Technische Universität München
TUM School of Engineering and Design

# Condition Monitoring of Machine Tool Feed Drives and Methods for the Estimation of Remaining Useful Life

**Maximilian Johann Florian Benker**

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktors der Ingenieurwissenschaften (Dr.-Ing.)**

genehmigten Dissertation.

**Vorsitz:** Prof. Dr.-Ing. Rüdiger Daub

**Prüfer*innen der Dissertation:**

1. Prof. Dr.-Ing. Michael Zäh
2. Prof. Kaan Erkorkmaz, Ph.D.
3. Prof. Chinedum Okwudire, Ph.D.

Die Dissertation wurde am 30.03.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 07.06.2023 angenommen.

# Editors' Preface

In times of global challenges, such as climate change, the transformation of mobility, and an ongoing demographic change, production engineering is crucial for the sustainable advancement of our industrial society. The impact of manufacturing companies on the environment and society is highly dependent on the equipment and resources employed, the production processes applied, and the established manufacturing organization. The company's full potential for corporate success can only be taken advantage of by optimizing the interaction between humans, operational structures, and technologies. The greatest attention must be paid to becoming as resource-saving, efficient, and resilient as possible to operate flexibly in the volatile production environment.

Remaining competitive while balancing the varying and often conflicting priorities of sustainability, complexity, cost, time, and quality requires constant thought, adaptation, and the development of new manufacturing structures. Thus, there is an essential need to reduce the complexity of products, manufacturing processes, and systems. Yet, at the same time, it is also vital to gain a better understanding and command of these aspects.

The research activities at the Institute for Machine Tools and Industrial Management (*iwb*) aim to continuously improve product development and manufacturing planning systems, manufacturing processes, and production facilities. A company's organizational, manufacturing, and work structures, as well as the underlying systems for order processing, are developed under strict consideration of employee-related requirements and sustainability issues. However, the use of computer-aided and artificial intelligence-based methods and the necessary increasing degree of automation must not lead to inflexible and rigid work organization structures. Thus, questions concerning the optimal integration of

ecological and social aspects in all planning and development processes are of utmost importance.

The volumes published in this book series reflect and report the results from the research conducted at *iwb*. Research areas covered span from the design and development of manufacturing systems to the application of technologies in manufacturing and assembly. The management and operation of manufacturing systems, quality assurance, availability, and autonomy are overarching topics affecting all areas of our research. In this series, the latest results and insights from our application-oriented research are published, and it is intended to improve knowledge transfer between academia and a wide industrial sector.

Rüdiger Daub Gunther Reinhart Michael Zäh

# Geleitwort der Herausgeber

Die Produktionstechnik ist in Zeiten globaler Herausforderungen, wie der Klimakrise, des Mobilitätswandels und der Überalterung der Gesellschaft in westlichen Ländern, für eine nachhaltige Weiterentwicklung unserer Industriegesellschaft von zentraler Bedeutung. Der Einfluss eines Industriebetriebs auf die Umwelt und die Gesellschaft hängt dabei entscheidend von den eingesetzten Produktionsmitteln, den angewandten Produktionsverfahren und der eingeführten Produktionsorganisation ab. Erst das optimale Zusammenspiel von Mensch, Organisation und Technik erlaubt es, alle Potenziale für den Unternehmenserfolg auszuschöpfen. Dabei muss größtes Augenmerk darauf gelegt werden, möglichst ressourcenschonend, effizient und resilient zu werden, um flexibel im volatilen Produktionsumfeld zu agieren.

Um in dem Spannungsfeld Nachhaltigkeit, Komplexität, Kosten, Zeit und Qualität bestehen zu können, müssen Produktionsstrukturen ständig neu überdacht und weiterentwickelt werden. Dabei ist es notwendig, die Komplexität von Produkten, Produktionsabläufen und -systemen einerseits zu verringern und andererseits besser zu beherrschen.

Ziel der Forschungsarbeiten des *iwb* ist die ständige Verbesserung von Produktentwicklungs- und Planungssystemen, von Herstellverfahren sowie von Produktionsanlagen. Betriebsorganisation, Produktions- und Arbeitsstrukturen sowie Systeme zur Auftragsabwicklung werden unter besonderer Berücksichtigung der Anforderungen des Personals sowie von Nachhaltigkeitsaspekten entwickelt. Die dabei eingesetzten rechnergestützten und Künstliche-Intelligenz-basierten Methoden und die notwendige Steigerung des Automatisierungsgrades dürfen jedoch nicht zu einer Verfestigung arbeitsteiliger Strukturen führen. Fragen der optimalen Einbindung ökologischer und sozialer Aspekte in alle Planungs- und Entwicklungsprozesse spielen deshalb eine sehr wichtige Rolle.

Die im Rahmen dieser Buchreihe erscheinenden Bände stammen thematisch aus den Forschungsbereichen des *iwb*. Diese reichen von der Entwicklung von Produktionssystemen über deren Planung bis hin zu den eingesetzten Technologien in den Bereichen Fertigung und Montage. Die Steuerung und der Betrieb von Produktionssystemen, die Qualitätssicherung, die Verfügbarkeit und die Autonomie sind Querschnittsthemen hierfür. In den *iwb*-Forschungsberichten werden neue Ergebnisse und Erkenntnisse aus der praxisnahen Forschung des Institutes veröffentlicht. Diese Buchreihe soll dazu beitragen, den Wissenstransfer zwischen dem Hochschulbereich und den Anwendenden zu verbessern.

Rüdiger Daub Gunther Reinhart Michael Zäh

# Acknowledgements

This dissertation was written while I was working as a research associate at the Institute for Machine Tools and Industrial Management (*iwb*) at the Technical University of Munich.

I would like to thank Prof. Dr.-Ing. Michael Zäh for giving me the chance to be part of this unique institute and the affirmation and guidance I received. I would also like to thank Prof. Kaan Erkorkmaz and Prof. Chinedum Okwudire for the careful review of my dissertation as well as Prof. Dr.-Ing. Rüdiger Daub for chairing the examination committee.

Special thanks to all my colleagues at the *iwb*, especially the ones in my research group "Machine Tools". It was such a great pleasure to work with so many highly motivated and brilliant people. Keep up the good work and let us keep in touch for many years to come. Thank you, Roman, Johannes and Max for your comments on the draft of this thesis. I also want to thank all the students that I had the privilege of supervising. You supported me with great diligence and commitment. Further thanks go to Prof. Dr. Olga Fink and her team, who provided me with a magnificent retreat at the Chair for Intelligent Maintenance Systems at ETH Zurich for three months, where the last building blocks of this thesis were created whilst having a view over the city into the distant mountains.

There are not enough words to express my gratitude to my family. My parents and siblings play an immense role in my life. You have always been there for me, gave me so much love, support and trust and finally taught me how to lead a happy and content life. Thank you so much for everything! Finally, thank you, Lena for not only putting up with me while writing this thesis, but also for supporting me every single day and enriching my life to such an extent that I can hardly wait to embark on our future together.

Munich, October 2023 Maximilian Benker

# Contents

# Nomenclature

The used acronyms, abbreviations, notations and symbols are listed here. Multiple mentioning may occur, since sometimes varying symbols were used to describe the same things in the different publications of this publication-based thesis.

## Acronyms and Abbreviations

| short | description |
|---|---|
| AE | auto encoder |
| AI | artificial intelligence |
| ANN | artificial neural network |
| ARIMA | auto-regressive integrated moving average |
| BPF | ball passing frequency |
| CBM | condition-based maintenance |
| CNC | computer numerical control |
| CNN | convolutional neural network |
| DES | discrete event simulation |
| DIN | German Institute for Standardisation (in German: Deutsches Institut für Normung) |
| DL | deep learning |
| DOI | digital object identifier |
| ELBO | evidence lower bound |
| EMA | experimental modal analysis |
| FFNN | feed forward neural network |
| FFT | fast Fourier transform |
| FRF | frequency response function |
| GP | Gaussian process |

*(Acronmys and Abbreviations continued)*

| | |
|---|---|
| GPC | Gaussian process classification |
| GPR | Gaussian process regression |
| HHT | Hilbert-Huang transform |
| HI | health indicator |
| HMC | Hamiltonian Monte Carlo |
| ISO | International Organization for Standardization |
| *iwb* | Institute for Machine Tools and Industrial Management of TU Munich (in German: Institut für Werkzeugmaschinen und Betriebswissenschaften der TU München) |
| KF | key finding |
| KL | Kullback-Leibler |
| LSCF | least-squares complex frequency domain |
| MAE | mean absolute error |
| MCMC | Markov chain Monte Carlo |
| MDoF | multi degree of freedom |
| MEMS | micro-electrochemical systems |
| MH | Metropolis-Hastings |
| ML | machine learning |
| ML-II | type II maximum likelihood |
| NLP | natural language processing |
| OMA | operational modal analysis |
| PCA | principal component analysis |
| PDF | probability density function |
| PdM | predictive maintenance |
| PHM | prognostics and health management |
| RBF | radial basis function |
| RBM | restricted Boltzmann machine |
| ReLU | rectified linear unit |
| RG | research gap |
| RMS | root mean square |
| RMSE | root mean squared error |
| RNN | recurrent neural network |
| RT | research target |
| RUL | remaining useful life |
| RVM | relevance vector machine |
| SDoF | single degree of freedom |

*(Acronmys and Abbreviations continued)*

| | |
|---|---|
| sq. | and the following one |
| sqq. | and the following ones |
| std | standard deviation |
| STFT | short-time Fourier transform |
| SVM | support vector machine |
| TUM | Technical University of Munich |
| VI | variational inference |
| WN | white noise |

## Notations

If not indicated otherwise, vectors are written as bold type (e.g. $\mathbf{x}$) and matrices are written capitalized (e.g. $X$). Unknown parameters estimated from data are denoted with a hat symbol. For example, $\hat{\omega}_r$ is an estimated natural frequency of mode $r$ and $\hat{\theta}$ is an estimated (hyper-)parameter of a machine learning model.

| **notation** | **description** |
|---|---|
| $a \coloneqq b$ | definition; a is defined to be equal to b |
| $\mathrm{Cov}(x, y)$ | covariance of the random variables $x$ and $y$; also denoted as $\sigma^2_{x,y}$ |
| $\mathcal{D}$ | data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with $i = 1, \ldots, N$ pairs of inputs $\mathbf{x}_i$ and corresponding outputs $y_i$ |
| $\det(X)$ | determinant of a matrix |
| $\mathrm{diag}(\mathbf{x})$ | diagonal matrix containing the elements of vector $\mathbf{x}$ |
| $\mathrm{diag}(x_i)$ | diagonal matrix containing the elements $x_i$ with $i = 1, \ldots, N$ |
| $\mathbb{E}(x)$ | expected value of a random variable $x$ |
| $\exp(x)$ | exponential function with argument $x$ |
| $(f * g)(t)$ | convolution of two time-dependent functions $f$ and $g$ |
| $(f \circ g)(x)$ | composition of functions $f$ and $g$ applied on $x$ |
| $\mathcal{GP}$ | Gaussian process: $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, the function $f$ is a random function from a Gaussian process with mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ |
| $I$ | identity matrix |
| $k(\mathbf{x}, \mathbf{x}')$ | covariance function of a Gaussian process |

*(Notations continued)*

| | |
|---|---|
| $K(X,X)$ | covariance matrix in the context of a Gaussian process |
| $\log(x)$ | natural logarithm of $x$ (base $e$) |
| $m(\mathbf{x})$ | mean function of a Gaussian process |
| $\mathcal{N}(\mathbf{x}\|\boldsymbol{\mu},\Sigma)$ | the variable $\mathbf{x}$ follows a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ |
| $\mathcal{N}\left(x\|\mu,\sigma^2\right)$ | the variable $x$ follows a Gaussian distribution with mean $\mu$ and variance $\sigma^2$ |
| $\mathrm{Var}(x)$ | variance of a random variable $x$; also denoted as $\sigma_x^2$ |
| $\|x\|$ | absolute value of a variable |
| $\|\|\mathbf{x}\|\|$ | Euclidean norm of a vector $\mathbf{x}$ |
| $\bar{x}$ | sample mean of a random variable $x$ |
| $\mathbf{x}^\top, X^\top$ | transpose of a vector $\mathbf{x}$ or matrix $X$ |
| $\ell$ | lengthscale parameter of the radial basis function (RBF) covariance function |

## Symbols

| Latin | description | unit |
|---|---|---|
| $a$ | acceleration | $\mathrm{m/s^2}$ |
| $A_{r,kl}$ | modal constant of the receptance frequency response function of mode $r$ with excitation at point $l$ and response at point $k$ | 1/kg |
| $a_x$ | acceleration in X-direction | $\mathrm{m/s^2}$ |
| $a_y$ | acceleration in Y-direction | $\mathrm{m/s^2}$ |
| $a_z$ | acceleration in Z-direction | $\mathrm{m/s^2}$ |
| $\tilde{b}$ | optimal correction factor for the predicted remaining useful life value | – |
| $b$ | bias of a neural network neuron | – |
| $b_s$ | sample skewness of a random variable $x$ | – |
| $C$ | viscous damping matrix | N s/m |
| $C_m$ | modal viscous damping matrix | N s/m |
| $C_r$ | basic dynamic radial load rating | N |
| $c_s$ | security factor to account for a user's risk averseness | – |

*(Symbols continued)*

| | | |
|---|---|---|
| $c_{sl}$ | slope of the leaky rectified linear unit (ReLU) activation function | – |
| $c_{total}$ | total expected costs | – |
| $d$ | dimensionality of the input vector $\mathbf{x}$ in publication ② | – |
| $d$ | number of selected features in publication ③ | – |
| $F$ | force | N |
| $\mathbf{F}$ | force vector | N |
| $\mathbf{F}_a$ | amplitude of a harmonic force vector | N |
| $H$ | hysteretic damping matrix | N/m |
| $H(j\omega)$ | experimental frequency response function | – |
| $h^{(l)}(x)$ | activation function of a neural network neuron on layer $l$ | – |
| $j$ | imaginary unit of a complex number $c = a + bj$ | – |
| $K$ | stiffness matrix | N/m |
| $K_m$ | modal stiffness matrix | N/m |
| $k_r$ | modal stiffness of mode $r$ | N/m |
| $L_{10}$ | nominal service life in revolutions | – |
| $L_w$ | window size of the sliding window pre-rpocessing step | – |
| $M$ | mass matrix | kg |
| $M_m$ | modal mass matrix | kg |
| $m_r$ | modal mass of mode $r$ | kg |
| $p_d$ | share of the last degraded observations of a run-to-failure sequence | – |
| $p_h$ | share of the first healthy observations of a run-to-failure sequence | – |
| $P_{mech}$ | mechanical power | W |
| $P_r$ | dynamic equivalent radial load | N |
| $s^2$ | sample variance of a random variable $x$ | – |
| $S_{FF}$ | power spectrum of the excitation $F$ | – |
| $S_{Fx}$ | cross-power spectrum between the excitation force $F$ and the response $x$ | – |
| $t$ | time subscript | s |
| $t_{RUL}$ | true remaining useful life value in publication ⑤ | h |
| $\hat{t}_{RUL}$ | predicted remaining useful life value in publication ⑤ | h |
| $w$ | weight between two neural network neurons | – |
| $\mathbf{x}$ | position vector | m |

*(Symbols continued)*

| | | |
|---|---|---|
| $\ddot{\mathbf{x}}$ | acceleration vector; second time derivative $d^2\mathbf{x}/dt^2$ | m/s$^2$ |
| $\dot{\mathbf{x}}$ | velocity vector; first time derivative $d\mathbf{x}/dt$ | m/s |
| $\mathbf{x}_a$ | time-independent amplitudes of a position signal | m |
| $\bar{x}_\omega$ | mean frequency of the spectrum of a time series | rad/s |
| $y$ | true remaining useful life value in publication ③ and publication ④ | cycles |
| $y^\star$ | predicted remaining useful life value in publication ③ and publication ④ | cycles |

| **Greek** | **description** | **unit** |
|---|---|---|
| $\Sigma$ | covariance matrix of a multivariate Gaussian distribution | – |
| $\Psi$ | modal matrix with mode shape vectors as columns | – |
| $\Omega$ | diagonal matrix of squared natural frequencies | (rad/s)$^2$ |
| $\alpha$ | scale parameter of a Weibull distribution | – |
| $\beta$ | shape parameter of a Weibull distribution | – |
| $\zeta_r$ | modal damping ratio of mode $r$ | – |
| $\eta_r$ | damping loss factor of mode $r$ | – |
| $\boldsymbol{\mu}$ | mean of a Gaussian distributon | – |
| $\rho$ | parameter of an auto-regressive integrated moving average (ARIMA) model | – |
| $\sigma$ | standard deviation of a Gaussian distribution | – |
| $\tau$ | prediction error $y^\star - y$ | cycles |
| $\boldsymbol{\psi}_r$ | mode shape vector (i.e. eigenvector) of mode $r$ | – |
| $\omega$ | angular frequency | rad/s |
| $\omega'_r$ | damped natural frequency of mode $r$ | rad/s |
| $\omega_C$ | center frequency of the spectrum of a time series | rad/s |
| $\omega_D$ | dominant frequency of the spectrum of a time series | rad/s |
| $\omega_r$ | natural frequency of mode $r$ | rad/s |
| $\epsilon$ | noise term | – |
| $\boldsymbol{\phi}_r$ | mass-normalized mode shape vector of mode $r$ | kg$^{-0.5}$ |

Chapter 1

# Introduction

According to a study by Deutsche Bundesbank (2021), the growth rates of the labour productivity, which serves as a measure of the increase of prosperity, have significantly declined in the European manufacturing sector. The study further states that one of the reasons for this situation is decelerated growth in the efficiency of production processes.

Because of that, policy-makers have undertaken different attempts to reverse the economic slowdown, one of which is the European Commission's initiative on artificial intelligence (AI) (European Commission 2018). In addition to treating socio-economic challenges and ethical issues related to AI, the initiative aims to boost the industrial AI uptake and ultimately increase productivity and economic growth (Annoni et al. 2018, p. 77). A study by Aghion et al. (2017) indeed supports the idea that further automating manufacturing activities with the help of AI can lead to permanent and stable future economic growth rates.

Therefore, it is no surprise that the machine tool industry has already placed the topic on its agenda. In a yearly market report, Verein Deutscher Werkzeugmaschinenfabriken e.V. (2020, p. 27) listed AI as a technological trend and described it as a "[...] logical consequence of a constantly advancing digiti[s]ation". The German Academic Association for Production Technology echoed this in its study by Krüger et al. (2019), in which the authors also acknowledged the potentially large performance-enhancing effects of AI on manufacturing systems.

Examples of especially promising use cases in the context of manufacturing are the concepts of condition monitoring and predictive maintenance (PdM), which are meant to be an attempt to increase machine operability and availability by assessing the current degradation level of a machine and anticipating future breakdowns. Based on this information, maintenance actions can be scheduled

just in time (PERES et al. 2020). In fact, companies which have already successfully implemented a PdM strategy reported machine availability gains of up to 9 % (BIGGIO and KASTANIS 2020).

One of the prerequisites for applying a PdM strategy is the availability of accurate information about the current wear condition and, at best, even accurate remaining useful life (RUL) estimates. The research field prognostics and health management (PHM) addresses, amongst other issues, the derivation of such information. Although the field has recently grown rapidly, and an increase in scientific publications was observed (LEI et al. 2018), a widespread application to machine tools and especially to machine tool feed drives has not yet occurred.

On the one hand, this is due to gaps in the state of the art, according to which experiments are mainly conducted on simplified single-axis test benches in laboratory setups and, on the other hand, industrial use cases are restrictive due to difficult circumstances, such as the absence of sufficient historic failure data (BUTLER et al. 2022). Hence, additional work is needed to improve the state of the art and help to eventually exploit condition monitoring and predictive maintenance systems for machine tool feed drives.

## 1.1 Objectives

The work presented in this publication-based thesis aims to contribute to the state of the art by investigating approaches for condition monitoring of machine tool feed drives and methods for the estimation of RUL values. In particular, the objective of this thesis is to derive a condition monitoring approach for machine tool feed drives, which is applicable to industrial machine tools and can assess the condition of previously unseen feed drive components. Furthermore, this thesis attempts to present novel approaches for the estimation of RUL values which, on the one hand, produce accurate prediction results and, on the other hand, fulfil the requirements from industrial use cases, for which historic run-to-failure data are often not available.

## 1.2 Thesis Structure

In order to achieve these aims, the thesis is structured as follows: Chapter 2 will lay out the theoretical foundations. This includes machine tool feed drives, the general PHM process and the relevant machine learning (ML) methods applied in this thesis. Chapter 3 will review the state of the art with respect to condition monitoring of machine tool feed drives and methods for the estimation of RUL values. Based on this review, research gaps (RGs) will be identified, and research targets (RTs) as well as a research approach will be derived in Chapter 4. The publications, which this thesis is based upon and which aim at addressing the identified RTs, will be summarised in Chapter 5 including a critical review of the results. In Chapter 6, the economic potential of the research approach will be investigated. In Chapter 7, a conclusion and an outlook will be given.

Chapter 2

# Theoretical Foundations

This chapter presents the theoretical foundations of the thesis. First, the components under investigation (i.e. the critical machine tool feed drive components) will be described in Section 2.1. Second, the general process of prognostics and health management (PHM), which is the underlying methodology used in this thesis, will be described in Section 2.2. Third, the machine learning (ML) algorithms applied in this thesis will be introduced in Section 2.3.

## 2.1 Machine Tool Feed Drives

Feed drives are key components of modern machine tools as they significantly determine the manufacturing accuracy and the economic efficiency. Their task is to realise the relative positioning of the workpiece and the tool (ALTINTAS et al. 2011). Requirements for feed drives are, among others, the following: jerk-free movement even at low feed velocities, high positioning accuracy, the accomplishment of small travel distances ($\ll 1\,\mathrm{mm}$) and high reliability.

In large-scale machine tools, rack-and-pinion, rack-and-worm and linear direct drives are commonly used. In small to medium-sized machine tools, feed drives with ball screws are most frequently used because of their favourable properties, such as low rolling friction, the absence of stick-slip effects, long service life and high positioning accuracy (WECK and BRECHER 2006b, pp. 8, 82 sqq.).

Since ball screw feed drives are the most widespread feed drive type (ALTINTAS et al. 2011), they were selected to be the subject of this work. In the following section, they will be described in more detail.

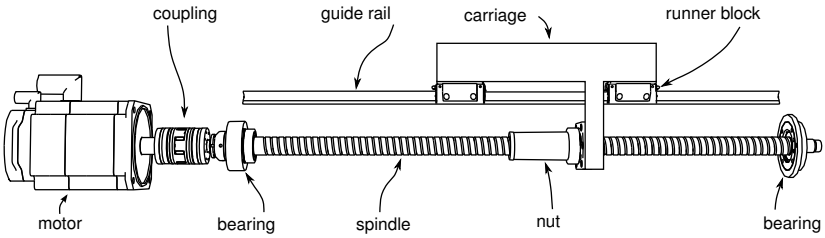



**Figure 2.1:** Schematic illustration of a machine axis with a ball screw feed drive

### 2.1.1 Structure of Ball Screw Feed Drives

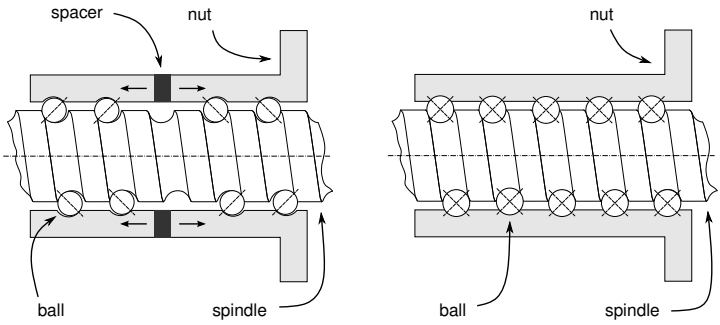
Ball screw feed drives generally consist of all the mechanical components that are within the force flux between the motor and the tool or workpiece. This includes couplings, ball screws, bearings, an optional gearbox and linear guides (WECK and BRECHER 2006b, p. 81). In Fig. 2.1, an exemplary machine axis with a ball screw feed drive is depicted.

The ball screw, which consists of a spindle and nut assembly, is usually mounted between bearings at each end of the spindle. It should be noted, that the bearing on the right-hand side of the depiction in Fig. 2.1 is optional. The nut has recirculating balls and is usually preloaded to avoid backlash and to ensure high rigidity (SPIESS 1970).

The linear guides allow motion in one degree of freedom only and hold a carriage via runner blocks and guide rails. This carriage is also connected to the ball screw nut. The ball screw spindle has a predefined pitch and is connected to a motor through a coupling. Hence, the ball screw feed drive system transforms the rotary motion of the motor into a linear motion of the carriage. Below, the two sub-assemblies ball screw and linear guide will be described in more detail.

#### Ball Screws

As mentioned earlier, ball screws are often preloaded to avoid backlash and to achieve higher rigidity within the feed drive system. The preload can be set either by creating an offset in the nut, by bracing two single nuts against each other (e.g. by introducing a spacer into the nut) or by using oversized balls (ALTINTAS et al. 2011). The latter two are depicted in Fig. 2.2 as examples. Some manufacturers also provide ball screws with an adjustable preload screw, which allows a user to flexibly set the preload level. The preload that is predefined by the manufacturer




**(a)** Ball screw nut with preload introduced by a spacer resulting in a two point contact between each ball, the spindle and the nut

**(b)** Ball screw nut with preload introduced by oversized balls resulting in a four point contact between each ball, the spindle and the nut

**Figure 2.2:** Schematic illustration of two exemplary ways of introducing preload to ball screws based on ALTINTAS et al. (2011)

is sometimes referred to as static preload. According to a study by VERL and FREY (2010), another load component, which linearly correlates with the rotational speed of the shaft, is added to the static preload during operation, resulting in the so-called effective preload on the ball screw nut. In order to retain the balls within the nut-screw-assembly, they are recirculated by either external return tubes, or internal return channels or internal recirculation caps.

The general design and calculation of ball screw drives are strongly use case dependent and are described in the International Organization for Standardization (ISO) standards ISO 3408-2 (2021), ISO 3408-3 (2006), ISO 3408-4 (2006), and ISO 3408-5 (2006).

#### Linear Guides

Similar to the ball screws, linear guides can be found in many designs. In general, common machine tool linear guide systems can be categorised into hydrodynamic, hydrostatic, aerostatic, electromagnetic and rolling element systems (WECK and BRECHER 2006a, pp. 219 sq.):

**Hydrodynamic linear guides** In hydrodynamic linear guides, a lubricant is applied to the contact area between the guide rails and the carriage without additional pressure. With relative movement between the contact surfaces, the lubricant becomes involved, and a lubricant film forms between the two surfaces.

**Hydrostatic and aerostatic linear guides** In contrast to hydrodynamic linear guides, hydrostatic and aerostatic systems provide a lubricant for the contact area with additional pressure generated from an external pressure system. This leads to a consistent lubricant film between the contact surfaces. In the case of a hydrostatic system, the lubricant is a liquid. In the case of an aerostatic system, the lubricant is a gas, mostly ambient air.

**Electromagnetic linear guide systems** They allow contact-free operation based on electromagnetic levitation. The advantages include friction-free operation, low maintenance costs and adjustable structural behaviour. They are still subject to active research and have only been introduced to machine tool prototypes for a long time (DENKENA et al. 2014; KRÜGER et al. 2022).

**Rolling element linear guides** They are the most frequently applied type of linear guides. This is because of their many advantages, such as the absence of stick-slip effects, high availability and low maintenance costs, as well as the presence of rolling friction, which results in smooth running characteristics. The available designs differ in terms of the rolling element type used (rollers or balls), the recirculation system design and the rail type used (profiled or non-profiled guideways).

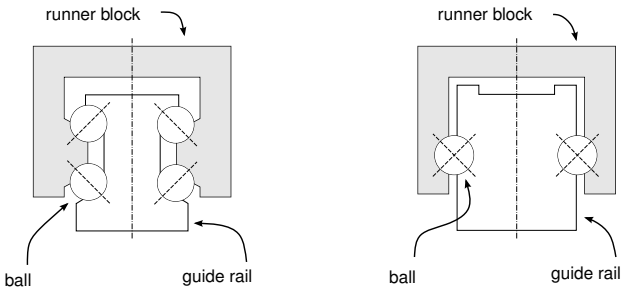
Linear guides can also be preloaded for increased rigidity, which is achieved by using oversized balls or by adjustable components of the rails. Depending on the linear guide design, this results in a two point or a four point contact between the balls, the guide rail and the runner block (see Fig. 2.3 for exemplary designs). The latter has the advantage of allowing designs with only two rows of balls and, therefore, has the potential of being more compact. (WECK and BRECHER 2006a, pp. 217 sqq.)

The general design and calculation of rolling element linear guides are described in the ISO standards ISO 12090-1 (2011) and ISO 12090-2 (2011).

### 2.1.2 Service Life and Wear of Ball Screw Feed Drives

For ball screws, a nominal service life $L_{10}$ (in revolutions) can be calculated. This is the value that 90 % of a group of similar components within similar operating conditions can reach. It is defined in ISO 281 (2007) and is given by

$$L_{10} = 10^6 \cdot \left( \frac{C_r}{P_r} \right)^p , \tag{2.1}$$




**(a)** Preloaded linear guide with four rows of two point contact balls

**(b)** Preloaded linear guide with two rows of four point contact balls

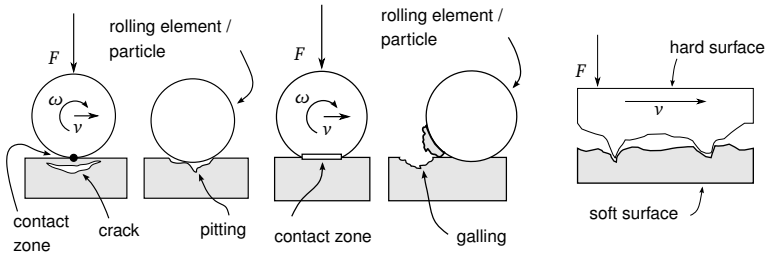
**Figure 2.3:** Schematic illustration of two exemplary linear guide designs based on WECK and BRECHER (2006a, p. 365)

where $C_r$ is the basic dynamic radial load rating and $P_r$ is the dynamic equivalent radial load. When balls are used as rolling elements, $p = 3$, whereas when rollers are used, $p = 10/3$. Although the calculation of the nominal service life of ball screw drives is widespread in the industry, feed drives can still unexpectedly fail due to different wear mechanisms, which will be described below.

Many definitions of wear and its patterns exist in the literature, which often leads to confusion (LUDEMA 1996). Therefore, a basic definition of wear for the purpose of this thesis will be given in this subsection, before the relevant wear mechanisms and patterns for machine tool feed drives will be described. The following content is a summary of the studies by HABERKERN (1998), MATE and CARPICK (2019), and SPATH et al. (1995). The reader is kindly referred to these three references for an in-depth study.

Generally, wear can be described as "[...] the removal of material when one solid surface rubs against another [...]" (MATE and CARPICK 2019, p. 389). Although this is a rather generic and broad description, it clearly illustrates the root cause of wear in ball screw feed drives. As both, the ball screw and the linear guide assembly rely on rolling elements, a mixture of rolling and sliding is present during operation, and wear naturally occurs. The relevant wear mechanisms and patterns for this thesis are mentioned below and are illustrated in Fig. 2.4.

**Fatigue** When metal friction partners slide or roll at low speeds, plastic deformation is the dominant wear mechanism. During this process, the mechanical stresses evoked by sliding or rolling exceed the yield stress of either one or both




**(a)** Illustration of fatigue resulting in the wear pattern pitting

**(b)** Illustration of adhesive wear resulting in the wear pattern galling

**(c)** Illustration of abrasive wear resulting in micro-cutting and the wear pattern scratches

**Figure 2.4:** Illustration of relevant wear mechanisms for ball screw feed drives

friction partners. When this contact stress is alternated by repetitively rolling over the contact areas between the ball screw spindle, the nut and the balls, embrittlement and subsurface cracks emerge. Such a deformation can grow over time and eventually lead to small incoherent fractions being severed from the material on the surface. This wear pattern is called pitting and leads to uneven running behaviour and eventually to the fracture of the ball screw nut.

**Adhesive wear** Another form of plastic deformation is adhesive wear, which occurs when the adhesive forces between two friction partners are high enough to sever fractions of material from one or both of them. The higher these adhesive forces are in the contact area, the higher is the risk of fragments of material being severed from the surface. These friction partners can be not only rolling elements and the ball screw spindle or the nut, but also pollutants from outside the ball screw assembly. Adhesive wear can occur during rolling and sliding friction, but it is more common for the latter. It leads to the segregation of larger coherent fractions of material from the surface. The resulting wear pattern is called galling.

**Abrasive wear** When the surface of one friction partner is substantially harder than the surface of the other, abrasive wear can occur. In this case, small particles are removed from the less hard friction partner. As a result, this wear mechanism is sometimes also referred to as micro-cutting. One can further differentiate between three-body abrasion, in which wear is caused by hard particles between the friction partners, and two-body abrasion, in which wear is caused by the harder friction partner directly. Abrasive wear occurs only during sliding. In ball

screw feed drives, abrasive wear leads to a scratched wear pattern on the surface of the ball screw spindle, the balls and the nut.

According to HABERKERN (1998), in practical applications, the three wear mechanisms highlighted above can lead to the following three main fault modes in ball screw drives:

**Early breakdown** After a fraction of the calculated lifetime, the ball screw drive shows uneven running behaviour, which eventually leads to its fracture. This is due to pitting damage starting to evolve in the recirculation system, where the motion of the balls is less characterised by rolling, but rather by sliding and sudden impulses. With ongoing operation, the damage spreads to the rolling elements and the spindle. This fault mode is facilitated by high spindle pitches and high rotational speeds and is caused by fatigue and adhesive wear.

**Late breakdown** This fault type usually occurs only when a ball screw is operated for longer than the calculated service lifetime according to Eq. (2.1). At the beginning, uneven running behaviour is observed, which gradually becomes worse, eventually leading to the fracture of the ball screw. This is due to pitting damage that begins on either the surface of the spindle or the surface of the rolling elements first, before spreading to other parts, such as the recirculation system. This fault type is caused by fatigue and adhesive wear.

**Slow preload loss** The preload set at the beginning of the lifetime gradually decreases with the operation of the ball screw. This is observed as a reduction in the rolling elements' diameter, which leads to a decrease in the stiffness and rigidity of the entire feed drive system, which can lead to chatter during machining and eventually to poor surface quality, problems with dimensional accuracy and backlash. This fault type is facilitated by high and oscillating acceleration and is induced by abrasive wear.

These three fault modes outlined above are regarded as the natural fault modes of ball screw feed drives, as they occur even in cases of correct assembly, optimal lubrication, protection against pollution and the absence of excess loads.

Designing and selecting suitable ball screw drives involve a trade-off. On the one hand, achieving high stiffness and rigidity through a high preload is desirable to fulfil dynamic requirements. On the other hand, a higher preload level directly leads to higher friction, higher wear, accelerated degradation and

changing dynamic behaviour (WEULE and GOLZ 1991). Hence, preload and wear directly influence the dynamics of machine tool feed drive systems. The following subsection will briefly introduce how to describe and quantify this dynamic behaviour.

### 2.1.3 Modal Analysis and Modal Testing for Machine Tools

A common method for determining the dynamic characteristics of engineering systems is *modal analysis*. It describes the dynamics of a system by quantifying the so-called *modal parameters*: natural frequencies (or eigenfrequencies), damping ratios and mode shapes. With the help of these modal parameters, a mathematical model, which describes the dynamic behaviour of a system, can be formulated. Modal analysis can be performed numerically and experimentally. The former assumes a physical model with certain mass, stiffness and damping properties. According to these properties, the system dynamics can be simulated and analysed either analytically or numerically, for example, by using modern finite element techniques. Experimental modal analysis, or sometimes also called *modal testing*, aims at identifying the modal model directly from experimental data. Both, modal analysis and modal testing are briefly described in the following. The following content is based on the studies by ALTINTAS (2012), EWINS (2000), and HE and FU (2001).

#### Modal Analysis

A machine tool can be described as a multi-degree-of-freedom system with a mass matrix $M$, a viscous damping matrix $C$, a hysteretic damping matrix $H$ and a stiffness matrix $K$ by the equation of motion[1]

$$M\ddot{\mathbf{x}}_t + C\dot{\mathbf{x}}_t + (K + jH)\,\mathbf{x}_t = \mathbf{F}_t, \tag{2.2}$$

where $\ddot{\mathbf{x}}_t$ is a vector of accelerations, $\dot{\mathbf{x}}_t$ is a vector of velocities, $\mathbf{x}_t$ is a vector of positions, $\mathbf{F}_t$ is a vector of applied forces, $t$ is the time subscript and $j$ is the imaginary unit.

[1] For better readability, a time-dependent vector $\mathbf{x}(t)$ is denoted as $\mathbf{x}_t$.

**Undamped system** Since the damping properties can make solving for Eq. (2.2) difficult, damping is often neglected for weakly damped structures. Setting the external forces $\mathbf{F}_t$ to zero results in

$$M\ddot{\mathbf{x}}_t + K\mathbf{x}_t = \mathbf{0}, \tag{2.3}$$

and its solution takes the form

$$\mathbf{x}_t = \mathbf{x}_a e^{j\omega t}, \tag{2.4}$$

where $\mathbf{x}_a$ is a vector of time-independent position amplitudes and $\omega$ is the angular frequency. Substituting Eq. (2.4) and

$$\ddot{\mathbf{x}}_t = -\omega^2 \mathbf{x}_a e^{j\omega t} \tag{2.5}$$

into Eq. (2.3) leads to

$$\left(K - \omega^2 M\right)\mathbf{x}_a e^{j\omega t} = \mathbf{0}. \tag{2.6}$$

If the trivial case $\mathbf{x}_a = \mathbf{0}$ is ignored, then the solutions must fulfil

$$\det\left(K - \omega^2 M\right) = 0. \tag{2.7}$$

Solving this eigenvalue problem leads to the natural frequencies (i.e. the square root of the eigenvalue) $\omega_r$ and mode shapes (i.e. eigenvectors) $\boldsymbol{\psi}_r$ of mode $r = 1, \ldots, N$. The eigenvalues $\omega_r^2$ and mode shapes $\boldsymbol{\psi}_r$ can be further combined by the eigenmatrix $\Omega$ and the so-called modal matrix $\Psi$:

$$\Omega = \operatorname{diag}(\omega_r^2) = \begin{bmatrix} \omega_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_N^2 \end{bmatrix}, \qquad \Psi = \begin{bmatrix} \boldsymbol{\psi}_1 & \boldsymbol{\psi}_2 & \cdots & \boldsymbol{\psi}_N \end{bmatrix}. \tag{2.8}$$

With the modal matrix $\Psi$, the modal stiffness matrix $K_m$ and the modal mass matrix $M_m$ can be calculated using

$$K_m = \Psi^\top K \Psi = \operatorname{diag}(k_r) \tag{2.9}$$

and

$$M_m = \Psi^\top M \Psi = \operatorname{diag}(m_r). \tag{2.10}$$

Therefore, the modal stiffness $k_r$ and the modal mass $m_r$ can be obtained for each mode separately with

$$k_r = \boldsymbol{\psi}_r^\top K \boldsymbol{\psi}_r \tag{2.11}$$

and

$$m_r = \boldsymbol{\psi}_r^\top M \boldsymbol{\psi}_r. \tag{2.12}$$

The ratio of the modal stiffness and modal mass provides the squared natural frequency

$$\omega_r^2 = \frac{k_r}{m_r}, \tag{2.13}$$

and the mass-normalizsed mode shapes can be calculated using

$$\boldsymbol{\phi}_r = \frac{\boldsymbol{\psi}_r}{\sqrt{m_r}}. \tag{2.14}$$

With the mass-normalised mode shapes, $\boldsymbol{\phi}_r^\top K \boldsymbol{\phi}_r = \omega_r^2$ and $\boldsymbol{\phi}_r^\top M \boldsymbol{\phi}_r = 1$ hold. Hence, in the absence of (or the presence of extremely low) damping, each mode can be described by its mass-normalised mode shape and its natural frequency.

From the results above, the resulting decoupled receptance frequency response function (FRF) for the harmonic force excitation $\mathbf{F}_t = \mathbf{F}_a e^{j\omega t}$ at point $l$ and response $\mathbf{x}_t = \mathbf{x}_a e^{j\omega t}$ at point $k$ is given by

$$\alpha_{kl}(\omega) = \frac{\mathbf{x}_{t,k}}{\mathbf{F}_{t,l}} = \sum_{r=1}^{N} \frac{A_{r,kl}}{\omega_r^2 - \omega^2} \tag{2.15}$$

with $A_{r,kl}$ being the so-called *modal constant*, which is defined as

$$A_{r,kl} = \phi_{r,k} \cdot \phi_{r,l}, \tag{2.16}$$

where $\phi_{r,k}$ and $\phi_{r,l}$ are the $k^{th}$ and $l^{th}$ element of the mass-normalised mode shape vector $\boldsymbol{\phi}_r$, respectively.

**Proportional viscously damped system** If viscous damping is added to the system described in Eq. (2.3), then the equation of motion changes to

$$M\ddot{\mathbf{x}}_t + C\dot{\mathbf{x}}_t + K\mathbf{x}_t = \mathbf{0}. \tag{2.17}$$

Assuming that the viscous damping matrix $C$ is proportional to the mass and stiffness matrices, it can be written as

$$C = \alpha M + \beta K, \quad \alpha, \beta \in \mathbb{R}^{+}. \tag{2.18}$$

Substituting Eq. (2.18) into Eq. (2.17) leads to an equation that can be decoupled with the same $\Psi$ as computed in Eq. (2.8). With the viscous modal damping matrix

$$C_m = \boldsymbol{\Psi}_r^{\top} C \boldsymbol{\Psi}_r = \mathrm{diag}(c_r) \tag{2.19}$$

containing the modal viscous damping $c_r$ of each mode, the damped system can be uncoupled to $N$ equations, each describing a single-degree-of-freedom system with

$$\omega_r' = \omega_r \sqrt{1-\zeta_r^2} \tag{2.20}$$

and

$$\zeta_r = \frac{c_r}{2\sqrt{k_r m_r}} = \frac{c_r}{2\omega_r m_r}, \tag{2.21}$$

where $\omega_r'$ is the damped natural frequency and $\zeta_r$ is the modal damping ratio. The resulting receptance FRF is given by

$$\alpha_{kl}(\omega) = \frac{\mathbf{x}_{t,k}}{\mathbf{F}_{t,l}} = \sum_{r=1}^{N} \frac{A_{r,kl}}{\left(\omega_r^2 - \omega^2\right) + j2\zeta_r \omega_r \omega}. \tag{2.22}$$

In this case, assuming a proportional damping matrix $C$ leads to a result, which is similar to the undamped case derived in Eq. (2.15). The main difference is that now the denominator is complex.

There are more elaborate ways for incorporating damping into the model from Eq. (2.2), such as adding hysteretic damping to the model or discarding the assumption of proportional damping. However, they are not within the scope of this thesis. For further details on this topic, the reader is kindly referred to EWINS (2000).

#### Modal Testing

In modal testing, the dynamic behaviour of a system can be characterised by experimentally measuring its FRF. This can be achieved by exciting the system with a certain dynamic force $F$, which leads to a dynamic system response $x$. In an idealised measuring setup without noise, an experimental FRF $H(j\omega)$ can be

estimated by dividing the Fourier transform of the response $x(j\omega)$ by that of the excitation $F(j\omega)$:

$$\hat{H}(j\omega) = \frac{x(j\omega)}{F(j\omega)}. \tag{2.23}$$

However, the measurements are usually subject to noise. With noisy measurement results, the FRF can be estimated with

$$\hat{H}_1(j\omega) = \frac{S_{Fx}(j\omega)}{S_{FF}(j\omega)} \tag{2.24}$$

or with

$$\hat{H}_2(j\omega) = \frac{S_{xx}(j\omega)}{S_{xF}(j\omega)}, \tag{2.25}$$

where $S_{Fx}$ and $S_{xF}$ are the cross-power spectra between the excitation and the response, $S_{FF}$ is the power spectrum of the excitation and $S_{xx}$ is the power spectrum of the response. Under ideal conditions without noise in the signals, the statement $\hat{H}(j\omega) = \hat{H}_1(j\omega) = \hat{H}_2(j\omega)$ holds.

The estimated FRF $\hat{H}(j\omega)$ serves as the basis for the extraction of the modal parameters (i.e. the natural frequencies $\omega_r$, the mode shapes $\boldsymbol{\psi}_r$ and the damping ratios $\zeta_r$). The estimation of the modal parameters from FRFs is an extensive field within the research community and is only briefly touched in this thesis.

In general, methods for modal parameter identification have been developed for both, the time and the frequency domain, as well as for single-degree-of-freedom and multi-degree-of-freedom systems.

A simple method for modal parameter identification in the frequency domain is the so-called *peak-amplitude method*, which identifies a system based on each mode separately. In EWINS (2000, pp. 306 sq.) the method is described as follows: first, the natural frequencies $\hat{\omega}_r$ are identified by picking the peaks in the FRF. Second, for each peak, the maximum absolute value of the FRF at the identified natural frequency $\left|\hat{H}(j\hat{\omega}_r)\right|$ is noted. Third, the frequency ranges $\Delta\omega = \omega_b - \omega_a$, which capture a decrease of the amplitude of the fraction $\left|\hat{H}(j\hat{\omega}_r)\right| / \sqrt{2}$, are determined. The resulting damping loss factor for each mode can then be estimated as

$$\hat{\eta}_r = 2\hat{\zeta}_r = \frac{\omega_b^2 - \omega_a^2}{2\hat{\omega}_r^2}, \tag{2.26}$$

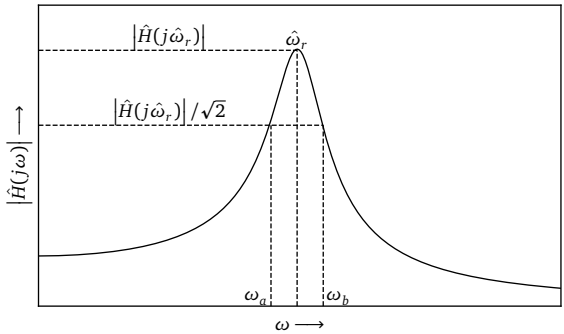



**Figure 2.5:** Illustration of the peak-amplitude method for the estimation of the modal parameters of a single-degree-of-freedom system based on EWINS (2000, p. 308)

and the modal constant for each mode can be calculated with

$$\hat{A}_r = \left|\hat{H}(j\hat{\omega}_r)\right| \hat{\omega}_r^2 \hat{\eta}_r. \tag{2.27}$$

The peak-amplitude method is illustrated in Fig. 2.5. Another exemplary and more advanced method for the identification of modal parameters for multi-degree-of-freedom systems is the *PolyMAX* method, which was originally presented by GUILLAUME et al. (2003). The method is a generalisation of the least-squares complex frequency domain (LSCF) method (VAN DER AUWERAER et al. 2001) and identifies modal parameters by solving a weighted least-squares problem.

An advantage of the PolyMAX method is its ability to identify modal parameters for multi-degree-of-freedom systems with close and even strongly overlapping modes.

## 2.2 Prognostics and Health Management

In order to ensure and conserve the productivity of a machine tool, maintenance needs to be conducted. This can be done with a reactive maintenance strategy to replace worn-out components after an unexpected machine breakdown. In contrast, planned maintenance is performed according to a fixed schedule. This avoids unplanned failures at the cost of potentially conducting maintenance too early and more frequently than necessary. Modern maintenance strategies include

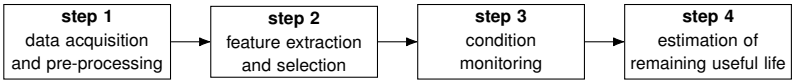



**Figure 2.6:** General PHM workflow as pursued in this thesis

condition-based maintenance (CBM) and predictive maintenance (PdM). The former initiates maintenance actions depending on the current health assessment of the machine, whereas the latter predicts the development of the machine's degradation level in the near future and schedules future maintenance actions accordingly. (FARRAR and WORDEN 2012, pp. 3 sq.)

Research on PHM focusses on the health assessment and the prediction of the degradation of a system of interest based on historic and real-time data collected during operation. This allows for the application of a CBM or even a PdM strategy, potentially leading to a reduction in maintenance costs. (KIM et al. 2017, pp. 1–2)

Historically, PHM techniques were first used in defence, space and aerospace applications. An early example is the so-called *Joint Strike Fighter PHM system*, whose goal was to enhance the safety of military air planes that are operated for longer than originally intended due to shrinking military budgets (HESS and FILA 2002). In another study, VOLPONI et al. (2004) provided an example for the aerospace industry. They developed a PHM system for aircraft engines. Nowadays, PHM techniques are used in many industries and applications, such as in wind power systems (LAU et al. 2012), lithium-ion batteries (MENG and LI 2019) and manufacturing (VOGL et al. 2019). Such widespread application of PHM can be explained by the necessity of further cutting operating costs on the one hand, and the availability of cost-effective sensors, cheap computation power and advanced algorithms, on the other hand. (SUN et al. 2012)

Many researchers have developed typical PHM workflows, such as JARDINE et al. (2006), LEI et al. (2018) and VOGL et al. (2019). While the designed workflows slightly differ in minor details, the general PHM workflow can be summarised and described by the following four steps (see also Fig. 2.6): data acquisition and pre-processing, feature extraction and selection, condition monitoring and the estimation of the remaining useful life (RUL). However, the latter two are often summarised as diagnosis and prognosis, hence resulting in three main steps. Each of the three steps will be briefly introduced in the following sections.

### 2.2.1 Data Acquisition and Pre-processing

The first step in every PHM workflow is the acquisition of data, upon which the diagnosis and prognosis can later be based. Depending on the investigated component, different measurements can be sensibly applied. A variety of common measurements are listed in ISO 13379-1 (2012). While the selection of a specific type of measurement and sensor is an individual decision dependent on the specific component under investigation, LEE et al. (2014) reviewed some of the most frequently applied ones and provided an overview thereof. They found that vibrations, acoustic emissions, temperatures, currents and voltages are widely used signals for monitoring mechanical and electrical components, such as gears, bearings, shafts, pumps and alternators.

After the raw sensor signals have been acquired, a pre-processing step is often necessary. This step is rather technical and includes all the actions needed to clean the recorded data. This can include the alignment of different sensor signals in time, the handling of missing or redundant data (ATAMURADOV et al. 2020) or simply the standardisation of the data formats (BEKAR et al. 2020).

### 2.2.2 Feature Extraction and Selection

With a data set at hand, the second step of a typical PHM workflow, which is the extraction and selection of wear-sensitive features, can be performed. This step is considered to be crucial, as the acquired raw data often consist of high-frequency time series and, therefore, contain redundant information and can thus be largely uninformative (FULCHER 2018). The goal of feature extraction is to find wear-sensitive representations of the original raw data that can serve as an input for diagnostic and prognostic models (KIMOTHO and SEXTRO 2014). In general, feature extraction techniques are often divided into the following three categories: *time domain*, *frequency domain* and *time-frequency domain* techniques (TSUI et al. 2015). After the construction of features, selecting only the most informative ones for further analysis can sometimes be reasonable. For example, after the construction of frequency-domain features by applying a fast Fourier transform (FFT) to the raw time series, it can be useful to further exploit only certain frequency ranges related to the failure or wear mechanism of interest (ELLINGER et al. 2019). Overall, the literature on feature extraction and selection methods is vast, and a comprehensive treatment of the topic is beyond the

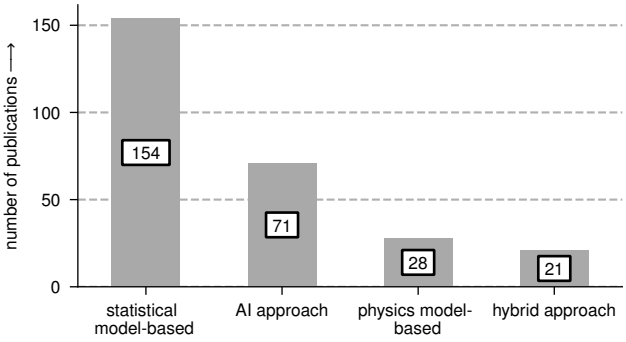



**Figure 2.7:** Overview of the different RUL prediction models as reviewed by (LEI et al. 2018); data-driven models were applied more often than physics model-based and hybrid approaches.

scope of this work. However, some simple feature extraction techniques will be described in more detail and with examples in Subsection 2.3.1.

### 2.2.3 Diagnosis and Prognosis

From the collected data and the extracted wear-sensitive features, a diagnosis of the current degradation state (also referred to as condition monitoring) and a prognosis of its future development can be made. The latter ultimately leads to an estimate of the RUL, which can be utilised within a predictive maintenance strategy. (LEE et al. 2014)

LEI et al. (2018) conducted an extensive review of the applied models for diagnosis and prognosis. They assigned the reviewed techniques to the following four categories: artificial intelligence (AI) approaches, statistical model-based, physics model-based and hybrid approaches. They also provided an overview of the frequency of the application of the respective techniques in the literature (see Fig. 2.7). Recently, AI approaches have received widespread popularity as a result of the availability of large data sets and affordable computing power (BIGGIO and KASTANIS 2020). Statistical model-based and AI approaches often share the same models and inference principles, and the boundaries between the two are debatable (GHAHRAMANI 2015). Therefore, both approaches can be combined into one category and viewed as data-driven approaches (TSUI et al. 2015). The resulting three categories will be briefly described below:

**Physics model-based approaches** Those approaches aim to model the failure mechanism directly with basic physical laws in order to derive an RUL estimate. Although this approach seems attractive because of its explainability, in many cases it cannot be applied, as either a complete understanding of the failure mechanism is missing or the system under investigation is too complex.

**Data-driven approaches** In contrast, data-driven approaches do not assume any knowledge about the underlying failure mechanism. Rather, they solely rely on historic data, which are fit to a probabilistic model (SI et al. 2011). This assumes the availability of informative historic data.

**Hybrid approaches** The combination of the advantages of physics model-based and data-driven approaches is referred to as hybrid approaches.

Especially deep learning (DL) models (i.e. deep artificial neural networks (ANNs)) have been established as reliable and powerful tools in the recent past (ZHANG et al. 2019). Driven by advances from other application areas, such as computer vision and speech recognition, DL provides a large number of promising model architectures along with powerful software ready for application within the field of PHM (FINK et al. 2020). The AI models applied in this thesis will be further described in the Subsections 2.3.2 and 2.3.3.

## 2.3 Machine Learning

In the following subsections, feature extraction techniques and the relevant ML models for this thesis will be introduced. Different features will be explained, since they often serve as inputs for an ML model and can strongly influence its prediction accuracy. Therefore, some exemplary ones will be introduced in Subsection 2.3.1. The relevant ML models for this thesis are ANNs, which will be described in Subsection 2.3.2, and Gaussian process (GP) models, which will be explained in Subsection 2.3.3.

### 2.3.1 Feature Extraction

Extracting features is a crucial step in any data analysis process because it can significantly influence the success of later applied ML models. The following

aspects describe the objectives of a feature extraction process (GUYON and ELISSEEFF 2006):

**General data reduction** Finding a set of features that can reduce or compress the original data can help to save storage space and ultimately speed up the training process of an ML model. However, compressing data may lead to an information loss and to a decline in the predictive performance of an ML model.

**Performance improvement** When the main goal is to make accurate predictions, feature extraction can transform the raw data into a more compact and informative data set. This is especially the case when the raw data contain many redundant values.

**Data understanding** When, instead of making accurate predictions, the main goal is to gain a better understanding of and knowledge about the data, feature extraction can be helpful. In fact, the modal analysis and modal testing techniques outlined in Subsection 2.1.3 are examples of a feature extraction process that facilitates the user's understanding of the data and the system under investigation.

Common features from time series can be categorised into time-domain, frequency-domain and time-frequency-domain features (LEI et al. 2020). Since in this thesis mainly the extraction of features from time series data is relevant, some features will be introduced and illustrated in the following Examples 2.1 to 2.3.

**Example 2.1** (Time-Domain Features)**.** Features can be directly extracted from the raw time series in the time domain. Simple time-domain features include estimates of statistical moments, such as the sample mean $\bar{x}$, the sample variance $s^2$ and the sample skewness $b_s$. For a univariate time series $\mathbf{x}_t = \begin{bmatrix} x_1 & \cdots & x_T \end{bmatrix}^\top$ with $T$ samples, they can be estimated using (HAMILTON 1994, pp. 740 sq.)

$$\bar{x} = \frac{1}{T}\sum_{i=1}^{T} x_i, \tag{2.28}$$

$$s^2 = \frac{1}{T-1}\sum_{i=1}^{T}(x_i - \bar{x})^2 \tag{2.29}$$

and (JOANES and GILL 1998)

$$b_s = \frac{1/T\sum_{i=1}^{T}(x_i - \bar{x})^3}{\left(1/(T-1)\sum_{i=1}^{T}(x_i - \bar{x})^2\right)^{3/2}}. \tag{2.30}$$

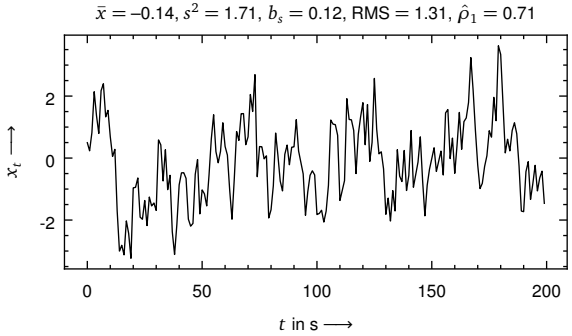



**Figure 2.8:** Illustration of an exemplary time series with 200 samples; the calculated time-domain features $\mathbf{z} = \begin{bmatrix}\bar{x} & s^2 & b_s & \text{RMS} & \hat{\rho}_1\end{bmatrix}^\top = \begin{bmatrix}-0.14 & 1.71 & 0.12 & 1.31 & 0.71\end{bmatrix}^\top$ can be used to represent the original time series $\mathbf{x}_t$. This leads to a data reduction and, depending on the use case, potentially enhances the predictive performance.

Another frequently utilised time-domain feature is the root mean square (RMS) value, which is defined as (TSUI et al. 2015)

$$\text{RMS} = \sqrt{\frac{1}{T}\sum_{i=1}^{T} x_i^2}. \tag{2.31}$$

Among the more complex time-domain features are the parameters of an auto-regressive integrated moving average (ARIMA)(p,d,q) model, where p, d and q are the model orders. In its simplest form it is the ARIMA(1,0,0) model, which is defined as (BOX et al. 2016, p. 54)

$$x_t = \rho_1 x_{t-1} + \epsilon_t, \tag{2.32}$$

where $\epsilon_t$ is a Gaussian error term and $\rho_1$ is the only unknown model parameter. It can be estimated from the raw time series as $\hat{\rho}_1$ and can serve as a time-domain feature. In contrast to the above-mentioned time-domain features, the ARIMA model parameters do not capture amplitude-related information but rather auto-regressive information about the time series (i.e. information about the correlation between the lagged time series' values). The mentioned features are exemplarily shown in Fig. 2.8.

In contrast, frequency-domain features are calculated on the basis of the frequency components of the raw time series. Therefore, the raw time series $\mathbf{x}_t$ should first be transformed to the frequency domain. This can be achieved by

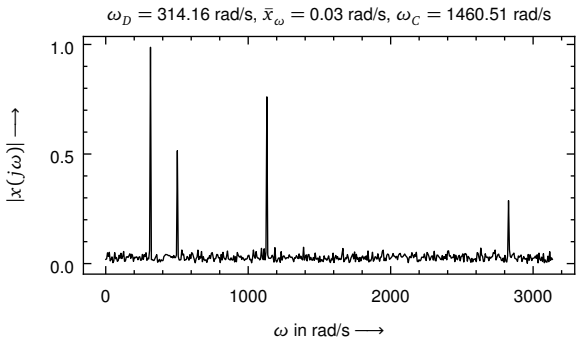



**Figure 2.9:** Illustration of an exemplary frequency spectrum of a time series; the calculated frequency-domain features are given by the vector $\mathbf{z} = \begin{bmatrix}\omega_D & \bar{x}_\omega & \omega_C\end{bmatrix}^\top = \begin{bmatrix}314.16 & 0.03 & 1460.51\end{bmatrix}^\top$ rad/s.

applying a fast Fourier transform (FFT) resulting in a symmetric complex vector from which the one-sided spectrum $x(j\omega)$ can be selected. (JARDINE et al. 2006)

**Example 2.2** (Frequency-Domain Features)**.** Various features can be extracted from the frequency-domain signal. Among the simple ones are the dominant frequency $\omega_D$, the mean frequency $\bar{x}_\omega$ and the frequency centroid $\omega_C$, which are defined as (LEI et al. 2010)

$$\omega_D = \underset{\text{w.r.t.}\omega}{\arg\max}\ |x(j\omega)|\,, \tag{2.33}$$

$$\bar{x}_\omega = \frac{1}{M}\sum_{i=1}^{M}|x(j\omega_i)| \tag{2.34}$$

and

$$\omega_C = \frac{\sum_{i=1}^{M}\omega_i|x(j\omega_i)|}{\sum_{i=1}^{M}|x(j\omega_i)|}, \tag{2.35}$$

where $M$ is the number of frequency bins and, therefore, $\omega_1$ and $\omega_M$ are the smallest and largest frequencies, respectively. An illustration of that is shown in Fig. 2.9. An example for a physically motivated and advanced frequency-domain feature extraction method was already given by the introduction of modal analysis in Subsection 2.1.3.

Finally, time-frequency-domain features are a combination of the features described above. They represent a raw time series in both, the time domain and

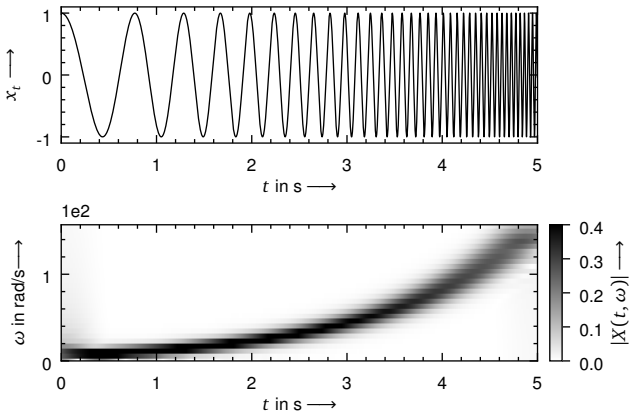



**Figure 2.10:** Illustration of exemplary time-frequency features; the upper plot shows a time series with an increasing frequency. The lower plot shows the spectrogram of the time series containing the squared absolute values of the STFT which are the time-frequency features. It can be seen that the change in frequency is emphasised by the spectrogram.

the frequency domain. One advantage of these features compared to the above-mentioned ones is that they can capture the non-stationary behaviour of a time series, such as changing frequency components over time.

**Example 2.3** (Time-Frequency-Domain Features)**.** A common method for extracting time-frequency-domain features is the short-time Fourier transform (STFT) (JARDINE et al. 2006). For a univariate time series $\mathbf{x}_t$ it can be computed with (OPPENHEIM et al. 1999, p. 717)

$$X(t,\omega) = \sum_{m=-\infty}^{\infty} x(m)w(m-t)e^{-j\omega m}, \tag{2.36}$$

which can be regarded as an FFT of the original discrete time series, weighted by the so-called *window function* $w(m)$. The STFT can be visualised in a spectrogram, as depicted in Fig. 2.10. It can be seen, that the STFT transforms the original one-dimensional (i.e. univariate) time series into a two-dimensional matrix of frequencies along time, allowing changes in frequencies to become visible.

In summary, the different feature extraction methods are applied according to the use case. For example, in the case of a stationary time series, simple time-domain features may already be sufficient to characterise the recorded data. When more complex, non-stationary signals with changing frequencies are considered, more elaborate tools, such as the STFT, can be applied. Generally,

feature extraction is widely used nowadays and modern software makes the extraction of many features from time series straightforward (CHRIST et al. 2018). However, manually designed features, like the ones introduced, also have some disadvantages: first, the user has to decide which features should be extracted. This decision requires domain knowledge, which may not be or which may only be partially present. Second, manually designed features are limited in terms of extracting all the available information. One remedy to both of these disadvantages is to *automatically* learn *optimal* features from the available data. This can be achieved using ANNs, which will be described in the next subsection.

### 2.3.2 Artificial Neural Networks

Artificial neural networks[2] are a large class of very diverse models that can be applied to supervised and unsupervised ML problems. In the former, historic inputs $\mathbf{x}$ and the corresponding outputs $y$ (sometimes also referred to as *labels* or *targets*) are available. In contrast to that, in the latter, only historic inputs $\mathbf{x}$ are present (MACKAY 2003, p. 470). This thesis focusses on the application of neural networks to supervised learning problems.

As stated by HASTIE et al. (2009, p. 392), "[t]here has been a great deal of hype surrounding neural networks, making them seem magical and mysterious. [...] [However,] they are just non-linear statistical models [...]". This means, that when neural networks are treated as statistical models, then the general reasoning comprises two steps (HASTIE et al. 2009, pp. 28 sqq.): first, the assumption is made, that a data-generating function $f_{\boldsymbol{\theta}}$ exists, that produces the observed outputs $y$ on the basis of the corresponding inputs $\mathbf{x}$ such that

$$y = f_{\boldsymbol{\theta}}(\mathbf{x}). \tag{2.37}$$

Second, if $i = 1, \ldots, N$ input and output pairs $(\mathbf{x}_i, y_i)$ of that function are observed and collected in a data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, then the true data-generating function $f_{\boldsymbol{\theta}}$ can be approximated with $f_{\hat{\boldsymbol{\theta}}}$ such that

$$y_i = f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}_i) + \epsilon_i, \tag{2.38}$$

where $f$ is the neural network, $\hat{\boldsymbol{\theta}}$ is the set of estimated network parameters (i.e. weights $w$ and biases $b$) and $\epsilon_i$ is random noise, which is independent of the

[2] The terms *artificial neural network* and *neural network* are used interchangeably in this thesis.

inputs $\mathbf{x}_i$ and which is defined as the difference between the observations $y_i$ and the predictions $\hat{y}_i$:

$$\epsilon_i = y_i - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}_i) = y_i - \hat{y}_i. \tag{2.39}$$

Neural networks can learn arbitrarily complex non-linear relationships from training data by using fairly simple algorithms (DUDA et al. 2001, p. 283). Thereby, a non-linear function is learned as a composition of several simpler functions (GOODFELLOW et al. 2016, p. 5). Below, two types of architectures of neural networks, a feed forward neural network and a convolutional neural network, will be introduced. Afterwards, methods for finding suitable parameters $\hat{\boldsymbol{\theta}}$, which minimise the prediction errors $\epsilon_i$, will be presented.

#### Feed Forward Neural Networks

The simplest possible neural network is the feed forward neural network (FFNN)[3] with one input layer of size $D$, one hidden layer of size $M$ and one output layer of size $K$. For the FFNN shown in Fig. 2.11a, an output value $y_k$ with $k = 1, \ldots, K$ is defined as (BISHOP 2006, p. 228)

$$y_k(\mathbf{x}, \boldsymbol{\theta}) = h^{(2)}\left(\sum_{j=1}^{M} w_{k,j}^{(2)} h^{(1)}\left(\sum_{i=1}^{D} w_{j,i}^{(1)} x_i + b_{j,0}^{(1)}\right) + b_{k,0}^{(2)}\right) \tag{2.40}$$

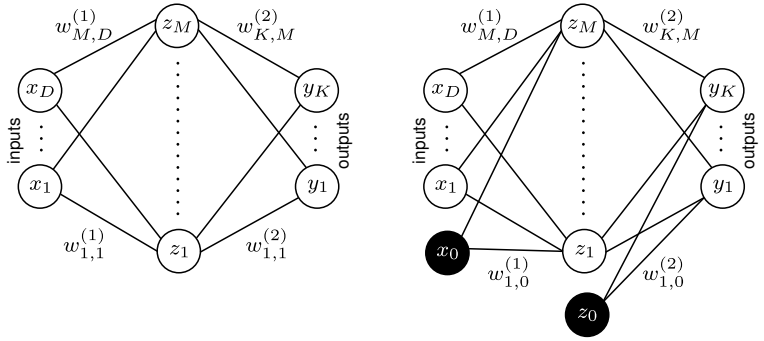
with the model parameters $\boldsymbol{\theta}$ being the weights $w_{j,i}^{(1)}$ and biases $b_{j,0}^{(1)}$ on layer one, and the weights $w_{k,j}^{(2)}$ and biases $b_{k,0}^{(2)}$ on layer two. The functions $h^{(1)}(x)$ and $h^{(2)}(x)$ are the so-called *activation functions* on layer one and two, respectively.

For a more compact notation, the biases can be absorbed by additional variables $x_0 = 1$ and $z_0 = 1$ (see Fig. 2.11b). This reduces Eq. (2.40) to (BISHOP 2006, p. 229)

$$y_k(\mathbf{x}, \boldsymbol{\theta}) = h^{(2)}\left(\sum_{j=0}^{M} w_{k,j}^{(2)} h^{(1)}\left(\sum_{i=0}^{D} w_{j,i}^{(1)} x_i\right)\right). \tag{2.41}$$

Notably, the number of output units $K$ can be set freely depending on the use case. In case of a classification problem with $K$ classes, it is common to define $K$ output units. However, in case of a regression problem with only one target variable, such as an RUL value, a single output unit is sufficient.

[3] Sometimes, a feed forward neural network is also called *multilayer perceptron* or *multilayer network*.




**(a)** FFNN with explicit biases; the ouput values can be calculated according to Eq. (2.40).

**(b)** FFNN with absorbed biases; the ouput values can be calculated according to Eq. (2.41).

**Figure 2.11:** Illustration of simple FFNNs with one hidden layer consisting of $M$ hidden units based on BISHOP (2006, p. 228); the neural network maps the input $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_D \end{bmatrix}^\top$ to the resulting output $\mathbf{y} = \begin{bmatrix} y_1 & \cdots & y_K \end{bmatrix}^\top$.

Dropping the biases entirely for the sake of a simpler notation allows to denote the model from Eq. (2.41) in a more general form as

$$\begin{aligned} \mathbf{y}(\mathbf{x}, \boldsymbol{\theta}) &= h^{(2)}\left(\begin{bmatrix} w^{(2)}_{1,1} & \cdots & w^{(2)}_{1,M} \\ \vdots & \ddots & \vdots \\ w^{(2)}_{K,1} & \cdots & w^{(2)}_{K,M} \end{bmatrix} h^{(1)}\left(\begin{bmatrix} w^{(1)}_{1,1} & \cdots & w^{(1)}_{1,D} \\ \vdots & \ddots & \vdots \\ w^{(1)}_{M,1} & \cdots & w^{(1)}_{M,D} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}\right)\right) \\ &= h^{(2)}\left(W^{(2)} h^{(1)}\left(W^{(1)}\mathbf{x}\right)\right) \\ &:= \left(\mathcal{H}^{(2)} \circ \mathcal{H}^{(1)}\right)(\mathbf{x}) \end{aligned} \tag{2.42}$$

with the weight matrices $W^{(l)}$ for each layer $l = 1, 2$. For more complex (i.e. deeper) FFNNs with $L$ layers, the general notation can be defined as follows:

$$\mathbf{y}(\mathbf{x}, \boldsymbol{\theta}) = \left(\mathcal{H}^{(L)} \circ \mathcal{H}^{(L-1)} \circ \ldots \circ \mathcal{H}^{(1)}\right)(\mathbf{x}). \tag{2.43}$$

One of the reasons for the success of neural networks in many engineering applications is their expressive power. Even the simple neural network depicted in Fig. 2.11 with only one hidden layer can represent *any* continuous function as long as sufficiently many units on the hidden layer are available (DUDA et al. 2001, p. 287).

Sometimes, however, the input data $\mathbf{x}$ of a neural network may slightly vary as a result of the data gathering process. For example, several time series recordings of a periodic signal may start at different points in time. Although the different recordings represent exactly the same signal, they look different from the perspective of a neural network. In these cases, the so-called *convolutional neural networks* can be advantageous, as they can extract local features from small subregions of the signal. This makes them less sensitive to the described minor changes in the input data (BISHOP 2006, p. 267). This type of neural network will be briefly described below.

#### Convolutional Neural Networks

Convolutional neural networks (CNNs) were originally introduced in a study by LECUN et al. (1989), whose goal was to classify handwritten postcode digits. Nowadays, CNNs are well known for their successful application in computer vision tasks. This is due to their ability to detect local features, such as edges and objects, in grid-like data. Images are an example of two-dimensional grid-like data, whereas time series are an example of one-dimensional grid-like data (GOODFELLOW et al. 2016, p. 321). In this thesis, CNNs are applied to time series data only. Therefore, they will be further described in that context.

CNNs are a type of FFNNs but with heavily constrained hidden layers. In contrast to the neural networks depicted in Fig. 2.11, in which all units of the hidden layer are connected with all units of the previous layer, convolutional layers are restricted to combine local inputs only.

This is achieved by a convolutional *kernel*, which slides over the input vector and conducts simple mathematical operations. An example of this is shown in Fig. 2.12, in which a kernel $\mathbf{k}$ slides over a time series $\mathbf{x}_t$, resulting in a feature vector $\mathbf{z}$.

According to LECUN et al. (2015), a typical CNN consists of several convolutional layers followed by so-called pooling layers (see Appendix A.1) and activation functions (see below). This subsequent application of many of these layers allows the neural network to learn from data that is composed by lower-level features. For example, a photograph can be regarded as a composition of edges and objects and a time series can be regarded as a composition of different periodic signals and noise.

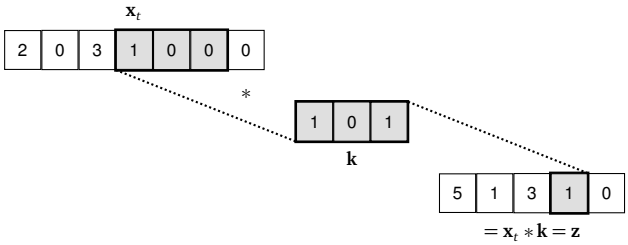



**Figure 2.12:** Illustration of an exemplary convolution of a one-dimensional time series $\mathbf{x}_t$ with a convolutional kernel $\mathbf{k} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$ and the resulting feature vector $\mathbf{z} = \mathbf{x}_t * \mathbf{k} = \begin{bmatrix} z_1 & \cdots & z_M \end{bmatrix}$; the depicted convolutional operation yields the result $z_4 = (1 \cdot 1) + (0 \cdot 0) + (1 \cdot 0) = 1$.

#### Activation Functions for Neural Networks

The non-linearity of ANNs is induced by the units' activation functions $h^{(l)}(x)$ on each layer $l$ (see Eqs. (2.40) to (2.42)). Figure 2.13 shows the ones used in this thesis, which are also formally defined below. The most common, and from a best practice perspective recommended, activation function is the so-called rectified linear unit (ReLU) function, which is non-linear only at the location $x = 0$ (GOODFELLOW et al. 2016, pp. 168 sq.):

$$h(x) = \max(0, x). \tag{2.44}$$

A slightly modified version of this activation function is the leaky ReLU function, which instead of being constant for $x \leq 0$ has a negative slope (MAAS et al. 2013):

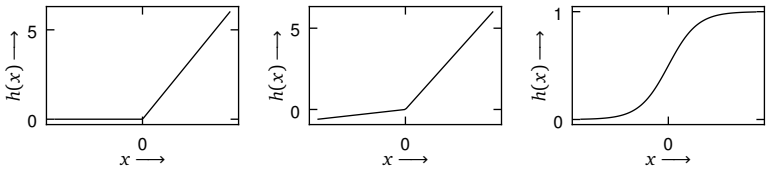
$$h(x) = \begin{cases} x & \text{if } x > 0 \\ c_{sl}\, x & \text{if } x \leq 0. \end{cases} \tag{2.45}$$

The slope $c_{sl}$ is usually set to a small value, such as $c_{sl} = 0.01$.

Sometimes, bounding the output of a unit can be helpful. This can be achieved with the *sigmoid* activation function, which is defined as (BISHOP 2006, p. 228)

$$h(x) = \frac{1}{1 + \exp(-x)}, \tag{2.46}$$

and which bounds the outputs $h(x)$ to the range from zero to one. This sigmoid function is especially helpful when the task is a binary classification problem with the two classes zero and one. In that case, it can be applied on the output layer. If there are more than just two classes but $K$ classes are present, then the




**(a)** ReLU **(b)** leaky ReLU **(c)** sigmoid

**Figure 2.13:** Three exemplary activation functions for ANNs

sigmoid function can be extended to the so-called *softmax* activation function (BISHOP 2006, p. 198)

$$h(x)_k = \frac{\exp(x_k)}{\sum_{j=1}^{K} \exp(x_j)}, \tag{2.47}$$

where $k = 1, \ldots, K$. Hence, for each class, a value between zero and one is computed. This value can be interpreted as the probability of class $k$ membership.

#### Training Neural Networks with Backpropagation

Equation (2.42) shows that even simple ANNs have many unknown parameters (i.e. weights and biases that can be summarised in a parameter vector $\boldsymbol{\theta}$). These parameters are usually found with gradient-based algorithms, which optimise a loss function $J(\boldsymbol{\theta})$ for an available training data set $\mathcal{D}$. This loss function differs depending on the use case. For regression tasks, it is common to minimise the sum-of-squares error function (BISHOP 2006, pp. 232 sq.)

$$J(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{N} (f_{\boldsymbol{\theta}}(\mathbf{x}_i) - y_i)^2, \tag{2.48}$$

whereas in classification tasks with $K$ classes it is common to minimise the cross entropy loss, which is defined as (BISHOP 2006, pp. 235 sq.)

$$J(\boldsymbol{\theta}) = -\sum_{i=1}^{N} \sum_{k=1}^{K} \left( y_{i,k} \log(f_{\boldsymbol{\theta}}(\mathbf{x}_i)_k) + (1 - y_{i,k}) \log(1 - f_{\boldsymbol{\theta}}(\mathbf{x}_i)_k) \right). \tag{2.49}$$

In order to estimate the parameters $\hat{\boldsymbol{\theta}}$ and solve

$$\hat{\boldsymbol{\theta}} = \underset{\text{w.r.t.}\boldsymbol{\theta}}{\arg\min} J(\boldsymbol{\theta}), \tag{2.50}$$

gradient-based algorithms can be used to iteratively compute

$$\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old} - \alpha \nabla J(\boldsymbol{\theta}^{old}) \tag{2.51}$$

until convergence. In Eq. (2.51), $\alpha$ is a hyper-parameter called the *learning rate* and $\nabla J(\boldsymbol{\theta}^{old})$ is the partial derivative of the cost function $J$ with respect to the parameters $\boldsymbol{\theta}^{old}$. The gradients $\nabla J(\boldsymbol{\theta})$ can be found using the so-called *backpropagation* algorithm (RUMELHART et al. 1986). Nowadays, they no longer need to be calculated manually, as software can now automatically calculate them (PASZKE et al. 2017). For an in-depth treatment of the training of neural networks, the reader is kindly referred to GOODFELLOW et al. (2016).

Generally, training neural networks with backpropagation has been well studied and has been found to be reliable in practical applications. However, the result obtained is only a single set of parameters, or in other words, a *point estimate* for the parameter vector $\boldsymbol{\theta}$. Therefore, the parameters found may not exactly be the true parameters of the data-generating function from Eq. (2.37). It remains unclear how *certain* the set of parameters is in approximating this true function. This *uncertainty* can be quantified by conducting a Bayesian inference instead of training with backpropagation. The general idea behind this will be briefly described below.

#### Bayesian Inference for Neural Networks

In a Bayesian approach, the unknown parameters $\boldsymbol{\theta}$ are assumed to be a random vector that follows the so-called posterior distribution $p(\boldsymbol{\theta} \mid \mathcal{D}, f)$, which can be computed using the Bayes' rule (LAMPINEN and VEHTARI 2001)

$$\underbrace{p(\boldsymbol{\theta} \mid \mathcal{D}, f)}_{\text{posterior}} = \frac{\overbrace{p(\mathcal{D} \mid \boldsymbol{\theta}, f)}^{\text{likelihood}} \overbrace{p(\boldsymbol{\theta} \mid f)}^{\text{prior}}}{\underbrace{p(\mathcal{D} \mid f)}_{\text{marginal likelihood}}}, \tag{2.52}$$

where the marginal likelihood is defined as (LAMPINEN and VEHTARI 2001)

$$p(\mathcal{D} \mid f) = \int_{\boldsymbol{\theta}} p(\mathcal{D} \mid \boldsymbol{\theta}, f)\, p(\boldsymbol{\theta} \mid f)\, d\boldsymbol{\theta}. \tag{2.53}$$

When the posterior distribution of the parameters is known, one can compute the posterior predictive distribution of a new input $\mathbf{x}^\star$ with (LAMPINEN and VEHTARI 2001)

$$p(y^\star \mid \mathbf{x}^\star, \mathcal{D}, f) = \int_{\boldsymbol{\theta}} p(y^\star \mid \mathbf{x}^\star, \boldsymbol{\theta}, f)\, p(\boldsymbol{\theta} \mid \mathcal{D}, f)\, d\boldsymbol{\theta}. \tag{2.54}$$

As $f$ is defined as an ANN (see Eqs. (2.37) and (2.38)), the posterior predictive distribution becomes (NEAL 1996, pp. 13 sq.)

$$p(y^\star \mid \mathbf{x}^\star, \mathcal{D}, f) = \int_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(\mathbf{x}^\star) p(\boldsymbol{\theta} \mid \mathcal{D}, f)\, d\boldsymbol{\theta}. \tag{2.55}$$

By sampling parameters $\boldsymbol{\theta}^{(m)}$ from $p(\boldsymbol{\theta} \mid \mathcal{D}, f)$, one can generate model predictions $f_{\boldsymbol{\theta}^{(m)}}(\mathbf{x}^\star)$ and compute statistical moments, such as the mean prediction which is given by (NEAL 1996, p. 23)

$$\mathbb{E}[y^\star] \approx \frac{1}{M} \sum_{m=1}^{M} f_{\boldsymbol{\theta}^{(m)}}(\mathbf{x}^\star). \tag{2.56}$$

In models with many parameters, such as ANNs, computing the marginal likelihood from Eq. (2.53) and ultimately the posterior distribution from Eq. (2.52) is rarely tractable and often cannot be done analytically (LAMPINEN and VEHTARI 2001). Therefore, several methods have been developed to approximate the posterior distribution. Two of these methods are applied in this thesis:

**Markov chain Monte Carlo** Markov chain Monte Carlo (MCMC) algorithms are a class of algorithms whose goal is to construct a Markov chain whose stationary distribution is equal to the posterior distribution from Eq. (2.52). When such a Markov chain is found, one can sample from it and calculate quantities of interest from the posterior, such as its sample mean and sample variance. A more technical explanation can be found in Appendix A.2.

**Variational inference** In contrast to MCMC, variational inference (VI) algorithms do not attempt to sample from the posterior distribution $p$ directly. Instead, they approximate it with a simpler and known variational distribution $q$. In this scenario, the parameters of $q$ are set such that the Kullback-Leibler (KL) divergence between the posterior and the variational distribution is minimised. Appendix A.3 provides more technical details on that.

In summary, the Bayesian treatment described above generates probabilistic predictions $p(y^\star \mid \mathbf{x}^\star, \mathcal{D}, f)$ based on a highly flexible and parametric function

(i.e. a neural network) $f_{\boldsymbol{\theta}}$, whose parameters $\boldsymbol{\theta}$ are treated as a random vector. A different Bayesian approach for generating probabilistic predictions is using GPs. In contrast to treating the unknown parameter vector $\boldsymbol{\theta}$ as random, the function $f$, which represents the data generating process from Eq. (2.37), is directly treated as random (i.e. as a *random function*). The details are explained in the next subsection.

### 2.3.3 Gaussian Processes

According to RASMUSSEN and WILLIAMS (2006, p. 13), "[a GP] is a collection of random variables, any finite number of which have a joint Gaussian distribution". A GP is fully described by its mean function $m(\mathbf{x})$ and its covariance function $k(\mathbf{x}, \mathbf{x}')$ (sometimes also referred to as the *kernel*), which are defined as follows:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{2.57}$$

and

$$k\left(\mathbf{x}, \mathbf{x}'\right) = \mathbb{E}\left[\left(f(\mathbf{x}) - m(\mathbf{x})\right)\left(f(\mathbf{x}') - m(\mathbf{x}')\right)\right]. \tag{2.58}$$

A function following a GP can be denoted as

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right). \tag{2.59}$$

This definition of a GP is indeed complete from a theoretical perspective but can be cumbersome for the purpose of building an intuition about GPs. Therefore, the simpler multivariate Gaussian distribution[4], which is a special case of a GP, will be introduced below.

A $D$-dimensional random vector $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_D \end{bmatrix}^\top$ follows a multivariate Gaussian distribution with the mean vector

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \\ \vdots \\ \mu_{x_D} \end{bmatrix} \tag{2.60}$$

[4]The terms *Gaussian* distribution and *normal* distribution are used interchangeably in this thesis.

and the covariance matrix

$$\Sigma = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) & \cdots & \mathrm{Cov}(x_1, x_D) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) & \cdots & \mathrm{Cov}(x_2, x_D) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(x_D, x_1) & \mathrm{Cov}(x_D, x_2) & \cdots & \mathrm{Var}(x_D) \end{bmatrix}, \tag{2.61}$$

when the probability density function (PDF) of the joint distribution is given by (MARDIA et al. 1979, p. 37):

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{2.62}$$

If $\mathbf{x}$ is normal, then this is denoted as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ or as $p(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma) = \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

The multivariate normal distribution has useful properties, which will be exploited later in this thesis, when the Gaussian process regression model will be introduced in Subsection 2.3.4.

**Partitioned notation** A normal random vector $\mathbf{x}$ can be partitioned into $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}^\top$ with

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1} \\ \boldsymbol{\mu}_{\mathbf{x}_2} \end{bmatrix} \tag{2.63}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{x}_1,\mathbf{x}_1} & \Sigma_{\mathbf{x}_1,\mathbf{x}_2} \\ \Sigma_{\mathbf{x}_2,\mathbf{x}_1} & \Sigma_{\mathbf{x}_2,\mathbf{x}_2} \end{bmatrix}. \tag{2.64}$$

This notation is a premise for the following properties.

**Marginal distribution** Based on the joint distribution, the marginal distribution of $\mathbf{x}_1$ is $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{x}_1}, \Sigma_{\mathbf{x}_1,\mathbf{x}_1}\right)$ and the marginal distribution of $\mathbf{x}_2$ is $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{x}_2}, \Sigma_{\mathbf{x}_2,\mathbf{x}_2}\right)$. As a special case, the marginal distribution $p(x_i \mid \mu_{x_i} \sigma_{x_i}^2)$ of a single variable $x_i$ (with $i = 1, \ldots, D$) is given by the univariate Gaussian distribution (RENCHER 2002, p. 87)

$$p(x_i \mid \mu_{x_i}, \sigma_{x_i}^2) = \frac{1}{\sqrt{2\pi\sigma_{x_i}^2}} \exp\left(-\frac{(x_i - \mu_{x_i})^2}{2\sigma_{x_i}^2}\right), \tag{2.65}$$

with $\sigma_{x_i}^2 = \mathrm{Var}(x_i)$.

**Conditional distribution** Given $\mathbf{x}_1$, the conditional distribution $p(\mathbf{x}_2 \mid \mathbf{x}_1)$ of the unobserved variables $\mathbf{x}_2$ is again Gaussian (RENCHER 2002, p. 88)

$$p(\mathbf{x}_2 \mid \mathbf{x}_1) = \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{x}_2|\mathbf{x}_1}, \Sigma_{\mathbf{x}_2|\mathbf{x}_1}\right) \tag{2.66}$$

with

$$\boldsymbol{\mu}_{\mathbf{x}_2|\mathbf{x}_1} = \boldsymbol{\mu}_{\mathbf{x}_2} + \Sigma_{\mathbf{x}_2,\mathbf{x}_1}\Sigma^{-1}_{\mathbf{x}_1,\mathbf{x}_1}\left(\mathbf{x_1} - \boldsymbol{\mu}_{\mathbf{x}_1}\right) \tag{2.67}$$

and

$$\Sigma_{\mathbf{x}_2|\mathbf{x}_1} = \Sigma_{\mathbf{x}_2,\mathbf{x}_2} - \Sigma_{\mathbf{x}_2,\mathbf{x}_1}\Sigma^{-1}_{\mathbf{x}_1,\mathbf{x}_1}\Sigma_{\mathbf{x}_1,\mathbf{x}_2}. \tag{2.68}$$

In the following example, the properties mentioned are illustrated for a special case with $D = 2$, which is the bivariate Gaussian distribution.

**Example 2.4** (Bivariate Gaussian Distribution)**.** When $\mathbf{x}$ is two-dimensional, namely $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$, the mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ become

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix} \tag{2.69}$$

and

$$\Sigma = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) \end{bmatrix}. \tag{2.70}$$

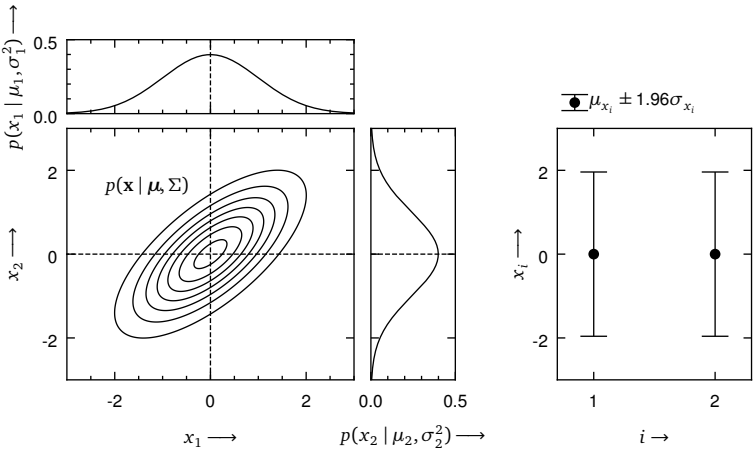
An example of a bivariate Gaussian distribution with the parameters

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{x_1} \\ \mu_{x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{2.71}$$

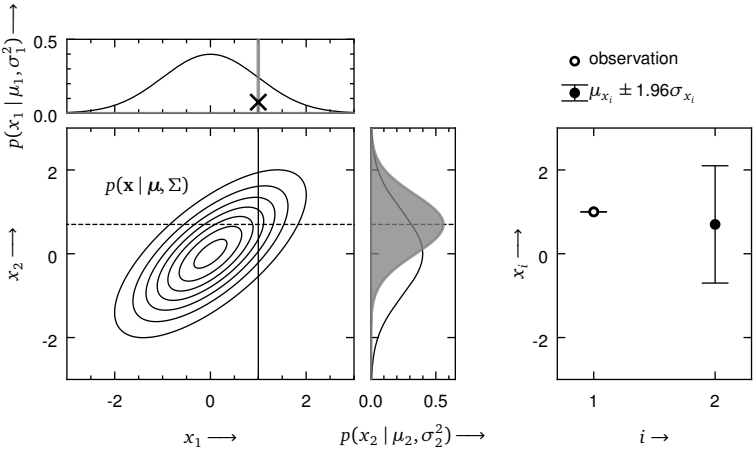
and

$$\Sigma = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) \end{bmatrix} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} \tag{2.72}$$

is depicted in Fig. 2.14a, in which the joint distribution and the marginal distributions for the two random variables $x_1$ and $x_2$ are shown. Before observing any data, the expected value of each variable is $\mathbb{E}(x_i) = \mu_{x_i} = 0$ and the variance of each variable is $\mathrm{Var}(x_i) = 1$. As soon as $x_1$ is observed, the conditional distribution can be computed using Eq. (2.66). This is depicted in Fig. 2.14b for the case where $x_1 = 1$, which leads to the conditional distribution $p(x_2 \mid x_1 = 1) = \mathcal{N}(0.7, 0.51)$. Notably, the new distribution of $x_2$ has not only a new mean value (i.e. 0.7 instead of 0), but it also has a smaller variance (0.51




**(a)** Bivariate normal distribution before observing data




**(b)** Bivariate normal distribution after observing $x_1 = 1$

**Figure 2.14:** Illustration of a bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ as defined in Eqs. (2.71) and (2.72); on the left side of panel (a), the joint distribution as well as the marginal distributions of the random variables $x_1$ and $x_2$ before observing any data are shown. On the right side of panel (a), the two random variables' mean value $\mu_i$ and 1.96 times their standard deviations $\sigma_i$, which approximately accounts for 95 % of the area under the normal distribution, are plotted against their index $i$. In panel (b), the same distribution is shown after observing $x_1 = 1$. The marginal distribution of $x_2$ is now a conditioned distribution depicted by the grey shaded area. The conditional distribution has a different mean and standard deviation. On the right side of panel (b), it can be seen that observing $x_1$ reduces the variance (i.e. the uncertainty) of $x_2$. This is due to the covariance between the two defined in $\Sigma$ in Eq. (2.72).

instead of 1). Hence, because a joint normal distribution is assumed and $x_1$ is observed, one is more certain about the location of $x_2$.

The fact, that the conditional distribution is simply another Gaussian distribution, is particularly useful for regression settings, which will be demonstrated with a higher-dimensional multivariate Gaussian distribution in the following example.

**Example 2.5** (Multivariate Gaussian Distribution)**.** Figure 2.15a shows an example of an eight-dimensional multivariate Gaussian distribution with the following parameters:

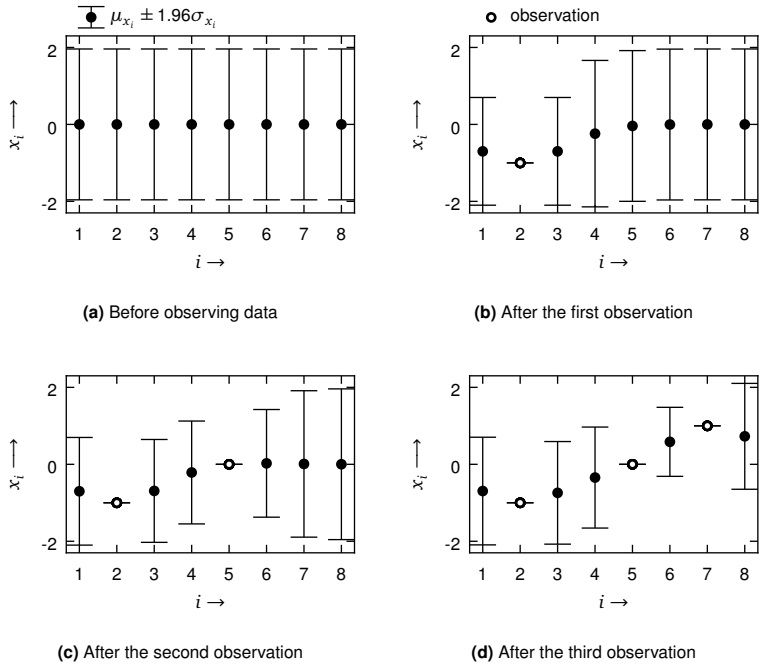
$$\boldsymbol{\mu} = \mathbf{0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^\top \tag{2.73}$$

and

$$\Sigma = \begin{bmatrix} 1 & 0.7 & 0.24 & 0.04 & 0 & 0 & 0 & 0 \\ 0.7 & 1 & 0.7 & 0.24 & 0.04 & 0 & 0 & 0 \\ 0.24 & 0.7 & 1 & 0.7 & 0.24 & 0.04 & 0 & 0 \\ 0.04 & 0.24 & 0.7 & 1 & 0.7 & 0.24 & 0.04 & 0 \\ 0 & 0.04 & 0.24 & 0.7 & 1 & 0.7 & 0.24 & 0.04 \\ 0 & 0 & 0.04 & 0.24 & 0.7 & 1 & 0.7 & 0.24 \\ 0 & 0 & 0 & 0.04 & 0.24 & 0.7 & 1 & 0.7 \\ 0 & 0 & 0 & 0 & 0.04 & 0.24 & 0.7 & 1 \end{bmatrix}. \tag{2.74}$$

Given the value $x_2 = -1$ in Fig. 2.15b, a new conditional distribution can be calculated for the remaining unobserved variables. The same effect as in the bivariate case in Fig. 2.14b can be observed: the neighbouring variables, namely $x_1$ and $x_3$, have a new mean value closer to the observation $x_2$ and a new, smaller variance. However, the values of the variables further away from $x_2$ do not change their mean values or variances as much. This is due to the covariance matrix $\Sigma$ defined in Eq. (2.74), in which variables with larger distances are defined to have a smaller covariance. In fact, the variables $x_6, x_7$ and $x_8$ are not affected by the observation of $x_2$ at all. This is because their covariance with $x_2$ is defined to be zero.

When the next value $x_5 = 0$ in Fig. 2.15c is observed, this effect is again observed for the neighbouring variables $x_4$ and $x_6$. Then, after the third observation $x_7 = 1$ in Fig. 2.15d, the conditional distribution of the remaining unobserved values shows even less variance. In addition, the observed values in combination with the new conditional means of the unobserved variables adopt the shape of a smooth, non-linear curve (i.e. function).




**(a)** Before observing data

**(b)** After the first observation

**(c)** After the second observation

**(d)** After the third observation

**Figure 2.15:** Illustration of a joint multivariate Gaussian distribution for an eight-dimensional random vector $\mathbf{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ as defined in Eqs. (2.73) and (2.74); in panel (a), the marginal distributions of all ten variables $x_i$ are shown before observing any data. In panels (b) – (d), the conditional distributions are shown after observing $x_2$, $x_5$ and $x_7$, respectively.

The example above shows that assuming a joint normal distribution on a vector $\mathbf{x}$ can be used to probabilistically model the values of the unobserved variables $x_i$. In this example, only the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\Sigma$ were required to be defined a priori. The unobserved variables were simply predicted by computing their conditional distributions.

The example also revealed some limitations: for instance, instead of function values $f(\mathbf{x})$, single values $x_i$ of $\mathbf{x}$ along an evenly spaced index $i$ were modelled. However, in regression problems, the goal is to model the unknown function $f$ directly. Furthermore, the input values of a function may not always be evenly spaced. A time series, for example, can have irregularly spaced values as a result of missing values or a special data acquisition routine.

The case outlined above is one in which the concept of a GP shows its strength: it allows to model the unknown function $f(\mathbf{x})$ directly as a Gaussian *random function*. In other words, a GP is a distribution over functions (as opposed to a distribution over a vector). This can be perceived as a generalisation of the multivariate normal distribution (RASMUSSEN and WILLIAMS 2006, p. 2). In the following Subsection 2.3.4, it will be shown how this can be exploited for regression and classification problems.

### 2.3.4 Gaussian Process Regression and Classification

In the following, the application of GPs in regression and classification problems will be described.

#### Gaussian Process Regression

Given a data set $\mathcal{D}$ of input and output pairs, the Gaussian process regression (GPR) model implements a function $f$ that maps the inputs $\mathbf{x}_i$ to the outputs $y_i$, such that

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right), \tag{2.75}$$

where $\epsilon_i$ represents random Gaussian noise and $f$ follows a GP. Hence, $f$ can be denoted as $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$. The mean function is often set to zero, that is $m(\mathbf{x}) = \mathbf{0}$, which is also adopted in this thesis. (RASMUSSEN and WILLIAMS 2006, p. 13)

As in the case of the multivariate Gaussian distribution in Eqs. (2.63) and (2.64), the joint distribution of the observed function values $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}^\top$ at input locations $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix}$ and unobserved function values $\mathbf{f}^\star = \begin{bmatrix} f_1^\star & f_2^\star & \cdots & f_M^\star \end{bmatrix}^\top$ at input locations $X^\star = \begin{bmatrix} \mathbf{x}_1^\star & \mathbf{x}_2^\star & \cdots & \mathbf{x}_M^\star \end{bmatrix}$ can be denoted as (RASMUSSEN and WILLIAMS 2006, p. 16)

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^\star \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_\epsilon^2 I & K(X,X^\star) \\ K(X^\star,X) & K(X^\star,X^\star) \end{bmatrix}\right), \tag{2.76}$$

where $I$ is the identity matrix and $K(\cdot,\cdot)$ is the covariance matrix[5]. For example, $K(X,X^\star)$ denotes the covariances between the observed values $X$ and the

[5] In order to be consistent with the relevant literature, $K(\cdot,\cdot)$ is used to denote the covariance matrix in the context of a Gaussian process instead of the $\Sigma$ notation which was used in the context of the multivariate Gaussian distribution.

unobserved values $X^\star$. The conditional distribution, which in this case is also the predictive distribution of the unknown values $\mathbf{f}^\star$, can be calculated with (RASMUSSEN and WILLIAMS 2006, p. 16)

$$p(\mathbf{f}^\star \mid X, \mathbf{y}, X^\star) = \mathcal{N}(\mathbf{f}^\star \mid \mathbb{E}(\mathbf{f}^\star \mid X, \mathbf{y}, X^\star), \mathrm{Var}(\mathbf{f}^\star \mid X, \mathbf{y}, X^\star)), \tag{2.77}$$

where

$$\mathbb{E}(\mathbf{f}^\star \mid X, \mathbf{y}, X^\star) = K(X^\star, X)\left(K(X,X) + \sigma_\epsilon^2 I\right)^{-1}\mathbf{y} \tag{2.78}$$

and

$$\mathrm{Var}(\mathbf{f}^\star \mid X, \mathbf{y}, X^\star) = K(X^\star, X^\star) - K(X^\star, X)\left(K(X,X) + \sigma_\epsilon^2 I\right)^{-1} K(X, X^\star). \tag{2.79}$$

Notably, the optimal prediction for any GP is given by the mean of the predictive distribution provided in Eq. (2.78) (HAMILTON 1994, p. 100). The following example illustrates the expressive power of GPs in a regression setting.

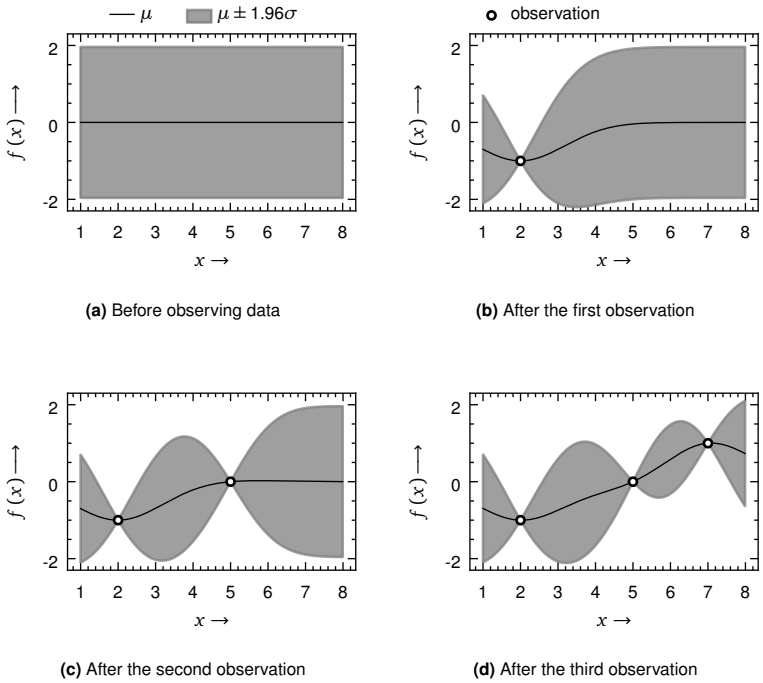
**Example 2.6** (Gaussian Process Regression)**.** Figure 2.16 shows the graph of a function of one-dimensional inputs $f(x)$ from a GP with a mean function $m(x) = 0$ and the following radial basis function (RBF) kernel:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right), \quad \sigma_f = 1, \quad \ell = 1.18. \tag{2.80}$$

The prior distribution is shown in Fig. 2.16a, and the conditional distributions after observing data are shown in Figs. 2.16b to 2.16d. With an increasing number of observations, the conditional distribution becomes a smooth, non-linear function. In fact, the covariance function from Eq. (2.80) was used to generate the covariance matrix defined in Eq. (2.74). Therefore, the conditional distributions of $f(x)$ at the input locations $x = 1, 2, \ldots, 8$ are equal to the conditional distributions of $x_i$ with the indices $i = 1, 2, \ldots, 8$ as shown in Fig. 2.15d.

This case highlights the advantage of the GPR model: the conditional distribution of the unknown function $f(x)$ can be evaluated *anywhere*: at locations lower than one, larger than eight and everywhere between the evenly spaced values of $x_i$. Furthermore, the GPR model produces flexible functions and sensible uncertainty bounds, which are smaller when close to observations and larger in areas farther away from observations.

The example outlined above highlights the ability of the GPR model to probabilistically model non-linear functions. The key ingredient to this is the covariance




**(a)** Before observing data

**(b)** After the first observation

**(c)** After the second observation

**(d)** After the third observation

**Figure 2.16:** Illustration of a Gaussian process; the black line represents the mean value of the function $f(x)$ and the grey shaded area represents 1.96 standard deviations from the mean, which approximately accounts for 95 % of the area under the normal distribution. In panel (a), the prior distribution is shown before observing data. In panels (b) – (d), the conditional (i.e. posterior) distribution is shown after observing $f(x = 2) = -1$, $f(x = 5) = 0$ and $f(x = 7) = 1$, respectively.

function, which defines the covariance between neighbouring function values $f(\mathbf{x})$ at input locations $\mathbf{x}$. Many valid covariance functions exist, which in turn can be recombined to form even more valid ones. Depending on the selected covariance function, the characteristics of the functions represented by the GPR model significantly change (DUVENAUD 2014, pp. 10 sqq.). Hence, the choice of the covariance function is crucial and considered as a way to incorporate expert knowledge into the GPR model.

Another noteworthy aspect of the GPR model is that the need for identifying the unknown parameters $\boldsymbol{\theta}$ of a parametric function (as it was necessary for ANNs shown in Subsection 2.3.2) is diminished by applying the rules of probability

and directly calculating a conditional Gaussian distribution for the unknown function values. The only two parameters that needed to be set in Example 2.6 were $\sigma$ and $\ell$ from the covariance function defined in Eq. (2.80). However, as the two are not parameters of the modelled function, they are referred to as hyper-parameters. Depending on their value, the GPR model may look completely different (see Fig. 2.17 for three different values of $\ell$). Fortunately, an efficient way to determine the values of these hyper-parameters exists, which will be briefly described below.

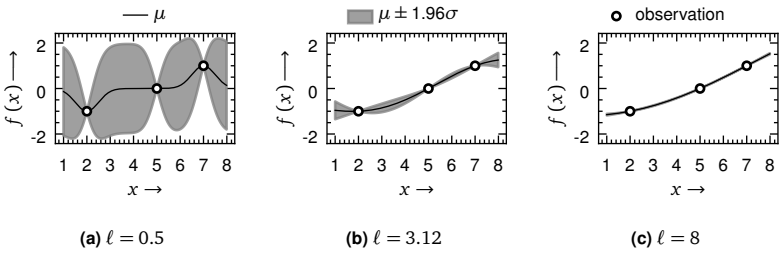
#### Model Selection for Gaussian Process Regression

In order to find a favourable set of hyper-parameters (i.e. a favourable GPR model), conducting the so-called type II maximum likelihood (ML-II) approximation is considered common practice. The goal of this approximation is to find the hyper-parameters that maximise the marginal likelihood, which in the case of the GPR model is defined as (RASMUSSEN and WILLIAMS 2006, pp. 109 sqq.)

$$\log\left(p(\mathbf{y} \mid X, \boldsymbol{\theta})\right) = -\frac{1}{2}\mathbf{y}^\top K_y^{-1}\mathbf{y} - \frac{1}{2}\log\left(\det\left(K_y\right)\right) - \frac{N}{2}\log(2\pi) \tag{2.81}$$

with $K_y = K(X,X) + \sigma_\epsilon^2 I$ and $N$ being the number of observations in the training data set $\mathcal{D}$. In this context, the marginal likelihood can be optimised by computing the gradient with respect to the hyper-parameters and applying gradient-based optimisation algorithms.

The optimal solution for $\ell$ in the GPR examples shown in Fig. 2.16 is given by $\ell = 3.12$. The resulting optimal GPR model is shown in Fig. 2.17b.




**(a)** $\ell = 0.5$ **(b)** $\ell = 3.12$ **(c)** $\ell = 8$

**Figure 2.17:** Illustration of three GPR models with different hyper-parameters $\ell$; the model in panel (b) is maximizing the marginal likelihood and can be regarded as the most suitable model for the three given observations.

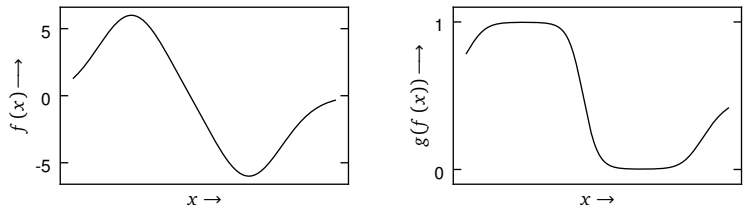
Although the explanations outlined earlier were concerned with the use of GPs in regression tasks, GPs can also be used in (binary) classification tasks, which will be discussed below.

#### Binary Gaussian Process Classification

In a binary classification task with two classes (i.e. class zero and class one), a GP can be used to model a latent function $f$, which governs the class membership. As shown in Fig. 2.18a, a GP is not bound on the interval $[0,1]$. This is problematic, as the value for the class membership of an observation is either zero or one. Hence, in order to map $f$ onto that interval, a so-called squashing function $g(\cdot)$ is applied so that the GP outputs values on the interval $[0,1]$, which can be read as the probability of an observation belonging to class one. This probability $p^\star$ can be calculated with (RASMUSSEN and WILLIAMS 2006, p. 40)

$$p^\star = p(y^\star = 1 \mid X, \mathbf{y}, X^\star) = \int g(f^\star)\, p(f^\star \mid X, \mathbf{y}, X^\star)\, df^\star. \tag{2.82}$$

Among the examples of the squashing function $g(\cdot)$ are the logistic and the probit functions (RASMUSSEN and WILLIAMS 2006, p. 40). In Fig. 2.18, an example of the application of the logistic function is depicted.




**(a)** Gaussian process sample $f$ **(b)** Squashed Gaussian process sample $g(f)$

**Figure 2.18:** Illustration of the effect of applying the logistic function $g(\cdot)$ to a sample $f$ from a Gaussian process

As both are non-linear functions applied to the Gaussian function $f$, the predictive distribution in Eq. (2.82) is not Gaussian anymore and usually cannot be computed analytically. Hence, approximations such as MCMC and VI need to be applied (RASMUSSEN and WILLIAMS 2006, p. 41). In particular, one variant

of VI, the *expectation propagation* algorithm introduced by MINKA (2001), has been found to be an accurate approximation of the predictive distribution for a Gaussian process classification (GPC) model (KUSS and RASMUSSEN 2005). More details on MCMC and VI can be found in the Appendices A.2 and A.3.

Chapter 3

# State of the Art

This chapter will summarise the relevant state of the art in two fields: first, the state of the art with respect to condition monitoring of ball screw feed drives will be presented in Section 3.1. Second, the state of the art for remaining useful life (RUL) estimation methods will be given in Section 3.2, where the focus will be more on the proposed methods rather than on the use cases to which they were applied. It will be shown that in both fields, a need for action can be identified, which will finally be summarised in Section 3.3.

## 3.1 Condition Monitoring of Ball Screw Feed Drives

It is notable that, although condition monitoring of machine tools with respect to the tool has been extensively reviewed in the literature (MOHANRAJ et al. 2020; REHORN et al. 2005; SERIN et al. 2020), review papers concerned with condition monitoring of machine tool feed drives, however, are rare. This was also emphasised in a review paper about prognostics and health management (PHM) for machine tools by BAUR et al. (2020), who found that, whilst considered an important topic, publications about condition monitoring of machine tool feed drives are scarce. In the following subsections, the relevant ones for this thesis will be summarised.

### 3.1.1 Ball Screws

An early approach for automated condition monitoring of ball screws was presented by SCHOPP (2009), who developed a single-axis ball screw feed drive test

bench. An artificial load was applied by bracing the ball screw nuts against each other with tensile rods. With regular execution of a defined test cycle during lifetime tests, Schopp observed a decrease in natural frequencies, which is a sign for preload loss. With ongoing service life, however, the identified natural frequencies also showed greater variance, resulting in a sequence of increasingly fluctuating natural frequencies, which ultimately exacerbated the assessment of the condition of the ball screws. This was due to the fact that the long-term trials also provoked pitting damage on the surface of the balls. This resulted in non-reproducible arrangements and jamming of the balls within the ball screw nut, finally leading to strongly varying measurement results of the natural frequencies.

Nevertheless, Schopp was able to observe a clear increase in the amplitude of vibrations measured between the opposing ball screw nuts with ongoing pitting damage. This was later validated on another test bench on which five ball screws could be degraded at once. There, an increase in sound amplitudes was also detectable. However, the progression of the increase differed significantly.

Finally, Schopp applied his approach to an industrial use case and monitored an axis of a machine tool that was used in the automotive industry. By reassembling a new and a worn-out ball screw from the above mentioned test bench, he showed that the increase in sound amplitudes was also observable in an industrially used machine tool. Based on those findings, a simple model for the diagnosis and prognosis of the RUL was designed. Although Schopp pioneered the condition monitoring of machine tool feed drives at that time by demonstrating that wear can be observed in the recorded signals, shortcomings include the need for costly additional sensors, as well as a missing validation of the proposed model for the diagnosis and prognosis of the RUL.

Verl et al. (2009) also designed a single-axis ball screw feed drive test bench on which they conducted run-to-failure experiments. While running a defined test cycle, which consisted of a constant-speed movement along the axis, they collected velocity signals from the rotary and linear encoder of the feed drive's computer numerical control (CNC) as well as signals from a laser-interferometer. Their experiments showed that several frequency-domain and time-domain features such as the vibration energy and the reversal positioning error correlated with wear.

Walther (2011) extended the work by Verl et al. (2009) by showing that the different wear mechanisms mentioned in Subsection 2.1.2 can lead to different

patterns in the vibration energy. Walther first periodically recorded the motor currents as well as the velocity and position signals from the rotary and linear encoder provided by the CNC. He subsequently constructed the vibration energy feature from the raw time series data and monitored the feature for trends. The whole data acquisition and monitoring process was implemented on common industrial hardware in order to illustrate that there are no technical restrictions for an industrial application. However, a validation of the approach with previously unseen components or long-term trials was discussed only as a topic for future work.

MAIER (2015) further extended the work by VERL et al. (2009) and WALTHER (2011) by additionally monitoring the torque of the feed drive motor. It could be observed that the torque of the motor for moving industrially decommissioned ball screws was strongly dependent on the axis position, whereas the torque for moving a new ball screw was almost constant along the entire axis length. Furthermore, several features were extracted from and transforms such as the Hilbert-Huang transform (HHT) were applied to the acquired signals. Overall, he demonstrated that wear in early stages could be accurately captured by the extracted features. Nevertheless, wear in later stages was not investigated and an approach for automatically classifying worn-out ball screws was not presented either.

JIA et al. (2019) retrofitted an industrial machine tool with an accelerometer placed on the ball screw nut. Their experiments lasted for 100 days and consisted of regularly operational modal analyses (OMAs) based on a free vibration response captured by the accelerometer. After the 100 days, the ball screw was reported to be worn out. From the OMA results, it could be observed that some of the identified natural frequencies and damping ratios suddenly showed an increasing variance towards the end of the lifetime. After replacing the worn-out ball screw, pitting damage on the spindle surface was detected. This indicated that the modal parameters were suitable features for assessing severe damage in the ball screw drive at later stages of its service life. This has previously been suggested by other authors as well (IMIELA 2006; MAIER 2015).

Instead of estimating modal parameters from vibration data and correlating them with preload loss, NGUYEN et al. (2019) directly calculated the preload level from a dynamic model designed for a single-axis feed drive test bench. After identifying the parameters of the dynamic model, deploying current sensors to the motor and designing and applying a test cycle for the excitation of the feed

drive, the authors showed how to directly estimate the preload level of the ball screw.

There are also researchers, who followed the strategy of designing novel sensors for facilitating condition monitoring. MÖHRING and BERTRAM (2012), for example, developed a sensory ball screw double nut system with integrated strain gauges which can directly measure the preload inside the ball screw nut. The additional components needed for processing the signals were mounted onto the flange of the nut. In a long-term experiment with a single ball screw, the sensory double nut was applied and could detect a loss in preload. However, the sensor was also sensitive to thermally induced preload differences, which hindered the assessment of wear.

SCHLAGENHAUF et al. (2019) presented an integrated camera system which was mounted to the ball screw nut and took photographs of the spindle surface. Based on those photographs, they identified pitting damage by using a convolutional neural network (CNN) which was able to correctly classify pitting damages with a mean accuracy of 91.50 %. In industrial use cases, where pollution due to lubrication and dust from the manufacturing process can distort the camera image, the approach has not yet shown to be effective.

The work presented by SCHLAGENHAUF et al. (2019) is not only an example for the development of novel sensors but also one for the recent trend in the literature to apply machine learning (ML) algorithms for solving the problem of feed drive condition monitoring. LI, JIA, et al. (2018), for example, conducted experiments on a single-axis feed drive test bench and used ML algorithms to detect wear from the acquired signals. Similarly to MAIER (2015), they acquired the velocities and motor torque signals from the CNC, as well as vibration signals from three accelerometers placed on each bearing and the carriage. The authors examined three ball screw drives and three linear guide systems with three different preload levels. The total of nine combinations were assembled onto the single-axis test bench for data acquisition. From this data basis, the authors extracted a large number of time-domain and frequency-domain features and selected the best performing ones with respect to correctly classifying the components' preload levels. It should be noted that the extraction of features and the classification were conducted on the same components (i.e. a strict separation of the training and the test data set on the component level was not conducted). As a consequence, no statement regarding the validity of the results for unseen components could be made.

DENKENA et al. (2021) proposed a similar approach as LI, JIA, et al. (2018) as they also performed experiments with ball screws with three different preload levels on a single-axis feed drive test bench. They acquired signals from the CNC and vibration signals from a single accelerometer mounted to the ball screw nut. Based on many predefined features from the time and the frequency domain, they conducted a principal component analysis (PCA) and kept the first two principal components for further analysis. In a last step, a classification model predicted the preload level based on the two principal components. This was done separately for the position error signals provided by the CNC and the vibration signals provided by the accelerometer. Although the prediction accuracies were equal to 100 % in almost all cases, it must be noted that the feature learning and the classification, which led to those high accuracies, were conducted with the same components. A strict separation of training and test data on the component level did not occur. Therefore, like in the case mentioned before, a statement regarding the validity of the results for unseen components could not be made.

In summary, it can be noted that the early work was mostly concerned with identifying potentially suitable signals for condition monitoring of ball screws. As a result, signals from the rotary encoder, the linear encoder, the torque from the feed drive motor, and vibration signals from additional accelerometers were selected and analysed, producing promising results in exploratory experiments. In a next step, many authors focused on constructing features from those signals, such that a stronger correlation with wear could be observed (i.e. constructing *wear-sensitive* features). Additionally, non-purely data-driven approaches, which directly estimate the preload level based on a dynamic model, were also presented. Those models have the advantage of better interpretability but are difficult to set up as identifying suitable parameters for them is difficult and must be repeated for each new feed drive system. The approaches, which integrated novel sensor systems, were indeed promising but led to additional costs. More recent data-driven approaches used the already available CNC signals and applied ML algorithms. The results were encouraging as the prediction accuracies with respect to correctly classifying preload loss were high. However, the presented results were not based on a strict separation between training and test data (i.e. the same components were used for training and testing the proposed ML approaches). Therefore, the validity of the reported results for previously unseen components is still unclear and has yet to be shown.

### 3.1.2 Linear Guides

The body of literature with respect to condition monitoring of linear guides is even smaller than the one on condition monitoring of ball screws. The four most relevant publications for this topic will be described below.

LI, JIA, et al. (2018), who were mentioned in the previous subsection already, can also be mentioned in this section, as their approach also covered monitoring the preload loss of linear guides. They collected CNC data and vibration signals, extracted features from those signals, and applied ML algorithms to classify linear guides by their respective preload level. As no strict training and test data split was conducted, the results cannot yet be regarded as transferable to unseen components.

TSAI et al. (2017) monitored preload loss of linear guides by regularly conducting an OMA based on vibration signals collected by three accelerometers mounted onto the carriage. The excitation of the feed drive was done by feeding a sine chirp signal to the motor current. They showed that their OMA-based approach could accurately yield the modal parameters which were also identified with a traditional experimental modal analysis (EMA). Using this approach for long-term tests and continuously monitoring the modal parameters was discussed only as a topic for future work.

KIM et al. (2021) conducted experiments on a test bench which only consisted of a single rail and runner block. They used a CNN for detecting faulty linear guides based on vibration signals collected by an accelerometer mounted onto the runner block. Their CNN was designed specifically to learn features which can be interpreted in the frequency domain, which facilitated the analysis and interpretation of the results. However, they used the same linear guide in their training and test data, limiting the transferability to unseen components.

SCHWARZENBERGER et al. (2022) proposed to use the so-called *isolation forests* anomaly detection method (LIU et al. 2008) for identifying wear. Their experiments were done on a life cycle test bench for linear guides, which was designed for accelerated lifetime tests, and a five-axis milling machine. In both cases, the data were acquired with affordable micro-electrochemical systems (MEMS) accelerometers mounted onto the carriage. During the experiments, pitting damage on the linear guide and the rolling elements of the runner blocks was observed. The isolation forests were trained on the data gathered from the life cycle test bench and performed well in detecting anomalies on the test data collected

from the machine tool. The anomalies were also detected when the model was trained on the machine tool data and tested on the life cycle test bench data demonstrating the transferability of the proposed approach. Furthermore, the detection of anomalies indeed occurred when the linear guides were damaged.

As a conclusion, it can be stated that monitoring the condition of linear guides is less explored in research. One reason for this might be that they do not fail as frequently as ball screws in many use cases. With the publication of SCHWARZENBERGER et al. (2022), however, work exists that demonstrated an approach which is accurate in detecting wear and can be successfully applied to unseen components. This has yet to be shown for condition monitoring of ball screws.

## 3.2 Methods for Remaining Useful Life Estimation

In contrast to the state of the art on condition monitoring of ball screw feed drives described in Section 3.1, a large number of review papers is available discussing general methods, which are independent of a specific use case, for RUL estimation.

LEI et al. (2018) published an extensive work reviewing methods for all four PHM process steps depicted in Fig. 2.6 (page 18). They found that most approaches for RUL estimation can be classified into the following categories: physics model-based approaches, statistical model-based approaches, artificial intelligence (AI) approaches and hybrid approaches, which combine the former three. They assessed that artificial neural networks (ANNs) were the most commonly used AI technique in the context of RUL estimation but concluded that there are still issues which must be addressed in the future. Two of these are especially important for this thesis: first, dealing with little or even no failure data; and second, managing the uncertainty of RUL estimates.

ZHANG et al. (2019) conducted a review on the application of deep ANNs (i.e. deep learning (DL)) in PHM (see Table 3.1). The main finding of their review was that DL can be universally applied to different sub-fields of PHM. Examples of such sub-fields are fault detection and diagnosis (i.e. condition monitoring) and the estimation of RUL. However, they also identified current challenges, one of which is especially relevant for this thesis: the data quality and data quantity in real-world use cases are often different from those presented in many scientific

| | number of publications per data type | | | |
|---|---|---|---|---|
| model | vibration | time series | image | structured |
| RNN | 5 | 5 | – | – |
| RBM | 5 | 1 | – | 7 |
| AE | 17 | – | – | 3 |
| CNN | 30 | 4 | 8 | – |

**Table 3.1:** Usage of the different DL model variants convolutional neural network (CNN), auto encoder (AE), restricted Boltzmann machine (RBM) and recurrent neural network (RNN) dependent on the available data type as reviewed by ZHANG et al. (2019); it can be seen that CNNs are the most frequently applied ones.

papers. Industrial data could be noisier, subject to environmental disturbances and, in general, scarce. The last is especially the case for failure data. Hence, methods addressing this issue must be developed in the future.

FINK et al. (2020) also provided a review about DL models in the context of PHM. In contrast to ZHANG et al. (2019), their review had a slightly wider perspective as they briefly reviewed the progress of DL in the ML community, the natural language processing (NLP) community and the computer vision community first, before they discussed the application of DL models in PHM applications. Finally, future research needs of DL in the context of PHM were identified, some of which are the following: first, domain adaptation models must be developed to handle tasks where the training data from the source domain differ from the data in the application domain. An example of this would be a machine tool feed drive on two different machines from different manufacturers. It is not guaranteed that collecting training data from one machine (i.e. the source domain) and applying a trained model to the other machine (i.e. the target domain) is successful. Today's DL approaches are vulnerable to such domain shifts. Second, the combination of DL and expert knowledge must be further investigated. Current approaches are still exploratory and there is no common consensus on how they can be transferred to practical use cases yet. Third, they found that the quantification and utilisation of uncertainty in RUL estimation must be addressed more comprehensively.

WEN et al. (2022) provided a more recent review of general data-driven approaches to PHM. Hence, not only DL models were reviewed but also more conventional ML methods such as support vector machines (SVMs) (BOSER et al. 1992) and decision trees (QUINLAN 1986). They emphasised that the availability of public failure and run-to-failure data sets for rotating machinery (e.g. bear-

ings) and aircraft systems has led to the fact that most publications describe the application of new methods to those data sets. In real industrial domains, they concluded, that the collection of data is still very costly, and failure data are often scarce or not available at all. This is especially the case for entire run-to-failure data sequences. Furthermore, they agreed with the above-mentioned authors and noted that the quantification of the uncertainty of RUL estimates is still an open research direction for future work.

Finally, BAUR et al. (2020), a source which was already introduced in Section 3.1, also concluded that, in the case of ball screw feed drives, future methods must reduce the need for run-to-failure data.

Summarising all that leads to the statement, that all recent and relevant review papers on the estimation of RUL in the context of PHM agree on the two shortcomings in the state of the art: first, many authors have developed methods which estimate the RUL as a single value (i.e. a point estimate) and only few have proposed methods for also quantifying the uncertainty of the RUL estimate. Having such an uncertainty estimate could facilitate making the decision when to replace a part before a failure occurs. Second, industrial use cases suffer from the absence or shortage of failure and run-to-failure data. As a result, many current methods cannot be applied to real use cases. Hence, methods are needed which can predict an RUL value based on little or even no run-to-failure data. The state of the art of both shortcomings will further be reviewed in more detail in the following subsections.

### 3.2.1 Quantifying and Utilising Uncertainty Information

Classical approaches for estimating a probabilistic RUL value, which inherently include uncertainty information, are statistical models for lifetime analysis. Those models assume that the lifetime of an object can be described by a structural degradation represented by a hazard function and noise. In so-called *proportional hazard models* (COX 1972), this idea is extended by integrating sensor data into the model, in order to explain the noise at least partially (LAWLESS 2003, pp. 8–36).

More recent approaches apply Bayesian ML models (as described in Subsections 2.3.2 and 2.3.3) to exploit the ML model's capability of explaining non-linear relationships between sensor signals and the associated RUL value. Fur-

thermore, Bayesian ML models automatically quantify the uncertainty of the RUL estimate (GHAHRAMANI 2015).

KRAUS and FEUERRIEGEL (2019) presented a model for predicting RUL values which consisted of three additive parts: first, a lifetime model as described above; second, a linear model which only relies on the sensor signals at the current time $t$; and third, a recurrent neural network (RNN) which relies on the entire sequence of sensor signals up to the current time $t$. They estimated the parameters with a variational inference (VI) algorithm (see Appendix A.3) and validated the approach with the C-MAPSS benchmark data set, which was originally published by SAXENA et al. (2008) (see Appendix C.1 for details). Although the obtained RUL prediction results were competitive compared to other results which were state-of-the-art at that time, the probabilistic nature of their approach was not fully exploited. It was not shown how to utilise and how to handle the uncertainty associated with the RUL estimates.

LI et al. (2020) also presented an approach which is based on both a lifetime distribution and a DL model. They used the sensor signals from a hydraulic system as inputs for an ANN which mapped them to the parameters of a lifetime distribution. Hence, their final RUL prediction was generated from the lifetime distribution whose parameters changed with increasing wear. In their experiments, they applied different lifetime distributions such as the Weibull distribution (see Appendix A.4) or the Gaussian distribution (see Subsection 2.3.3). The unknown parameters of the ANN were found by conducting VI (see Appendix A.3). Their approach had the advantage that the uncertainty could be separated into the following two sources: first, the *aleatoric* uncertainty, which describes the uncertainty of the DL model's parameters and is reflected in the probabilistic outcomes of the ANN; and second, the *epistemic* uncertainty, which describes the uncertainty due to random measurement errors and is reflected by the lifetime distribution. However, the authors did not discuss how to benefit from those uncertainties to enhance the decision-making process for scheduling maintenance actions.

PENG et al. (2020) made a comparison between conventional, non-probabilistic ANNs trained with backpropagation, as described in Subsection 2.3.2, and their Bayesian counterparts trained with VI. The comparison was done with the C-MAPSS turbofan engine data set and the FEMTO bearing data set (NECTOUX et al. 2012), which are further described in Appendices C.1 and C.2, respectively. The experiments showed, that the Bayesian and conventional models lead to

acceptable results with respect to the RUL prediction accuracy. Although the authors claimed that the additional information from the Bayesian model (i.e. the uncertainty) could be leveraged in decision-making, they did not outline how this can be achieved.

BIGGIO et al. (2021) proposed to apply a deep Gaussian process (GP) model, which sequentially stacks Gaussian process regression (GPR) models to represent more non-linear relationships between the inputs and outputs than a standard GPR model as described in Subsection 2.3.4 (DAMIANOU and LAWRENCE 2013). They validated their deep GP model on the N-CMAPSS data set, which is a more recent variant of the C-MAPSS data set published by ARIAS CHAO et al. (2021), and compared the prediction accuracy with conventional DL models and Bayesian DL models trained with VI. As a result, they demonstrated that the deep GP models yielded comparable prediction accuracies while providing sensible uncertainty estimates. Also in this case, the utilisation of the uncertainty information was not discussed.

In summary, the referenced authors have investigated the application of Bayesian ML models to the task of RUL estimation. Three types of models were applied: first, statistical lifetime models, which were extended with more flexible ML models; second, Bayesian DL models; and third, GP models. All of them have the property that they not only predict a point estimate but an entire probability density for the RUL value. All authors reported competitive prediction accuracies compared to the state of the art. It is notable, however, that all authors have used VI for model fitting and none has used Markov chain Monte Carlo (MCMC). Furthermore, an approach for utilising the derived uncertainties was not presented.

### 3.2.2 Data-Efficient Estimation of Remaining Useful Life Values

In industrial use cases, failure data and especially run-to-failure data, which reflect sensor data along the entire life cycle of a system, are scarce. Therefore, data-efficient methods are critical for the general applicability of RUL prediction methods. Although this insight was emphasised by the review papers presented in Section 3.2, there are hardly any authors who address this open issue. The few who did are summarised below.

LV et al. (2020) proposed a special neural network architecture which they called *sequence adaption adversarial network*, which was pre-trained on an auxiliary

training data set (i.e. data from the source domain) to reduce the amount of training data needed for the actual training data set of the use case of interest (i.e. the target domain). They conducted experiments with the C-MAPSS data set and showed that they could generate competitive prediction accuracies compared to the state of the art. Although they could reduce the training data set size, they still needed 16 run-to-failure sequences to demonstrate good results.

ZHANG et al. (2018) proposed to apply *transfer learning* to reduce the amount of training data needed. They used the C-MAPSS data set and showed that, in some cases, their transfer learning approach could indeed yield good RUL prediction accuracies, when it was trained on one subset of the dataset and applied (i.e. transferred) to another subset (see Appendix C.1).

Although the two above-mentioned approaches are more data-efficient than conventional DL models, they still rely on run-to-failure data. This is due to the fact that they map the sensor signals *directly* to an associated RUL value. In contrast to that, *indirect* approaches map the sensor signals to a health indicator (HI) value first. This HI value can be extrapolated into the future. With a threshold for the HI, an RUL value can be derived (see Fig. 3.1). Consequently, the estimated RUL values are not directly dependent on historical run-to-failure data but are only dependent on a historic sequence of HI values. This can be exploited to eliminate the need for entire run-to-failure sequences. The resulting task, however, is to design a wear-sensitive HI. In the following paragraphs, examples from the literature of indirect approaches for designing HIs and predicting RUL values will be presented.

NGUYEN and MEDJAHER (2020) suggested finding an optimal HI by first extracting time-domain features from the raw time series and recombining them according to predefined mathematical operations. For example, those operations included addition and subtraction. The multi-objective function of the optimisation problem aimed at optimising several evaluation criteria at once. For example, the resulting HI should have correlated with wear, followed a trend and exhibited monotonicity. The optimisation problem was solved with a genetic algorithm (ZITZLER et al. 2001). Their approach was tested with the C-MAPSS and FEMTO data sets (see Appendices C.1 and C.2, respectively) and showed good results with respect to maximising the objective function. However, they did not thoroughly investigate the influence of their optimised HIs on the RUL prediction performance.

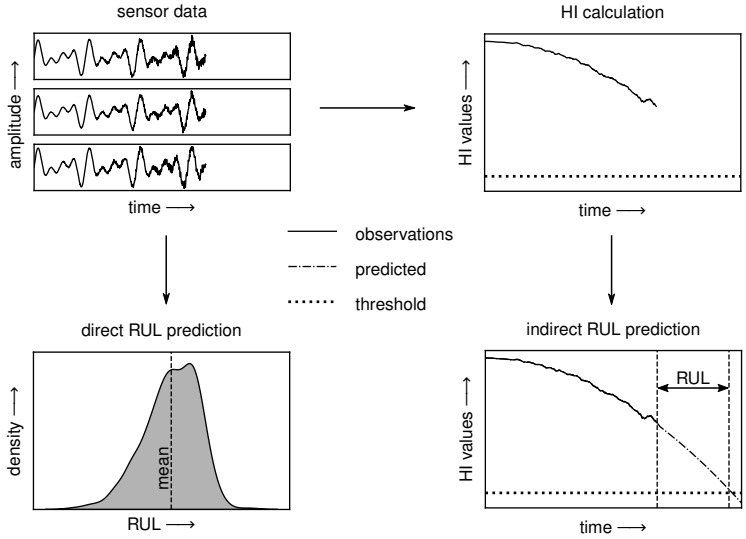



**Figure 3.1:** Illustration of direct and indirect RUL prediction approaches; on the left hand side, a direct approach is depicted, which calculates a probability density of the RUL based on the raw sensor signals directly. On the right hand side, an indirect approach is illustrated, which first calculates an HI. In a second step the HI is extrapolated. With the help of a threshold, an RUL prediction can be derived.

Finally, Li, Jia, et al. (2018), who were already mentioned in Subsections 3.1.1 and 3.1.2 as an example for condition monitoring of ball screws and linear guides, can also be referenced in the context of data-efficient RUL estimation. Based on a large number of features, which were extracted from the raw time series, they selected the most useful ones and constructed a simple HI with a linear regression model. Each parameter of the regression model was associated with one of the selected features. The regression model was fitted with data from the beginning and the end of the life cycle of a ball screw from their test bench. At the beginning, the HI was defined to be one and at the end of the lifetime the HI was defined to be zero. With this approach, entire run-to-failure sequences were not necessary any more, as one could simply use data from a new and a degraded component. They validated their approach with a ball screw run-to-failure experiment on their test bench and showed that their HI indeed had a trending, accelerated degradation pattern. The HI was extrapolated with a GPR model with the radial basis function (RBF) kernel denoted in Eq. (2.80). Although the approach by Li, Jia, et al. (2018) can be regarded as a big step towards an applicable and data-efficient method for the estimation of RUL values for industrial use cases,

such as machine tool feed drives, several open questions remain: first, the paper does not clearly explain what data were used for constructing and training the HI (i.e. the linear regression model). It is unclear whether a strict training and test data split was conducted. Second, the applied RBF kernel for the extrapolation with the GPR model is a stationary kernel (i.e. suitable for data, that does not follow a trend). However, the HI values do follow a trend by definition. Hence, there is still room for improvement regarding the GPR model.

Summarising all that leads to the conclusion, that the state of the art provides approaches which aim at estimating RUL values based on little or even no run-to-failure data. The latter is achieved by indirect approaches, which only rely on observations at the beginning and at the end of a component's life cycle, making entire run-to-failure sequences dispensable. With the work by LI, JIA, et al. (2018), even an approach exists for indirect RUL estimation for ball screw feed drives, which is also the main focus of this thesis. Nevertheless, open questions have been identified and gaps in the state of the art still exist. In the following section they will be summarised and a need for action will be derived.

## 3.3 Summary and Research Gaps

The review of the state of the art presented in the Sections 3.1 and 3.2 revealed that, although previous authors have contributed significantly to the field of condition monitoring for machine tool feed drives and methods for the estimation of RUL, there are still research gaps (RGs), which must be addressed. These RGs can be divided into the following two research categories: first, there are open questions regarding the practical application of methods for condition monitoring of machine tool feed drives:

**RG1** Most authors have conducted experiments on single-axis test benches, only. This is perfectly reasonable in order to develop a basic understanding about condition monitoring of machine tool feed drives and the relevant signals. Nevertheless, ball screw feed drives in industrial applications are part of larger and more complex systems, such as multi-axis machine tools. Hence, it must be doubted that the transfer of the findings made on single-axis feed drives to more complex systems can succeed. However, this is an unconditional prerequisite for applying a condition monitoring system in industrial machine tools.

**RG2** It has not yet been shown that the condition of previously unseen ball screws can be accurately predicted. Most of the known approaches for condition monitoring of machine tool feed drives are data-driven. Although they have the advantage of being very flexible, they can lead to over-fitting and subsequently to a model which fails to explain wear for previously unseen components. This would obstruct the transferability of a data-driven approach to industrial use cases and, hence, make it useless for practical applications. According to the state of the art, this transferability has only been previously shown for linear guides and has yet to be shown for ball screws.

Second, there are methodological gaps in the literature regarding suitable approaches for the estimation of RUL values for machine tool feed drives, a use case which goes hand in hand with the scarcity of failure observations and often even the complete absence of run-to-failure data:

**RG3** The methods proposed for RUL prediction according to the state of the art achieved good overall performances with respect to the mean squared error between the true and predicted RUL values. However, those approaches usually predicted a single RUL value (i.e. a point estimate). It is almost certain that this single predicted RUL estimate differs from the true RUL value in most cases. This *uncertainty* can be quantified, which has recently been demonstrated by several authors. Nevertheless, the state of the art falls short of an approach which can quantify the RUL prediction's uncertainty and, at the same time, utilise this uncertainty for better decision-making, namely for answering the question what point in time is the optimal one for replacing a component before it fails.

**RG4** Finally, the state of the art falls short of a data-efficient approach which is able to accurately predict the RUL values of a system for which historic run-to-failure data are not available and historical failure observations are scarce. This is especially critical for the estimation of RUL values in industrial use cases. Although some authors have proposed indirect approaches to the estimation of the RUL, which can spare the necessity of entire historic run-to-failure sequences, most of them still relied on a large amount of historic failure data. For machine tool feed drives, this amount of historic failure data is restrictively expensive to collect.

Based on the four RGs above, the next chapter derives research questions and a research approach for answering the open questions and extending the state of the art.

Chapter 4

# Research Approach

This chapter will explain the overall research approach. First, in Section 4.1, research targets (RTs) will be derived based on the identified research gaps (RGs) from the state of the art. Second, the proposed approach for addressing the RTs and ultimately closing the RGs will be presented in Section 4.2.

## 4.1 Research Targets

In Section 3.3, four RGs from two categories were identified. In order to further specify the need for action and find suitable targets, an RT will be derived for each identified RG. The two identified research categories from Section 3.3 will be retained.

#### Condition Monitoring of Machine Tool Feed Drives

**RG1** was defined as the lack of publications demonstrating a condition monitoring approach for machine tool feed drives on an industrial test bench, as most authors conducted their experiments on a simplified single-axis setup. Based on this RG, the following RT can be proposed:

**RT1** Derive a condition monitoring test cycle for an industrial test bench enabling a user to obtain reproducible measurements which are robust against environmental influences. This condition monitoring test cycle should integrate the insights gained from the state of the art. That includes the recording of the signals that were identified as relevant and the excitations that have shown to be successful in generating wear-sensitive signals.

**RG2** revealed that the data-driven condition monitoring approaches from the state of the art have not yet demonstrated good prediction results with respect to different failure modes for previously unseen ball screw drives. This lack of transferability is especially problematic as it can prevent a user from believing in a successful practical application, which ultimately hinders the adoption in industry. Therefore, the following RT can be noted:

**RT2** Deliver a data-driven condition monitoring approach which can successfully assess different fault modes of previously unseen ball screws. As the application to machine tool feed drives implies the scarcity of historic failure observations, such an approach should produce good results even when it is trained with only few historic failure observations.

#### Methods for the Estimation of Remaining Useful Life

**RG3** noted that the state of the art did not show how to utilise uncertainty information for supporting the decision of when to conduct a maintenance action. Hence, it has not been demonstrated how to use this additional information for determining a better end-of-life time of a component, which ultimately defines the last possible time for conducting a maintenance action. Due to that, the following RT arises:

**RT3** Develop a method for utilising the uncertainty information of the estimated remaining useful life (RUL) value of a probabilistic machine learning model. The method should achieve a more accurate prediction of the end-of-life time of a component, ultimately leading to an optimised time for a maintenance action.

**RG4** stems from the conclusion that data-efficient methods for the estimation of RUL values have not yet been sufficiently developed according to the state of the art. Past work on RUL estimation has heavily relied on the availability of historic failure observations and run-to-failure sequences. In many industrial applications, run-to-failure data of a component are scarce or do not even exist. Hence, the following RT must be achieved:

**RT4** Develop a method for the estimation of RUL values based on only few historic failure observations. The method should be able to predict RUL values as accurately as, or close to the accuracy of, state-of-the-art methods, which rely on many historic failure observations or run-to-failure sequences.

By achieving the RTs defined above, the RGs are to be closed. The next section will introduce the publications within this publication-based dissertation.

## 4.2 Proposed Approach

The proposed approach of this thesis consists of four publications addressing the RTs defined in the previous section. Additionally, the economic potential of the proposed approach was investigated and the results of that were described in a fifth publication. Below, all publications will be briefly introduced and their contributions to the RTs will be highlighted. Furthermore, each publication led to findings which will further be summarised in a single key finding (KF) for each of them in Section 7.1.

**RT1** is pursued by **publication** ①, which was concerned with the derivation of a condition monitoring test cycle for machine tool feed drives on a fairly realistic, industrial test bench (see Appendix B.1). The resulting test cycle enables the recording of reproducible measurements which are robust against environmental influences and capture wear-sensitive signals of machine tool feed drives.

**RT2** is addressed by **publication** ②, in which a data-driven condition monitoring approach trained on only few historic failure observations was investigated. The approach was tested on previously unseen ball screws and the data were recorded using the test bench and test cycle derived in publication ①. Special attention was paid to the approach's ability of correctly predicting different fault modes of previously unseen ball screws.

**RT3** is achieved by **publication** ③, in which a new method for the utilisation of the uncertainty of RUL estimates was introduced. The novel method was compared to the state of the art with respect to accurately predicting RUL values. The comparison was carried out with the C-MAPSS benchmark data set (see Appendix C.1), and it was investigated whether the new method could enhance the decision-making process for selecting a more suitable time for maintenance actions.

**RT4** is accomplished by **publication** ④, in which a novel method for RUL estimation was presented which can be trained with only few historic failure observations. Based on the trained model, RUL values for previously unseen components can be calculated. The method was validated with the C-MAPSS and

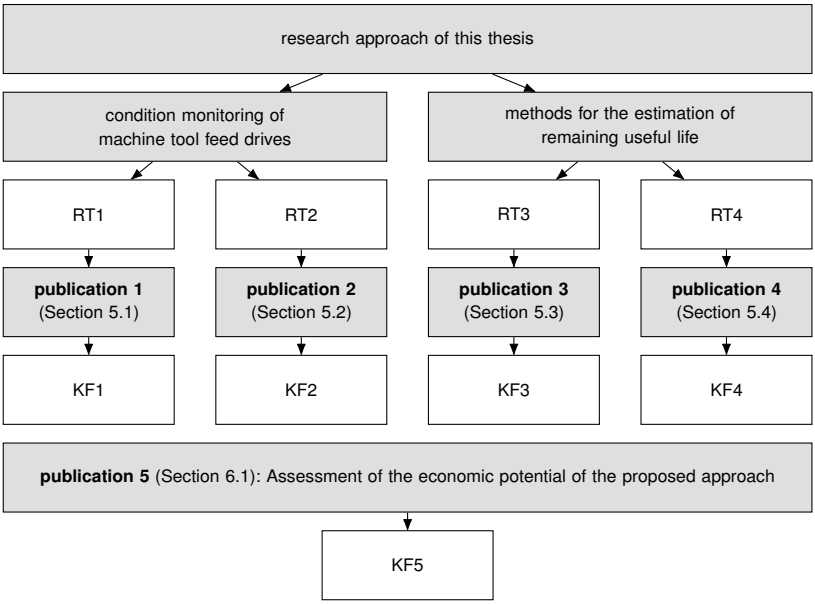



**Figure 4.1:** Research approach of this thesis

FEMTO data sets (see Appendices C.1 and C.2) and was compared to state-of-the-art methods, which rely on a large number of entire run-to-failure sequences. In particular, the impact of the reduced data set with respect to RUL prediction accuracies was quantified.

Additionally, **publication** ⑤ assessed the economic potential of the overall research approach of this thesis.

All publications are listed below with their full bibliographic data and the research approach is depicted in Fig. 4.1.

**Publication** ① BENKER, M., JUNKER, S., ELLINGER, J., SEMM, T., and ZAEH, M. F., (2022). "Experimental Derivation of a Condition Monitoring Test Cycle for Machine Tool Feed Drives". In: *Production Engineering* 16.1, pp. 55–64. DOI: 10.1007/s11740-021-01085-9

**Publication** ② BENKER, M. and ZAEH, M. F., (2022). "Condition Monitoring of Ball Screw Feed Drives Using Convolutional Neural Networks". In: *CIRP Annals* 71.1, pp. 313–316. DOI: 10.1016/j.cirp.2022.03.017

**Publication ③** BENKER, M., FURTNER, L., SEMM, T., and ZAEH, M. F., (2021). "Utilizing Uncertainty Information in Remaining Useful Life Estimation via Bayesian Neural Networks and Hamiltonian Monte Carlo". In: *Journal of Manufacturing Systems* 61, pp. 799–807. DOI: 10.1016/j.jmsy.2020.11.005

**Publication ④** BENKER, M., BLIZNYUK, A., and ZAEH, M. F., (2021). "A Gaussian Process Based Method for Data-Efficient Remaining Useful Life Estimation". In: *IEEE Access* 9, pp. 137470–137482. DOI: 10.1109/ACCESS.2021.3116813

**Publication ⑤** BENKER, M., ROMMEL, V., and ZAEH, M. F., (2022). "An Investigation into the Economic Efficiency of Different Maintenance Strategies Based on a Discrete Event Simulation". In: *Procedia CIRP* 107, pp. 428–433. DOI: 10.1016/j.procir.2022.05.003

An overview and a categorisation of the first four publications with respect to the application they aim at and the steps of the general prognostics and health management (PHM) process they are concerned with are given in Fig. 4.2. It can be seen that publications ① and ② are exclusively concerned with the application to machine tool feed drives, whereas publications ③ and ④ are methodological in nature and therefore are not restricted to specific applications or machines or components.

The next chapter will describe the details and findings of publications ① – ④. The content of publication ⑤ will be summarised in Chapter 6.

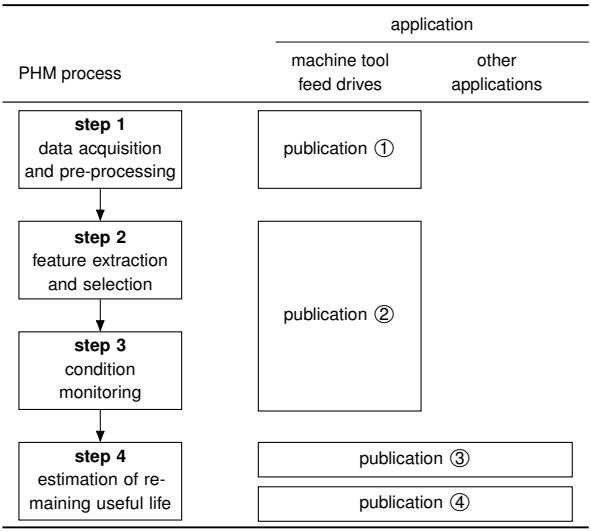



**Figure 4.2:** Overview and categorisation of the publications; each publication can be categorised by one or several PHM process steps it addresses and by the application it aims at.

Chapter 5

# Research Results

This chapter will present the publications addressing the research targets (RTs) defined in Section 4.1. In Sections 5.1 to 5.4, the content of each publication will be summarised[1], the resulting findings will be stated and assessed with respect to achieving their respective RT.

## 5.1 Publication 1: Derivation of a Condition Monitoring Test Cycle

The content of publication ① will be summarised in Subsection 5.1.1, whereas the findings will be stated in Subsection 5.1.2. The overall objective was to derive a condition monitoring test cycle that is robust against environmental influences and which is able to record reproducible data for condition monitoring of machine tool feed drives.

### 5.1.1 Summary

The overall objective of publication ① was to first derive a novel condition monitoring test cycle for machine tool feed drives and, second, demonstrate its applicability and ability to acquire data for a more realistic system than a single-axis test bench, which was used in the state-of-the-art publications. In order to achieve this objective, an experimental test bench, which is closer to a real machine tool, was set up at the Institute for Machine Tools and Industrial

[1]The notations of the original publications, which can slightly differ from those used in Chapter 2, will be retained. However, all notations will be clarified in their context.

Management of TU Munich (in German: Institut für Werkzeugmaschinen und Betriebswissenschaften der TU München) (*iwb*). The test bench consisted of a DMG DMC duo Block 55H milling machine without a housing, a spindle system and a rotary axis (see Appendix B.1 for details). In addition, ball screws and runner blocks with different wear states were available for the investigated machine's X-axis (see Tables B.1 and B.2). All measurements were conducted with the machine tool's Heidenhain iTNC530 computer numerical control (CNC) and three additional external accelerometers mounted to the ball screw nut and onto two runner blocks. The external accelerometers provided vibration signals, whereas the CNC delivered position signals from the rotary and linear encoder as well as the motor current and torque. Hence, all signals considered useful for condition monitoring by the authors mentioned in the chapter on the state of the art were recorded.

The final condition monitoring test cycle (see Fig. 5.1) was designed so that the system reactions to the following three excitations were captured after a warm-up phase of 60 min:

**Constant speed excitation** During this excitation, the carriage was moved back and forth along the entire axis length from $-1300$ mm to $-700$ mm[2] with a commanded feed rate of 24 000 mm/min. Both signals from the internal oscilloscope of the CNC and vibrations from the external accelerometers were recorded during the excitation.

**Direction change excitation** The carriage was moved back and forth for only a small travel distance from $-1000$ mm to $-999$ mm with a commanded feed rate of 24 000 mm/min. This excitation was conducted around the half way position of the ball screw spindle, and both signals from the internal oscilloscope of the CNC and vibrations from the accelerometers were recorded.

**Sine sweep excitation** In order to obtain a frequency response function (FRF), a sine sweep signal from 1 Hz to 500 Hz was fed to the motor. The excitation was implemented as a velocity set point of the motor. Signals from the rotary and linear encoders were recorded with the Heidenhain TNCopt software tool. They were treated as excitation and response, respectively, and an FRF was calculated by dividing the Fourier transform of the response by the Fourier transform of the excitation. In addition, the vibrations from the external accelerometers were also

[2] Due to the machine's reference position, the values are negative. Position $-1300$ mm is at the left end of the axis, position $-700$ mm is at the right end of the axis and position $-1000$ mm is in the middle of the axis.

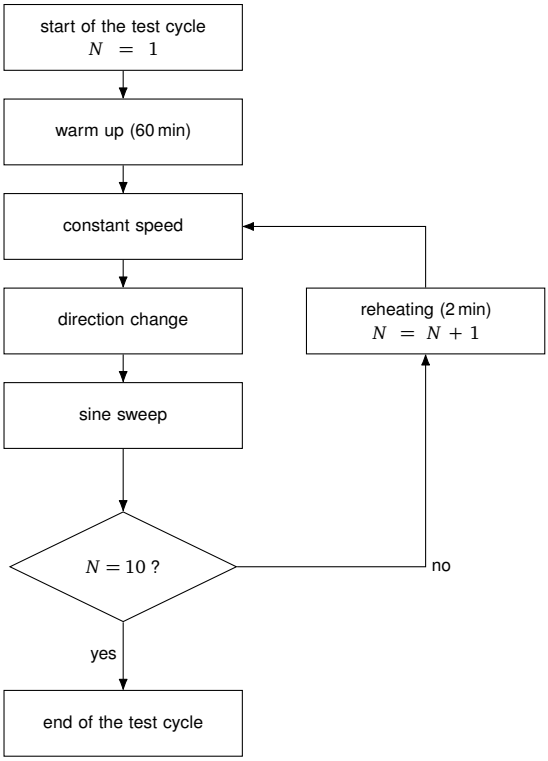



**Figure 5.1:** Flow chart of the resulting test cycle as described in BENKER, JUNKER, et al. (2022)

recorded. The sine sweep excitation was performed at an axis position close to the motor (i.e. at the X-axis position $-1300\,\mathrm{mm}$).

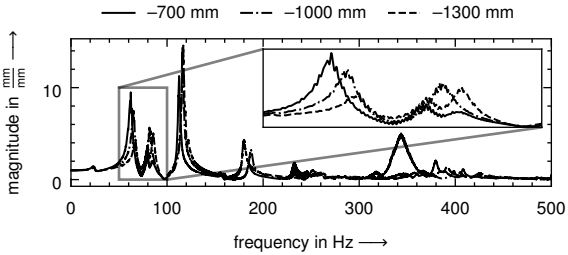
In a second step, the main factors of disturbance on the measurements were identified in order to control them and to ensure that the test cycle generated reproducible measurements. The three main factors of disturbance were identified as the following:

**Axis position of the investigated feed drive** Preliminary investigations by authors of state-of-the-art research papers and by the author of this thesis have indicated that modal parameters (see Subsection 2.1.3) are useful features for condition monitoring, as they change with wear (ELLINGER et al. 2019). Hence, accurately identifying and monitoring them is considered a promising strategy for condition monitoring. For this purpose, the axis position of the feed drive had

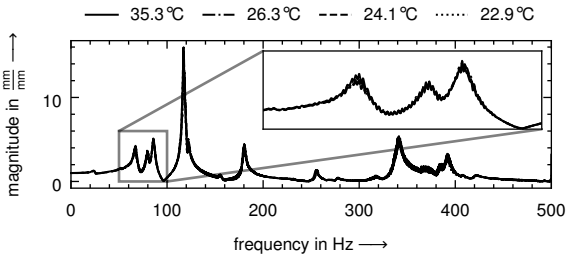
to be controlled because different axis positions lead to different distributions of the involved masses, changed dynamic characteristics and ultimately different modal parameters. This can be seen in Fig. 5.2a, where three FRFs from the investigated feed drive dependent on the X-axis positions are shown. In particular, a mode shift became visible. Hence, for the purpose of condition monitoring, the axis position must always be the same to receive comparable and reproducible measurements from the test cycle.

**Temperature of the feed drive** Thermal effects on machine tools have been extensively studied in the past and are generally considered an important influence (MAYR et al. 2012). For example, LEE and DONMEZ (2007) showed that with higher temperatures of the tool-holder-spindle its natural frequencies change. Hence, for high precision machinery, the temperature is a disturbing factor. Therefore, the temperature's influence on the data collected with the proposed condition monitoring test cycle had to be investigated and ultimately controlled, too. This was done in experimental investigations, in which the considered feed drive was heated up by traversing the carriage back and forth along the entire axis length from −1300 mm to −700 mm with a commanded feed rate of 24 000 mm/min for approximately 90 min. During this time, the temperature measured on the surface of the ball screw nut was observed to converge to 36 °C. Afterwards, the test cycle was conducted every ten minutes (i.e. regularly whilst the feed drive was cooling back down to ambient temperature). In Fig. 5.2b, it can be seen that in a temperature range from 22.9 °C to 35.3 °C, the FRFs were not significantly influenced by the temperature, as they are almost perfectly aligned.

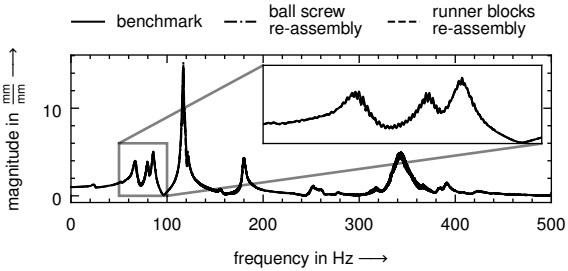
**Re-assembly of the feed drive components** A potentially major disturbance factor is the re-assembly of components. In order to avoid this, a standardised re-assembly procedure according to the manufacturer's assembly instructions was followed. To investigate the influence of re-assembly actions, the test cycle was conducted before and after a ball screw and runner block re-assembly. Figure 5.2c shows that the FRFs did not significantly change after a ball screw re-assembly nor after a runner block re-assembly. This was not the case for the external accelerometer measurements, which exhibit changes in the frequency ranges from 1 Hz to 50 Hz and from 420 Hz to 500 Hz. Hence, those frequency ranges had to be excluded from this signal for further analysis.




**(a)** FRFs dependent on the X-axis position




**(b)** FRFs with the temperature measured at the surface of the ball screw nut as parameter




**(c)** FRFs dependent on re-assemblies of the ball screw and the runner blocks

**Figure 5.2:** Sample measurements from the final test cycle as in BENKER, JUNKER, et al. (2022); in panel (a) it can be seen, that the calculated FRF from the test cycle strongly depends on the X-axis position. In panels (b) and (c), the recorded data is almost perfectly aligned demonstrating that the developed test cycle enables the acquisition of reproducible data independent of temperature changes and the re-assembly of components. The FRFs were calculated from the velocity signals from the rotary and linear encoder.

### 5.1.2 Findings

For publication ①, a novel condition monitoring test cycle was successfully designed and implemented on a test bench, which, in contrast to the state of the art, is comparable in complexity to a real machine tool. The derived test cycle enables the recording of different signals during various excitations. Some of them have already been identified as promising for the condition monitoring of machine tool feed drives by other authors in state-of-the-art research. Furthermore, it was shown that the test cycle is robust against environmental disturbance factors. Hence, reproducible measurements can be recorded. In summary, the findings of publication ① are the following:

- A novel test cycle for the acquisition of condition monitoring data for a realistic machine tool test bench on industrially available hardware was defined. The test cycle considers knowledge from the state of the art by recording signals which were already found useful for condition monitoring of feed drives by previous authors.
- The derived test cycle is robust against the following two disturbing factors: temperature of the axis and re-assembly of components. It is sensitive to the machine's axis position, which can be controlled, however. Hence, the test cycle enables the recording of reproducible measurements. It can serve as a basis for data acquisition and subsequent development of novel condition monitoring approaches.

Overall, it can be noted that all objectives declared in **RT1** were fully addressed and achieved.

## 5.2 Publication 2: Condition Monitoring of Ball Screw Feed Drives

The following subsections will summarise the content and findings of publication ②. The overall objective was to develop an approach to assess the wear state of a given, and previously unseen, ball screw with respect to preload loss and pitting damage.

### 5.2.1 Summary

The second publication in this thesis describes the application of the condition monitoring test cycle derived in publication ① (see Section 5.1) on the same test bench described in Appendix B.1 for the ball screws listed in Table B.1. All ball screws were subsequently assembled into the X-axis of the test bench. After that, the condition monitoring test cycle was conducted for data acquisition. As depicted in Fig. 5.1, this implied a warm-up for 60 min and subsequent recording of signals for the following three excitation patterns: constant speed, direction change and sine sweep excitation. The recorded signals were provided by the machine tool's CNC and external accelerometers. This resulted in a data set with measurements for all available ball screws.

The ball screws were provided with different preload levels from the manufacturer. Furthermore, some had a pitting damage, which had been provoked before on special test benches by the manufacturer. Hence, the ball screws showed different fault modes (i.e. pitting damage and preload loss), and the experiments were defined as classification tasks for the following three class definitions:

**Preload loss** In this setup, the ball screws with pitting damage were discarded from the experiments. The remaining ball screws with a high preload level (denoted as C3 by the manufacturer) were labelled *healthy*, and ball screws with a low preload level (denoted as C1 by the manufacturer) were labelled *degraded*.

**Pitting damage** For this class definition, all ball screws were considered for the experiments. The ones without a pitting damage were labelled *healthy*, and the ones with pitting damage were labelled *degraded*.

**Indistinct defect** Assuming a user does not want to differentiate between a loss in preload and pitting damage and must conduct a maintenance action in both cases, this class definition labelled ball screws with both a high preload level and no pitting damage as *healthy* and ball screws with a low preload level or pitting damage as *degraded*.

In each experiment, binary labels were assigned to each component dependent on the respective class setup (i.e. the label 0 for *healthy* and the label 1 for *degraded* ball screws; see Table B.1 in Appendix B.1).

The data set was pre-processed according to the following steps: first, the training and test observations were separated on the component level such that one

component from each class (i.e. class 0 and class 1) was available for training. The remaining components were set aside for testing. Second, the recorded time series were cut into equally sized sub-series which covered an entire excitation. This resulted in each original time series being cut into four sub-series with $d = 18\,750$ samples for the constant speed excitation and 40 sub-series with $d = 1875$ samples for the direction change excitation. Since the recordings of the original time series did not always start at exactly the same position, the resulting sub-series were slightly shifted. This was accepted on purpose to impede the training process and to avoid over-fitting. The sub-series $\mathbf{x}_i$ and their associated label were fed into a convolutional neural network (CNN) model, which will be described below.

For each combination of excitation and signal, the CNN model, as originally presented by ROMBACH et al. (2021), was applied. The model consists of an encoder (i.e. feature learner) and a classifier. The encoder received a sub-series as input vector $\mathbf{x}_i$. The architecture of the encoder is a composition of successive convolutional layers with kernel output sizes 64, 32, 16 and 8. Each convolutional layer is followed by a leaky rectified linear unit (ReLU) activation function with slope 0.1, a one-dimensional max pooling layer and a batch normalisation layer. For better regularization, the last convolutional layer is followed by a dropout layer with probability 0.5. The output of the encoder is reshaped to be a one-dimensional vector, which is mapped into a two-dimensional latent space vector $\mathbf{z}_i$ with the help of a fully connected layer. The classifier uses the latent space representation, normalises the latent space vector such that $\|\mathbf{z}_i\|_2 = 1$, and maps it to an output vector $\hat{\mathbf{y}}_i$ of size two with the help of a single fully connected layer (see Fig. 5.3 for an illustration of the approach). The model was trained for 100 epochs with the Adam optimiser (KINGMA and BA 2017) and with an initial learning rate of 0.01. In case the training loss did not decrease by a value of at least 0.0001 after ten successive epochs, the learning rate was halved. The results of the experiments on accurately predicting the class labels on the test data set are summarised in Table 5.1.

Because the test data set was slightly imbalanced (i.e. there were not equal numbers of observations from each class in the test data set), the reporting of the conventional accuracy, which represents the share of the correctly predicted observations, would have given a biased result. Therefore, the *balanced accuracy*, which accounts for an imbalanced data set, was applied. It can be seen that the different combinations of signals and excitations led to different accuracies

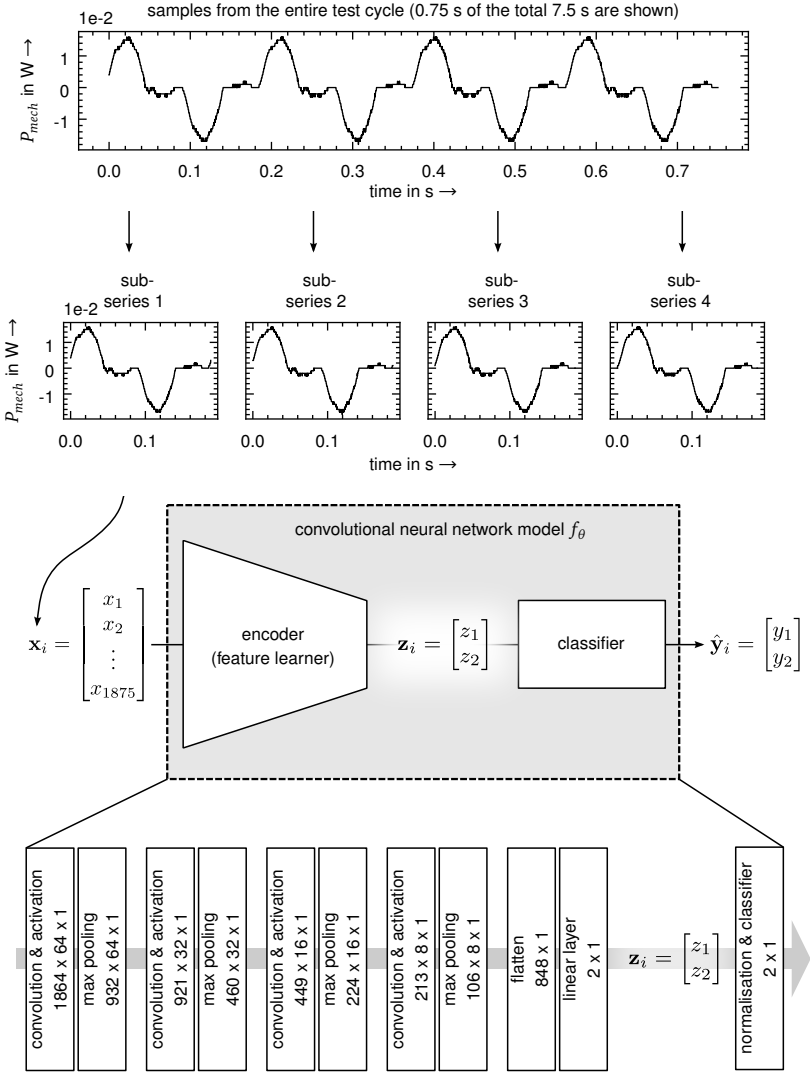



**Figure 5.3:** Illustration of the approach followed in publication ② based on BENKER and ZAEH (2022); the signal $P_{mech}$ from the ball screw C11 (see Table B.1) during a direction change excitation is shown. It was cut into equally sized sub-series, which served as inputs for a CNN model. At the bottom of the figure, the neural network's architecture is illustrated and the output sizes of the tensors after each layer are given.

| | | | balanced accuracy in % | |
|---|---|---|---|---|
| class setup | signal | excitation | mean | standard deviation |
| preload loss | $P_{mech}$ | direction change | 91.90 | 11.76 |
| | $a_z$ | direction change | 62.95 | 5.14 |
| | $P_{mech}$ | constant speed | 52.83 | 13.25 |
| pitting damage | $a_y$ | constant speed | 73.07 | 7.23 |
| | $a_z$ | constant speed | 66.21 | 22.59 |
| | $a_x$ | direction change | 57.72 | 27.58 |
| indistinct defect | $P_{mech}$ | direction change | 98.84 | 1.16 |
| | $P_{mech}$ | constant speed | 68.20 | 7.78 |
| | $a_x$ | direction change | 51.55 | 11.42 |

**Table 5.1:** The prediction results as in BENKER and ZAEH (2022) sorted by the highest mean balanced accuracy of the experiments are shown. Each mean and standard deviation (std) is based on five repetitions with random training and test data splits.

for the three class setups. The pitting damage was difficult to detect with the proposed approach, reaching only a mean balanced accuracy of 73.07 %. In contrast to that, a loss in preload could be detected with a mean balanced accuracy of 91.90 %. In the case where a distinction between a pitting damage and a loss in preload was not necessary, the mean balanced accuracy even reached 98.84 %.

### 5.2.2 Findings

In publication ②, a data-driven approach for condition monitoring of ball screw feed drives was proposed. It relies on data recorded with the test cycle from publication ① and predicted the loss in preload and pitting damage of ball screws with the help of a CNN model. A notable feature of the proposed approach is that it could achieve high prediction accuracies, although it only had access to a single component of each class during training. Hence, the approach can be regarded as data-efficient, which encourages a practical application. In summary, the findings of publication ② are as follows:

- A purely data-driven approach is capable of accurately predicting the current wear state of previously unseen ball screws when it is based on data recorded with the test cycle derived in publication ①.
- The proposed approach can be trained on a single component per class. Even with such little historic failure data, high prediction accuracies can be reached.

- The overall best performing combination of signal and excitation is the $P_{mech}$ signal recorded during the direction change excitation in the indistinct defect class setup. In this case, the balanced accuracy of five random experiments has a mean value of 98.84 % and a very low standard deviation of 1.16 %. Hence, in the case where a distinction between a loss in preload and pitting damage is not necessary, which holds true as in both cases maintenance actions must be conducted, the proposed approach is particularly effective.

In conclusion, **RT2** was fully addressed and the goals were almost reached in full. Only the detection of pitting damage with the proposed approach must be further investigated in the future.

## 5.3 Publication 3: Uncertainty Information in Remaining Useful Life Estimation

The following subsections will summarise the content and findings of publication ③. The overall objective was to present an approach for quantifying the uncertainty of remaining useful life (RUL) predictions, and utilising this uncertainty information to determine better end-of-life times of a component under investigation. In order to be able to compare the results to the state of the art, all experiments were conducted on the widely used C-MAPSS benchmark data set (LEI et al. 2018), which consists of several subsets of simulated run-to-failure sequences of turbofan engines (see Appendix C.1 for details).

### 5.3.1 Summary

For the publication, two aspects were investigated: first, it compared Bayesian neural networks trained with variational inference (VI) and Markov chain Monte Carlo (MCMC) (see Appendices A.2 and A.3) for the estimation of RUL prediction values. Second, a novel approach for utilising the uncertainty information provided by a Bayesian neural network was proposed and applied.

The C-MAPSS data set was pre-processed according to LI, DING, et al. (2018): first, a label rectification step limited the maximum value of the historic RUL values to 125 cycles, which had demonstrated to enhance the RUL prediction

| | subset | | | |
|---|---|---|---|---|
| | FD001 | FD002 | FD003 | FD004 |
| observations in the training data set | 17 731 | 48 819 | 21 820 | 57 763 |
| observations in the test data set | 100 | 259 | 100 | 248 |
| number of selected features $d$ | 14 | 24 | 14 | 24 |
| time window length $L_w$ | 30 | 20 | 30 | 15 |

**Table 5.2:** Summary of pre-processed C-MAPSS training and test data as in BENKER, FURTNER, et al. (2021)

accuracy. Second, the sensors 1, 5, 6, 10, 16, 18 and 19 have shown to provide little to no information about the RUL values in the subsets FD001 and FD003 and were excluded from further analysis in those subsets. Third, the data was normalised using a *min-max normalisation* to ensure that the different sensor signals were on the same scale. Finally, the run-to-failure sequences from the training data set were cut into equally sized sub-series with the sliding window approach presented by SATEESH BABU et al. (2016). By applying a sliding window of size $L_w$ onto the original time series, sub-series of size $L_w \times d$ resulted, where $d$ is the number of selected features (i.e. sensors). Since the training data consisted of entire run-to-failure sequences, it was possible to assign an RUL value to each sub-series. In Table 5.2, the pre-processed data sets for the different subsets are summarised.

In the experiments, two neural network architectures were applied, which will be briefly described below:

**D3** The D3 model was designed as a baseline model with a one-dimensional input layer and three fully connected hidden layers (sometimes also referred to as dense layers (GÉRON 2019, p. 282), and hence the name D3). Each of the hidden layers has 100 neurons with sigmoid activation functions. The output layer consists of a single neuron which returns the RUL estimate.

**C2P2** This model was originally proposed by SATEESH BABU et al. (2016) and consists of two successive convolutional and average pooling layers (hence the name C2P2). The convolutional layer output sizes are 8 and 14, respectively. The convolutional kernel has the size $5 \times 14$ in the first layer and $2 \times 1$ in the second layer, where $d$ is the number of input features. Each average pooling layer has the kernel size $2 \times 1$ and is followed by a sigmoid activation function. After the two convolutional layers, the output is flattened and mapped to a single output neuron, which returns the RUL estimate.

Both of the two model architectures were trained with the following four inference schemes:

**Backpropagation and stochastic gradient descent (BP)** For a baseline comparison to the Bayesian neural networks, classic neural network training with backpropagation and gradient descent, as described in Subsection 2.3.2, was conducted. The loss function was defined to be the *mean squared error*, which can be regarded as a variant of the *sum-of-squares error* defined in Eq. (2.48). The optimisation was conducted with the Adam optimiser (KINGMA and BA 2017) with an initial learning rate of 0.001. After 200 epochs, the learning rate was reduced to 0.0001 for another 50 epochs. As a result, a single set of neural network parameters $\boldsymbol{\theta}$ was obtained.

**Hamiltonian Monte Carlo (H1)** The Hamiltonian Monte Carlo (HMC) algorithm applied in this publication was originally proposed by NEAL (1996) and is also often called *hybrid Monte Carlo* (HOFFMAN and GELMAN 2014). In contrast to classical MCMC algorithms, HMC is faster in exploring high-dimensional spaces and, therefore, more suitable for sampling from the unknown neural network parameter vector $\boldsymbol{\theta}$. The HMC implementation of the *hamiltorch* open-source software presented by COBB and JALAIAN (2021) was used. Four hyper-parameters had to be set: first, the initial step size was defined as 0.0012. Second the number of leapfrog steps was set to 10. Third, the variance of the Gaussian prior over the parameters $p(\boldsymbol{\theta})$ was set to 100, and, finally, the variance of the observations was set to the sample variance of all RUL values in the training data set. In total, 500 samples were generated, from which the first 250 were discarded[3].

**Hamiltonian Monte Carlo pre-trained (H2)** This was an extension of the H1 sampling scheme introduced above. Instead of randomly initializing the mean of the prior of the parameter vector $p(\boldsymbol{\theta})$, the prior mean was set to the result of the BP inference described above. In total, 50 samples were generated, from which the first 25 were discarded[3].

**Variational inference (VI)** In order to compare the HMC inference approach to a VI approach, an algorithm from the open-source software package *BLiTZ* (ESPOSITO 2020) was used. The optimisation was conducted for 250 epochs with an Adam optimiser (KINGMA and BA 2017), a mean squared error loss function and an initial learning rate of 0.001, which was decreased to 0.0001

[3]When sampling from a Markov chain, it is good practice to discard the initial samples to avoid a bias towards them. This is also referred to as *burn-in* (ANDRIEU et al. 2003).

after 200 epochs. The prior over the unknown parameters $p(\boldsymbol{\theta})$ was set to be Gaussian with a standard deviation $\sigma = \log(1 + \exp(-1))$. For approximating the loss function in each step, 10 samples were randomly drawn from the posterior.

The three Bayesian inference schemes (i.e. H1, H2 and VI) produced probabilistic RUL predictions, and in those cases, the uncertainty of the RUL prediction was quantified. Furthermore, in publication ③, a novel approach for utilising this uncertainty was proposed. The approach is illustrated in Fig. 5.4 and will be briefly explained: first, it can be assumed, that too early maintenance actions are less costly than unexpected failures, which is reflected in the following cost function (Saxena et al. 2008):

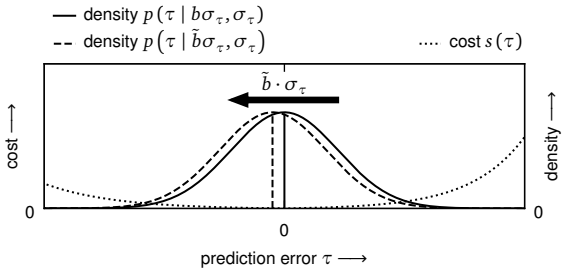
$$s(\tau) = \begin{cases} s_1(\tau) = e^{-\tau/13} - 1 & \forall\ \tau < 0 \\ s_2(\tau) = e^{\tau/10} - 1 & \forall\ \tau \geq 0, \end{cases} \tag{5.1}$$

where $\tau$ is the prediction error between the predicted RUL $y^\star$ and the true RUL $y$. This cost function is shown in Fig. 5.4a, where it becomes visible that early replacements cause lower costs than late replacements after unexpected failures. This asymmetric nature of the cost function can be exploited when the distribution of the prediction error is assumed to be Gaussian according to

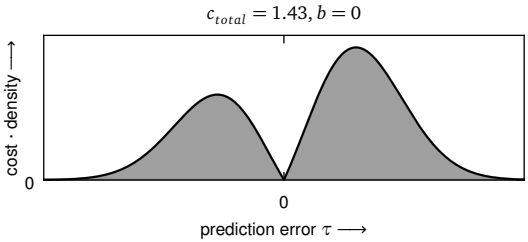
$$p(\tau \mid \mu_\tau, \sigma_\tau) = p(\tau \mid b\sigma_\tau, \sigma_\tau) = \mathcal{N}(b\sigma_\tau, \sigma_\tau), \tag{5.2}$$

where the mean $\mu_\tau$ is re-parametrised and defined to be $b\sigma_\tau$. In this case, artificially shifting the distribution of $\tau$ to the left (i.e. smaller mean values) might increase the prediction error but decrease the overall costs. This is shown in Fig. 5.4. Without artificially shifting the distribution of the prediction error (i.e. $b = 0$) the distribution of all costs results in the illustration of Fig. 5.4b. When shifting the distribution of the prediction error by a factor $b = \tilde{b}$, the distribution of the total costs differs (see Fig. 5.4c). The area under both curves, which can be interpreted as the total expected costs $c_{total}$, is smaller when the distribution is shifted. In the publication, it was further shown how to formulate an optimisation problem for finding an optimal value for $\tilde{b}$. The optimisation problem was solved numerically for some values of $\sigma_\tau$ and applied to the predictions from all Bayesian models described above.

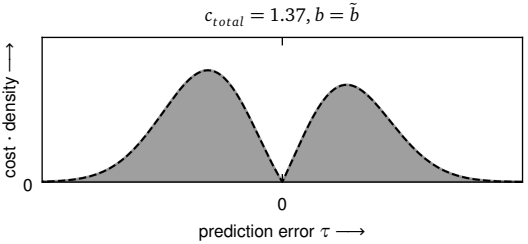
The results for the baseline model D3 in combination with the four different inference schemes are summarised in Table 5.3. The RMSEs, the MAEs, the




**(a)** Adjusting the RUL prediction leads to an adjusted distribution of the prediction error $\tau$.




**(b)** Total expected costs without an adjusted RUL prediction.




**(c)** Total expected costs with an adjusted RUL prediction.

**Figure 5.4:** Illustration of the proposed approach of publication ③ for utilising uncertainty in RUL estimation based on BENKER, FURTNER, et al. (2021); in panel (a), the RUL prediction error $\tau$, which is assumed to be Gaussian with $p(\tau \mid b\sigma_\tau, \sigma_\tau)$ and $b = 0$, and the cost function $s(\tau)$ are shown. Due to the asymmetric nature of the cost function, unexpected failures (i.e. $\tau > 0$) are more expensive than too early maintenance actions (i.e. $\tau < 0$). This can be seen in panels (b) and (c). In panel (b) no adjustment to the prediction is made resulting in total expected costs of $c_{total} = 1.43$. In panel (c) the prediction is adjusted by a factor $\tilde{b}$ times the variance $\sigma_\tau$ resulting in lower total expected costs of $c_{total} = 1.37$.

| | | BP | | H1 | | H2 | | VI | |
|---|---|---|---|---|---|---|---|---|---|
| subset | metric | mean | std | mean | std | mean | std | mean | std |
| FD001 | RMSE | **13.84** | 0.25 | 14.63 | 1.01 | 13.98 | 0.57 | 14.01 | 0.10 |
| | MAE | **9.69** | 0.22 | 11.07 | 1.49 | 9.86 | 0.53 | 10.07 | 0.03 |
| | score1$\times 10^{-2}$ | 4.27 | 0.47 | 4.12 | 0.39 | 4.39 | 0.47 | **3.93** | 0.15 |
| | score2$\times 10^{-2}$ | – | – | 3.71 | 0.33 | 4.31 | 0.46 | **3.64** | 0.13 |
| FD002 | RMSE | 20.74 | 0.99 | 37.99 | 4.86 | 20.69 | 0.92 | **19.40** | 0.22 |
| | MAE | **14.83** | 0.70 | 33.14 | 4.04 | 14.86 | 0.67 | 15.19 | 0.19 |
| | score1$\times 10^{-2}$ | 194.00 | 143.64 | 370.29 | 230.67 | 204.07 | 150.50 | **28.98** | 2.74 |
| | score2$\times 10^{-2}$ | – | – | 336.46 | 231.58 | 201.20 | 149.81 | **25.52** | 2.26 |
| FD003 | RMSE | 14.41 | 6.00 | 16.11 | 1.69 | **14.18** | 6.06 | 14.90 | 0.27 |
| | MAE | **10.05** | 4.55 | 12.68 | 1.72 | **10.05** | 4.61 | 10.88 | 0.18 |
| | score1$\times 10^{-2}$ | 29.77 | 83.47 | 5.71 | 1.69 | 29.02 | 81.80 | **5.69** | 0.58 |
| | score2$\times 10^{-2}$ | – | – | **4.80** | 1.21 | 28.82 | 81.45 | 5.07 | 0.53 |
| FD004 | RMSE | **22.73** | 0.74 | 39.53 | 3.80 | 22.76 | 0.58 | 23.10 | 0.14 |
| | MAE | **16.46** | 0.39 | 33.88 | 3.03 | 16.58 | 0.34 | 17.92 | 0.10 |
| | score1$\times 10^{-2}$ | 103.76 | 65.21 | 621.22 | 322.91 | 98.17 | 65.03 | **66.95** | 5.57 |
| | score2$\times 10^{-2}$ | – | – | 546.63 | 327.44 | 94.84 | 61.79 | **55.26** | 4.72 |

**Table 5.3:** The root mean squared error (RMSE), mean absolute error (MAE), score1 (i.e. without utilising the uncertainty information) and score2 (i.e. with utilising the uncertainty information) of all D3 models for the test data are shown (BENKER, FURTNER, et al. 2021). Each mean and std was calculated from ten runs with different random seeds. The best results are marked bold.

score without utilisation of uncertainty information (score1) and the score with utilisation of the uncertainty information (score2) are reported. It can be seen that the Bayesian neural networks perform as well as their non-Bayesian counterparts with respect to RMSE and MAE. However, in terms of the costs (i.e. score1 and score2), they perform better. This is even more the case when the uncertainty information is utilised with the proposed approach. As the C2P2 model performed worse than the D3 model, its prediction results are only reported in the publication and are omitted in this thesis.

### 5.3.2 Findings

In publication ③, a novel approach for utilising uncertainty information in RUL estimation was presented and applied to the C-MAPSS benchmark data set. Furthermore, Bayesian neural networks were implemented and compared to their non-Bayesian counterparts with respect to the RMSE, the MAE and the resulting costs. The proposed approach achieved state-of-the-art prediction accuracies and led to a reduction in costs in all cases. Summarising, the findings of publication ③ are the following:

- It could be demonstrated that Bayesian neural networks perform equally well with respect to RUL prediction accuracies compared to their non-Bayesian counterparts, whilst at the same time providing uncertainty information. In the experiments, the Bayesian neural networks trained with VI resulted in more accurate RUL predictions than the ones trained with MCMC.
- The novel approach for utilising the additional information about the uncertainty of the RUL estimate led to lower total costs in every single experiment. Hence, the application of Bayesian neural networks and the utilisation of uncertainty enhance the decision-making process in a predictive maintenance strategy.

Consequently, the objectives defined in **RT3** can be regarded as fully addressed and the goals as achieved.

## 5.4 Publication 4: Data-Efficient Remaining Useful Life Estimation

In the following subsections, the content and findings of publication ④ will be summarised. The overall objective was to present an approach for data-efficient RUL estimation, which can achieve state-of-the-art prediction results without the presence of historic run-to-failure sequences.

### 5.4.1 Summary

Instead of using run-to-failure data, only few measurements at the beginning and the end of the lifetime of the system under investigation were used. In order to compare the proposed method to state-of-the-art approaches which rely on many run-to-failure sequences, the experiments were done with the C-MAPSS (see Appendix C.1) and FEMTO (see Appendix C.2) benchmark data sets. Hence, state-of-the-art results were available with which the novel approach could be compared.

Both data sets were pre-processed first to reflect the targeted situation, in which historic run-to-failure sequences are not available but only a few measurements

at the beginning and end of the entire life cycle are. Hence, each of the training run-to-failure sequences was manipulated in the following way: the first share $p_h$ of the observations and the last share $p_d$ of the observations were cut from the entire sequence and labelled as *healthy* and *degraded*, respectively. For example, in the case of the C-MAPSS data set, the first 7.5 % of the observations of each training run-to-failure sequence were selected and labelled *healthy* and the last 2 % were labelled *degraded*.

The C-MAPSS data set was further pre-processed by standardising the data. In contrast, the FEMTO data set had to be pre-processed more intensively: since the raw data consisted of high-frequency acceleration signals, different features from the time domain, the frequency domain and the time-frequency domain were extracted (see Subsection 2.3.1). The feature extraction techniques were adopted from state-of-the-art approaches originally presented by KIM et al. (2016) and SUTRISNO et al. (2012).

In all experiments, the run-to-failure sequences from only two randomly chosen training instances[4] were used. As outlined above, the first share $p_h$ of observations and the last share $p_d$ of each instance were selected and labelled *healthy* and *degraded*, respectively. As a result, a small training data set with only a few *healthy* and *degraded* observations was available for training.

Based on this training data set, two models were consecutively applied. The purpose of the first one was to construct a health indicator (HI) whereas the purpose of the second one was to extrapolate the HI into the future and derive an RUL value. Both are further described below and the general approach is depicted in Fig. 5.5.

**Health indicator construction** A Gaussian process classification (GPC) model was trained based on the available training data set described above. The model's output was the probability of an input coming from a healthy component. Hence, it computed values close to one for healthy components and values close to zero for degraded components. In other words, the GPC model served as a tool to translate the sensor measurements into an HI value. The main advantages of the GPC model are its data-efficiency (the model could be successfully trained on the basis of only a few observations) and the fact that it outputs bounded values on the interval $[0, 1]$, which enables the second model to predict an RUL value.

[4] In case of the C-MAPSS data set, an instance is a certain turbofan engine. In case of the FEMTO data set, an instance is a certain bearing.

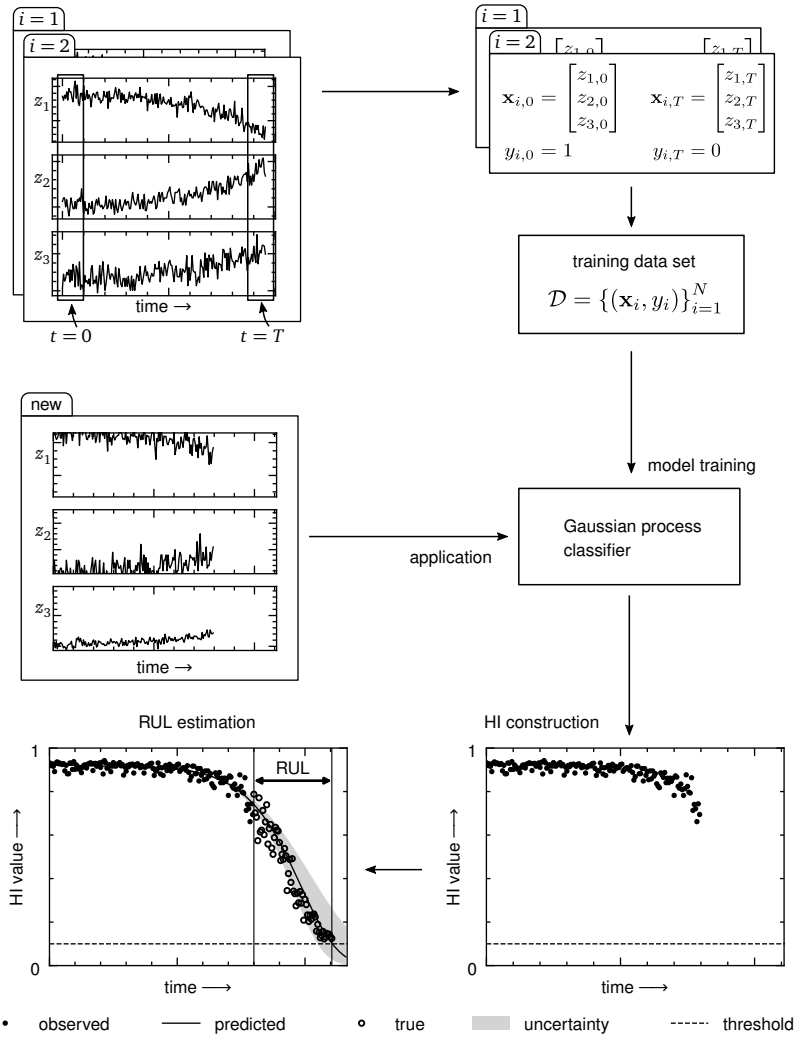



**Figure 5.5:** Illustration of the proposed approach presented in publication ④ based on BENKER, BLIZNYUK, et al. (2021); instead of using the entire run-to-failure sequence, only the first few observations at time $t = 0$, and the last few observations at time $t = T$ were selected for training a Gaussian process classification (GPC) model. During the application of the model, a new sequence of measurements of an unseen component was first translated into a health indicator (HI) value before those HI values were extrapolated with a Gaussian process regression (GPR) model. With the help of a threshold, an RUL value could be derived.

| | C-MAPSS | FEMTO | | |
|---|---|---|---|---|
| | subset | subset | | |
| hyper-parameter | all | C1 | C2 | C3 |
| $p_h$ | 7.5 % | 6.5 % | 7.5 % | 7 % |
| $p_d$ | 2 % | 1.3 % | 1.7 % | 3 % |
| threshold | 0.1 | 0.1 | 0.1 | 0.1 |
| training instances | 2 | 2 | 2 | 2 |

**Table 5.4:** Chosen hyper-parameters for the experiments with the C-MAPSS and FEMTO data sets as in BENKER, BLIZNYUK, et al. (2021)

**Estimation of remaining useful life** A warped, monotonic Gaussian process regression (GPR) model was used for predicting an RUL value. The enforced monotonicity according to RIIHIMÄKI and VEHTARI (2010) led to monotonically falling HI predictions. The conducted warping according to JENSEN et al. (2013) enforced the monotonic prediction to be mapped onto the interval $[0, 1]$, which is the codomain of the HIs. The HI threshold was set to 0.1 and the end-of-life time was defined to be the point in time, at which the prediction fell below this threshold. Finally, the resulting RUL value was computed as the difference between the end-of-life time and the time of the latest measurement.

The chosen hyper-parameters for the different data sets are summarised in Table 5.4.

In some cases, the estimated RUL values were very high (in fact, too high). This is especially the case when the component under investigation is at the beginning of its life cycle. In this situation, the HI does not exhibit a trend yet, and the GPR model simply extrapolates a constant value, resulting in an obviously too large RUL estimate. Therefore, a strategy for limiting the RUL value was introduced and demonstrated for the C-MAPSS data set, where the maximum RUL estimate was set to 125 cycles (where cycles are the time units of the C-MAPSS data set).

The experiments for the C-MAPSS data set were assessed with respect to the RMSE, the MAE and the score (i.e. costs; see Eq. (5.1)). In the case of the FEMTO data set, the relative error

$$\text{error} = 100 \cdot \left( \frac{y - y^\star}{y} \right), \tag{5.3}$$

where $y$ is the true RUL and $y^\star$ is the predicted one, was calculated. The results for the C-MAPSS data set are shown in Table 5.5. Comparing the results to those

| | RMSE | | MAE | | score | |
|---|---|---|---|---|---|---|
| subset | mean | std | mean | std | mean | std |
| FD001 | 32.61 | 219.00 | 16.06 | 1.16 | $1.25 \times 10^2$ | $3.18 \times 10^3$ |
| FD002 | 68.54 | 1434.94 | 35.19 | 4.31 | $7.01 \times 10^4$ | $1.78 \times 10^5$ |
| FD003 | 49.89 | 1013.54 | 24.81 | 5.82 | $3.51 \times 10^3$ | $3.71 \times 10^3$ |
| FD004 | 63.93 | 317.72 | 33.45 | 1.82 | $1.09 \times 10^4$ | $2.70 \times 10^3$ |

**Table 5.5:** Results of the experiments conducted with the C-MAPSS data set as reported in BENKER, BLIZNYUK, et al. (2021); each mean and std was calculated based on ten random experiments.

in Table 5.3 from publication ③, it becomes clear that the RUL prediction results of the data-efficient approach are less accurate in terms of the RMSE. This was to be expected as it does not rely on any run-to-failure sequences at all and uses only a tiny fraction of the original data instead. In fact, for the C-MAPSS data set, less than 1 % of the original data was used. Only the mean score for the subset FD001 outperforms the deep learning (DL) approaches from publication ③. Furthermore, the results vary widely as indicated by the high values of the standard deviations. This reveals a strong sensitivity of the approach to the selection of the training data. This can be explained as follows: the recordings of the C-MAPSS training run-to-failure sequences do not all start at the beginning of an instance's lifetime. Some do, others start somewhere in the middle of the lifetime, and some start towards the end of the lifetime. When two instances (see Footnote 4 on page 86) are chosen randomly, unfavourable selections of the ones with recordings starting at the end of their lifetime cannot be ruled out. In practical applications, however, this can be controlled.

The prediction accuracies for the FEMTO data set are listed in Table 5.6. It can be seen that large prediction errors were produced, which is in line with the benchmark results from SUTRISNO et al. (2012), who won the FEMTO data challenge of the International Conference on Prognostics and Health Management in 2012. This emphasises that the data set is more difficult for RUL prediction in general due to the fact that the extraction of wear-sensitive features for this data set is more complicated. However, the proposed approach is close to the state of the art whilst relying on only a fraction of the original data, which demonstrates its efficiency.

In summary, the proposed data-efficient approach from publication ④ showed good RUL prediction results, although the accuracies were slightly lower than those produced by state-of-the-art approaches, which relied on many run-to-failure sequences. The advantage of the proposed approach, however, is that,

| condition | bearing | error in % | benchmark error in % |
|---|---|---|---|
| C1 | 3 | 83.16 | 37 |
| | 4 | –44.21 | 80 |
| | 5 | 91.99 | 9 |
| | 6 | 91.48 | –5 |
| | 7 | 73.77 | –2 |
| C2 | 3 | 97.21 | 64 |
| | 4 | 10.46 | 10 |
| | 5 | 97.54 | –440 |
| | 6 | 0.87 | 49 |
| | 7 | –98.73 | –317 |
| C3 | 3 | 86.26 | 90 |
| mean absolute error in % | | 70.52 | 100.27 |

**Table 5.6:** Results of the experiments for the FEMTO data set as reported in BENKER, BLIZNYUK, et al. (2021); the benchmark results are taken from SUTRISNO et al. (2012).

in industrial use cases, measurements in different degradation states can be acquired more easily than run-to-failure sequences. This can be done, for example, by targeted replacements of degraded components and making measurements (see publication ②).

### 5.4.2 Findings

In publication ④, a novel approach for data-efficient RUL estimation was presented together with its application to the C-MAPSS and FEMTO benchmark data sets. The proposed approach achieved good prediction accuracies close to the state-of-the-art ones without relying on run-to-failure data. This is especially useful for industrial applications, where run-to-failure data are usually not available. Hence, the findings of publication ④ are as follows:

- A novel data-efficient method for the estimation of RUL values was presented. The method does not rely on a single historic run-to-failure sequence, strongly enabling the application in industrial use cases. The method only relies on a few observations of a component under investigation in a healthy and degraded state. The former can be recorded with a new component. The latter can be recorded by artificially inducing wear.

- The proposed approach led to good prediction accuracies compared to the state-of-the-art results. Although the predictions were not as accurate due to the fact that only a tiny fraction of the data was used, the pre-

diction accuracies were still within the range reported by other authors. Nevertheless, the absence of run-to-failure data led to a systematic loss in prediction accuracy.

- It became visible, that the proposed approach is sensitive to the presence of failure-sensitive features. In the case of the FEMTO data set, where failure-sensitive features were only partially available, this led to low prediction accuracies.

In conclusion, **RT4** was only partially addressed, as the novel approach resulted in a loss of prediction accuracy. The usefulness of the novel method for practical applications is indisputable, however, such that **RT4** can be regarded as largely addressed.

Chapter 6

# Analysis of the Economic Potential of the Research Results

This chapter is concerned with the investigation of the economic potential of the research results from Chapter 5. The economic potential of a predictive maintenance (PdM) strategy, in contrast to a planned and reactive maintenance strategy, was quantified with the help of a discrete event simulation (DES) model as described in publication ⑤. The following Section 6.1 will outline the content of the publication and Section 6.2 will give a summary on the economic potential of the research results.

## 6.1 Publication 5: Economic Potential

In order to assess the economic potential of different maintenance strategies, a job shop operated in single-shift operation (8 h) with three machines and three product variants was simulated. Each product variant was assumed to have different processing times on each machine and having to pass a different machine sequence. The three machines were assumed to be in one of the following states:

1. **Running:** in this state, a machine is producing as planned.
2. **Maintenance:** a planned maintenance action is conducted in this state.
3. **Repairing:** the machine unexpectedly fails and has to be repaired.
4. **Idle:** a machine is in idle state when it is ready for production and is waiting for the next job.

Furthermore, the following three maintenance strategies were implemented and compared:

**Reactive maintenance strategy** This is the least complex maintenance strategy implemented as there is no maintenance action planned at all. The machines unexpectedly fail and are subsequently repaired. Hence, when this maintenance strategy was simulated, the state *maintenance* was never taken by any of the machines.

**Planned maintenance strategy** Maintenance actions for each machine are planned based on a fixed schedule (e.g. every 500 h). The fixed time for conducting a maintenance action had to be set before starting the simulation. When the planned maintenance action for a machine is scheduled before the true failure, it goes into the state *maintenance* for 10 h. Otherwise, it unexpectedly fails and goes into the state *repairing* for 20 h.

**Predictive maintenance strategy** The implemented PdM strategy consists of two steps: first, a remaining useful life (RUL) estimate $\hat{t}_{RUL}$ is sampled from a normal distribution with a mean equal to the true RUL value $t_{RUL}$ and a standard deviation $\sigma$. The latter is also referred to as the uncertainty of the RUL estimation. Hence, the RUL estimate $\hat{t}_{RUL}$ follows a normal distribution:

$$\hat{t}_{RUL} \sim \mathcal{N}\left(t_{RUL}, \sigma^2\right). \tag{6.1}$$

In other words, it is assumed that the expected values of the RUL estimate $\hat{t}_{RUL}$ is equal to the true RUL value $t_{RUL}$. As the RUL estimates are uncertain, a security factor $c_s \in ]0,1]$, by which the RUL estimates are multiplied, is introduced in the second step. The security factor can be regarded as a user's degree of risk averseness. The lower $c_s$, the more risk averse a user is. The result of this multiplication is used as the time for the next planned maintenance action. If that happens before the true machine failure, the machine goes into the state *maintenance* for 10 h. Otherwise, it unexpectedly fails and goes into the state *repairing* for 20 h.

At the beginning of each week of the simulation, a certain fictitious customer demand was generated such that approximately an entire work week, which lasted for 40 h, was filled with production orders. Together with the planned maintenance actions, a scheduling algorithm determined an optimal maintenance integrated production schedule with respect to minimal makespan for

| strategy | state | time in h |
|---|---|---|
| reactive | running | 7833.47 |
| | maintenance | 0.00 |
| | repairing | 272.23 |
| | idle | 3894.30 |
| | total | 12 000 |
| planned | running | 7902.64 |
| | maintenance | 197.51 |
| | repairing | 149.43 |
| | idle | 3750.42 |
| | total | 12 000 |
| predictive | running | 8236.77 |
| | maintenance | 160.85 |
| | repairing | 0.68 |
| | idle | 3601.70 |
| | total | 12 000 |

**Table 6.1:** Results of all experiments as reported in BENKER, ROMMEL, et al. (2022); the mean values of 500 runs are reported. The results of the planned maintenance strategy were generated with a fixed maintenance action every 500 h. The reported results of the PdM strategy were generated with $c_s = 0.95$ and $\sigma = 0$.

the upcoming week. It should be noted that the times for maintenance actions determined by the planned and PdM strategies were allowed to be moved forward to an earlier time by the scheduler if this reduced the total makespan. The scheduler was not allowed, however, to postpone a maintenance action to a later time. In each simulation run, 100 weeks were simulated and the event time step of the model was set to one hour. Machine failures were sampled from a Weibull distribution (see Appendix A.4) with $\alpha = 919.5$ and $\beta = 1.9$, as this can be assumed to be a representative failure distribution of machine tools (DAI et al. 2003; YANG et al. 2015).

In order to account for statistical fluctuations, each experiment was conducted 500 times. The results for all three maintenance strategies are shown in Table 6.1 and are reported as mean values calculated from those 500 repetitions. It can be seen that the PdM strategy led to the highest value of total running time of all three machines. In fact, the running time is 4 % higher than the one of the planned maintenance strategy and 5 % higher than the one of the reactive maintenance strategy. This amounts to 2.8 weeks and 3.4 weeks more available production capacity within the simulated time, respectively.

However, the results are sensitive towards the choice of maintenance and repairing times. The previous results assumed a maintenance action to last for 10 h and a repair action to take 20 h. With a change in the ratio between the two, the

results are expected to change as well. Furthermore, the question arises of how the permissible range of viable combinations of $c_s$ and $\sigma$ evolves with the different assumptions. In other words, it is unclear whether a PdM is still economically beneficial if the uncertainty of the RUL predictions and the repair times after an unexpected failure are high. This is depicted and investigated in Fig. 6.1, where it can be seen that, with higher repairing times, the area of viable combinations of the security factor $c_s$ and the RUL prediction uncertainty $\sigma$ decreases. This is due to the fact that inaccurate RUL estimates led to late planned maintenance actions, which in turn led to expensive, unexpected failures in the simulation. It can also be seen, however, that the security factor $c_s$ is a parameter which gives a user flexibility in coping with more uncertain RUL predictions. The less accurate they become (i.e. the higher $\sigma$ is), the lower is an optimal $c_s$. In other words, even when the RUL predictions are highly uncertain and the repair times after an unexpected failure are high, a user can viably operate a PdM strategy by setting an appropriate security factor.

## 6.2 Summary of the Economic Potential of the Research Results

Publication ⑤ showed that a PdM strategy is economically superior to other, more traditional maintenance strategies (i.e. reactive and planned maintenance). This is even the case when the RUL estimates, upon which a PdM strategy is based, are highly uncertain and deviate from the true RUL values. As a user can account for high uncertainties by being more risk averse, the effect of inaccurate RUL predictions can be attenuated.

Summarising, in a small fictitious job shop with only three machines, the simulated application of a PdM strategy led to an increase in productivity of up to 5 % (compared to a reactive maintenance strategy), which was equal to 403.3 h or 3.4 weeks of single-shift production. The latter can be expected to be even higher for more complex production sites. Hence, it becomes clear that if the research results of this thesis are applied in industry in the context of a PdM strategy, considerable economic advantages can be exploited.

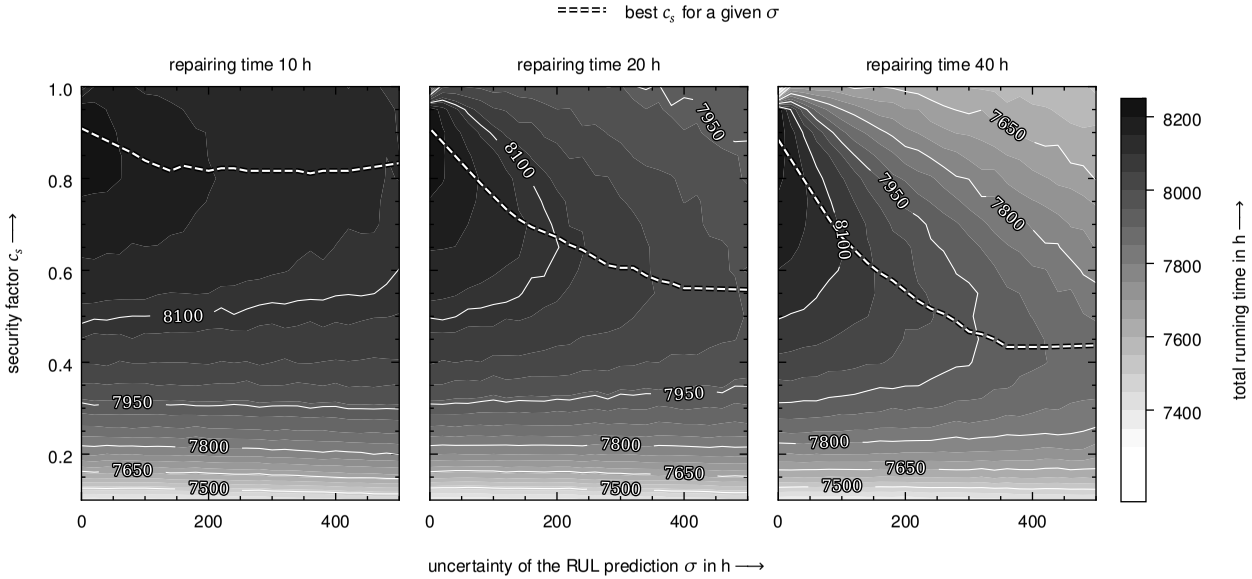



**Figure 6.1:** Investigation into the sensitivity of the economic viability of a PdM strategy based on BENKER, ROMMEL, et al. (2022); it can be seen that with increasing repairing times, the area of viable combinations of $c_s$ and $\sigma$ decrease. Furthermore, it can be seen that the optimal security factor $c_s$ decreases with an increasing RUL prediction uncertainty $\sigma$. The shrinking of the area of viable combinations is faster when the repairing times are higher. The time for planned maintenance actions was set to be 10 h in all experiments.

Chapter 7

# Conclusions

In order to continue to secure prosperity in industrial societies, the efficiency of production processes must be continuously improved. This includes an increase of the utilisation of production resources. Since machine tools are the backbone of industrial manufacturing (SCHWENKE et al. 2009), ensuring their high utilisation is especially important. With ongoing digitisation in factories, the application of condition monitoring and novel maintenance strategies, such as predictive maintenance, will become increasingly feasible.

Nevertheless, it can be stated that a widespread use of condition monitoring for machine tools and predictive maintenance strategies has not yet been observed (BUTLER et al. 2022). In part, this is due to gaps in the state of the art, which still hinder an application today.

This thesis aimed at contributing to the literature and the state of the art of condition monitoring for machine tool feed drives and methods for the estimation of remaining useful life (RUL). First, this was done by thoroughly reviewing the state of the art in Chapter 3 and subsequently identifying research gaps (RGs). Based on these, suitable research targets (RTs) were formulated in Chapter 4. The fulfilment of the RTs was the objective of four publications (see Appendix E). They were outlined and summarised in Chapter 5 with the help of all findings for each publication. The summary included a critical comparison to the established RTs. Finally, the economic potential of the findings was investigated in a fifth publication, which was discussed up in Chapter 6.

In the following Section 7.1, the key findings (KFs) of this thesis will be briefly outlined and an outlook will be given in Section 7.2.

## 7.1 Key Findings

The contributions and findings of publications ① – ⑤ can be further summarised as five KFs (see Fig. 4.1), which are the essential contributions of this thesis and which are given below:

**KF1** In publication ①, a novel condition monitoring test cycle was derived, which enables the recording of reproducible condition monitoring data from industrial machine tools. Furthermore, it was shown that the test cycle is robust against environmental disturbing factors.

**KF2** In publication ②, it was demonstrated that the degradation of ball screws can be assessed with a purely data-driven approach when the data are recorded with the test cycle from publication ①. Notably, the proposed approach could accurately assess wear of previously unseen ball screws.

**KF3** In publication ③, a general issue connected to RUL prediction was addressed: as most state-of-the-art methods produce a single value for the RUL (i.e. a point estimate), users might be reluctant to base a critical decision, such as the time for the next maintenance action, on such a prediction. In publication ③, it was demonstrated how to not only predict a single RUL value, but also how to quantify the uncertainty of this prediction. Furthermore, a novel approach was presented, which utilises this uncertainty information for determining a better and more reliable end-of-life time, and which ultimately enhances the decision for scheduling the next maintenance action.

**KF4** In publication ④, a specific but especially restrictive requirement from industrial use cases was picked up: usually, historic run-to-failure data are not available. Nevertheless, many state-of-the-art methods for the estimation of RUL values assume the exact opposite – the presence of large run-to-failure data sets. In order to resolve this discrepancy, publication ④ proposed a novel approach for the estimation of RUL values, which is not based on any run-to-failure data. Although the novel approach came with a loss in accuracy, the prediction accuracies were still within the range of those reported in the state of the art based on conventional methods.

**KF5** Finally, publication ⑤ demonstrated how to build a discrete event simulation model of a job shop with different maintenance strategies. Notably,

it was shown that a predictive maintenance strategy is viable, even in cases where the predictions are highly uncertain. This is because a user can introduce a security factor by which the RUL prediction is reduced and which subsequently decreases the risk of unexpected machine failures. In this way, job shops operated with a predictive maintenance strategy become economically superior with respect to maintenance costs compared to those operated with a reactive or planned maintenance strategy.

In summary, it can be stated that the thesis significantly contributed to the state of the art with the above-mentioned KFs. Nevertheless, there are still open issues regarding condition monitoring of machine tool feed drives and the estimation of the RUL. The next section provides an outlook into possible future research directions.

## 7.2 Outlook

From a methodological perspective, the review papers already mentioned in Section 3.2 (e.g. FINK et al. (2020)) are a good guidance for future research directions. One of the recurring issues with data-driven approaches is their sensitivity towards a *domain shift*. A data-driven model trained on one machine is not necessarily valid on another, similar machine. The other machine could be operated differently or have slightly varying components assembled. Handling such changes with domain adaptation models is a promising research direction with the potential to significantly contribute to the task of condition monitoring for any application and particularly for machine tool feed drives.

Another promising future research direction is fusing flexible machine learning (ML) models such as artificial neural networks (ANNs) or Gaussian process regressions (GPRs) models with physical models. As the physical models are usually a very good representation of the system under investigation, this approach could reduce the number of historic observations required and enhance the robustness of the ML model predictions. Especially in the research field of machine tools, a deep physical understanding of the machine tool system has already been established. Combining this knowledge with ML models should be explored in future work.

Finally, from an exploitation perspective, higher degrees of digitisation will be needed in the future to enable a wide-spread application of condition monitoring and predictive maintenance systems. Industrial software such as numeric control software of machine tools is often restrictive and exacerbates the extraction of data and the integration of novel algorithms, such as the ones presented in this thesis. Therefore, future work should also focus on further developing and establishing smart system architectures and dedicated industrial standards. One recent example showing how such an intelligent system architecture can be developed and what this architecture could look like is given in a study by SCHMUCKER et al. (2021). In conclusion, this thesis contributed to the state of the art and laid the foundation for future research in the directions outlined above.

# Bibliography

The references are sorted alphabetically according to the names of the first authors, and the volume and the issue of an article are separated with the period symbol. For example, *CIRP Annals* 60.2 refers to issue 2 of volume 60 of the CIRP Annals journal. In cases, where the issue or the digital object identifier (DOI) is not indicated, they do not exist or are not available to the author of this thesis.

AGHION, P., JONES, B., and JONES, C., (2017). *Artificial Intelligence and Economic Growth*. Tech. rep. w23928. Cambridge, MA: National Bureau of Economic Research, w23928. DOI: 10.3386/w23928.

ALTINTAS, Y., VERL, A., BRECHER, C., URIARTE, L., and PRITSCHOW, G., (2011). "Machine Tool Feed Drives". In: *CIRP Annals* 60.2, pp. 779–796. DOI: 10.1016/j.cirp.2011.05.010.

ALTINTAS, Y., (2012). *Manufacturing Automation: Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design*. 2nd ed. Cambridge, UK ; New York, NY: Cambridge University Press. ISBN: 978-0-521-17247-9.

ANDRIEU, C., DE FREITAS, N., DOUCET, A., and JORDAN, M. I., (2003). "An Introduction to MCMC for Machine Learning". In: *Machine Learning* 50, pp. 5–43. DOI: 10.1023/A:1020281327116.

ANNONI, A., BENCZUR, P., BERTOLDI, P., DELIPETREV, P., DE PRATO, G., FEIJOO, C., FERNANDEZ MACIAS, E., GOMEZ, E., IGLESIAS, M., JUNKLEWITZ, H., LÓPEZ COBO, M., MARTENS, B., NASCIMENTO, S., NATIVI, S., POLVORA, A., SANCHEZ, I., TOLAN, S., TUOMI, I., and VESNIC ALUJEVIC, L., (2018). *Artificial Intelligence: A European Perspective*. Ed. by CRAGLIA, M. European Commission / Joint Research Centre 29425 EN. Luxembourg: Publications Office of the European Union. ISBN: 978-92-79-97219-5.

Arias Chao, M., Kulkarni, C., Goebel, K., and Fink, O., (2021). "Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics". In: *Data* 6.1, p. 5. DOI: 10.3390/data6010005.

Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., and Zerhouni, N., (2020). "Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation". In: *International Journal of Prognostics and Health Management* 8.3. DOI: 10.36001/ijphm.2017.v8i3.2667.

Baur, M., Albertelli, P., and Monno, M., (2020). "A Review of Prognostics and Health Management of Machine Tools". In: *The International Journal of Advanced Manufacturing Technology* 107.5-6, pp. 2843–2863. DOI: 10.1007/s00170-020-05202-3.

Bekar, E. T., Nyqvist, P., and Skoogh, A., (2020). "An Intelligent Approach for Data Pre-Processing and Analysis in Predictive Maintenance with an Industrial Case Study". In: *Advances in Mechanical Engineering* 12.5. DOI: 10.1177/1687814020919207.

Benker, M., Bliznyuk, A., and Zaeh, M. F., (2021). "A Gaussian Process Based Method for Data-Efficient Remaining Useful Life Estimation". In: *IEEE Access* 9, pp. 137470–137482. DOI: 10.1109/ACCESS.2021.3116813.

Benker, M., Furtner, L., Semm, T., and Zaeh, M. F., (2021). "Utilizing Uncertainty Information in Remaining Useful Life Estimation via Bayesian Neural Networks and Hamiltonian Monte Carlo". In: *Journal of Manufacturing Systems* 61, pp. 799–807. DOI: 10.1016/j.jmsy.2020.11.005.

Benker, M., Junker, S., Ellinger, J., Semm, T., and Zaeh, M. F., (2022). "Experimental Derivation of a Condition Monitoring Test Cycle for Machine Tool Feed Drives". In: *Production Engineering* 16.1, pp. 55–64. DOI: 10.1007/s11740-021-01085-9.

Benker, M., Rommel, V., and Zaeh, M. F., (2022). "An Investigation into the Economic Efficiency of Different Maintenance Strategies Based on a Discrete Event Simulation". In: *Procedia CIRP* 107, pp. 428–433. DOI: 10.1016/j.procir.2022.05.003.

BENKER, M. and ZAEH, M. F., (2022). "Condition Monitoring of Ball Screw Feed Drives Using Convolutional Neural Networks". In: *CIRP Annals* 71.1, pp. 313–316. DOI: 10.1016/j.cirp.2022.03.017.

BIGGIO, L. and KASTANIS, I., (2020). "Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead". In: *Frontiers in Artificial Intelligence* 3, p. 578613. DOI: 10.3389/frai.2020.578613.

BIGGIO, L., WIELAND, A., CHAO, M. A., KASTANIS, I., and FINK, O., (2021). "Uncertainty-Aware Prognosis via Deep Gaussian Process". In: *IEEE Access* 9, pp. 123517–123527. DOI: 10.1109/ACCESS.2021.3110049.

BISHOP, C. M., (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. New York, NY: Springer. ISBN: 978-0-387-31073-2.

BLEI, D. M., KUCUKELBIR, A., and MCAULIFFE, J. D., (2017). "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: 10.1080/01621459.2017.1285773.

BOSER, B. E., GUYON, I. M., and VAPNIK, V. N., (1992). "A Training Algorithm for Optimal Margin Classifiers". In: *The Fifth Annual Workshop on Computational Learning Theory*. Pittsburgh, Pennsylvania, United States: ACM Press, pp. 144–152. DOI: 10.1145/130385.130401.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., and LJUNG, G. M., (2016). *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley Series in Probabilit and Statistics. Hoboken, NJ: John Wiley & Sons, Inc. ISBN: 978-1-118-67492-5.

BUTLER, Q., ZIADA, Y., STEPHENSON, D., and ANDREW GADSDEN, S., (2022). "Condition Monitoring of Machine Tool Feed Drives: A Review". In: *Journal of Manufacturing Science and Engineering* 144.10. 100802. DOI: 10.1115/1.4054516.

CHRIST, M., BRAUN, N., NEUFFER, J., and KEMPA-LIEHR, A. W., (2018). "Time Series FeatuRe Extraction on Basis of Scalable Hypothesis Tests (Tsfresh – A Python Package)". In: *Neurocomputing* 307, pp. 72–77. DOI: 10.1016/j.neucom.2018.03.067.

COBB, A. D. and JALAIAN, B., (2021). "Scaling Hamiltonian Monte Carlo Inference for Bayesian Neural Networks with Symmetric Splitting". In: *Proceedings of Machine Learning Research*. Vol. 161. Online Conference, pp. 675–685.

Cox, D. R., (1972). “Regression Models and Life-Tables”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2, pp. 187–202.

Dai, Y., Zhou, Y.-f., and Jia, Y.-z., (2003). “Distribution of Time between Failures of Machining Center Based on Type I Censored Data”. In: *Reliability Engineering & System Safety* 79.3, pp. 377–379. doi: 10.1016/S0951-8320(02)00243-0.

Damianou, A. and Lawrence, N. D., (2013). “Deep Gaussian Processes”. In: *International Conference on Artificial Intelligence and Statistics*. Vol. 31. Scottsdale, AZ: PMLR, pp. 207–215.

Denkena, B., Bergmann, B., and Schmidt, A., (2021). “Preload Monitoring of Single Nut Ball Screws Based on Sensor Fusion”. In: *CIRP Journal of Manufacturing Science and Technology* 33, pp. 63–70. doi: 10.1016/j.cirpj.2021.02.006.

Denkena, B., Guemmer, O., and Floeter, F., (2014). “Evaluation of Electromagnetic Guides in Machine Tools”. In: *CIRP Annals* 63.1, pp. 357–360. doi: 10.1016/j.cirp.2014.03.130.

Deutsche Bundesbank, (2021). “The Slowdown in Euro Area Productivity Growth”. In: *Monthly Report* 73.1, pp. 15–46.

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D., (1987). “Hybrid Monte Carlo”. In: *Physics Letters B* 195.2, pp. 216–222. doi: 10.1016/0370-2693(87)91197-X.

Duda, R. O., Hart, P. E., and Stork, D. G., (2001). *Pattern Classification*. 2nd ed. New York, NY: Wiley. isbn: 978-0-471-05669-0.

Duvenaud, D. K., (2014). “Automatic Model Construction with Gaussian Processes”. PhD thesis. University of Cambridge.

Ellinger, J., Semm, T., Benker, M., Kapfinger, P., Kleinwort, R., and Zah, M., (2019). “Feed Drive Condition Monitoring Using Modal Parameters”. In: *MM Science Journal* 2019.04, pp. 3206–3213. doi: 10.17973/MMSJ.2019_11_2019072.

Esposito, P., (2020). *BLiTZ - Bayesian Layers in Torch Zoo (a Bayesian Deep Learing Library for Torch)*. Available at http://github.com/piEsposito/blitz-bayesian-deep-learning (2022/08/31).

EUROPEAN COMMISSION, (2018). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe*. Communication COM(2018) 237. Brussels.

EWINS, D. J., (2000). *Modal Testing: Theory, Practice, and Application*. 2nd ed. Mechanical Engineering Research Studies 10. Baldock, UK ; Philadelphia, PA: Research Studies Press. ISBN: 978-0-86380-218-8.

FARRAR, C. R. and WORDEN, K., (2012). *Structural Health Monitoring: A Machine Learning Perspective*. Chichester, UK: John Wiley & Sons, Ltd. ISBN: 978-1-119-99433-6.

FINK, O., WANG, Q., SVENSÉN, M., DERSIN, P., LEE, W.-J., and DUCOFFE, M., (2020). "Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications". In: *Engineering Applications of Artificial Intelligence* 92. DOI: 10.1016/j.engappai.2020.103678.

FULCHER, B. D., (2018). "Feature-Based Time-Series Analysis". In: *Feature Engineering for Machine Learning and Data Analytics*. Boca Raton, FL: CRC Press, pp. 87–116. ISBN: 978-1-315-18108-0.

GELMAN, A., (2014). *Bayesian Data Analysis*. 3rd ed. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton, FL: CRC Press. ISBN: 978-1-4398-4095-5.

GÉRON, A., (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd ed. Beijing ; Sebastopol, CA: O'Reilly Media, Inc. ISBN: 978-1-4920-3264-9.

GHAHRAMANI, Z., (2015). "Probabilistic Machine Learning and Artificial Intelligence". In: *Nature* 521.7553, pp. 452–459. DOI: 10.1038/nature14541.

GOODFELLOW, I., BENGIO, Y., and COURVILLE, A., (2016). *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press. ISBN: 978-0-262-03561-3.

GUILLAUME, P., VERBOVEN, P., VANLANDUIT, S., VAN DER AUWERAER, H., and PEETERS, B., (2003). "A Poly-Reference Implementation of the Least-Squares Complex Frequency-Domain Estimator". In: *21st IMAC Conference and Exposition 2003*. Society for Experimental Mechanics, pp. 183–192.

GUYON, I. and ELISSEEFF, A., (2006). "An Introduction to Feature Extraction". In: *Feature Extraction*. Ed. by GUYON, I., NIKRAVESH, M., GUNN, S., and ZADEH, L. A. Vol. 207. Berlin, Heidelberg: Springer, pp. 1–25. ISBN: 978-3-540-35487-1.

HABERKERN, A., (1998). "Leistungsfähigere Kugelgewindetriebe durch Beschichtung". PhD thesis. Universität Karlsruhe (TH).

HAMILTON, J. D., (1994). *Time Series Analysis*. Princeton, N.J: Princeton University Press. ISBN: 978-0-691-04289-3.

HASTIE, T., TIBSHIRANI, R., and FRIEDMAN, J. H., (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York, NY: Springer. ISBN: 978-0-387-84857-0.

HASTINGS, W. K., (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, pp. 97–109. DOI: 10.1093/biomet/57.1.97.

HE, J. and FU, Z.-F., (2001). *Modal Analysis*. Oxford, UK ; Boston MA: Butterworth-Heinemann. ISBN: 978-0-7506-5079-3.

HESS, A. and FILA, L., (2002). "The Joint Strike Fighter (JSF) PHM Concept: Potential Impact on Aging Aircraft Problems". In: *Proceedings, IEEE Aerospace Conference*. Vol. 6. Big Sky, MT: IEEE, pp. 6-3021–6-3026. DOI: 10.1109/AERO.2002.1036144.

HOFFMAN, M. D. and GELMAN, A., (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research* 15, pp. 1593–1623.

IMIELA, J., (2006). *Verfügbarkeitssicherung von Werkzeugmaschinenachsen mit Kugelgewindetrieb durch modellbasierte Verschleißüberwachung*. Berichte aus dem IFW / Institut für Fertigungstechnik und Werkzeugmaschinen 01/2006. Garbsen: PZH, Produktionstechn. Zentrum. ISBN: 978-3-939026-04-4.

ISO 12090-1, (2011). *Rolling bearings – Profiled rail guides for linear motion – Part 1: Boundary dimensions and tolerances for series 1, 2 and 3*. Standard. Geneva: International Organization for Standardization.

ISO 12090-2, (2011). *Rolling bearings – Profiled rail guides for linear motion – Part 2: Boundary dimensions and tolerances for series 4 and 5*. Standard. Geneva: International Organization for Standardization.

ISO 13379-1, (2012). *Condition monitoring and diagnostics of machines — Data interpretation and diagnostics techniques – Part 1: General guidelines*. Standard. Geneva: International Organization for Standardization.

ISO 281, (2007). *Rolling bearings – Dynamic load ratings and rating life*. Standard. Geneva: International Organization for Standardization.

ISO 3408-2, (2021). *Ball screws – Part 2: Nominal diameters, leads, nut dimensions and mounting bolts — Metric series*. Standard. Geneva: International Organization for Standardization.

ISO 3408-3, (2006). *Ball screws – Part 3: Acceptance conditions and acceptance tests*. Standard. Geneva: International Organization for Standardization.

ISO 3408-4, (2006). *Ball screws – Part 4: Static axial rigidity*. Standard. Geneva: International Organization for Standardization.

ISO 3408-5, (2006). *Ball screws – Part 5: Static and dynamic axial load ratings and operational life*. Standard. Geneva: International Organization for Standardization.

JARDINE, A. K., LIN, D., and BANJEVIC, D., (2006). "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance". In: *Mechanical Systems and Signal Processing* 20.7, pp. 1483–1510. DOI: 10.1016/j.ymssp.2005.09.012.

JENSEN, B. S., NIELSEN, J. B., and LARSEN, J., (2013). "Bounded Gaussian Process Regression". In: *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. Southampton, UK: IEEE, pp. 1–6. DOI: 10.1109/MLSP.2013.6661916.

JIA, P., RONG, Y., and HUANG, Y., (2019). "Condition Monitoring of the Feed Drive System of a Machine Tool Based on Long-Term Operational Modal Analysis". In: *International Journal of Machine Tools and Manufacture* 146, p. 103454. DOI: 10.1016/j.ijmachtools.2019.103454.

JOANES, D. N. and GILL, C. A., (1998). "Comparing measures of sample skewness and kurtosis". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.1, pp. 183–189. DOI: 10.1111/1467-9884.00122.

KIM, M. S., YUN, J. P., and PARK, P., (2021). "An Explainable Convolutional Neural Network for Fault Diagnosis in Linear Motion Guide". In: *IEEE Transactions on Industrial Informatics* 17.6, pp. 4036–4045. DOI: 10.1109/TII.2020.3012989.

KIM, N.-H., AN, D., and CHOI, J.-H., (2017). *Prognostics and Health Management of Engineering Systems*. Cham: Springer International Publishing. ISBN: 978-3-319-44740-7.

KIM, S., PARK, S., KIM, J.-W., HAN, J., AN, D., KIM, N. H., and CHOI, J.-H., (2016). "A New Prognostics Approach for Bearing Based on Entropy Decrease and Comparison with Existing Methods". In: *Annual Conference of the PHM Society* 8.1. DOI: 10.36001/phmconf.2016.v8i1.2540.

KIMOTHO, J. K. and SEXTRO, W., (2014). "An Approach for Feature Extraction and Selection from Non-Trending Data for Machinery Prognosis". In: *Proceedings of the European Conference of the PHM Society 2014*. Nantes: PHM Society, p. 8. DOI: 10.36001/phme.2014.v2i1.1462.

KINGMA, D. P. and BA, J., (2017). "Adam: A Method for Stochastic Optimization". In: *arXiv:1412.6980 [cs]*. arXiv: 1412.6980 [cs].

KRAUS, M. and FEUERRIEGEL, S., (2019). "Forecasting Remaining Useful Life: Interpretable Deep Learning Approach via Variational Bayesian Inferences". In: *Decision Support Systems* 125, p. 113100. DOI: 10.1016/j.dss.2019.113100.

KRÜGER, J., FLEISCHER, J., FRANKE, J., and GROCHE, P., (2019). *WGP-Standpunkt: KI in Der Produktion*. Tech. rep. Hannover: WGP Wissenschaftliche Gesellschaft für Produktionstechnik e.V.

KRÜGER, R., BERGMANN, B., and DENKENA, B., (2022). "Electromagnetic Ultra-Precision Linear Guide". In: *Ultra-Precision High Performance Cutting*. Ed. by BRINKSMEIER, E. and SCHÖNEMANN, L. Cham: Springer International Publishing, pp. 75–106. ISBN: 978-3-030-83764-8.

KUSS, M. and RASMUSSEN, C. E., (2005). "Assessing Approximate Inference for Binary Gaussian Process Classification". In: *Journal of Machine Learning Research* 6.57, pp. 1679–1704.

LAMPINEN, J. and VEHTARI, A., (2001). "Bayesian Approach for Neural Networks - Review and Case Studies". In: *Neural Networks* 14.3, pp. 257–274. DOI: 10.1016/S0893-6080(00)00098-8.

LAU, B. C. P., MA, E. W. M., and PECHT, M., (2012). "Review of Offshore Wind Turbine Failures and Fault Prognostic Methods". In: *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference*. Beijing: IEEE, pp. 1–5. DOI: 10.1109/PHM.2012.6228954.

LAWLESS, J. F., (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed. Wiley Series in Probability and Statistics. Hoboken, N.J: Wiley-Interscience. ISBN: 978-0-471-37215-8.

LECUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W., and JACKEL, L. D., (1989). "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1.4, pp. 541–551. DOI: 10.1162/neco.1989.1.4.541.

LECUN, Y., BENGIO, Y., and HINTON, G., (2015). "Deep Learning". In: *Nature* 521.7553, pp. 436–444. DOI: 10.1038/nature14539.

LEE, J., WU, F., ZHAO, W., GHAFFARI, M., LIAO, L., and SIEGEL, D., (2014). "Prognostics and Health Management Design for Rotary Machinery Systems—Reviews, Methodology and Applications". In: *Mechanical Systems and Signal Processing* 42.1-2, pp. 314–334. DOI: 10.1016/j.ymssp.2013.06.004.

LEE, K.-J. and DONMEZ, M. A., (2007). "Repeatability Analysis on the Tool Point Dynamics for Investigation on Uncertainty in Milling Stability". In: *ASME International Mechanical Engineering Congress and Exposition*. Seattle, WA, pp. 1–7.

LEI, Y., LI, N., GUO, L., LI, N., YAN, T., and LIN, J., (2018). "Machinery Health Prognostics: A Systematic Review from Data Acquisition to RUL Prediction". In: *Mechanical Systems and Signal Processing* 104, pp. 799–834. DOI: 10.1016/j.ymssp.2017.11.016.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A. K., (2020). "Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap". In: *Mechanical Systems and Signal Processing* 138, p. 106587. doi: 10.1016/j.ymssp.2019.106587.

Lei, Y., Zuo, M. J., He, Z., and Zi, Y., (2010). "A Multidimensional Hybrid Intelligent Method for Gear Fault Diagnosis". In: *Expert Systems with Applications* 37.2, pp. 1419–1430. doi: 10.1016/j.eswa.2009.06.060.

Li, G., Yang, L., Lee, C.-G., Wang, X., and Rong, M., (2020). "A Bayesian Deep Learning RUL Framework Integrating Epistemic and Aleatoric Uncertainties". In: *IEEE Transactions on Industrial Electronics*. doi: 10.1109/TIE.2020.3009593.

Li, P., Jia, X., Feng, J., Davari, H., Qiao, G., Hwang, Y., and Lee, J., (2018). "Prognosability Study of Ball Screw Degradation Using Systematic Methodology". In: *Mechanical Systems and Signal Processing* 109, pp. 45–57. doi: 10.1016/j.ymssp.2018.02.046.

Li, X., Ding, Q., and Sun, J.-Q., (2018). "Remaining Useful Life Estimation in Prognostics Using Deep Convolution Neural Networks". In: *Reliability Engineering & System Safety* 172, pp. 1–11.

Liu, F. T., Ting, K. M., and Zhou, Z.-H., (2008). "Isolation Forest". In: *2008 Eighth IEEE International Conference on Data Mining*. Pisa, Italy: IEEE, pp. 413–422. doi: 10.1109/ICDM.2008.17.

Ludema, K. C., (1996). *Friction, Wear, Lubrication: A Textbook in Tribology*. Boca Raton, FL: CRC Press. isbn: 978-0-8493-2685-1.

Lv, H., Chen, J., and Pan, T., (2020). "Sequence Adaptation Adversarial Network for Remaining Useful Life Prediction Using Small Data Set". In: *IEEE 18th International Conference on Industrial Informatics*. Warwick, UK: IEEE, pp. 115–118. doi: 10.1109/INDIN45582.2020.9442160.

Maas, A. L., Hannun, A. Y., and Ng, A. Y., (2013). "Rectifier Nonlinearities Improve Neural Network Acoustic Models". In: *Proceedings of the 30th International Conference on Machine Learning*. Atlanta, GA.

MacKay, D. J. C., (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK ; New York, NY: Cambridge University Press. isbn: 978-0-521-64298-9.

MAIER, D., (2015). *Sensorlose online Zustandserfassung von Vorschubantriebskomponenten in Werkzeugmaschinen*. Berichte aus dem Institut für Maschinenelemente, Antriebstechnik, CAD, Dichtungen, Zuverlässigkeit 157. Stuttgart: IMA. ISBN: 978-3-936100-58-7.

MARDIA, K. V., KENT, J. T., and BIBBY, J. M., (1979). *Multivariate Analysis*. Probability and Mathematical Statistics. London, UK ; New York, NY: Academic Press. ISBN: 978-0-12-471250-8.

MATE, C. M. and CARPICK, R. W., (2019). *Tribology on the Small Scale: A Modern Textbook on Friction, Lubrication and Wear*. 2nd ed. Oxford Graduate Texts. Oxford, UK: Oxford University Press. ISBN: 978-0-19-960980-2.

MAYR, J., JEDRZEJEWSKI, J., UHLMANN, E., ALKAN DONMEZ, M., KNAPP, W., HÄRTIG, F., WENDT, K., MORIWAKI, T., SHORE, P., SCHMITT, R., BRECHER, C., WÜRZ, T., and WEGENER, K., (2012). "Thermal Issues in Machine Tools". In: *CIRP Annals* 61.2, pp. 771–791. DOI: 10.1016/j.cirp.2012.05.008.

MENG, H. and LI, Y.-F., (2019). "A Review on Prognostics and Health Management (PHM) Methods of Lithium-Ion Batteries". In: *Renewable and Sustainable Energy Reviews* 116, p. 109405. DOI: 10.1016/j.rser.2019.109405.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., and TELLER, E., (1953). "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092. DOI: 10.1063/1.1699114.

MINKA, T. P., (2001). "Expectation Propagation for Approximate Bayesian Inference". In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers Inc., pp. 362–369.

MOHANRAJ, T., SHANKAR, S., RAJASEKAR, R., SAKTHIVEL, N., and PRAMANIK, A., (2020). "Tool Condition Monitoring Techniques in Milling Process – a Review". In: *Journal of Materials Research and Technology* 9.1, pp. 1032–1042. DOI: 10.1016/j.jmrt.2019.10.031.

MÖHRING, H.-C. and BERTRAM, O., (2012). "Integrated Autonomous Monitoring of Ball Screw Drives". In: *CIRP Annals* 61.1, pp. 355–358. DOI: 10.1016/j.cirp.2012.03.138.

NEAL, R. M., (1996). *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics 118. New York, NY: Springer. ISBN: 978-0-387-94724-2.

NECTOUX, P., GOURIVEAU, R., MEDJAHER, K., RAMASSO, E., CHEBEL-MORELLO, B., ZERHOUNI, N., and VARNIER, C., (2012). "PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests." In: *IEEE International Conference on Prognostics and Health Management*. Denver, Colorado, United States: IEEE, pp. 1–8.

NGUYEN, K. T. P. and MEDJAHER, K., (2020). "An Automated Health Indicator Construction Methodology for Prognostics Based on Multi-Criteria Optimization". In: *ISA Transactions*. DOI: 10.1016/j.isatra.2020.03.017.

NGUYEN, T. L., RO, S.-K., and PARK, J.-K., (2019). "Study of Ball Screw System Preload Monitoring during Operation Based on the Motor Current and Screw-Nut Vibration". In: *Mechanical Systems and Signal Processing* 131, pp. 18–32. DOI: 10.1016/j.ymssp.2019.05.036.

OPPENHEIM, A. V., SCHAFER, R. W., and BUCK, J. R., (1999). *Discrete-Time Signal Processing*. 2nd ed. Upper Saddle River, N.J: Prentice Hall. ISBN: 978-0-13-754920-7.

PASZKE, A., GROSS, S., CHINTALA, S., CHANAN, G., YANG, E., DEVITO, Z., LIN, Z., DESMAISON, A., ANTIGA, L., and LERER, A., (2017). "Automatic Differentiation in PyTorch". In: *31st Conference on Neural Information Processing Systems*. Long Beach, CA.

PENG, W., YE, Z.-S., and CHEN, N., (2020). "Bayesian Deep-Learning-Based Health Prognostics Toward Prognostics Uncertainty". In: *IEEE Transactions on Industrial Electronics* 67.3, pp. 2283–2293. DOI: 10.1109/TIE.2019.2907440.

PERES, R. S., JIA, X., LEE, J., SUN, K., COLOMBO, A. W., and BARATA, J., (2020). "Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook". In: *IEEE Access* 8, pp. 220121–220139. DOI: 10.1109/ACCESS.2020.3042874.

QUINLAN, J. R., (1986). "Induction of Decision Trees". In: *Machine Learning* 1.1, pp. 81–106. DOI: 10.1007/BF00116251.

RASMUSSEN, C. E. and WILLIAMS, C. K. I., (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press. ISBN: 978-0-262-18253-9.

REHORN, A. G., JIANG, J., and ORBAN, P. E., (2005). "State-of-the-Art Methods and Results in Tool Condition Monitoring: A Review". In: *The International Journal of Advanced Manufacturing Technology* 26.7-8, pp. 693–710. DOI: 10.1007/s00170-004-2038-2.

RENCHER, A. C., (2002). *Methods of Multivariate Analysis*. 2nd ed. Wiley Series in Probability and Mathematical Statistics. New York, NY: J. Wiley. ISBN: 978-0-471-41889-4.

RIIHIMÄKI, J. and VEHTARI, A., (2010). "Gaussian Processes with Monotonicity Information". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Vol. 9. Proceedings of Machine Learning Research, pp. 645–652.

ROMBACH, K., MICHAU, G., and FINK, O., (2021). "Contrastive Learning for Fault Detection and Diagnostics in the Context of Changing Operating Conditions and Novel Fault Types". In: *Sensors* 21.10, p. 3550. DOI: 10.3390/s21103550.

RUMELHART, D. E., HINTON, G. E., and WILLIAMS, R. J., (1986). "Learning Representations by Back-Propagating Errors". In: *Nature* 323.6088, pp. 533–536. DOI: 10.1038/323533a0.

SATEESH BABU, G., ZHAO, P., and LI, X.-L., (2016). "Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life". In: *Database Systems for Advanced Applications*. Lecture Notes in Computer Science. Ed. by NAVATHE, S. B., WU, W., SHEKHAR, S., DU, X., WANG, X. S., and XIONG, H., pp. 214–228. DOI: 10.1007/978-3-319-32025-0_14.

SAXENA, A., GOEBEL, K., SIMON, D., and EKLUND, N., (2008). "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation". In: *2008 International Conference on Prognostics and Health Management*. Denver, CO: IEEE, pp. 1–9. DOI: 10.1109/PHM.2008.4711414.

SCHLAGENHAUF, T., FEURING, C.-P., HILLENBRAND, J., and FLEISCHER, J., (2019). "Camera Based Ball Screw Spindle Defect Classification System". In: *Production at the Leading Edge of Technology*. Ed. by WULFSBERG, J. P., HINTZE, W., and BEHRENS, B.-A. Lecture Notes in Production Engineering. Berlin: Springer, pp. 503–512. ISBN: 978-3-662-60416-8.

SCHMUCKER, B., TRAUTWEIN, F., SEMM, T., LECHLER, A., ZAEH, M. F., and VERL, A., (2021). "Implementation of an Intelligent System Architecture for Process Monitoring of Machine Tools". In: *Procedia CIRP* 96, pp. 342–346. DOI: 10.1016/j.procir.2021.01.097.

SCHOPP, M., (2009). *Sensorbasierte Zustandsdiagnose und -prognose von Kugelgewindetrieben*. Forschungsberichte aus dem wbk, Institut für Produktionstechnik, Karlsruher Institut für Technologie (KIT) 152. Aachen: Shaker. ISBN: 978-3-8322-8733-7.

SCHWARZENBERGER, M., DROWATZKY, L., WIEMER, H., and IHLENFELDT, S., (2022). "Transferable Condition Monitoring for Linear Guidance Systems Using Anomaly Detection". In: *Production at the Leading Edge of Technology*. Ed. by BEHRENS, B.-A., BROSIUS, A., DROSSEL, W.-G., HINTZE, W., IHLENFELDT, S., and NYHUIS, P. Lecture Notes in Production Engineering. Cham: Springer, pp. 497–505. ISBN: 978-3-030-78423-2.

SCHWENKE, H., SCHMITT, R., JATZKOWSKI, P., and WARMANN, C., (2009). "On-the-Fly Calibration of Linear and Rotary Axes of Machine Tools and CMMs Using a Tracking Interferometer". In: *CIRP Annals* 58.1, pp. 477–480. DOI: 10.1016/j.cirp.2009.03.007.

SERIN, G., SENER, B., OZBAYOGLU, A. M., and UNVER, H. O., (2020). "Review of Tool Condition Monitoring in Machining and Opportunities for Deep Learning". In: *The International Journal of Advanced Manufacturing Technology* 109.3-4, pp. 953–974. DOI: 10.1007/s00170-020-05449-w.

SI, X.-S., WANG, W., HU, C.-H., and ZHOU, D.-H., (2011). "Remaining Useful Life Estimation – A Review on the Statistical Data Driven Approaches". In: *European Journal of Operational Research* 213.1, pp. 1–14. DOI: 10.1016/j.ejor.2010.11.018.

SPATH, D., ROSUM, J., HABERKERN, A., and WEULE, H., (1995). "Kinematics, Frictional Characteristics and Wear Reduction by PVD Coating on Ball Screw Drives". In: *CIRP Annals* 44.1, pp. 349–352. DOI: 10.1016/S0007-8506(07)62340-3.

SPIESS, D., (1970). "Das Steifigkeits- Und Reibungsverhalten Unterschiedlich Gestalteter Kugelschraubtriebe Mit Vorgespannten Und Nicht Vorgespannten Mutternsystemen". PhD thesis. Technische Universität Berlin.

SUN, B., ZENG, S., KANG, R., and PECHT, M. G., (2012). "Benefits and Challenges of System Prognostics". In: *IEEE Transactions on Reliability* 61.2, pp. 323–335. DOI: 10.1109/TR.2012.2194173.

SUTRISNO, E., OH, H., VASAN, A. S. S., and PECHT, M., (2012). "Estimation of Remaining Useful Life of Ball Bearings Using Data Driven Methodologies". In: *2012 IEEE Conference on Prognostics and Health Management*. Denver, CO: IEEE, pp. 1–7. DOI: 10.1109/ICPHM.2012.6299548.

TSAI, P. C., CHENG, C. C., and CHENG, Y. C., (2017). "A Novel Method Based on Operational Modal Analysis for Monitoring the Preload Degradation of Linear Guideways in Machine Tools". In: *Mechanical Engineering Journal* 4.2, pp. 16-00480-16–00480. DOI: 10.1299/mej.16-00480.

TSUI, K. L., CHEN, N., ZHOU, Q., HAI, Y., and WANG, W., (2015). "Prognostics and Health Management: A Review on Data Driven Approaches". In: *Mathematical Problems in Engineering* 2015, pp. 1–17. DOI: 10.1155/2015/793161.

VAN DER AUWERAER, H., GUILLAUME, P., VERBOVEN, P., and VANLANDUIT, S., (2001). "Application of a Fast-Stabilizing Frequency Domain Parameter Estimation Method". In: *Journal of Dynamic Systems, Measurement, and Control* 123.4, pp. 651–658. DOI: 10.1115/1.1410369.

VEREIN DEUTSCHER WERKZEUGMASCHINENFABRIKEN E.V., (2020). *Marktbericht 2020*. Tech. rep. Frankfurt am Main.

VERL, A. and FREY, S., (2010). "Correlation between Feed Velocity and Preloading in Ball Screw Drives". In: *CIRP Annals* 59.1, pp. 429–432. DOI: 10.1016/j.cirp.2010.03.136.

VERL, A., HEISEL, U., WALTHER, M., and MAIER, D., (2009). "Sensorless Automated Condition Monitoring for the Control of the Predictive Maintenance of Machine Tools". In: *CIRP Annals* 58.1, pp. 375–378. DOI: 10.1016/j.cirp.2009.03.039.

VOGL, G. W., WEISS, B. A., and HELU, M., (2019). "A Review of Diagnostic and Prognostic Capabilities and Best Practices for Manufacturing". In: *Journal of Intelligent Manufacturing* 30.1, pp. 79–95. DOI: 10.1007/s10845-016-1228-8.

VOLPONI, A., BROTHERTON, T., and LUPPOLD, R., (2004). "Development of an Information Fusion System for Engine Diagnostics and Health Management". In: *AIAA 1st Intelligent Systems Technical Conference*. Chicago, IL: American Institute of Aeronautics and Astronautics. DOI: 10.2514/6.2004-6461.

WALTHER, M., (2011). *Antriebsbasierte Zustandsdiagnose von Vorschubantrieben*. ISW/IPA Forschung und Praxis 183. Heimsheim: Jost-Jetter. ISBN: 978-3-939890-77-5.

WECK, M. and BRECHER, C., (2006a). *Werkzeugmaschinen 2: Konstruktion und Berechnung*. 8th ed. VDI-Buch. Berlin, Heidelberg: Springer. ISBN: 978-3-540-22502-7.

WECK, M. and BRECHER, C., (2006b). *Werkzeugmaschinen 3: Mechatronische Systeme, Vorschubantriebe, Prozessdiagnose*. 6th ed. VDI-Buch. Berlin, Heidelberg: Springer. ISBN: 978-3-540-22506-5.

WEN, Y., FASHIAR RAHMAN, M., XU, H., and TSENG, T.-L. B., (2022). "Recent Advances and Trends of Predictive Maintenance from Data-Driven Machine Prognostics Perspective". In: *Measurement* 187. DOI: 10.1016/j.measurement.2021.110276.

WEULE, H. and GOLZ, H., (1991). "Preload-Control in Ball Screws — A New Approach for Machine Tool Building?" In: *CIRP Annals* 40.1, pp. 383–386. DOI: 10.1016/S0007-8506(07)62012-5.

YANG, Z., CHEN, C., WANG, J., and LI, G., (2015). "Reliability Assessment of CNC Machining Center Based on Weibull Neural Network". In: *Mathematical Problems in Engineering* 2015, pp. 1–8. DOI: 10.1155/2015/292197.

ZHANG, A., WANG, H., LI, S., CUI, Y., LIU, Z., YANG, G., and HU, J., (2018). "Transfer Learning with Deep Recurrent Neural Networks for Remaining Useful Life Estimation". In: *Applied Sciences* 8.12, p. 2416. DOI: 10.3390/app8122416.

ZHANG, L., LIN, J., LIU, B., ZHANG, Z., YAN, X., and WEI, M., (2019). "A Review on Deep Learning Applications in Prognostics and Health Management". In: *IEEE Access* 7, pp. 162415–162438. DOI: 10.1109/ACCESS.2019.2950985.

ZITZLER, E., LAUMANNS, M., and THIELE, L., (2001). "SPEA2: Improving the Strength Pareto Evolutionary Algorithm". In: *TIK-report* 103. DOI: 10.3929/ethz-a-004284029.

Appendix A

# Supplementary Theory

This appendix will present supplementary theory and details on the topics discussed in Chapter 2.

## A.1 Pooling Layers

Pooling layers are common parts of convolutional neural networks (CNNs) (LeCun et al. 2015). They aggregate filtered data and keep only a fraction of the information. In a maximum pooling layer, for example, only the maximum value of a local number of units is kept. This results in the combination of semantically similar features, which in turn makes the model more robust against shifts in the raw data (e.g. different starting and end points of a time series or different locations of an object in an image). Figure A.1 shows the pooling operation of a maximum pooling layer.

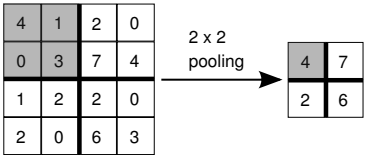



**Figure A.1:** Illustration of a maximum pooling layer, which only keeps the maximum value at each operation; in the illustrated operation, the maximum value in the grey shaded area is four.

## A.2 Markov Chain Monte Carlo

The application of probabilistic models to practical prediction problems often involves a posterior distribution

$$\underbrace{p(\boldsymbol{\theta} \mid \mathcal{D}, f)}_{\text{posterior}} = \frac{\overbrace{p(\mathcal{D} \mid \boldsymbol{\theta}, f)}^{\text{likelihood}} \overbrace{p(\boldsymbol{\theta} \mid f)}^{\text{prior}}}{\underbrace{p(\mathcal{D} \mid f)}_{\text{marginal likelihood}}}, \tag{A.1}$$

which is not tractable anymore (BISHOP 2006, p. 523). As a result, the posterior distribution of the unknown parameters $\boldsymbol{\theta}$ cannot be computed. To solve this issue, Markov chain Monte Carlo (MCMC) algorithms are used to directly sample from the unknown posterior distribution and calculate different statistics, such as the sample mean and variance. MCMC is a large class of algorithms involving various approaches. In the following, to illustrate the general idea of MCMC algorithms, a simple approach called the Metropolis-Hastings (MH) algorithm (HASTINGS 1970; METROPOLIS et al. 1953) will be briefly explained. The following explanations are based on the study by ANDRIEU et al. (2003), which can be used as a reference for an in-depth treatment of the topic.

MCMC generates samples $x^{(i)}$ from the target distribution $p(x)$ using a Markov chain mechanism. The goal is to find a Markov chain which has a stationary distribution matching the unknown distribution $p(x)$. For this approach to work, it must be possible to evaluate $p(x)$ up to a normalising constant. In the case of Eq. (A.1), for example, this means that $p(\boldsymbol{\theta} \mid \mathcal{D}, f) \propto p(\mathcal{D} \mid \boldsymbol{\theta}, f) p(\boldsymbol{\theta} \mid f)$ can be evaluated and the normalising term $p(\mathcal{D} \mid f)$ can be omitted. Hence, the normalising constant of the denominator in Eq. (A.1) (i.e. the marginal likelihood) does not need to be known.

For example, a discrete Markov chain consists of states $x = x_1, x_2, \ldots, x_s$ and a transition matrix, which provides the probability for the transition between each of the states by $p\left(x^{(i+1)} \mid x^{(i)}\right)$. Hence, the evolution to the future state $x^{(i+1)}$ depends only on the previous state $x^{(i)}$. In addition, the distribution of being in the different states $p\left(\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_s \end{bmatrix}^\top\right)$ is the so-called *stationary distribution*, which in some cases can be calculated analytically. In practise, a Markov chain is simulated until convergence to find $p(\mathbf{x})$.

With a one-dimensional variable $x$, the MH algorithm comprises two main steps: first, a sample $x^\star$ from the proposal distribution $q(x^\star \mid x)$ is drawn. Second, the Markov chain moves to the new state $x^\star$ with the acceptance probability

$$\mathcal{A}(x, x^\star) = \min\left\{1, \frac{p(x^\star)q(x \mid x^\star)}{p(x)q(x^\star \mid x)}\right\}. \tag{A.2}$$

Repeating this process $M$ times can result in a stationary distribution, which matches the unknown target distribution. Algorithm 1 outlines the pseudo code of the MH algorithm.

**Algorithm 1** Metropolis-Hastings algorithm

initialise $x^{(0)}$
**for** $i \leftarrow 0$ to $M-1$ **do**
  sample $u \sim \mathcal{U}_{[0,1]}$
  sample $x^\star \sim q\left(x^\star \mid x^{(i)}\right)$
  **if** u $< \mathcal{A}\left(x^{(i)}, x^\star\right) = \min\left\{1, \frac{p(x^\star)q\left(x^{(i)} \mid x^\star\right)}{p(x^{(i)})q(x^\star \mid x^{(i)})}\right\}$ **then**
    $x^{(i+1)} \leftarrow x^\star$
  **else**
    $x^{(i+1)} \leftarrow x^{(i)}$
  **end if**
**end for**

The design of the proposal distribution $q(\cdot)$ is considered to be particularly critical for a successful application of the MH algorithm. A multivariate Gaussian with mean $x^{(i)}$ and a small variance is often chosen as a proposal distribution.

In high-dimensional settings (e.g. when attempting to sample from a large neural network parameter vector $\boldsymbol{\theta}$) the MH algorithm is usually prohibitively slow, which is why one must resort to other sampling schemes. In such high-dimensional use cases, Hamiltonian Monte Carlo (HMC) algorithms, which are sometimes also referred to as *hybrid Monte Carlo* algorithms, can be applied. In contrast to MH algorithms, HMC algorithms make use of the posterior distribution's gradient, which in turn allows a more efficient exploration of higher-dimensional spaces (Hoffman and Gelman 2014). HMC was introduced by Duane et al. (1987) and was pioneered for applications in artificial neural networks (ANNs) by Neal (1996). For an in-depth examination of HMC, the reader is kindly referred to Gelman (2014, pp. 301 sqq.).

## A.3 Variational Inference

The following explanations are based on a review paper on variational inference (VI) by BLEI et al. (2017), which can serve as an elaborate reference for the reader. This section outlines the general ideas.

In contrast to the MCMC approach presented in Appendix A.2, VI approaches do not attempt to sample from the unknown distribution $p(x)$ directly. Rather, they use optimisation techniques to find a simpler and known *variational distribution* $q(x)$ so that

$$q^\star(x) = \underset{\text{w.r.t.} q(x)}{\arg\min}\, \mathrm{D}_{\mathrm{KL}}(q(x) \parallel p(x \mid y)), \tag{A.3}$$

with $p(x \mid y)$ being the general form of Eq. (A.1)

$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{p(y)}, \tag{A.4}$$

and $\mathrm{D}_{\mathrm{KL}}(q(x) \parallel p(x \mid y))$ being the Kullback-Leibler (KL) divergence, which is defined as (BLEI et al. 2017)

$$\begin{aligned} \mathrm{D}_{\mathrm{KL}}(q(x) \parallel p(x \mid y)) &= \mathbb{E}[\log(q(x))] - \mathbb{E}[\log(p(x \mid y))] \\ &= \mathbb{E}[\log(q(x))] - \mathbb{E}[\log(p(x, y))] + \log(p(y)). \end{aligned} \tag{A.5}$$

However, the KL divergence, which is to be optimised, contains the unknown target distribution $p(x \mid y)$. Hence, Eq. (A.5) cannot be computed. Instead, the so-called evidence lower bound (ELBO), which is defined as (BLEI et al. 2017)

$$\mathrm{ELBO}(q) = \mathbb{E}[\log(p(x, y))] - \mathbb{E}[\log(q(x))], \tag{A.6}$$

is maximised. The fact that maximising the ELBO is indeed minimising the KL divergence from Eq. (A.5) becomes evident when combining Eqs. (A.5) and (A.6) and considering that

$$\mathrm{D}_{\mathrm{KL}}(q(x) \parallel p(x \mid y)) = \log(p(y)) - ELBO(q). \tag{A.7}$$

The *expectation propagation* algorithm introduced by MINKA (2001) is a special variant of VI algorithms which is used in this thesis for the Gaussian process classification (GPC) model.

Overall, the choice between MCMC and VI for approximate Bayesian inference is mainly a trade-off between cost and accuracy. This is because MCMC is slow but guarantees to asymptotically produce exact samples from the target distribution, whereas VI is fast but offers no such guarantees.

## A.4 Weibull Distribution

A random variable $x$ follows a Weibull distribution when its probability density function (PDF) is given by

$$f(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) & \forall\ x \geq 0 \\ 0 & \forall\ x < 0. \end{cases} \tag{A.8}$$

The scale parameter $\alpha$ and the shape parameter $\beta$ determine the form of the resulting PDF (see Fig. A.2). As the Weibull distribution is considered suitable for modelling failures of machine tools (DAI et al. 2003; YANG et al. 2015), it was applied for publication ⑤ (see Section 6.1) to analyse the economic efficiency of different maintenance strategies.

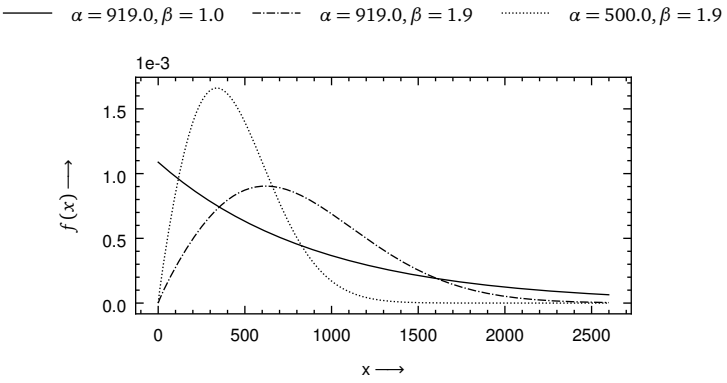



**Figure A.2:** Illustration of several PDFs with different parameters; it can be seen that the scale and shape parameters $\alpha$ and $\beta$ determine the overall form of the PDF.

Appendix B

# Experimental Setup

In the following sections, the experimental setup used for publications ① and ② (see Sections 5.1 and 5.2, respectively) will be described. In Appendix B.1, the test bench and the investigated components will be presented, and in Appendix B.2, the measurement equipment will be described.

## B.1 Test Bench

The investigated test bench is a DMG DMC duo Block 55H five-axis milling machine without housing, without spindle, without workpiece table and with only the linear axes installed (see Fig. B.1). The advantage of this reduced system lies in the accessibility enabling a quick disassembly and re-assembly of different components. Nevertheless, the test bench is more realistic and complex than a single-axis feed drive test bench, which is the basis of many experiments in the existing literature.

The investigated components were Bosch Rexroth FEM-E-C 40x16Rx6-4 ball screws and Bosch Rexroth RWA-045-FNS-C1-U-2 runner blocks, which were available in different preload levels as well as with and without pitting damage. The components are listed in Tables B.1 and B.2. The runner blocks F3 with the preload level C1 were mounted on the guide rails for the experiments in publication ② (see Section 5.2).

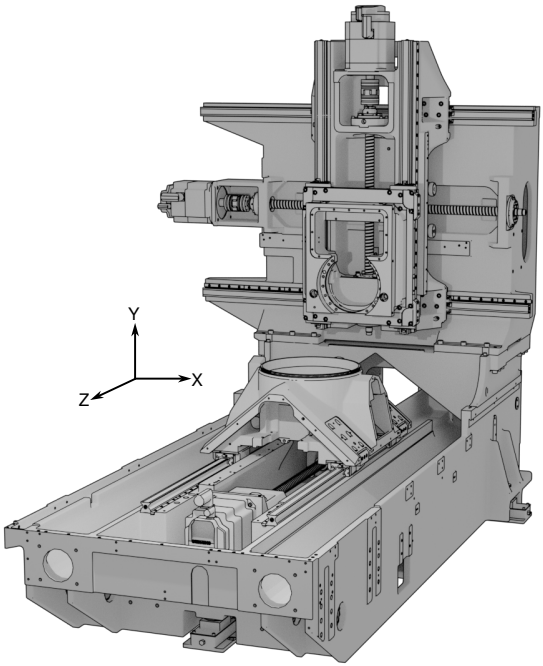



**Figure B.1:** Render of the DMG DMC duo Block 55H test bench; the feed drive of the X-axis was investigated in this thesis.

| components | | | class labels | | |
|---|---|---|---|---|---|
| ball screw | preload level | pitting | preload loss | pitting damage | indistinct defect |
| P1 | C3 (2070 N) | yes | – | 1 | 1 |
| P2 | C3 (2160 N) | yes | – | 1 | 1 |
| C11 | C1 (950 N) | no | 1 | 0 | 1 |
| C12 | C1 (845 N) | no | 1 | 0 | 1 |
| C31 | C3 (2390 N | no | 0 | 0 | 0 |
| C32 | C3 (2328 N) | no | 0 | 0 | 0 |
| C33 | C3 (2031 N) | no | 0 | 0 | 0 |

**Table B.1:** Overview of the ball screws investigated in the publications of this thesis and the respective class labels for the following three class definitions used in publication ②: preload loss, pitting damage and indistinct defect

| runner block | preload level | pitting |
|---|---|---|
| C1 | C1 (4060 N) | no |
| C2 | C1 (4430 N) | no |
| C3 | C1 (4430 N) | no |
| F1 | C1 (3880 N) | no |
| A9 | C3 (13 470 N) | no |
| A10 | C3 (14 530 N) | no |
| A12 | C3 (12 840 N) | no |
| D3 | C3 (12 840 N) | no |

**Table B.2:** Overview of the runner blocks investigated for the publications of this thesis; in each preload class, four runner blocks were available.

## B.2 Utilised Sensors, Measurement Equipment and Measurement Software

During the experiments at the test bench described in Appendix B.1, the measurements were conducted with the following hardware: all vibration measurements were acquired with Kistler 8762A10 piezo-electric accelerometers at a frequency of 20 kHz. For data acquisition, the National Instruments™ cDAQ-9198 system in combination with an *iwb* data acquisition software toolbox based on MATLAB® was applied. The machine tool from Fig. B.1 was equipped with a Heidenhain iTNC530 numeric control. The Heidenhain software tool TNCopt v. 8.0.125.1 was used to excite the machine with a defined sine sweep (see Fig. 5.1). The Heidenhain software tool TNCscope v. 4.2.37 was deployed to acquire signals from the numeric control. The temperature measurements were conducted with type K thermocouples.

Appendix C

# Benchmark Data Sets

This part of the appendix will present the two benchmark data sets utilised for publications ③ and ④.

## C.1 C-MAPSS Turbofan Engine Data Set

The C-MAPSS data set was first published by SAXENA et al. (2008) and comprises simulated data from realistically large commercial turbofan engines. It consists of four subsets, each of which is composed of a training data set and a test data set of simulated run-to-failure sequences. The training data set provides entire run-to-failure sequences, such that, for every point in time, the true remaining useful life (RUL) of an engine is known. In contrast, the test data set contains sequences which randomly stop. However, the true RUL values for the test data are provided. Each run-to-failure sequence is composed of different sensor signals. The four subsets vary in the number of training sequences, test sequences, fault modes and operating conditions (see Table C.1). An exemplary run-to-failure sequence for the engine number 43 from the subset FD001 is depicted in Fig. C.1.

| | subset | | | |
|---|---|---|---|---|
| | FD001 | FD002 | FD003 | FD004 |
| number of training data sequences | 100 | 260 | 100 | 249 |
| number of test data sequences | 100 | 259 | 100 | 248 |
| operating conditions | 1 | 6 | 1 | 6 |
| fault modes | 1 | 1 | 2 | 2 |

**Table C.1:** Overview of the different C-MAPSS subsets based on BENKER, FURTNER, et al. (2021)

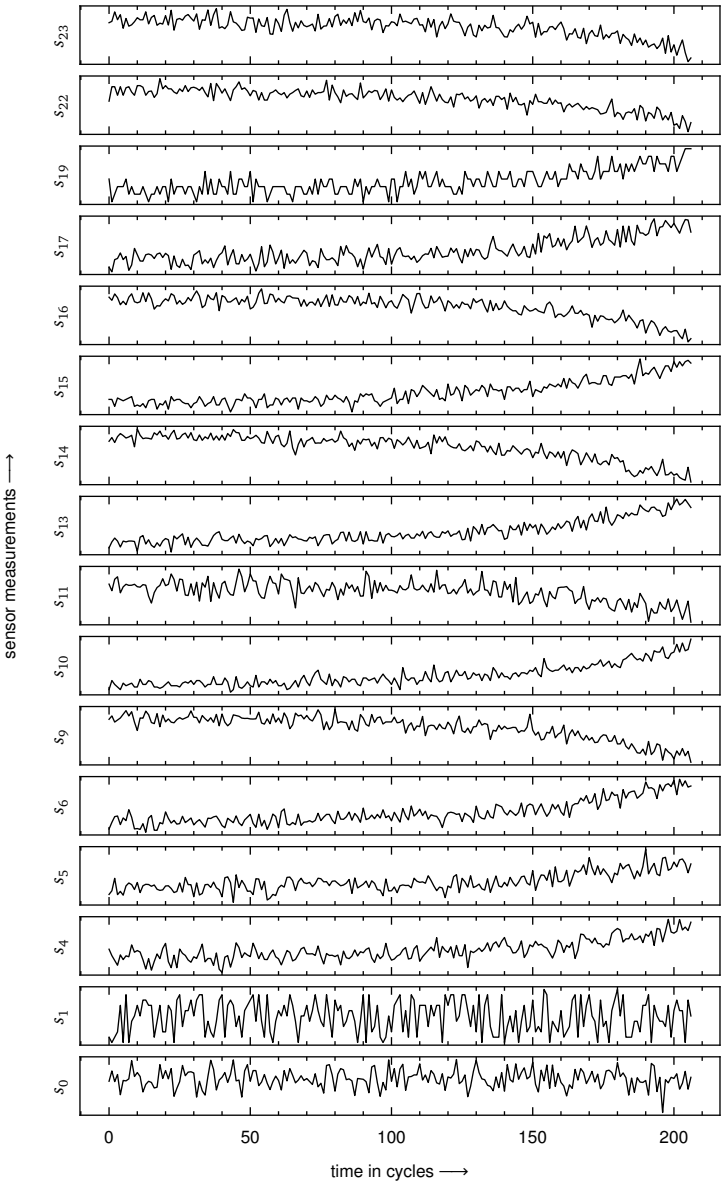



**Figure C.1:** Exemplary C-MAPSS training data of engine number 43 from subset FD001; sensors $s_2$, $s_3$, $s_7$, $s_8$, $s_{12}$, $s_{18}$, $s_{20}$ and $s_{21}$ were omitted, as their signals are either constant or only change slightly along time.

## C.2 FEMTO Bearing Data Set

The FEMTO data set, originally published by NECTOUX et al. (2012), consists of run-to-failure sequences recorded at a bearing test bench. The rolling bearings were clamped on a bearing support shaft, which was driven by a motor. The bearings were operated without artificially inducing failures until the end of their lifetime. During operation, a radial force was introduced by a hydraulic actuator, which was varied for the different run-to-failure experiments. Hence, different operating conditions were recorded for different bearings. During the experiments, vibrations were recorded every 10 s for a duration of 0.1 s with a sampling frequency of 25.6 kHz. Additionally, temperatures were measured continuously with a sampling frequency of 10 Hz.

For each of the three operating conditions, at least two run-to-failure sequences are available for training. For the conditions one and two, five sequences are provided for testing. For condition three, one sequence is available for testing. In Fig. C.2, an exemplary recording of the vibration data from the FEMTO data set is shown.

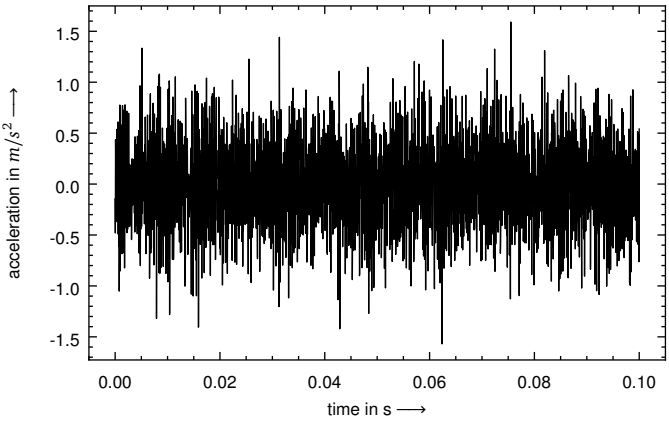



**Figure C.2:** Exemplary FEMTO training data recorded with the accelerometer

**Appendix D**

# List of Supervised Student Theses

In the context of the research performed by the author, various student theses (see Table D.2) were intensively supervised with regard to the methodology, problem statements, objectives, and research approach, along with the interpretation and documentation of all results. The supervision took place in the years 2018 to 2022 at the Institute for Machine Tools and Industrial Management of TU Munich (in German: Institut für Werkzeugmaschinen und Betriebswissenschaften der TU München) (*iwb*). The findings and results of the student theses have contributed to this dissertation. The author would like to express his gratitude for the commitment and contributions of all supervised students.

| name | title | submission |
|---|---|---|
| Álvaro Gómez Asensio | Unsupervised Learning for Structural Health Monitoring | May 2019 |
| Manuel Friedlhuber | Condition Monitoring von Kugelgewindetrieben mit Methoden des maschinellen Lernens | June 2019 |
| Florian Trocker | Prädiktion von Montagefehlern in der Automobilindustrie mit Methoden des maschinellen Lernens | November 2019 |
| Tobias Bremicker | Untersuchung der thermischen Stabilität des dynamischen Verhaltens eines Vorschubantriebs | June 2020 |
| Sebastian Junker | Feature Engineering and Machine Learning for Condition Monitoring of Machine Tools | August 2020 |
| Lukas Furtner | Application of Bayesian Neural Networks for Remaining Useful Life Prediction using Hamiltonian Monte Carlo and Variational Inference | September 2020 |
| Artem Bliznyuk | Validation and Development of a Probabilistic Method for Remaining Useful Life Estimation for Machine Tool Ball Screws | October 2020 |
| Leonie Müller | Development of a Method to Evaluate the Economic Feasibility of Predictive Maintenance and Definition of Requirements on its Profitable Use | October 2020 |
| Nico Schneucker | Condition Monitoring of Machine Tool Feed Drives via Probabilistic Machine Learning | April 2021 |
| Benedikt Tegethoff | Life Cycle Costs of Industrial Milling and Honing Machines and Identification of Risk Transfer Solution Eligibility Criteria Based on a Case Study | July 2021 |
| Hans Stadlbauer | Zustandsüberwachung von Kugelgewindetrieben mittels maschineller Lernverfahren | August 2021 |
| Victor Rommel | Discrete Event Simulation for the Economic Assessment of Predictive Maintenance Strategies | September 2021 |
| Maria Stroganova | Setpoint Optimization within Process Control: A Case Study Towards Production Waste Minimization | October 2021 |
| Leopold Beck | Automatisierung der experimentellen Modalanalyse für Werkzeugmaschinen | April 2022 |
| Fabian Kolb | Intelligent Ball Screw Fault Diagnosis Using Deep Learning Based Domain Adaptation and Transfer Learning | September 2022 |

**Table D.2:** Supervised student theses

Appendix E

# Publications and Contributions of the Author

This publication-based dissertation is based on the following five papers:

**Publication ①**
BENKER, M., JUNKER, S., ELLINGER, J., SEMM, T., and ZAEH, M. F., (2022). "Experimental Derivation of a Condition Monitoring Test Cycle for Machine Tool Feed Drives". In: *Production Engineering* 16.1, pp. 55–64. DOI: 10.1007/s11740-021-01085-9

**Publication ②**
BENKER, M. and ZAEH, M. F., (2022). "Condition Monitoring of Ball Screw Feed Drives Using Convolutional Neural Networks". In: *CIRP Annals* 71.1, pp. 313–316. DOI: 10.1016/j.cirp.2022.03.017

**Publication ③**
BENKER, M., FURTNER, L., SEMM, T., and ZAEH, M. F., (2021). "Utilizing Uncertainty Information in Remaining Useful Life Estimation via Bayesian Neural Networks and Hamiltonian Monte Carlo". In: *Journal of Manufacturing Systems* 61, pp. 799–807. DOI: 10.1016/j.jmsy.2020.11.005

**Publication ④**
BENKER, M., BLIZNYUK, A., and ZAEH, M. F., (2021). "A Gaussian Process Based Method for Data-Efficient Remaining Useful Life Estimation". In: *IEEE Access* 9, pp. 137470–137482. DOI: 10.1109/ACCESS.2021.3116813

**Publication ⑤**

BENKER, M., ROMMEL, V., and ZAEH, M. F., (2022). "An Investigation into the Economic Efficiency of Different Maintenance Strategies Based on a Discrete Event Simulation". In: *Procedia CIRP* 107, pp. 428–433. DOI: 10.1016/j.procir.2022.05.003

The author of this thesis is the main author of all five publications. Co-author Johannes Ellinger has contributed as a colleague at the Institute for Machine Tools and Industrial Management of TU Munich (in German: Institut für Werkzeugmaschinen und Betriebswissenschaften der TU München) (*iwb*) by brainstorming, discussing and auditing the research results. Thomas Semm was the head of the department for machine tools and reviewed the manuscripts. Sebastian Junker, Lukas Furtner, Artem Bliznyuk and Victor Rommel contributed in their role as supervised students. They supported the experimental activities, assisted during the programming tasks and reviewed the manuscripts. Prof. Dr.-Ing. Michael F. Zaeh contributed as the head of the *iwb* and the supervisor of this dissertation project.

In Table E.1, the contributions of the author of this thesis are shown. These are further divided into developing the ideas, execution of the experiments and writing the publications.

| | idea | execution | publication | total |
|---|---|---|---|---|
| publication ① | 70 % | 65 % | 75 % | 70 % |
| publication ② | 90 % | 100 % | 90 % | 93 % |
| publication ③ | 80 % | 50 % | 80 % | 70 % |
| publication ④ | 75 % | 50 % | 80 % | 68 % |
| publication ⑤ | 80 % | 35 % | 85 % | 67 % |

**Table E.1:** Summary of the contributions of the author of this thesis to the five publications