SAMPS	Integer	Number of sampled data per channel
DATA	None	Indicates that the following information is unique to this device and is different from the DADiSP format.
DEVICE	Categorical	Fixed at GX-1
CH1_	Categorical	Channel number, type of amplifier range setting, and filter setting follow the under bar

CH_SLOT	Categorical	Number of channels per each slot. "1" is appended when voice memo is recorded in the MIX mode
CLOCK	Categorical	Source of sampling clock
MARK	Categorical	Number of scans at the moment an event mark is attached
GX-1_VOICE_MEMO	Categorical	Indicate the recording mode for voice memo.

Table 8: Description of the dataset variables.

Variable	Data type	Description	Units	Scale
CH1_Moment	Integer variable	Drilling torque or torsional moment	Nm	Ordinal discrete variable
CH2_Force	Integer variable	Force in feed direction	N	Ordinal discrete variable
CH3_SyncSig or CH4_SyncSig	Integer variable	Synchronous signal	V	Ordinal discrete variable
CH4_Acoustic or CH6_Acoustic	Integer variable	Sound / noise of machine	Pa	Ordinal discrete variable
CH5_a1_WSAS	Integer variable	Acceleration of the drilling head in lateral direction	m/s^2	Ordinal discrete variable
CH6_a2_WSAF	Integer variable	Acceleration of the drilling head in frontal direction	m/s^2	Ordinal discrete variable
CH7_a3_BOZA or CH5_a3_BOZA	Integer variable	Acceleration of the drilling oil supply apparatus	m/s^2	Ordinal discrete variable
CH3_Bending	Integer variable	Bending moment	m/s^2	Ordinal discrete variable
CH7_a4_Drilling	Integer variable	Acceleration sensor mounted at the drilling bar	m/s^2	Ordinal discrete variable

Table 9: Descriptiom of the dataset variables

Chapter	SLOPE 1	SLOPE 2	Y_OFFSET
CH1_Moment	0.07200000	0.04840000	0
CH2_Force	1.91600000	1.14800000	0
CH3_SyncSig	0.00080000	0.00080000	0
CH4_Acoustic	0.00020000	0.00039526	0
CH5_a1_WSAS	0.00800000	0.00800000	0
CH6_a2_WSAF	0.00800000	0.00800000	0
CH7_a3_BOZA	0.04000000	0.01000000	0

Table 10: SLOPE and Y_OFFSET of V* header files

Chapter	SLOPE 3	Y_OFFSET
CH1_Moment	0.02836000	0
CH2_Force	1.07200000	0
CH3_Bending	0.00768000	0
CH4_SyncSig	0.00080000	0
CH5_a3_BOZA	0.01000000	0
CH6_Acoustic	0.00056610	0
CH7_a4_Drilling	0.12000000	0

Table 11: SLOPE and Y_OFFSET of D* header files