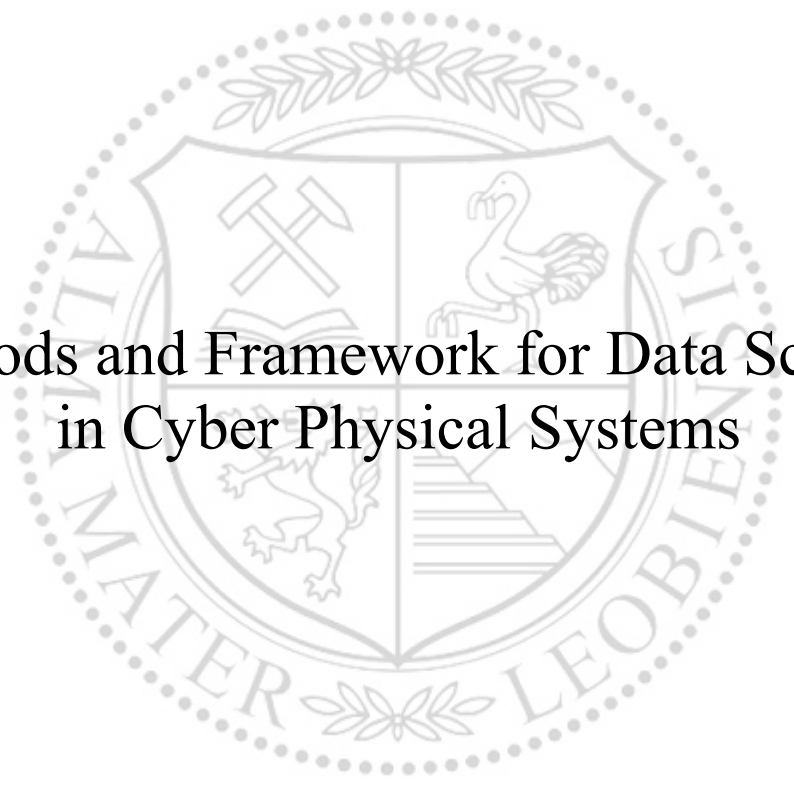




Chair of Automation

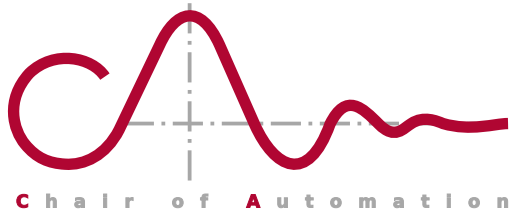
Doctoral Thesis

The background features a large, faint watermark of the University of Leoben seal. The seal is circular and contains a shield with four quadrants: top-left shows crossed hammers, top-right shows a stork, bottom-left shows a rampant lion, and bottom-right shows a staircase. The text 'UNIVERSITAS LEOBENSIS' is written around the top and 'MATER LEOBENSIS' around the bottom of the seal.

Methods and Framework for Data Science
in Cyber Physical Systems

Dipl.-Ing. Roland Ritt, BSc

April 2019



Copyright © 2019

Roland Ritt

Chair of Automation
Department Product Engineering
Montanuniversitaet Leoben
Peter-Tunner Straße 25
8700 Leoben, Austria

W: automation.unileoben.ac.at
E: automation@unileoben.ac.at
T: +43(0)3842/402-5301
F: +43(0)3842/402-5302

Cite this thesis as

```
@phdthesis{Ritt2019Thesis,  
  Author = {Ritt, Roland},  
  School = {Montanuniversitaet Leoben,  
            Department Product Engineering,  
            Chair of Automation},  
  Title  = {Methods and Framework for Data Science  
            in Cyber Physical Systems},  
  Year   = {2019}  
}
```

All rights reserved.

Last compiled on April 29, 2019; the document has been compiled 2042 times in total.

This thesis was typeset using L^AT_EX (PDF_Latex, Bib_Tex and Make_Index from MiK_TeX 2.9). Formulas and equations are formatted according to ISO80000-2. Computations were performed using The MathWorks MATLAB (2016a, 2017a, 2018b). Illustrations were designed using CorelDRAW X7. All registered trademarks and trade names are the property of their respective holders.

AFFIDAVIT

I declare on oath that I wrote this thesis independently, did not use other than the specified sources and aids, and did not otherwise use any unauthorized aids.

I declare that I have read, understood, and complied with the guidelines of the senate of the Montanuniversität Leoben for "Good Scientific Practice".

Furthermore, I declare that the electronic and printed version of the submitted thesis are identical, both, formally and with regard to content.

Date 30.04.2019

Signature Author
Roland, Ritt
Matriculation Number: 01035329

Acknowledgements

First and foremost, I would like to thank my supervisor and mentor Paul O’Leary. He always took the time needed to discuss personal and technical details, although he had to lead and manage the Chair of Automation. His inner force, his broad knowledge in nearly all fields and his creativity to solve scientific problems inspired me to become a researcher and to take the challenge of writing this thesis. Paul’s sensitivity to pull or push me in the right moment helped me to find my own way and to overcome difficulties. Secondly, I want to thank Matthew Harker, who supported me to find the answer to nearly every scientific question. Last but not least, I want to thank Peter Lee for taking the challenge to be my second supervisor.

For me it would have been impossible to be a Ph.D. candidate without a suitable environment and working atmosphere. Therefore, I want to thank Roland Schmidt, whom I shared an office with for nearly four years and so over time we established a mutual friendship. Besides technical discussions, he always offered me his help wherever he could. Furthermore, I need to mention the coffee table at our Chair – a crucial meeting point. Without it, and of course all my Ph.D. and Master colleagues sitting around the table, some crazy ideas, either technical or personal in nature, would have been never born. Another big “Thank you” goes to Gerold Probst and Petra Hirtenlehner, who always knew how to handle the jungle of bureaucracy and IT.

Music plays an important role in my life. Therefore, I want to thank my friends and band colleagues from “Schnopsidee”, who converted several nights into days, either during our music sessions or afterwards, sharing a glass of beer and lively discussions about god and the world. This was important to clear the mind and free it up for new thoughts.

Also, I want to express my sincere gratitude to my family, who always encourages me to follow my dreams and support me wherever they can.

A final, but very important “Thank you” goes to Elisabeth, my amazing girlfriend. She is not only the one who makes me smile every single day but also manages to make difficult phases in life much easier – simply the best support during my studies.

Kurzfassung

Die vorliegende Arbeit erforscht mathematische und computergestützte Methoden, die zur Analyse von Daten geeignet sind, welche von großen cyber-physikalischen Systemen stammen. Durch die Einbettung der dem Systemverhalten zugrundeliegenden Gleichungen, insbesondere der Dynamik, werden Lösungen abgeleitet, welche mit den physikalischen Grundgesetzen des Systems kompatibel sind. Bei den entwickelten Methoden werden dabei die Messunsicherheiten, welche grundsätzlich die Daten überlagern, in der Fehlerabschätzung berücksichtigt.

Basierend auf Ideen aus dem Gebiet der symbolischen Datenanalyse werden Ansätze entwickelt, welche automatisch und unüberwacht Strukturen in multivariaten Zeitreihen identifizieren können. Dabei werden Elemente und Methoden, die in der Entwicklung der natürlichen Sprache eine wesentliche Rolle spielen, computergestützt nachgeahmt. Die Funktionsweise wird anhand eines Beispiels gezeigt, in welchem automatisch unterschiedliche Betriebszustände erkannt werden. Besonders interessant ist in diesem Zusammenhang die Identifikation der menschlichen Interaktion mit dem System, welche zu einer Struktur in den Zeitreihen führt und somit erkannt werden kann.

Darüber hinaus wird in dieser Arbeit die Charakterisierung von Sensoren und die Quantifizierung ihres Verhaltens behandelt, wodurch deren Messunsicherheit abgeleitet und modelliert werden kann. Dies ist von grundlegender Bedeutung, da sich Fehler, die bereits bei der Interpretation von Sensordaten entstehen, fortpflanzen und dadurch den gesamten Analysezyklus beeinflussen.

Die eingeführten und präsentierten Techniken und Methoden werden in ein Framework integriert, welches sämtliche Schritte der Datenanalyse – von der Datenerfassung bis zur Aufbereitung der Ergebnisse – unterstützt. Ein in dieser Arbeit entwickeltes Softwaretool erweitert den Funktionsumfang des Frameworks durch das Bereitstellen von Werkzeugen zur Handhabung, Analyse und Visualisierung von großen multivariaten Zeitreihen, wodurch die Arbeit des Datenanalysten unterstützt wird.

Die vorliegende Dissertation fasst die durchgeführte Forschung als eine Sammlung von Publikationen zusammen, welche mit einleitenden Texten und Erweiterungen zu einem durchgängigen Dokument verknüpft wird.

Schlagwörter

Datenwissenschaften; cyber-physikalisches System; inverses Problem; diskrete orthogonale Polynome; symbolische Zeitreihenanalyse; Polynomapproximation

Abstract

This work investigates mathematical and computational methods suitable for analysing data emanating from large cyber physical systems. Embedding the governing equations for the system behaviour, especially dynamics, ensures analysis solutions which are consistent with the physics of the system. The developed methods also deal with the implicit uncertainty fundamentally associated with perturbed data.

Symbolic data analysis is investigated as a means of establishing a consistent computational approach to perform automatic unsupervised identification of structures in multi-channel time series data. This is achieved by mimicking techniques from the evolution of natural language. The validity of the approach is demonstrated in an application to automatic operations recognition. Particularly interesting in this context is the identification of human interaction with the system via structure embedded in the data.

Additionally, this thesis considers the issue of characterizing sensors and quantifying their behaviour, in particular modelling their uncertainty. This is fundamental since errors entering via the interpretation of sensor data will propagate through the entire analysis cycle.

The established methods and techniques are integrated into a framework to support end-to-end applications, i.e. from the data acquisition to the presentation of the results. A software tool, developed within this work, extends the framework to support the data analyst in the handling, analysis and visualization of large multi-dimensional time series together with the computational results.

The conducted research is presented as a collection of papers woven together with introductory texts and some extensions to form a complete thesis.

Index Terms

Data science; cyber physical system; inverse problem; discrete orthogonal polynomials; symbolic time series analysis; polynomial approximation

Table of Contents¹

Affidavit	V
Acknowledgements	VII
Abstract	IX
Table of Contents	XI
1 Introduction	1
1.1 Motivation	1
1.2 Structure and Synopsis of the Thesis	2
1.3 Contribution	5
1.4 Remarks	6
I Prerequisites	7
2 Data Science in Cyber Physical Systems	8
2.1 Data Science	8
2.2 Cyber Physical Systems	10
2.3 A Structured Approach to Data Analytics and Knowledge Discovery .	13
2.3.1 Knowledge Discovery Process Models	13
2.3.2 From Data to Knowledge	15
2.3.3 Fundamental Premiss Behind Data Analytics in Sensor Data .	17
3 Data Ingestion Framework	19
3.1 Data Flow Structure	19
3.2 Data Acquisition and Ingestion Process	21
3.2.1 Data Acquisition	22
3.2.2 Data Ingestion	23
3.3 Contiguous Data Model	24
3.3.1 Data On-Demand – Interface to the Global Data Warehouse .	25
4 Data Analytics Framework	26
4.1 Data Handling Framework	28
4.1.1 Core Attributes	28
4.1.2 Key Methods	29

¹The section numbering within the papers is according to the original format.

4.2	Decorative Objects	31
4.2.1	Symbolic Time Series Representation	31
4.2.2	Segments	33
4.2.3	Events	33
4.3	Additional Functions and Toolboxes	34
4.4	Data Visualization	35
4.4.1	Visualizing Multi-Dimensional Time Series Data	35
4.4.2	Data Decimation	35
4.4.3	Decorative Overlays	37
 II Polynomial Methods		 39
5	Synopsis	40
6	Paper: Constrained Polynomial Approximation for Inverse Problems in Engineering	42
1	Introduction	43
2	Notation and definition of constraint types	45
3	Polynomial approximation with constraining roots	46
3.1	Algebraic formulation	47
3.2	Algorithmic implementation	48
3.3	Example and interpretation	49
4	Polynomial approximation with constraining values	51
4.1	Algebraic formulation	52
4.2	Numerical example	53
5	Polynomial approximation with generalized constraints	53
5.1	Algebraic formulation	55
5.2	Algorithmic implementation	57
5.3	Numerical example	58
5.4	Extension of this work	59
6	Coefficient constrained polynomial approximation	59
6.1	Example coefficient constraints	60
7	Conclusions	61
	Acknowledgments	61
	References	61
7	Paper: Hierarchical Decomposition and Approximation of Sensor Data	63
1	Introduction	64
2	Methodology and Algebraic Framework	65
2.1	Weighted Local Polynomial Approximation - Hierarchy Level 1	67
	Weighting Functions	67
	Spatioal Weighted Local Regression	69
	Covariance Propagation	70
2.2	Hermite Approximation - Hierarchy Level 2 and above	71
	Algebraic Formulation	71

	Local Hermite Approximation	73
	Performance Test	74
2.3	Data Reconstruction	75
	Taylor Expansion	75
	Generalized Hermite Interpolation	75
3	Numerical Testing	77
4	Conclusion	79
	Acknowledgments	80
	References	81
8	Paper: Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials	83
I	Motivation	85
II	Review of Literature	86
III	Theoretical Framework	86
	A Modelling of Measured Values	86
	B Approximation of Values and Derivatives	86
	C Synthesis of Weighted Discrete Orthogonal Basis	87
	D Covariance Propagation	89
IV	Numerical Example	89
V	Numerical Quality of Basis	89
VI	Conclusion	91
	Acknowledgment	91
	References	91
III	Symbolic Time Series Analysis	93
9	Synopsis	94
	9.1 Local Linear Differential Operator	95
10	Paper: Mining Sensor Data in Larger Physical Systems	97
1	Introduction	98
2	A Structured Approach to Data Analytics/Mining	99
3	Data Collection and Management	99
4	Linear Differential Operators	100
5	Single Channel Information	100
6	Epistemology and the Emergence of Speech	101
	6.1 Parallel Channels	101
7	Example Applications of the System	101
	7.1 Commissioning Support	102
	7.2 Fleet Management	102
	7.3 Logistics and Preventative Maintenance	102
	7.4 System Identification	102
8	Conclusions	102
	References	102

11 Paper: Advanced Symbolic Time Series Analysis in Cyber Physical Systems	104
1 Local Linear Differential Operators (LDO)	105
2 Symbolic Time Series Analysis	107
3 Conclusion	110
References	111
12 Paper: Symbolic Analysis of Machine Behaviour and the Emergence of the Machine Language	112
1 Preamble	113
2 Introduction and Related Work	115
3 Methodology	117
3.1 Linear Differential Operator (LDO)	117
3.2 Advanced Symbolic Time Series Analysis (ASTSA)	118
3.3 Hierarchical Compounding of Words	119
4 Background - Relation to Natural Language	120
5 Experimental Evaluation	121
6 Conclusion and Future Work	122
Acknowledgments	122
References	123
IV Applied Data Analytics in Cyber Physical Systems	125
13 Synopsis	126
14 Paper: MEMS Based Inclinometers: Noise Characteristics and Suitable Signal Processing	128
I Introduction	129
II Analysis of the SCA103T-D04	130
A Histograms and distributions for $x(t)$ and $y(t)$	130
B Correlation in the perturbations of $x(t)$ and $y(t)$	131
III Analysis of the SCA830-D07	133
IV Conclusion	133
Appendix: Cauchy-Lorentz Distribution	134
References	134
15 Paper: Force Based Tool Wear Detection using Shannon Entropy and Phase Plane	135
I Introduction	136
II Measurement Setup	137
III Segmentation	137
IV Statistical Central Moments and Time Histogram	138
V Entropy and Information	139
A Segment based Entropy	139
B Local Entropy	139
VI Phase Diagram and Joint Entropy	140

VII	Conclusion	141
VIII	Acknowledgment	141
	References	141
16	Paper: Real-Time-Data Analytics in Raw Materials Handling	142
1	Introduction	143
2	System Premiss	144
3	Data Ingestion	145
4	Systems Currently Being Monitored	148
5	Exemplary Data Evaluations	149
5.1	Incident Analysis	150
5.2	Long-Term Logistics Optimisation	150
6	Conclusions	151
	References	151
17	Paper: Condition Monitoring of Hydraulics in Heavy Plant and Machinery	153
	Introduction	155
	Methodology	156
	Hydraulics Monitoring	157
	Statistics	159
	Results and Conclusion	160
	References	161
V	Discussion and Appendices	163
18	Conclusion and Outlook	164
A	List of Figures	168
B	List of Tables	169
C	List of Algorithms	169
D	List of Author's Publications	170
E	References	172

1 | Introduction

1.1 Motivation

During the work on projects with companies building and operating large mining machines, i.e. large physical systems, the people involved often expressed the wish of collecting data from their machines and to analyse this data. Their goals being to:

1. Improve their machines, i.e. to obtain engineering feedback.
2. To find and analyse incidents, i.e. to detect misbehaviour of the machine and to determine the cause of this behaviour. This is particularly prevalent for incidents with serious or catastrophic consequences.
3. Estimate the wear of parts with the aim of making maintenance and/or contingency plans, i.e. enable predictive maintenance.
4. Operate equipment more efficiently through continuous monitoring, i.e. automatic report generation.
5. Characterize operating procedures automatically, i.e. automatic operations recognition. This is a key aspect in achieving efficiency.

To support these activities, there is a need for a structured and secure data collection-, archiving- and analysing-system, which is easy to interact with. In other words, the person who is working on the data should focus on the analysis and not be concerned about collecting the data or how they are stored. Although this task seems simple, a lot of companies which tried to implement such systems failed due to the complexity and diversity of tasks involved.

Besides this issue, the available data analytics software is normally based on statistics (i.e. it is used as a black box) and therefore does not support the embedding of a-priori knowledge of the system within calculations. Since large physical systems have to follow the laws of physics, the systems cannot be operated randomly, i.e. not purely stochastic. Statistics alone cannot lead to semantics based on the physical behaviour of a system, as system models are required to ensure causality. Therefore, including the system dynamics is important to analyse the data more precisely.

To investigate the available data, various levels of abstraction need to be generated to “read” and interpret the data. Therefore, meaning must be associated with the data hierarchically. Starting from a global overview, the meaning is refined accordingly to obtain a more detailed view. In general *meaning* can be expressed by using language. Since the monitored machines are operated from humans acting non-analytically, the idea of interpreting the data using human readable text, i.e. words and symbols, was developed.

Based on these considerations the *thesis* of this dissertation can be summarized as:

It is possible to formulate a framework based on consistent data structures, mathematical and statistical models and methods to enable structured analysis of large data sets emanating from cyber physical systems.

Within this work, this statement is substantiated by the included publications and content introduced in additional chapters. The structure therefore is presented in the following section.

1.2 Structure and Synopsis of the Thesis

As the title of this document suggests, this thesis investigates multiple topics important for data analytics in large physical systems.

In Section 1.1 the motivation which lead to this work is presented together with the thesis statement.

The main body of the thesis is structured into four parts (Parts I to IV) dealing with four different areas of work, whereby Parts II to IV contain ten papers, which can be seen as the major contribution of this thesis. Each of these parts is preceded by a detailed synopsis (see Chapters 5, 9 and 13), weaving the publications together within the topic. An overview of the author’s contribution to the included publications is given in Section 1.3.

The parts of the thesis are:

Part I: Prerequisites. This one presents prerequisites needed to perform data science in cyber physical systems. To get an overview of the entire topic, Chapter 2 investigates the terms *Data Science* and *Cyber Physical Systems* in Section 2.1 and Section 2.2. Additionally, the structured approach to data analytics followed within this thesis is presented in Section 2.3.

Since the basis for data analytics is the *data* itself, Chapter 3 presents the means and methods for data collection and ingestion into a global data warehouse. Therefore, the overall data flow, the data acquisition and ingestion structure are introduced in Section 3.1 and Section 3.2. A continuous view onto the data present on the global

data warehouse is enabled by using a contiguous data model, which is established in Section 3.3.

To handle data locally and perform computations, so called “data on-demand” services map requested data to a structure (object) which fits into the data analytics framework. This is presented in Chapter 4. It describes the core functionality needed to set the focus on data and development of new algorithms rather than on data handling.

Part II: Polynomial Methods. The Weierstrass approximation theorem [1] proves that polynomials can model or approximate *any* function. As a result, polynomials have become a central tool in analysing data; remember the Fourier bases area are also polynomials. Therefore, three publications dealing with polynomials are collected and presented in Part II.

In the first paper, Chapter 6, a consistent mathematical framework for the approximation with polynomials, which have to fulfil constraints, is presented together with the covariance propagation. In this manner, the systematic behaviour of the system can be characterized together with the uncertainty. Fundamentally, there will always be some uncertainty involved when establishing models from perturbed data. The constraints addressed in this paper are zero-, value- and general derivative constraints and constraints on the coefficients.

Motivated by the idea of how the constraints are included in the computation, the ideas in the paper presented in Chapter 7 were developed. In this publication, a time series is approximated hierarchically by first calculating the state vectors for given intervals using weighted local polynomial approximations. To approximate the state vectors in the next hierarchical levels, a new method was developed, which takes both, value and derivative information into account. It uses geometric polynomials (i.e. Vandermonde basis) and their analytical derivatives to simultaneously approximate the state vectors. Covariance weighting is used to establish a metric relationship between values and derivatives. Additionally, the temporal behaviour of the states, which can be analysed in the state space, is characteristic for the dynamics of a system.

Since geometric polynomials may become numerically unstable for high degrees, discrete orthogonal polynomial methods are developed in the work presented in Chapter 8 to address this issue. It uses covariance weighting in the three term recurrence relation to synthesize an orthogonal basis function set, whereby the covariance weighting establishes a metric in the state space leading to a valid approximation. It is shown that this basis is advantageous compared to the Vandermonde basis, especially for high degree polynomial approximation.

Part III: Symbolic Time Series Analysis. In symbolic time series analysis a stream of data is quantized and transformed into a stream of symbols which is in general a compressed representation of the original data. This idea is adapted and used in the paper appended in Chapter 10. There, meaning is associated with the

symbols/words by including the dynamics of the system within the symbolization step. Therefore, a stream of data can now be seen as a stream of words similar to natural language.

The presented methods build the basis for the publication presented in Chapter 11. To analyse multi-dimensional time series ideas from natural language are taken up, i.e. the symbols from two different channels are merged to form polysyllabic words for describing more complex behaviour. Additionally, frequency dictionaries are used to identify different operation modes from data emanating from a bucket-wheel-excavator.

To automatically reveal structure with various degree of detail within such multi-dimensional data, a linguistic mechanism called *compounding* was mimicked in the paper introduced in Chapter 12. In the presented technique, common sequences of symbols/words are merged iteratively yielding new words, similar to natural language. Using this method, a given data set is automatically segmented hierarchically, revealing structures and their substructures in an unsupervised manner.

Part IV: Applied Data Analytics in Cyber Physical Systems. The publications within the above mentioned parts are mainly introducing new data science concepts and methods with focus on including the physics of the system within the models and subsequent calculations. In this part, publications which address applications throughout the full data science cycle are collected.

Using a sensor to observe the behaviour of a system already includes the first assumption for data analysis (i.e. indicator hypothesis), since the data emanating from this specific sensor is tagged implicitly as an important source of information – otherwise the data would not be collected. Although the obtained data may not contain significant information all the time, sensors build the major source of information within cyber physical systems. Thus, the precision of the sensor is linked to the precision of subsequent computations and results. In Chapter 14 the precision and characterization of inclinometer sensors with two sensing elements in opposite directions are investigated. As a result, a not perfect alignment of the sensing elements was found by analysing the bivariate histogram of the individual signals. Additionally, it was discovered that the distribution of the perturbations are well modelled by a Cauchy-Lorenz distribution, which must be taken into account in further computations and considerations.

The idea of using the information content for segmenting a stream of data is introduced in the paper presented in Chapter 15. Shannon's entropy is used to detect regions of interest within production processes, e.g. drilling or milling. Furthermore, time varying histograms are used to detect changes of the system in observation. For example, this enabled the detection of a tool malfunction during milling by analysing the force signals observed on the tool holder.

Several use cases for the data analysis framework to data emanating from raw materials handling machines are presented in the publication within Chapter 16.

Additionally, the main structure of a data collection and analytics framework as well as a structured approach to data analytics is presented. Exemplary data evaluations show the added value during the full life cycle of the machines.

Applying the data analytics framework to analyse the parallel hydraulic system of mining machines is presented in the paper introduced in Chapter 17. A defective sensor was identified using time-varying histograms. Additionally, investigating the statistics of the signals revealed the presence of negative pressures within the system, indicating cavitation. Avoiding this behaviour is to be considered in future designs of the machine.

Part V: Discussion and Appendices. Within Chapter 18 the insights gained during the herein presented research are used to draw a conclusion and give a direction for possible future research.

This thesis closes with the appendices which comprise the list of figures, the list of tables, a complete list of the author's publication as well as a list of references¹.

1.3 Contribution

The main part of the thesis is built by the ten papers included and addressed herein. Besides that, an extensive introduction for a structured approach and the needed environment to perform data science in large cyber physical systems is given in Part I. In Table 1.1 the contribution of the author to the peer-reviewed papers is summarized. With the progress of time, the contribution has moved increasingly from contributing to primary author.

As visible in the structure of the thesis, the papers can be grouped into three areas: Polynomial Methods [P1–P3, P8], Symbolic Time Series Analysis [P4, P5, P9] and Applied Data Analytics in Cyber Physical Systems [P6, P7, P10, P11]. A detailed breakdown of the areas can be found in Section 1.2.

¹This list does not contain the literature cited within the papers, since each paper includes its own list of references.

Tab. 1.1 Contribution of the Author to the publications collected in the thesis in percent.

Paper	Chapter	Conception and planning	Experiments	Analysis and interpretation	Manuscript preparation
[P1][P2] ²	8	80	100	90	95
[P3]	7	80	100	90	95
[P4]	12	85	95	95	95
[P5]	11	75	95	90	95
[P6]	17	60	50	60	30
[P7]	16	50	50	50	40
[P8]	6	40	40	40	30
[P9]	10	20	20	25	20
[P10]	15	30	30	35	20
[P11]	14	25	25	30	25

1.4 Remarks

Within this thesis two types of citation marks are used:

1. **Numbered citations, e.g. [1]:** This form of citation marks is used to cite the literature listed in the References section at Page 172 ff.
2. **Prefixed and numbered citations, e.g. [P1]:** The Prefix “*P*” within citation marks indicates papers with significant contribution of the author. These papers build the main part of this thesis and are presented in separate chapters. A complete list can be found in the List of Author’s Publications at Page 170 ff.

The author’s papers are included as “stand-alone” documents in their final version. Consequently, each paper has its own bibliography. Note: This literature is not included within the References list at Page 172 ff. of this thesis.

All the author’s papers (except the preprint [P1]) are peer-reviewed. This preprint [P1], in a slightly modified version, has passed peer-review and is accepted for publication as [P2].

²The paper [P2] is accepted for publication but not published at the time of submitting this thesis. The preprint version of this paper is [P1].

Part I

Prerequisites

2 | Data Science in Cyber Physical Systems

The presented thesis deals with a broad range of topics spanning data collection, data structures and data analytics from different technical fields, especially large physical systems. This is summarized in the title of this thesis *Methods and Framework for Data Science in Cyber Physical Systems*. Since there is no common body of knowledge for the terms in use, the following sections summarize the definitions given in literature and point out the relevance to this work.

2.1 Data Science

Data Science is nowadays a widely used term in academia as well as in industry. Various definition can be found in literature, e.g. [2–7], fitting to certain scopes.

The most general definition, which may explain the popularity of data science, is given in the description of the *Journal of Data Science* [5]. It states that data science is almost everything dealing with *data* spanning: data collection, data analytics and data modelling. Although this “definition” includes the topics dealt within this thesis, it is rather nebulous.

Several authors of the above cited papers try to give more precise definitions for the abilities important in data science. They state that data science is a multidisciplinary field which needs a depth knowledge in various areas. Therefore, it is normally approached by a team to cover all aspects [2].

Common core components important for data science and relevant for this thesis are:

Mathematics: This is the most basic skill needed for data science. It gives you the profound basis to approach, model and solve problems.

Statistics: This component is important to characterize given data as well as to identify correlations and may predict the future. It is heavily used in *Big Data Analytics* to build the basis of information [8]. It is important to not rely only

on statistics since correlation is not a measure for causality. Especially when working with large physical systems, a causal link to the physics of the system is important, which is done by using mathematical models, i.e. differential equations. Furthermore, the difference between *uncertainty* and *confidence* needs to be addressed within this area.

Computer Science (often referred to as programming or hacking skills): This includes the efficient implementation and use of mathematical and statistical methods and algorithms, the use and knowledge of how data are stored and handled, and the use of special tools, e.g. machine learning, artificial intelligence and optimization¹. In addition, the task of visualizing information and knowledge to transport insights belongs to this area.

Domain-Specific Knowledge (also named as *Substantive Expertise* [2]). It is important to include the available knowledge and all boundary conditions within data analysis and to formulate and reduce the scientific question to meet the required needs. This yields inverse problems to be solved in a regularized manner, due to the nature of the data addressed herein.

Creativity: Since the information hidden in data is not straight forward to retrieve, creativity is needed to combine knowledge from various fields to transform the data to reveal the needed insights.

A summed up definition for those components is given by Dedge Parks in [3]. He defines data science as a methodology using statistics, scientific rigour and systemic capabilities to ensure that an answer to a data question is accurate. A similar definition can be found in [4], which states that

... “data science” refers to the statistical, technical, and domain-specific knowledge required to ensure that the analysis is done properly.

Dhar in [6] defines data science as

... the study of the generalizable extraction of knowledge from data.

As it can be seen, depending on the context data science is used in, the focus of the definition changes. A big difference is also visible whether the term is used in academia or in the industrial environment. This can be found in the extensive study performed in [2].

The above definitions cover what the author refers to as data science in this work. Summarized, in this thesis mathematics and statistics are used to develop and implement efficient algorithms to extract knowledge and understanding in a specific domain (large cyber physical systems) and transport this knowledge to others (using data visualization) to generate added value.

¹Machine learning and artificial intelligence are not considered within this thesis, because it is beyond the scope of this work.

2.2 Cyber Physical Systems

Although, *cyber physical systems* (CPS) found their way into curricula of universities, development plans of governments and scientific communities, there is no unified definition what constitutes a CPS. Following, a short review of various definitions is given, which are important to this work.

As the name *cyber physical system* suggests, the most general definition, which builds the basis in literature, is: A CPS is a system with a strong coupling of *cyber* aspects (including hardware and software) with the physical aspects from systems [9–11]. In other words, computational systems (virtual world) work together with physical systems (real world) to improve efficiency. The cyber aspects are often referred to as computation, communication and control [12–16]. The physical aspect deals with physical processes which are observed by sensors and controlled by actuators [10, 11, 17, 18]. Therefore, a CPS often contains a feedback loop (control). In this manner, the physical processes affect the computations and vice versa [15, 17, 19–22].

Kagermann *et al.* in [22] state that communication is not only between the physical system (physio-space) and the cyber-system (cyber-space), it also effects the socio-space (the social environment), since CPS contain various human-machine interfaces [16, 17, 22] to interact with [23]. Sometimes the interaction with humans is seen as a major part of the system, i.e. human in the loop [10].

A stronger definition for CPS includes that a CPS is a complex system, which consists of multiple subsystems (each with a closed control loop), which interact and communicate within a network (wired or wireless) [15–17, 22]. Therefore, a CPS not only uses information given directly, but requests mutual information from other connectors or the *internet of things* (IoT). Huang *et al.* in [14] and Liu *et al.* in [12] state that CPS can realize real-time perception and dynamic control of multi-dimensional complex systems (CPS networks).

To establish the strong coupling between the cyber-space and the physio-space, CPS are often seen as embedded systems with a communication core and extended capabilities, e.g. efficiency, safety, complexity [9, 11, 15, 18, 22, 23]. Additionally, some authors do not restrict a CPS to be within a local network. They state that a CPS uses data available worldwide (global IoT). Therefore, the CPS should also be able to store data [16]. It is important to notice that data analytics (computational aspect of CPS) is performed directly on the CPS [17], which is also referred to as decentralized control [24].

A good definition of what constitutes a CPS can be found in [20, 25]. The authors defined that the core goal of CPS is the study of the joint dynamics of physical processes, software and networks, since it is about the intersection and not the union of the cyber- and physical aspects. In this definition the importance of the physics of the system is addressed. Lee in [19] described this as:

In the physical world, the passage of time is inexorable and concurrency is intrinsic.

Park *et al.* in [15] state that a CPS should incorporate the key characteristics of the application domain within the computations. This is also seen by others, who point out the importance of the physics of the system being included within computations to establish a causal link between observations and the cause [26, 27]. The definition they gave is:

A cyber physical system is a system with the coupling of the cyber aspects of computation and communications with the physical aspects of dynamics and engineering, that must abide by the laws of physics.

This definition is the one which is used throughout this thesis, as the importance of including the physical behaviour of the system in observation is essential. Although CPS are seen as a subclass within IoT and Industry 4.0 [11, 23, 25], the fact that a CPS includes a physical system differentiates them clearly from IoT and Industry 4.0. Therefore, a major task in CPS (which is also addressed in this thesis) is to solve inverse problems, since performing measurements are fundamentally inverse problems, especially if the dynamics of the system is modelled using a causal link.

A more formal comment for modelling the physics of the system is given by Letichevsky *et al.* in [28]. They described that the basic mathematical models for CPS are build by hybrid automaton, which describe continuous dynamics as a linear or piecewise linear problem. Linear inequalities are then discrete transitions – a change in the behaviour of the system. This idea is also formulated by Lee in [20, 21], who described that the dynamics of the real world is reduced to sequences of state changes without temporal semantics in the cyber world. Within this thesis the temporal semantics is not neglected. It is taken into account by using symbolic time series analysis [P4].

In general, the possibility of transforming the real world to the cyber world by introducing physical models enables the exchange of mutual feedback (by solving the associated forward and inverse problems). This builds the basis of a digital twin [29].

The applications for CPS can be split up into six groups [17, 18, 30]:

1. infrastructure and mobility,
2. health and living,
3. energy and resources,
4. production and logistics,
5. monitoring and control,
6. military and defence².

²The author wants to strictly distance from CPS used in military.

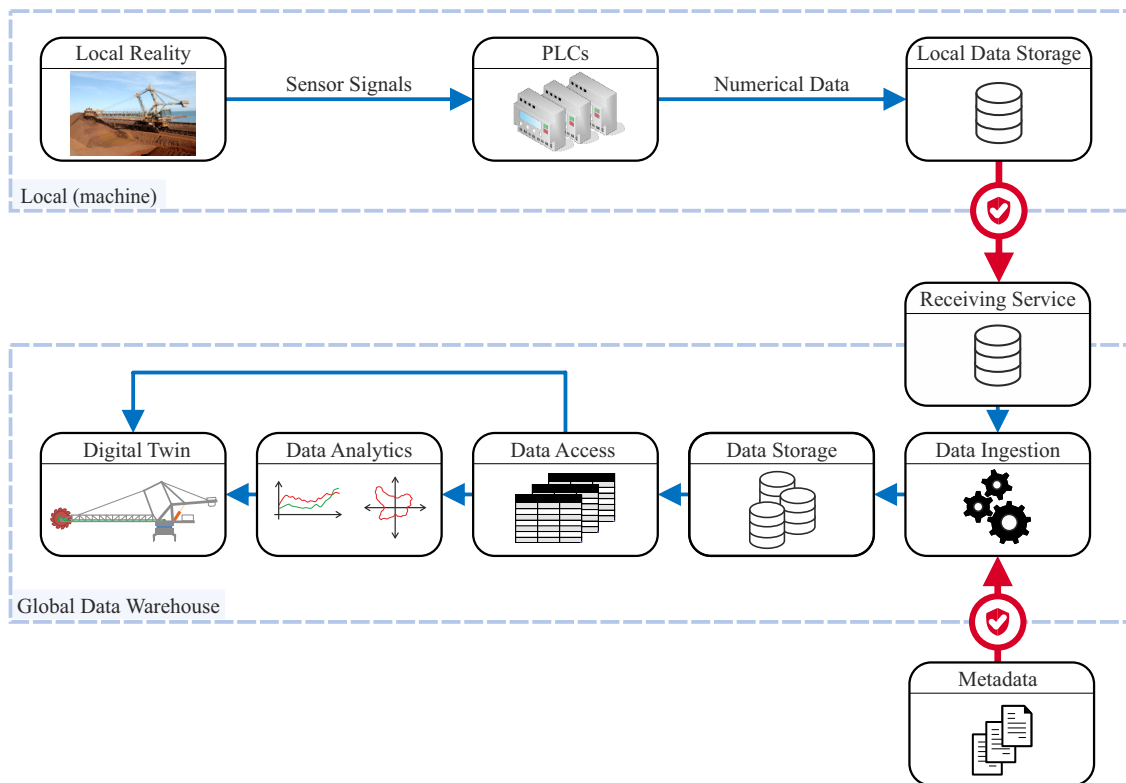


Fig. 2.1 Digital Twin for Large Physical Systems

The presented thesis is within the scope of monitoring and control which is at some points extended to the field of production and logistics and the field of energy and resources. The consistent modelling of the system dynamics builds the basis for computations done within the CPS. A numerical efficient design of the herein developed methods and algorithms is sought to make them suitable for near real-time computations. Additionally, the interaction with the global world, i.e. global data warehouse, is established (see Chapter 3). Therefore, the CPS implemented in this work form digital twins (see Part IV). The structure is shown in Fig. 2.1. The physical behaviour of the system is observed by sensors, sampled in real-time using a programmable logic controller (PLC). The data is collected and transformed using a local device and put into a local data storage (data base). The local device performs computations and interacts with the machine, e.g. trigger alarms. Using communication techniques, the data is mirrored to a global data warehouse to make them available to other services, e.g. automatic report generation, data analysis or data on-demand. A more detailed explanation is given in Chapter 3.

2.3 A Structured Approach to Data Analytics and Knowledge Discovery

During literature research it became clear that the terms *data mining* and *knowledge discovery* are used in the same context [31]. Again there are no clear definitions for these terms.

Since in our understanding the goal of both, data analytics and data mining, is to extract knowledge from given data, the following sections address this issue in a structured manner. The aim is to include models for the physics of the system within the computation. This is an issue which is currently insufficiently addressed in literature.

2.3.1 Knowledge Discovery Process Models

In literature various knowledge discovery process models exist (e.g. [32–40]), which are mainly used for commercial data with the goal of extracting knowledge in a structured way. A good overview can be found in [41]. The author pointed out that the cross-industry standard process for data mining (*CRISP-DM*) [34, 42] builds the basis for the generic structure he found.

Although the CRISP-DM makes no proposals how a specific task can be performed, it does have value because of its generic nature, i.e. it describes successfully the generic processes which need to be dealt with, independent of the nature of the project being addressed. It has more the nature of being a *reminder* of what should be not forgotten. The six process phases, their description and their generic tasks given by Kurgan *et al.* and Chapman *et al.* in [41, 42] are:

Business Understanding:

Description: Understanding of business objectives and requirements, which are converted into a data mining problem definition.

Generic tasks: Determine business objectives, assess situation, determine data mining goals, produce project plan.

Data Understanding:

Description: Identification of data quality problems, data exploration and selection of interesting data subsets.

Generic tasks: Collect initial data, describe data, explore data, verify data quality.

Data Preparation:

Description: Preparation of the final dataset, which will be fed into data mining tools and includes data and attribute selection, cleaning, construction of new attributes, and data transformation.

Generic tasks: Select data, clean data, construct data, integrate data, format data.

Modeling:

Description: Calibration and application of data mining methods to the prepared data.

Generic tasks: Select modelling techniques, generate test design, build model, assess model.

Evaluation:

Description: Evaluation of the generated knowledge from the business perspective.

Generic tasks: Evaluate results, review process, determine next steps.

Deployment:

Description: Presentation of the discovered knowledge in a customer-oriented way. Performing deployment, monitoring, maintenance, writing final report.

Generic tasks: Plan deployment, plan monitoring and maintenance, produce final report, review project.

The generic tasks and their interaction is shown in the CRISP-DM reference model in Fig. 2.2. As it can be seen, knowledge discovery is a cyclic and iterative process. This process can be taken as a basis and adapted to mining sensor data in CPS.

Although, one of the most valuable aspects of CRISP is the clear description of the tasks together with their outputs [34, 42], some very fundamental issues are ignored in the CRISP model. The first issue is that before starting knowledge discovery one must determine what data needs to be collected to ensure that sufficient information is available to establish semantics. Secondly, the most serious issue is: that one must design and install data collection prior to starting this process. Furthermore, in physical systems metadata has a greater significance than in evaluating commercial data since the sensor data has no meaning without the metadata.

In addition, in many companies the task of *business understanding* is not emphasized sufficiently. As a consequence there is commonly a divergence in expectations as a project proceeds. However, there is a *chicken and egg* situation when mining sensor data from CPS, i.e. it is not possible to determine a-priori what can be achieved or what the data will reveal. Alongside this the question of how success is

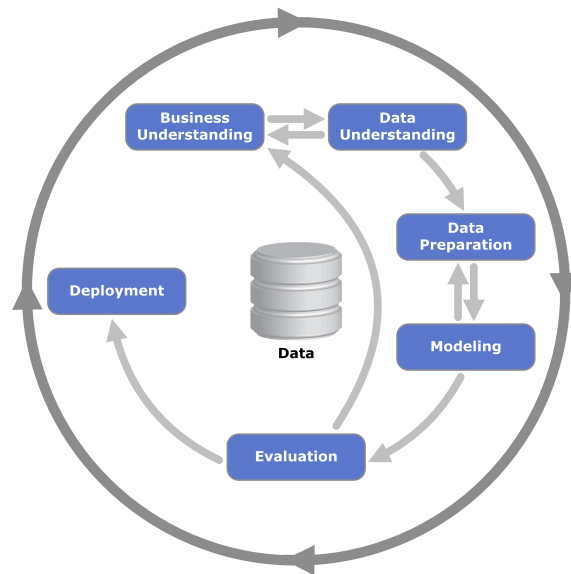


Fig. 2.2 CRISP Data Mining Cycle [42]. Image by Kenneth Jensen [43], distributed under a CC BY-SA 3.0 license³.

measured is often not addressed. This question becomes important in the evaluation step, especially when there is no “training data” available. This issue should be investigated at the very beginning of a knowledge discovery task.

Additionally, there is the fundamental question of whether an associated inverse problem can be solved, i.e. there is a significant difference between explanatory models (e.g. explaining an incident) and predictive models (e.g. predict an incident). For this very reason, performing an extensive feasibility study on CPS prior to committing to a major data mining system development is recommended. This involves exploratory data analytics, a topic which is also within the scope of this thesis.

2.3.2 From Data to Knowledge

As mentioned above, the described knowledge discovery process deals with the entire business perspective. In literature the special nature of the sensor data (its relation to physical systems) is rarely taken into account when performing data analytics. Present data mining techniques mostly rely on correlation (in some manner) being a reliable measure for significance. However, the solutions computed from the sensor data should/must obey the equations modelling the physics of the system being observed – this is fundamentally an inverse problem and requires the modelling of the system dynamics. Unfortunately, the issue of inverse problems is not addressed in literature on *mining sensor data*, see for example [44–47].

³<https://creativecommons.org/licenses/by-sa/3.0/deed.en>

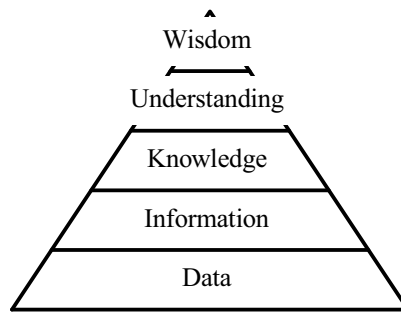


Fig. 2.3 Data Mining Pyramid as proposed by Mark Embrechts [48].

In a proper approach the inverse solution of the model-equations is required for the digital twin to establish the semantic reference between the sensor observation and its cause. Without this semantic reference to causality there can be no physical based knowledge discovery. Data analytics from CPS is still a research topic, for which there are only a few recognized standard procedures and there probably will not be any new standards in the near future, since the modelling required is application and domain specific.

To overcome this issue, this thesis follows a structured data analytics approach which is also presented in [P7, P9] but revisited here for consistency. This approach is based on the work from Embrechts *et al.* [48]. The authors proposed the pyramid of data mining as shown in Figure 2.3, which was an extension of Ackoff's work [49]. This pyramid is often cited in data mining, in particular in temporal data mining, as the valid structure for implementations. Embrechts offers no definitions for the terms *information*, *knowledge*, *understanding* and *wisdom* in his work, while Ackoff offers intuitive but rather nebulous inaccurate definitions. The pyramid and the terms used have positive connotations⁴; however, they do not provide a scientific basis for the implementation of mining sensor data.

Nevertheless, the hierarchy does provide a possible structuring for approaching the questions of what one wishes to extract from the processing of large data sets. Based on this data mining pyramid the fundamental premiss followed within this thesis is framed in the following section.

⁴ *Wisdom*, just as the word *creativity*, have positive connotations but resist any formal definition, see [50] for a discussion of this issue. Without formal definition they do not form the basis for objective data analytics.

2.3.3 Fundamental Premiss Behind Data Analytics in Sensor Data

In Fig. 2.4 the fundamental premiss behind data analytics in large physical systems is presented. Similar to the CRISP-DM model it forms a closed loop indicating multiple iterations, since in exploratory data analytics there is in general no straight forward way how to solve a certain problem. As one can see, the steps up to *understanding* from the data mining pyramid Fig. 2.3 are part of this loop.

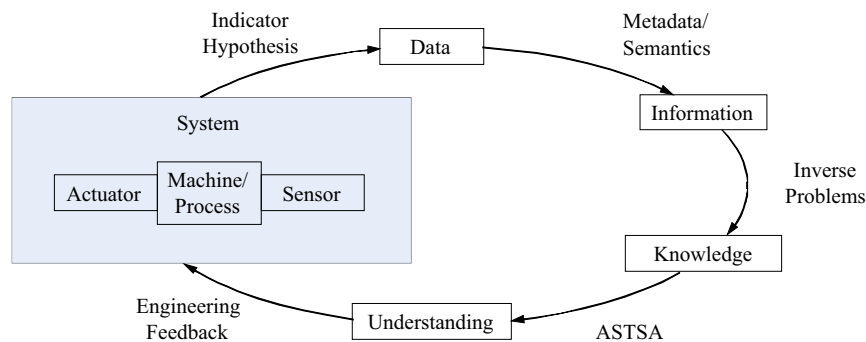


Fig. 2.4 Fundamental premiss behind data analytics [P7].

The relationships between the steps [P7, P9] are:

1. An indicator hypothesis is required, otherwise there is no basis for the collection of *data*. Selecting a specific sensor is already an implicit indicator hypothesis, i.e. conditions measured by the sensor are relevant. Therefore, one paper of this thesis deals with characterizing sensors to improve and support the indicator hypotheses [P11].
2. The output of the data acquisition is simply a stream of numbers. Metadata is necessary to add *meaning* to the data resulting in *information*. Furthermore, *context* is additionally required to define *significance*, e.g. a temperature measurement of $T = 39.8^\circ\text{C}$ has a different significance if it is the temperature of hydraulic oil or human body temperature – clearly a strong fever. This topic together with data handling is addressed in [P7] and Chapter 3. Additionally, Shannon [51] provided a mathematical definition for information content. Although there is no causal link to *significance*, this idea can be used for a first segmentation of the data by identifying points where the information content changes. This idea is used in this thesis in the paper [P10].
3. To establish semantics based on physical results, a causal link between the observation (measurement data) and its possible cause must be built. This requires system models and the solution of the corresponding inverse problems⁵. The results of the inverse solutions are dubbed *knowledge* in this context.

⁵In general, inverse problems do not have unique solutions. It is necessary at this point to embed a-priori knowledge into the system to ensure that the desirable solution is found.

Inverse problems and their solution are addressed in Part II of this thesis in the papers [P1, P3, P8, P9].

4. To gain *understanding* of the behaviour of the complete system, the effects of the human *operator-machine interaction* must be included. Human behaviour is not strictly deterministic. Thus, human operated machines are hybrid systems, since stochastic physical processes are combined with non-analytical human interaction. For this reason a new research approach is proposed based on the emergence of language as modelled by the philosophy of phenomenology. This process is called advanced symbolic time series analysis (ASTSA) [P4, P5, P9], see Part III.

The basic approach is to assign symbols to actions related to derivatives – these symbols are likened to verbs; similarly states are modelled by symbols – nouns. Additionally, the actions and states are predicated with adverbs and adjectives. Finally, different pauses are likened to punctuation. In this manner the time series is automatically converted to a sequence of symbols, opening the door to the use of symbolic query methods to explore the data.

5. The extracted understanding can now be fed back into the whole process as engineering feedback. Thus, the next level of understanding can be extracted in the next iterations by improving the data collection and/or the monitored system [P6, P7, P10, P11], see Part IV.

To follow this premiss, the necessity for a data collection and data analysis framework is given to support the mentioned transitions. These topics are addressed in Chapter 3 and Chapter 4 of this thesis.

3 | Data Ingestion Framework

To perform data analytics on data from different types of machinery with the focus on extracting information rather than data handling and manipulation, a structured approach to collect and store data needs to be established, which is partially introduced in the papers [P7, P9]. This is the step prior to establishing a local data analytics framework to actually work on the data, which is described in Chapter 4. The structured data handling is often underestimated, but an important prerequisite. Since the focus of this thesis is on data analytics, the following data handling structures are described conceptually with the aim of transporting the main ideas rather than being a complete implementation and specification guideline.

The first step to establish a working data collection and handling system is to define the data flow, with a focus on data security. The data is normally collected on site directly from the controlling device, i.e. from a programmable logic controller (PLC). This is done using an edge-device, which collects the data and transmits it over a secure channel to the global data storage. This is described in Section 3.1.

The structured ingestion of time series data collected on site into the global infrastructure is crucial. Therefore, Section 3.2 explains the steps implemented and used to feed the data (analysed in this thesis) into such a structure. The data is quality checked and after authentication the data is transferred to its own virtual destination.

To support data analysis, a contiguous data model is used to provide data on-demand. Therefore, various data storage models are possible, each with its own benefits. This is investigated in Section 3.3.

3.1 Data Flow Structure

This section deals with the aspects of how the data from machines and cyber physical systems (CPS), information and results of analyses are passed through the system. Since CPS are stand-alone devices, which are often globally distributed, the data transmission is mostly wireless. This fact, and the fact that the emanating data probably contains sensitive informations, a special focus in the design is laid on data security.

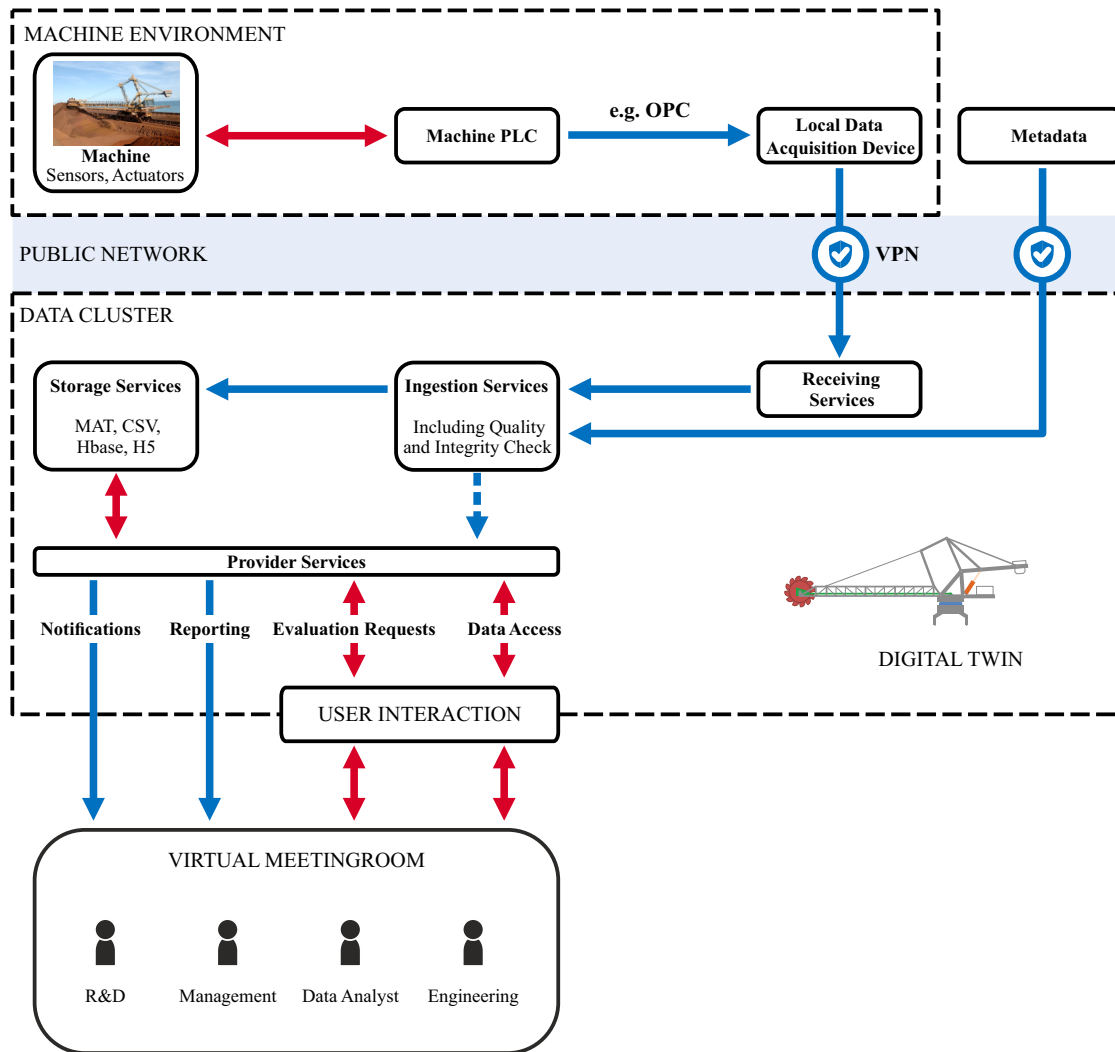


Fig. 3.1 Data Flow and Ingestion.

The entire data flow on a global scale is shown schematically in Fig. 3.1. Starting on site, the data emanating from one machine is collected by a local industrial edge-device (or an industrial PC; iPC). Since a large CPS is normally equipped with a PLC, which collects and processes sensor data, the most common way is to use an interface such as OPC-UA¹ [52] with a publish and subscribe mechanism (which will be explained in more detail in the subsequent section). If no PLC is available the data may be delivered by smart sensors or directly acquired by sensors attached to the iPC.

Within the iPC the data is collated – and if implemented – preprocessed. Thus, the iPC can act as part of the CPS. In the next step, the data is transmitted over a secure path to the *data receiving service*. This service is located at the data processing centre and acts as the entry to the global data warehouse. It is used to receive data

¹<https://opcfoundation.org/about/opc-technologies/opc-ua/>

OPC-UA is an open-source standard, which is acknowledged and used widely in industry. Therefore, it is often used to replace proprietary systems and protocols.

from numerous machines located on multiple sites from various customers. It is the only entry-point to get machine data into the data warehouse. Thus, this is the point of interest for cyber-attacks and needs to have a high level of security. This is established using encryption and certificate exchange. Additionally, the receiving service only accepts data from known IP-addresses, which is an additional security layer. Each time a data receive is triggered, a virtual server instance (*receiving server*) is booted up which subsequently performs the authentication, i.e. the certificate check.

If the data is authorized, it is forwarded and put onto the *raw data* partition of a dedicated storage. For each machine (each closed entity) there is a separate storage. The pointer to the correct location is part of the certificate. This partition is exclusively served by the receiving server. This adds the next level of security.

If new data is deposited on this *raw data* partition, the *data ingestion service* is triggered. This again boots up a new virtual server instance (*data ingestion server*). This server now takes the deposited data, merges it with the metadata, optionally performs preprocessing steps and deposits the result either file based on the dedicated data partition or in the data base of this machine. A detailed description of the data ingestion can be found in Section 3.2.2.

Access to the data is granted via the *provider services*. An important service is the *data on-demand service* to request data to process it locally (see Chapter 4). To perform the work done in the thesis, this is the mostly used service. Further, services which are an optional part of the provider-services include, e.g. *Evaluation Requests* for incident analysis and automatic evaluations, *Reporting* to automatically provide daily/weekly/monthly reports, and *Notifications* to trigger warnings and alarms in case of unwanted behaviour of the system.

3.2 Data Acquisition and Ingestion Process

During the work with companies starting data analytics on a global scale it became clear that the most security concerns arise in regard to the transmission of the data from the machine to the cloud, the main fear being that an attacker takes over the control of the machine. This section gives a detailed description of how the data acquisition was designed and used to collect the data analysed in this thesis.

3.2.1 Data Acquisition

Since the large cyber physical systems monitored in this thesis are controlled by PLC, the first step of data acquisition is to collect the data from the PLC using a local industrial edge-device, i.e. iPC. OPC-UA [52] is used to implement a mono-directional data transfer using the publish and subscribe mechanism [53]. This ensures that the iPC, or an unwanted attacker who is in control of the iPC, can't write data or malicious code to the PLC. Thus, the data collection device cannot be used as a back-door to control the machine. In addition, only the data which is *published* on the PLC can be collected from the data collection device. Using this mechanism, secure data is not visible outside the device if not published.

The collected data is now mapped to a local data base (e.g. SQL) on the iPC. Subsequently, batched files (e.g. *.**csv** files, *.**json** files) are generated. These files contain the collected data either in a *on-change* or *full-table* format.

On-change: The data record consists of the triplets [**time-stamp**, **sensor tag**, **sensor value**], which is generated each time a sensor value changes. The single records are appended to form a list of records. This format is efficient for systems such as ship loaders. These types of machines have long time periods, where sub-systems and families of sensors do not change and/or are not active. As a result, only a few records for the active sensors are generated.

Full-table: Full table data is shaped, as the name suggests, like a table. For each time-stamp, the sensor values for every sensor in the local system are collected. One time-stamp forms a row of the table. Therefore, each column of the table represents a single sensor in the system. Optionally, the header line contains the sensor tags. This type of files is more efficient if most of the sensors within the monitored system change continuously. If on-change records are used in such a case, a record would be generated for each sensor for only a single time-stamp. This would clearly include much more overhead due to the fact that the sensor tag and the time-stamp is included in each record.

Note: in both cases the sensor tag can be a hashed value so as to exclude information of the sensor within the data files. In this manner, one can only see numbers without association to the sensor. Thus, the attacker cannot interpret the data, since no information on the meaning of the data is contained within the data file. Consequently, no knowledge discovery is possible.

To add a physical layer to security, two separated Ethernet ports are used on the iPC: one is connected to the PLC to collect the data using OPC-UA, the other port is used to connect to a modem through a firewall. The two ports ensure that the internal network used to control the machine cannot be inhibited by external network traffic. In this manner, *Denial of Service* attacks do not affect the operating machine.

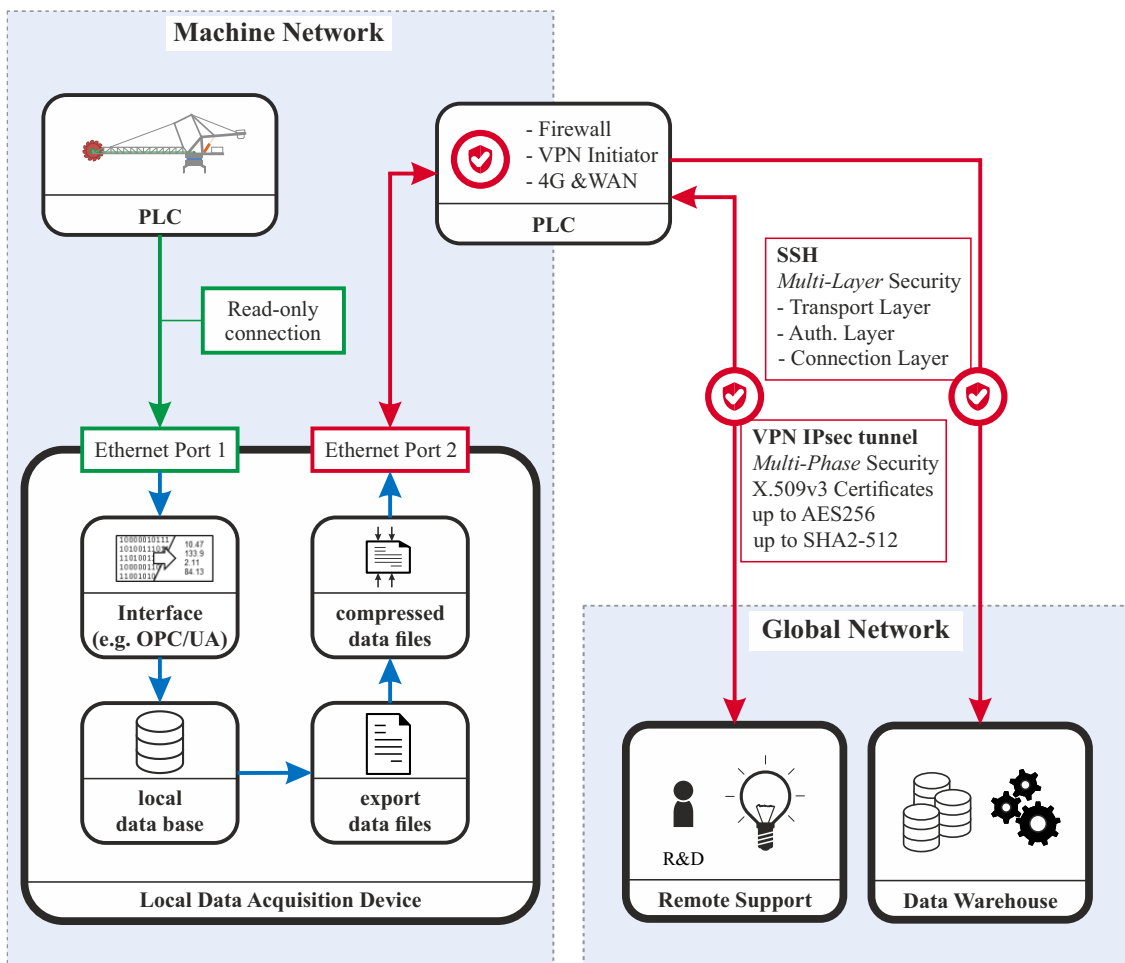


Fig. 3.2 Data collection and secure data transfer.

Subsequently, the modem transfers the data encrypted and mono-directional to the global data warehouse (cloud). This is done using a secure multi-layer connection including certificate exchange. As described above, the receiving service of the global data warehouse is now responsible for the next steps.

The full architecture of the data collection is shown schematically in Fig. 3.2.

3.2.2 Data Ingestion

As mentioned above, after the data is received and the authentication is performed (which is provided by the data receiving server), the raw data is put onto the dedicated raw data storage partition and triggers the ingestion service. Archiving the raw data enables a regeneration of the processed data at a later date if required for any reason.

In the first step, the data ingestion service converts the incoming data to a full-table matrix since multi-dimensional time series are most suitable for further calculations. During this step the data is checked for consistency using the metadata. The metadata

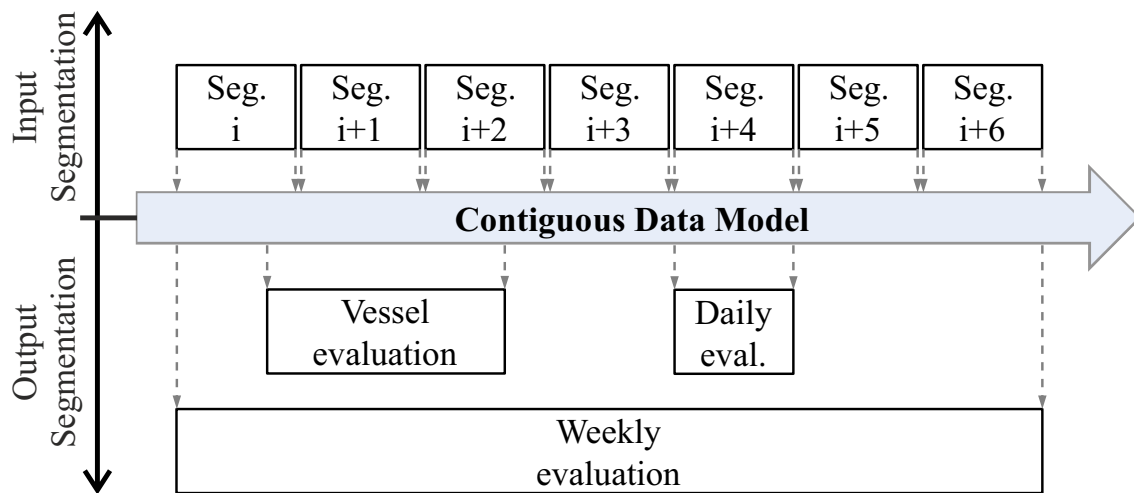


Fig. 3.3 Contiguous Data Model.

includes the sensor definitions, such as the sensor names, sensor ID, description, limits and units. If a sensor is not defined within the metadata or a value exceeds the predefined limits, a notification is triggered. Additionally, if data is sampled at a fixed sampling rate, the consistency to previous data is checked to support the contiguous data model (described below).

If the consistency checks are successful, the data is put on the dedicated data partition of the specific machine. This is either done file based (e.g. `*.csv`, `*.mat` or `*.json`), i.e. data from one day/week/month are within one file, or on a data base system, e.g. SQL. The benefit of the file based system is the quick data on-demand in case the full range of a file is requested.

3.3 Contiguous Data Model

Within the global data warehouse, a contiguous data model is established for each individual machine to enable data on-demand. Independent of the input segmentation, the data forms a contiguous data stream on the data server. This enables a separated output segmentation, which can be used to automatically perform, e.g. daily or weekly reports, trigger evaluations on a vessel by vessel basis in case of a ship loader, or return data from a variable time span. Additionally, this supports the data scientist working on the data, since he can focus on the data itself and does not have to care about data handling and the underlying storage system nor the storage structure. The contiguous data model is shown schematically in Fig. 3.3. As it can be seen, the data on the warehouse appears as a single stream of (multi-dimensional) time series data for the actual user. This contiguous data model is a prerequisite to establish a structured access to the data and to implement data on-demand services.

3.3.1 Data On-Demand – Interface to the Global Data Warehouse

To support fast exploratory data analysis, data on-demand services together with the local data analytics framework need to be established. Therefore, an interface to the global data warehouse is implemented, which mirrors the data from the server to an instance of the local data model (**MdtsClass**, see next chapter). Since data from multiple machines from various companies on various sites are available, a structured request for data is necessary.

In the presented framework, the request for data (i.e. RESTfull) consists of the following input-parameters:

Company:

The name of the company to which the machine data belong to.

Site:

The name of the site the machine is located on.

Machine name:

A unique name of the machine the data is emanating from.

Start time:

The time (UTC) the requested time period starts.

End time:

The time (UTC) the requested time period ends.

Tags:

The tags of the sensor channels to be requested.

The data is returned as a ***.json**, which is used to instantiate the local data handling framework (i.e. **MdtsClass**), as described within the next chapter.

4 | Data Analytics Framework

Along with the herein developed methods, a consistent data analytics framework was designed and implemented to support data analysis and knowledge discovery for cyber physical systems. This supports the data scientist in the handling of data and visualization of large multi-dimensional data sets. Additionally, the framework implements various newly developed methods and functionality as presented in later parts of this thesis. The code, which is implemented using MATLAB^{®1}, together with the documentation has been made available online². Since the data analysis framework is only supporting the main topics of this thesis (which is data analytics and the development of new algorithms), the following sections describe only the key aspects behind the framework to transport the basic concept rather than being a complete implementation and code documentation³.

The framework is designed and implemented using an object-oriented approach for most of the components. Unit testing and a distributed version-control system are used to support the continuous development. In general, the components can be grouped into the *Data Handling Framework*, supportive components (*Decorative Classes*) and *Additional Functions and Toolboxes*⁴.

The schematic class diagram in Fig. 4.1 shows the relations between certain classes and presents the core attributes and key methods.

¹Version 2018b — <https://de.mathworks.com/>

²<https://github.com/RolandRitt/Matlab-mdtsToolbox>

³Some names of classes/methods/attributes are modified slightly within this document for the ease of reading.

⁴Currently, not all the additional functions are yet available as source code in the public domain. They are being reviewed and it is expected that they will be made available in the near future.

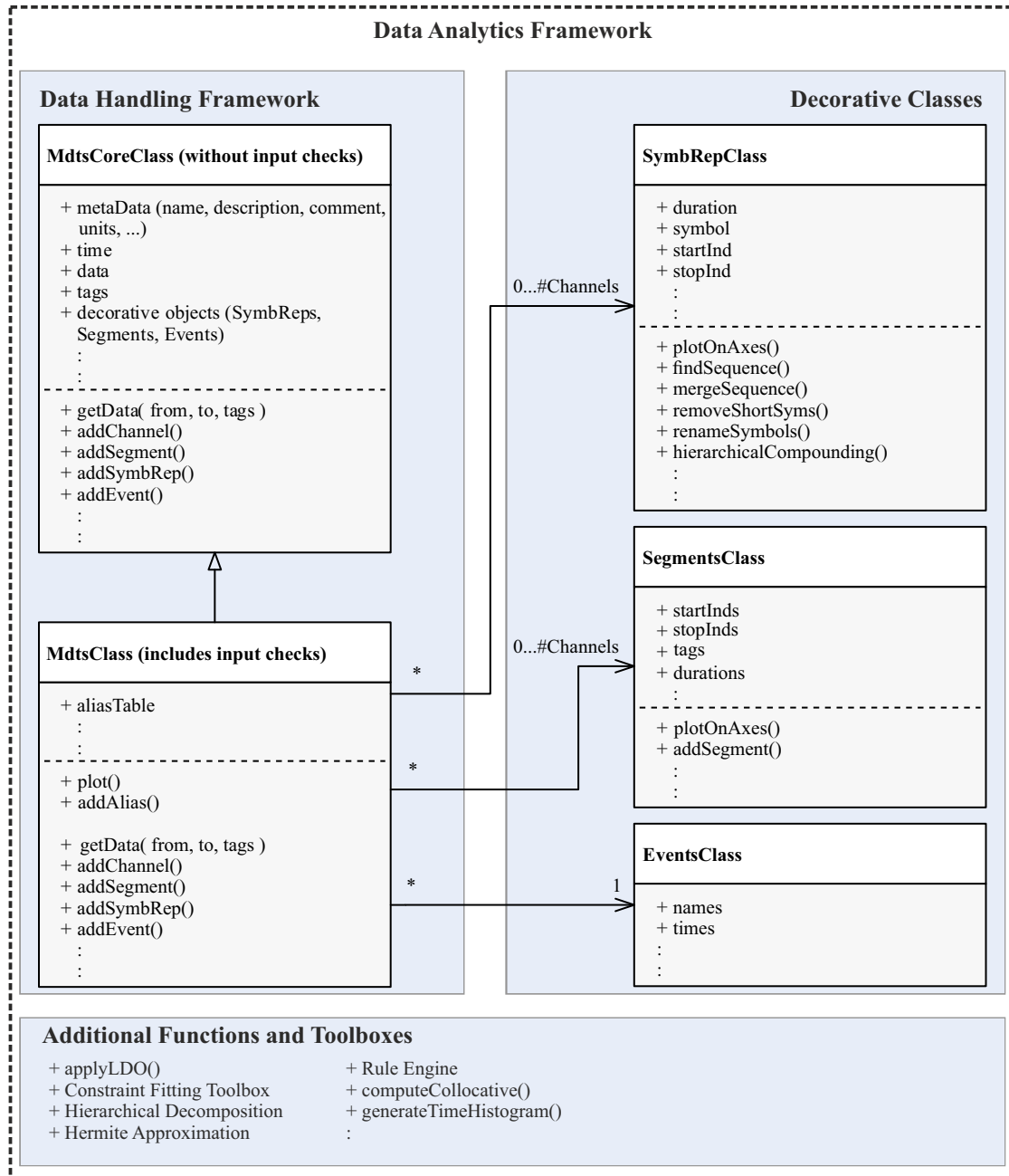


Fig. 4.1 Data analytics framework represented in the form of a schematic class diagram.

4.1 Data Handling Framework

The data handling classes form one of the core components of the analytics framework. It can be seen as a data container for multi-dimensional time series (mdts), with all the necessary functionality to easily interact with the data and to handle it.

The framework is split into two main classes: the core-class (**MdtsCoreClass**) and the user-class (**MdtsClass**). In general these classes have common attributes (i.e. *core attributes*) and implement the same functions (i.e. *core functions*), since the user-class inherits the core-class.

The core-class implements efficient data handling methods without input checks. This reduces the overhead but decreases the usability. Therefore, instances of these class are in general only used in special applications, with a focus on efficiency.

For the general use of this framework, where the overhead is not dominant, instances of the user-class are used. This is a wrapper class with the aim of simplifying usability. The class performs all the necessary input checks and the task of indexing prior to calling the methods from the core-class. Additionally, multiple helper methods and supplementary functionality are implemented in this class.

Splitting up the framework in this manner the focus of the implementation task is clearly defined, yielding a structured development.

4.1.1 Core Attributes

The mandatory elements of the data handling framework are: the **data** matrix, the **time** vector and the **tags**. The **data** matrix contains the raw data collected from the machine in observation. Each column of the matrix represents the observations of one sensor channel. Each row comprises the observations of all channels at a certain point in time, which is represented as a time-stamp. Therefore, the data matrix represents a multi-dimensional time series which is consistent with the contiguous data model on the global data warehouse. This structure is chosen to be consistent with matrix vector calculus. Each column (i.e. column vector) forms a time series in its own right, which can be directly used in calculations, e.g. apply linear differential operators or solve inverse problems.

Together with the data matrix, the **tags**, i.e. the identifiers of the individual channels, are used to index the columns of the data matrix. These are the names provided as metadata to the global data warehouse. Although the tags can contain useful information if a structured naming is chosen (e.g. if the tags contain information about the assembly group they belong to), a long tag name can be confusing and make the framework difficult to use. This can also cause problems if the same data analysis is applied to data from machines of the same type but different tag naming. To overcome this problem, aliases can be assigned to the tags (i.e. **alias table**).

Indexing a column can now be done using either the tag or its alias. Using this functionality, data analysis procedures can be implemented generically, using aliases instead of tags.

In the presented design, the time-stamps of the observations are collected in the **time** vector, which is separated from the data matrix. This vector is used to index rows of the data, e.g. to extract data spanning a certain time interval. Since the framework is mainly used with time series, various representations of the time vector are available, e.g. absolute time, relative time with respect to the first time-stamp or a number (e.g. UNIX time-stamps). The values of the time vector are in general used as abscissa values when visualizing time series. Therefore, this vector can be used to generate linear differential operators or (polynomial) function bases as presented in Part II and Part III.

Metadata, e.g. the name of the dataset, the units associated with the channels or a description of the machine the data is emanating from are stored along with the data. These informations are mainly used for the automatic visualization of the time series. Additionally, the metadata gives the data meaning.

Decorative objects can be added/linked to the time series to support enhanced analysis techniques, e.g. symbolic time series analysis. These objects are *Events*, *Segments* and *Symbolic Time Series Representations*. A detailed description of those objects is given below.

A visual representation of the above described attributes is given in Fig. 4.2.

4.1.2 Key Methods

The most important functionality of the data handling framework is the extraction of a subset of the data (**getData()**). Therefore, one has to define a time span (i.e. start and end) and the tags which should be returned. This returns a new instance of the class in use. This method is implemented for multiple types of indexes, e.g. direct indexing and time indexing for the rows and direct indexing, tags and aliases for column indexing.

The possibility to add new channels to the time series (**addChannel()**) is important, especially when performing collocative calculations. After adding a new channel, this channel can be indexed exactly the same way as the raw data.

To decorate the data, methods for linking decorative objects to the time series are implemented, e.g. **addSegment()**, **addEvent()** and **addSymbRep()**. These objects have their own functionality, but they are handled within this framework especially when subsets of data are extracted.

One of the most important methods is the visualization method (**plot()**). This is used to plot the multi-dimensional time series. A detailed description can be found

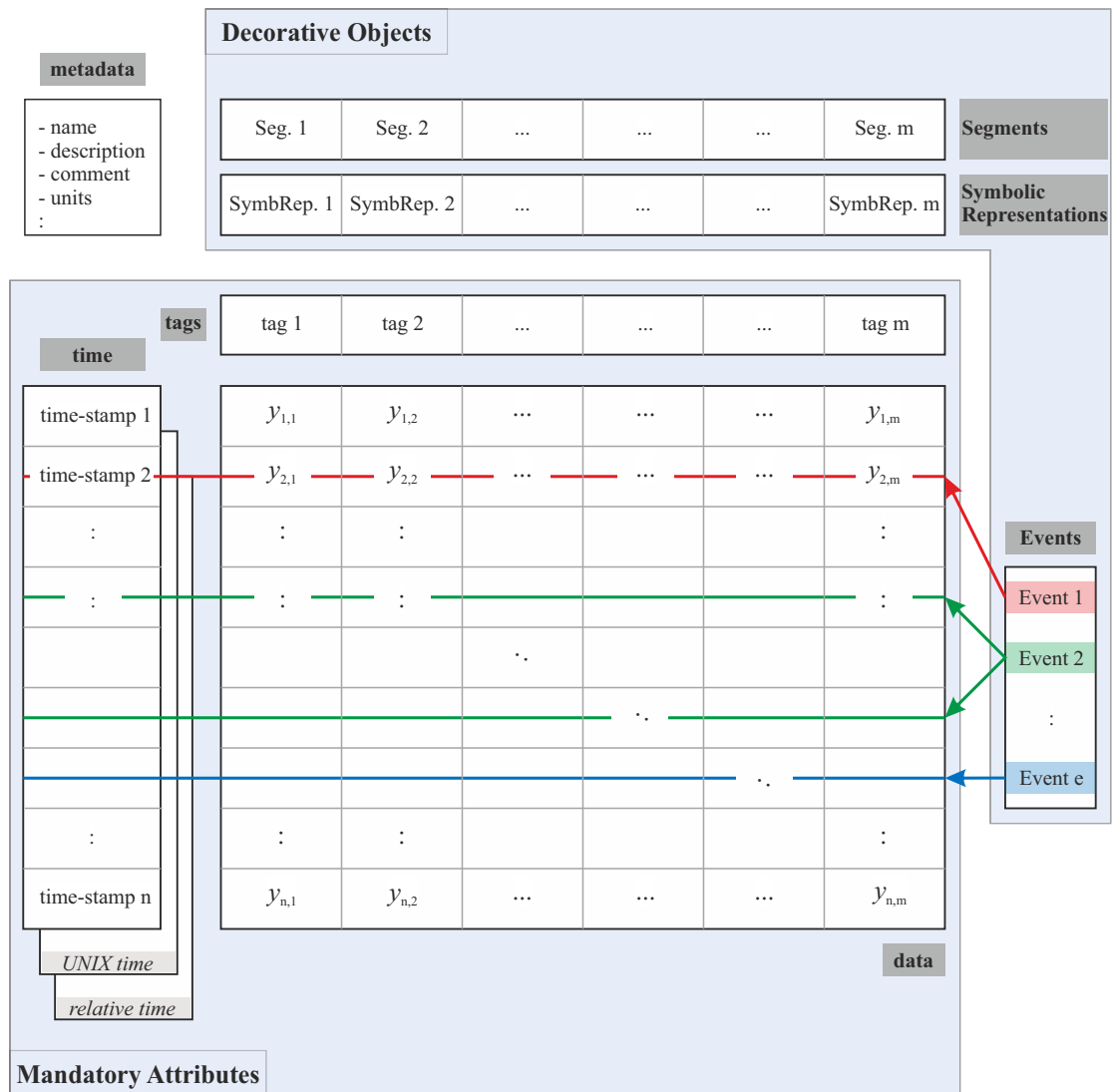


Fig. 4.2 Visual representation of the core attributes of the data handling framework.

below in Section 4.4.1. This method handles the visualization of the time series along with the visualization of the decorative objects.

4.2 Decorative Objects

Decorative objects⁵ are instances of additional classes, which are linked to the time series to expand either the functionality and/or contain additional information, i.e. other representations of the time series. Decorative objects comprise *Events*, *Segments* and *Symbolic Time Series Representations*. These are either known a-priori (e.g. certain events) or emanate as the result of computations, e.g. applying the rule engine delivers segments and transforming a time series into a symbolic time series as proposed in Part III results in a symbolic time series representation.

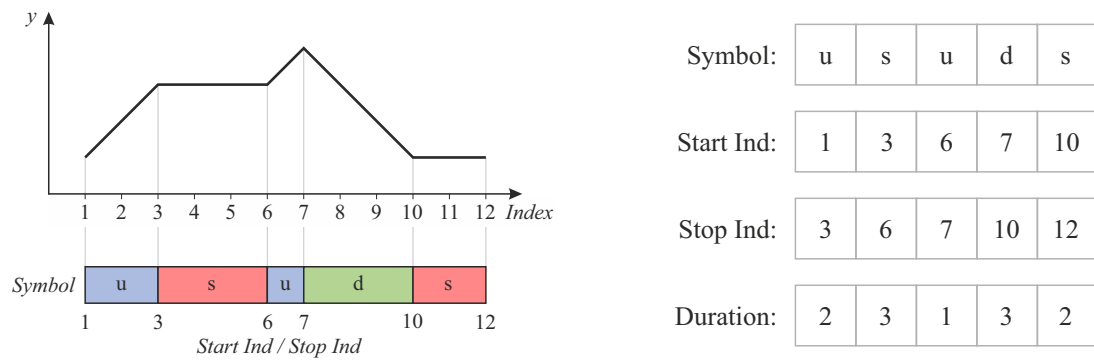
4.2.1 Symbolic Time Series Representation

Symbolic time series analysis is an acknowledged method in time series analysis (see Part III for details). In literature (e.g. [54]), the time series is divided into sub-segments of equal length. Subsequently, a symbol (from a finite set of symbols) is assigned to each sub-segment based on a derived value describing the sub-segment (e.g. mean, slope). Sub-sequences with the same behaviour/shape are assigned with the same symbol. As a result, the time series is transformed into a stream of symbols. Since each symbol spans a certain range, the dimensionality of the time series is reduced. This makes it suitable for tasks such as similarity search, finding discords, classification, e.g. [55–61].

Within this thesis, large physical systems are investigated: therefore, symbols/words are assigned based on the values obtained after processing the data with a linear differential operator (LDO). The combination of symbolic analysis with linear differential operators is a significant extension of the analysis paradigm since it now introduces symbolic analysis which includes system dynamics (see [P4, P5, P9]). This is tantamount to having the pseudo phase space available for symbolic analysis. Compared to the method described above, this results in sub-sequences of different length, see Fig. 4.3(a). The borders between the symbols and/or segments normally serve to locate change in the behaviour of the system.

In the shown example, the first derivative of the signal is used to symbolize the time series. The word **up** (abbreviated by the symbol **u**) is assigned to all values with a positive first derivative, the word **down** (abbreviated by the symbol **d**) for negative derivative values respectively. In the case the first derivative is zeros, the word **stationary** (with the symbol **s**) is assigned.

⁵Not to be confused with the idea of decorators in programming languages, such as Python.



(a) Symbolization of a time series based on its slope.

(b) Structure of the symbolic representation of the time series.

Fig. 4.3 Graphical visualization of the symbolic representation of a time series.

Instances of the class **SymbRepClass** are used to store and handle symbolic time series. In Fig. 4.3(b) the structure used for the implementation is shown. Additionally, this class implements various methods used for advanced symbolic time series, e.g.:

removeShortSyms(): This method is used to delete symbols (sequences) which are shorter than a given length. Outliers caused by perturbations of the signal can be removed using this method, yielding a “smoothed” version of the symbolic time series.

findSequence(): This method is used to find sequences of symbols within the stream of symbols. Using this method, hypotheses can be formulated to find certain patterns in the time series.

mergeSequence(): This method is used to merge sequences of symbols and assign a new symbol to the sequence. In this manner, a more complex behaviour can be described using a single new word/symbol. This builds the basis for hierarchical compounding.

hierarchicalCompounding(): Applying this method performs hierarchical compounding as proposed in [P4]. This can be used to automatically identify hierarchical patterns in the time series.

plotOnAxes(): To visualize symbolic time series this method is used. It uses semi-transparent patches as overlay to already existing time series plots. This is described in Section 4.4.1. Additionally, the symbols and their duration are annotated on the patches, see Fig. 4.7.

Each signal channel from a **MdtsClass** instance can hold a link to an instance of a **SymbRepClass**. This provides the basis to perform more advanced computations, e.g. multi-channel symbolic time series analysis.

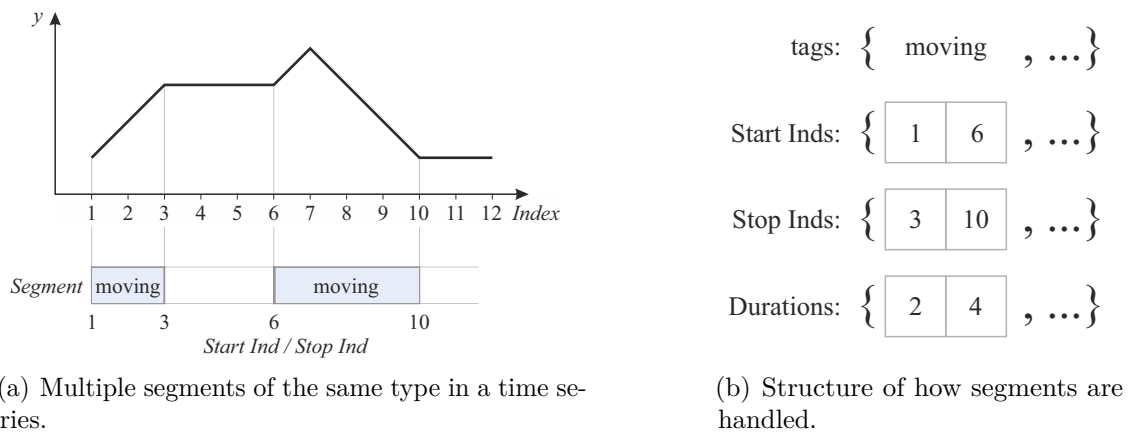


Fig. 4.4 Graphical visualization of segments.

4.2.2 Segments

Segments are similar to the symbolic representation as mentioned before, with the exception that not each data point must belong to a segment. They are used for marking ranges of data with similar behaviour, e.g. to mark where the machine is in active operation (see Fig. 4.4(a)) or to mark incidents (see [P7]). This is useful to extract data and visualize time intervals of interest.

The class **SegmentsClass** defines and implements the structure used to store segments. The structure of how segments are stored within instances of this class is shown in Fig. 4.4(b). Segments with the same properties are grouped together using a single tag. The start and stop indices of each single segment of the group are stored along with their duration. Multiple segments of a different type can be hold by a single instance of the **SegmentsClass**.

To visualize the segments on already existing time series plots, the method **plotOnAxes()** can be used. This is achieved using semi-transparent patches, similar to the symbolic time series representation, see Fig. 4.7.

Again, each signal channel of an **MdtsClass** instance can refer to one instance of a **SegmentsClass**.

4.2.3 Events

Events are used to mark something special on certain time-stamps (see Fig. 4.2), e.g. breakdown/replacement of a certain part. Each event consists of the tuple `[event-name, time-stamp(s)]`. If the same event occurs multiple times, there is a link to multiple time-stamps established. Different events are collected and grouped in a single instance of the **EventsClass**. The tuples are stored in the fields **names** and **times**.

An instance of the **MdtsClass** can hold exactly one **EventsClass** instance since events are treated to be valid for the entire time series. As it can be seen in Fig. 4.7, events are displayed as line-markers in each channel.

4.3 Additional Functions and Toolboxes

The herein presented data analytics framework for cyber physical systems additionally comprises newly developed algorithms and functions collected in toolboxes. Some of those are developed along with the papers implementing the presented algorithms⁶.

The functions and toolboxes most pertinent to this thesis are:

Rule Engine: The rule engine is used to formulate hypotheses which are to be found within the multi-dimensional time series. An example for a hypothesis would be: **channel1 > 100 & channel2 <= 0**. As a result, segments where this hypothesis is valid are returned. This can be seen as the first primitive of a domain specific language. The rule engine is heavily used for incident analysis, e.g. [P7].

applyLDO(): Linear differential operators (LDO) are used to include the physics of the system within calculations. This is used in the presented methods for symbolic time series analysis, see Part III. The function applies a given LDO matrix to channels of the **MdtsClass** instance. For example, this can be used to smooth a channel, calculate a regularized derivative or solve the associated inverse problem.

Hierarchical Decomposition: This toolbox implements the “hierarchical decomposition” algorithm as presented in [P3].

Hermite Approximation: This toolbox implements Hermite approximation as presented in [P1, P2].

Constraint Fitting Toolbox: The constraint fitting toolbox implements constrained polynomial approximation as presented in [P8].

computeCollocative(): This function is used to perform collocative computations for channels of a **MdtsClass** instance. In this case values are computed at collocated time-stamps, i.e. this results in a new channel with data at the same time-stamps as the channels in the **MdtsClass** instance. This can be seen as derived measurements. Different types of calculations can lead to a derived channel. The most common are:

1x1: In this case the derived channel is based on exactly one existing channel, e.g. from a hydraulic pressure signal the force signal is calculated (with

⁶Currently, not all the additional functions are yet available as source code in the public domain. They are being reviewed, and it is expected that they will be made available in the near future.

the use of the piston area of a hydraulic cylinder). Another example would be the calculation of a local derivative of one channel using a local linear differential operator (see Section 9.1).

nx1: Here a new channel is computed based on n existing channels, e.g. from the rod- and piston-side pressures of a cylinder the actual working force is calculated.

This pool of functions and toolboxes will increase in future and some are planned to be implemented as methods of the **MdtsClass** for the ease of use.

4.4 Data Visualization

A large part of this thesis deals with the development of new methods for analysing data emanating from large physical systems. To gain new insights, exploratory data analysis (EDA) [62] is an important tool. Different types of time series visualizations, e.g. [63, 64], reveal different insights leading to novel data analysis algorithms.

In the presented data analytics framework a data visualization supporting the above mentioned data handling framework is implemented. The goal is to support the formulation of hypotheses and the development of new algorithms.

4.4.1 Visualizing Multi-Dimensional Time Series Data

Since the multi-dimensional time series represents multiple channels sampled at the same time-stamps, the implemented visualization is based on stacked plots (axes) sharing a common abscissa (time axis), see Fig. 4.5. Each signal channel is visualized in one sub-plot. There are various representations available for displaying the data. The most common ones are scatter-plots and line-plots. The analysed time series are temporally ordered sets, therefore the default representation for displaying the data are line-plots. In this manner the order of the points gets visible.

4.4.2 Data Decimation

In exploratory data analytics a seamless and fast interaction with the data is important. The implemented data visualization uses the MATLAB[®] graphics-engine. If a figure is generated, the figure holds the entire data which can yield a poor interaction performance in the case of big data sets.

In the presented framework, data decimation is implemented to support plotting multi-dimensional time series. The focus is to minimize the amount of data to be handled by the graphics-engine and simultaneously preserving the visual perception

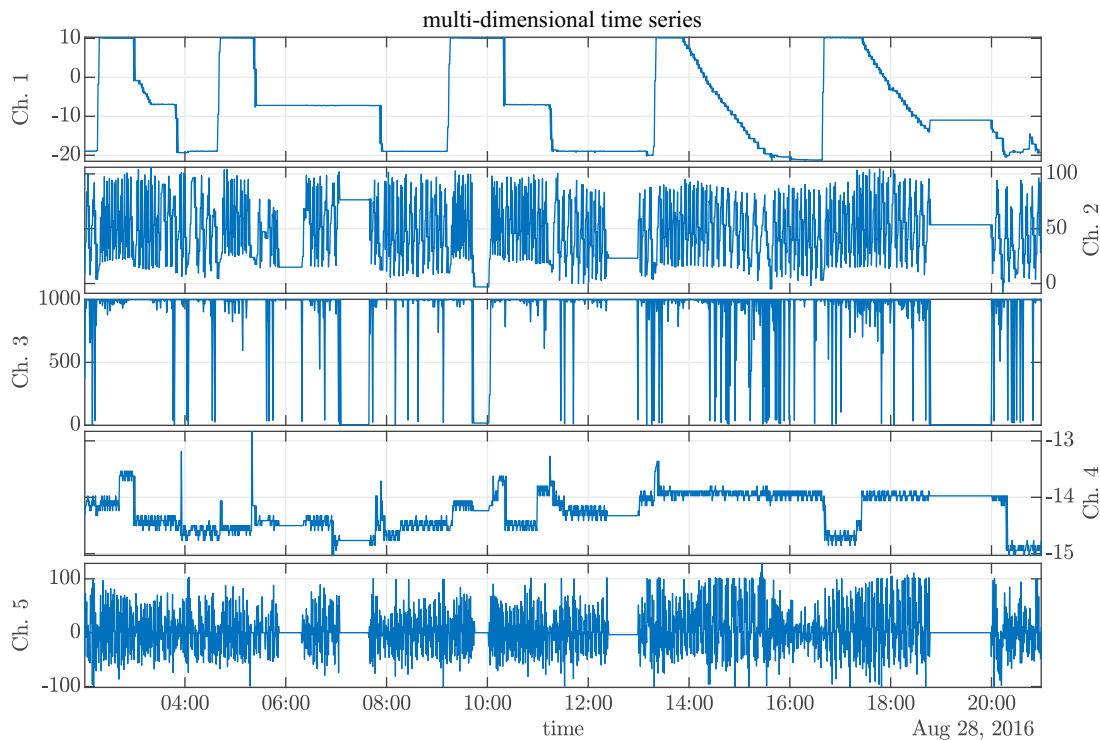


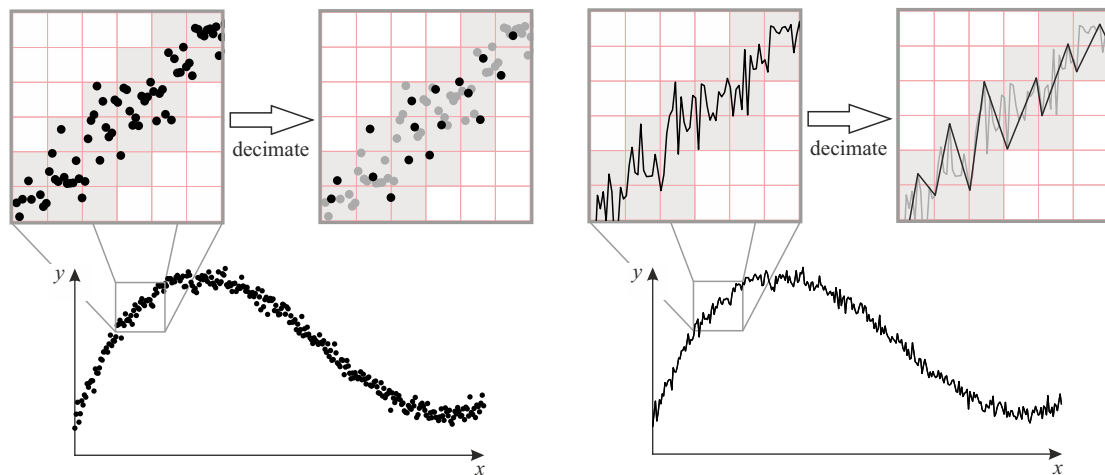
Fig. 4.5 A multi-dimensional time series visualized as a stacked plot with a common time axis. Each channel is plotted as one sub-plot.

(shape) of the time series. Various work is published in this area, e.g. [65–67]. The basic idea implemented herein is, to decimate the data based on a given resolution (e.g. screen resolution). On a screen it makes no sense to generate graphics with a resolution better than one data point per one pixel. If multiple data points are within one pixel of the screen, not all points need to be displayed [68–70].

This idea is shown in Fig. 4.6 for scatter-plots and line-plots. In Fig. 4.6(a) the data decimation for scatter plots is presented schematically. In the shown case, multiple data points activate the same pixel yielding an overdrawing. In other words, although there are multiple data points within one pixel, the information content cannot be represented by this pixel. Therefore, it is enough to keep only one of the points.

The same idea is valid for line-plots as presented in Fig. 4.6(b). Since the MATLAB[®] graphics-engine automatically connects two data points with a line, the points kept for plotting are the data points with the highest and lowest y -value within each column of pixels.

In general, the reduced data set is computed in a preprocessing step for a given monitor resolution and a given range of data to be plotted. Since those parameters change during interaction this preprocessing step is triggered on each interaction (e.g. zooming or resizing of figure). To improve speed, only the data within the needed range is decimated, stored temporarily and provided to the MATLAB[®] graphics-engine.



(a) Schematic of decimation used in scatter-plots. (b) Schematic of decimation used in line-plots.

Fig. 4.6 Schematic of decimation; Multiple data points within the same pixel are reduced to a single data point, since sub-pixel details cannot be displayed on a monitor. The grey squares indicate the activated pixels.

4.4.3 Decorative Overlays

To visualize the herein presented decorative objects, i.e. events, segments and symbolic time series, the multi-dimensional time series plots as presented in Section 4.4.1 are overlaid with the “decorative” information. This is shown in Fig. 4.7. Events (**EventsClass**) are visualized using a vertical marker (vertical line, spanning the full y -range). Different colours represent various types of events. Segments (**SegmentsClass**) and symbolic time series representations (**SymbRepClass**) are represented as semi-transparent patches spanning the given time-intervals and the full y -range. Again, different colours indicate different *symbols/words* in the case of symbolic time series or segments of various behaviour. By defining the degree of opacity, multiple overlays can overlap by simultaneously preserving the shown information. These overlays are heavily used in Part III.

As described above, to handle and visualize data from complex systems each signal channel can hold a link to his decorative objects. Therefore, each sub-plot (channel) can show different representations, which is necessary to support the identification of sub-sequences. A visualization of a time series with the mentioned overlays is shown in Fig. 4.7.

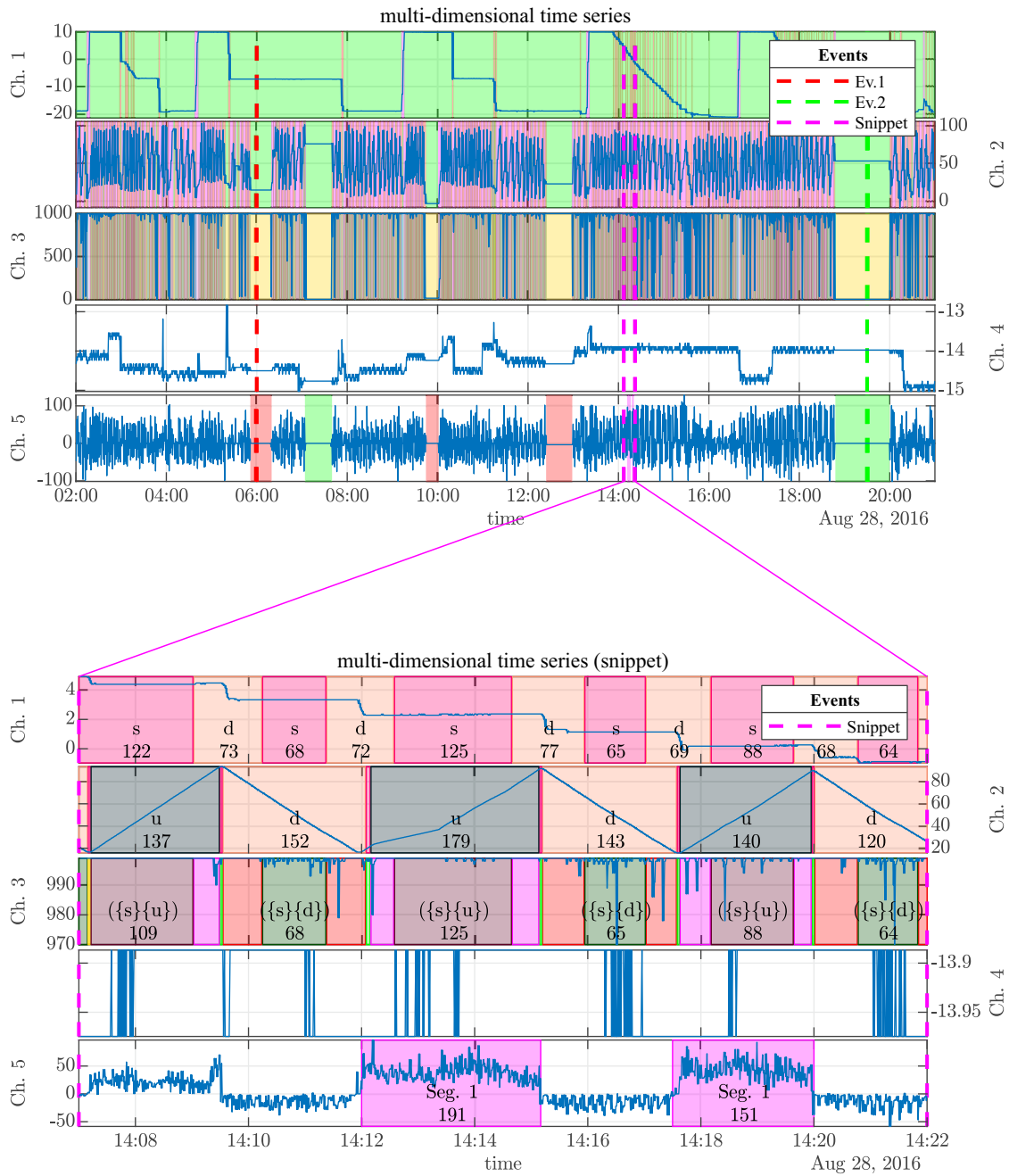


Fig. 4.7 Visualization of decorative objects using overlays. The three top channels are overlaid with *symbolic representations*. The bottom channel is overlaid with *segments*. Additionally, *events* are shown as dashed lines on each channel. The plot on the bottom shows a zoomed in snippet with additional annotations regarding the length and the name/symbol of the segments/symbolic time series.

Part II

Polynomial Methods

5 | Synopsis

Polynomials are famous and studied extensively in literature due to their nature and simplicity. This is, amongst other reasons, because every continuous function within a closed interval can be approximated with a polynomial so that it does not exceed a given approximation error (the Weierstrass approximation theorem[1]). Therefore, polynomials are the main topic of this part of the thesis, since they play an essential role in data analytics.

The first paper [P8] (see Chapter 6) presents a framework for approximating data with a polynomial which fulfils certain constraints. These constraints can either be *zero constraints*, *value constraints*, *general constraints* or *constraints on the coefficients* of the polynomial. In the first step, the solution for zero constraints is presented which yields the homogeneous solution. Based on that, Vandermonde vectors and their analytical derivatives are used to model the value/general constraints yielding the particular solution. This solution is subtracted from the noisy data and a polynomial is approximated in a homogeneous manner yielding the homogeneous portion of the solution. Throughout all types of constraints, the mathematical formulation, the numerical implementation and the covariance propagation are shown.

The ideas in paper [P3] (see Chapter 7) were developed based on the ability of using analytical derivatives of the Vandermonde vectors within polynomial approximation. In this paper, weighted polynomial approximation is used to derive the state vectors (value and derivatives) at a reduced set of abscissa values, since state space variables are known to provide a good description for the dynamics of a system. The next level of compression is achieved by polynomial approximation of the state vectors, i.e. Hermite approximation. To establish a metric for simultaneous least-squares approximation in both domains (value and derivative domain), covariance weighting together with concatenated Vandermonde matrices is used. Additionally, covariance propagation is performed throughout all compression levels. This idea can be used in cyber physical systems, since the weighted polynomial regression can be calculated on the edge-device. Subsequently, only the state vectors are transmitted to other devices which can be used to rebuild an approximation of the original signal. The reconstruction can be used to perform further evaluations.

As Weierstrass' theorem stated, every function within an interval can be approximated up to the needed precision using polynomials with a "high" enough degree. Although the Vandermonde basis is analytically linear independent, it becomes

degenerated in numerical computations at higher degrees. To perform high degree polynomial approximation for state vectors (i.e. given function values and derivatives) a method using discrete orthogonal polynomials was developed in paper [P1] (see Chapter 8). This method is numerically stable for high degree polynomials, since re-orthogonalization is used within the derivation of the method. In the case the abscissa-values do not change for multiple approximations, the basis can be calculated a-priori. Thus, the approximation problem reduces to a simple vector-matrix multiplication making it suitable to be processed on an edge-device.

6 | Constrained Polynomial Approximation for Inverse Problems in Engineering

Originally appeared as:

P. O’Leary, R. Ritt, and M. Harker, “Constrained Polynomial Approximation for Inverse Problems in Engineering,” in *Proceedings of the 1st International Conference on Numerical Modelling in Engineering*, Lecture Notes in Mechanical Engineering, M. A. Wahab, Ed., vol. NME2018, Springer Singapore, 2019, pp. 225–244. DOI: 10.1007/978-981-13-2273-0_19. [Online]. Available: http://link.springer.com/10.1007/978-981-13-2273-0%7B%5C_%7D19

BibT_EX:

```
@incollection{OLeary2019,
  author      = {O'Leary, Paul and Ritt, Roland and Harker,
                Matthew},
  booktitle   = {Proceedings of the 1st International Conference
                on Numerical Modelling in Engineering},
  doi         = {10.1007/978-981-13-2273-0_19},
  editor      = {Wahab, Magd Abdel},
  issn        = {21954364},
  number      = {Lecture Notes in Mechanical Engineering},
  pages       = {225--244},
  publisher   = {Springer Singapore},
  title       = {{Constrained Polynomial Approximation for Inverse
                Problems in Engineering}},
  url         = {http://link.springer.com/10.1007/978-981-13-2273-
                0{\_}19},
  volume      = {NME2018},
  year        = {2019}
}
```

Constrained Polynomial Approximation for Inverse Problems in Engineering

Paul O’Leary, Roland Ritt and Matthew Harker

Chair of Automation, University of Leoben, Austria,
paul.oleary@unileoben.ac.at,
WWW home page: automation.unileoben.ac.at

Abstract. This paper presents the derivation, implementation and testing of a series of algorithms for the least squares approximation of perturbed data by polynomials subject to arbitrary constraints. These approximations are applied to the solution of inverse problems in engineering applications. The generalized nature of the constraints considered enables the generation of vector basis sets which correspond to admissible functions for the solution of inverse initial-, internal- and boundary-value problems. The selection of the degree of the approximation polynomial corresponds to spectral regularization using incomplete sets of basis functions. When applied to the approximation of data, all algorithms yield the vector of polynomial coefficients α , together with the associated covariance matrix Λ_α . A matrix algebraic approach is taken to all the derivations. A numerical application example is presented for each of the constraint types presented. Furthermore, a new approach to performing constrained polynomial approximation with constraints on the coefficients is presented.

Keywords: constrained polynomial approximation, conditional least squares

1 Introduction

In 1964 Klopfenstein published a paper on *Conditional Least Squares Polynomial Approximation* [7]. In this paper he proposed a method which should implement a least squares approximation

$$y(x) = p(x, \alpha) = \sum_{i=0}^d \alpha_i x^i \quad (1)$$

with a set of m generalized constraints of the form,

$$y^{(r_i)}(c_i) = b_i, \quad (2)$$

i.e., a set of generalized derivative constraints. This is a very important class of equations since it corresponds to the problems which need to be solved when addressing inverse initial-, inner-, and boundary value problems, where we may have Neumann or Dirichlet boundary conditions. His work has been cited in numerous papers as a means of implementing such constraints [5, 8]. However, as we shall see with a closer examination this is not exactly the problem he has solved.

Now using Klopfenstein's notation, he proposes minimizing the cost function

$$S_m = \sum_{k=1}^m w_k \{y_k - Q_n(x_k)\}^2 \quad (3)$$

where $Q_n(x)$ is an n^{th} degree polynomial, there are l prescribed constraints on $Q_n(x)$ at one or more points x_j of the form,

$$Q_n^{(r)}(x_j) = b_{rj}. \quad (4)$$

This is exactly the problem we wish to solve. The proposed model polynomial is of the form,

$$Q_n(x) = P_{l-1}(x) + \Pi_l(x) Q_{n-l}(x) \quad (5)$$

where,

$$\Pi_l(x) = \prod_{j=1}^l (x - x_j)^{1+R_j}. \quad (6)$$

This definition for $\Pi_l(x)$ is where the error lies: defining the derivative constraint as multiple zeros at the location x_j not only places a constraint on the derivative but also on the discriminant [3]. For example the single constraint

$$Q_n^{(1)}(x_j) = b_{rj} \quad (7)$$

results in,

$$\Pi(x) = (x - x_j)^2. \quad (8)$$

This constraint implies that both the value of the function $\Pi_l(x) = 0$ and its first derivative $\dot{\Pi}_l(x) = 0$ are both zero, i.e., its discriminant is zero at these locations. This is not the same as only requiring $Q_n^{(1)}(x) = 0$. Consequently, the algorithm is not actually solving the case of a generalized derivative constraint. A further consequence is the product,

$$\Pi_l(x) Q_{n-l}(x) \quad (9)$$

is not of degree n as required but of degree $n + 1$. Indeed, the method is perfectly correct for value constraints, but not for generalized derivative constraints. Consequently, there is still the need for a generalized solution.

A further issue with the proposed calculation method is associated with the transformation

$$y' = \frac{y_k - P_{l-1}}{\Pi_l(x_k)}. \quad (10)$$

This transformation requires different solutions for collocated and interstitial constraints. Collocated constraints will have the value $\Pi_l(x_k) = 0$, which requires the elimination of the point y_k from the data set; whereas this is not required for interstitial constraints. Klopfenstein proposes eliminating the points which are collocated; however, this will not solve the issue with an interstitial constraint which is very close to a location of x_i . In this case the computation becomes numerically unstable. Also, the computation of the covariance propagation is by no means trivial.

The authors have in the past worked on covariance weighted implementation of constraints [10]. This method does not permit the determination of admissible functions in such a general manner and the work was restricted to first order derivatives. The work presented in [9, 11] solve a similar set of problems using discrete orthogonal polynomials however the constraints were restricted to be collocated. The interstitial case was not dealt with.

There are a number of other papers relating to what is called *constrained* polynomial approximation [1,4,6]; however, these papers refer to constraints on the residual and not on the approximating function. As a consequence they are not relevant to the solution of this problem.

In this paper we present a series of methods for constrained polynomial approximation which permit the solution of the above problem among others. A consistent algebraic formulation is derived which enables the computation of the coefficients α of the minimizing polynomial $p(x, \alpha)$, together with the associated covariances Λ_α . The paper is structured so that each algorithm is presented separately with a motivation why and for which problems it is required; an algebraic derivation of the method; an algorithmic implementation and a demonstration on a numerical example¹.

2 Notation and definition of constraint types

In this paper we shall use a matrix-vector formulation of the solutions. The following notation is used:

1. A polynomial of degree d is denoted as $p_d(x, \alpha)$,

$$p_d(x, \alpha) = \sum_{i=0}^d \alpha_i x^i. \quad (11)$$

2. The coefficient vectors, e.g., $\alpha = [\alpha_d, \dots, \alpha_0]^T$ are denoted using Greek letters. Their corresponding covariance matrix is denoted as Λ_α .
3. The sets of observations x_i, \hat{y}_i are denoted by vectors \mathbf{x} and $\hat{\mathbf{y}}$, the *hat* denotes that these values are perturbed.
4. The approximating values of $\hat{\mathbf{y}}$ are denoted as \mathbf{y} .
5. The vector \mathbf{v} , is an exception as it is a row vector, it corresponds to a row in the Vandermonde matrix \mathbf{V} , e.g., $\mathbf{v} = [x_i^d, x_i^{d-1}, \dots, x_i, 1]$

The types of constraints considered are shown in Figure 1, they are, polynomials:

1. with constraining zeros.
2. with constraining values, and
3. with generalized constraints of the form,

$$p^{(k)}(c, \alpha) = h, \quad (12)$$

i.e., the k^{th} derivative of the polynomial at the location c is h .

¹A full implementation of each method is available at <https://www.mathworks.com/matlabcentral/profile/authors/3977359-matthew-harker-paul-o-leary>

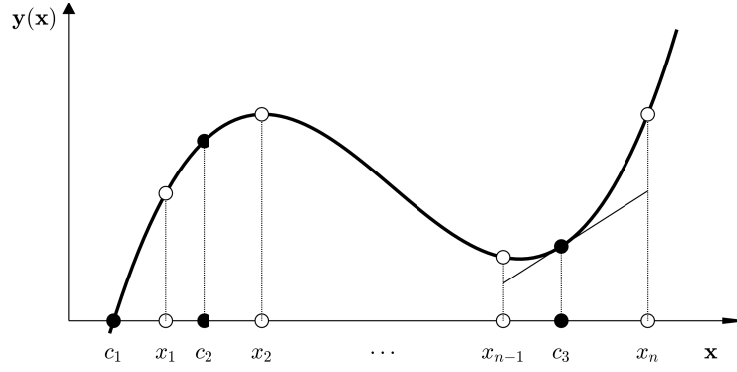


Fig. 1. The values of the polynomial are generated at the locations x_i . The three types of constraints considered are: a known zero located at c_1 ; a value constraint located at c_2 and a differential constraint located at c_3 . The location of the constraints may be either collocated with the data points x_i or interstitial.

3 Polynomial approximation with constraining roots

The task here is to approximate a set of n points x_i, \hat{y}_i by a polynomial, $p(x, \alpha)$ of degree d , while fulfilling the constraints $p(c_i, \alpha) = 0 \forall c_i \in \mathbf{c}$. The vector $\mathbf{c} \triangleq [c_1, \dots, c_m]^T$ contains the locations of the m constraints. These types of constraints are motivated from the fact that many physical and engineering systems exhibit eigenfunctions which have known zeros. In particular systems which are described by Sturm-Liouville equations commonly have such constraints. Consider, for example the simply supported beam shown in Figure 2: its bending modes have zeros at both ends. If we were using sensors to measure the deformation of the beam, it would be desirable to have a model which takes advantage of the a-priori knowledge about the solution.

The example in Figure 2 is a boundary value problem, i.e., the constraints are at the boundaries of the system. However, in this paper, we consider the more general case where zeros can be at any location, either within or outside the support; in this manner initial-, interior- and boundary-value problems can be solved. The model we are defining is a polynomial $p_d(x, \alpha)$ of degree d with at least m roots at the locations $\mathbf{c} = [c_1, \dots, c_m]^T$. The proposed polynomial $p_d(x, \alpha)$ model is the product of two polynomials,

$$p_d(x, \alpha) = p_m(x, \gamma) p_{d-m}(x, \beta), \quad (13)$$

whereby, $p_m(x, \gamma)$ is the polynomial with exactly m roots at the locations $\mathbf{c} = [c_1, \dots, c_m]^T$. This polynomial is known prior to the computation of the approximation, since it is fully defined by the locations of the roots,

$$p_m(x, \gamma) = \prod_{k=1}^m (x - c_k). \quad (14)$$

Given γ , the task of the least squares approximation is now to determine the values for β which minimize the residual in a least squares sense. Then to map $\beta \mapsto \alpha$ to obtain the required coefficients and also compute the covariance Λ_α .

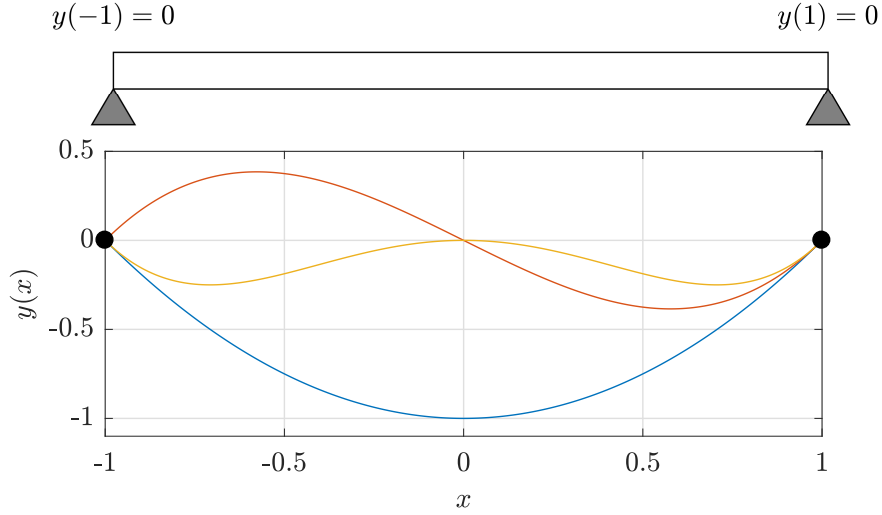


Fig. 2. Example of a simply supported beam (top) and the first three constrained polynomials which fulfil the constraints. These polynomials can be used to approximate the bending modes (bottom) of the beam.

3.1 Algebraic formulation

The values of the model are computed as

$$\mathbf{y} = \mathbf{V} \boldsymbol{\alpha}, \quad (15)$$

The coefficient vector $\boldsymbol{\alpha}$ results from the product of two polynomials, $p_m(x, \boldsymbol{\gamma})$ and $p_{d-m}(x, \boldsymbol{\beta})$. Given the coefficients $\boldsymbol{\gamma}$ we can compute the corresponding convolution matrix $\boldsymbol{\Gamma}$ such that,

$$\boldsymbol{\alpha} = \boldsymbol{\Gamma} \boldsymbol{\beta}. \quad (16)$$

Substituting into Equation (15) we obtain,

$$\mathbf{y} = \mathbf{V} \boldsymbol{\Gamma} \boldsymbol{\beta}. \quad (17)$$

The residual vector \mathbf{r} , is computed as,

$$\mathbf{r} = \hat{\mathbf{y}} - \mathbf{y} \quad (18)$$

$$= \hat{\mathbf{y}} - \mathbf{V} \boldsymbol{\alpha} \quad (19)$$

$$= \hat{\mathbf{y}} - \mathbf{V} \boldsymbol{\Gamma} \boldsymbol{\beta} \quad (20)$$

The weighted cost function $\varepsilon_w(\boldsymbol{\beta})$ is defined as

$$\varepsilon_w(\boldsymbol{\beta}) = \{\hat{\mathbf{y}} - \mathbf{V} \boldsymbol{\Gamma} \boldsymbol{\beta}\}^T \mathbf{W} \{\hat{\mathbf{y}} - \mathbf{V} \boldsymbol{\Gamma} \boldsymbol{\beta}\} \quad (21)$$

where $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ is a diagonal matrix, with $\mathbf{w} = w(x)$ containing the values of the weighting function $w(x)$. The weighting function $w(x)$ must be positive definite. The normal equations are obtained by solving,

$$\frac{d\varepsilon_w(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = 0. \quad (22)$$

This is a well known problem, the solution is,

$$\beta = (\Gamma^T V^T W V \Gamma)^{-1} \Gamma^T V^T W \hat{y}. \quad (23)$$

For clarity² we shall consider the case $W = I$,

$$\beta = (\Gamma^T V^T V \Gamma)^{-1} \Gamma^T V^T \hat{y} \quad (24)$$

$$= (V \Gamma)^+ \hat{y} \quad (25)$$

where given the matrix A , then A^+ denotes the Moore and Penrose pseudo inverse of this matrix. Then substituting into Equation (16) yields,

$$\alpha = \Gamma (V \Gamma)^+ \hat{y}. \quad (26)$$

Defining,

$$K \triangleq \Gamma (V \Gamma)^+, \quad (27)$$

yields,

$$\alpha = K \hat{y}, \quad (28)$$

and from the statistics of linear operators we obtain the covariance of α as,

$$\Lambda_\alpha = K \Lambda_{\hat{y}} K^T. \quad (29)$$

This further simplifies under the assumption that the model is free from bias and the input is i.i.d. perturbed with a standard deviation σ . In this case,

$$\Lambda_{\hat{y}} = \sigma^2 I, \quad (30)$$

and consequently,

$$\Lambda_\alpha = \sigma^2 K K^T. \quad (31)$$

and

$$\Lambda_y = V \Lambda_\alpha V^T \quad (32)$$

$$= V K \Lambda_{\hat{y}} K^T V^T. \quad (33)$$

3.2 Algorithmic implementation

Equation (13) can be written as,

$$p_d(x, \alpha) = \prod_{k=1}^m (x - c_k) \sum_{j=0}^{d-m} \beta_j x^j. \quad (34)$$

Since c is known we can compute the vector,

$$y_c = \prod_{k=1}^m (x - c_k) \quad (35)$$

²This simplifies the equations and makes the structure more visible. There is no principle change in the methods for generalized weighting.

this corresponds to $\Pi_l(x_k)$ in Klopfenstein's paper. He proposes dividing the left-hand-side of his equation [7, Equation 4] by these values. However, since \mathbf{y}_c can contain values of zero or very close to zero this computation becomes numerically unstable. Alternatively, we can compute the Hadamard product of \mathbf{y}_c with the monomials of the free polynomial $p(\mathbf{x}, \boldsymbol{\beta})$ to obtain,

$$p_d(\mathbf{x}, \boldsymbol{\alpha}) = \sum_{j=0}^{d-m} \beta_j \mathbf{y}_c \circ \mathbf{x}^j. \quad (36)$$

Now defining the basis function $\mathbf{b}_j \triangleq \mathbf{y}_c \circ \mathbf{x}^j$ and $\mathbf{B} \triangleq [\mathbf{b}_{d-m}, \dots, \mathbf{b}_1, \mathbf{b}_0]$ we obtain

$$\mathbf{B} \equiv \mathbf{V}\boldsymbol{\Gamma}, \quad (37)$$

leading to

$$\mathbf{y} = \mathbf{B}\boldsymbol{\beta}. \quad (38)$$

There are a number major advantages associated with this approach:

1. Most importantly, the numerical instability of the Klopfenstein's methods is eliminated.
2. There is no further need to distinguish between collocated and interstitial constraints.
3. The matrix \mathbf{B} can be computed in a numerically efficient manner using a Horner form starting from the vector \mathbf{y}_c , see m-code Listing 1.1.

Listing 1.1. Code snippet to compute \mathbf{B} .

```
B = zeros( length(x), nrCfsBeta );
% perform the synthesis
B(:, nrCfsBeta) = yz;
for k=(nrCfsBeta - 1):-1:1
    B(:, k) = x.*B(:, k+1);
end
```

The pseudo-inverse of $\mathbf{B} = \mathbf{Q}\mathbf{R}$ is computed using QR decomposition, so that $\boldsymbol{\beta}$ is obtained as,

$$\boldsymbol{\beta} = \mathbf{R}^+ \mathbf{Q}^T \hat{\mathbf{y}}, \quad (39)$$

and $\boldsymbol{\alpha}$ is obtained by computing $\boldsymbol{\alpha} = \boldsymbol{\Gamma}\boldsymbol{\beta}$.

3.3 Example and interpretation

Polynomials form a vector space \mathcal{P} . The columns \mathbf{b}_i of the matrix \mathbf{B} form a vector basis set, which spans the sub-space of \mathcal{P} that is the polynomials of degree d which have the prescribed roots. Consequently, any linear combination of the columns of \mathbf{B} also fulfills the constraints. The matrix \mathbf{B} has the same polynomial ordering as \mathbf{V} , since $\boldsymbol{\Gamma}$ is a convolution matrix and by definition it has a band diagonal structure. An example of such vector bases is shown in Figure 3.

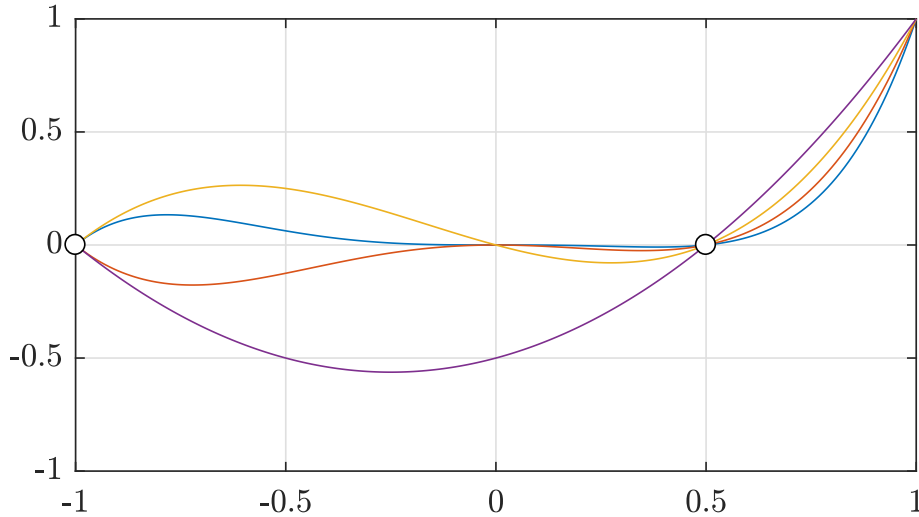


Fig. 3. Example of a set of polynomials, of degrees $\mathbf{d} = [2, 3, 4, 5]$, with prescribed roots $\mathbf{c} = [-1, 0.5]^T$. These polynomials form a vector basis set for the space of all polynomial up to degree $d = 5$ which fulfil the constraints.

To test the algorithm a synthetic set of data points were generated: the constraints are defined by $\mathbf{p}(\mathbf{c}, \boldsymbol{\alpha}) = \mathbf{0}$ with $\mathbf{c} = [-0.1, 1]^T$; an arbitrary *free* polynomial $y_f(x) = (-x^3 + 0.6x^2 + x)$ was assumed. The resulting polynomial used for the synthesis of the data is $y_g(x) = (x + 0.1)(x - 1)(-x^3 + 0.6x^2 + x)$ and iid. noise with $\sigma = 0.03$ was added. The corresponding coefficient vector $\boldsymbol{\alpha}_g$ and the results of the approximation $\boldsymbol{\alpha}_f$ are shown in Table 1. The data, the results of the approximation and the 1σ error bound for the data prediction are shown graphically in Figure 4. The variance of the solution resulting from the uncertainty in the coefficients is shown in Figure 5. The approximating polynomial is $\mathbf{p}_5(x, \boldsymbol{\alpha})$.

Table 1. Example of a polynomial with predefined zero locations, $\boldsymbol{\alpha}_g$ are the coefficients of the polynomial used to generate the data; whereby iid. noise with $\sigma = 0.03$ was used during the synthesis; the constraints are defined by $\mathbf{p}(\mathbf{c}, \boldsymbol{\alpha}) = \mathbf{0}$ and $\mathbf{c} = [-0.1, 1]^T$; an arbitrary *free* polynomial $y_f(x) = (-x^3 + 0.6x^2 + x)$ was assumed. The resulting polynomial used for the synthesis of the data is $y_g(x) = (x + 0.1)(x - 1)(-x^3 + 0.6x^2 + x)$. The coefficients $\boldsymbol{\alpha}_f$ were obtained from the fitting algorithm.

	$\boldsymbol{\alpha}_g$	$\boldsymbol{\alpha}_f$
α_5	-1.00	-0.9909
α_4	1.50	1.4956
α_3	0.56	0.5291
α_2	-0.96	-0.9511
α_1	-0.10	-0.0841
α_0	0.00	0.0015

Table 2. Covariance matrix Λ_{α} for the coefficients α resulting from the approximation: all values have been scaled by 10^3 .

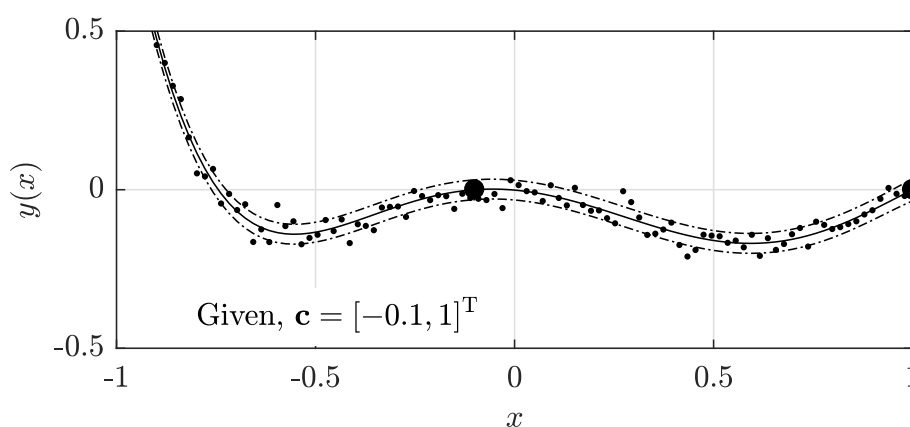
	α_5	α_4	α_3	α_2	α_1	α_0
α_5	3.7504	-0.2748	-4.5183	-0.0682	1.0134	0.0976
α_4	-0.2748	0.7678	0.0643	-0.6316	0.0618	0.0125
α_3	-4.5183	0.0643	5.7272	0.2785	-1.4133	-0.1384
α_2	-0.0682	-0.6316	0.2785	0.5709	-0.1312	-0.0185
α_1	1.0134	0.0618	-1.4133	-0.1312	0.4267	0.0426
α_0	0.0976	0.0125	-0.1384	-0.0185	0.0426	0.0043

This polynomial is equivalent to,

$$y(x) = (\gamma_2 x^2 + \gamma_1 x + \gamma_0) (\beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta_0), \quad (40)$$

$$y(x) = (x + 0.1) (x - 1) (\beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta_0), \quad (41)$$

$$y(x) = (x^2 - 0.9x - 0.1) (\beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta_0). \quad (42)$$

**Fig. 4.** Example data set and results as summarized in Table 1. The upper and lower curves shown correspond to a 1σ error bound for the prediction of a measurement value. Note the approximating polynomial fulfils the constraints exactly.

4 Polynomial approximation with constraining values

The task here is to approximate a set of n points x_i, \hat{y}_i by a polynomial, $p(x, \alpha)$ of degree d , while fulfilling the constraints $p(c_i, \alpha) = a_i \forall c_i, a_i$. This problem is the natural extension of approximation with polynomials having known roots. This type of problem

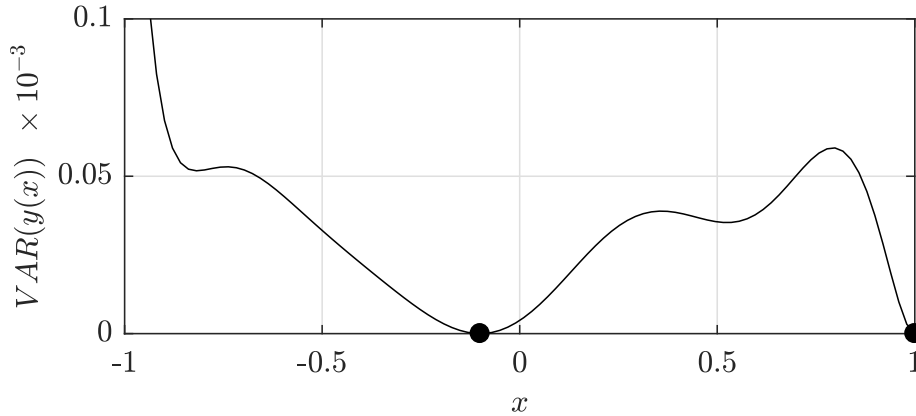


Fig. 5. Model prediction variance due to covariance of the polynomial coefficients, see Table 1. Note the approximating polynomial fulfils the constraints exactly, i.e., there is zero variance at the locations of the zeros.

is encountered in many inverse problems in engineering and science. The vectors \mathbf{c} and \mathbf{a} contain the locations and the corresponding values of the constraints. The task of approximating can be split into separate tasks:

1. Use polynomial interpolation to determine the coefficients $\boldsymbol{\delta}$ of $y_p(x) = \mathbf{p}(x, \boldsymbol{\delta})$ which fulfil the m constraints, i.e., $\mathbf{p}_{m-1}(\mathbf{c}, \boldsymbol{\delta}) = \mathbf{a}$. We call this portion the *particular solution* it is denoted by \mathbf{y}_p , since it changes with the particular values of the constraints a_i . Given m constraints the polynomial is of degree $m - 1$.
2. Residualize the observations $\hat{\mathbf{y}}$ wrt. the particular solution, i.e. $\hat{\mathbf{y}}_h = \hat{\mathbf{y}} - \mathbf{p}(x, \boldsymbol{\delta})$.
3. Approximate $\hat{\mathbf{y}}_h$ with a polynomial fulfilling the constraints in a homogeneous manner; this problem was solved in the Section 3. This portion, denoted by \mathbf{y}_h , is called the homogeneous portion of the solution, it is independent of the particular values of a_i and has the structure

$$y_h(x) = \mathbf{p}_m(x, \boldsymbol{\gamma}) \mathbf{p}_{d-m}(x, \boldsymbol{\beta}). \quad (43)$$

The resulting polynomial model being considered is:

$$\mathbf{p}_d(x, \boldsymbol{\alpha}) = \underbrace{\mathbf{p}_m(x, \boldsymbol{\gamma}) \mathbf{p}_{d-m}(x, \boldsymbol{\beta})}_{y_h(x)} + \underbrace{\mathbf{p}_{m-1}(x, \boldsymbol{\delta})}_{y_p(x)}. \quad (44)$$

4.1 Algebraic formulation

The least squares approximation of data by a polynomial with a-priori value constraints decomposes into a number of simple algebraic steps. The solutions for each of these steps are already available:

1. Determine the coefficients $\boldsymbol{\delta}$ of the interpolating polynomial $\mathbf{p}(x, \boldsymbol{\delta})$: we may use any method suitable to compute the interpolating polynomials, as far it delivers the coefficients with respect to a sum of monomials.

2. Then we proceed to compute, $\hat{y}_h(x_i) = \hat{y}_i - p(x_i, \delta)$, the approximation to this data should fulfil the constraints in a homogeneous manner.
3. Compute, γ the coefficients of the polynomial which has roots located at the values of c , i.e. $p(c, \gamma) = \mathbf{0}$
4. Determine the values for β using the methods previously presented in Section 3.
5. Compute α as follows:

$$\alpha = \gamma * \beta + \delta, \quad (45)$$

Note δ must be padded with leading zeros to obtain the same dimension as $\gamma * \beta$.

With respect to covariance propagation: the vector δ is error free since it represents the polynomial which interpolates the constraining values exactly. Additionally the constraining values are considered to be free from error. The covariance propagation comes from computing β from \hat{y} , this is exactly the same mechanism as in Section 3; consequently, the task of computing the covariance matrix Λ_α is also solved.

The numerical implementation consists primarily in the concatenation of a series of already available functions.

4.2 Numerical example

A synthetic data set was generated to test the algorithm: the constraints were defined as: $p(c, \alpha) = \mathbf{a}$ with $c = [1, 0, 0.9]^T$ and $\mathbf{a} = [-0.5, 0.3, -0.5]^T$. A free polynomial $y_f(x) = -x^3 + x^2 - 1$ was assumed and iid. noise with $\sigma = 0.03$ was added. The various coefficients involved are presented in Table 3.

Table 3. The values of the coefficients γ , β and δ used to generate the synthetic data. Additionally an iid Gaussian noise component with $\sigma = 0.03$ was added.

	β		γ		δ	
β_3	-1	γ_3	1.0			
β_2	1	γ_2	0.1	δ_2	-0.8889	
β_1	0	γ_1	-0.9	δ_1	-0.0889	
β_0	-1	γ_0	0	δ_0	0.3000	

The data set, results of the approximation, a 1σ error bound for the measurement prediction and the interpolation $p(x, \delta)$ are shown in Figure 6. For completeness the residualized data \hat{y} and the homogeneous approximation are shown in Figure 7

5 Polynomial approximation with generalized constraints

The task now is to perform polynomial approximation given a more general form of constraint,

$$\left. \frac{d^k y(x)}{dx^k} \right|_c = y^{(k)}(c) = a. \quad (46)$$

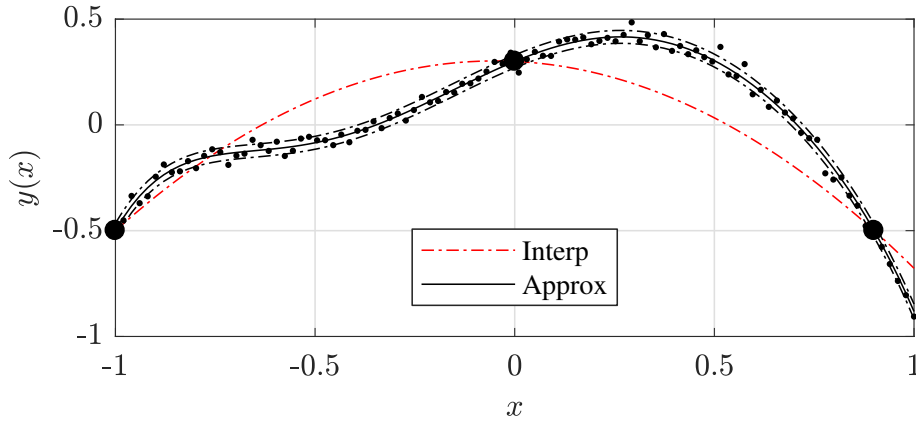


Fig. 6. Result of the approximation for the test data set defined by the values given in Table 3. The interpolating polynomial and the approximation are shown. The upper and lower curves shown correspond to a 1σ error error bound for the prediction of a measurement point.

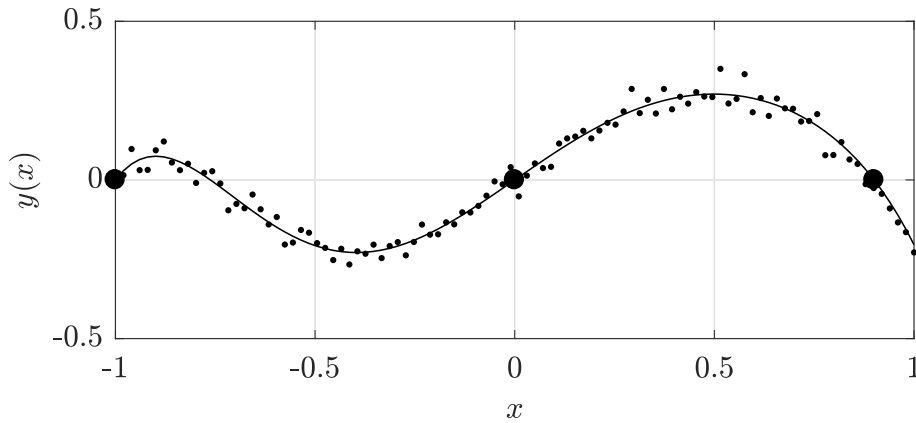


Fig. 7. The homogeneous approximation after $\hat{y}(x)$ has been residualized on the interpolation polynomial $p(x, \delta)$.

This is the case Klopfenstein claimed to solve; however, his method makes no provision for the case $y^{(k)}(c_i) = a_i$ but there being no constraints on $y^{(j < k)}(c_i) = a_i$. His method placed multiple roots to implement the constraints, this implies that the discriminant will also be zero.

Each constraint of this form is fully defined by the triplet of values: c the location of the constraints; k the order of the derivative associated with the constraint and a the value of the constraint. We define the triplet $t \triangleq [c, k, a]^T$ as the representation for the constraint. Some examples of constraints and their defining triplets are given in Table 4

Furthermore, given m constraints the i^{th} is defined as a triplet of values,

$$t_i = [c_i, k_i, a_i]^T \quad (47)$$

and the m definitions are concatenated to form the matrix $T \triangleq [t_1, \dots, t_m]$. This is a very general definition for constraints. It permits the implementation of initial, interior and

Table 4. Some examples of generalized constraints and their representation as triplets.

Constraint	Triplet
$y(0) = 0$	$\mathbf{t} = [0, 0, 0]^T$
$y^{(1)}(0.9) = 1.3$	$\mathbf{t} = [0.9, 1, 1.3]^T$
$y^{(2)}(-1) = -0.1$	$\mathbf{t} = [-1, 2, -0.1]^T$

boundary value conditions: Dirichlet, $y(c) = a$ and Neumann, $y^{(1)}(c) = a$, are simply special cases.

This more general definition of a constraint enables the application of the polynomials to a large class of inverse problems related to the observation of systems governed by ordinary differential equations. Additionally the solution algorithm will enable the generation of vector basis sets which span the sub-space of all polynomials which fulfill the constraints in a homogeneous manner. These correspond to admissible functions when solving differential equations using the Rayleigh-Ritz method [2, 12, 13]. Virtually all engineering systems have some known solution conditions. Using these constraints the a-priori knowledge can be embedded in the solution; this provides for regularization which is consistent with the known behaviour of the system.

5.1 Algebraic formulation

Let us consider the polynomial $p_d(x, \alpha)$: defining the *Vandermonde vector* \mathbf{v}_d of degree d as³,

$$\mathbf{v}_d(x) \triangleq [x^d, x^{d-1}, \dots, x, 1], \quad (48)$$

and the coefficient vector α , then we can define the value of $y(x)$ as,

$$y(x) = \mathbf{v}_d(x) \alpha. \quad (49)$$

Now taking the first derivative of this polynomial with respect to x ,

$$\dot{y}(x) = d \alpha_d x^{d-1} \dots + 2 \alpha_2 x^1 + \alpha_1 x^0. \quad (50)$$

One possibility is to define a vector $\dot{\alpha} \triangleq [0, d \alpha_d, \dots, 2 \alpha_2, \alpha_1]^T$ such that,

$$\dot{y}(x) = \mathbf{v}_d(x) \dot{\alpha}. \quad (51)$$

The relationship between the entries of $\dot{\alpha}$ and α can be written as,

$$\dot{\alpha}(k-1) = k \alpha(k). \quad (52)$$

³More stringently we should define all vectors as column vectors and use the transpose operation to obtain the corresponding row vector. However, to maintain consistency here we define the Vandermonde vector as one row of the Vandermonde matrix,

That is, each coefficient is scaled by k and shifted by one location in the coefficient vector. We now need to formulate the derivative $\dot{y}(x)$ in terms of $\dot{\mathbf{v}}_d(x)$ and $\boldsymbol{\alpha}$, i.e.,

$$\dot{y}(x) = \dot{\mathbf{v}}_d(x) \boldsymbol{\alpha}. \quad (53)$$

Let us start by differentiating the Vandermonde vector, to obtain,

$$\dot{\mathbf{v}}_d(x) \triangleq [dx^{d-1}, (d-1)x^{d-2}, \dots, 1, 0]. \quad (54)$$

We now wish to determine a matrix M such that,

$$\dot{\mathbf{v}}_d(x) = \mathbf{v}_d(x) M. \quad (55)$$

Observing the consequences of performing the derivative: each power of x is scaled by the degree of the monomial being differentiated. Additionally the monomials are shifted up one location in the vector. We can define a scaling vector \mathbf{s} associated with the derivative of each entry,

$$\mathbf{s} = [d, (d-1), \dots, 2, 1]. \quad (56)$$

Then from \mathbf{s} we generate the matrix S required to scale the entries of \mathbf{v}_d , i.e.,

$$S = \text{diag}\{\mathbf{s}\}. \quad (57)$$

Secondly we observe that the powers of x are all shifted one place to the left and a zero is inserted where the constant term (1) was present. The scaling and shifting are both achieved by applying the matrix M ,

$$M = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ S & \mathbf{0} \end{bmatrix} \quad (58)$$

With this we obtain,

$$\dot{y}(x) = \mathbf{v}_d(x) M \boldsymbol{\alpha}. \quad (59)$$

From this equation we can determine the two relationships,

$$\dot{\mathbf{v}}_d(x) = \mathbf{v}_d(x) M \quad (60)$$

$$\dot{\boldsymbol{\alpha}} \triangleq M \boldsymbol{\alpha} \quad (61)$$

This approach generalizes to derivatives of i^{th} order as follows,

$$\mathbf{v}_d^{(i)}(x) = \mathbf{v}_d(x) M^i \quad (62)$$

$$\boldsymbol{\alpha}^{(i)} = M^i \boldsymbol{\alpha} \quad (63)$$

With this we have all the equations required so that we can implement the triplets. Now returning to the m constraints defined as triplets.

The aim now is to define and compute a matrix C such that

$$C \boldsymbol{\alpha} = \mathbf{a}. \quad (64)$$

This is achieved using Equation (62) to generate C with one row per constraint, i.e.,

$$C \triangleq \begin{bmatrix} \mathbf{v}_d^{(k_1)}(c_1) \\ \vdots \\ \mathbf{v}_d^{(k_m)}(c_m) \end{bmatrix} \quad (65)$$

The $r_C = \text{rank}\{C\}$ gives the number of independent constraints. If $r_C < m$, then there are redundant constraints. Secondly, the values of the constraints \mathbf{a} must lie in the range of C , i.e., $\mathbf{a} \in \text{range}\{C\}$. If this is not the case then the constraints are inconsistent.

Solving Equation 64 for α yields,

$$\alpha = C^+ \mathbf{a} + N_C \gamma. \quad (66)$$

Whereby, N_C is an ortho-normal basis function set which spans the null space of C , i.e., the subspace of polynomials which fulfil the constraints. Consequently,

$$\mathbf{y} = V \{C^+ \mathbf{a} + N_C \gamma\}. \quad (67)$$

Expanding shows that the equation splits into the particular and the homogeneous portion which is used to perform the least squares approximation.

$$\mathbf{y} = \underbrace{V C^+ \mathbf{a}}_{\mathbf{y}_p} + \underbrace{V N_C \gamma}_{\mathbf{y}_h} \quad (68)$$

5.2 Algorithmic implementation

With Equation (68) we now have all the prerequisites to perform the least squares approximation. The algorithm can be summarized as follows:

1. Compute the constraining matrix C by concatenating the Vandermonde vectors corresponding to the definitions of the constraints represented by their triplets.
2. Determine the rank $r_C = \text{rank}\{C\}$, pseudo-inverse C^+ and a vector basis set for the null space of C , i.e., N_C .
3. Test for uniqueness and consistency of the defined constraints.
4. Compute the particular solution $\alpha_p = C^+ \mathbf{a}$ and $\mathbf{y}_p = V C^+ \mathbf{a}$.
5. Residualize the observations $\hat{\mathbf{y}}$ on the particular solution,

$$\hat{\mathbf{y}}_h = \hat{\mathbf{y}} - \mathbf{y}_p. \quad (69)$$

6. Compute γ ,

$$\gamma = \{V N_C\}^+ \hat{\mathbf{y}}_h. \quad (70)$$

7. Compute α ,

$$\alpha = \alpha_p + N_C \gamma. \quad (71)$$

8. Compute Λ_α by defining,

$$K \triangleq N_C \{V N_C\}^+. \quad (72)$$

then

$$\Lambda_\alpha = K \Lambda_{\hat{\mathbf{y}}} K^T. \quad (73)$$

5.3 Numerical example

A synthetic data set was generated to test the algorithm, the constraints used are defined in Table 5. In this example the coefficients of the *free* polynomial are chosen to have the coefficients $\beta = [1, 1, 0.8, 0.9]^T$ and iid. Gaussian noise with $\sigma = 0.03$ was added.

Table 5. Definition of the constraints used to synthesize the test data and their representations as triplets.

Constraint	Triplet
$y^{(0)}(-0.8) = 1$	$\mathbf{t}_1 = [-0.8, 0, 1]^T$
$y^{(0)}(0.1) = 0,$	$\mathbf{t}_2 = [0.1, 0, 0]^T$
$y^{(1)}(1) = 1$	$\mathbf{t}_3 = [1, 1, 1]^T$

The null space of C , i.e., N_C , forms a vector basis set for the space of the coefficients of all polynomials which fulfil the constraints in a homogeneous manner. Consequently, $B = VN_C$ is a vector basis set of polynomials for the space itself. The columns of B correspond to admissible functions for this portion of the problem. The basis functions and their first derivatives are shown in Figure 8; as can be seen, the basis functions fulfil all the constraints in a homogeneous manner as expected. Furthermore, the derivative constraint located at $x = 1$, places no restrictions on the values of the basis functions at this location.

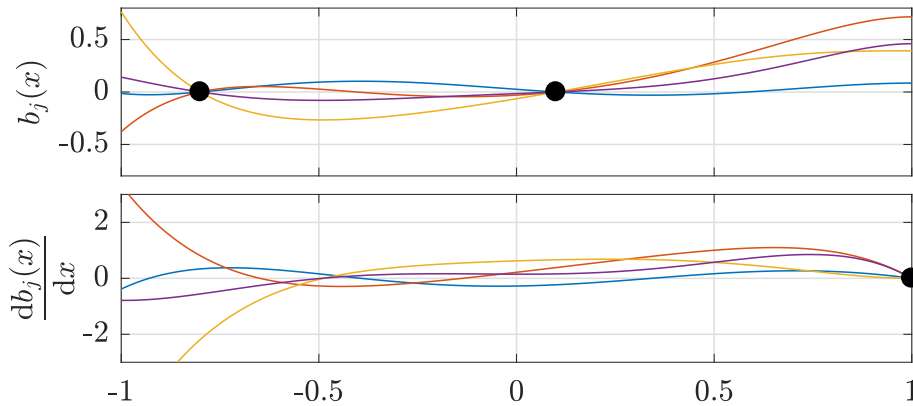


Fig. 8. Top: $B = VN_C$ and bottom $\dot{B} = VMN_C$ show the basis functions and their first derivatives respectively. It can be seen that all basis functions fulfil the constraints in a homogeneous manner as required. Furthermore, the derivative constraint places no restrictions on the values of the polynomials at this location.

The result of the least squares approximation are shown in Figure 9. The raw data together with the constrained polynomial approximation. The upper and lower curves

shown correspond to a 1σ error bound for the prediction of a measurement point. Additionally, the first derivative of the approximation is presented, this is to show that the derivative constraint has been fulfilled.

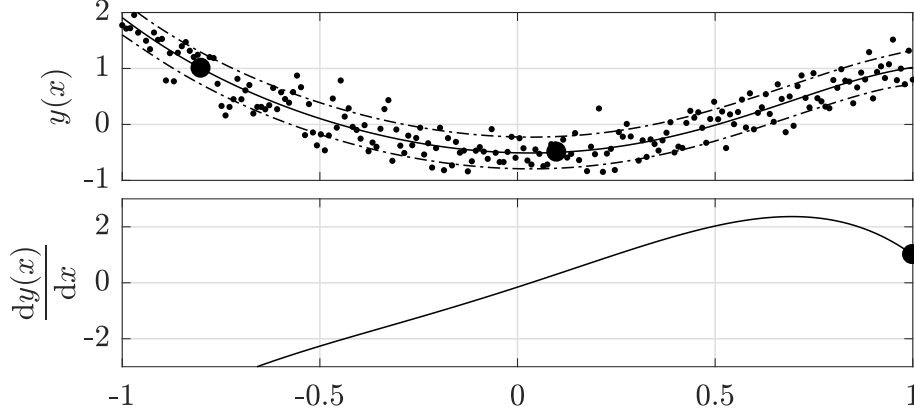


Fig. 9. Top: The raw data together with the constrained polynomial approximation. The upper and lower curves shown correspond to a 1σ error bound for the prediction of a measurement point. Bottom: The first derivative of the approximation; this is to show that the derivative constraint has been fulfilled.

5.4 Extensions of this work

In this paper we have demonstrated it with the geometric polynomials. However, it could just as well be performed using the Bernstein polynomials. In this case we would define C_B to be C in terms of the Bernstein polynomials

$$C_B(u)\beta = \mathbf{a}. \quad (74)$$

and perform a coordinate transformation,

$$u = \frac{x - x_{min}}{x_{max} - x_{min}}. \quad (75)$$

The Bernstein polynomials can be computed generically, as can their derivatives,

$$\mathbf{b}_{nk} = \binom{k}{n} u^k (1-u)^{n-k}, \quad k = 0, \dots, n. \quad (76)$$

Consequently, we can form $C_B(u)$ so that Equation (74) is fulfilled. The rest of the algorithm remains unchanged. This method works with any basis function set for which we can compute their values and derivatives as required.

6 Coefficient constrained polynomial approximation

The above constraints have all involved placing constraints on the function. However, we may also wish to put constraints on the values or relationships between coefficients.

To achieve this we define a matrix G which maps between an auxiliary vector of coefficients γ and the primary set of coefficients α . Starting from,

$$\mathbf{y} = V\alpha \quad (77)$$

Now defining the mapping between γ and α ,

$$\alpha = G\gamma, \quad (78)$$

we obtain,

$$\mathbf{y} = VG\gamma. \quad (79)$$

From the algebra in this paper we know the approximating solution given this model is,

$$\gamma = (VG)^+ \hat{\mathbf{y}} \quad (80)$$

$$K \triangleq G(VG)^+ \quad (81)$$

$$\alpha = K\hat{\mathbf{y}} \quad (82)$$

$$\Lambda_\alpha = K\Lambda_{\hat{\mathbf{y}}}K^T. \quad (83)$$

As we can see the solution has exactly the same structure as the solution in Section 3.

6.1 Example coefficient constraints

Consider the curve $y(u)$ defined by the coefficients of the Bernstein polynomials,

$$y(u) = (1-u)^3 \beta_0 + 3(1-u)^2 u \beta_1 + 3(1-u)u^2 \beta_2 + u^3 \beta_3. \quad (84)$$

Defining the individual basis functions as,

$$\mathbf{b}_{30} = (1-u)^3 \quad \mathbf{b}_{31} = 3(1-u)^2 u \quad \mathbf{b}_{32} = 3(1-u)u^2 \quad \mathbf{b}_{33} = u^3 \quad (85)$$

and $B \triangleq [\mathbf{b}_{30}, \mathbf{b}_{31}, \mathbf{b}_{32}, \mathbf{b}_{33}]$ then we can write,

$$\mathbf{y} = B\beta \quad (86)$$

Now let us consider some possible constraints, to perform symmetric approximation we may wish to define $\beta_0 = \beta_3 = \gamma_0$ and $\beta_1 = \beta_2 = \gamma_1$, this would require defining,

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix} \quad (87)$$

Then the formulation,

$$\mathbf{y} = BG\gamma, \quad (88)$$

yields the necessary equations to solve the approximation. Formulated, in this manner any linear relationship between the coefficients can be used as a constraining condition. The matrix G may also be rank deficient, as in the case of eliminating matrices.

7 Conclusions

This paper has presented a series of algorithms for constrained polynomial approximation. The methods enable constraints as roots of a polynomial, as value constraints and as generalized differential constraints. The generic matrix algebraic formulation permits the computation of the optimizing coefficient values as well as their covariance, in a direct and simple manner. Additionally, a new approach to applying generic linear constraints to the coefficients of polynomials is presented.

Acknowledgments

Partial funding for this work was provided by:

1. The Austrian research funding association (FFG) under the scope of the COMET program within the K2 center “IC-MPP” (contract number 859480). This programme is promoted by BMVIT, BMDW and the federal states of Styria, Upper Austria and Tyrol.
2. The Center of Competence for Recycling and Recovery of Waste 4.0 (acronym ReWaste4.0) (contract number 860884) under the scope of the COMET — Competence Centers for Excellent Technologies — is financially supported by BMVIT, BMWFW, and the federal state of Styria, managed by the FFG.

References

1. Akima, H.: A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM* **17**(4), 589–602 (1970). DOI 10.1145/321607.321609. URL <http://doi.acm.org/10.1145/321607.321609>
2. Gander, M.J., Wanner, G.: From Euler, Ritz, and Galerkin to Modern Computing. *SIAM Review* **54**(4), 627–666 (2012). DOI 10.1137/100804036. URL <http://epubs.siam.org/doi/10.1137/100804036>
3. Gelfand, I., Kapranov, M., Zelevinsky, A.: Discriminants, Resultants, and Multidimensional Determinants. *Modern Birkhäuser Classics*. Birkhäuser Boston (2008). URL <https://books.google.at/books?id=2zgxQVU1hFAC>
4. Giorgio, C., Michel, B.: *Constrained Polynomial Approximation*. Wiley-Blackwell (2016)
5. Handscomb, D.: *Methods of numerical approximation: Lectures delivered at a Summer School held at Oxford University, September, 1965*. Pergamon Press (2014)
6. J.M., M.: *Polynomial Approximations with Special Constraints*. In: *Elementary Functions*. Birkhäuser, Boston, MA (2016)
7. Klopfenstein, R.W.: Conditional least squares polynomial approximation. *Mathematics of Computation* **18**(88), 659–662 (1964). URL <http://www.jstor.org/stable/2002954>
8. Newbery, A.C.R.: Trigonometric interpolation and curve-fitting. *Mathematics of Computation* pp. 869–876 (1970)
9. O’Leary, P., Harker, M.: A framework for the evaluation of. In: *Inclinometer Data in the Measurement of Structures* URL. doi:10.1109/TIM.2011.2180969 <http://ieeexplore.ieee.org/document/6162983/> IEEE, pp. 61–1237 (2012)

10. O’Leary, P., Harker, M.: Inverse boundary value problems with uncertain boundary values and their relevance to inclinometer measurements. In: 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, pp. 165–169. IEEE (2014). DOI 10.1109/I2MTC.2014.6860725. URL <http://ieeexplore.ieee.org/document/6860725/>
11. O’Leary, P., Harker, M., Gugg, C.: Ieee. : An inverse problem approach to approximating sensor data in cyber physical systems. In: In: International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings,, pp. 1717–1722 (2015)
12. Pierce, J.G., Varga, R.S.: Higher Order Convergence Results for the Rayleigh Ritz Method Aplied to Eigenvalue Problems. I Estimates Relating Rayleigh-Ritz and Galerkin Approximations to Eigenfunctions.pdf. SIAM Journal on Numerical Analysis **9**(1), 137–151 (1972)
13. Ritz, W.: Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. Journal fur die Reine und Angewandte Mathematik **1909**(135), 1–61 (1909). DOI 10.1515/crll.1909.135.1. URL <http://eudml.org/doc/149295><https://www.degruyter.com/view/j/crll.1909.issue-135/crll.1909.135.1/crll.1909.135.1.xml>

7 | Hierarchical Decomposition and Approximation of Sensor Data

Originally appeared as:

R. Ritt, P. O’Leary, C. J. Rothschedl, A. Almasri, and M. Harker, “Hierarchical Decomposition and Approximation of Sensor Data,” in *Proceedings of the 1st International Conference on Numerical Modelling in Engineering*, Lecture Notes in Mechanical Engineering, M. A. Wahab, Ed., vol. NME2018, Springer Singapore, 2019, pp. 351–370. DOI: 10.1007/978-981-13-2273-0_27. [Online]. Available: http://link.springer.com/10.1007/978-981-13-2273-0%7B%5C_%7D27

BibT_EX:

```
@incollection{Ritt2019,
  author      = {Ritt, Roland and O'Leary, Paul and Rothschedl,
                Christopher Josef and Almasri, Ahmad and Harker,
                Matthew},
  booktitle   = {Proceedings of the 1st International Conference
                on Numerical Modelling in Engineering},
  doi         = {10.1007/978-981-13-2273-0_27},
  editor      = {Wahab, Magd Abdel},
  number      = {Lecture Notes in Mechanical Engineering},
  pages       = {351--370},
  publisher   = {Springer Singapore},
  title       = {{Hierarchical Decomposition and Approximation of
                Sensor Data}},
  url         = {http://link.springer.com/10.1007/978-981-13-2273-
                0{\_}27},
  volume      = {NME2018},
  year        = {2019}
}
```

Hierarchical Decomposition and Approximation of Sensor Data

Roland Ritt , Paul O’Leary, Christopher Josef Rothschedl, Ahmad Almasri, and Matthew Harker

Chair of Automation, University of Leoben, Austria,
roland.ritt@unileoben.ac.at,
WWW home page: automation.unileoben.ac.at

Abstract. This paper addresses the issue of hierarchical approximation and decomposition of long time series emerging from the observation of physical systems. The first level of the decomposition uses spatial weighted polynomial approximation to obtain local estimates for the state vectors of a system, i.e., values and derivatives. Covariance weighted Hermite approximation is used to approximate the next hierarchy of state vectors by using value and derivative information from the previous hierarchy to improve the approximation. This is repeated until a certain rate of compression and/or smoothing is reached. For further usage, methods for interpolation between the state vectors are presented to reconstruct the signal at arbitrary points. All derivations needed for the presented approach are provided in this paper along with derivations needed for covariance propagation. Additionally, numerical tests reveal the benefits of the single steps. The proposed hierarchical method is successfully tested on synthetic data, proving the validity of the concept.

Keywords: signal decimation, hermite approximation, hermite interpolation, covariance propagation, signal reconstruction, weighted regression

1 Introduction

Currently, much effort is being put into the collection of data, in particular in conjunction with IoT and smart sensors. With the rise of cyber physical systems (CPS) many data is collected from machines which, by nature, must abide by the laws of physics (e.g. dynamic systems). To analyse the behaviour of the system monitored, techniques for the approximation of the observed signal in presence of noise are necessary. A lot of research is done in the area of streaming algorithms, i.e., local regression problems [3, 7, 11, 13, 15, 16, 22, 23], for smoothing and approximation of data. In [24], a method for identifying patterns in dynamical systems using phase space was introduced. The analysis of the phase space is also used in [8] to detect fatigue damage based on ultrasonic data. To transform signals into the phase space, it is necessary to approximate data and derivatives, that is, to compute time series estimates for the state vector. In the past, the authors published work dealing with reconstruction of curves given its derivatives with the requirement to fulfil additional constraints [18, 19].

In this paper a new framework for the approximation of large time series data emanating from physical systems is presented. The main contributions of the paper are:

1. The proposal of a hierarchical approach for approximating large time series data which can be used in signal decimation;
2. The derivations for generating state vectors using spatial weighted polynomial approximation. This improves the quality of approximation by diminishing Runge's phenomenon;
3. A formulation for the approximation of data given value and derivative information (i.e. Hermite approximation). Covariance weighting is used to achieve a consistent metric used in the least squares approximation. This improves the quality of subsequent approximations;
4. The proposal to use a two-point expansion for the reconstruction of the signal and its derivatives based on the state vectors;
5. A consistent formulation of covariance propagation for the proposed derivations.

This paper is structured as follows: In Sect. 2 the framework for the hierarchical approximation is presented. For the calculation of state vectors, spatial weighted polynomial approximation is presented in Sect. 2.1. Different weighting functions are investigated and a matrix approach for the calculation of the state vectors and their according covariance matrix is presented. Section 2.2 presents a novel method for approximating data given collocated value and derivative information (state vectors). Covariance weighting and optional spatial weighting is used to achieve a consistent metric for least squares Hermite approximation. To reconstruct the signal from its decimated version (given its state vectors), a two-point expansion is proposed in Sect. 2.3 which performs better than a single point Taylor expansion. Finally, in Sect. 3 the performance of the proposed framework is tested on synthetic data. Different stages of the hierarchical approximation are presented and discussed.

2 Methodology and Algebraic Framework

A physical process $y(x)$ is observed at discrete points x_i (e.g. time, location). The observations made at these points are denoted as \hat{y}_i . Note: The *hat* indicates that the observation is perturbed by noise. This is said to be level 0 of the hierarchy (L_0). The n observations (signal) and locations are collected in the vector $\hat{\mathbf{y}} = [\hat{y}_1, \dots, \hat{y}_n]^T$ and $\mathbf{x} = [x_1, \dots, x_n]^T = [x_{(0),1}, \dots, x_{(0),n}]^T$, whereby $x_i = x_{(0),i}$ denotes the i -th location (i -th point) in level 0.

In the first level (L_1) of the proposed hierarchical approach the signal and its derivatives up to a certain order d , i.e., the state vectors \mathbf{s}_j , are approximated at collocated points x_j . These local estimates for the state vectors of the signal to be monitored are defined as

$$\mathbf{s}_j = \mathbf{s}_{(1),j} \triangleq \left[y_{(1),j}, \dot{y}_{(1),j}, \ddot{y}_{(1),j}, \dots, y_{(1),j}^{(d)} \right]^T = \left[y(x_j), \dot{y}(x_j), \ddot{y}(x_j), \dots, y^{(d)}(x_j) \right]^T,$$

where $x_j = x_{(1),j} = x_{(0),i=j \times l_1}$ denotes the j -th point¹ in level 1. In this level the first decimation takes place, since only points with a spacing l_1 are approximated. In ot-

¹To simplify readability, x_i is used for points in L_0 , x_j for points in L_1 and x_k for points in L_2 and above if not defined in another way. The subscripts (0), (1) and (2) denote the different levels.

her words, only the points $x_{(0),i=j \times l_1}$ from L_0 are approximated in L_1 . Additionally, the matrix Λ_j corresponding to the covariance of s_j can be computed along with the approximation. To improve the quality of the fit, a local weighting function $w_j(x)$ is used to perform weighted regression. Local weighting is used to obtain behaviour similar to splines; that is, input data only influence the result of the approximation in a finite region. The use of weighting functions which limit both, the values and derivatives at the end of the interval also reduce the Runge phenomenon. This is very closely related to Gibbs error and windowing in Fourier analysis [10].

The levels 0 and 1 (L_0 and L_1) of the hierarchical process is shown in Fig. 1.

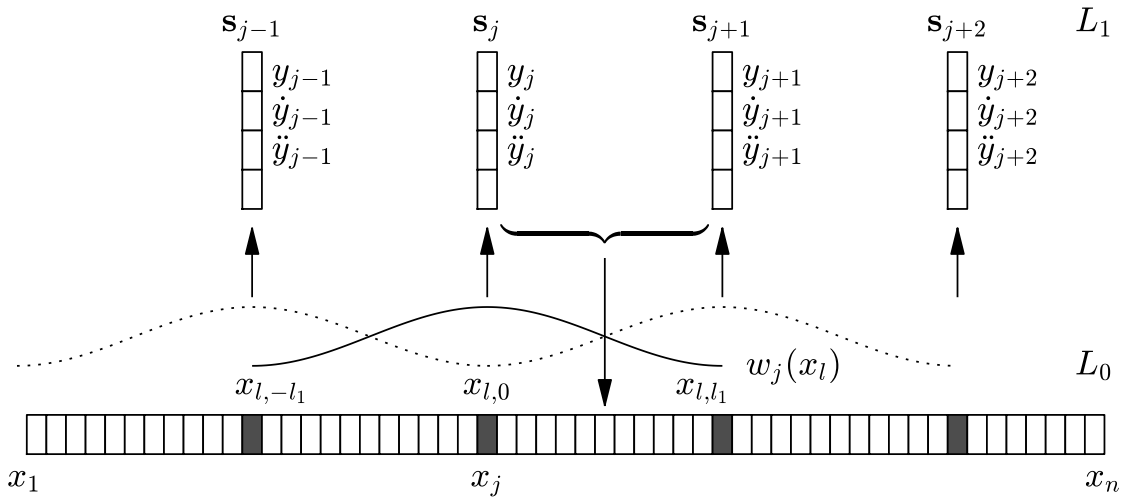


Fig. 1: Level 0 and 1 of the hierarchical approach: decimation is performed using weighted local polynomial approximation for obtaining local estimates for the state vectors s_j (and their covariances Λ_j).

Based on the state vectors s_j and their covariances Λ_j , a new method of covariance weighted Hermite approximation is used to perform the next level of hierarchical approximation. In this step, decimation can be implemented as well, if the approximated function is only evaluated at certain points x_k . As a result, you get the collocated state vectors $s_k = s_{(2),k}$ of level 2 (L_2). Since the derivatives are included in the approximation, a better confidence in the approximation is reached. Additional, spatial weighting can be implemented as well. This hierarchical process (Hermite approximation) can be repeated until the needed level of abstraction and smoothing (and/or decimation) of the signal is reached. This repetitive process is visualized in Fig. 2.

After this, the state vectors (i.e. the decimated signal) can be used to reconstruct (interpolate) the signal at the original positions (or somewhere in-between) using some form of expansion. or interpolation. Since there is derivative information available, an interpolation of higher degree is possible, resulting in better reconstruction. The necessary algebraic formulations are collected in the next sections.

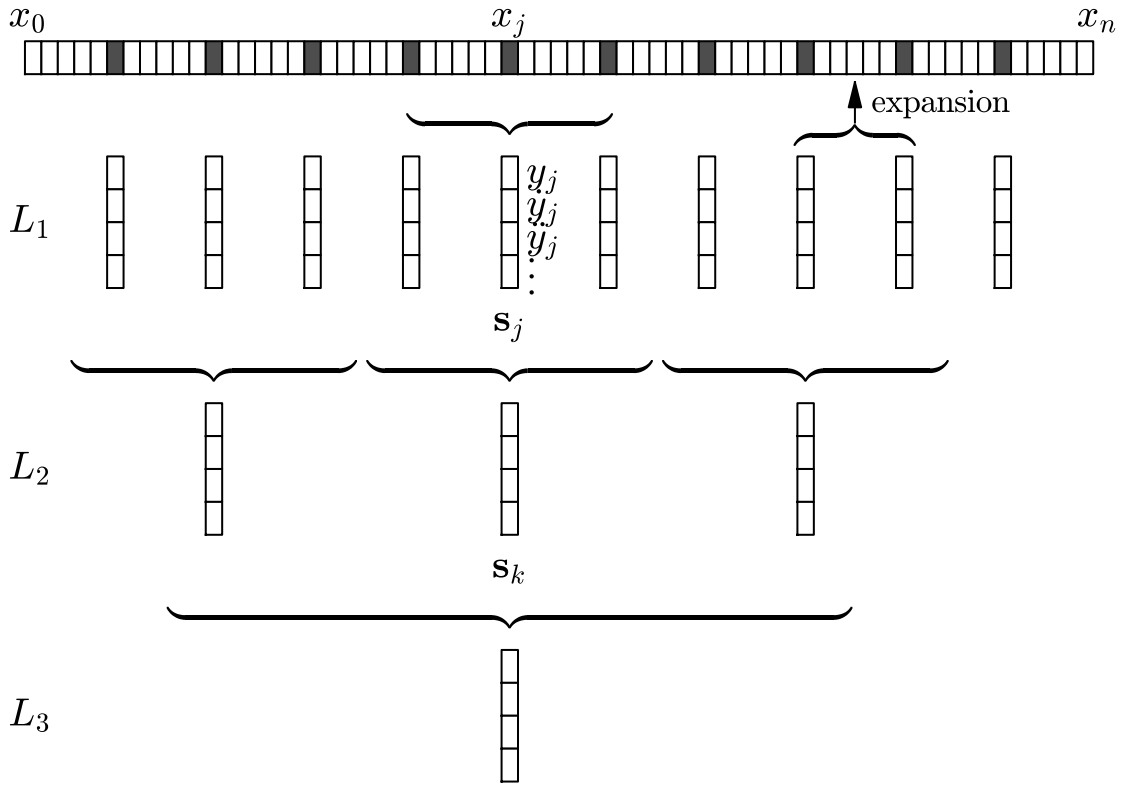


Fig. 2: Schematic of the higher levels of hierarchical approximation by using repetitive covariance weighted Hermite approximation.

2.1 Weighted Local Polynomial Approximation - Hierarchy Level 1

Local weighted regression is well-known in literature [3,5,13,15,17,20,22] with different studies on the weighting function to be used. In this work we investigate different weighting functions suitable for the herein presented hierarchical approach.

Weighting Functions: As [3] proposed, a local weighting function $w_j(u)$ should fulfil the following properties:

1. $w_j(u) > 0$ for $|u| < 1$: the weighting function influences only points in a certain range $u \in]-1, 1[$;
2. $w_j(-u) = w_j(u)$: it is symmetric around $u = 0$;
3. $w_j(u)$ is a non-increasing function for $|u| \geq 0$: the weighting function decreases with increasing distance to the point of interest;
4. $w_j(u) = 0$ for $|u| \geq 1$: everything outside the local window does not influence the approximation.

In our proposed framework, the weighting function w_j is shifted to the point of interest x_j and scaled, so that only a certain number of points $n_{w,1} = 2l_1 - 1$ is within the local window,

$$u = 2 \frac{x - x_j}{x_{j+1} - x_{j-1}}. \quad (1)$$

The number of points in the local window used to generate the state vectors in level 1 are denoted by $n_{w,1}$. Note: In the case of evenly spaced points, the weighting function is already 0 at the points x_{j-1} and x_{j+1} , i.e., $w_j(x_{j-1}) = w_j(x_{j+1}) = 0$. In this work two considerations are made with respect to the weighting functions $w_j(u)$:

1. Overlapping weighting functions (from the neighbouring local approximations) should form a partition of unity ($\sum_{j=1}^{n_j} w_j(x) = 1$). This ensures that all points in the input stream contribute with the same total weighting to the result. Some possible weighting functions are shown in Fig. 3. Of special interest is the *raised cosine*

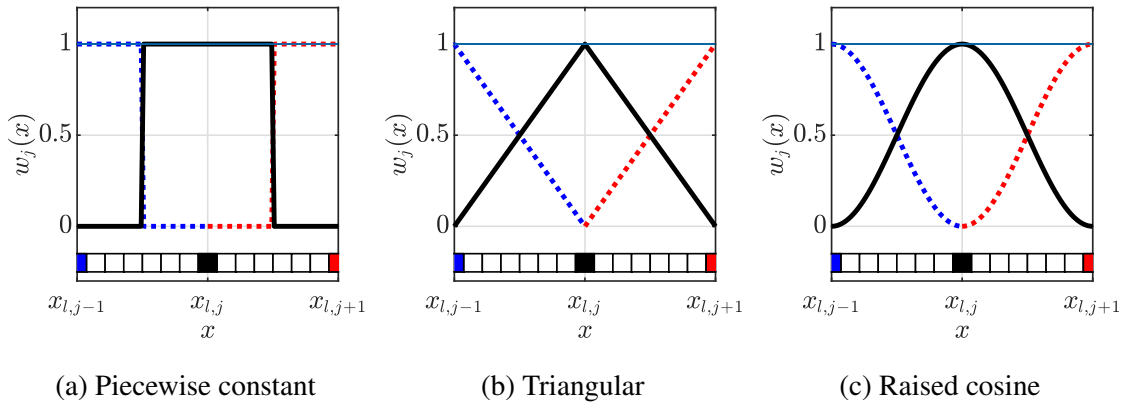


Fig. 3: Weighting functions forming a partition of unity; *black*: local weighting function $w_j(x)$ for approximating s_j ; *blue*: sum of the local weighting functions; *blue, red dotted*: parts of the weighting functions $w_{j-1}(x)$ and $w_{j+1}(x)$ for generating the neighbouring state vectors s_{j-1} and s_{j+1} ; *bottom*: schematic visualization of the position of original data points to be weighted.

function

$$w_j(u) = \frac{1 + \cos(\pi u)}{2}. \quad (2)$$

This function is known as the *Hanning* window in Fourier analysis which is used to diminish the Gibbs error [10]. This function also provides a first derivative of 0 at the end of the interval and at the centre point, i.e., $\frac{dw_j}{du} \Big|_{u=-1} = \frac{dw_j}{du} \Big|_{u=1} = \frac{dw_j}{du} \Big|_{u=0} = 0$; this is advantageous with respect to the Runge phenomenon.

2. Alternatively, we may wish to define a weighting function $w_j(u)$ such that its values and derivatives up to the k^{th} order tend to zero at the ends of the support; that is,

$$\lim_{|u| \rightarrow 1} w^{(0)}(u) \rightarrow 0, \quad \dots, \quad \lim_{|u| \rightarrow 1} w^{(k)}(u) \rightarrow 0. \quad (3)$$

This can be achieved using the polynomials,

$$w(u) = (u-1)^{(k+1)} (u+1)^{(k+1)}. \quad (4)$$

The polynomials computed using this weighting function are special cases of the Jacobi polynomials. These functions do not directly form a partition of unity but approximate it. The weighting function for $k = 1$ is shown in Fig. 4.

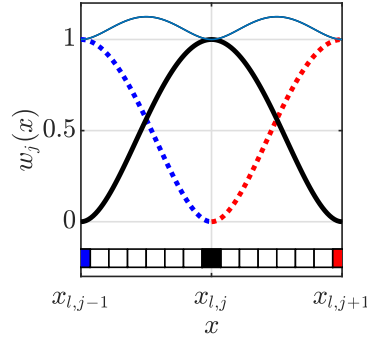


Fig. 4: Jacobi polynomial with $k = 1$; *black*: local weighting function $w_j(x)$ for approximating s_j ; *blue*: sum of the local weighting functions; *blue, red dotted*: parts of the weighting functions $w_{j-1}(x)$ and $w_{j+1}(x)$ for generating the neighbouring state vectors s_{j-1} and s_{j+1} ; *bottom*: schematic visualization of the position of original data points to be weighted.

Spatial Weighted Local Regression: After choosing a weighting function, weighted regression is performed. The points and the observed values within the segment j are denoted as $x_{l,j}$ and $\hat{y}_{l,j}$. The according weightings are collected in the vector $w_j = w_j(x_{l,j})$. This vector is expanded to form the diagonal weighting matrix $W_j = \text{diag}\{w_j\}$. If a linear model of the form

$$y_{l,j} = B_j \alpha_j \quad (5)$$

is used to model the signal, the cost function to be minimized can be written as

$$\varepsilon = (\hat{y}_{l,j}^T - \alpha_j^T B_j^T) W_j (\hat{y}_{l,j} - B_j \alpha_j), \quad (6)$$

which is the sum of the weighted squared errors. B_j is a basis function set and α_j are the according coefficients for modelling the signal. Now the matrix square root of W is defined as

$$U_j = W_j^{1/2} \quad \text{with} \quad U_j U_j = W_j. \quad (7)$$

Note: if W_j is positive semi-definite and symmetric, the numerically more stable Cholesky factorization ($W_j = U_j^T U_j$) can be used instead of the matrix square root. Minimizing (6) with respect to α_j delivers the least-squares solution, i.e.,

$$\alpha_j = \{U_j B_j\}^+ U_j \hat{y}_{l,j}, \quad (8)$$

where $\{U_j B_j\}^+$ denotes the pseudoinverse of $U_j B_j$.

To calculate the state vector s_j , the model and the derivatives of the model up to order d are evaluated at the point x_j , yielding

$$s_j = \begin{bmatrix} y_j^{(0)} \\ \vdots \\ y_j^{(d)} \end{bmatrix} = \begin{bmatrix} B_j^{(0)}(x_j) \\ \vdots \\ B_j^{(d)}(x_j) \end{bmatrix} \alpha_j = \tilde{B}_j \alpha_j, \quad (9)$$

where $y_j^{(m)}$ is the approximation of the m -th derivative of the signal at the point x_j , and $B_j^{(m)}(x_j)$ is the m -th derivative of the basis function set B_j evaluated at x_j , so that

$$y_j^{(m)} = B_j^{(m)}(x_j)\alpha_j. \quad (10)$$

\tilde{B}_j is the concatenated matrix containing the basis functions and their derivatives evaluated at the point of interest.

Covariance propagation: Additionally, given the covariance $\Lambda_{l,j}$ associated with the original data in the segment j , the covariance for the state vector $\Lambda_{s,j}$ can be propagated using first order covariance propagation, i.e.,

$$\Lambda_{s,j} = \tilde{B}_j A_j \Lambda_{l,j} A_j^T \tilde{B}_j^T \quad (11)$$

with

$$A_j = \{U_j B_j\}^+ U_j. \quad (12)$$

The covariance matrix $\Lambda_{l,j}$ is either known a priori (e.g. knowing the error associated with the sensor used for observation) or can be calculated from the data, e.g., via the norm of the residual vector (especially when using a polynomial basis).

The impact of the weighting function onto the covariance of the approximation is shown in Fig. 5. As a basis function set a Vandermonde basis (polynomial basis) of degree 4 is chosen along with an i.i.d noise on the input. Clearly it can be seen that the piecewise constant weighting function performs badly outside its support due to the Runge phenomenon [6, 25].

The above calculation is repeated for each segment $x_{l,j}$ to approximate the decimated signal at the points x_j . If the signal is equally spaced and the window size does not change, the matrices W_j , B_j and A_j are the same for each interval, i.e. $W_j = W$, $B_j = B$ and $A_j = A$ for $j = 1 \dots n_j$. This is of major advantage, since these matrices can be calculated a priori and the approximation of the state vectors (see Eq. 8) reduces to the single linear mapping $\alpha_j = A \hat{y}_{l,j}$, which can be easily implemented on embedded devices and smart sensors collecting the data. This makes the presented method suitable for collecting real-time machine data. As shown later, the quality of approximation is improved using state vectors instead of standard approximation techniques using only value information.

The right choice of the basis functions used for approximation depends on the observed system. The most common basis function set, also used in this work, is the Vandermonde basis (polynomial basis). To improve stability, it is referred to discrete orthogonal basis functions [16, 18]. The availability of the state vectors opens the door to analyse signals in the so-called *pseudo phase space* (e.g. investigating Poincaré recurrence times [21]). This is especially interesting for analysing dynamic systems.

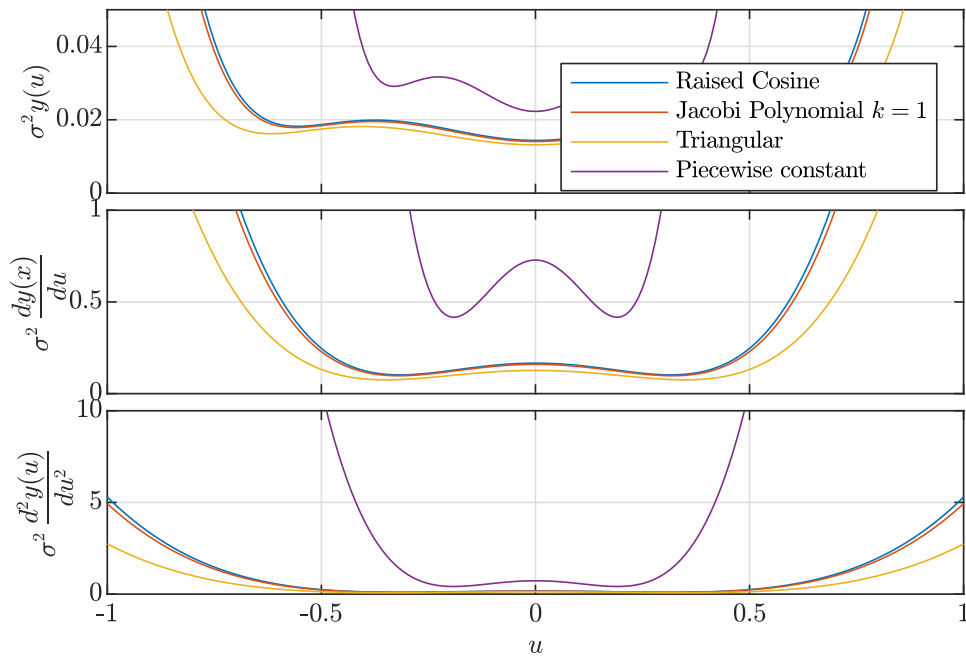


Fig. 5: This figure shows from top to bottom the propagated variances of $y(u)$, $y^{(1)}(u)$ and $y^{(2)}(u)$ respectively, for some weighting functions. Note that in all cases there is a significant improvement in the variance compared to the piecewise constant window.

2.2 Hermite Approximation - Hierarchy Level 2 and above

In the first level of the presented hierarchical approximation the signal and its derivatives (concatenated in the state vector \hat{s}_j)² are approximated at points x_j . Additionally, the covariance associated with each state vector is available. The approximation of such signals (given the values and their derivatives) is not common in literature. It is related to Hermite polynomials [2]. In [1] Hermite approximation is used for multidimensional surface approximation. Hermite weighting functions are used in [12] to perform local polynomial approximation. An application using two-point Hermite approximation for solving initial value and boundary value problems can be found in [14]. The idea of reconstructing signals given the first derivatives and additional value information at some points is presented in [19]. Based on this work, we introduce a new method of doing least squares Hermite approximation using covariance weighting which extends the work to higher order derivative information. It is used to approximate higher levels of the hierarchical process presented in this paper.

Algebraic Formulation Given a system model in terms of a linear combination of basis functions, the goal is to find the according coefficients that approximate the given observations (values as well as derivatives) best, i.e., minimizing the residuals in a least squares sense. Therefore, a consistent measure for the residuals in both, the value and

²The $\hat{\cdot}$ notation is used here, since the given state vector is the 'noisy' input for the next level of approximation.

derivative domain, is necessary. Since the covariance information for the value and the derivatives is available at each point, it can be used to weight the residual according to this. That is, if a given information is precise, it has more impact on the solution.

First, the matrix H_j is assembled as

$$H_j = \begin{bmatrix} \mathbf{h}(x_j) \\ \mathbf{h}^{(1)}(x_j) \\ \vdots \\ \mathbf{h}^{(d)}(x_j) \end{bmatrix}, \quad (13)$$

where $\mathbf{h}(x_j)$ denotes the basis function vector (e.g. Vandermonde vector). Each column of this vector contains one of the basis functions evaluated at the point of interest x_j . $\mathbf{h}^{(m)}(x_j)$ is the m -th derivative of the basis function evaluated at x_j and d is the number of derivatives given in the state vector $\hat{\mathbf{s}}_j$. The local observed state vector $\hat{\mathbf{s}}_j$ is now approximated by $\mathbf{s}_j = H_j \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ denotes the coefficient vector. Consequently, the local residual vector \mathbf{r}_j is given by,

$$\mathbf{r}_j = \hat{\mathbf{s}}_j - \mathbf{s}_j \quad (14)$$

$$= \hat{\mathbf{s}}_j - H_j \boldsymbol{\beta}. \quad (15)$$

Additionally, for each $\hat{\mathbf{s}}_j$ we have the corresponding covariance $\Lambda_{s,j}$; consequently we can define an inverse covariance weighted computation of the error $\boldsymbol{\varepsilon}_j$,

$$\boldsymbol{\varepsilon}_j = \mathbf{r}_j^T \Lambda_{s,j}^{-1} \mathbf{r}_j \quad (16)$$

$$= \mathbf{r}_j^T W_{C,j} \mathbf{r}_j \quad (17)$$

$$= (\hat{\mathbf{s}}_j - \mathbf{s}_j)^T W_{C,j} (\hat{\mathbf{s}}_j - \mathbf{s}_j) \quad (18)$$

$$= (\hat{\mathbf{s}}_j - H_j \boldsymbol{\beta})^T W_{C,j} (\hat{\mathbf{s}}_j - H_j \boldsymbol{\beta}). \quad (19)$$

$W_{C,j}$ denotes the covariance weighting matrix for the point j . All the individual local estimates for the state variables can be vertically concatenated to obtain the normal equations for the minimization problem. Let us start by defining

$$\Lambda \triangleq \begin{bmatrix} \Lambda_{s,1} & 0 & \dots & 0 \\ 0 & \Lambda_{s,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_{s,n} \end{bmatrix}, \quad H \triangleq \begin{bmatrix} H_1 \\ \vdots \\ H_n \end{bmatrix}, \quad \hat{\mathbf{s}} \triangleq \begin{bmatrix} \hat{\mathbf{s}}_1 \\ \vdots \\ \hat{\mathbf{s}}_n \end{bmatrix}. \quad (20)$$

Given Λ , the weighting matrix W_C is computed as,

$$W_C \triangleq \Lambda^{-1} = \begin{bmatrix} \Lambda_{s,1}^{-1} & 0 & \dots & 0 \\ 0 & \Lambda_{s,2}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_{s,n}^{-1} \end{bmatrix} \quad (21)$$

Since all data is approximated with the same underlying function, β is the same for each point and the sum of the covariance weighted errors can be written as

$$\varepsilon = \sum_{j=1}^n \varepsilon_j = \left(\begin{bmatrix} \hat{s}_1 \\ \vdots \\ \hat{s}_n \end{bmatrix} - \begin{bmatrix} H_1 \\ \vdots \\ H_n \end{bmatrix} \beta \right)^T \begin{bmatrix} \Lambda_{s,1}^{-1} & 0 & \cdots & 0 \\ 0 & \Lambda_{s,2}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda_{s,n}^{-1} \end{bmatrix} \left(\begin{bmatrix} \hat{s}_1 \\ \vdots \\ \hat{s}_n \end{bmatrix} - \begin{bmatrix} H_1 \\ \vdots \\ H_n \end{bmatrix} \beta \right). \quad (22)$$

This is again a weighted regression problem of the form

$$\min_{\beta} \varepsilon = \min_{\beta} (\hat{s} - H\beta)^T W_C (\hat{s} - H\beta) \quad (23)$$

with the solution,

$$\beta = \left(W_C^{1/2} H \right)^+ W_C^{1/2} \hat{s}. \quad (24)$$

Now defining,

$$L \triangleq \left(W_C^{1/2} H \right)^+ W_C^{1/2}, \quad (25)$$

we also obtain the covariance for β , i.e.,

$$\Lambda_{\beta} = L \Lambda L^T. \quad (26)$$

Local Hermite Approximation The algebraic formulation given above is a generalized version for approximating data from given values and its derivatives. For the proposed hierarchical algorithm this approximation is applied locally to a window k which spans only a certain number of the state vectors approximated in level 1. This is shown in Fig. 2. As a result we get approximations for the coefficients for each window denoted as β_k . Additionally, spatial weighting can be added as described in Sect. 2.1. When using spacial weighting captured in the weighting matrix $W_{S,k}$ along with the covariance weighting $W_{C,k}$, the matrix W_C in (21)-(25) is replaced by the matrix product of the two local weighting matrices $W_{SC,k} = W_{S,k} W_{C,k}$. To implement decimation also in this hierarchical level, the approximation of the state vectors is only done for the centre x_k of the local window, yielding s_k , i.e.,

$$s_k = H_k L_k \hat{s}_k. \quad (27)$$

H_k denotes the matrix of basis functions (and their derivatives) evaluated at the point x_k as given in (13), L_k denotes the local version of (25) and \hat{s}_k is the concatenation of all state vectors derived in level 1 within the local window. Again, the covariance for this level of state vectors can be propagated as

$$\Lambda_{s,k} = H_k L_k \Lambda_k L_k^T H_k^T, \quad (28)$$

where Λ_k is the local version of Λ .

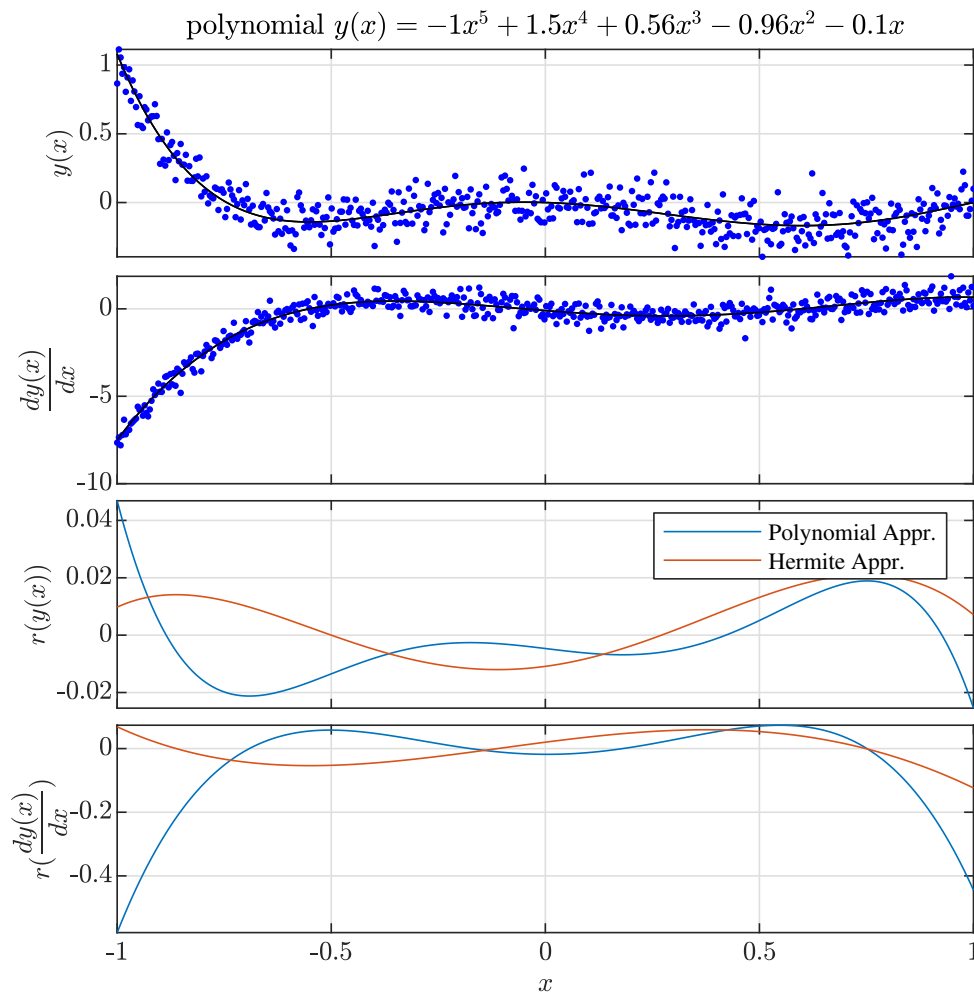


Fig. 6: Synthetic data used for Monte-Carlo simulation. The two plots on the top show the function and its derivative with the addition of Gaussian noise. The two plots on the bottom show residuals for the approximation of the values and the derivatives for both, the Hermite approximation and a standard polynomial fit, with twice the number of samples (same amount of data).

Performance Test A Monte-Carlo experiment with $n = 10000$ iterations revealed that using Hermite approximation is of advantage, if both, the signal and the derivatives, should be approximated well. Therefore, a synthetic polynomial data set (see Fig. 6) was generated with known values and first derivatives. A Gaussian noise with $\sigma_y = 0.1$ was added to the values and noise with $\sigma_{dy} = 0.4$ was added to the first derivatives. The covariance weighted regression was performed without spatial weighting. The result was compared to standard polynomial fitting with the doubled number of samples but without derivative information (to provide an identical amount information in both methods, to make them comparable). As a measure, the weighted norm of the residual vector $\mathbf{r}_y = \mathbf{y}_{\text{approx}} - \mathbf{y}_{\text{orig}}$ and $\mathbf{r}_{dy} = \dot{\mathbf{y}}_{\text{approx}} - \dot{\mathbf{y}}_{\text{orig}}$ was taken. The results are presented in Table 1. As it can be seen, Hermite approximation performs around 2 times better than the standard polynomial fit for approximating derivatives. On the other hand, this

Table 1: Result of the Monte-Carlo experiment.

	$\frac{1}{\sigma_y} \ \mathbf{r}_y\ _2$	$\frac{1}{\sigma_{dy}} \ \mathbf{r}_{dy}\ _2$
Hermite approximation	0.0282	0.0262
Standard polynomial fit	0.0245	0.0443

leads to a slightly worse performance in approximating values. As to be expected, the norm of the residual for the values and derivatives are nearly the same for the Hermite approximation, due to the fact of the covariance weighting. This is not the case for standard polynomial fitting. Note: the behaviour at the ends of the interval is also better for Hermite approximation (see Fig. 6).

As a result, it is suggested to provide state vectors sampled at a lower frequency instead of using only value information sampled with full sampling frequency. The amount of data does not change, whereas the quality of approximating derivatives is improved. This fact can be considered in future design of smart sensors and IoT devices collecting data of dynamic systems.

2.3 Data Reconstruction

In this section, methods for the reconstruction of the signal are investigated. During the presented hierarchical approximation, the signal is decimated, and state vectors are only available at certain points x_k . In some applications, it is necessary to describe the signal analytically to provide the possibility to do calculations at arbitrary locations and not only at discrete points. This is known as *Interpolation*. In literature, especially in digital signal processing, a lot of methods are well-established [4]. Basically, there are two main categories:

1. Local interpolation: e.g, piecewise constant, linear or spline interpolation. For each segment, a different interpolating function is used. If one point changes, only neighbouring segments are affected. Note: this is not fully true for splines, since a change in one point (knot) does affect a wider range of segments, based on the degree of continuity to be fulfilled.
2. Global interpolation: this is related to approximation, e.g., polynomial interpolation, trigonometric interpolation. A change in one point does affect the whole range to be interpolated. These types also suffer from problems which arise due to overfitting.

Taylor Expansion: In the presented hierarchical method, state vectors, containing value and derivative information, are available. A straight forward possibility is to use a one-point expansion (i.e Taylor expansion) to interpolate between two given points, i.e.,

$$f_k(x) = y_k + y_k^{(1)} (x - x_k) + \dots + \frac{y_k^{(d)}}{d!} (x - x_k)^d, \quad (29)$$

where $f_k(x)$ denotes the interpolating function based on the state vector \mathbf{s}_k at the position x_k . The order of derivatives available in the state vector is denoted by d . Since the interpolating function is a polynomial, the derivatives are simple to calculate as well. For q discrete points x_i , the interpolated points \mathbf{y}_k can be calculated as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{bmatrix} = \begin{bmatrix} (x_1 - x_k)^d & \dots & (x_1 - x_k) & 1 \\ (x_2 - x_k)^d & \dots & (x_2 - x_k) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ (x_q - x_k)^d & \dots & (x_q - x_k) & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{d!} & 0 & \dots & 0 \\ 0 & \frac{1}{(d-1)!} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} y_k^{(d)} \\ y_k^{(d-1)} \\ \vdots \\ y_k \end{bmatrix}. \quad (30)$$

or in matrix vector equation as

$$\mathbf{y}_k = \mathbf{V} \mathbf{F} \mathbf{s}_k, \quad (31)$$

where \mathbf{V} denotes the Vandermonde matrix for the expansion around x_k and \mathbf{F} holds the scaling values resulting from Taylor approximation. For interpolating derivatives of degree d , one can use

$$\mathbf{y}_k^{(d)} = \mathbf{V}^{(d)} \mathbf{F} \mathbf{s}_k, \quad (32)$$

where $\mathbf{V}^{(d)}$ is the d -th derivative of the Vandermonde matrix.

A major problem of using this method is that discontinuities occur where the interpolating functions $f_k(x)$ and $f_{k+1}(x)$ meet (see Fig. 7).

Generalized Hermite Interpolation: To overcome the problem of discontinuities, we take the idea of Hermite Interpolation [9] and extend it to higher degrees. Given n state vectors, each with d states (order of derivatives) at each point x_k , the polynomial fulfilling all the given constraints would be at most of degree $p = nd - 1$. The coefficients $\boldsymbol{\rho}$ for this are computed by solving

$$\min_{\boldsymbol{\rho}} \left\| \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(d)} \end{bmatrix} - \begin{bmatrix} \mathbf{V}_p \\ \mathbf{V}_p^{(1)} \\ \vdots \\ \mathbf{V}_p^{(d)} \end{bmatrix} \boldsymbol{\rho} \right\|_2^2 = \min_{\boldsymbol{\rho}} \|\tilde{\mathbf{y}} - \tilde{\mathbf{V}}_p \boldsymbol{\rho}\|_2^2, \quad (33)$$

where $\mathbf{y}, \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(d)}$ are the vectors which hold the values and the derivatives for each point x_k used for interpolation. \mathbf{V}_p denotes the Vandermonde matrix of degree p and $\mathbf{V}_p^{(i)}$ denotes the i -th derivative of this Vandermonde matrix evaluated at the interpolating points. If $\tilde{\mathbf{V}}_p$ is full rank, the interpolating polynomial is unique. In this case, the solution is given as,

$$\boldsymbol{\rho} = \tilde{\mathbf{V}}_p^{-1} \tilde{\mathbf{y}}. \quad (34)$$

If the covariances of the state vectors are available, covariance propagation can be calculated as in the above methods.

The presented hierarchical method proposes to use this generalized Hermite interpolation to interpolate between two neighbouring points, which is a two-point expansion. Thus, a change of one state vector does only influence the two neighbouring segments.

With this method, the resulting curve (which is piecewise polynomial) is at least C^d -continuous. This type of interpolation is closely related to splines. In Fig. 7 both, the Taylor expansion and the generalized Hermite interpolation, are demonstrated. Both methods use a cubic polynomial for interpolation. As it can be seen in the plots, the Taylor expansion is discontinuous at points where two functions from neighbouring intervals meet. This is not the case for the Hermite interpolation. At the end of the interval, the Taylor expansion performs better due to the fact of being a single-point expansion with the same degree as the two-point expansion of the Hermite interpolation.

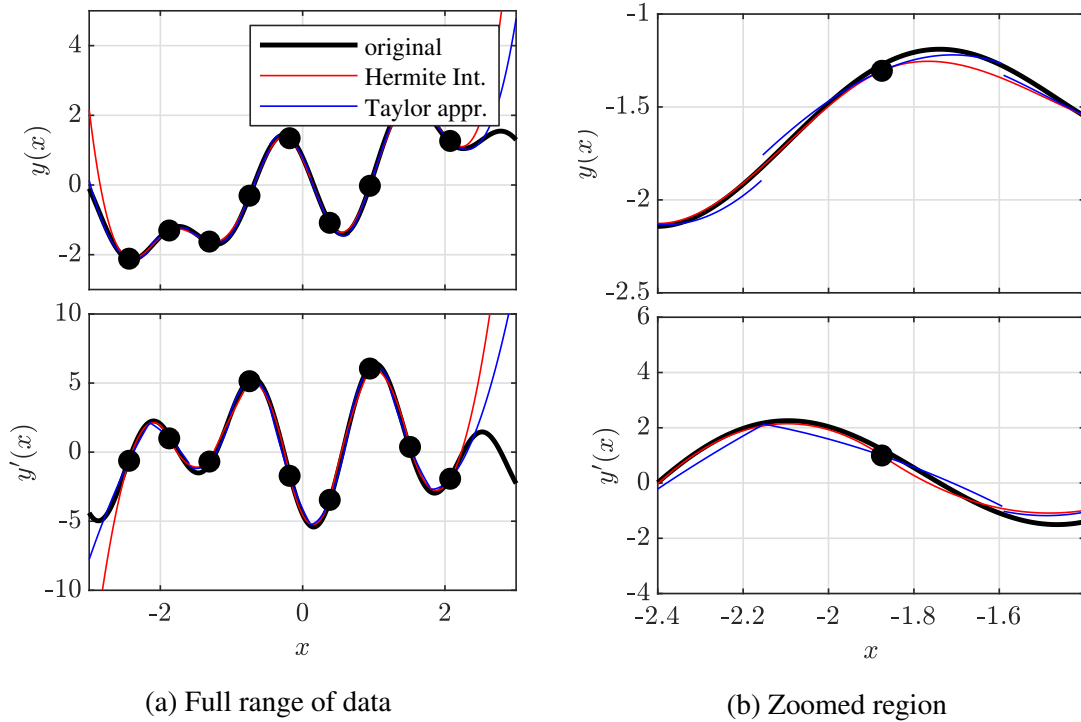


Fig. 7: Hermite interpolation vs Taylor expansion.

If these methods are used to interpolate from L_1 to L_0 , a consistency check can be made simultaneously to identify discontinuities in the sampled data. This can be used to trigger additional state vector samples at these point. However, this is not in the scope of this paper.

3 Numerical Testing

To show the abilities of the herein proposed hierarchical approximation of data, the method is tested on a synthetic dataset. The test data originate from the function

$$y(x) = \sin(x) + \sin(3 + 2.5x) + \sin(15 + 4x) + \frac{x^3}{25} \quad (35)$$

with the analytical first derivative

$$\frac{dy}{dx}(x) = \cos(x) + 2.5 \cos(3 + 2.5x) + 4 \cos(15 + 4x) + \frac{3x^2}{25}. \quad (36)$$

The function is sampled at $n = 2001$ equidistant locations. A Gaussian noise with σ_y was added to the function. To generate the first level of the hierarchical approximation (L_1), a local window covering $n_{w,1} = 75$ points from the sampled signal is chosen. This results in a decimated signal with a distance of $l_1 = 37$ data points between the approximated state vectors s_j . As a local model, a polynomial of degree 1 is chosen, which can be modelled using a Vandermonde matrix ($B_j = V_1$). Returned are state vectors of dimension $d = 1$ containing the approximated value and first derivative. As a weighting function the piecewise constant and a raised cosine (weighted) are used for demonstration. The results are shown in Fig. 8. As it can be seen, the raised cosine

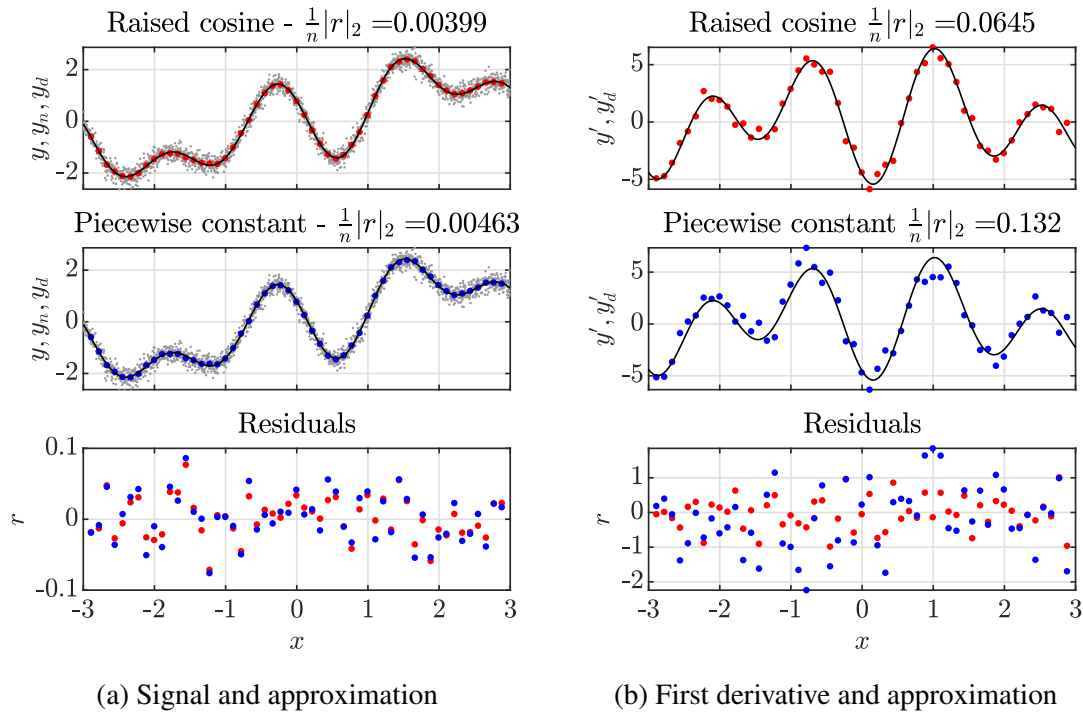


Fig. 8: L_1 Spatial weighted approximation; *top* and/or *red*: raised cosine weighting; *middle* and/or *blue*: piecewise constant weighting; *bottom*: residuals of approximation; *black line*: original function values; *gray*: noisy data.

weighting performs better than the piecewise constant, which is expected. Especially the approximation of the first derivative is of better quality.

After generating L_1 , covariance weighted Hermite approximation is used to generate the subsequent levels (L_2 and L_3). Therefore, a local window, containing $n_{w,2} = n_{w,3} = 5$ state vectors was used, resulting in a decimated signal with a distance of $l_2 = 74$ and $l_3 = 148$ data points in terms of the original signal. This corresponds to a compression c -ratio= 74 for level 2 and c -ratio= 148 for level 3. For the approximation, a polynomial

of degree 2 is used. The results for both, L_2 and L_3 , are shown in Fig. 9. Although the

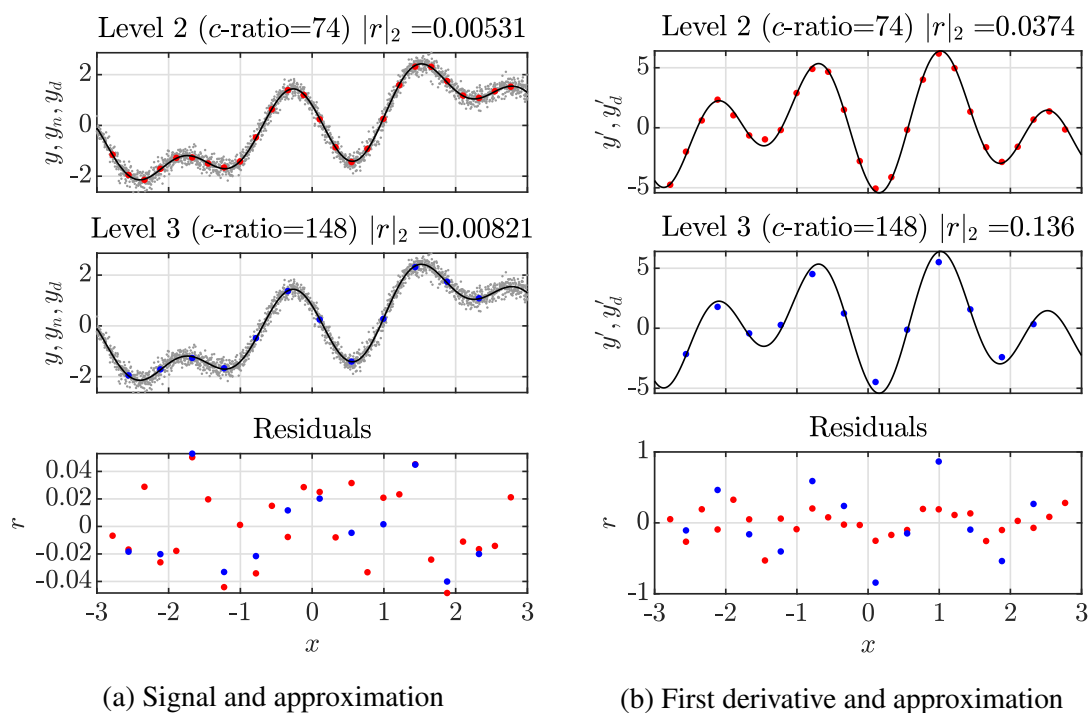


Fig. 9: L_2 and L_3 Hermite approximation (covariance and spatial weighted); *top* and/or *red*: L_2 approximation; *middle* and/or *blue*: L_3 approximation; *bottom*: residuals of approximation; *black line*: original function values; *gray*: noisy data.

compression ratio is high, the hierarchical approximation delivers good results for approximating the values and also the derivatives. This is important, if the approximation of the derivative is used for further calculation.

In Fig. 10 the given state vectors from L_2 and L_3 are used to interpolate the signal at the original locations. The proposed two-point expansion using generalized Hermite interpolation was used. Additionally, covariance propagation was performed through all levels. The resulting variance after interpolation is visualized in the figure as well. Note: different magnification gains are used to make the variance visible in the plots. Again, the proposed method delivers suitable results. As it can be seen, the covariance propagation is getting worse at the end of the interval, which is due to the fact that at the ends the signal is extrapolated, since the supporting points are missing.

4 Conclusion

This paper has presented a new method for hierarchical approximation of sensor data along with all derivations. In the first level spatial weighting was used to approximate state vectors at collocated locations. Different weighting functions have been investigated by analysing their covariance propagation. It has been shown that a piecewise

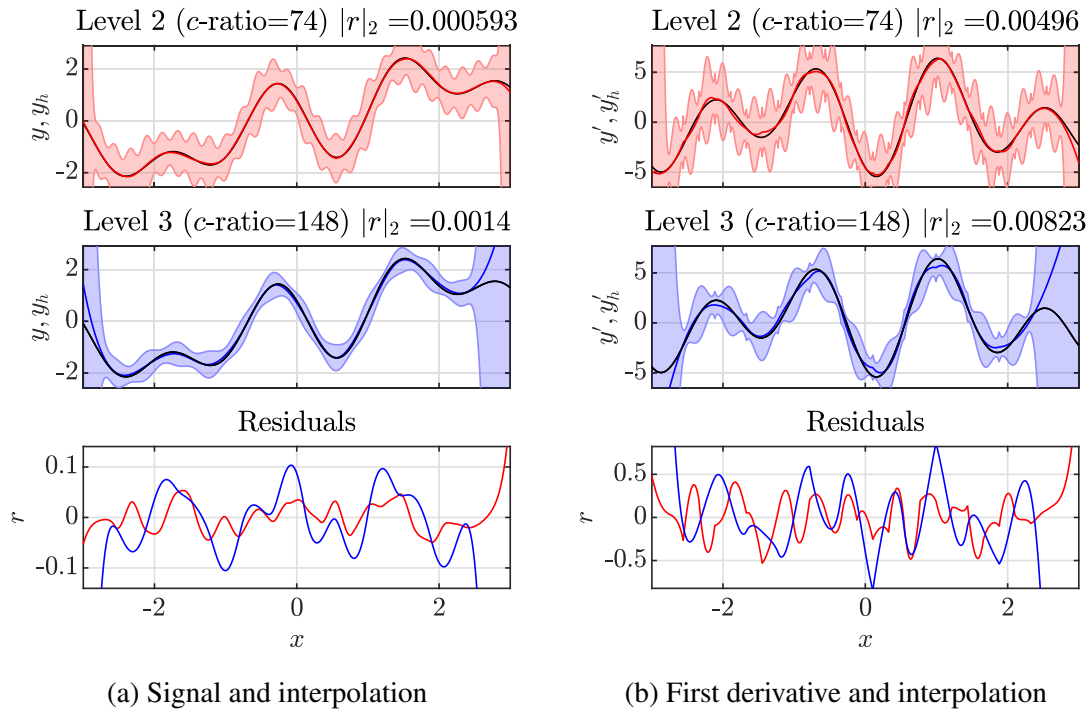


Fig. 10: Hermite interpolation based on L_2 and L_3 ; *top* and/or *red*: L_2 interpolation; *middle* and/or *blue*: L_3 interpolation; *bottom*: residual of interpolation; *black line*: original function values; *shaded areas*: propagated covariances (magnified with $g = 2e3$ on the left, $g = 8e1$ top-right and $g = 4e2$ middle-right).

constant weighting function is not the method of choice. In subsequent levels of the hierarchy, a new method of covariance weighted Hermite approximation is proposed to approximate the signal by the given value and derivative information. This yields a decimated signal which maintains derivative information. This method was compared to a standard fitting method and revealed a large improvement in approximating derivatives by only a minor decrease of quality in approximating values. Based on this, it can be concluded that future sensors should deliver the state vector instead of a higher frequent signal without derivative information. To make use of the approximated state vectors, two interpolation methods are presented which approximate the original signal in a continuous sense. The presented generalized Hermite interpolation proved to be the method of choice, since the signal and its derivatives are continuous within the whole interval, which is beneficial for further derivations based on the signal. A successful test on synthetic data showed the correct functionality of the proposed hierarchical method and delivered good results for the approximation of large data sets, especially for the approximation of derivatives.

Acknowledgments

Partial funding for this work was provided by the Austrian research funding association (FFG) under the scope of the COMET program within the K2 center “IC-MPP”

(contract number 859480). This programme is promoted by BMVIT, BMDW and the federal states of Styria, Upper Austria and Tyrol.

References

1. Bajaj, C.L.: Multi-dimensional Hermite Interpolation and Approximation for Modelling and Visualization. In: Proceedings of the IFIP TC5/WG5.2/WG5.10 CSI International Conference on Computer Graphics: Graphics, Design and Visualization, pp. 335–348. North-Holland Publishing Co., Amsterdam, The Netherlands, The Netherlands (1993). URL <http://dl.acm.org/citation.cfm?id=645465.653690>
2. Burden, R.L., Faires, J.D.: Numerical Analysis, 872 edn. Cengage Learning. Thomson Brooks/Cole (2005). URL <http://books.google.at/books?id=wmcL0y2avuUC>
3. Cleveland, W.S.: Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**(368), 829 (1979). DOI 10.2307/2286407. URL <http://www.jstor.org/stable/2286407>
4. Crochiere, R.E., Rabiner, L.R.: Multirate digital signal processing. Prentice-Hall (1983). URL <https://de.scribd.com/doc/82733332/Multirate-Digital-Signal-Processing-Crochiere-Rabiner>
5. Eilers, P.H.C.: A Perfect Smoother. *Analytical Chemistry* **75**(14), 3631–3636 (2003). DOI 10.1021/ac034173t
6. Epperson, J.F.: On the Runge Example. *The American Mathematical Monthly* **94**(4), 329 (1987). DOI 10.2307/2323093. URL <https://www.jstor.org/stable/2323093?origin=crossref>
7. Grabocka, J., Wistuba, M., Schmidt-Thieme, L.: Scalable Classification of Repetitive Time Series Through Frequencies of Local Polynomials. *IEEE Transactions on Knowledge and Data Engineering* **27**(6), 1683–1695 (2015). DOI 10.1109/TKDE.2014.2377746. URL <http://ieeexplore.ieee.org/document/6975152/>
8. Gupta, S., Ray, A., Keller, E.: Symbolic time series analysis of ultrasonic data for early detection of fatigue damage. *Mechanical Systems and Signal Processing* **21**(2), 866–884 (2007). DOI 10.1016/j.ymssp.2005.08.022. URL <http://linkinghub.elsevier.com/retrieve/pii/S0888327005001329>
9. Hermite, C.: Sur la formule d’interpolation de Lagrange. (Extrait d’une lettre de M. Ch. Hermite à M. Borchardt). *Journal für die reine und angewandte Mathematik* **84**, 70–79 (1877). URL <http://eudml.org/doc/148345>
10. Jerri, A.J.: The Gibbs phenomenon in Fourier analysis, splines, and wavelet approximations. Kluwer Academic Publishers (1998)
11. Joldes, G.R., Chowdhury, H.A., Wittek, A., Doyle, B., Miller, K.: Modified moving least squares with polynomial bases for scattered data approximation. *Applied Mathematics and Computation* **266**, 893–902 (2015). DOI 10.1016/j.amc.2015.05.150. URL <http://dx.doi.org/10.1016/j.amc.2015.05.150>
12. Komargodski, Z., Levin, D.: Hermite type moving-least-squares approximations. *Computers & Mathematics with Applications* **51**(8), 1223–1232 (2006). DOI 10.1016/j.camwa.2006.04.005. URL <http://linkinghub.elsevier.com/retrieve/pii/S0898122106000757>
13. Marron, J.S., Hill, C.: Local Polynomial Smoothing Under Qualitative Constraints. *COMPUTING SCIENCE AND STATISTICS* pp. 647—652 (1997)
14. Mennig, J., Auerbach, T., Hälgl, W.: Two point hermite approximations for the solution of linear initial value and boundary value problems. *Computer Methods in Applied Mechanics and Engineering* **39**(2), 199–224 (1983). DOI 10.1016/0045-7825(83)90021-X
15. Moore, A.W., Schneider, J., Deng, K.: Efficient locally weighted polynomial regression predictions. *International Conference on Machine Learning* pp. 236–244 (1997)

16. O’Leary, P., Harker, M.: Discrete polynomial moments and Savitzky-Golay smoothing. World Academy of Science, Engineering and Technology; International Journal of Computer and Information Engineering **72**, 439–443 (2010). URL <https://waset.org/publications/12268/discrete-polynomial-moments-and-savitzky-golay-smoothing>
17. O’Leary, P., Harker, M.: Surface Modelling Using Discrete Basis Functions for Real-Time Automatic Inspection. In: 3-D Surface Geometry and Reconstruction, pp. 216–264. IGI Global (2010). DOI 10.4018/978-1-4666-0113-0.ch010. URL <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-0113-0.ch010>
18. O’Leary, P., Harker, M.: A Framework for the Evaluation of Inclinator Data in the Measurement of Structures. IEEE Transactions on Instrumentation and Measurement **61**(5), 1237–1251 (2012). DOI 10.1109/TIM.2011.2180969. URL <http://ieeexplore.ieee.org/document/6162983/>
19. O’Leary, P., Harker, M.: Inverse boundary value problems with uncertain boundary values and their relevance to inclinometer measurements. In: 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, pp. 165–169. IEEE (2014). DOI 10.1109/I2MTC.2014.6860725. URL <http://ieeexplore.ieee.org/document/6860725/>
20. O’Leary, P., Harker, M., Neumayr, R.: Savitzky-Golay smoothing for multivariate cyclic measurement data. 2010 IEEE International Instrumentation and Measurement Technology Conference, I2MTC 2010 - Proceedings pp. 1585–1590 (2010). DOI 10.1109/IMTC.2010.5488242
21. Poincaré, H.: Sur le probleme des trois corps et les équations de la dynamique. Acta Mathematica **13**(1), 5–7 (1890). DOI 10.1007/BF02392506. URL <http://link.springer.com/10.1007/BF02392506>
22. Proietti, T., Luati, A.: Low-pass filter design using locally weighted polynomial regression and discrete prolate spheroidal sequences. Journal of Statistical Planning and Inference **141**(2), 831–845 (2011). DOI 10.1016/j.jspi.2010.08.006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378375810003769>
23. Racine, J.S.: Local Polynomial Derivative Estimation: Analytic or Taylor? In: Essays in Honor of Aman Ullah (Advances in Econometrics), vol. 36, chap. 18, pp. 617–633. Emerald Group Publishing Limited (2016). DOI 10.1108/S0731-905320160000036027. URL <http://www.emeraldinsight.com/doi/10.1108/S0731-905320160000036027>
24. Rajagopalan, V., Ray, A., Samsi, R., Mayer, J.: Pattern identification in dynamical systems via symbolic time series analysis. Pattern Recognition **40**(11), 2897–2907 (2007). DOI 10.1016/j.patcog.2007.03.007
25. Runge, C.: Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. Zeitschrift für Mathematik und Physik **46**, 224–243 (1901)

8 | Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials

Information: The appended paper is accepted for publication in the proceedings of *2019 IEEE International conference on Industrial Cyber-Physical Systems (ICPS 2019) / 2019 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2019)*¹. The proceedings will be later available on *IEEE Xplore*².

Appended version will appear as:³

R. Ritt, M. Harker, and P. O’Leary, “Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials,” in *2019 IEEE Industrial Cyber-Physical Systems (ICPS)*, Taipei: IEEE, 2019

BibTeX:³

```
@inproceedings{Ritt2019b,
  author      = {Ritt, Roland and Harker, Matthew
                and O'Leary, Paul},
  booktitle   = {2019 IEEE Industrial Cyber-Physical Systems (
                ICPS)},
  publisher   = {IEEE},
  title       = {{Simultaneous Approximation of
                Measurement Values and Derivative Data using
                Discrete Orthogonal Polynomials}},
  year        = {2019}
}
```

¹<http://icps19.org/>

²<https://ieeexplore.ieee.org/>

³Since this paper was not published at the time this thesis is finished, the citation may change.

A preprint of this paper originally appeared on arXiv⁴ as:

R. Ritt, M. Harker, and P. O’Leary, “Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials,” *arXiv Open Access Journal Article*, Mar. 2019. arXiv: 1903.10810. [Online]. Available: <http://arxiv.org/abs/1903.10810>

BibTeX for Preprint:

```
@article{Ritt2019a,
  archivePrefix = {arXiv},
  arxivId      = {1903.10810},
  author       = {Ritt, Roland and Harker, Matthew
                 and O'Leary, Paul},
  journal      = {arXiv Open Access Journal Article},
  month        = {mar},
  title        = {{Simultaneous Approximation of
                 Measurement Values and Derivative Data using
                 Discrete Orthogonal Polynomials}},
  url          = {http://arxiv.org/abs/1903.10810},
  year        = {2019}
}
```

⁴<https://arxiv.org/>

Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials

Roland Ritt

Chair of Automation
University of Leoben
Leoben, Austria
roland.ritt@unileoben.ac.at

Matthew Harker

Chair of Automation
University of Leoben
Leoben, Austria
matthew.harker@unileoben.ac.at

Paul O’Leary

Chair of Automation
University of Leoben
Leoben, Austria
paul.oleary@unileoben.ac.at

Abstract—This paper presents a new method for polynomial approximation using the fusion of value and derivative information emanating from different sources, i.e., sensors. Therefore, the least-squares error in both domains is simultaneously minimized. A covariance weighting is used to introduce a metric between the value and derivative domain, to handle different noise behaviour. Based on a recurrence relation with full re-orthogonalization, a weighted polynomial basis function set is generated. This basis is numerically more stable compared to other algorithms, making it suitable for the approximation of data with high degree polynomials. With the new method, the fitting problem can be solved using inner products instead of matrix-inverses, yielding a computational more efficient method, e.g., for real-time approximation.

A Monte Carlo simulation is performed on synthetic data, demonstrating the validity of the method. Additionally, various tests on the basis function set are presented, showing the improvement on the numerical stability.

Index Terms—multisensor fusion, optimization, discrete orthogonal polynomials, basis functions, Hermite approximation

I. MOTIVATION

Measurements with a fusion of value and derivative information are common in geotechnical monitoring. There are measurement cases where observations from different measurement devices must be fused to obtain optimal results. An example in case is, when total stations (theodolites) and inclinometers are used to monitor structural degradation, e.g. Fig. 1 and [1]. In this example the total station measures position as a function of time and the inclinometers measure the first spatial derivatives of position as a function of time. Consequently, the function and its first derivative are measured independently and thus are not correlated in the common case. The task now at hand is to compute an optimal estimate for the deflection from the two independent observation sources with the presence of perturbations, to detect unwanted changes or subsidence. Thus, there is an explicit need for methods which perform approximation based on function values and their measured derivatives.

In the case shown in Fig. 1 the total station measurements are significantly more accurate and stable with respect to the

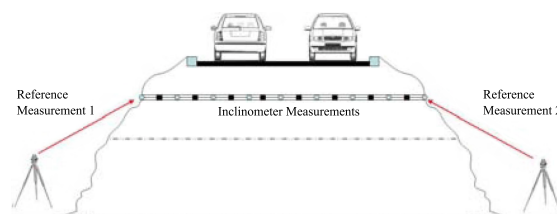


Fig. 1. Schematic of geotechnical monitoring using inclination measurements in combination with reference measurements (total stations, theodolites)

inclinometer data so that they were considered as reference measurements. Subsequently, they were used as constraints in the approximation. The generalization of the above issue is the development of an approximation method which utilizes independent observations of a function and its derivatives. This is a special case of sensor data fusion where there is no a-priori relationship between the two sources of data.

Especially in engineering problems, polynomials are the model of choice for the approximation of such data, due to their properties and their relation to the underlying physical model [2]. Hermite [3] provided a method of performing interpolation given function values $y(x)$ and its derivatives $y'(x)$. However, he did not address the issue of approximation. Due to the Runge phenomenon it is not a trivial task to extend his methods directly to approximation. As discussed in the following section, there is some literature relating to Hermite approximation of analytical functions. However, there is no method with respect to measurement data.

Within this work, we develop a framework for the approximation with polynomials of given values and derivatives (Hermite approximation) with different noises characteristics using covariance weighted discrete orthogonal polynomials. The method has been verified using synthetic data with known properties. It proves to be suitable for high degree polynomial approximation, since it is numerically more stable than other methods.

The main contribution of this paper span:

- 1) The derivation of a novel methodology for the generation of a discrete orthogonal polynomial basis function set which can be used to approximate values and derivatives which are perturbed by noise of different characteristics. It uses a three-term recurrence relation with full re-orthogonalization (to increase numerical stability) together with covariance weighting of the residual (for introducing a metric between the value and derivative domain).
- 2) The derivation of covariance propagation for the coefficients and the resulting approximation.
- 3) The introduction of various measures for the evaluation of the numerical stability.

The paper is structured as follows: A review of literature is found in Sec. II. In Sec. III all the derivations for the generation of a covariance weighted discrete orthogonal polynomial basis function set are introduced. The derivation of the covariance propagation is presented in Sec. III-D. To demonstrate the validity of the novel method, a Monte Carlo experiment for high degree polynomial approximation is performed on a synthetic data set in Sec. IV. It is shown that the residual between the data and the approximated values is reduced to noise with the same parameters as used in the generation of the synthetic noise. The numerical stability of the generated basis is validated in Sec. V demonstrating the novel method to be advantageous compared to using Vandermonde type basis functions, especially for higher degrees. Various quality measures are tested and presented for both, complete and incomplete basis function sets.

II. REVIEW OF LITERATURE

The use of value and derivative information for polynomial approximation (Hermite approximation) is not common in literature, whereas the use value and derivative information for interpolation is well-known in literature.

Based on the idea of a Newton type interpolation, Hermite introduced a similar methodology for the interpolation of values and derivatives. The original idea can be found in [3]. Based on this, several books, e.g. [4], introduce this idea to a broader audience.

A generalized version for arbitrary orders of derivatives at given points can be found in [5]. This idea is extended to bivariate functions in [6]. An extensive study on multivariate Hermite interpolation can be found in [7].

Since the calculation of those interpolating polynomials is straight forward, Hermite type interpolating polynomials gained popularity for the approximation of functions using a two point approach rather than Taylor expansion [8], [9]. They are well studied in terms of error bounds [10], [11] and compared to other standard methods [12]. They are used to solving ordinary differential equation [13], [14] and partial differential equation [15]–[17] by approximating the equations using Hermite type interpolation. It is pointed out, that using this method lead to higher order of approximation, which improves step-size for iterative methods compared to standard methods.

Another use of the Hermite interpolation is presented in [18]. In this work it is used to perform a moving-least-squares approximation based on local basis functions which use value and derivative information.

The idea of including derivative information within the approximation of data is mostly found in constrained polynomial approximation, e.g., [1], [19]–[21]. There it is assumed, that some information (e.g. reference points) is 100% certain. Approximation of data with uncertain value and derivatives is introduced in [22].

The idea of including a covariance weighting on the residuals to get a valid metric between the value and derivative domain is introduced in [1] and extended to higher order derivatives in [23].

Within this paper we use this idea of covariance weighting within the generation of a discrete orthogonal polynomial basis function set to improve the numerical stability, making it suitable for high degree polynomial fitting.

III. THEORETICAL FRAMEWORK

The herein presented framework addresses the approximation of a polynomial model given noisy values and collocated derivatives using discrete orthogonal polynomials, i.e., systems of polynomials that satisfy a discrete orthogonality constraint.

A. Modelling of Measured Values

The measured noisy values and derivatives are collected in the vectors $\hat{\mathbf{y}} = [\hat{y}_1 \ \hat{y}_2 \ \dots \ \hat{y}_n]^T$ and $\hat{\mathbf{y}}' = [\hat{y}'_1 \ \hat{y}'_2 \ \dots \ \hat{y}'_n]^T$. They are sampled at the positions $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$. For modelling the measurement, we assume that we measure the true value plus covariant noise. This covariant noise is generated from a vector of gaussian random variables with zero mean and unit variance (i.i.d. noise) together with the according covariance matrix. Mathematically this is described as

$$\hat{\mathbf{y}} = \mathbf{y} + \Lambda_{\mathbf{y}}^{\frac{1}{2}} \mathbf{s} \quad (1)$$

$$\hat{\mathbf{y}}' = \mathbf{y}' + \Lambda_{d\mathbf{y}}^{\frac{1}{2}} \mathbf{t}, \quad (2)$$

where \mathbf{y} and \mathbf{y}' are the true values, $\Lambda_{\mathbf{y}}$ and $\Lambda_{d\mathbf{y}}$ are the associated covariance matrices and \mathbf{s} and \mathbf{t} are vectors of i.i.d. noise.

B. Approximation of Values and Derivatives

The goal is now to approximate the given data in a least-squares sense, using a set of discrete basis functions collected in the columns of the matrix \mathbf{B} and their derivatives \mathbf{B}' . The true values \mathbf{y} and derivatives \mathbf{y}' are modelled as

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} \quad (3)$$

$$\mathbf{y}' = \mathbf{B}'\boldsymbol{\gamma}, \quad (4)$$

where \mathbf{B}' denotes the first derivative of the basis functions \mathbf{B} with respect to x and $\boldsymbol{\gamma}$ is the coefficient vector. To apply Gauss's least-squares theorem, the involved errors must be

i.i.d., so we solve for the gaussian random variables s and t using Eqn. (1)-(4), yielding

$$s = \Lambda_{\mathbf{y}}^{-\frac{1}{2}} (\hat{\mathbf{y}} - \mathbf{B}\gamma) \quad (5)$$

$$t = \Lambda_{\frac{d\mathbf{y}}{dy}}^{-\frac{1}{2}} (\hat{\mathbf{y}}' - \mathbf{B}'\gamma). \quad (6)$$

This results in the following functional to be minimized,

$$E(\gamma) = \|s\|_2^2 + \|t\|_2^2 = \quad (7)$$

$$= \|\Lambda_{\mathbf{y}}^{-\frac{1}{2}} (\hat{\mathbf{y}} - \mathbf{B}\gamma)\|_2^2 + \|\Lambda_{\frac{d\mathbf{y}}{dy}}^{-\frac{1}{2}} (\hat{\mathbf{y}}' - \mathbf{B}'\gamma)\|_2^2. \quad (8)$$

Clearly, this introduces covariance weighting on the residual and thus a metric between value and derivative domain is established as stated in [1] and [23]. Substituting $\mathbf{W}_{\hat{\mathbf{y}}} \triangleq \Lambda_{\mathbf{y}}^{-\frac{1}{2}}$ and $\mathbf{W}_{\frac{d\hat{\mathbf{y}}}{d\mathbf{y}}} \triangleq \Lambda_{\frac{d\mathbf{y}}{dy}}^{-\frac{1}{2}}$ and minimizing Eqn. (7) leads to the normal equations for weighted regression [24],

$$(\mathbf{B}^T \mathbf{W}_{\mathbf{y}} \mathbf{B} + \mathbf{B}'^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \mathbf{B}') \gamma = \mathbf{B}^T \mathbf{W}_{\mathbf{y}} \hat{\mathbf{y}} + \mathbf{B}'^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \hat{\mathbf{y}}'. \quad (9)$$

To solve this equation for γ one can use standard methods, e.g, inverting $(\mathbf{B}^T \mathbf{W}_{\mathbf{y}} \mathbf{B} + \mathbf{B}'^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \mathbf{B}')$ which is known to be computational costly.

To overcome this problem, we developed a method to find a polynomial basis function set \mathbf{P} and its derivative \mathbf{P}' , which fulfil

$$\mathbf{P}^T \mathbf{W}_{\mathbf{y}} \mathbf{P} + \mathbf{P}'^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \mathbf{P}' = \mathbf{I}, \quad (10)$$

i.e., a discrete orthogonality condition. The matrices $\mathbf{P} = [\mathbf{p}_0 \ \mathbf{p}_1 \ \dots \ \mathbf{p}_i \ \dots \ \mathbf{p}_d]$ and $\mathbf{P}' = [\mathbf{p}'_0 \ \mathbf{p}'_1 \ \dots \ \mathbf{p}'_i \ \dots \ \mathbf{p}'_d]$ are a collection of discrete polynomial basis functions and their derivatives. Each column \mathbf{p}_i represents a polynomial of degree i and \mathbf{p}'_i is the first derivative of that discrete polynomial. They are sorted on increasing degrees yielding a basis function set of degree d . Respectively a linear combination of those basis functions yield a polynomial of degree d at most.

The coefficients γ for the approximating polynomial can then be easily calculated from Eqn. (9) using,

$$\gamma = \mathbf{P}^T \mathbf{W}_{\mathbf{y}} \hat{\mathbf{y}} + \mathbf{P}'^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \hat{\mathbf{y}}', \quad (11)$$

which are inner products of the basis functions and the covariance weighted measurements. That is, the coefficient γ_i for a certain basis function \mathbf{p}_i of degree i can be directly calculated as

$$\gamma_i = \mathbf{p}_i^T \mathbf{W}_{\mathbf{y}} \hat{\mathbf{y}} + \mathbf{p}'_i^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \hat{\mathbf{y}}', \quad (12)$$

which is similar to the calculation of discrete Fourier series. Thus, the computational efficiency is improved compared to solving the fitting problem using standard algorithms including matrix inverses. To calculate the approximated values $\hat{\mathbf{y}}$ and $\hat{\mathbf{y}}'$, we use the estimated parameters γ within the model equation (3) and (4), yielding

$$\hat{\mathbf{y}} = \mathbf{P}\gamma \quad (13)$$

$$\hat{\mathbf{y}}' = \mathbf{P}'\gamma, \quad (14)$$

C. Synthesis of Weighted Discrete Orthogonal Basis

In this section, a novel method for the synthesis of a set of weighted discrete orthogonal polynomials fulfilling Eqn. (10) is presented.

Two important conditions can be directly derived from the identity in Eqn. (10). These are the normal condition

$$\mathbf{p}_{k+1}^T \mathbf{W}_{\mathbf{y}} \mathbf{p}_{k+1} + \mathbf{p}'_{k+1}^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \mathbf{p}'_{k+1} = 1 \quad (15)$$

and the orthogonality condition

$$\mathbf{p}_k^T \mathbf{W}_{\mathbf{y}} \mathbf{p}_{k+1} + \mathbf{p}'_k{}^T \mathbf{W}_{\frac{d\mathbf{y}}{dy}} \mathbf{p}'_{k+1} = 0, \quad (16)$$

which requires, that a basis function of degree $k+1$ is orthogonal to all basis functions of lower degree. The matrices $\mathbf{P}_k = [\mathbf{p}_0 \ \mathbf{p}_1 \ \dots \ \mathbf{p}_k]$ and $\mathbf{P}'_k = [\mathbf{p}'_0 \ \mathbf{p}'_1 \ \dots \ \mathbf{p}'_k]$ collect the basis vectors up to degree k .

For the generation of a set of suitable polynomial basis functions \mathbf{P} and their derivatives \mathbf{P}' up to a certain degree d , we use a recurrence relation with full re-orthogonalisation (as studied in [19], [25]) together with covariance weighting.

Starting from the classical *three-term recurrence relation* from functional analysis for orthogonal polynomials, e.g. [26], also known as Gram-Schmidt process [24], a polynomial $p_{k+1}(x)$ of degree $k+1$ can be generated from lower degree polynomials using

$$p_{k+1}(x) = \alpha x p_k(x) - \beta p_k(x) - \gamma p_{k-1}(x). \quad (17)$$

The derivative of this polynomial with respect to x is calculated as

$$p'_{k+1}(x) = \alpha x p'_k(x) + \alpha x' p_k(x) - \beta p'_k(x) - \gamma p'_{k-1}(x). \quad (18)$$

This is the definition for the continuous case.

The discrete formulations of Eqn. (17) and (18) for a given vector $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$ are¹

$$\mathbf{p}_{k+1} = \alpha \mathbf{x} \circ \mathbf{p}_k - \beta \mathbf{p}_k - \gamma \mathbf{p}_{k-1} \quad (19)$$

and

$$\mathbf{p}'_{k+1} = \alpha \mathbf{x} \circ \mathbf{p}'_k + \alpha \mathbf{x}' \circ \mathbf{p}_k - \beta \mathbf{p}'_k - \gamma \mathbf{p}'_{k-1}. \quad (20)$$

The vectors $\mathbf{p}_i = [p_i(x_1) \ p_i(x_2) \ \dots \ p_i(x_n)]^T$ and $\mathbf{p}'_i = [p'_i(x_1) \ p'_i(x_2) \ \dots \ p'_i(x_n)]^T$ are the polynomial of degree i and its derivative evaluated at the points \mathbf{x} . As shown in [27], the normal three-term recurrence relation is numerically unstable, so we use a complete re-orthogonalization as suggested in [25]. Using this improvement, Eqn. (19) and (20) read as

$$\mathbf{p}_{k+1} = \alpha \mathbf{x} \circ \mathbf{p}_k - \mathbf{P}_k \boldsymbol{\beta} \quad (21)$$

and

$$\mathbf{p}'_{k+1} = \alpha \mathbf{x} \circ \mathbf{p}'_k + \alpha \mathbf{x}' \circ \mathbf{p}_k - \mathbf{P}'_k \boldsymbol{\beta}. \quad (22)$$

Using the substitutions

$$\mathbf{u}_k = \mathbf{x} \circ \mathbf{p}_k \quad (23)$$

¹The operation with the symbol \circ denotes the Hadamard product, i.e., the element-wise product.

and

$$\mathbf{v}_k = \mathbf{x} \circ \mathbf{p}'_k + \mathbf{x}' \circ \mathbf{p}_k, \quad (24)$$

Eqn. (21) and (22) read as

$$\mathbf{p}_{k+1} = \alpha \mathbf{u}_k - \mathbf{P}_k \boldsymbol{\beta} \quad (25)$$

and

$$\mathbf{p}'_{k+1} = \alpha \mathbf{v}_k - \mathbf{P}'_k \boldsymbol{\beta}. \quad (26)$$

Using the orthogonality condition (16) together with (25) and (26) yields

$$(\mathbf{P}_k^T \mathbf{W}_y \mathbf{P}_k + \mathbf{P}_k^T \mathbf{W}_{dy} \mathbf{P}'_k) \boldsymbol{\beta} = \alpha (\mathbf{P}_k^T \mathbf{W}_y \mathbf{u}_k + \mathbf{P}_k^T \mathbf{W}_{dy} \mathbf{v}'_k). \quad (27)$$

Since, per definition, the previous generated basis functions have to fulfil the orthogonality and normal condition

$$\mathbf{P}_k^T \mathbf{W}_y \mathbf{P}_k + \mathbf{P}_k^T \mathbf{W}_{dy} \mathbf{P}'_k = \mathbf{I}, \quad (28)$$

$\boldsymbol{\beta}$ can be expressed as

$$\boldsymbol{\beta} = \alpha (\mathbf{P}_k^T \mathbf{W}_y \mathbf{u}_k + \mathbf{P}'_k^T \mathbf{W}_{dy} \mathbf{v}_k). \quad (29)$$

Using this, Eqn. (25) and (26) can be rewritten as

$$\mathbf{p}_{k+1} = \alpha (\mathbf{u}_k - \mathbf{P}_k \mathbf{P}_k^T \mathbf{W}_y \mathbf{u}_k - \mathbf{P}_k \mathbf{P}'_k^T \mathbf{W}_{dy} \mathbf{v}_k) = \alpha \mathbf{c}_{k+1} \quad (30)$$

and

$$\mathbf{p}'_{k+1} = \alpha (\mathbf{v}_k - \mathbf{P}'_k \mathbf{P}_k^T \mathbf{W}_y \mathbf{u}_k - \mathbf{P}'_k \mathbf{P}'_k^T \mathbf{W}_{dy} \mathbf{v}_k) = \alpha \mathbf{c}'_{k+1}. \quad (31)$$

\mathbf{c}_{k+1} and \mathbf{c}'_{k+1} represent a basis function and (its derivative) which is orthogonal to all previous basis functions but not yet normed. To fulfil the norm condition in Eqn. (15), the scaling factor α is calculated as

$$\alpha = \sqrt{\frac{1}{\mathbf{c}_{k+1}^T \mathbf{W}_y \mathbf{c}_{k+1} + \mathbf{c}'_{k+1}^T \mathbf{W}_{dy} \mathbf{c}'_{k+1}}}. \quad (32)$$

This scaling factor and the coefficient vector $\boldsymbol{\beta}$ have to be calculated for each newly generated basis function.

To start the recurrence, the first basis functions \mathbf{p}_0 , \mathbf{p}_1 and their derivatives have to be defined in an initial step, to meet the above mentioned conditions. The first basis function of degree $d = 0$ is defined as

$$\mathbf{p}_0 = \frac{\mathbf{e}}{\sqrt{\mathbf{e}^T \mathbf{W}_y \mathbf{e}}}, \quad (33)$$

which is a normalized constant vector. \mathbf{e} denotes a vector of ones. The first derivative of this basis function is the zero vector

$$\mathbf{p}'_0 = \mathbf{0}. \quad (34)$$

For the second basis function \mathbf{p}_1 we first generate the vector

$$\mathbf{u}_1 = \mathbf{x} \circ \mathbf{p}_0, \quad (35)$$

which we project onto the orthogonal complement of \mathbf{p}_0 to meet the orthogonality condition, yielding

$$\hat{\mathbf{p}}_1 = (\mathbf{I} - \mathbf{p}_0 \mathbf{p}_0^T \mathbf{W}_y) \mathbf{u}_1. \quad (36)$$

This is a scaled version of \mathbf{p}_1 . To get the slope for the derivatives of the basis functions right, the vector \mathbf{x}' is calculated as

$$\mathbf{x}' = \sqrt{\mathbf{e}^T \mathbf{W}_y \mathbf{e}} \begin{pmatrix} \hat{p}_{1,n} - \hat{p}_{1,1} \\ x_n - x_1 \end{pmatrix} \mathbf{e}. \quad (37)$$

Using this, we calculate a scaled version of \mathbf{p}'_1 based on Eqn. (24) and (26), yielding

$$\hat{\mathbf{p}}'_1 = \mathbf{x}' \circ \mathbf{p}_0. \quad (38)$$

To generate the basis functions fulfilling the norm conditions, we calculate the scaling factor

$$\alpha_1 = \sqrt{\frac{1}{\hat{\mathbf{p}}_1^T \mathbf{W}_y \hat{\mathbf{p}}_1 + \hat{\mathbf{p}}_1^T \mathbf{W}_{dy} \hat{\mathbf{p}}'_1}}. \quad (39)$$

From this we calculate the second pair of basis functions

$$\mathbf{p}_1 = \alpha_1 \hat{\mathbf{p}}_1 \quad (40)$$

and

$$\mathbf{p}'_1 = \alpha_1 \hat{\mathbf{p}}'_1. \quad (41)$$

This is the prerequisite to start the synthesis of higher order basis functions. The final set of basis functions $\mathbf{P} \triangleq \mathbf{P}_d$ and $\mathbf{P}' \triangleq \mathbf{P}'_d$ of a certain degree d can now be used in Eqn. (11), (13) and (14) to calculate the coefficients and to finally approximate perturbed values and its derivatives with a polynomial of degree d . A set of basis functions is shown in Fig. 2.

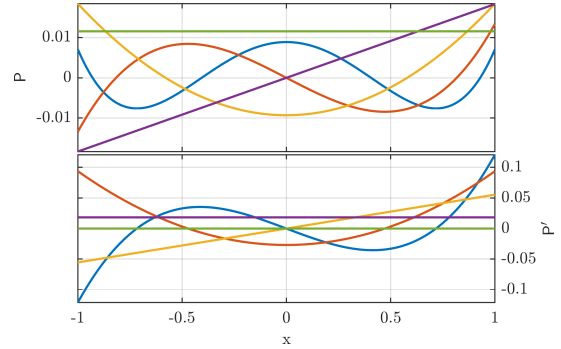


Fig. 2. A set of discrete orthogonal basis functions \mathbf{P} of degree $d = 4$ and its derivative \mathbf{P}' generated for $n = 300$ equally spaced points with $\sigma_{y_i} = \sigma_y = 0.2$ and $\sigma_{d y_i} = \sigma_{d y} = 0.8$

The generation of a basis function set using the presented method is a generalized Gram-Schmidt process. As it can be seen, the generation of the basis depends only on the relative locations of the x values² and not on the values itself. If the abscissa values x do not change, the basis function set can be calculated a priori and the solution to the weighted fitting problem reduces to a simple matrix-vector multiplication (see

²To improve numerical stability, \mathbf{x} is transformed to be centered at the origin and scaled to unit norm.

Eqn. (11)). This is advantageous when implemented in smart sensors or low power controllers. A further advantage is, that the weighting matrices W_y and W_{dy} can be rank-deficient, e.g., points can be weighted with 0 if they should not be considered. This can be helpful to suppress outliers.

D. Covariance Propagation

Since we are dealing with noisy data, covariance propagation is a prerequisite for making assumptions about the quality of the approximated values. Based on Eqn. (11) the covariance Λ_γ for the coefficients γ is calculated as

$$\Lambda_\gamma = P_c^T W_c \Lambda_c W_c^T P_c \quad (42)$$

with the block matrices

$$\Lambda_c = \begin{bmatrix} \Lambda_y & 0 \\ 0 & \Lambda_{dy} \end{bmatrix}, W_c = \begin{bmatrix} W_y & 0 \\ 0 & W_{dy} \end{bmatrix}, P_c = \begin{bmatrix} P \\ P' \end{bmatrix}. \quad (43)$$

Similarly, the covariance matrices for the approximated values and derivatives can be propagated as

$$\Lambda_{\hat{y}} = P P_c^T W_c \Lambda_c W_c^T P_c P^T \quad (44)$$

and

$$\Lambda_{d\hat{y}} = P' P_c^T W_c \Lambda_c W_c^T P_c P'^T. \quad (45)$$

Since $W_c \triangleq \Lambda_c^{-1}$ the middle-term results in $W_c \Lambda_c W_c^T = W_c$. Together with the identity from Eqn. (10), the above equations simplify to

$$\Lambda_\gamma = I, \quad (46)$$

$$\Lambda_{\hat{y}} = P P^T \quad (47)$$

and

$$\Lambda_{d\hat{y}} = P' P'^T. \quad (48)$$

These covariance matrices for the approximated coefficients and values, can be used to calculate confidence intervals or predictions intervals. Note: As it can be seen in Eqn. (46), the presented method decorrelates the noise to i.i.d. noise, to accord with Gauss's theorem.

IV. NUMERICAL EXAMPLE

To test the validity of the herein presented method to approximate a polynomial given perturbed values and derivatives, a synthetic dataset is generated. The underlying function is defined as

$$f(x) = \cos(5x) \quad (49)$$

with its analytical first derivative

$$\frac{df(x)}{dx} = -5 \sin(5x). \quad (50)$$

The function and its derivative are evaluated in the range $[-2\pi, 2\pi]$ at $n = 500$ equally spaced nodes, yielding the vectors of values and derivatives y and y' . Gaussian noise with different gains $\sigma_{y_i} = \sigma_y = 0.1$ and $\sigma_{dy_i} = \sigma_{dy} = 2$ is added to those vectors yielding the noisy measurement vectors \hat{y} and \hat{y}' . A polynomial of degree $d = 35$ is used for approximating the noisy data set. Although the test function seems simple, i.e., a cosine function, it is a difficult task to approximate

perturbed data corresponding to multiple cycles of a cosine by a polynomial. Using geometric polynomials (i.e. Vandermonde basis) would be numerically unstable due to the high degree required.

To test the developed method, a Monte Carlo simulation is performed with $n_{iter} = 1000$ iterations. As a measure, the standard deviation of the residuals $\text{std}\{r_y\} = \text{std}\{y - \hat{y}\}$ and $\text{std}\{r_{dy}\} = \text{std}\{y' - \hat{y}'\}$ are calculated in each run. Since the presented method uses covariance weighting, the standard deviation of the result should be the same as $\sigma_y \approx \text{std}\{y - \hat{y}\}$ and $\sigma_{dy} \approx \text{std}\{y' - \hat{y}'\}$. The mean value of the standard deviations over all runs is shown in Fig. 3. As it can be seen, although the noise gains are very different, the presented method which uses covariance weighted approximation delivers the correct results for both, values and derivatives demonstrating the method to be valid.

As expected, the method using the Vandermonde basis as presented in [23], is not stable for such a high degree. As one can inspect, the approximation does not follow the signal.

V. NUMERICAL QUALITY OF BASIS

To verify the numerical quality of herein presented method, a meaningful measure has to be found. As Wilkinson [28] pointed out, a posteriori estimation of error bounds is preferred to a priori error predictions in such cases. The identity in Eqn. (10) can be written in terms of the block matrices as

$$P_c^T W_c P_c = I. \quad (51)$$

Rewriting in terms of a unitary matrix U yields

$$U^T U = I, \quad (52)$$

with

$$U = W_c^{\frac{1}{2}} P_c. \quad (53)$$

The residual matrix

$$R = I - P_c^T W_c P_c \approx 0 \quad (54)$$

should ideally be equivalent zero, i.e., all entries should be zero. However, numerical limitations during the synthesis may lead to errors, which can be correlated. The residual matrix R is shown in Fig. 4. Although there are errors, it is important to notice, that the errors made are within the range of $\pm 5 \times 10^{-14}$. Additionally, the errors show no significant structure within the residual. The uniformity of the residual is important, since the approximation is not exhibited to structural errors. In Fig. 5 a zoomed-in section with the most significant errors is shown, but again the errors are only in the range mentioned above. The revealed pattern indicates a small correlation between basis functions with even and odd degrees. To summarize the numerical quality of the generated basis, the following error measures are tested in order to find the appropriate measure:

- 1) **Maximum norm.** The maximum norm is the largest single element within the residual matrix, i.e., $\epsilon_{\max} = \|R\|_{\max} = \max\{|r_{ij}|\}$. Since this norm depends only on one specific entry, this may lead to wrong conclusions.

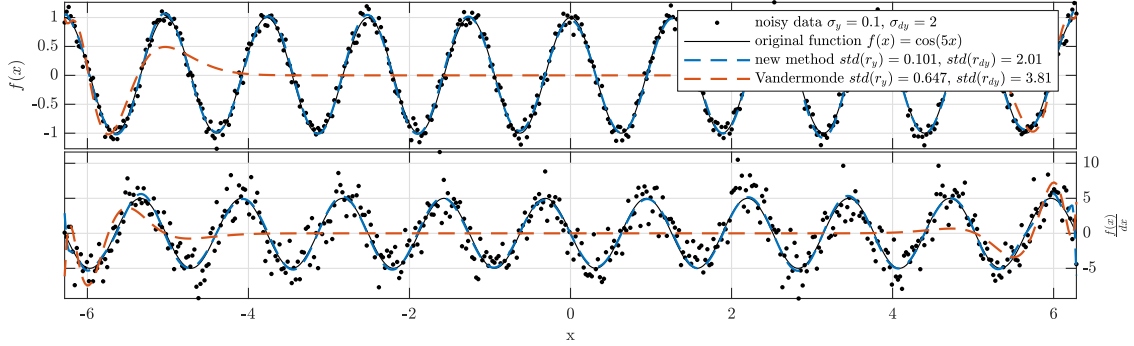


Fig. 3. Approximation of a polynomial of degree $d = 35$ to synthetic data generated from $f(x) = \cos(5x)$ with $\sigma_{y_i} = \sigma_y = 0.1$ and $\sigma_{dy_i} = \sigma_{dy} = 2$

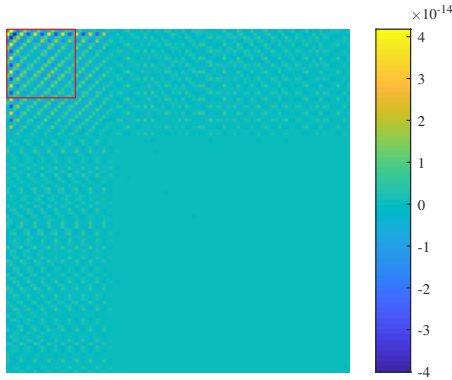


Fig. 4. The structure of a residual matrix $R = 1 - P_c^T W_c P_c$ for $d = 100$ and $n = 50$ with $\sigma_{y_i} = \sigma_y = 0.2$ and $\sigma_{dy_i} = \sigma_{dy} = 0.8$. The red box indicates the zoomed in region shown in Fig. 5.

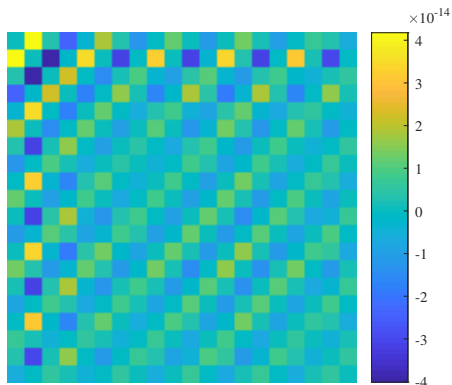


Fig. 5. Zoomed in region of the residual matrix as shown in Fig. 4. The checkerboard structure reveals the correlation, be they very small, between the even and odd degree of the synthesized basis functions.

- 2) **Frobenius norm.** The Frobenius norm is the square root of the sum of the squares of all entries in the residual matrix, i.e., $\epsilon_F = \|R\|_F = \sqrt{\sum_i \sum_j r_{ij}^2}$. This norm is a measure for the total error.
- 3) **Determinant.** The determinant of a matrix is a theoretical quality measure and equals 1 for an ideal unitary matrix. The variation from this, i.e., $\epsilon_{\det} = 1 - \det\{U\}$ is a measure for the quality of the tested basis function set.
- 4) **Condition number.** The condition number of a matrix is connected to the error propagation. For a unitary matrix the condition number is 1. Thus, the measure tested is $\epsilon_{\text{cond}} = 1 - \text{cond}\{U\}$.
- 5) **Rank.** Since the presented method generates a weighted orthogonal basis function set, the rank of U should be full rank, i.e., no linear dependencies. Due to round off errors this can be used to find linear dependencies, i.e., $\epsilon_{\text{rank}} = n - \text{rank}\{U\}$.

Based on those measures the approximate number of significant digits η is calculated as

$$\eta_m = -\log_{10}(\epsilon_m). \quad (55)$$

These measures are calculated for a complete basis function set where the number of basis functions equals the number of data points (values and derivatives). Therefore, the degree of the resulting polynomial is $d = 2n - 1$. In Fig. 6 the different measures are presented for varying degrees. As it can be seen, the Frobenius norm and the condition number are the most meaningful measures, since they show the highest dependency on the degree of the resulting polynomial. The rank measure is not visible, since the proposed algorithm generates full-rank basis function sets, so there is no error visible. Since the Frobenius norm is a measure for the total error, this measure is chosen to compare to other algorithms in the following.

The new method is compared to the one presented in [23]. It uses a Vandermonde basis function set to solve the same problem. As [19] pointed out, the Vandermonde basis for normal polynomial regression gets degenerate at high degrees.

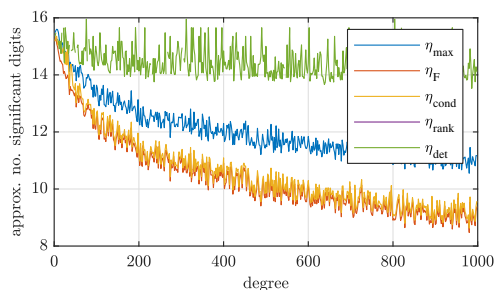


Fig. 6. Comparison of different error measures for a complete basis function set with $d = 2n - 1$ and $\sigma_{y_i} = \sigma_y = 0.2$ and $\sigma_{d_{y_i}} = \sigma_{d_y} = 0.8$

This behaviour can also be inspected within this paper. As it can be seen in Fig. 7, the new method generates a more stable result also for high degrees.

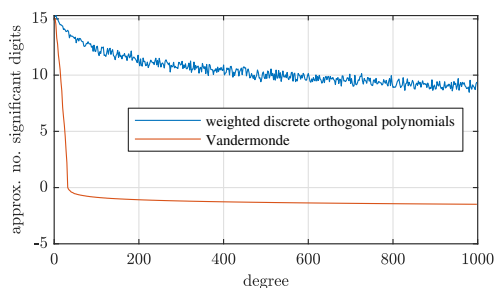


Fig. 7. Numerical quality of Vandermonde basis compared to the new method using a complete basis with $\sigma_{y_i} = \sigma_y = 0.2$ and $\sigma_{d_{y_i}} = \sigma_{d_y} = 0.8$

Since this new method can also be used for approximation (overdetermined system of equations), the quality of the basis is determined for an incomplete basis function set for $n = 1000$. The result for various degrees of polynomial is visualized in Fig. 8, showing that the new method performs also better for an incomplete basis.

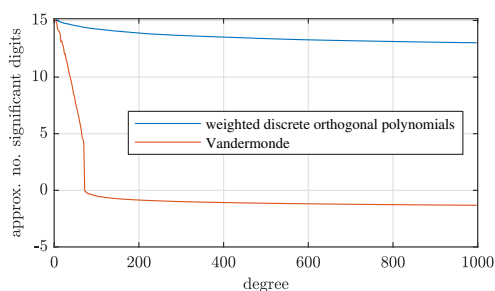


Fig. 8. Numerical quality of Vandermonde basis compared to the new method using an incomplete basis with $\sigma_{y_i} = \sigma_y = 0.2$, $\sigma_{d_{y_i}} = \sigma_{d_y} = 0.8$ and $n = 1000$

VI. CONCLUSION

The herein presented method introduces a novel polynomial fitting framework for the approximation of value and derivative data. Including both sources of information within the fitting procedure improves the quality of the fitted polynomial improving both, reconstruction of values and derivatives. The method uses a recurrence relation with full re-orthogonalization together with covariance weighting for introducing a metric between value and derivative domain, yielding a set of discrete orthogonal polynomials. As it is shown, the generated basis function set is numerically more stable compared to other methods, especially for high degree polynomials. Using this basis, the fitting problem is reduced to inner products, which is beneficial in terms of computational efficiency. Due to the covariance weighting, the noise associated with the channels is decorrelated to i.i.d. noise, to accord with Gauss's theorem. Thus, there is no bias based on different noise parameters. The validity of the method is tested and presented on a numerical example, where a polynomial of degree $d = 35$ is fitted to a periodic function. In future research, this method will be adapted for fitting data given noisy constraints, based on discrete orthogonal polynomials.

ACKNOWLEDGMENT

This work was partially funded under the auspices of the EIT - KIC Raw materials program within the project "Maintained Mine and Machine" (MaMMA) with the grant agreement number: [EIT/RAW MATERIALS/SGA2018/1]

REFERENCES

- [1] P. O'Leary and M. Harker, "Inverse boundary value problems with uncertain boundary values and their relevance to inclinometer measurements," in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. IEEE, may 2014, pp. 165–169. [Online]. Available: <http://ieeexplore.ieee.org/document/6860725/>
- [2] K. Weierstrass, "Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen," pp. 633–639 and 789–805, 1885. [Online]. Available: <https://www.math.auckland.ac.nz/hat/fpapers/we4.pdf>
- [3] C. Hermite, "Sur la formule d'interpolation de Lagrange. (Extrait d'une lettre de M. Ch. Hermite à M. Borchardt)." *Journal für die reine und angewandte Mathematik*, vol. 84, pp. 70–79, 1877. [Online]. Available: <http://eudml.org/doc/148345>
- [4] A. Ralston and P. Rabinowitz, *A first course in numerical analysis*. Dover Publications, 2001.
- [5] A. Spitzbart, "A Generalization of Hermite's Interpolation Formula," *The American Mathematical Monthly*, vol. 67, no. 1, p. 42, jan 1960. [Online]. Available: <https://www.jstor.org/stable/2308924?origin=crossref>
- [6] M. M. Chawla and N. Jayarajan, "A generalization of Hermite's interpolation formula in two variables," *Journal of the Australian Mathematical Society*, vol. 18, no. 04, p. 402, dec 1974. [Online]. Available: https://www.jstor.org/stable/pdf/2308924.pdf?%0Ahttp://www.jstor.org/stable/2308http://www.journals.cambridge.org/abstract/_%S1446788700029074
- [7] T. Sauer and Y. Xu, "On multivariate Hermite interpolation," *Advances in Computational Mathematics*, vol. 4, no. 1, pp. 207–259, dec 1995. [Online]. Available: <http://link.springer.com/10.1007/BF03177515>
- [8] V. V. Shustov, "Approximation of functions by two-point Hermite interpolating polynomials," *Computational Mathematics and Mathematical Physics*, vol. 55, no. 7, pp. 1077–1093, 2015. [Online]. Available: <http://link.springer.com/10.1134/S0965542515040156>

- [9] —, “Approximation of functions by asymmetric two-point hermite polynomials and its optimization,” *Computational Mathematics and Mathematical Physics*, vol. 55, no. 12, pp. 1960–1974, dec 2015. [Online]. Available: <http://link.springer.com/10.1134/S0965542515120155>
- [10] A. Varma and K. Katsifarakis, “Optimal error bounds for Hermite interpolation,” *Journal of Approximation Theory*, vol. 51, no. 4, pp. 350–359, dec 1987. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/0021904587900438>
- [11] R. P. Agarwal and P. J. Wong, “Optimal error bounds for the derivatives of two point hermite interpolation,” *Computers & Mathematics with Applications*, vol. 21, no. 8, pp. 21–35, 1991. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/0898122191900489>
- [12] M. Sharifi, F. Soleymani, M. Khan, M. A. Gondal, and I. Hussain, “On two-point Hermite interpolation: An application of Newton’s Theorem,” *World Applied Sciences Journal*, vol. 13, no. 12, pp. 2451–2454, 2011.
- [13] J. Mennig, T. Auerbach, and W. Hälgl, “Two point hermite approximations for the solution of linear initial value and boundary value problems,” *Computer Methods in Applied Mechanics and Engineering*, vol. 39, no. 2, pp. 199–224, 1983.
- [14] R. E. Grundy, “Hermite interpolation visits ordinary two-point boundary value problems,” *The ANZIAM Journal*, vol. 48, no. 04, p. 533, apr 2007. [Online]. Available: http://www.journals.cambridge.org/abstract/_S1446181100003205
- [15] R. Grundy, “Polynomial representations for initial-boundary-value problems involving the inviscid Proudman-Johnson equation,” *The Quarterly Journal of Mechanics and Applied Mathematics*, vol. 59, no. 4, pp. 631–650, oct 2006. [Online]. Available: <https://academic.oup.com/qjmam/article-lookup/doi/10.1093/qjmam/hbl019>
- [16] R. E. Grundy, “The application of Hermite interpolation to the analysis of non-linear diffusive initial-boundary value problems,” *IMA Journal of Applied Mathematics (Institute of Mathematics and Its Applications)*, vol. 70, no. 6, pp. 814–838, 2005.
- [17] R. Grundy, “The analysis of initial-boundary-value problems using Hermite interpolation,” *Journal of Computational and Applied Mathematics*, vol. 154, no. 1, pp. 63–95, may 2003. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0377042702008117>
- [18] Z. Komargodski and D. Levin, “Hermite type moving-least-squares approximations,” *Computers & Mathematics with Applications*, vol. 51, no. 8, pp. 1223–1232, apr 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0898122106000757>
- [19] M. Harker and P. O’Leary, “A Matrix Framework for the Solution of ODEs: Initial-, Boundary-, and Inner-Value Problems,” apr 2013. [Online]. Available: <http://arxiv.org/abs/1304.3312>
- [20] C. Gugg, M. Harker, and P. O’Leary, “Structural deformation measurement via efficient tensor polynomial calibrated electro-active glass targets,” in *IS&T/SPIE Electronic Imaging, Image Processing: Machine Vision Applications VI*, P. R. Bingham and E. Y. Lam, Eds., vol. 8661, mar 2013, p. 86610F. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2003679>
- [21] P. O’Leary, R. Ritt, and M. Harker, “Constrained Polynomial Approximation for Inverse Problems in Engineering,” in *Proceedings of the 1st International Conference on Numerical Modelling in Engineering*, M. A. Wahab, Ed. Springer Singapore, 2019, vol. NME2018, no. Lecture Notes in Mechanical Engineering, pp. 225–244. [Online]. Available: http://link.springer.com/10.1007/978-981-13-2273-0_{_}19
- [22] O. Roderick, M. Anitescu, and P. Fischer, “Polynomial Regression Approaches Using Derivative Information for Uncertainty Quantification,” *Nuclear Science and Engineering*, vol. 164, no. 2, pp. 122–139, feb 2010. [Online]. Available: <https://www.tandfonline.com/doi/full/10.13182/NSE08-79>
- [23] R. Ritt, P. O’Leary, C. J. Rothschedl, A. Almasri, and M. Harker, “Hierarchical Decomposition and Approximation of Sensor Data,” in *Proceedings of the 1st International Conference on Numerical Modelling in Engineering*, M. A. Wahab, Ed. Springer Singapore, 2019, vol. NME2018, no. Lecture Notes in Mechanical Engineering, pp. 351–370. [Online]. Available: http://link.springer.com/10.1007/978-981-13-2273-0_{_}27
- [24] G. H. Golub and C. F. Van Loan, *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*, 3rd ed. The Johns Hopkins University Press, 1996.
- [25] P. O’Leary and M. Harker, “An Algebraic Framework for Discrete Basis Functions in Computer Vision,” in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE, dec 2008, pp. 150–157. [Online]. Available: <http://ieeexplore.ieee.org/document/4756064/>
- [26] W. Gautschi, *Orthogonal polynomials : computation and approximation*. Oxford University Press, 2004.
- [27] P. O’Leary, “Discrete polynomial moments for real-time geometric surface inspection,” *Journal of Electronic Imaging*, vol. 18, no. 1, p. 013015, jan 2009. [Online]. Available: <http://electronicimaging.spiedigitallibrary.org/article.aspx?doi=10.1117/1.2987725>
- [28] J. H. Wilkinson, “Modern Error Analysis,” *SIAM Review*, vol. 13, no. 4, pp. 548–568, oct 1971. [Online]. Available: <http://www.jstor.org/stable/2029191http://epubs.siam.org/doi/10.1137/1013095>

Part III

Symbolic Time Series Analysis

9 | Synopsis

The basic idea behind symbolic time series analysis as proposed in the following papers is that symbols/words are assigned to states of the monitored machine. Additionally, linear differential operators (LDO) are used to model the dynamics of the system in observation, e.g. calculate regularized derivatives or solve the associated inverse problems. To not only work on a global scale with LDO, the use of local linear differential operators is investigated in Section 9.1. Applying a LDO to a single signal can be seen as supporting physical based knowledge discovery by taking dynamics into account while calculating the cause of the observations. To express this knowledge in a compact form, a symbol/word is assigned to portions of the signal with the same cause (e.g. the word *up* is assigned to each observation with a positive first derivative, indicating that the signal has an upwards trend, whereas the word *down* is assigned to values with a negative first derivative). By combining run-length of the same word to a single word predicated with its length (number of subsequent occurrences), a discrete signal is transformed into a symbolic time series, including the first layer of compression. The physics of the system is respected by applying LDO prior to symbolization and the idea of assigning *meaningful* words to the signal is a major contribution of this thesis, since standard methods perform symbolization based on equal probability rather than on a-priori knowledge.

Combining the symbolic time series of multiple channels in form of polysyllabic words, a more complex state of the machine can be described and automatically separates/segments multi-dimensional time series. By analysing the distribution of such polysyllabic words, i.e. building frequency dictionaries, operation modes and rare events can be identified, which can be used to perform a higher level of segmentation. A detailed description of those ideas can be found in [P5, P9] (see Chapter 10 and Chapter 11).

Since large physical systems are often operated from humans, these machines can be seen as hybrid systems. In this context, a hybrid machine performs operations which are a sequence of physical processes connected involving non-analytical human interaction. The physical processes can be modelled using LDO. To model the non-analytical portion, the metaphor of language is used in combination with the above presented symbolic time series approach. The idea of *compounding* (i.e. give common repetitions of word combinations a new word) is used iteratively in a hierarchical

manner to automatically reveal operation modes, which are a combination of common word sequences. This idea is described in [P4] (see Chapter 12).

9.1 Local Linear Differential Operator

The data analysed in this thesis is emanating from large physical systems driven by human operators and is operated 24/7. This results in large multi-dimensional time series (TS) to be dealt with. Each channel builds a continuous stream of discrete data $\mathbf{y} = \mathbf{y}(\mathbf{t})$ which are collections of values at specific time points \mathbf{t} , the so called data points. These TS work well with linear differential operators (LDO) as described in detail in the papers below. As the TS grows (number of data points increases) the matrices involved in the computations grow accordingly and slow down the computation. The global view on the entire TS is not necessary in a lot of applications, especially when talking about hybrid system with non-analytical human interaction. A local view (window) on the data may be sufficient for numerous tasks and allows a more efficient calculation. When talking about real-time data mining, one has to deal with TS with a constant sampling rate (which leads to equally spaced nodes). This allows setting up the local LDO-matrix \mathbf{L} only for a window length l_s and to compute the solution for each window with this local matrix. Since \mathbf{L} remains constant if the sampling rate is constant, it can be calculated a-priori.

In general, the calculation of an operated signal \mathbf{w}_{LDO} of a snippet $\mathbf{w} = \mathbf{y}(s : s + l_s)$ of the TS \mathbf{y} is done via the matrix multiplication

$$\mathbf{w}_{\text{LDO}} = \mathbf{L}\mathbf{w}. \quad (9.1)$$

This is, row j of \mathbf{L} multiplied with \mathbf{w} calculates the operated signal $w_{\text{LDO},j}$ at the position of data point w_j , which is a linear combination of all values w_k in this window. If l_s is odd, the middle row of \mathbf{L} is used for calculating the operated signal at the centre node of the window, see Fig. 9.1(a). Note: The number of points taken into account to the left and to the right of the actual viewpoint need to be equal.

If the window is incrementally moved along the TS, it is only necessary to use the centre row of \mathbf{L} for computing the entire operated signal \mathbf{y}_{LDO} , only the $\frac{l_s-1}{2}$ endpoints at the beginning and the end have to be corrected using the top and lower portion of the local LDO-matrix, see Fig. 9.1(b). The entire computation for the TS can be either done using an expanded version of the local LDO-matrix, which results in a band-diagonal matrix \mathbf{L}_{full} , see [71], or by using a convolutional approach, see Fig. 9.1(c) and Fig. 9.1(d).

Using the expanded version \mathbf{L}_{full} , the calculation of the entire operated signal is given by

$$\mathbf{y}_{\text{LDO}} = \mathbf{L}_{\text{full}}\mathbf{y}. \quad (9.2)$$

The expanded matrix \mathbf{L}_{full} is built by creating a squared matrix with zeros of size $[n \times n]$, where n is the length of the TS \mathbf{y} . In the upper left corner the first $\frac{l_s-1}{2}$

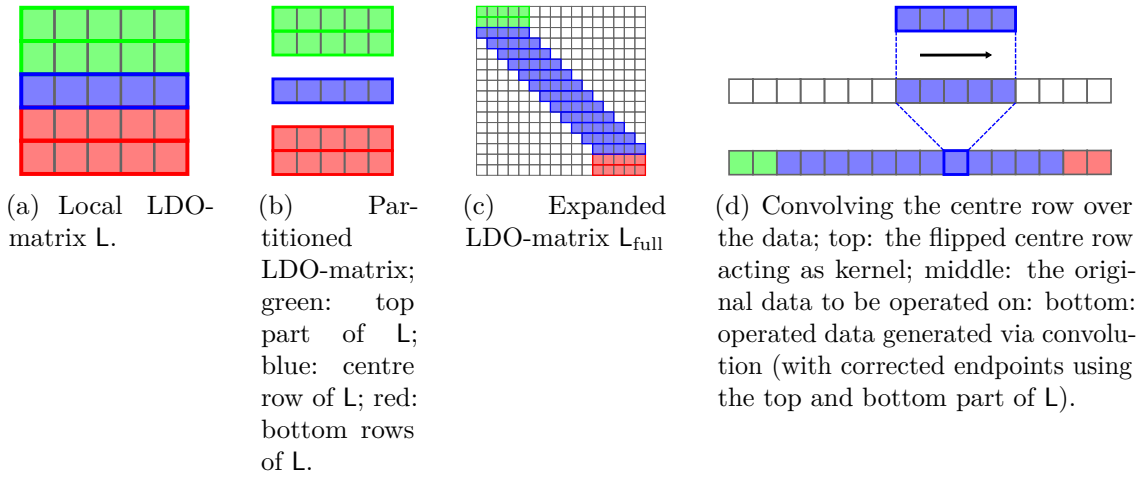


Fig. 9.1 Graphical visualization of the LDO-matrices and the convolutional approach.

rows of the local linear operator matrix L , and in the lower right corner the last $\frac{l_s-1}{2}$ rows of L are placed. The core part of L_{full} is filled diagonally with the centre row of L , which results in the structure shown in Fig. 9.1(b) and Fig. 9.1(c). This computation requires setting up a matrix with size $[n \times n]$ which is memory-costly. Although the band-diagonal structure of L_{full} allows the use of sparse matrices, the computation itself can be expensive with large time series. Furthermore, the full TS must be available at the point of analysis.

To overcome this problem, a convolutional approach is proposed for local calculations. Here the signal \mathbf{y} is convolved with the flipped middle row of the local linear differential operator (the so called kernel). Only the endpoints on both sides of the TS have to be corrected using the top and the bottom parts of L . This approach is shown in Alg. 9.1 and for a visual representation see Fig. 9.1(d). Since the convolution part acts as sliding window, this methodology is also applicable for streaming data and high performance analysis, especially in real-time data analytics.

Alg. 9.1: Convolutional approach for collocational local linear differential computations.

```

Input :  $L$  // local linear differential operator matrix (LDO-matrix)
           $\mathbf{y}$  // signal to be operated on
Output :  $\mathbf{y}_{\text{LDO}}$  // processed signal:

[1]  $i_m := \frac{l_s-1}{2}$ ; // find index of centre line
[2]  $T := L(1:i_m, :)$ ; // Extract top portion of  $L$ 
[3]  $B := L(\text{end} - i_m + 1 : \text{end}, :)$ ; // Extract bottom portion of  $L$ 
[4]  $\mathbf{m}_L := L(l_s, :)$ ; // Extract middle row of  $L$ 
[5]  $\mathbf{y}_{\text{LDO}} := \text{CONV}(\mathbf{y}, \text{FLIPLR}(\mathbf{m}_L), \text{'same'})$ ; // perform convolution with middle row
[6]  $\mathbf{y}_{\text{LDO}}(1:i_m) = T\mathbf{y}(1:i_m)$ ; // Correct endpoints at the beginning
[7]  $\mathbf{y}_{\text{LDO}}(\text{end} - i_m + 1 : \text{end}) = B\mathbf{y}(\text{end} - i_m + 1 : \text{end})$ ; // Correct endpoints at the
    end

```

10 | Mining Sensor Data in Larger Physical Systems

Originally appeared as:

P. O’Leary, M. Harker, R. Ritt, M. Habacher, K. Landl, and M. Brandner, “Mining Sensor Data in Larger Physical Systems,” *IFAC-PapersOnLine*, vol. 49, no. 20, pp. 37–42, 2016, ISSN: 24058963. DOI: 10.1016/j.ifacol.2016.10.093. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S2405896316316561>

Bib_TE_X:

```
@article{OLeary2016MiningSensorData,
  author = {O'Leary, Paul and Harker, Matthew and Ritt,
            Roland and Habacher, Michael and Landl, Katharina
            and Brandner, Michael},
  doi = {10.1016/j.ifacol.2016.10.093},
  issn = {24058963},
  journal = {IFAC-PapersOnLine},
  keywords = {Data mining,data mining,entropy,lexical analysis,
             linear differential operators},
  number = {20},
  pages = {37--42},
  title = {{Mining Sensor Data in Larger Physical Systems}},
  url = {http://linkinghub.elsevier.com/retrieve/pii/
         S2405896316316561},
  volume = {49},
  year = {2016}
}
```

Mining Sensor Data in Larger Physical Systems

Paul O’Leary, Matthew Harker, Roland Ritt,
Michael Habacher, Katharina Landl, Michael Brandner

* *Institute for Automation, University of Leoben, Leoben, Austria.*
(e-mail: roland.ritt@unileoben.ac.at).

Abstract: This paper presents a framework for the collection, management and mining of sensor data in large cyber-physical systems. Particular emphasis has been placed on mathematical methods, data structures and implementations which enable the real-time solution of inverse problems associated with the system in question. That is, given a system model, to obtain an estimate for the phenomenological cause of the sensor observation. This enables the use of causality, rather than mere correlation, when computing measures of significance during machine learning and knowledge discovery in very large data sets. The model is an abstract representation of a real physical system establishing the relationships between cause and effects. The pertinent behaviour of the model is captured in the form of equations, e.g., differential equations. The inverse solution of these model-equations, within certain constraints, permit us to establish the semantic reference between the sensor observation and its cause. Without this semantic reference there can be no physically based knowledge discovery.

Embrechts pyramid of knowledge is addressed and shown that it will not suffice for future developments. The issue of information content is addressed more formally than in most data mining literature. Additionally the Epistemology for the emergent-perceptive portion of speech is presented and a prototype implementation with experimental results in data mining are presented. A lexical symbolic analysis of sensor data is implemented.

Keywords: Data mining, entropy, linear differential operators, lexical analysis.

1. INTRODUCTION

This paper addresses issues involved in mining sensor data in cyber physical systems (CPS), see e.g. O’Leary et al. (2015b). The advent of cyber physical systems has brought about a significant change in the architecture of sensor and measurement systems. The change is in terms of the numbers of sensors involved, the complexity of the system being addressed, the spatial distribution of the sensors within the systems and the global nature of the system itself. This is facilitated by the use of network components.

There are many different definitions for what constitutes a cyber physical system (CPS) Baheti and Gill (2011); Geisberger and Broy (2012); IOSB (2013); Lee (2008); NIST (2012); Park et al. (2012); Spath et al. (2013a,b); Tabuada (2006). The most succinct and pertinent to this paper is the definition given by the IEEE Baheti and Gill (2011) and ACM¹:

*A CPS is a system with a coupling of the cyber aspects of computing and communications with the physical aspects of dynamics and engineering that **must abide by the laws of physics**. This includes sensor networks, real-time and hybrid systems.*

The solutions computed from the sensor data **must** obey the equations modelling the physics of the system being

observed - this is fundamentally an inverse problem and requires the modeling of the system dynamics. Unfortunately, the issue of inverse problems is not addressed in literature on *mining sensor data*, see for example Esling and Agon (2012); Fuchs et al. (2010); Keogh and Kasetty (2003); Last et al. (2004). Actually, none of the standard books available, e.g. Aggarwal (2013) on mining sensor data, take the special nature of the sensor data into account. Present data mining techniques rely on correlation (in some manner) as being a reliable measure for significance. However, the inverse solution of the model-equations is required to establish the semantic reference between the sensor observation and its cause. Without this semantic reference to causality there can be no physically based knowledge discovery. The main contributions of this paper are:

- (1) A structured approach based on *data, information, hypothesis, evidence, truth and fact* is proposed and demonstrated.
- (2) A mechanism for the solution of ordinary differential equations (ODE) is introduced. The numerical methods for the linear differential operators (LDO) have been developed in such a manner that they can be integrated into standard data mining environments such as Hadoop Shvachko et al. (2010); White (2009). This permits the implementation of embedded simulation (forward problems), the solution of inverse problems and regularizing computations.

¹ ACM/IEEE International Conference on Cyber-Physical Systems (ICCCPS) (icccps.acm.org)

- (3) The extension of the symbolic aggregate approximation (SAX) Keogh and Kasetty (2003) to non-linear intervals while maintaining the lower bound property. The approximation with series of symbols permits the use of regular expressions to perform symbolic searches in the real sensor data.

2. A STRUCTURED APPROACH TO DATA ANALYTICS/MINING

Embrechts et al. (2005) proposed the pyramid of data mining as shown in Figure 1, which was an extension of Ackoff's work Ackoff (1989). Embrechts offers no definitions for the terms *information*, *knowledge*, *understanding* and *wisdom* in his work, while Ackoff offers intuitive but rather nebulous inaccurate definitions. The pyramid and the terms used have positive connotations²; however, they do not provide a scientific basis for mining sensor data.

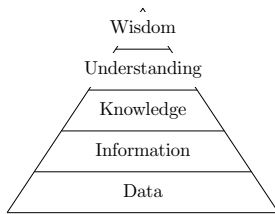


Fig. 1. The data mining wisdom pyramid as proposed by Mark Embrechts.

Shannon provided a formal mathematical framework for computing *information* content in a data stream. It is important to note that this measure of information is neither a measure for *significance* nor of *meaningfulness*. It is however a powerful tool in identifying temporal locations where the information content in a data stream changes.

The addition of *meta-data* is required to give a data stream *meaning*. For example, the stream of data from a digital thermometer is meaningless, i.e. its just a stream of numbers, without the meta-data for the measurement device. Correct meta-data should ensure that a *datum* is a statement about a *fact* and can so be considered as a low level of *knowledge*. In the context of this paper we shall define *understanding* as the relationship between facts, *constructs* (models) and *temporal context* which is indicative of a specific operation (*behaviour*)³. Understanding permits the prediction of the behaviour of a system under similar circumstances and *wisdom* can be considered as advantageous behaviour in a certain circumstance given the specific state of the available resources.

To implement data analytics and mining, it is necessary to develop tools which support the *exploration* of data which supports the formulation of *hypotheses*. Then we

² *Wisdom* just as the word *creativity* have positive connotations but resist any formal definition, see von Hentig (1998) for a discussion of this issue. Without formal definition they do not form the basis for objective data analytics.

³ We are using the word *behaviour* with some concern at this point, since it reflects the subjective behaviour of the operator. Large plant and machinery can generate data sequences which have a very strong operator dependency.



Fig. 2. Photograph of the port and ship loader used as an example to characterize the data flows involved in continuous monitoring.

need a means of representing the hypotheses to enable the extraction of *evidence* from the data with the aim of supporting or refuting the hypothesis. This is the process of *knowledge discovery*. It requires a representation for knowledge which enables its later utilization — this mechanism must also include the possibility to encode a-priori knowledge, e.g., coming from the design process. *Statistical evidence* plays a very significant role in mining sensor data, since the complexity of the systems may preclude the complete *observability* of the process, i.e., there is no possibility to have a conclusive *induction* of the creating process from the observation represented by the sensor data in a finite time.

3. DATA COLLECTION AND MANAGEMENT

Two modes of data collection are supported: *full-table* mode whereby all channels are provided at each time stamp. This mode is suitable for systems where there is a high degree of activity and no long periods during which the system is dormant; *on-change* in this mode values for the channels are only transmitted when they change, each message consists of a *time-stamp*, a *tag* and a *value*. In on-change mode it is necessary to introduce a calibration mechanism which initiates the transmission of all *time-stamp: tag: value triplets* at predefined times, otherwise there is the danger of undefined data states. At the server the sensor data is always stored in full-table format.

We have characterized the data flows from a ship loader currently being monitored, see Fig. 2. Currently, there are $n = 150$ sensors being monitored with the sampling period $t_s = 1s$, in addition there are four vibration channels acquired with $f_s = 2.5kHz$. The sensor data is transmitted as on-change and the vibration data as full-table, but segmented into observation periods of $t_p = 25s$. The summary of the data flows for this plant is given in Table 1. The volume of data transmitted varies from day to day due to the use of on-change format. The maximum file size observed over a time period of $m = 147$ days was $s_d = 62Mb$ and the average over this period was $35.2Mb$. we have chosen the HFD5 file format to store the sensor data, since in this manner the data can be efficiently stored as binary in double precision. In this work we are using MATLAB[®] for exploratory work and Python for established methods. The consequence of this investigation is, that approximately $s_y \approx 2.6Gb$ of storage is required, with level nine compression, to archive the complete sensor

Table 1. Summary of data from the ship loader: for $n = 150$ sensors, with a sampling interval of $t_s = 1s$ over a period of 24 hours. The data is transmitted *on change* and the full table is reconstructed on the server.

File format	Structure	Size	read time
CSV (ascii) maximum	on change	62Mb	
CSV (ascii) average	on change	35.2Mb	
HDF5 (level 0 - binary)	full table	72Mb	0.3 s
HDF5 (level 9 - binary)	full table	7Mb	0.3...0.6 s



Fig. 3. Photograph of the reclaimer from which the data for Table 2 were acquired.

Table 2. Summary of data from the reclaimer: for $n = 12$ sensors, with a sampling interval of $t_s = 1s$ over a period of 24 hours. The data is transmitted directly in full table format.

Read times in seconds	Local	Network
HDF5 High level	0.0028 s	0.1683 s
HDF5 Low level	0,0024 s	0.0032 s
mysql		2.1394 s
HBase via Java		1.2570 s
HBase via MATLAB		4.5835 s

data for one year in HDF5 format; this is by no means prohibitive.

A second set of tests were performed on data from a reclaimer (see Fig. 3) with the aim of determining data retrieval times: in this case $n = 12$ sensors are being monitored with a sampling period of $t_s = 1s$ and blocked into observation periods of 24 hours. The results for retrieving the data from local HDF5, SQL and HBase are shown in Table 2; the retrieval times are significantly faster for HDF5. Finally, Fig. 4 shows the comparison of local vs remote retrieval as a function of number of rows loaded. This would suggest that the final system should support a three level cache: in memory for the most actual data, since these are the most commonly queried data set; local HDF5 files for recent data and remote data base supported data archiving for the complete data sets.

4. LINEAR DIFFERENTIAL OPERATORS.

The dynamics of physical and engineering systems are, in general, modelled using differential equations. Whereby, the sensor data is a measure of the response of the system to some phenomena affecting the system. Consequently, correlation with other signals may not be reveal *causality*; since, to relate the response of the system to causing phenomena we must solve the inverse problem. Furthermore,

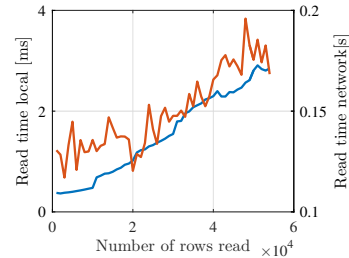


Fig. 4. HDF5 data retrieval times: blue is for local i.e. left axis and red is for the remote retrieval.

we must first determine if the system demonstrates an ergodic⁴ Walters (1982) behavior. the question of ergodic behaviour is particularly important when dealing with fleet management. The measurement model considered here is an ODE, (see O’Leary et al. (2015a) for more details):

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_1(x)y^{(1)} + a_0(x)y = g(x), \quad (1)$$

where y is a function of x , $y^{(n)}$ is the notation for the n^{th} derivative of y with respect to x and $g(x)$ is the exciting function, in this case the sensor data. The nature of the constraints determines if the system is considered to be an inner-, initial- or boundary-value problem; whereby any mixture is also admissible in the proposed method.

We use the general notion of a linear differential operator Lanzcos (1997) D ; such that, $D^{(n)}y(x) = y^{(n)}$: using this notation in Equation 1 yields,

$$a_n(x)D^{(n)}y + a_{n-1}(x)D^{(n-1)}y + \dots + a_1(x)Dy + a_0(x)y = g(x). \quad (2)$$

Factoring y to the right yields and defining the linear differential operator

$$L \triangleq a_n(x)D^{(n)} + a_{n-1}(x)D^{(n-1)} + \dots + a_1(x)D + a_0(x) \quad (3)$$

Consequently, Equation 1 can be written as,

$$Ly = g(x). \quad (4)$$

The matrix L can in general be partitioned into two portions: a band diagonal portion with a Toeplitz structure and a correction factor for the first and last l_s entries, where l_s is the support length chosen. The storage of the sensor data as contiguous full table has been chosen to enable the computation of the product of the Toeplitz matrix and the data vector as a convolution, for which there are particularly fast implementations.

5. SINGLE CHANNEL INFORMATION

Shannon (1948) provided a mathematical framework for the computation of information in a sequence of symbols being transmitted between systems, see Gallager (1968). It is important to note that information content computed as entropy is a simple manipulation of symbols

⁴ Ergodic theory is a branch of mathematics that studies dynamical systems with an invariant measure. Its initial development was motivated by problems of statistical physics. Broadly speaking, the term is used to denote a system that has the same behavior averaged over time as averaged over the space of all the system’s states.

and their probabilities, the result is neither a measure of significance nor meaningfulness. However, when mining very large data sets, we may be interested in inexact measures which identify sub-portions of the code which justify further investigation. Time varying histograms are a very efficient and valuable tool in this area of obtaining statistical evidence; they also permit the very efficient computation of local entropy in a signal. An example is shown in Fig 5. The concept is extended by computing the regularized first derivative of entropy as a means of identifying events, which result in a perdurant change in information content.

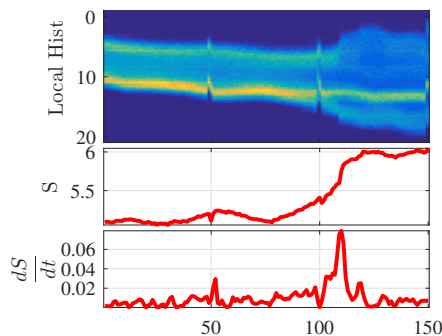


Fig. 5. Top: time varying histogram of the chuck torque during the milling of a metal part. Mid: Entropy as a function of time $S(t)$. Bottom: the regularized first differential of entropy $\dot{S}(t)$. The time point of the peak in the bottom plot corresponds to a fracturing of the milling tool's cutting edge.

6. EPISTEMOLOGY AND THE EMERGENCE OF SPEECH

The Merriam Webster dictionary defines epistemology as: *the study or a theory of the nature and grounds of knowledge especially with reference to its limits and validity.* This is the question we need to address in mining sensor data. This leads us naturally to consider phenomenology. To cite Husserl in brief: *All knowledge is derived from the experience of phenomena.* Phenomenology as a branch of physics is defined as:

A body of knowledge that relates empirical observations of phenomena to each other, in a way that is consistent with fundamental theory, but is not directly derived from theory.

This definition is much closer to what we need to establish. However, it does not say anything about how the empirical observations of phenomena are performed and how the models are established so that they are consistent with fundamental theory. The implementation of the LDO as part of the mining process is the first step in ensuring that the laws of physics governing the system are observed.

In this portion of the paper we shall address the Asian model on the emergence of human speech and its relevance to data mining. It is considered that when we observe repetitions in our sensory experience that we wish to assign

names (symbols) to this experience, as abbreviations. There are three relevant types of repetition in data mining:

- (1) Value repetitions. This would correspond to modes in a histogram in a technical application.
- (2) Repetition of temporal sequences, and
- (3) Repetition in combinations within the senses.

The first corresponds to a modified symbolic aggregation approximation Lin et al. (2003) (SAX). The advantage of the SAX approach is that there is a formal proof for the lower bounding theorem, i.e., there is a positive semi-definite distance measure, this implies that comparisons of sequences are valid. The symbols are compressed by replacing a multitude of the same symbol as a single symbol predicated by its length. This also solves the problem of dynamic time warping (DTW), since the sequence remains unchanged, under time dilation, only the predicates need to be compared. This process converts a real valued sensor data stream to a sequence of symbols and predicates enabling symbolic searching.

We present the concept of a single channel lexical analyser (SCLA) consisting of: a linear differential operator LDO; a symbolic SAX approximation and a lexical compression, see Fig 7 for an example of its application to the slewing data of a reclaimer. We call it lexical analysis since we can use regular expressions to search for symbols for specific types of events in the real sensor data.

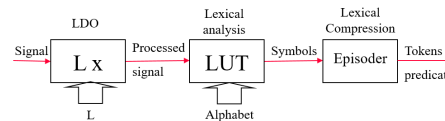


Fig. 6. Block diagram for the SCLA.

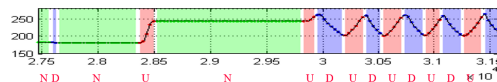


Fig. 7. Example of the application of an SCLA to the slewing data of the reclaimer shown in Fig 3. It has segmented the data into symbol segments shown as underlying colours.

6.1 Parallel Channels

The sequence of symbols and predicates delivered by the SCLA modules are grouped, then parallel sequences are automatically identified, see Fig. 8. As is well known from the A-Priori algorithm the number of combinations does not explode, since the machines are not being operated in a random fashion. The search reveals which sequences occur with which frequencies, this corresponds to the concept of *frequency dictionaries* in language. They are ideal for identifying which operations need to be considered during optimization.

7. EXAMPLE APPLICATIONS OF THE SYSTEM

The concepts presented here have been applied to a number of different case studies:

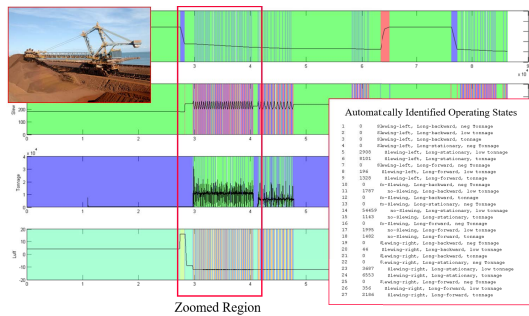


Fig. 8. Example of knowledge discovery in a multi sensor system as the generation of a frequency dictionary form combinations of the SAX approximation of each channel.

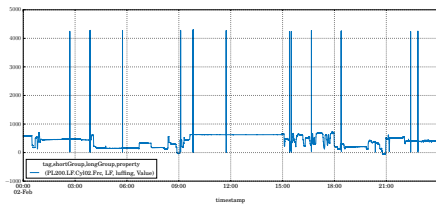


Fig. 9. Example of corrupted data.

7.1 Commissioning support

The system has been used to support engineers during commissioning of a ship loader; Fig. 9 shows a 24 hour sequence of data with intermittent corruptions of the data, each with a duration of approximately 3s. An SCLA was implemented which automatically identifies such short interruptions. This type of error is very difficult for a field engineer to identify.

7.2 Fleet Management

The comparative monitoring of two ship loaders is shown in Figure 10. The time varying histograms are generated after the SCLA's has automatically identified comparable operating conditions for each machine. The automatic identification of operations is an important issue when comparing the performance of machines, otherwise meaningless statistical results are obtained.

7.3 Logistics and Preventative Maintenance

In this application the total tonnage produced as a function of the reclaimers angular position is determined and the statistics generated for one complete years production, see Figure 11. Clearly, the life time of the slew-bearing is a function of the cumulative loading. The very slow rotational speeds of slew-bearings in reclaimers make them susceptible to local damage. The high loading of one specific quadrant is due to the instruction on the positioning of material is the stock yard; consequently, the life time of the bearing can be extended significantly simply by changing the operational instructions.

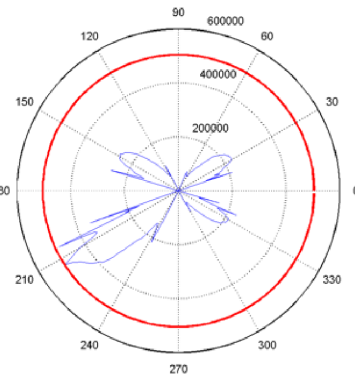


Fig. 11. Total tonnage produced as a function of the angular position on the slew-bearing.

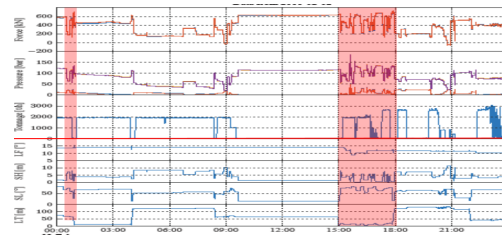


Fig. 12. Channels above the red line are sensor data and below actor data, the data is for a 24 hour period acquired with a sampling interval of 1s. The parallel analysis of multiple SCLA's is used here to identify regions with a very high interaction between the actor and sensor signals.

7.4 System Identification

This example demonstrates the use of parallel channel analysis to identify time intervals where there is a very high interaction between the actor and sensors data, see Figure 12. Such intervals are particularly important when performing system identification.

8. CONCLUSIONS

This paper has presented some new approaches to data mining in cyber physical systems and some of the background considerations behind these developments. These are emerging technologies and the descriptions are not as tight as is the case with established technologies. Nevertheless, very impressive results are achieved with the first prototypes. The use of HDF5 as a file format permits the loading of very large data sets in modest times. The presented methods are fundamentally parallel in nature and are well suited for implementation with the MapReduce paradigm.

REFERENCES

Ackoff, R.L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
 Aggarwal, C.C. (ed.) (2013). *Managing and Mining Sensor Data*. Springer.

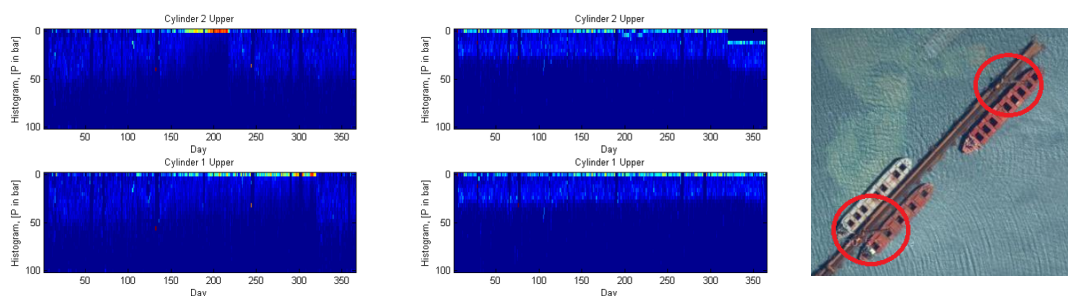


Fig. 10. Example of fleet management: The time varying histograms show the complete data for one year for two identical ship loader operating on one and the same quay. It permits their direct comparison.

- Baheti, R. and Gill, H. (2011). Cyber-physical systems. *The Impact of Control Technology*, 161–166.
- Embrechts, M., Szymanski, B., and Sternickel, K. (2005). *Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing*. John Wiley and Sons, New York.
- Esling, P. and Agon, C. (2012). Time-series data mining. *ACM Comput. Surv.*, 45(1), 12:1–12:34. doi:10.1145/2379776.2379788. URL <http://doi.acm.org/10.1145/2379776.2379788>.
- Fuchs, E., Gruber, T., Pree, H., and Sick, B. (2010). Temporal data mining using shape space representations of time series. *Neurocomputing*, 74(13), 379 – 393. doi: <http://dx.doi.org/10.1016/j.neucom.2010.03.022>. URL <http://www.sciencedirect.com/science/article/pii/S0925231210002237>. Artificial Brains.
- Gallager, R.G. (1968). *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA.
- Geisberger, E. and Broy, M. (2012). *agendaCPS: Integrierte Forschungsagenda Cyber-Physical Systems*, volume 1. Springer.
- IOSB, F. (2013). Industry 4.0 information technology is the key element in the factory of the future. Press Information.
- Keogh, E. and Kasetty, S. (2003). On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Min. Knowl. Discov.*, 7(4), 349–371. doi:10.1023/A:1024988512476. URL <http://dx.doi.org/10.1023/A:1024988512476>.
- Lanczos, C. (1997). *Linear Differential Operators*. Dover books on mathematics. Dover Publications. URL <http://books.google.at/books?id=WLEVEY-3gIAC>.
- Last, M., Kandel, A., and Bunke, H. (2004). *Data Mining in Time Series Databases*. Series in machine perception and artificial intelligence. World Scientific. URL <http://books.google.at/books?id=f38wqKjyBm4C>.
- Lee, E.A. (2008). Cyber physical systems: Design challenges. In *Object Oriented Real-Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on*, 363–369. IEEE.
- Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 2–11. ACM, New York, NY, USA.
- NIST (2012). Cyber-physical systems: Situation analysis of current trends, technologies, and challenges. Technical report, National Institute of Standards and Technology (NIST). URL www.nist.gov.
- O’Leary, P., Harker, M., and Gugg, C. (2015a). An inverse problem approach to approximating sensor data in cyber physical systems. In *Instrumentation and Measurement Technology Conference (I2MTC), 2015 IEEE International*, 1717–1722.
- O’Leary, P., Harker, M., and Gugg, C. (2015b). A position paper on: Sensor-data analytics in cyber physical systems, from husslerl to data mining. In *SensorNets 2015, Le Cresout, France*.
- Park, K.J., Zheng, R., and Liu, X. (2012). Cyber-physical systems: Milestones and research challenges. *Computer Communications*, 36(1), 1–7.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.
- Shvachko, K., Kuang, H., Radia, S., and Chansler, R. (2010). The hadoop distributed file system. In *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, MSST ’10, 1–10. IEEE Computer Society, Washington, DC, USA. doi:10.1109/MSST.2010.5496972. URL <http://dx.doi.org/10.1109/MSST.2010.5496972>.
- Spath, D., Gerlach, S., Hämmerle, M., Schlund, S., and Strölin, T. (2013a). Cyber-physical system for self-organised and flexible labour utilisation. *Personnel*, 50, 22.
- Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., and Schlund, S. (2013b). *Produktionsarbeit der Zukunft-Industrie 4.0*. Fraunhofer IAO Stuttgart.
- Tabuada, P. (2006). Cyber-physical systems: Position paper. In *NSF Workshop on Cyber-Physical Systems*.
- von Hentig, H. (1998). *Kreativität: hohe Erwartungen an einen schwachen Begriff*. Hanser.
- Walters, P. (1982). *An Introduction to Ergodic Theory*. Springer.
- White, T. (2009). *Hadoop: The Definitive Guide*. O’Reilly Media, Inc., 1st edition.

11 | Advanced Symbolic Time Series Analysis in Cyber Physical Systems

Originally appeared as¹:

R. Ritt, P. O’Leary, C. J. Rothschedl, and M. Harker, “Advanced Symbolic Time Series Analysis In Cyber Physical Systems,” in *Proceedings - International work-conference on Time Series, ITISE 2017*, O. Valenzuela, F. Rojas, H. Pomares, and I. Rojas, Eds., vol. 1, Granada: University of Granada, Feb. 2017, pp. 155–160, ISBN: 9788417293017. arXiv: 1802.00617. [Online]. Available: <http://arxiv.org/abs/1802.00617>http://itise.ugr.es/proceedings/ITISE%7B%5C_%7D2017.zip

BibT_EX:

```
@inproceedings{Ritt2017,
  author      = {Ritt, Roland and O'Leary, Paul and Rothschedl,
                Christopher Josef and Harker, Matthew},
  booktitle   = {Proceedings - International work-conference on
                Time Series, ITISE 2017},
  editor      = {Valenzuela, Olga and Rojas, Fernando and Pomares,
                H{\e}ctor and Rojas, Ignacio},
  eprint      = {1802.00617},
  isbn       = {9788417293017},
  keywords    = {cyber physical system, linear differential
                operator, single channel lexical analyser, symbolic
                time series analysis, time series},
  pages       = {155--160},
  publisher   = {University of Granada},
  title       = {{Advanced Symbolic Time Series Analysis In Cyber
                Physical Systems}},
  url         = {http://arxiv.org/abs/1802.00617},
  volume      = {1},
  year        = {2017}
}
```

¹A preprint of this paper is available on arXiv: <http://arxiv.org/abs/1802.00617>

Advanced Symbolic Time Series Analysis in Cyber Physical Systems

Roland Ritt, Paul O’Leary, Christopher Josef Rothschedl, and Matthew Harker

Institute for Automation, University of Leoben, Leoben, Austria

`roland.ritt@unileoben.ac.at`

Keywords: symbolic time series analysis, single channel lexical analyser, time series, cyber physical system, linear differential operator

This paper presents advanced symbolic time series analysis (ASTSA) for large data sets emanating from cyber physical systems (CPS). The definition of CPS most pertinent to this paper is: *A CPS is a system with a coupling of the cyber aspects of computing and communications with the physical aspects of dynamics and engineering that must abide by the laws of physics. This includes sensor networks, real-time and hybrid systems* [1]. To ensure that the computation results conform to the laws of physics a linear differential operator (LDO) is embedded in the processing channel for each sensor. In this manner the dynamics of the system can be incorporated prior to performing symbolic analysis. A non-linear quantization is used for the intervals corresponding to the symbols. The intervals are based on observed modes of the system, which can be determined either during an exploratory phase or on-line during operation of the system. A complete processing channel (see Fig. 2) is called a single channel lexical analyser; one is made available for each sensor on the machine being observed.

The implementation of LDO in the system is particularly important since it enables the establishment of a causal link between the observations of the dynamic system and their cause. Without causality there can be no semantics and without semantics no knowledge acquisition based on the physical background of the system being observed. Correlation alone is not a guarantee for causality¹

This work was originally motivated from the observation of large bulk material handling systems, see Fig. 1 for three examples of such systems. Typically, there are $n = 150 \dots 250$ sensors per machine, and data is collected in a multi rate manner; whereby general sensors are sampled with $f_s = 1 Hz$ and vibration data being sampled in the kilo-hertz range.

1 Local Linear Differential Operators (LDO)

Although processing the entire ‘large’ time series is a common practice in exploratory data analysis, reliable local computations (implemented as streaming

¹ Consider an exothermic system with a high activation energy. We must include the exothermic model if we are to establish causality, correlation alone will lead to erroneous interpretation.



Fig. 1. Examples of machines to which the analysis is applied. Image courtesy: Sandvik

algorithms) are preferred in on-line data processing. Since in this work we deal with time series emanating from cyber physical systems new techniques for local computations including the physics of the system (described by differential equations) have to be developed.

An ordinary differential equation (ODE) of the form

$$a_d(t) y^{(d)}(t) + a_{d-1}(t) y^{(d-1)}(t) + \dots + a_0(t) y^{(0)}(t) = g(t) \quad (1)$$

can be described using a linear differential operator (LDO) D [2] such that $D^{(i)}y(t) = y^{(i)}(t)$ where y is a function of t , $y^{(i)}$ is the n -th derivative with respect to t and $g(t)$ is the exciting function, in our case the noisy sensor data. This yields to the notation [3]

$$a_d(t) D^{(d)}y^{(d)}(t) + a_{d-1}(t) D^{(d-1)}y^{(d-1)}(t) + \dots + a_0(t) D^{(0)}y^{(0)}(t) = g(t). \quad (2)$$

Factoring $y(t)$ leads to the compact formulation of the model

$$Ly(t) = g(t), \quad (3)$$

with

$$L \triangleq a_d(t) D^{(d)} + a_{d-1}(t) D^{(d-1)} + \dots + a_0(t) D^{(0)}. \quad (4)$$

In the discrete case (3) can be formulated as matrix equation. Solving this equation for y is an inverse problem which can be solved numerically in a discrete sense by

$$\mathbf{y} = \mathbf{L}^+ \mathbf{g} + \mathbf{N}_L \boldsymbol{\alpha}, \quad (5)$$

where \mathbf{y} is the solution to the inverse problem, \mathbf{L}^+ is the pseudo-inverse of \mathbf{L} , \mathbf{N}_L is an orthonormal basis function set of the null space of \mathbf{L} , $\boldsymbol{\alpha}$ is a coefficient vector for the null space (computed by initial- and/or the boundary-values) and \mathbf{g} is the noisy time series data vector. Algebraic implementations for the solution of such problems can be found in [4–7].

The LDO, and their inverses, can be implemented as local operators and efficiently computed using a convolutional approach. This is basically a streaming-algorithm and thus suitable for big-data processing.

Furthermore, the covariance of the solution (5) is simply propagated as

$$\Lambda_y = \mathbf{L}^+ \Lambda_g (\mathbf{L}^+)^T. \quad (6)$$

Using Λ_y as an estimate for the covariance in conjunction with the student- t and/or F-distribution permits the estimation of a confidence interval over the complete solution and allows the computation of a prediction interval for future values.

That is, the approach presented here to implementing linear differential operators not only permits the solution of embedded system dynamics but also yields a confidence interval for the predicted values of the dynamics.

2 Symbolic Time Series Analysis

The availability of the sensor signals, their regularized derivative and/or the application of a LDO permits the implementation of an advanced symbolic time series analysis (ASTSA) which includes the modelling of the system dynamics. As a result the time series (TS) can be discretized and compressed using unique symbols for different intervals (the so called alphabet). This step is named lexical analysis. A number of methods for the selection of the symbol intervals based on, e.g. , equal probability, variance or entropy can be found in literature [8–12]. Here, in a new approach, we define the intervals to correspond to the modes of the dynamic system in operation, i.e. each symbol corresponds to a mode or portion of a mode which should be identified. Commonly controllers are designed to operate optimally in a number of specific but distinct modes of the dynamic system.

In a next step, connected sequences with the same symbol can be compressed to a single symbol predated with its length. The combination of applying a LDO, lexical analysis of the derived signal and compression is called single channel lexical analyser (SCLA), see Fig. 2. Combining the output of multiple SCLA

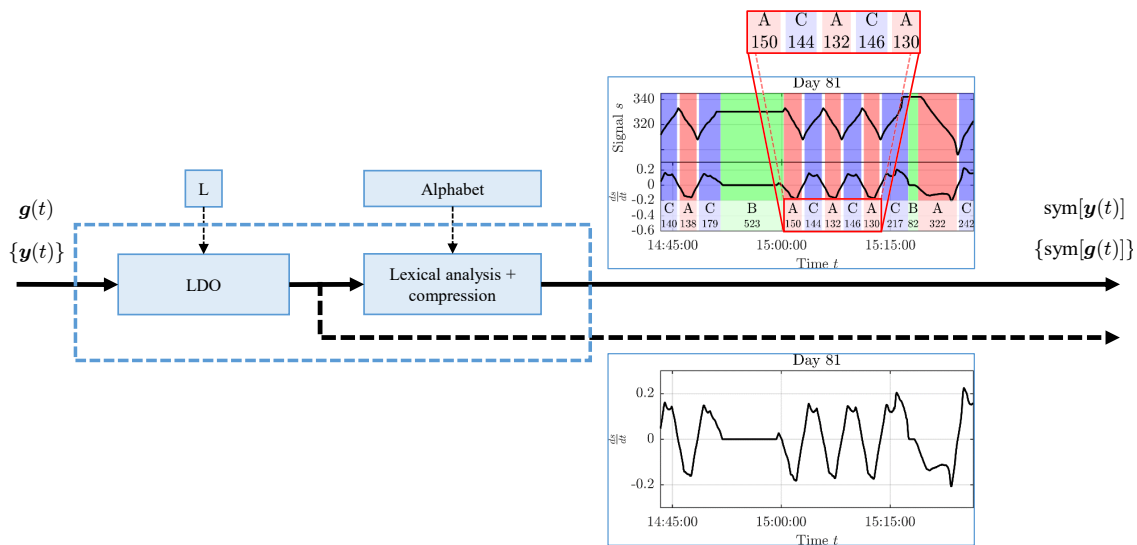


Fig. 2. A single channel lexical analyser (SCLA)

is called multi channel lexical analyser (MCLA). Two examples of symbolic time series analysis using MCLA are demonstrated in Fig. 3 and Fig. 4). For signal 1 and signal 2 the alphabet consists of the three symbols [u, s, d] assigned to the direction of the signal (up, stationary, down). The figures show two operation modes from the same machine. It can be clearly seen, that the operation modes of the machine have a different symbolic representation (visualized as different shaded colours in the plots) and allow a fast intuitive inspection and characterization of the signal. The signal range from the first dashed-blue line to the

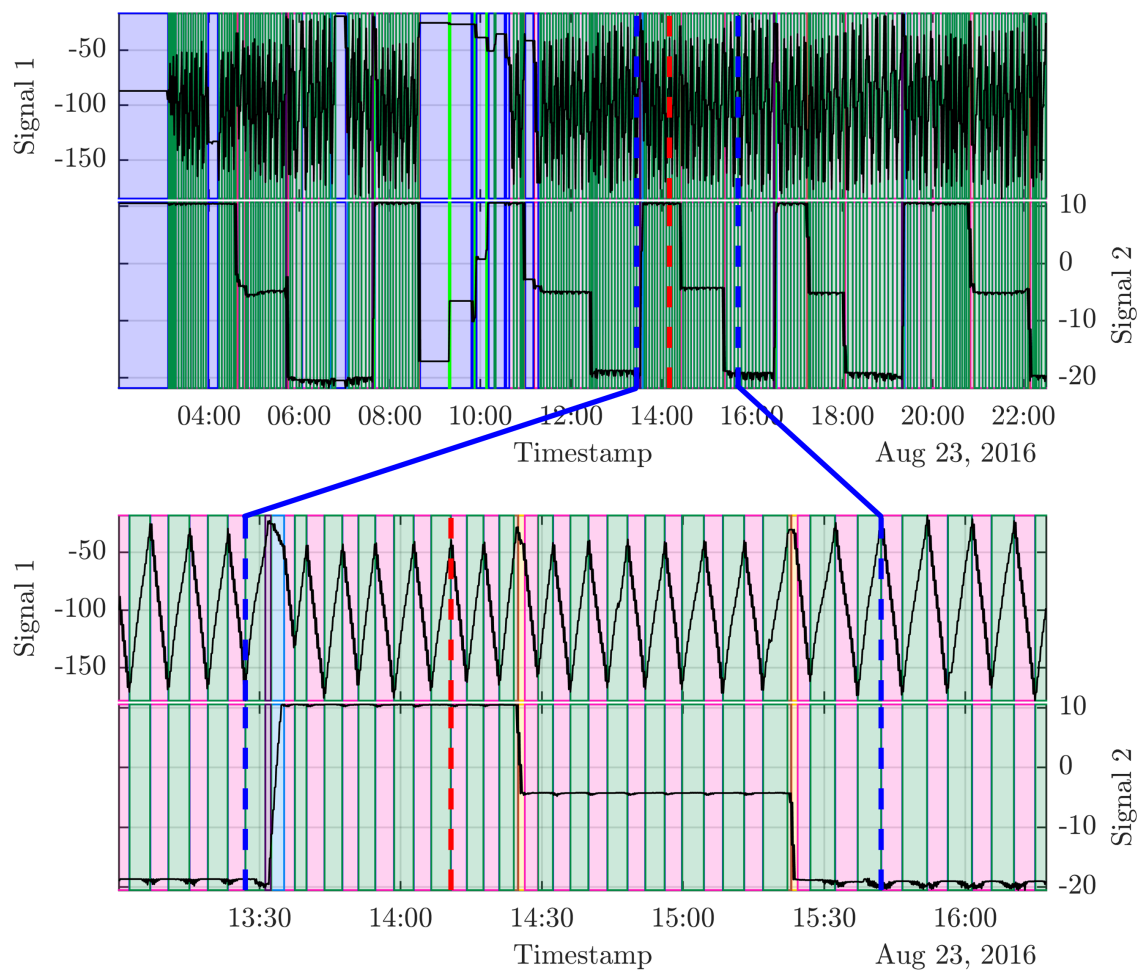


Fig. 3. Operation mode 1; the coloured areas illustrate the output of the MCLA; different colours represent different combinations of symbols from the SCLA of each channel (in this case two channels); the alphabet used for signal 1 and 2 consists of the three symbols [u, s, d]. Top: machine working in operation mode 1 with longer interrupts in-between (light blue area - both signals are stationary); Bottom: snippet of the signal showing the typical repeating pattern of operation mode 1.

dashed-red line (marked in both plots) have the same symbolic representation in both modes, whereas the portion of the signal after the dashed-red line shows a different colour-code for each mode.

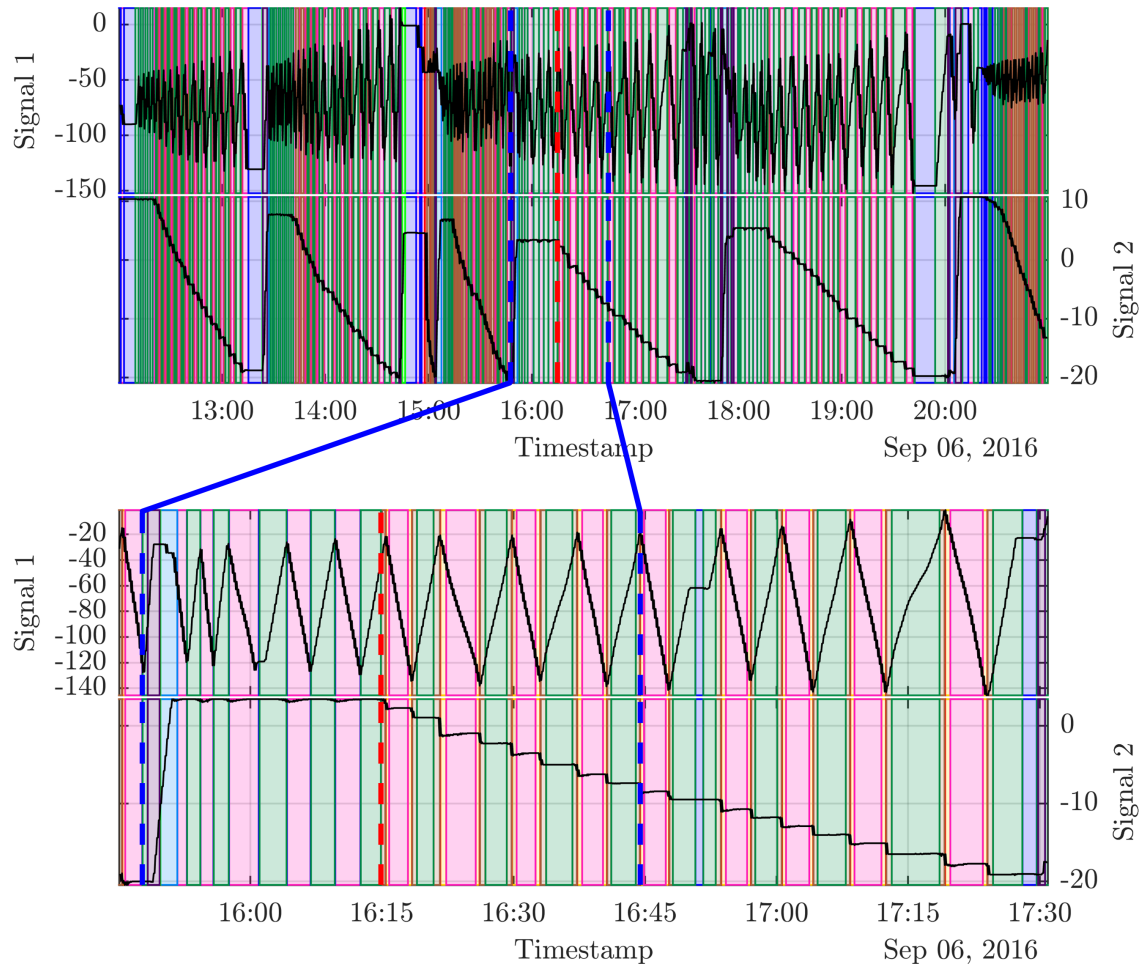


Fig. 4. Operation mode 2; the coloured areas illustrate the output of the MCLA; different colours represent different combinations of symbols from the SCLA of each channel (in this case two channels); the alphabet used for signal 1 and 2 consists of the three symbols [u, s, d]. Top: machine working in operation mode 2 with interrupts in-between (light blue area - both signals are stationary); Bottom: snippet of the signal showing the typical repeating pattern of operation mode 2.

The generated symbolic representation is used for further analyses. Building up histograms for occurring symbol combinations offers an insight in the overall behaviour of the system, see Fig. 5). This allows inter-machine comparison and comparison of different signal portions/ranges as well as classification of the operation mode. On top of Fig. 5 the histograms of the entire signal ranges shown in Fig 3 (top) and Fig 4 (top) are presented. The histograms for the typical repeating snippets, shown in Fig 3 (bottom) and Fig 4 (bottom), are visualized on the bottom. Since the machine is interrupted several times in both operating modes, the bins for the stationary state (**ss**) are more visible for the entire signal sequences (top). Excluding these bins, the statistics (histograms) of the shown snippets can act as representatives (motifs) for the operating modes. It can be seen that the histograms differ whether the machine is operating in

mode 1 (left) or mode 2 (right). Especially the occurrences of *dd* and *ud* reveal the differences. In future investigations the definition of a similarity measure for such histograms is planned to compare them qualitatively and may use this for automatic operation recognition and finding motifs. Note: sorting the histograms in decreasing order of occurrences will yield a classical frequency dictionary.

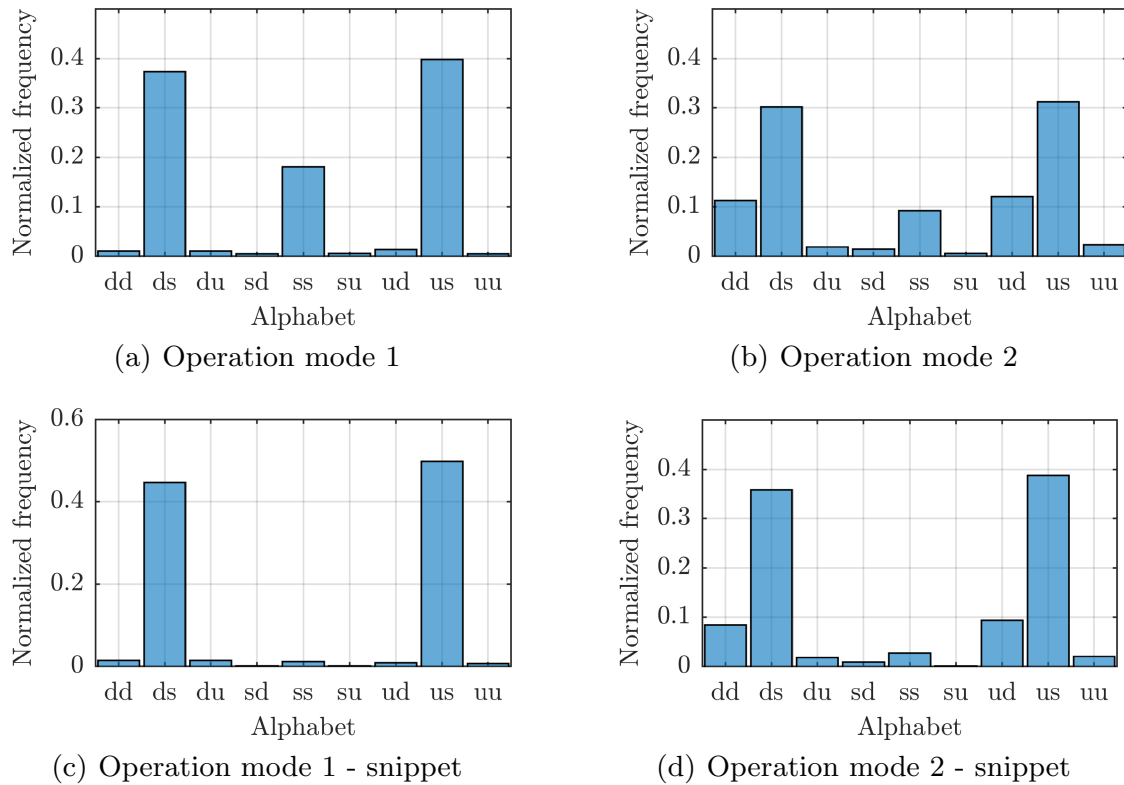


Fig. 5. Histograms of occurring symbol combinations of a machine in two different operation modes. Top: Histograms for the entire time range shown in Fig 3 (top) and Fig 4 (top); Bottom: Histograms for the signal snippets presented in Fig 3 (bottom) and Fig 4 (bottom).

A big advantage of the presented symbolic time series analysis is, that the sequence of symbols - either single or multi channel - can now be addressed with techniques more common to computational linguistics (e.g. *regex*) [13], which is a growing field of research.

3 Conclusion

Successful data analytics in large physical systems must embed the modelling of the individual component and complete system dynamics. This has been addressed by providing for a linear differential operator or its inverse in each and every signal- or derived-data-channel. A multi-variate symbolic time series analysis has

been introduced. It permits a symbolic view of the system and its dynamics. The concept of frequency dictionaries has been applied to automatic operation recognition; this functions for operation types which are characterised by a specific distribution of symbols. A major advantage of the proposed method is its intrinsic multi-scale property. This enables the identification of very short events in very large data sets. Currently, we are performing research on the relationships between the sequences of symbols and the metaphor of language. Initial results indicate that this opens the door to take advantage of new methods emerging in computational linguistics.

References

1. Baheti, R., Gill, H.: Cyber-physical systems. *The Impact of Control Technology* (2011) 161–166
2. Lanczos, C.: *Linear differential operators*. SIAM (1961)
3. O’Leary, P., Harker, M., Gugg, C.: An inverse problem approach to approximating sensor data in cyber physical systems. In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. Volume 2015-July., IEEE (may 2015) 1717–1722
4. Gugg, C., Harker, M., O’Leary, P., Rath, G.: An Algebraic Framework for the Real-Time Solution of Inverse Problems on Embedded Systems. In: *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*. Volume V., IEEE (aug 2015) 1097–1102
5. Harker, M., O’Leary, P.: *Discrete Orthogonal Polynomial Toolbox - Matlab File Exchange*
6. Gugg, C.: *An Algebraic Framework for the Solution of Inverse Problems in Cyber-Physical Systems*. Phd thesis, Montanuniversitaet Leoben (2015)
7. O’Leary, P., Harker, M.: An algebraic framework for discrete basis functions in computer vision. In: *Proceedings - 6th Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP 2008*, IEEE (dec 2008) 150–157
8. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery* **15**(2) (aug 2007) 107–144
9. Veenman, C., Reinders, M., Backer, E.: A maximum variance cluster algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(9) (sep 2002) 1273–1280
10. Chau, T., Wong, A.: Pattern discovery by residual analysis and recursive partitioning. *IEEE Transactions on Knowledge and Data Engineering* **11**(6) (1999) 833–852
11. Keogh, E., Lonardi, S., Chiu, B.Y.c.: Finding surprising patterns in a time series database in linear time and space. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’02*, New York, New York, USA, ACM Press (2002) 550
12. Daw, C.S., Finney, C.E.A., Tracy, E.R.: A review of symbolic analysis of experimental data. *Review of Scientific Instruments* **74**(2) (feb 2003) 915–930
13. Clark, A., Fox, C., Lappin, S.: *The Handbook of Computational Linguistics and Natural Language Processing*. Volume XXXIII. Wiley-Blackwell (2010)

12 | Symbolic Analysis of Machine Behaviour and the Emergence of the Machine Language

Originally appeared as:

R. Ritt and P. O’Leary, “Symbolic Analysis of Machine Behaviour and the Emergence of the Machine Language,” in *Theory and Practice of Natural Computing*, Springer International Publishing, 2018, pp. 305–316. DOI: 10.1007/978-3-030-04070-3_24. [Online]. Available: http://link.springer.com/10.1007/978-3-030-04070-3%7B%5C_%7D24

BibT_EX:

```
@incollection{Ritt2018a,
  author      = {Ritt, Roland and O'Leary, Paul},
  booktitle   = {Theory and Practice of Natural Computing},
  doi         = {10.1007/978-3-030-04070-3_24},
  keywords    = {compounding, cyber physical system, emergence,
                knowledge discovery, of language, segmentation,
                symbolic time series},
  pages       = {305--316},
  publisher   = {Springer International Publishing},
  title       = {{Symbolic Analysis of Machine Behaviour and the
                Emergence of the Machine Language}},
  url         = {http://link.springer.com/10.1007/978-3-030-04070-
                3{\_}24},
  year        = {2018}
}
```


Symbolic Analysis of Machine Behaviour and the Emergence of the Machine Language

Roland Ritt^[0000-0002-2519-8303] and Paul O’Leary

Chair of Automation – Departement Product Engineering
University of Leoben
Peter-Tunner-Straße 25, A-8700 Leoben, Austria
{roland.ritt,paul.oleary}@unileoben.ac.at

Abstract. This paper takes a fundamental new approach to symbolic time series analysis of real time data acquired from human driven mining equipment, which can be seen as stochastic physical systems with non analytic human interaction (hybrid systems). The developed framework uses linear differential operators (LDO) to include the system dynamics within the analysis, whereas the metaphor of language is used to mimic the human interaction. After applying LDO, the multidimensional data stream is converted into a single symbolic time series yielding a more abstract but highly condense representation of the original data. Inspired by natural language, the presented algorithm combines iteratively symbol pairs (word pairs) which occur frequently to new symbols/words; a machine-specific language emerges in a hierarchical manner, which automatically structures the dataset into segments and sub-segments. As a demonstration, the automatic recognition of operation modes of a bucket-wheel excavator is presented, proving the metaphor of language to be valuable in such hybrid systems.

Keywords: Knowledge discovery · Symbolic time series · Emergence of language · Compounding · Segmentation · Cyber physical system · Hybrid systems

1 Preamble

This paper addresses issues involved in the analysis of data from large plant and heavy machinery. In particular, we consider systems which are implemented as cyber physical systems¹ (CPS). In general this involves working with real time multidimensional time series. The inclusion of physical and chemical systems implies that the issues of dynamical systems must be considered. The determination of causes from observation in such systems is fundamentally an inverse problem. Consequently, linear differential operators are an integral part of the proposed

¹ The definition for CPS assumed here is: a system with a coupling of the cyber aspects of computing and communications with the physical aspects of dynamics and engineering that must abide by the laws of physics. This includes sensor networks, real time and hybrid systems.

processing. Additionally, there is significant human manual interaction with the systems being considered. The goal of the human interaction is to implement processes through procedures involving combinations of operations. However, the human behaviour cannot be modelled analytically, nor is it well described by techniques such as: finite state machines or hidden Markov models. The combination of stochastic dynamical systems and non analytical human behaviour implies that we are dealing fundamentally with hybrid systems. Consequently, new hybrid analysis techniques are required.

The continuous monitoring of plant and machinery through CPS techniques, makes very large volumes of data, with high temporal resolution, available for analysis. The availability of very large data sets, at the beginning of the 21st century, lead to people seriously predicting the "end of theory" [3], the suggestion being that learning techniques would replace scientific method. However, prominent failures of learning from very large data sets, e.g. Google Flue Trends [17], lead to the insight that it is essential to embed *understanding* and good scientific methods into data analysis relating to physical phenomena. The lack of causality, in the models implied by learning at that time, were the prominent cause of failure. In the mean time hybrid analysis systems, which combine theory and learning, have become a major focus of research [7,14]. However, as of yet the work has only looked at combining theory with learning.

In this paper we take a fundamentally new approach — inspired by the metaphor of emergence of language — and present a new algorithm for the automatic detection of hierarchical structure in data. The inspiration comes from the Asian model of phenomenology [20], which takes the view that language structures human thinking and is almost definitive for behaviour. The new approach presented here combines: LDO to enable the modelling of dynamics; symbolization of data to reduce numerosity and *a new algorithm which mimics the mechanism of compounding as observed in language*. Furthermore, this opens the door to applying techniques from computational linguistics to the analysis of sensor data from physical systems. We consider these methods to be complementary to theory driven learning.

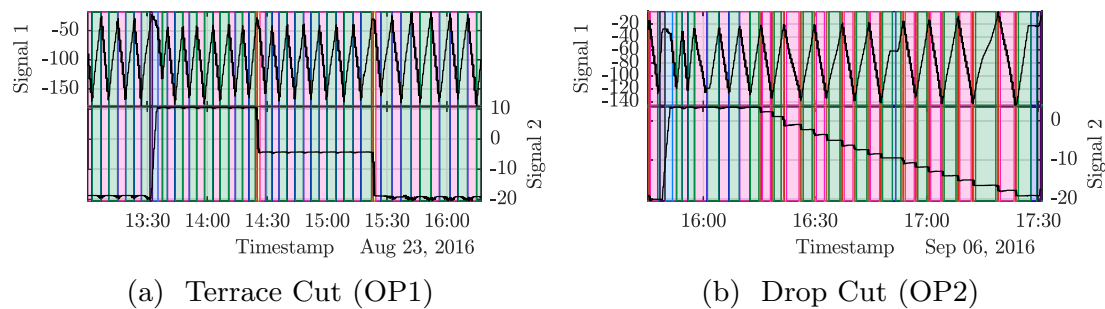


Fig. 1. Two different operation modes of a bucket-wheel excavator. The colours highlight different sub-operation modes

Within this paper we focus on investigating the performance of the new algorithm and its ability to detect hierarchical structure in data streams coming from real machinery. As a possible example, two specific operation modes from a bucket-wheel excavator are shown in Figure 1a and 1b. These modes are build from various sub-modes (color-shaded areas). To improve data analytics in such hybrid systems, these (and various other) operation modes have to be identified automatically in a multidimensional data stream, which is shown exemplarily in Figure 3b. In additional subsequent analysis the sub-modes may be used in further investigations.

2 Introduction and Related Work

The first — and most common — step in finding structure in time series² is to segment the data streams into relevant portions. Most techniques are based on identifying temporal locations of discontinuity in characteristic channels of the time series. A good overview of segmentation techniques for univariate data can be found in [26]. Good results have also been achieved using algorithms based on dynamic programming, e.g. [11,13]; however, with the drawback of high computational costs. Additionally, good results have been obtained using techniques based on heuristics, such as Top-Down, Bottom-Up and Sliding-Window algorithm, e.g. [15,19] and derived algorithms, e.g. [5,15]. The heuristics are normally less computationally intensive compared to dynamic programming. For multidimensional data, algorithms based on principal component analysis (PCA) and singular value decomposition (SVD) are available, e.g. [4,8,32].

Polynomial modelling of segments, see for example [9,10], is used in cases where simple straight line models are insufficient. Additionally, local Taylor approximations — implemented via polynomial approximation — can be applied to obtain information w.r.t. the derivative behaviour of a segment, e.g. [22]. This information can be used to determine the temporal locations of local maxima, minima or points of inflection which act as possible segmentation points, e.g., [26]. In the case of physical system it is advantageous to embed a priori knowledge about the underlying process/system, e.g. integrating system models [1,8,12].

The concept of transforming time series into a *symbolic time series*, e.g. [18], was introduced to reduce the numerosity. Based on this [33] developed a mechanism to address dynamic time warping and find segments with the same shape but different length. This concept will prove important in this paper since in human driven machines, the same operation may be performed spanning different durations; this is tantamount to dynamic time warping. A further interesting concept, especially for dynamic systems, is to symbolize a signal within it's phase-space as presented in [27]. This embeds additional information in the form of derivatives in the analysis, which is also intended by the authors of this paper.

² There is much work on time series for the analysis of data relating to financial transactions. However, financial transactions are not bound to any physical laws; consequently much of these techniques are not applicable to CPS

Multidimensional symbolic time series are investigated in [6,23,21] and contain the first idea of combining symbols along multiple channels to obtain a new data stream containing combinations of symbols. This stream is then analysed to find frequent patterns, using algorithms such as the *Apriori algorithm* [2]. Although the previous authors use symbols/words to discretize a time series, there is no meaning associated to the symbols in use. [22] used this idea in his work to segment data based on the shape in a tree-like structure. More advanced techniques used for finding structure within data are inspired by natural language and grammars, e.g. [24,30,29].

A brief discussion of *discontinuity* when performing discrete time observations of continuous physical systems is necessary at this point: discrete time-series are discontinuous at every point by their very nature. Consequently, we need to introduce some measure related to the physical system to define and detect characteristics which are to be considered as a discontinuity. In this paper we introduce the embedding of linear differential operators (LDO) [16] to model the behaviour of dynamical systems from which the time series are emanating. The dynamics of the system determine the computation of maximum values for derivatives which have physical meaning. This in turn permits us to determine the required sampling rate, so that the dynamics can be correctly identified and not misinterpreted as discontinuities. Sampling rate in the range of ten's of hundred's of milliseconds may be required to characterize the dynamics; whereas human operation and processes are significantly slower. Consequently, high temporal resolution over long periods of time are commonly required. The high temporal resolution over extended time periods will lead to very long symbolic series. Consequently, naive methods such as the Apriori algorithm — which take no advantage of hierarchy — suffer from the extremely high number of symbol combinations and permutations. Consequently, we consider it to be essential to introduce some form of hierarchical structure detection to deal with CPS in an efficient manner. In this paper we shall concentrate on one such mechanism: this work is inspired by the mechanism of *compounding* as observed in the emergence of natural language. That is, common sequences of words are compounded (or sometimes contracted) to form new polysyllabic words. These new words are then represented using new single symbols. This compounding is performed in a hierarchical manner, revealing implicit structure in the data.

To address the afore mentioned problems, the main contributions of this paper are:

1. The embedding of LDO to model dynamical systems in real time. This enables the integration of a priori knowledge about the physics of the system being observed.
2. Development of a new algorithm to segment multidimensional time series data in a hierarchical manner to detect the underlying structure and sub-structures based on symbolic time series analysis. We propose the use of the metaphor of the emergence of a natural machine language, as a helpful support tool, e.g. the mechanisms of natural language become accessible.

3. Using a new way of symbolizing time series data, which includes a priori knowledge about the underlying system. The combination of LDO with symbols permits the association of *meaning* with the symbols in a physical sense. In this manner human readable sequences of symbols can be generated.
4. Using the metaphor of language and the mechanism of *compounding* — inspired from natural language evolution — to combine frequent combinations of symbols (words) to form new words. Run-lengths of the same word are combined and predicated with its length which addresses the problem of dynamic time warping.
5. The proposed mechanisms are verified with time series data emanating from the high resolution monitoring of a large piece of mining equipment.

3 Methodology

The process we are investigating can be segmented into a series of computational aspects:

1. Apply linear differential operators (LDO) to include the dynamics of the system into the analysis.
2. Symbolize multiple channels, and combine them to get a stream of polysyllabic words.
3. Iteratively combine frequent word combinations within the stream of words to find the underlying structure and reveal ‘the machine language’. This is called hierarchical compounding.

3.1 Linear Differential Operator (LDO)

Modelling the dynamical behaviour of physical systems is done using differential equations. Using linear differential operators (LDO, see [16]) an ordinary differential equation (ODE) of the form

$$a_d \frac{d^d}{dt^d} y(t) + a_{d-1} \frac{d^{d-1}}{dt^{d-1}} y(t) + \cdots + a_1 \frac{d}{dt} y(t) + a_0 y(t) = g(t) \quad (1)$$

can be rewritten as

$$Ly(t) = g(t) \quad \text{with} \quad L = \sum_{i=0}^d a_i \left(\frac{d}{dt} \right)^{(i)}, \quad (2)$$

where the function of interest $y(t)$ is a function of t , $g(t)$ is the excitation function, $\left(\frac{d}{dt} \right)^{(i)}$ is the i -th derivative operator and a_i is the according coefficient.

A discrete implementation for solving ODE in this manner can be found in [25]. Using this, Equation 2 can be written as matrix-vector equation

$$Ly = g, \quad (3)$$

where L is a linear differential operator matrix, the vector \mathbf{y} corresponds to sampling the function $y(t)$ at a set of locations $\mathbf{t} \triangleq [t_1 \dots t_n]^T$, similarly for \mathbf{g} and $g(t)$.

In our framework the LDO is applied to the sensor channels as appropriate. This is, the LDO is selected to model the ODE approximating the dynamics of the system or as a generic means of computing regularized derivatives; which yield estimates for the state vectors. This step is a precursor to the assignment of symbols and approaching the metaphor of language.

3.2 Advanced Symbolic Time Series Analysis (ASTSA)

The transition to *language* starts with the assignment of symbols to the signal — to support the metaphor of language we shall call the symbols words. The SAX algorithm [18] uses a linear quantization of the input signal to what they call an alphabet. Here we take a fundamentally different approach: during an *exploratory phase* the statistical modes of the signal are determined. The modes may be based on histograms or distributions of entropy. A *dictionary* is defined and a symbolic name, i.e., a word is assigned to each mode. The dictionary contains the value ranges and the associated words.

In the *operative phase* various LDO are applied and the resulting time series are converted to symbolic time series, i.e. a sequence of words. The assignment of the corresponding word to each sample of each channel is performed according to the dictionary associated with the sensor channel. At this point we have time series of words. There may be multiple sequential occurrences of the same word in a symbolic times series. These cases can be *contracted* to a single occurrence of the word and an associated *predicate* corresponding to the number of occurrences. This yields a very high compression ratio, which is nevertheless lossless, i.e., we can reconstruct the original sequence of words without error [28].

This process is shown exemplarily on the left side of Figure 2. Here an LDO is chosen which acts as a derivative operator. To all positive values of the resulting signal the word up (u), to all negative values the word down (d) and to all values which are zero the word stationary (s) is assigned. In this manner meaning (about the trend of the signal) is associated to the original signal. Note: If only the words without the predicated lengths are used to compare signals, the problem of dynamic time-warping is solved directly, as it can be inspected in the plot.

In a subsequent step, parallel sequences of words from a collection of multiple channels can be automatically combined to form a new single-channel polysyllabic time series, see Figure 2 (top-right). This step is beneficial for analysing multidimensional time series and is the precursor for using existing single-channel symbolic analysis methods. As a result, the first level of segmentation of multidimensional time series is achieved with this method. For getting deeper insights in different levels of the underlying structure, the following algorithm for hierarchical compounding of words is developed.

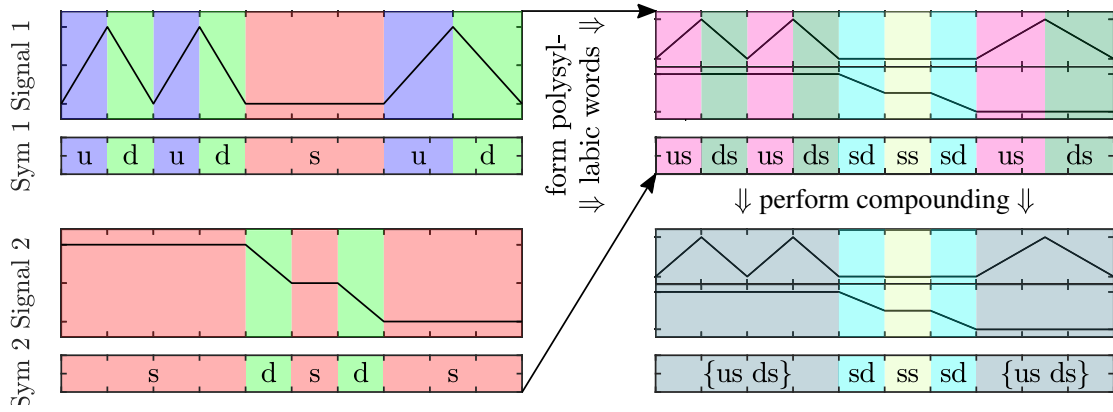


Fig. 2. Process of compounding of words; left: two signals are symbolized (based on their derivatives: *u*...up, *d*...down, *s*...stationary, colours represent the different words); right-top: the symbols are combined to form polysyllabic words; right-bottom: frequently recurring word-pairs (*us-ds*) are compounded

3.3 Hierarchical Compounding of Words

Here we focus on explicating the hierarchical compounding of words, since this is the most significant new contribution in this paper. A formal definition of the algorithm is given in Algorithm 1. As is the case in natural language, we do not expect random combinations and permutations of word sequences [31]. This is simply a reflection of the fact that the machine is not being operated in a random manner, nor do the automatic controllers, which are in essence Turing machines, generate random sequences.

Given the j^{th} . channel we have the dictionary \mathcal{D}_j containing the definitions for m words; $\mathcal{D}_j(i)$ is used to denote the i^{th} . entry in the dictionary. We define M_j an $m \times m$ matrix for the j^{th} . channel, whereby at the location $M_j(i, k)$ the counts $n_{i,k}$ for the observed occurrences of the bigram sequence $[\mathcal{D}_j(k), \mathcal{D}_j(i)]$ are entered.

$$M_j \triangleq \begin{bmatrix} 0 & n_{1,2} & \dots & n_{1,m} \\ n_{2,1} & & \ddots & \vdots \\ \vdots & \ddots & & n_{m-1,m} \\ n_{m,1} & \dots & n_{m,m-1} & 0 \end{bmatrix} \quad (4)$$

The previous contraction ensures that the diagonal of this matrix contains zeros. M_j can be used to find the most common word combinations - but this is not a measure of coherence, since the absolute count is taken as measure. A more scaleable approach is to normalize M_j such that the sum of entries in each column is 1; we obtain the Markov probability matrix \hat{M}_j . That is, $\hat{M}_j(i, k)$ is the probability of $\mathcal{D}_j(k)$ being followed by $\mathcal{D}_j(i)$. Note: if the sum of one column in M_j is one, this word combination must not be merged. This is, if a word occurs only once in the stream the probability of the combination of this word with the subsequent word is 100%. This causes a growth of this combination, which is not desirable.

Subsequently, a new single word is defined corresponding to the bigram with the highest probability and added to the dictionary. All occurrences of this bigram in the symbolic time series are now replaced by the new word; yielding a further compression of the sequence. A contraction is performed following each compounding step. Note: Both, simultaneous repetitions and also repetitions identified by the compounding lead to polysyllabic words.

The compounding and contraction process is repeated till a stopping criterion is met (e.g. no more bigrams to merge or the alphabet-size is bigger than number of words in the time series). This is effectively always a trade off between the size of the dictionary and the compression ratio for the sequence. At some point adding new words no longer yields a significant compression of the sequence.

Each iteration forms a new level of hierarchical compounding revealing different structures and substructures within the data. The mechanism of contraction

Algorithm 1: Hierarchical Compounding of Words

```

Input  :  $sts$                                      // a symbolic time series
            $\mathcal{D}$                                      // a finite set of words (symbols) building the dictionary  $\mathcal{D}$ 
Output:  $sts_c$                                        // hierarchical collection of compressed symbolic time series
            $\mathcal{D}_n$                                      // expanded dictionary including compounds which are performed

[1]  $i := 0$ ;                                           // initialize counter
[2]  $sts_c := []$ ;                                       // initialize  $sts_c$ 
[3]  $\mathcal{D}_n = \mathcal{D}$ ;                                     // initialize  $\mathcal{D}_n$  with  $\mathcal{D}$ 
[4] repeat
[5]    $sts_c(i) := \text{PERFORMCONTRACTION}(sts)$ ;         // perform contraction
[6]    $S := \text{FINDSEQTOCOMPOUND}(sts_c(i), \mathcal{D}_n)$ ;       // find sequence of words
                                                    // to be compounded
[7]    $\mathcal{D}_n := [\mathcal{D}_n, S]$ ;                             // add sequence  $S$  as new word to  $\mathcal{D}_n$ 
[8]    $sts := \text{MERGESEQ}(sts_c(i), S)$ ;               // replace the found sequences in the symbolic
                                                    // time series with the new compound
[9]    $i := i + 1$ ;                                       // increase counter
[10] until  $\text{STOPPINGCRITERIAMET}(sts) \neq \text{true}$ ;
[11]  $sts_c(i) := \text{PERFORMCONTRACTION}(sts)$ ;         // perform contraction for last level

```

and compounding are common in linguistics. As it will be seen in the following application, the metaphor of language is proving powerful, e.g., operation modes can be identified automatically. The emerging structure of the symbols we shall regard as the emergence of a machine's own language.

In Figure 2 (bottom-right) it can be seen, that after the compounding, all ranges where signal 1 is oscillating and signal 2 is stationary are identified as having the same underlying structure ($\{us\ ds\}$; note: the curly braces indicate compounding). Due to the contraction process, the number of oscillations does not matter. This is advantageous to other algorithms, since the structure of the signal is identified and not only strictly similar repeated patterns.

4 Background - Relations to Natural Language

The fundamental premiss being investigated in this work is the Yogācāra phenomenological model on the emergence of human speech [20]. It proposes that

observed repetitions in our sensory excitation, which are significant to our situation, are assigned a language representation, i.e., words. Consequently, meaning is simultaneously experiential and contextual. Sensory excitation in multiple senses are merged, by the portion of mind known as *mano-vijhana*, to perceptions of objects (nouns) and activities (verbs). In Proto-Indo-European languages each syllable has a specific meaning and more complex experiences are expressed as poly-syllabic words. Furthermore, predicates — primarily adjectives and adverbs — emerge to define properties of objects and activities. Punctuation, is a relatively late development, it first became necessary with the wider availability of books, where the natural pauses of the spoken language, which give meaning, were not available. In particular, Latin liturgical texts were more carefully punctuated than the vernacular, the aim was to circumvent misreading which might lead to heretical meanings. In this work we investigate the usefulness of the metaphor of language when analysing real time machine data — particularly when they are being operated by humans. However, we shall extend these concepts by proceeding them with a preprocessing of the sensor signals with linear differential operators. The idea of this work is, that words derived from the original signal are considered to constitute nouns and from derivatives, i.e, activity in a signal, verbs. The lengths from the contraction process are predicates, i.e, adjectives and adverbs. In a more advanced language, the states where the machine is not operated are defined to be punctuations. Defined in this manner, the symbolization converts the data stream into a readable text. However, this work only deals with the aspect of compounding frequent word combinations to form new words. These words describe more complex machine states and thus reveal the underlying structure which can be used for segmentation of the original data.

5 Experimental Evaluation

The above presented algorithm for hierarchical compounding of symbolic time series is tested on real time data emanating from large human driven mining machinery (e.g. bucket-wheel excavators). The slewing and luffing angle (Signal 1 & 2) of an excavator are used to identify different operating modes (see Figure 1a and 1b). These signals are sampled with $f_s = 1$ Hz.

As an LDO, a regularized first derivative operator is used locally on each channel with a support length of $l_s = 111$ samples. For the approximation of the derivative a polynomial of degree $d = 2$ is used. The resulting signals are symbolized: the word stationary (s) is assigned to values within the range of $-0.0231^\circ \text{s}^{-1}$ to $0.0231^\circ \text{s}^{-1}$ for slewing and between $-0.0081^\circ \text{s}^{-1}$ to $0.0231^\circ \text{s}^{-1}$ for luffing. The word down (d) and up (u) are assigned to values below and above those limits, indicating that the trend of the signal is downwards or upwards. After this, the two symbolic time series are combined to form a single polysyllabic time series (see Figure 3a). On this time series, the herein presented algorithm is tested. The result after 17 iterations is shown in Figure 3b.

As it can be seen, the two operation modes presented above are automatically identified.

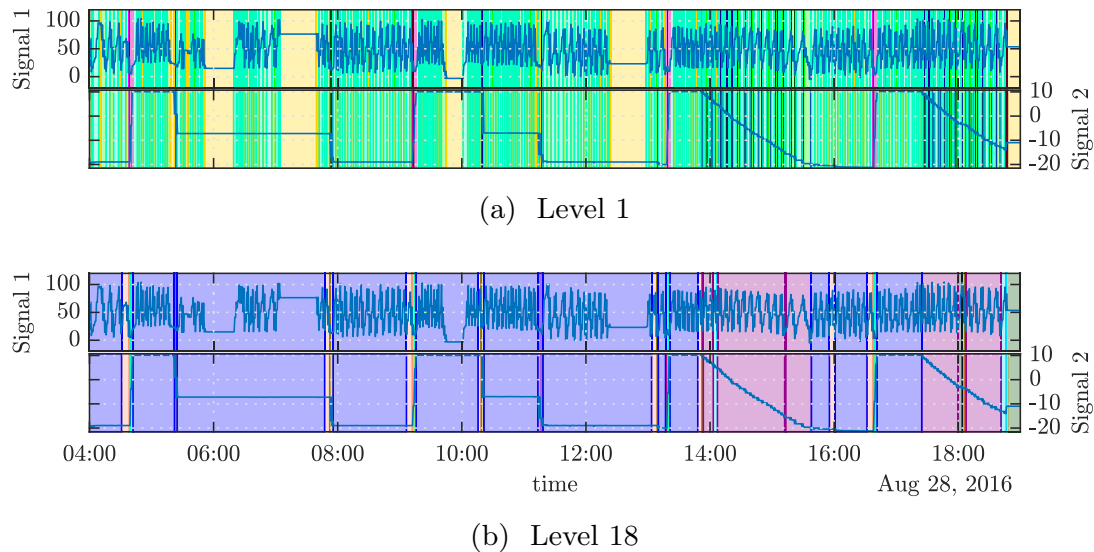


Fig. 3. Data from a bucket-wheel excavator analysed with the hierarchical compound-ing algorithm; Top: symbolized data without compounding; Bottom: symbolized data after 17 iterations - the operation modes are revealed and can be used for segmentation

6 Conclusion and Future Work

The presented methods have proved, that including the metaphor of language in the analysis of data emanating from large physical systems is powerful and of major advantage. Transforming a time series into a symbolic time series using ASTSA includes the dynamics of the system by applying an LDO and thus includes a priori knowledge in further analysis. A new algorithm is proposed which identifies implicit structure in real time machine data. Via the application of LDO and the metaphor of language this emergent structure can be likened to machine having its own language. This language automatically partitions large data sets into segments with the same behaviour, which can be used in further investigations. In future research, the composition of the machines own language is to be improved using different techniques for finding recurring sequences as well as using the more advanced definitions for language as introduced in Section 4 of this paper.

Acknowledgments

This work was partially funded under the auspices of the EIT - KIC Raw materials program within the project "Maintained Mining Machine" (MaMMa) with the grant agreement number: [EIT/RAW MATERIALS/SGA2018/1]

References

1. Agarwal, R., Gotman, J.: Adaptive segmentation of electroencephalographic data using a nonlinear energy operator. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems VLSI **4**, 199–202 (1999)
2. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules in Large Databases. *Journal of Computer Science and Technology* **15**(6), 487–499 (1994)
3. Anderson, C.: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete (2008), <https://www.wired.com/2008/06/pb-theory/>
4. Banko, Z., Dobos, L., Abonyi, J.: Dynamic Principal Component Analysis in Multivariate Time-Series Segmentation. *Conservation, Information, Evolution* **1**(1), 11–24 (2011)
5. Borenstein, E., Ullman, S.: Combined top-down/bottom-up segmentation. *IEEE transactions on pattern analysis and machine intelligence* **30**(12), 2109–25 (dec 2008)
6. Esmael, B., Arnaout, A., Fruhwirth, R.K., Thonhauser, G.: Multivariate Time Series Classification by Combining Trend-Based and Value-Based Approximations. In: *Computational Science and Its Applications – ICCSA 2012*, vol. 7336 LNCS, pp. 392–403. Springer, Berlin, Heidelberg (2012)
7. Faghmous, J.H., Banerjee, A., Shekhar, S., Steinbach, M., Kumar, V., Ganguly, A.R., Samatova, N.: Theory-Guided Data Science for Climate Change. *Computer* **47**(11), 74–78 (nov 2014)
8. Feil, B., Abonyi, J., Nemeth, S., Arva, P.: Monitoring process transitions by Kalman filtering and time-series segmentation. *Computers and Chemical Engineering* **29**(6 SPEC. ISS.), 1423–1431 (2005)
9. Fuchs, E., Gruber, T., Nitschke, J., Sick, B.: Online Segmentation of Time Series Based on Polynomial Least-Squares Approximations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(12), 1–15 (dec 2009)
10. Grabocka, J., Wistuba, M., Schmidt-Thieme, L.: Scalable Classification of Repetitive Time Series Through Frequencies of Local Polynomials. *IEEE Transactions on Knowledge and Data Engineering* **27**(6), 1683–1695 (jun 2015)
11. Guo, H., Liu, X., Song, L.: Dynamic programming approach for segmentation of multivariate time series. *Stochastic Environmental Research and Risk Assessment* **29**(1), 265–273 (2015)
12. Han, Z., Chen, H., Yan, T., Jiang, G.: Time Series Segmentation to Discover Behavior Switching in Complex Physical Systems. In: *2015 IEEE International Conference on Data Mining*. pp. 161–170. IEEE (nov 2015)
13. Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., Toivonen, H.: Time series segmentation for context recognition in mobile devices. *Proceedings 2001 IEEE International Conference on Data Mining* pp. 203–210 (2004)
14. Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V.: Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering* **29**(10), 2318–2331 (oct 2017)
15. Keogh, E., Chu, S., Hart, D., Pazzani, M.: Segmenting Time Series: A Survey and Novel Approach. *Data Mining in Time Series Databases* pp. 1–21 (2003)
16. Lanczos, C.: *Linear differential operators*. SIAM (1961)
17. Lazer, D., Kennedy, R., King, G., Vespignani, A.: The Parable of Google Flu: Traps in Big Data Analysis. *Science* **343**(6176), 1203–1205 (mar 2014)

18. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery* **15**(2), 107–144 (aug 2007)
19. Lovrić, M., Milanović, M., Stamenković, M.: Algorithmic methods for segmentation of time series: an overview. *Journal of Contemporary Economic and Business Issues* **1**(1), 31–53 (2014)
20. Lusthaus, D.: *Buddhist Phenomenology: A Philosophical Investigation of Yogacara Buddhism and the Ch'eng Wei-shih Lun*. Curzon critical studies in Buddhism, Routledge Curzon (2002)
21. Minnen, D., Isbell, C., Essa, I., Starner, T.: Detecting Subdimensional Motifs: An Efficient Algorithm for Generalized Multivariate Pattern Discovery. In: *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. pp. 601–606. IEEE (oct 2007)
22. Minnen, D., Starner, T., Essa, I., Isbell, C.: Discovering Characteristic Actions from On-Body Sensor Data. In: *2006 10th IEEE International Symposium on Wearable Computers*. pp. 11–18. IEEE (oct 2006)
23. Minnen, D., Starner, T., Essa, I., Isbell, C.: Discovering Characteristic Actions from On-Body Sensor Data. In: *2006 10th IEEE International Symposium on Wearable Computers*. pp. 11–18. IEEE (oct 2006)
24. Nevill-Manning, C.G., Witten, I.H.: Identifying Hierarchical Structure in Sequences: A linear-time algorithm. *Journal of Artificial Intelligence Research* **7**(1), 67–82 (aug 1997)
25. O'Leary, P., Harker, M., Gugg, C.: An inverse problem approach to approximating sensor data in cyber physical systems. In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. vol. 2015-July, pp. 1717–1722. IEEE (may 2015)
26. Panagiotou, V.: *Blind segmentation of time-series*. Ph.D. thesis, Delft University of Technology (2015)
27. Rajagopalan, V., Ray, A., Samsi, R., Mayer, J.: Pattern identification in dynamical systems via symbolic time series analysis. *Pattern Recognition* **40**(11), 2897–2907 (2007)
28. Ritt, R., O'Leary, P., Rothschedl, C.J., Harker, M.: Advanced Symbolic Time Series Analysis In Cyber Physical Systems. In: Valenzuela, O., Rojas, F., Pomares, H., Rojas, I. (eds.) *Proceedings - International work-conference on Time Series, ITISE 2017*. vol. 1, pp. 155–160. University of Granada, Granada (feb 2017)
29. Senin, P., Lin, J., Wang, X., Oates, T., Gandhi, S., Boedihardjo, A.P., Chen, C., Frankenstein, S.: GrammarViz 3.0: Interactive Discovery of Variable-Length Time Series Patterns. *ACM Transactions on Knowledge Discovery from Data* **12**(1), 1–28 (feb 2018)
30. Senin, P., Lin, J., Wang, X., Oates, T., Gandhi, S., Boedihardjo, A.P., Chen, C., Frankenstein, S., Lerner, M.: GrammarViz 2.0: A Tool for Grammar-Based Pattern Discovery in Time Series. In: *Machine Learning and Knowledge Discovery in Databases*. pp. 468–472. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)
31. Shannon, C.E.: A Mathematical Theory of Communication. *Bell System Technical Journal* **27**(3), 379–423 (jul 1948)
32. Spiegel, S., Gaebler, J., Lommatzsch, A., Luca, E.D., Albayrak, S.: Pattern Recognition and Classification for Multivariate Time Series. *Time* pp. 34–42 (2011)
33. Sun, C., Stirling, D., Ritz, C., Sammut, C.: Variance-wise segmentation for a temporal-adaptive SAX. *Conferences in Research and Practice in Information Technology Series* **134**(AusDM), 71–77 (2012)

Part IV

Applied Data Analytics in Cyber Physical Systems

13 | Synopsis

This part of the thesis contains the papers which investigate topics within the full data analytics cycle (as presented in Section 2.3.3) starting from the sensor-level up to applying the data analytics framework. As it can be seen, a broad field of knowledge is necessary to deal with the different steps involved. Profound knowledge on the sensor-level is important to collect the needed information and make a causal link to their origin. To extract knowledge from the data, information content is an indicator for which data may be of importance and relevant for further investigations. These ideas and the availability of a functional data collection and analysis framework than serve as a support for various applications.

The importance of knowledge in the sensor-level to ensure correct measurements is presented in [P11] (see Chapter 14). In this paper, the noise behaviour of inclinometers with two sensing elements in opposite directions is characterized. To identify the correlation of the sensing elements, singular value decomposition is used. Based on this, the misalignment of the sensors can be corrected to perform measurements with an improved confidence interval. Although Gaussian noise is assumed in most of the available data analytics algorithms, this paper shows that the noise of the used sensors is more correctly modelled using a Cauchy Lorenz distribution. This has to be considered in subsequent use of the data.

Although every available source of information should be used in data analytics, a common and sometimes necessary step is to separate data into portions with significant and insignificant information content, especially for analysing large volumes of data. In [P10] (see Chapter 15) Shannon Entropy is used to analyse the forces measured on a tool holder with the goal of identifying tool wear in manufacturing processes (drilling and milling). Local entropy is used to divide the data into individual drilling/milling operations. These segments can then be used to calculate the segment-entropy, which acts as an indicator for the health of the tool. Additionally, the idea of time-varying histograms is introduced which is also used in [P6].

In paper [P7] (see Chapter 16) an overview of the relevance of the full data analytics cycle in the context of mining and raw material handling is given. Within this paper the basic idea of the fundamental premiss of sensor data analytics is presented. Based on that, the means of collecting data on large physical machines used in the mining environment is presented. Those topics are revisited in detail in Chapter 2.

Various fields of applications are identified and investigated which are relevant for both, machine constructors and operators.

Hydraulic sensors and actuators are widely used in mining equipment, and thus are an important source of information when analysing those machines. Malfunctioning of the hydraulic system can lead to unwanted behaviour and damage to the machine. Therefore, the herein presented data analytics framework (see Chapter 4) is used to analyse the data of a bucket-wheel excavator [P6] (see Chapter 17). The monitored machine uses a parallel hydraulic system for the main lifting boom. By investigating the statistics of the pressure signals, the existence of negative pressure was verified, which can cause cavitation and normally results in the damage to system components. This indicates a possible design error in the hydraulic system, which can be reported back as engineering feedback. Additionally, time-varying histograms are used to detect changes in the system behaviour. This was used to find a flawed sensor. With the addition of metadata, the pressure signals are used to calculate sum and difference forces of the parallel hydraulics. Investigating these signals show that there is a continuous torsion on the main boom due to the skewed bucket-wheel.

To conclude: following the steps involved in the data analytics premiss is the prerequisite to base data analytics on causality. Therefore, including theoretical profound knowledge within each step is important. The availability of the data analytics framework supports the data scientist to focus on development of new methods and the analysis itself, which is valuable for inspecting large multi-dimensional time series.

14 | MEMS Based Inclinometers: Noise Characteristics and Suitable Signal Processing

Originally appeared as:

R. Schmidt, P. O’Leary, R. Ritt, and M. Harker, “MEMS based inclinometers: Noise characteristics and suitable signal processing,” in *I2MTC 2017 - 2017 IEEE International Instrumentation and Measurement Technology Conference, Proceedings*, IEEE, May 2017, pp. 1–6, ISBN: 9781509035960. DOI: 10.1109/I2MTC.2017.7969830. [Online]. Available: <http://ieeexplore.ieee.org/document/7969830/>

BibTeX:

```
@inproceedings{Schmidt2017,  
  author      = {Schmidt, Roland and O'Leary, Paul and Ritt,  
                Roland and Harker, Matthew},  
  booktitle   = {I2MTC 2017 - 2017 IEEE International  
                Instrumentation and Measurement Technology  
                Conference, Proceedings},  
  doi         = {10.1109/I2MTC.2017.7969830},  
  isbn        = {9781509035960},  
  month       = {may},  
  pages       = {1--6},  
  publisher   = {IEEE},  
  title       = {{MEMS based inclinometers: Noise characteristics  
                and suitable signal processing}},  
  url         = {http://ieeexplore.ieee.org/document/7969830/},  
  year        = {2017}  
}
```


This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication

MEMS Based Inclinometers: Noise Characteristics and Suitable Signal Processing

Roland Schmidt, Paul O’Leary, Roland Ritt and Matthew Harker
Chair of Automation, Department Product Engineering
University of Leoben, Styria, Austria
Email: automation@unileoben.ac.at

Abstract—This paper presents a detailed modelling, analysis and experimental verification of the noise behaviour for two families of MEMS based inclinometers. These MEMS use two accelerometers in opposing directions to measure inclination. Large sample sizes have been used to ensure reliable statistical results. The results reveal a differential susceptibility of the two sensors to ambient vibrational noise. It is shown that singular value decomposition can be used to orthogonalize the data emanating from the two sensors, yielding a significant improvement in the confidence interval of the measurement result. It is experimentally verified that the noise observed in the application of the investigated MEMS system has a Cauchy-Lorentz distribution. This leads to the fundamental necessity for non-linear signal processing if reliable performance is to be obtained in safety relevant applications.

I. INTRODUCTION

Inclinometers can be found in many applications [1]–[10]. In the monitoring of civil-structures [1], [4], [5], [10], [11] and ground-subsidence [3]. This paper presents a detailed modelling, analysis and experimental verification of the noise behaviour of micro-electro-mechanical system (MEMS) based inclinometers. When monitoring structures on large building cites, e.g. [12], ambient vibrational noise and impulses result in perturbations of the acquired signal and must be considered in the noise analysis.

Detailed mathematical approaches to the analysis of data from multiple inclinometers have been provided in the past [7]. Additionally, the startup behaviour of such MEMS has been considered in [9]. The noise behaviour of MEMS based inertial sensors was presented in [13], the Gauss-Markov (GM) and Auto-Regressive (AR) methods presented there assume gaussian perturbations. A Kalman based approach was proposed in [14]; once again this assumes a Gaussian noise model to obtain a maximum likelihood prediction. Newer research on inclinometer implementations can be found in [15]. None of the above literature reports detailed statistical analysis which document that noise observed with MEMS-based inclinometers is best modelled by a Cauchy-Lorentz distribution. This is highly significant; since, maximum likelihood predictors are not appropriate in this case; actually, most classical filter design procedures will fail to produce the desired results.

The main contributions of this paper are:

- 1) The experimental acquisition of long term data to determine the noise behaviour of two families of MEMS based

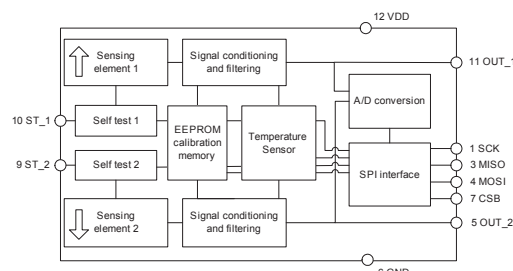


Fig. 1: Schematic of the SCA103T MEM inclinometer sensor.

inclinometers. The data is representative of both laboratory and realistic application conditions.

- 2) A detailed statistical modelling and analysis of the noise behaviour of dual-sensor differential inclinometers. Although each of the sensors individually exhibit a Gaussian noise behaviour, the difference signal, i.e., the inclination has a Cauchy-Lorentz distribution. This behaviour has not been previously reported in literature and no detailed analysis was available up until now.
- 3) A differential sensing model with individual sensor gains is proposed: $i(t) = \alpha \cdot x(t) - \beta y(t)$ and $a(t) = \beta \cdot x(t) + \alpha y(t)$, together with a singular value decomposition based orthogonalization procedure to obtain optimal values for α and β . The procedure yields minimum entropy and maximum likelihood results for $i(t)$ and $a(t)$ respectively. This leads to a significant improvement in the width of the probability distributions of $i(t)$ and $a(t)$, which reflects an improvement in the uncertainty bound for the final measurement.

In this paper we analyse data from the MEMS based SCA103T-D04 and SCA830-D07 families of devices from Murata. The schematic diagrams for the respective devices are shown in figures 1 and 2.

A laboratory test-rig was developed to enable the static positioning of the inclinometer sensors for deflections in the range $\pm 25^\circ$ with an accuracy of 0.25° , (i.e., approx 1%, see Fig. 3). This accuracy $\varepsilon = 1\%$ was considered sufficient, since it is the goal of the measurements to characterize the noise at different inclinations, but not to determine the absolute accuracy of the sensors.

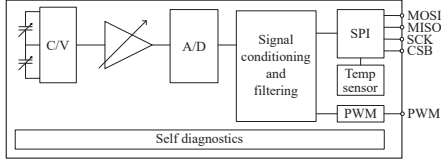


Fig. 2: Schematic of the SCA830-D07 MEM inclinometer sensor.

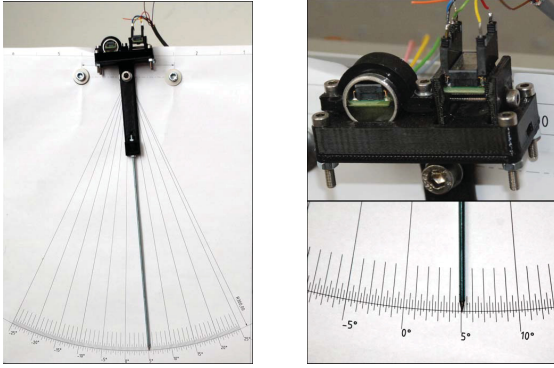


Fig. 3: Above left: shows the experimental setup as a whole. Above right upper: shows the sensor bracket, in this case with two different sensor holders. The cylindrical sensor holder (left) is preferred in field applications. Above right lower: the pointer and scale from which the inclination is read. Note the stability is relevant for the noise characterization and not the absolute accuracy.

II. ANALYSIS OF THE SCA103T-D04

The SCA103T, see Fig. 1, is a MEMS based inclinometer sensor. It consists of two accelerometer sensors arranged in diametrically opposing directions yielding the signals,

$$x(t) = g + i(t) + v(t) + n_x(t), \quad (1)$$

$$y(t) = g - i(t) + v(t) + n_y(t) \quad (2)$$

where g is due to gravitation, $i(t)$ is the inclination as a function of time, $v(t)$ is the component due to ambient vibrational noise, $n_x(t)$ and $n_y(t)$ are the individual noise components of each sensor. Defining the differential $d_m(t)$ and sum $s_m(t)$ measurement signals as:

$$d_m(t) = x(t) - y(t) = 2i(t) + n_x(t) - n_y(t), \quad (3)$$

$$s_m(t) = x(t) + y(t) = 2g + 2v(t) + n_x(t) + n_y(t). \quad (4)$$

Consequently, it should be possible to determine both the extraneous vibrations and inclinations from the sensor data. According to the manufacturer the inclination, scaled to mg , can be computed as,

$$i_\phi(t) = \frac{d_m(t) - d_0(t)}{s} \quad (5)$$

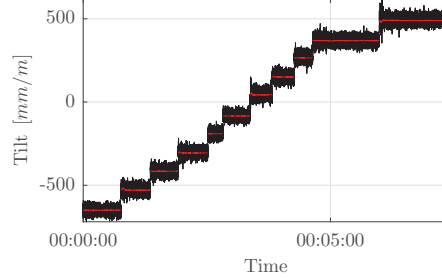


Fig. 4: The signal acquired from the SCA103T sensor after evaluation to *tilt*: (black) signal prior to filtering and (red) after filtering. A total of $n = 183311$ samples are used for the ensuing statistical analysis.

where $d_0(t)$ is the zero measurement and s is the sensitivity of the device. The proposed computations assume that both sensor chains, i.e., sensors, amplifiers and ADCs, have identical gain. As we shall see later this is not the case and leads to a significant loss in sensor performance.

The SCA103T was positioned at $m = 11$ different deflections and a total of $n = 183311$ samples were acquired. The raw data¹ and a filtered result are shown in Fig. 4. The histograms for each of the $m = 11$ segments together with a Gaussian and Cauchy Lorentz distribution approximations are shown in Fig. 5 and the numerical values for the distribution parameters are given in Table I. The χ^2 test for the Cauchy-Lorentz distribution in each of the $n = 11$ segments is an order of magnitude better than the corresponding Gaussian approximation; this is consistent with the visual inspection. These results justify the further and more detailed investigation of the nature of the noise in the system. Furthermore, the data from the segments show that the distribution width γ is independent of the specific inclination x_0 ; consequently, we may conclude that the distribution of the noise is independent of the deflection angle.

A. Histograms and distributions for $x(t)$ and $y(t)$.

Given the fact that there is no correlation between x_0 and γ , it is permissible to concatenate the median free data from each segment to obtain a larger sample size for the determination of the noise characteristics. This has been performed for the signals $x(t)$ and $y(t)$; their respective histograms and PDFs are shown in Fig. 6, while the deviation from a Gaussian CDF is given in Fig. 7. Clearly, the signals $x(t)$ and $y(t)$ are subject to Gaussian noise. It is important to note that $\sigma_x = 17.4$ and $\sigma_y = 19.2$ have different values and indicate that the sensor

¹A brief note on nomenclature: All data presented in the histograms have integer values corresponding to the digital reading of the individual sensor. Consequently, the width of the distributions are all in LSB. PDF refers to probability distribution function and CDF to cumulative distribution function. The notation $p(\alpha)$ refers to the probability of the variable α having a given value. The symbol σ_α refers to the standard deviation of α and γ_α is the half-width at half-maximum value of the Cauchy Lorentz distribution for α .

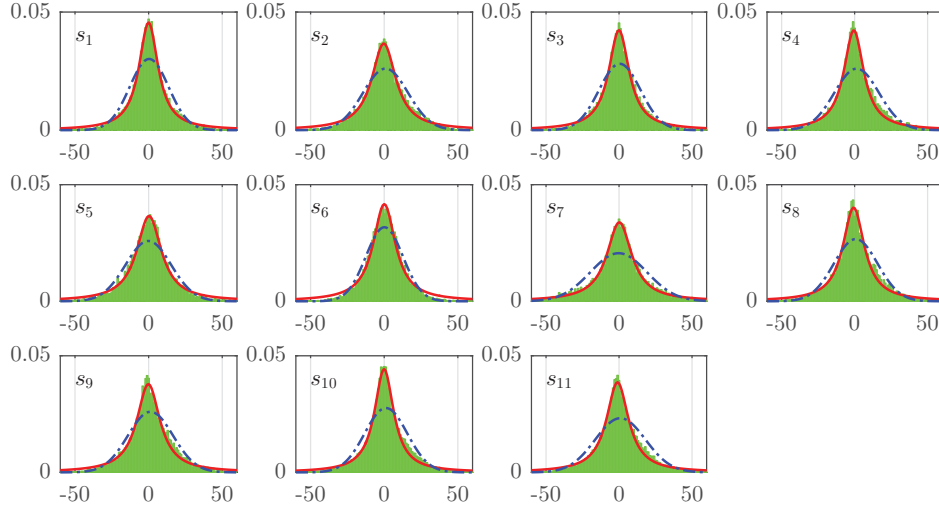


Fig. 5: Histogram for $x - y$ in each of the segments $i \in 1 \dots 11$ denoted by s_i , as shown in Fig. 4. The PDF for the Cauchy and Gaussian are shown in red and blue respectively. The corresponding coefficients for the distributions are given in Table I. The histograms have been centered around the median value of each segment, this simplifies the comparison of the results in each segment.

	n	x_o	γ_{x-y}	I_o	χ_C^2	μ	σ	χ_G^2
S_1	19126	-650	7.89	4.56	1.42	-649.65	13.24	20.21
S_2	14648	-529	10.00	3.67	2.07	-528.10	15.31	12.00
S_3	13812	-416	8.51	4.23	1.76	-415.48	14.20	18.63
S_4	14755	-305	8.27	4.23	2.68	-303.60	15.32	25.73
S_5	7789	-191	10.21	3.66	2.35	-191.20	15.43	11.94
S_6	15155	-84	9.18	4.16	3.11	-83.86	12.57	7.79
S_7	11265	43	10.35	3.39	1.21	42.74	19.35	21.33
S_8	11166	151	8.91	4.02	2.94	151.76	14.96	18.90
S_9	9347	265	9.54	3.79	3.15	266.25	15.35	15.72
S_{10}	32523	368	7.89	4.43	2.24	369.18	14.46	25.01
S_{11}	33725	490	9.17	3.86	2.00	490.90	17.17	24.14

TABLE I: Statistics for each of the $n = 11$ segments, as shown in Fig. 4. Whereby: n is the number of samples in the segment, x_o , γ and I_o (scaled by $\times 100$) are the parameters of the Cauchy-Lorentz distribution, χ_C^2 for its PDF (scaled by $\times 1E4$), μ and σ as the coefficients for the Gaussian and χ_G^2 for the PDF (scaled by $\times 1E4$).

channels may have different gains with respect to ambient vibrational noise.

Now proceeding to the computation of the statistics for $d_m(t)$ and $s_m(t)$: the respective histograms and Gaussian models are shown in Fig. 8. Additionally for the signal $d_m(t)$ a Cauchy-Lorentz PDF has been computed. The fact that $d_m(t)$ has Cauchy-Lorentz distribution implies that the perturbations of $x(t)$ and $y(t)$ must be correlated, since the difference of two Gaussians is also a Gaussian if the signals are not correlated. Furthermore, given $\sigma_x = 17.4$ and $\sigma_y = 19.2$ we would expect $\sigma_{x+y} \approx 25.91$; however, we observe $\sigma_{x+y} \approx 33.35$, obviously ignoring the relative gains of the sensor chains is degrading

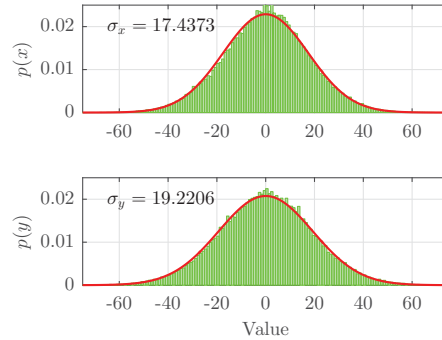


Fig. 6: Histogram of the values $x(t)$ and $y(t)$ with their corresponding Gaussian approximations with respective standard deviations.

the quality of the result.

B. Correlation in the perturbations of $x(t)$ and $y(t)$

To investigate the correlations in the perturbations of $x(t)$ and $y(t)$ a bivariate histogram has been computed, see Fig. 9. The correlation between the signals is clearly visible. The orientation of the dominant axis is not at 45° a further indication that the sensor chains have differing gains. If the gain of the sensor chains were equal the dominant axis would be at 45° .

Singular value decomposition (SVD) is now used to determine the dominant axes and the distributions of the data with respect to these axes. Defining the matrix $D \triangleq [x, y]$ where x

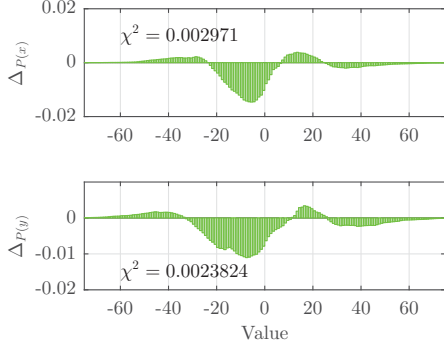


Fig. 7: Deviation of the cumulative distribution functions $P(x)$ and $P(y)$ from their ideal Gaussian models and the respective χ^2 values.

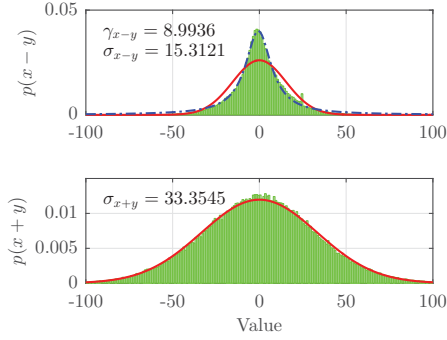


Fig. 8: Histogram of the values $p(x-y)$ and $p(x+y)$ with Gaussian approximations. Additionally for $p(x-y)$ a Cauchy-Lorentz distribution has been approximated.

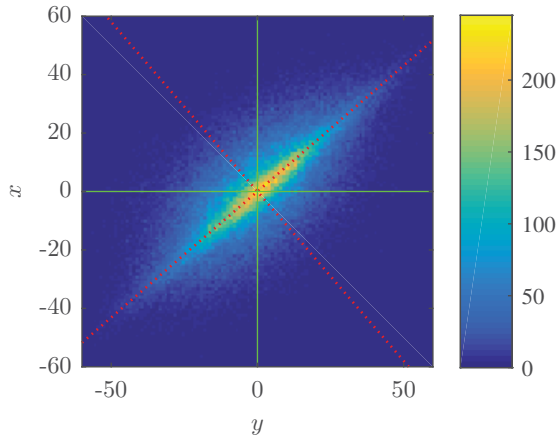


Fig. 9: Bivariate histogram for $x(t)$ and $y(t)$, the color indicates the frequency and is proportional to $p(x,y)$.

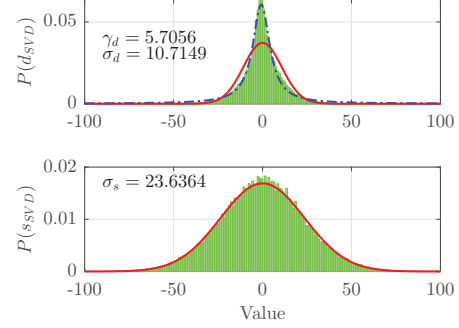


Fig. 10: Histograms for the orthogonalized values d_{SVD} and s_{SVD} together with the Gaussian and Cauchy-Lorentz PDF.

is the column vector of the values of $x(t)$ (similarly for y and $y(t)$). The SVD for a matrix D is defined as,

$$D = USV^T. \quad (6)$$

The matrix V forms an ortho-normal vector basis set for the span $\{D\}$, in the 2D case this corresponds to a rotation matrix from which we can determine the orientation of the major and minor axes. S is a diagonal matrix containing the singular values, i.e., the 2-norm distance of the points in D to the vector basis set. U is the scale free orthogonal projection of D onto V . Consequently, we now obtain,

$$d_{SVD} = U(:,1)S(1,1) \quad (7)$$

$$s_{SVD} = U(:,2)S(2,2) \quad (8)$$

as orthogonalized estimates for d_m and s_m . The matrix V has the values,

$$V = \begin{bmatrix} 0.6530 & 0.7574 \\ 0.7574 & -0.6530 \end{bmatrix}. \quad (9)$$

corresponding to the angle $\phi = 40.77^\circ$ and a relative gain for the sensors of $g_r = 0.86$. The histograms and respective probability distribution functions for d_{SVD} and s_{SVD} are shown in Fig. 10. Note: the observed standard deviation $\sigma_s = 23.64$ now corresponds very closely to the predicted value when the relative gain is taken into account,

$$\sigma = \sqrt{\sigma_x^2 + (g_r \sigma_y)^2} = 24.01. \quad (10)$$

This indicates that the application of SVD has performed the correct orthogonalization of the signals $x(t)$ and $y(t)$. Furthermore, the γ value for the Cauchy-Lorentz distribution has also been reduced. These results indicate that both vibration and inclination can be measured with a better confidence interval when orthogonalization is applied.

The results of computing γ_{x-y} and γ_d after applying orthogonalization to each segment of the data from Fig. 4 are shown in Table II. A mean reduction in γ of $r_\gamma \approx 0.66$ has been achieved.

	γ_{x-y}	γ_d	r
S_1	7.89	5.01	0.63
S_2	10.00	6.54	0.65
S_3	8.51	5.47	0.64
S_4	8.27	5.27	0.64
S_5	10.21	6.97	0.68
S_6	9.18	5.45	0.59
S_7	10.35	6.94	0.67
S_8	8.91	5.86	0.66
S_9	9.54	6.42	0.67
S_{10}	7.89	5.09	0.64
S_{11}	9.17	6.09	0.66

TABLE II: The values of γ_{x-y} and γ_d obtained after applying orthogonalization to each segment of the data shown in Fig. 4. A mean reduction in γ of $r_\gamma = 0.66$ has been achieved.

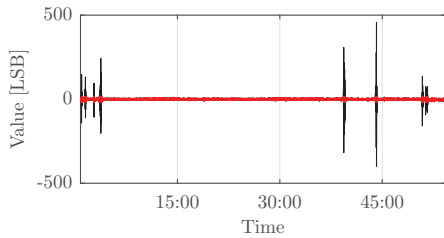


Fig. 11: The full signal acquired with the SCA830 sensor over a period of approximately one hour. The portions marked in red correspond to all values lying within 1% (percentile). This data set contains $n = 400693$ samples. The values are in LSB acquired directly from the ADC. The device as a 16 bit ADC for the range ± 1 g.

III. ANALYSIS OF THE SCA830-D07

The second MEMS based inclinometer considered in this paper is the SCA830 device, see Fig. 2 for the schematic diagram. Unfortunately, with this device the individual sensor signals are not available. Consequently, it is not possible to perform orthogonalization prior to signal processing. Data from this sensor was collected over prolonged periods of time. A sample data set collected over a period of approximately one hour is shown in Fig. 11.

The peaks observed in Fig. 11 caused much concern and lead to extensive testing of the sensor at night during periods when there was little or no activity in the building. We finally came to the conclusion that the observed peaks are due to disturbances within the building, but are unavoidable in any real application. They lead to wide tails in the probability distributions. In particular when working on construction sites, as reported in [12], such perturbations must be considered. They are simply a fact when measuring with these devices in the vicinity of heavy machinery.

The histogram, Gaussian and Cauchy-Lorentz approximations to the signal from the SCA830 device are shown in Fig. 12. This result has been obtained using the complete data sequence. Once again the Cauchy-Lorentz distribution provides a good model for the noise behaviour.

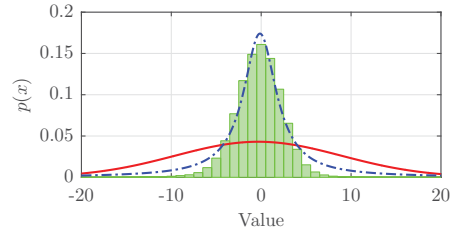


Fig. 12: Histogram of the SCA830 Signal, with the Gaussian and Cauchy-Lorentz models for the complete data set.

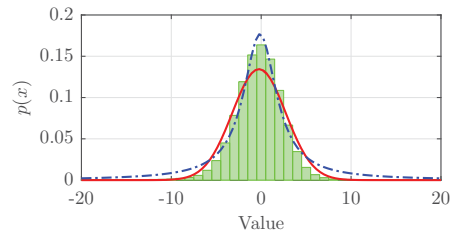


Fig. 13: Histogram of the SCA830 Signal, with the Gaussian and Cauchy-Lorentz models for the centered 98% percentile of the data set.

It is now interesting to consider the statistics for $p(x)$ in the range $p(x) \in 1 \dots 99\%$. This 98 percentile signal is shown in Fig. 11 — it basically corresponds to eliminating local outliers. The histogram, Gaussian and Cauchy-Lorentz approximations for this signal are shown in Fig. 13. This is an interesting result, since in this case the Gaussian provides a possibly satisfactory noise model. Consequently, after eliminating outliers, it would be possible to apply classical signal processing techniques — although they implicitly assume Gaussian perturbations — to obtain satisfactory results. However, for precision measurements modelling the Cauchy-Lorentz nature of the perturbations is unavoidable.

IV. CONCLUSIONS

The large sample sizes used in the analysis presented here ensure reliable statistical results. The first conclusion we can draw from these results is: the noise performance of the vibration and inclination measurements can be improved by applying orthogonalization to the dual-sensors data (when available). The width of the Cauchy-Lorentz distribution is reduced by a factor of $r_\gamma \approx 0.66$.

The influence of ambient vibrational noise and impulses lead to the perturbations having a long-tailed distribution; the distributions are well modelled by a Cauchy-Lorentz distribution. The nature of these distributions **must** be taken into account when designing the signal filtering. Ignoring this fact will lead to significant errors in the signal processing, for example, a sliding average will not reduce the width of a Cauchy-Lorentz distribution. For non-critical applications a good approximation to Gaussian noise is obtained by extracting

```

1 % Median as estimate for x0
2 x0 = median( x );
3 % Compute the histogram counts
4 [cts, edges] = histcounts(x-x0,...
5     'Normalization', 'pdf' );
6 % Determine the corresponding bins
7 bins = edges(1:end-1) + diff( edges ) / 2;
8 % Apply a Savitzky Golay filter
9 spanMin = 5;
10 factor = 50;
11 % use 2% span length but not less than 5
12 span = max( [round(length(cts)/factor),
13     spanMin] );
14 ctsSmooth = smooth(cts, span, 'sgolay', 0);
15 % Use the peak as an estimate for I0
16 I0 = max( ctsSmooth );
17 % Use I0 = 1 / ( pi gamma ) for initial
18 % gamma
19 gamma0 = 1 / ( pi * I0 );
20 % Start values for the optimization
21 cfs = [x0, gamma0, I0];
22 % Anonomous function for the norm
23 cauchyPDFn = @(cfs) norm(cauchyPDF( bins,
24     cfs ) - cts);
25 % Optimize
26 [cfs, ~] = fminsearch(cauchyPDFn, cfs);

```

Listing 1: M-code to perform a non-linear approximation of the PDF of x by a Cauchy-Lorentz distribution.

the center 98% percentile of the signals, prior to filtering. This fundamentally leads to non-linear signal processing methods.

Median filters are one potential non-linear filtering technique that one may consider applying. However, you need to be aware that the median is not always a good estimate for the mode of a samples Cauchy-Lorentz distribution [16], [17].

APPENDIX: CAUCHY-LORENTZ DISTRIBUTION

The three term Cauchy-Lorentz distribution is defined as,

$$p(x) \triangleq I_0 \frac{\gamma^2}{(x-x_0)^2 + \gamma^2}. \quad (11)$$

In all computations performed here: the histogram has been normalized to yield probabilities and a least square approximation of these probabilities by the Cauchy-Lorentz distribution is performed. The trapezoidal rule has been used to implement a discrete approximation to integration. In this manner we ensure that

$$\int_{-\infty}^{\infty} p(x) dx \approx 1. \quad (12)$$

There is a significant volume of literature on estimating the parameters of Cauchy-Lorentz distributions, see for example [16], [17]. To avoid any misunderstandings and to simplify the verification of our results, we have chosen to present the m-code used to perform the nonlinear least squares estimation of the parameters of Cauchy-Laurentz distribution, see Listing 1.

REFERENCES

- [1] X. Hou, X. Yang, and Q. Huang, "Using inclinometers to measure bridge deflection," *Journal of Bridge Engineering*, vol. 10, no. 5, pp. 564–569, 2005. [Online]. Available: <http://link.aip.org/link/?QBE/10/564/1>
- [2] J. Van Cranenbroeck, "Continuous beam deflection monitoring using precise inclinometers," in *FIG Working Week 2007*, Hong Kong, SAR, 13..17 May, 2007.
- [3] G. Machan and V. G. Bennett, "Use of inclinometers for geotechnical instrumentation on transportation projects," *Transportation Research E-Circular*, vol. E-C129, 2008. [Online]. Available: <http://worldcat.org/issn/00978515>
- [4] J. Golser, "Fallbeispiel zur Bauwerksberwachung mittels online Neigungssensoren," in 25. *Cristian Veder Kolloquium*, 2010.
- [5] P. O'Leary, M. Harker, and J. Golser, "Direct discrete variational curve reconstruction from derivatives and its application to track subsidence measurements," in *Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE*, May 2011, pp. 1–6.
- [6] P. O'Leary and M. Harker, "A framework for the evaluation of inclinometer data in the measurement of structures," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1237–1251, May 2012.
- [7] —, "Inverse boundary value problems with uncertain boundary values and their relevance to inclinometer measurements," in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, May 2014, pp. 165–169.
- [8] M. Makwana and G. S. Gupta, "Inclinometer based low-cost biofeedback balanceboard for injury rehabilitation," in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, May 2014, pp. 551–556.
- [9] M. Vágner and P. Benes, "Start-up response improvement for a mems inclinometer," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, May 2015, pp. 1363–1366.
- [10] P. O'Leary, M. Harker, and C. Gugg, "An inverse problem approach to approximating sensor data in cyber physical systems," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, May 2015, pp. 1717–1722.
- [11] D. W. Ha, H. S. Park, S. W. Choi, and Y. Kim, "A wireless mems-based inclinometer sensor node for structural health monitoring," *Sensors*, vol. 13, no. 12, pp. 16 090–16 104, 2013. [Online]. Available: <http://www.mdpi.com/1424-8220/13/12/16090>
- [12] R. Schmidt, P. O'Leary, and M. Harker, "Precision inclinometer measurement system with a wireless gateway," in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, May 2016, pp. 1–6.
- [13] M. Park and Y. Gao, "Error and performance analysis of mems-based inertial sensors with a low-cost gps receiver," *Sensors*, vol. 8, no. 4, pp. 2240–2261, 2008. [Online]. Available: <http://www.mdpi.com/1424-8220/8/4/2240>
- [14] B. Yongqiang, H. Junhui, and Q. Xianghai, "Data processing algorithm of mems inclinometer based on improved Sage-Husa adaptive Kalman filter," in *Control Conference (CCC), 2012 31st Chinese*, July 2012, pp. 3702–3707.
- [15] R. Levy, B. Bourgeteau, J. Guerard, and P. Lavenus, "A high precision quartz crystal mems accelerometer based 2 axis inclinometer," in *2016 Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS (DTIP)*, May 2016, pp. 1–3.
- [16] D. Bloch, "A note on the estimation of the location parameter of the cauchy distribution," *Journal of the American Statistical Association*, vol. 61, no. 315, pp. 852–855, 1966.
- [17] C. B. T. Thomas J. Rothenberg, Franklin M. Fisher, "A note on estimation from a cauchy sample," *Journal of the American Statistical Association*, vol. 59, no. 306, pp. 460–463, 1964.

15 | Force Based Tool Wear Detection using Shannon Entropy and Phase Plane

Originally appeared as:

W. Kollment, P. O'Leary, R. Ritt, and T. Klünsner, "Force based tool wear detection using Shannon entropy and phase plane," in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, IEEE, May 2017, pp. 1–6, ISBN: 978-1-5090-3596-0. DOI: 10.1109/I2MTC.2017.7969765. [Online]. Available: <https://ieeexplore.ieee.org/document/7969765/>

Bib_TE_X:

```
@inproceedings{Kollment2017a,  
  author      = {Kollment, Werner and O'Leary, Paul and Ritt,  
                Roland and Kl{"u"}nsner, Thomas},  
  booktitle   = {2017 IEEE International Instrumentation and  
                Measurement Technology Conference (I2MTC)},  
  doi         = {10.1109/I2MTC.2017.7969765},  
  isbn       = {978-1-5090-3596-0},  
  month      = {may},  
  pages      = {1--6},  
  publisher   = {IEEE},  
  title      = {{Force based Tool Wear detection using Shannon  
                Entropy and Phase Plane}},  
  url        = {http://ieeexplore.ieee.org/document/7969765/},  
  year       = {2017}  
}
```

This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication

Force based Tool Wear Detection using Shannon Entropy and Phase Plane

Werner Kollment*, Paul O'Leary*, Roland Ritt*, Thomas Klünsner†

* Chair of Automation, University of Leoben, Austria, werner.kollment@unileoben.ac.at

† Materials Center Leoben Forschung GmbH, Leoben, Austria, thomas.kluensner@mcl.at

Abstract—In this paper, an approach for the detection of tool wear occurring during a machining process is presented. An instrumented chuck is used to measure the tool forces. The measured signals are partitioned into machining operation and auxiliary movement. Each machining operation is analysed using the statistical central moments: mean, variance, skewness and kurtosis. The machining operations are represented as time histogram for further examination. From the time histogram the concept of segment based entropy is derived. A method for a local computation of the entropy based on probability mapping is presented. Based on the joint probability derived from the phase diagram the method of local joint entropy is developed.

I. INTRODUCTION

In metal machining, such as drilling and milling, worn tools reduce the quality of parts via surface degenerations or deviation from the ideal geometry. This can lead to high cost due to full rejection of the part, expensive refinishing and in the worst case claims from the customers and bad reputation. These costs are avoidable with a reliable tool condition monitoring system. Based on such systems, methods for predicting tool damage can be developed. Furthermore, the influence of different tool and process parameter can be analysed. The results can be used to validate process simulations [1] and also to optimize tool parameters [2], [3].

In the summaries given by Lauro et al [4] and Jantunen et al [5], it can be seen that four measurement types are favoured for tool monitoring: force, vibration, acoustic emission and motor current. Force sensors measure the tools cutting forces but they are difficult to integrate in existing machines. Vibration and acoustic emission sensors are simple to install but they are susceptible for perturbation. Monitoring motor current requires no changes of the machine but it is not sufficient sensitive for every monitoring task. Rizal et al [6] states that the cutting force is the most sensitive indicator of wear.

This paper focuses on the detection of wear and anomalies in the cutting force signals. The forces are measured with an instrumented tool holder equipped with a wireless transmission system. The measured data is statistical analysed. Additionally to the statistics the local information content of the signals are investigated. The aim of the investigation is to detect local changes in the information content caused by an external event such as cutting edge breakage or wear.



Fig. 1: The instrumented chuck is equipped with four strain gauges, enabling the measurement of: bending moments M_x and M_y in orthogonal direction; the torque T and the holding force F_a . In addition, there is a temperature sensor. The chuck is also equipped with a wireless data connection to enable a remote real-time observation of the measured values. The magnification shows the damage of one cutting edge at the end of the life time experiment.

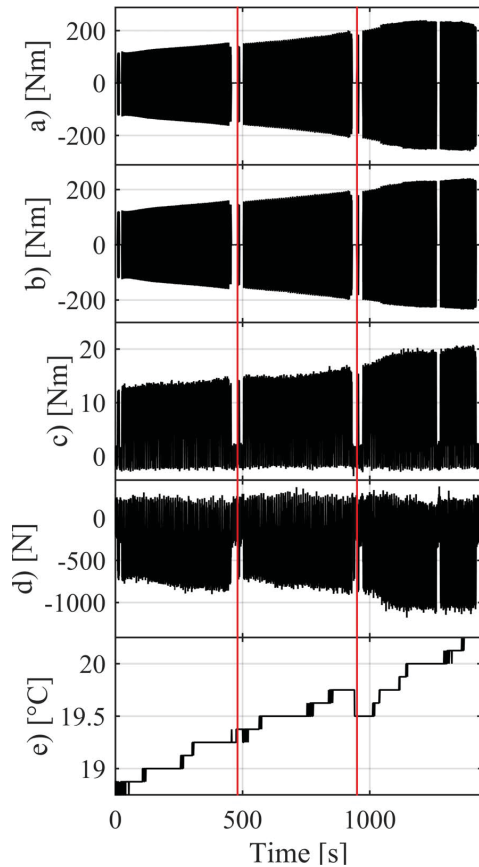


Fig. 2: Signals collected during the end mills life time experiment. The signals are: a) bending moment M_x , b) bending moment M_y , c) torque T , d) axial force F_a , e) temperature at the strain gauge location. The vertical red lines indicate interruptions for the optical measurement of the cutting edge wear.

II. MEASUREMENT SETUP

A direct measurement of the force at the cutting edge is not possible but the reaction force is measurable. These reaction forces are measured with an instrumented chuck¹ with a standardized machine interface and is shown in Fig.1 top. It is instrumented with four full bridge strain gauge circuits. Each strain gauge circuit measures a component of the reaction force. The reaction forces are the axial force F_a , the torque T , the bending moment M_x about the x-axis, the bending moment M_y about the y-axis and the temperature T at the strain gauges location. From these reaction force components, it is possible to reconstruct the cutting force using the tools

¹It is available at pro-micron GmbH & Co. KG under the name SPIKE®.

Tool	Hard metal
Cutting Conditions	Cutter diameter : 10 mm
	Radial depth of cut: 8 mm
	Depth of cut: 10 mm
	Feed per tooth: 0.060 mm/tooth
	Surface cutting speed: 125 m/min
Milling method	climb milling
	Material
	X155CrMoV12
	Tensile strength: 800 N/mm

TABLE I: Milling Parameter used for the life time experiment of the end mill.

diameter and lever arm. Also the bending moment magnitude M can be derived from its components M_x and M_y .

The measured strain signals are converted to digital signal with a 15 bit analog-digital converter at a sampling rate of 1.6 kHz. The digital signal is transmitted wireless over the 2.45 GHz frequency band to a stationary receiver that is connected to a personal computer.

The data, presented in the first part of this paper, are collected from life time experiments for a hard metal end mill with four teeth. During the experiments all process parameters, listed in TABLE I, are kept constant to avoid interference by them. The experiment is interrupted in predefined intervals to measure the wear with an optical system. Fig.2 shows an example of the signals occurring during these experiments. Interruptions for the reference measurements cause a block structure in the signals. The life time experiment has been aborted after the cutting edge damage exceeded the predefined width of 0.4 mm. The bottom of Fig.1 shows the damage of the cutting edge after the termination of the experiment.

The data, presented in the last two sections of the paper, are obtained from a drilling operation. During this operation twelve holes with a diameter of 5 mm and a depth of 25 mm were drilled with a spindle speed of 15000 rpm.

III. SEGMENTATION

Manufacturing processes have a typical signal pattern as shown in Fig.3. A common cause of these patterns are repeating events e.g. the repeated pass of the tool over the workpiece. To analyse such events, a separation of the events is of advantage. It is convenient to use the feature transitions as natural separation points. Therefore a threshold based segmentation algorithm was developed in [7] which extract the indices of the segments rather than the data itself. This avoids redundancy of the data which is important for an efficient handling with large data sets. Additionally the indices allow a simultaneous segmentation of other time synchronous signals. This approach enables the storage of large data set in no-SQL data bases such as HBase [8], while maintaining the segmentation relations in a meta-data structure, whereby the indices form the key row values in HBase.

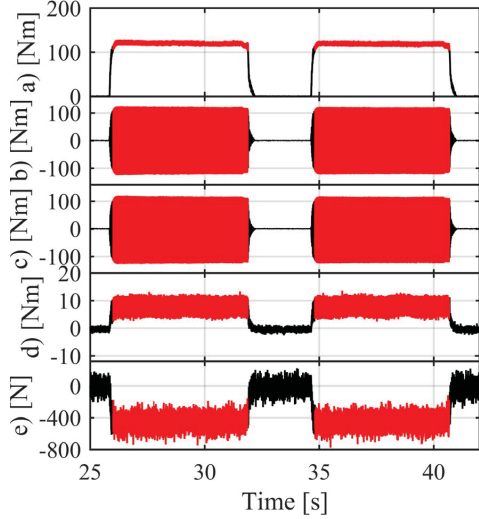


Fig. 3: Magnification of two passes of the miller for following signals: a) bending moment M , b) bending moment M_x , c) bending moment M_y , d) torque T e) axial force F_a . The extracted segments are marked in red.

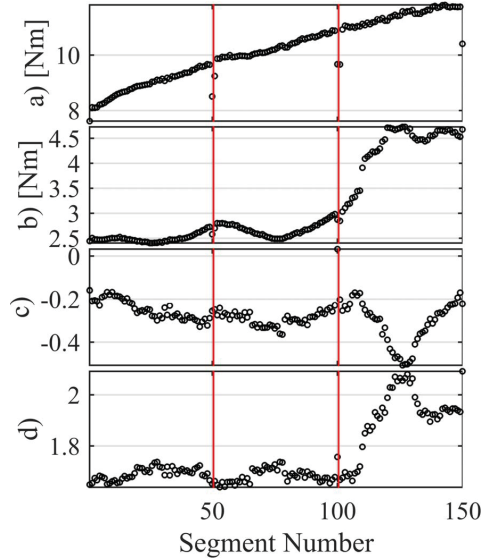


Fig. 4: Statistical central moments a) mean μ_1 , b) variance μ_2 , c) skewness μ_3 and d) kurtosis μ_4 of the segments of the torque signal T shown in Fig.2. The vertical red lines indicate the interruption by the optical wear measurement.

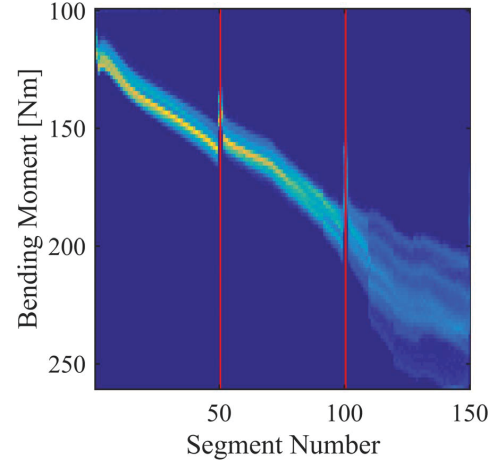


Fig. 5: The time histogram of the bending moment M is a collection of histograms. Increasing wear leads to a significant change in the distribution shape of the data. The vertical red lines indicate the interruption by the optical wear measurement.

IV. STATISTICAL CENTRAL MOMENTS AND TIME HISTOGRAM

A statistical analysis of a segment is possible if its number of data points exceeds the necessary samples size. The data contained in each segment have an individual probability distribution. This distribution can be characterized by its central moments, whereby only the first four orders are of practical relevance. The first and the second central moments μ_1 , μ_2 are the mean and the variance. The third moment μ_3 (see Eqn.1) is the skewness and is a measure for the asymmetry of a probability distribution about the mean μ_1 . Positive values indicate a skew to the left and negative values a skew to the right. The fourth central moment μ_4 (see Eqn.2) is the kurtosis and measures the spikiness of a probability distribution. If a normal distribution is used as reference, its kurtosis value, three, is subtracted from the computed kurtosis and is called excess kurtosis. Skewness and kurtosis are dimensionless.

$$\mu_3 = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N s^3} \quad (1)$$

$$\mu_4 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{N s^4} - 3. \quad (2)$$

Fig.4 shows the central moments of the signal segments. The sudden change in the central moments indicates a cutting edge damage.

Each signal value x_i has an occurrence probability P_i . For each segment of signals these occurrence probabilities are derived from its histogram. The probability for a single histogram bin j is calculated by

$$P_j = \frac{n_j}{\Delta x_j N}, \quad (3)$$

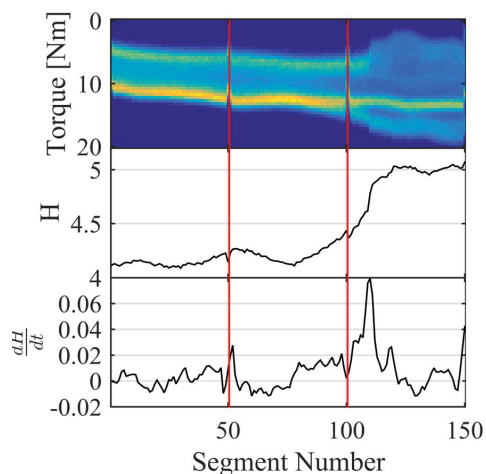


Fig. 6: Top: Time histogram computed from the segments of the torque signal T . The vertical red lines indicate the interruption for the wear measurement. Middle: Shannon entropy H computed from the columns of the time histogram. Bottom: The peaks in the derivative of the Shannon entropy indicate a significant change in the information content.

where n_j and Δx_j are the frequency and the width of the bin j and N is the total number of samples. It is possible to store the histogram of each segment as single column in a matrix i.e. the so called time histogram. This method requires the same bin intervals for all segment histograms and it has to be defined a priori. The time histogram can be used to detect temporal changes and anomalies in the data as shown in Fig.5. Time histograms can be easily presented as images or 3D-plots. The time histogram can be used as a data exploration tool for large data sets of time series such as time frames of a fluid simulation [9]. This method offers an effective way to explore large data sets by a human operator.

V. ENTROPY AND INFORMATION

In "A mathematical theory of communication" [10], Claude Shannon introduced the concept of entropy H , see Eqn.4, as a measure for the information content. He interpreted the Shannon entropy H for the basis $B = 2$ as minimum number of bits to which a symbolic sequence can be compressed. Due to the lack of systematic pattern random sequences require more bits than systematic ones.

$$H = - \sum_{i=1}^N P_i \log_B(P_i) \quad (4)$$

The Shannon entropy was developed for discrete random variables but measurements are continuous random variables which would require a integral formula for the entropy. Fortunately, an approximation of the continuous distribution is possible by a histogram to overcome this problem.

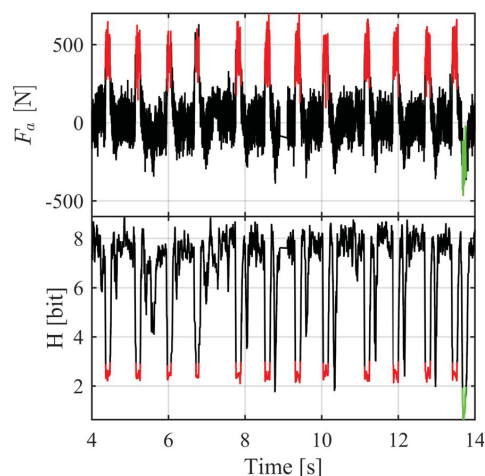


Fig. 7: Top: Axial force recorded from a drilling process. Bottom: Local entropy computed of the signal. The regions with low entropy, marked in red, match with the high levels of the feature. The detected outlier is marked in green.

A. Segment based Entropy

Additional to the central moments, the Shannon entropy H can be used to characterize discrete probability distributions i.e. histograms. Therefore, the probability of the bin P_j is treated as an occurrence probability P_i to compute the Shannon entropy according Eqn.4. This method can be applied to time histograms to compute an entropy sequence. Changes in this sequence indicate changes in the information content. These changes can be localized by the first derivative of the sequence. In physical systems a change in the information content is correlated with the change of its physical behaviour. In Fig.6 middle, the entropy is applied to the signal of a milling process. The entropy change correlates with the tool damage and can be clearly identified by the maxima of the entropies derivative shown in Fig.6 bottom. The entropy H , shown in Fig.6, and the standard deviation $\sigma = \sqrt{\mu_2}$, shown in Fig.4 b, are linked via their probability distribution. For a Gaussian distribution this link is given for a basis $B = e$ by Eqn.5.

$$H = \log_B(\sqrt{2\pi e}) + \log_B(\sigma) \quad (5)$$

The probability P_i , necessary to calculate the entropy, can be calculated from different sources than the occurrence frequency of the signal values. Hence, every property, that can be represented as probability P_i , can be used for the computation of the Shannon entropy. Examples for these sources are: the spectra of a short time fourier transform [11], wavelet energy [12], symbols [10] and permutations of number sequences [13].

B. Local Entropy

Every data point x_i is a sample of the signals probability distribution and has an individual occurrence probability P_i . Un-

fortunately, the exact analytic distribution is unknown but it is possible to approximate this distribution with a histogram. The probability of the histogram bins can be mapped to the data points inside the bins range. The result of this mapping is a vector of occurrence probabilities $\mathbf{p}^T = [P_1, \dots, P_N]$ of every data point. The probability vector \mathbf{p} is used to compute a vector of summands $\mathbf{h}^T = [-P_1 \log_B(P_1), \dots, -P_N \log_B(P_N)]$ of the Shannon entropy using Eqn.6²

$$\mathbf{h} = -\mathbf{p} \circ \log_B(\mathbf{p}). \quad (6)$$

The Shannon entropy for any region of interest can be computed by multiplying the summand vector \mathbf{h} with a windowing vector \mathbf{w} . Alternative the convolution of the summand vector \mathbf{h} with window $\mathbf{w}^T = [1, \dots, 1]$ can be used to compute the local entropy vector \mathbf{H} , see Eqn.7. Every entry of \mathbf{H} represent the entropy .i.e. information content of all data points inside the range of sliding windows \mathbf{w} .

$$\mathbf{H} = \mathbf{h} * \mathbf{w} \quad (7)$$

Deterministic signal parts show lower entropy values than stochastic parts. This can be used to extract the deterministic parts of a signal.

In Fig.7 the axial force acting on a drill and the derived local entropy is presented. The signal shows a sequence of individual drilling operations which can be identified by thresholding its local entropy. By defining a certain entropy limit, outlier exceeding this limit can be located.

VI. PHASE DIAGRAM AND JOINT ENTROPY

Differential equations are used to model physical system eg. dynamic systems. Differential equation of higher order are typically represented in the phase space, whereby every derivative corresponds to a phase space variable. This implies that a measured phase space variable is correlated with its derivative. Therefore, the derivative also contains information about the system and should be considered in a signal analysis.

The phase diagram is a method to analyse the phase space of a system, whereby, the phase diagram describes a two dimensional subspace of the phase space. Systems described by differential equation of order two, such as dynamic systems, the complete phase space coincides with the phase diagram. Hence, the phase diagram is used to analyze motions [14] or detect changes in the motion caused by changes of the system eg. defects [15].

The phase diagram is computed using the Savitzky-Golay [16] smoothed signal and its numerically computed derivative. Fig.8 shows the phase diagram of the axial force of the same drilling process as shown in Fig.7.

From the points in the phase diagram a two dimensional histogram can be computed as shown in the bottom of Fig.8. This histogram provides the joint probability of the data points and its derivatives. This joint probability can be mapped to the data points and results in a joint probability vector \mathbf{p}_{xy} . This joint probability vector \mathbf{p}_{xy} can be used to compute the joint

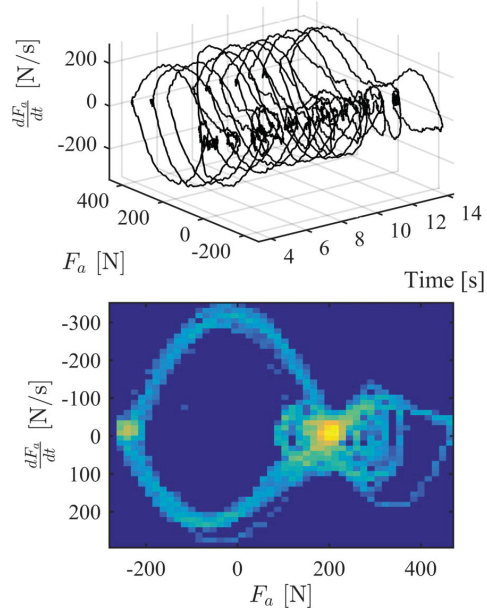


Fig. 8: Top: Phase diagram computed from Savitzky-Golay smoothed axial force and its derivative with a support length $l_s = 101$. Bottom: Two dimensional histogram computed from the phase diagram.

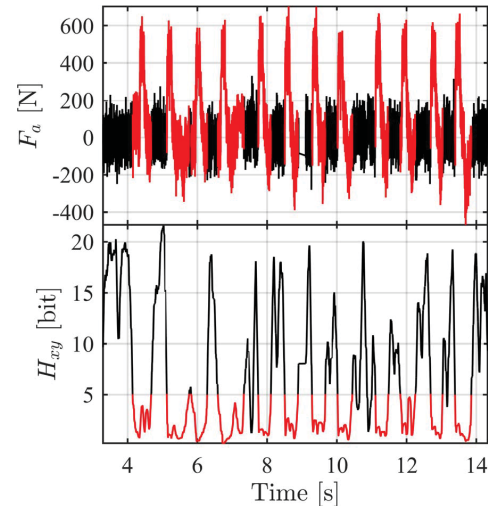


Fig. 9: Top: Axial force recorded from a drilling process. Bottom: Joint entropy computed from the phase space histogram. The region with lower joint entropy matches the drilling features marked in red.

²The symbol \circ represents the Hadamard product

local entropy vector H_{xy} using the methods described in the previous section.

In the described method the derivative is used as a second source of information. Thus, also information which is not covered by the local entropy is contained in joint local entropy.

Applying this method to the same axial force signal as presented in Fig.7 allows an improved identification of the individual drilling operations as shown in Fig.9.

Beside of the presented method also other entropy definitions, such as the Permutation entropy [13], use the phase space to detect anomalies of dynamical systems.

VII. CONCLUSION

Statistical central moments are a suitable tool to identify anomalous region of the data. The time histogram provides a visual summary of the data and takes advantage of the humans natural ability of pattern recognition. Also, it provides deeper insight in the nature of the data than the central moments. Anomalous regions show a different amount of information than the rest. They can be identified using the Shannon entropy as a measure for information. Entropy is a transformation of the probability and the source of the probability, used for its computation, decides about meaning of the detected information. A probability mapping enables a detection of local region with different information content. Linear transformations are able to uncover information parts that are impossible to capture in the original signal. In combination with the signal information an accurate detection of events is possible.

VIII. ACKNOWLEDGMENT

Financial support by the Austrian Federal Government (in particular from BMVIT) represented by Österreichische Forschungsförderungsgesellschaft GmbH (FFG) and the Styrian and the Tyrolean Provincial Government, represented by Steirische Wirtschaftsförderungsgesellschaft (SFG) GmbH and Standortagentur Tirol, within the framework of the COMET Funding Programme is gratefully acknowledged.

REFERENCES

- [1] S. Eck, H.-P. Gänser, S. Marsoner, and W. Ecker, "Error analysis for finite element simulation of orthogonal cutting and its validation via quick stop experiments," *Machining Science and Technology*, vol. 19, no. 3, pp. 460–478, 2015.
- [2] T. Tepperneegg, T. Klünsner, P. Angerer, C. Tritremmel, C. Czettl, J. Keckes, R. Ebner, and R. Pippan, "Evolution of residual stress and damage in coated hard metal milling inserts over the complete tool life," *International Journal of Refractory Metals and Hard Materials*, vol. 47, pp. 80–85, 2014.
- [3] I. Krajinović, W. Daves, M. Tkadletz, T. Tepperneegg, T. Klünsner, N. Schalk, C. Mitterer, C. Tritremmel, W. Ecker, and C. Czettl, "Finite element study of the influence of hard coatings on hard metal tool loading during milling," *Surface and Coatings Technology*, vol. 304, pp. 134–141, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0257897216305333>
- [4] C. H. Lauro, L. C. Brandão, D. Baldo, R. A. Reis, and J. P. Davim, "Monitoring and processing signal applied in machining processes—a review," *Measurement*, vol. 58, pp. 73–86, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0263224114003546>
- [5] E. Jantunen, "A summary of methods applied to tool condition monitoring in drilling," *International Journal of Machine Tools and Manufacture*, vol. 42, no. 9, pp. 997–1010, 2002.
- [6] M. Rizal, J. A. Ghani, M. Z. Nuawi, and C. H. C. Haron, "A review of sensor system and application in milling process for tool condition monitoring," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 7, no. 10, pp. 2083–2097, 2014. [Online]. Available: <http://www.airitilibrary.com/Publication/alDetailedMesh?docid=20407467-201403-201506300028-201506300028-2083-2097>
- [7] W. Kollment, P. O'Leary, M. Harker, U. Osberger, and S. Eck, "Towards condition monitoring of railway points: Instrumentation, measurement and signal processing," in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. IEEE, 2016, pp. 1–6.
- [8] R. C. Taylor, Ed., *An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics*, vol. 11 Suppl 12, 2010.
- [9] R. Kosara, F. Bendix, and H. Hauser, "Time histograms for large, time-dependent data," 2004.
- [10] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
- [11] H. Misra, S. Ikbali, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust asr," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 17–21 May 2004, pp. 1–193–6.
- [12] W.-X. Ren and Z.-S. Sun, "Structural damage identification by using wavelet entropy," *Engineering Structures*, vol. 30, no. 10, pp. 2840–2849, 2008.
- [13] Y. Cao, W.-W. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, Eds., *Detecting dynamical changes in time series using the permutation entropy*, vol. 70, 2004.
- [14] L. W. Campbell and A. F. Bobick, "Recognition of human body motion using phase space constraints," in *IEEE International Conference on Computer Vision*, 20–23 June 1995, pp. 624–630.
- [15] W. J. Wang, J. Chen, X. K. Wu, and Z. T. Wu, "The application of some non-linear methods in rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 15, no. 4, pp. 697–705, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0888327000913165>
- [16] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

16 | Real-Time-Data Analytics in Raw Materials Handling

Originally appeared as¹:

C. J. Rothschedl, R. Ritt, P. O’Leary, M. Harker, M. Habacher, and M. Brandner, “Real-Time-Data Analytics in Raw Materials Handling,” in *Proceedings of Real-Time Mining, International Raw Materials Extraction Innovation Conference*, T. van Gerwe and D. Höbelbarth, Eds., Amsterdam: Prof. Dr.-Ing. Jörg Benndorf, 2017, pp. 144–153, ISBN: 9783938390412. arXiv: 1802.00625. [Online]. Available: <http://arxiv.org/abs/1802.00625>

Bib_TE_X:

```
@inproceedings{Rothschedl2017,
  address    = {Amsterdam},
  author     = {Rothschedl, Christopher Josef and Ritt, Roland
               and O'Leary, Paul and Harker, Matthew and Habacher,
               Michael and Brandner, Michael},
  booktitle  = {Proceedings of Real-Time Mining, International
               Raw Materials Extraction Innovation Conference},
  editor     = {van Gerwe, Thom and H{"o"}{"s"}elbarth, Diana},
  eprint     = {1802.00625},
  isbn      = {9783938390412},
  pages     = {144--153},
  publisher  = {Prof. Dr.-Ing. J{"o"}rg Benndorf},
  title     = {{Real-Time-Data Analytics in Raw Materials
               Handling}},
  url       = {http://arxiv.org/abs/1802.00625},
  year      = {2017}
}
```

¹A preprint of this paper is available on arXiv: <http://arxiv.org/abs/1802.00625>

Real-Time-Data Analytics in Raw Materials Handling

Christopher Rothschedl, Roland Ritt, Paul O'Leary,
Matthew Harker, Michael Habacher and Michael Brandner

Chair of Automation, University of Leoben, Leoben, Austria.
(e-mail: roland.ritt@unileoben.ac.at).

1 Introduction

This paper proposes a system for the ingestion and analysis of real-time sensor and actor data of bulk materials handling plants and machinery. It references issues that concern mining sensor data in cyber physical systems (CPS) as addressed in O'Leary et al. [2015].

The advance of cyber physical systems has created a significant change in the architecture of sensor and actor data. It affects the complexity of the observed systems in general, the number of signals being processed, the spatial distribution of the signal sources on a machine or plant and the global availability of the data. There are different definitions for what constitutes cyber physical systems Baheti and Gill [2011], Geisberger and Broy [2012], IOSB [2013], Lee [2008], NIST [2012], Park et al. [2012], Spath et al. [2013a,b], Tabuada [2006]: the most succinct and pertinent to the work shown in this paper is the definition given by the IEEE Baheti and Gill [2011] and ACM¹:

A CPS is a system with a coupling of the cyber aspects of computing and communications with the physical aspects of dynamics and engineering that *must abide by the laws of physics*. This includes sensor networks, real-time and hybrid systems.

Results computed from sensor and actor data **must** obey the equations used for modelling the physics of the observed system — this fundamentally poses an *inverse problem*. Such problems are not covered sufficiently by literature addressing *mining of sensor data*, see for example Esling and Agon [2012], Fuchs et al. [2010], Keogh and Kasetty [2003], Last et al. [2004]. Even available standard books, such as Aggarwal [2013] on mining sensor data, do not discuss the special nature of sensor data. Typically, present approaches of mining data rely on correlation as being a sole, reliable measure for significance. It is not taken into account that the inverse solutions to the model-describing equations are required to establish a semantic link between a sensor observation and its precedent cause. Without this link — without *causality* — there can be no physics based knowledge discovery.

¹ ACM/IEEE International Conference on Cyber-Physical Systems (ICCPs) (iccps.acm.org)

The underlying data analytics problem can be described generally by the following statements:

1. The momentum of what is called Industry 4.0 promotes an increasing amount and availability of data. A suitable data ingestion system becomes necessary to acquire real-time sensor and actor data on a global scale. The fundamental concept on how to acquire, transport, ingest, and provide data needs to be sufficiently secure and adaptable enough to accommodate data of mining machines that may be located in remote areas.
2. Mathematical tasks are required to apply data analytics to industrial data sets, such as the solution of inverse problems and optimal-control-type problems.
3. Complex systems are modelled mathematically by following principles gained from modelling simple engineering systems, e.g., a vibrating string or a vibrating beam. These can be modelled using differential equations, ordinary and partial. More sophisticated mathematical models will be required to conquer the expanding complexity of modern mechatronic systems.
4. Data analytics will determine the particular causes to specific behaviour witnessed by sensor and actor data. Inverse problems are fundamental to accomplish such tasks. Additional metadata is required to accurately interpret the results of inverse models, as inverse problems do not have unique solutions per definition.
5. Extracting knowledge from data lies beyond simple information extraction. A more profound view on the philosophy of science points towards the necessity of assigning semantic information to data channels to establish such investigations. The metaphorical parallels between machine behaviour and natural language provide a form of knowledge extraction. It can be shown that machines have their own specific polysyllabic language. Once identified, it can be efficiently queried for symbolic patterns of normal or anomalous behaviour.

2 System Premiss

As an extension of Ackoff's work (Ackoff [1989]), Embrechts (Embrechts et al. [2005]) proposes the data mining pyramid consisting of the terms *data*, *information*, *knowledge*, *understanding* and *wisdom*. Embrechts does not provide any definitions for these terms, Ackoff offers intuitive but rather nebulous definitions; both do not provide a scientific basis for mining sensor data. Based on the integral idea we propose the fundamental concept behind the data analytics in Fig. 1.

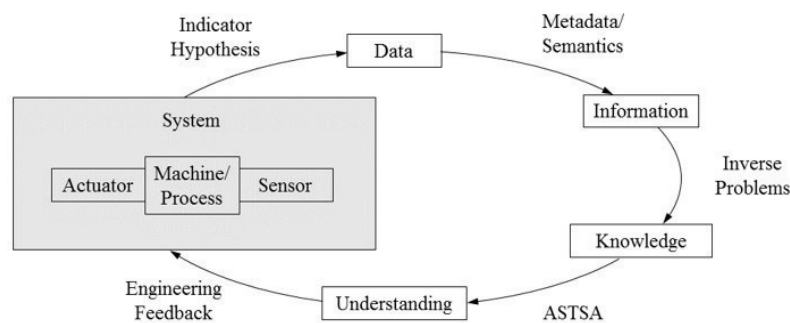


Fig. 1: The process behind the data analysis system.

The presented hierarchy illustrates how the questions of processing large data sets can be approached in a coherent and structured manner. The fundamental relationships of this premiss are:

1. A suitable indicator hypothesis builds the basis for the collection of data. If a specific sensor is chosen, an implicit indicator hypothesis has been selected as well, i.e., a temperature sensor defines that temperature is of relevance for the task.
2. Once acquired, data is only present as a simple stream of numbers; *metadata* adds meaning to the data. Beyond that, *context* is required to establish *significance*: a temperature value can have entirely contrasting significances for measurements of two different sources.
3. System models and the solution of the corresponding inverse problems are required to establish a causal link between measurement data and its possible cause. In general, there are no unique solutions to inverse problems.
4. Hence, a-priori knowledge is necessary to find the desired solution. These results of the inverse problems (the causes) constitute *knowledge*.
5. Effects of *human-machine interaction* must be considered to gain *understanding* of the whole system behaviour. Our approach, *Advanced Symbolic Time Series Analysis (ASTSA)*, is based on the emergence of language as it is modelled by the philosophy of phenomenology. The basic principle consists of symbols that are assigned to actions — *verbs*. The symbols for states are *nouns*. *Adverbs* and *adjectives* are used to predicate the verbs and nouns. *Punctuation* represents different lengths of pauses. Following such a segmentation, the time series is automatically converted into a sequence of symbols, enabling symbolic querying.
6. The whole process serves the understanding of what was originally only a stream of numbers. *Engineering feedback* can be derived from understanding the system response behaviour to certain loads and circumstances. Existing systems can be optimised and future revisions benefit from this as well.

3 Data Ingestion

A versatile data handling system is necessary to conquer large sets of time series data in a structured and efficient manner. Before such a system is able to provide any data, it has to ingest data following a specific workflow. In the course of the ingestion process, data is collected, quality-checked, and merged with corresponding metadata before it is prepared to fit a consistent data model. Sensor values are handled in the same way as derived measurements, i.e., the force of a hydraulic cylinder calculated from its dimensions (metadata) and its pressure values (time series from sensors).

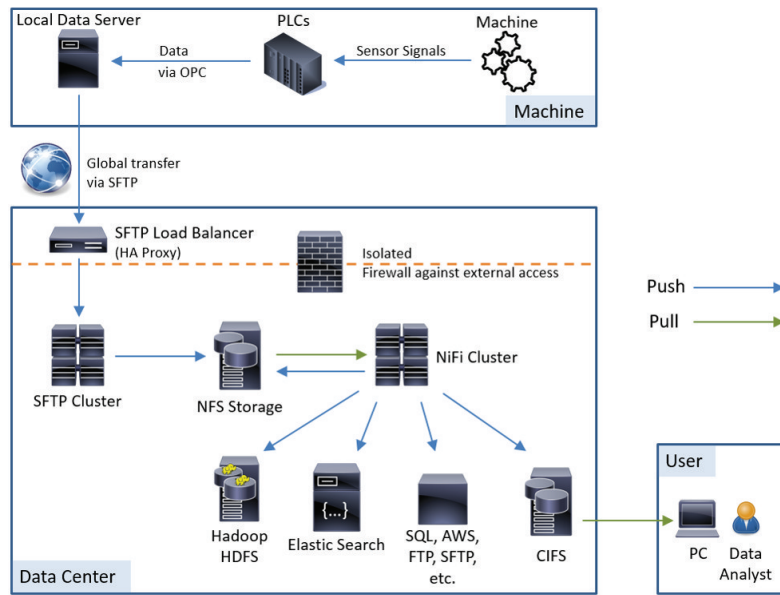


Fig. 2: This illustration shows the main processes of data ingestion. The top section corresponds to the machine or plant on which data is being collected, while the bottom part represents the data center located at a different location. The data is provided in several formats after it has been ingested.

The concept describing the data ingestion process is illustrated in Fig. 2. Data of a machine's sensors is collected from its main programmable logic controller (PLC) and stored on a local data server before it is transported via a secured connection to the data center. After passing quality control, the data is stored permanently according to the data model and specified data manipulation workflows can be triggered on the cluster. Ultimately, the data is made available to consumers (data analysts, report recipients, domain experts, etc.) in different formats: this ensures that all users are independent in their choice of working environment.

The data is stored as a contiguous data stream as a result of the data ingestion process, see Fig. 3. The data input can be split, e.g., as daily exports of a buffering database running on the local data server at the machine's location. The data of all packets are merged to a contiguous, multi-channel stream of time series. When a user requests data from the system, they have the experience of querying the machine directly and in real-time. This opens the door to evaluations spanning time ranges varying from days to months and years. Furthermore, time ranges fitting a machine's operation characteristics can be queried, such as the time for loading a vessel in the case of analysing a ship loader. This permits a complete differentiation between input and output segmentation.

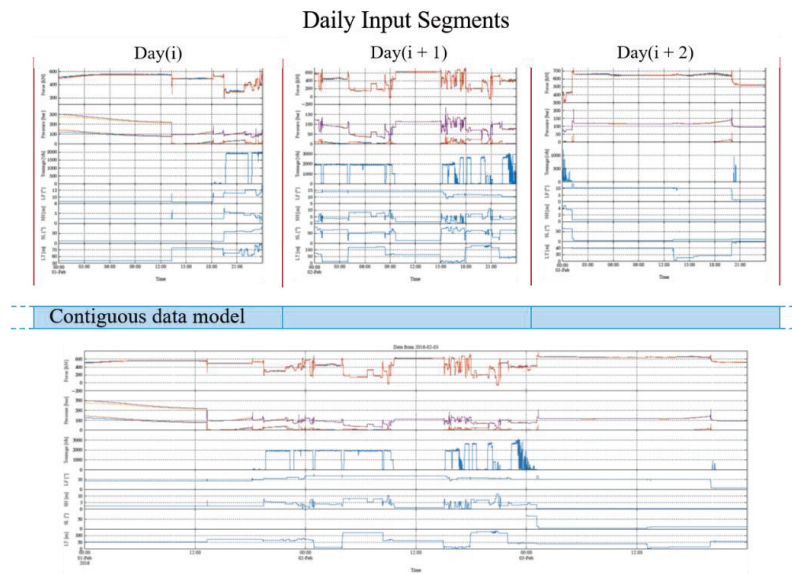


Fig. 3: Three single days of data are assembled to a contiguous data stream. The illustrated contiguous section corresponds to the time portion a ship loader needs to load a vessel: this enables evaluations based on time ranges that are significant to particular fields of interest.

4 Systems Currently Being Monitored

Four mining machines that are currently being monitored using the approach presented in this paper are shown in Fig. 4. Data of these systems is collected constantly with a sampling interval of 1s. Typically, 50 to 850 sensor signals are collected, depending on the complexity of the monitored system.

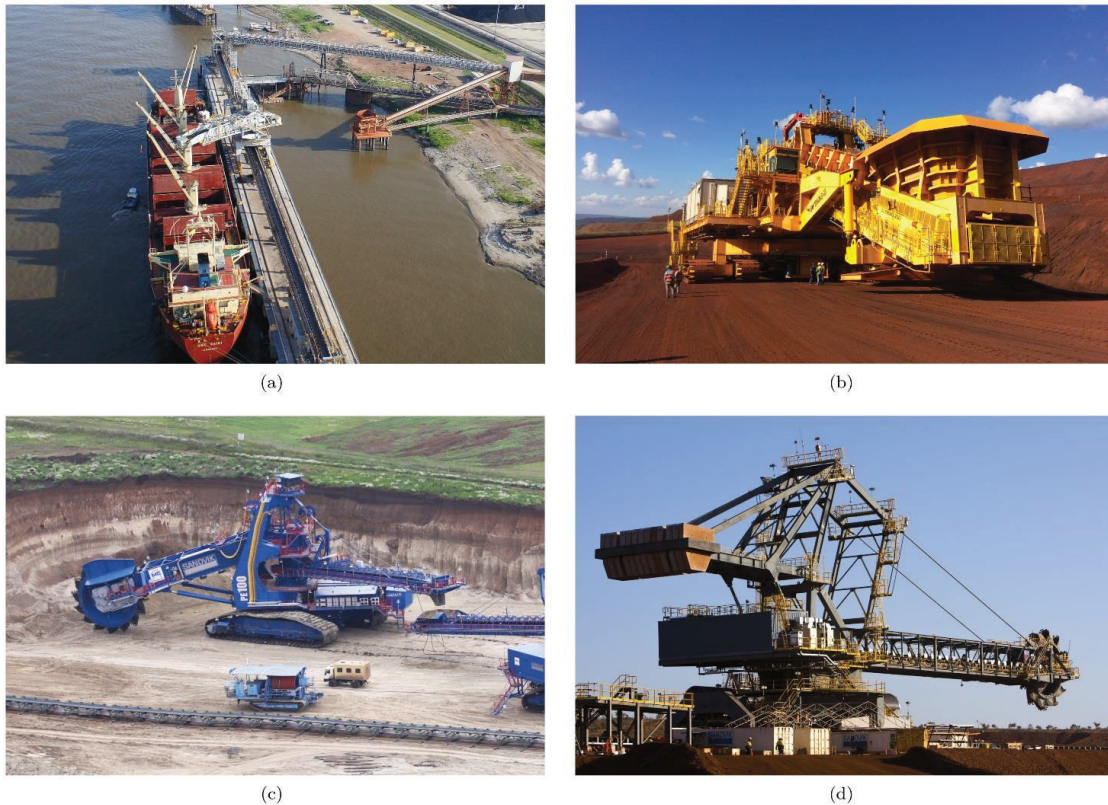


Fig. 4: Examples of four systems that are currently being monitored using the described approach: a) ship loader, b) mobile sizing rig, c) bucket-wheel excavator, d) bucket-wheel reclaimer. The sensor channels of these systems are monitored with a sampling interval of 1s. (Sources: (a) – http://www.flickrriver.com/photos/impalaterminals_images/17557941415/, retrieved on 2016-02-08; (b), (c), (d) – Courtesy of Sandvik.)

5 Exemplary Data Evaluations

The collection and analysis of data can be used for many different aspects of evaluating a machine during its life-cycle:

Condition Monitoring: Undoubtedly, data analytics can be used to address questions regarding condition monitoring or preventative maintenance, see Rothschedl [2016]. However, in this work we focus on issues that have received less attention in literature, e.g., incident analysis.

Commissioning: If data is already collected during the commissioning phase of a machine, analysing it can support shortening the time needed for this phase. Controlled tests can be verified with manageable effort and unexpected response behaviour to specific load scenarios can be detected. On

several occasions, it was possible to identify sensors that delivered erroneous values for only a few samples a day. Judging from the nature of such error patterns, it would not be possible for a commissioning engineer to spot these defective sensors without such a system.

Fleet management: Insights gained from analysing one machine can support understanding the behaviour of other machines of similar design. For example, two identical bucket-wheel excavators were monitored which are operated in the same mine, handling the same type of material. The characteristics of both machines matched in many aspects. In contrast, two similar ship loaders exhibited behaviour that was significantly different. This raises the question whether these machines fulfil the conditions required to be ergonomic systems.

Automatic Operations Recognition: With ASTSA, several data channels can be combined to define machine states. Sequences of these states refer to corresponding operation modes which can be used to characterise how a machine is being controlled. These sequences support the identification of inappropriate operations that may lead to damages or to missing performance goals.

Incident Analysis: Incidents with equipment in mining environments bear serious financial and legal issues. Unplanned maintenance and repair work in such environments and locations quickly reach immense financial dimensions, also because associated materials handling processes are interrupted, provoking serious follow-up costs. Liability for injury and damages are the main concerns from the legal point of view. The analysis of real-time operational data prior to incidents supports the determination of the possible causes for their occurrences and, hence, can provide more certainty to the financial and legal claims. Although this form of analysis can shed light on the clarification of far-reaching issues, this topic has been rarely mentioned in literature. It is evident that incident analysis plays a major role when working with mining machines.

Logistics Optimisation: The analysis of long-term time series allows evaluations based on aggregated data: the distribution of conveyed material over the full slewing range of a machine over a long period of time can support identifying unevenly distributed component utilisation. Such problems can often be avoided or mitigated if the logistics of a machine are adapted.

Two exemplary evaluations are presented:

5.1 Incident Analysis

The figures below (Fig. 5 and Fig. 6) show the results of performing incident analysis for a bucket-wheel excavator. The analysis shows a large number of events distributed over time and conspicuous times during which no events occurred: this is with most certainty operator-dependent behaviour of the system as a whole.

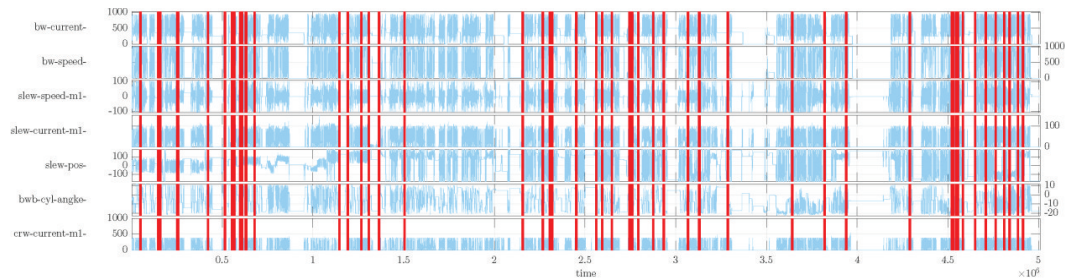


Fig. 5: This example of incident analysis shows data for a time period of two months, acquired with a sampling time of 1s. Each vertical line corresponds to an event; 63 events were found in total by using Advanced Symbolic Time Series Analysis (ASTSA). Every event corresponds to an inappropriate operation of the machine: the data can be zoomed in on automatically for every single event to perform local analysis, i.e., in the seconds and minutes right before the occurrence of the event (see Fig. 6).

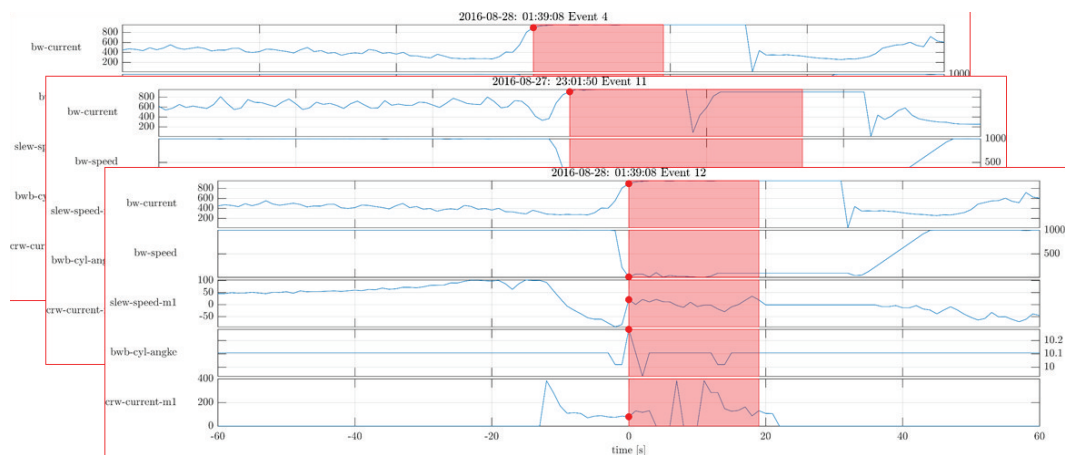


Fig. 6: Plots of the identified events with 1s resolution for three of the 63 events reported in Fig. 5.

5.2 Long-Term Logistics Optimisation

The data shown in Fig. 7 is the polar histogram of loading on the slew bearing of a bucket-wheel reclaimer. The data has been aggregated with $t_s=1s$ over an observation period of one year. Interestingly, the overloading in one quadrant is not visible on a daily basis. The higher loading, evident from aggregated long-term data in the figure, has significant consequences on the life span of the bearing.

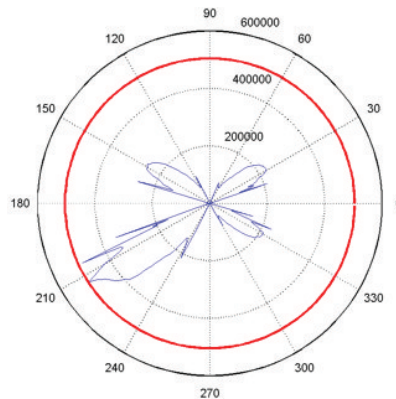


Fig. 7: Polar histogram of loading on the slew bearing of a bucket-wheel reclaimer. The data has been aggregated with a sampling time of 1s over an observation period of one year.

6 Conclusions

The collection of very large real-time data series from plant and machinery is highly relevant in a mining context. A strongly structured approach is required, if the best use is to be made of the data. The results are relevant for both, machine constructors and also their operators. It is significantly more than just preventative maintenance.

REFERENCES

- Ackoff, R.L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3–9.
- Aggarwal, C.C. (ed.) (2013). *Managing and Mining Sensor Data*. Springer.
- Baheti, R. and Gill, H. (2011). Cyber-physical systems. *The Impact of Control Technology*, 161–166.
- Embrechts, M., Szymanski, B., and Sternickel, K. (2005). *Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing*. John Wiley and Sons, NewYork.
- Esling, P. and Agon, C. (2012). Time-series data mining. *ACM Comput. Surv.*, 45(1), 12:1–12:34. doi:10.1145/2379776.2379788. URL <http://doi.acm.org/10.1145/2379776.2379788>.
- Fuchs, E., Gruber, T., Pree, H., and Sick, B. (2010). Temporal data mining using shape space representations of time series. *Neurocomputing*, 74(13), 379 – 393. doi:<http://dx.doi.org/10.1016/j.neucom.2010.03.022>. URL <http://www.sciencedirect.com/science/article/pii/S0925231210002237>. *Artificial Brains*.
- Geisberger, E. and Broy, M. (2012). *agendaCPS: Integrierte Forschungsagenda Cyber-Physical Systems*, volume 1. Springer.
- IOSB, F. (2013). *Industry 4.0 information technology is the key element in the factory of the future*. Press Information.

Keogh, E. and Kasetty, S. (2003). On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Min. Knowl. Discov.*, 7(4), 349–371. doi:10.1023/A:1024988512476. URL <http://dx.doi.org/10.1023/A:1024988512476>.

Last, M., Kandel, A., and Bunke, H. (2004). *Data Mining in Time Series Databases*. Series in machine perception and artificial intelligence. World Scientific. URL <http://books.google.at/books?id=f38wqKjyBm4C>.

Lee, E.A. (2008). Cyber physical systems: Design challenges. In *Object Oriented Real-Time Distributed Computing (ISORC)*, 2008 11th IEEE International Symposium on, 363–369. IEEE.

NIST (2012). *Cyber-physical systems: Situation analysis of current trends, technologies, and challenges*. Technical report, National Institute of Standards and Technology (NIST). URL www.nist.gov.

O’Leary, P., Harker, M., and Gugg, C. (2015). A position paper on: Sensor-data analytics in cyber physical systems, from Husserl to data mining. In *SensorNets 2015*, Le Cresout, France.

Park, K.J., Zheng, R., and Liu, X. (2012). Cyber-physical systems: Milestones and research challenges. *Computer Communications*, 36(1), 1–7.

Rothschedl, C.J. (2016). *Condition Monitoring of Large-Scale Slew Bearings in Bucket-Wheel Boom-Type Reclaimers*. Diploma Thesis, University of Leoben.

Spath, D., Gerlach, S., Hämmerle, M., Schlund, S., and Strölin, T. (2013a). Cyber-physical system for self-organised and flexible labour utilisation. *Personnel*, 50, 22.

Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., and Schlund, S. (2013b). *Produktionsarbeit der Zukunft-Industrie 4.0*. Fraunhofer IAO Stuttgart.

Tabuada, P. (2006). Cyber-physical systems: Position paper. In *NSF Workshop on Cyber-Physical Systems*.

17 | Condition Monitoring of Hydraulics in Heavy Plant and Machinery

Originally appeared as: S. F. Nussdorfer, R. Ritt, and C. J. Rothschedl, “Condition Monitoring of Hydraulics in Heavy Plant and Machinery,” in *6th International Congress on Automation in Mining*, V. Babarovich, D. Haro, M. Gajardo, and F. Gómez, Eds., Santiago, Chile: Gecamin, 2018, ch. 6, pp. 417–424, ISBN: 978-956-397-000-5

BibT_EX:

```
@inproceedings{Nussdorfer2018,
  address   = {Santiago, Chile},
  author    = {Nussdorfer, Stefan Franz and Ritt, Roland and
              Rothschedl, Christopher Josef},
  booktitle = {6th International Congress on Automation in
              Mining},
  chapter   = {6},
  editor    = {Babarovich, V{\'}ctor and Haro, Daniel and
              Gajardo, Mall{\'}n and G{\'}mez, Freddy},
  isbn      = {978-956-397-000-5},
  pages     = {417--424},
  publisher = {Gecamin},
  title     = {{Condition Monitoring of Hydraulics in Heavy
              Plant and Machinery}},
  year      = {2018}
}
```

Condition Monitoring of Hydraulics in Heavy Plant and Machinery

Stefan Franz Nussdorfer¹, Roland Ritt^{1*}, Christopher Josef Rothschedl¹
and Paul O'Leary¹

1. *Chair of Automation, Department Product Engineering, University of Leoben, Austria*

ABSTRACT

This paper presents a new approach to remote condition monitoring of hydraulic actuators as used in mining equipment. The condition of the hydraulic system has a major influence on the life span of machine parts and the equipment's performance in general. The aim is to use real-time monitoring to extract a better understanding of the state of the machine, its behaviour, and how it is being operated. Specifically, the goal is to generate added value along the complete life cycle of the equipment.

Within an existing framework, all available sensor and actuator data is collected. The hydraulic analysis module uses the pressures measured on the rod side $p_{r,i}$ and on the piston side $p_{p,i}$ of each hydraulic cylinder, with i indicating the i -th cylinder. Additionally, the values are combined with the metadata of the rod and piston areas $a_{r,i}$ and $a_{p,i}$, yielding the operating force $f_{o,i}$. The exemplary system consists of two hydraulic cylinders working in parallel. By analysing the sum f_s and difference f_d forces, the total lifting force and torsion exerted on the machine's boom are determined. For each of the variables $p_{r,i}$, $p_{p,i}$ and resulting forces, time series data is available for a period of 93 days, with the data being collected at sampling intervals of 1 second. Consequently, there are 8.035.200 samples per signal. The large number of samples ensures a well-defined confidence interval in the statistical evaluation of the data.

The time-varying histograms for 24 hour intervals make both short and long term changes visible. Whereby the large number of samples enables a reliable separation of systematic and stochastic components in the signal.

Operational analysis is presented for a bucket-wheel excavator which demonstrates the monitoring capabilities and specific results. The monitoring system has enabled the automatic detection of various defects in the observed hydraulics, e.g., a defective sensor was identified, as was the systematic occurrence of negative pressures.

INTRODUCTION

The condition of remotely located heavy plant and machinery, especially mining equipment, is vital to the processes they are part of. This paper focuses on the monitoring of the hydraulic systems of such machines, which are needed to carry out the luffing movements of a machine’s boom. The condition of these hydraulic systems has a major influence on the life span of the machine and, hence, has an important influence on the materials handling process. An unplanned shutdown can be the consequence of an unexpected failure of a crucial machine component: subsequent processes need to be stopped for the duration of the time-consuming maintenance works, which can quickly accumulate to high costs (Rothschedl, 2016).

The presented tool is part of a data analytics framework that supports the decision-making processes of mining and machine experts (domain experts): using their input, improved estimations of the component condition can be given. To accomplish the tasks necessary for this, a flexible and sufficiently secure framework is required; Figure 1 illustrates the implemented setup of such a framework. (Rothschedl et al., 2017)

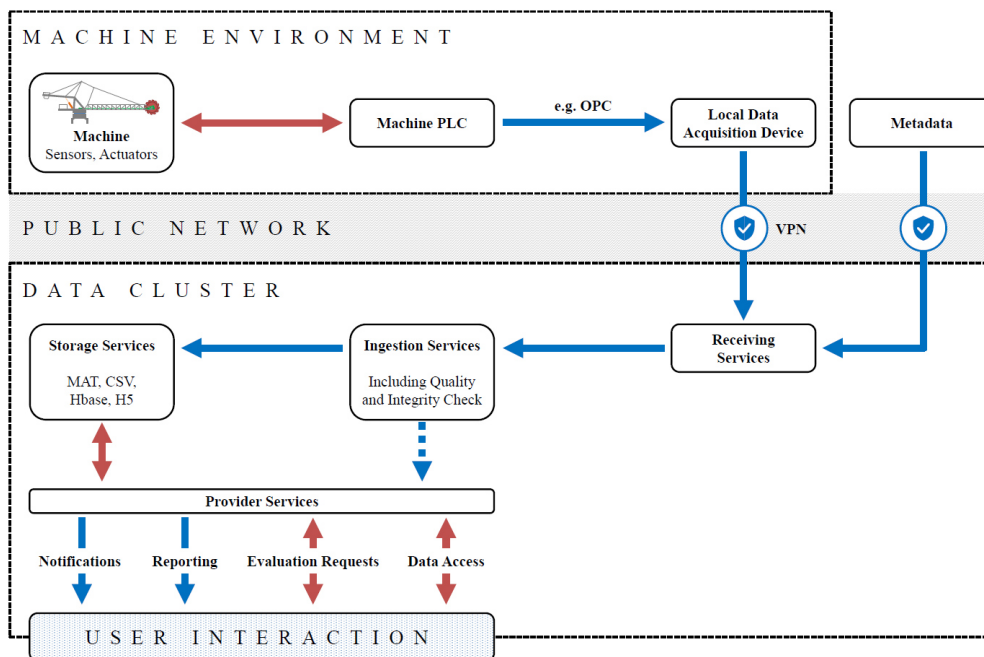


Figure 1 Schematic of the data ingestion system

It consists of a central data cluster that ingests, handles and processes operational data from currently 15 mobile mining machines, which are primarily located in remote locations and are equipped with a local data acquisition device. The ingestion process running on the data cluster ensures a consistent data format, the incoming values are accommodated in a contiguous data model. Such a model provides a distinct separation between input and output segmentation: data can be queried for an arbitrary period of time, even if the data is input on a daily basis only. Different analysing tools are available, which can be triggered manually by a user request or automatically on a regular basis. Metadata is integrated into the system via a secured connection that is entirely separated from the numerical sensor data. This is an additional security layer, as the numerical data is considered worthless without the corresponding metadata – no knowledge discovery is possible (O’Leary et al., 2016).

The available time series of the existing machines span several years, between 150 and 600 actuator and sensor signals are monitored per machine at a sampling rate of 1 Hz. Relevant channels of a bucket-wheel excavator were chosen to demonstrate the methods used in this paper.

METHODOLOGY

The herein presented analysis module gives an overview of tools and methods used for analysing hydraulics data to gain insights into machine behaviour. The aim is to support domain experts in monitoring the hydraulic systems of machines and in figuring out system improvement potential.

The developed analysis module addresses the hydraulics of boom luffing systems of machines used in mining: two cylinders are used to luff the boom up and down. The signal channels relevant for the exemplary evaluations are the rod and piston side pressures ($p_{r,i}$, $p_{p,i}$) of the cylinders in the hydraulic system. In a preliminary step, additional signal channels are used to partition the data into sequences to identify periods when the machine is operating and not operating. This is an important step, since the conclusions of evaluations may differ significantly based on the operation modes, e.g., the sequences where the machine is not operating can be used to characterise the hydraulic system, whereas sequences where the machine is operating can be used to monitor the operational behaviour.

Based on the available data, additionally derived channels can be included in the investigations, e.g., geometric metadata of the cylinders, operational forces produced by the cylinders can be calculated. They, again, can support the derivation of statistical properties, e.g., local entropy (Kollment, O’Leary, Ritt, & Klünsner, 2017).

Hydraulics Monitoring

In Figure 2, exemplary data of a bucket-wheel excavator is presented, showing the channels most relevant for the evaluations. Selected are: the luffing and slewing positions (*luff*, *slew*) to identify the current operation mode, and the pressure signals from the cylinders holding the bucket-wheel boom. Since there are two cylinders working in parallel, there are four signals: two for the piston side ($p_{p,1}$, $p_{p,2}$) and two for the rod side ($p_{r,1}$, $p_{r,2}$).

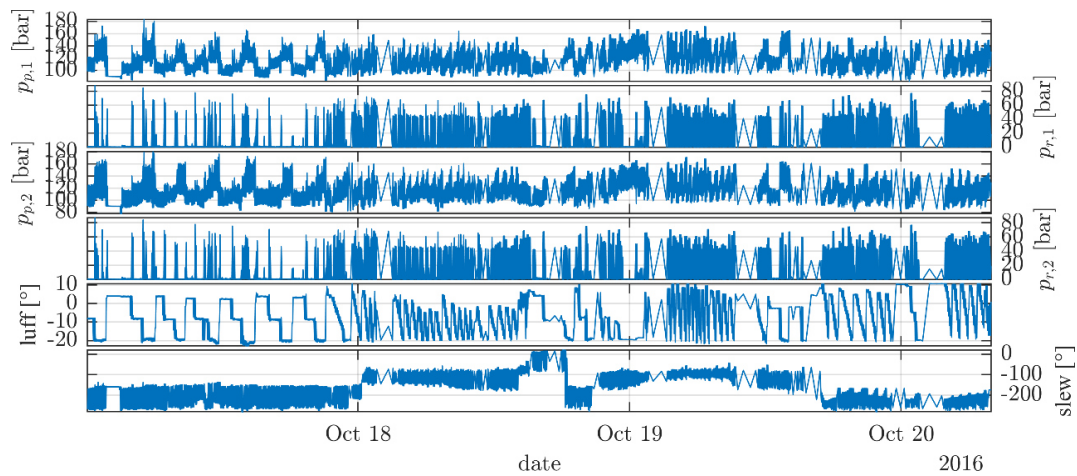


Figure 2 Relevant data channels for a duration of 4 days

In combination with the metadata of the rod and piston areas ($a_{r,i}$, $a_{p,i}$), the operating forces $f_{o,i}$ are calculated as

$$f_{o,i} = p_{p,i} a_{p,i} - p_{r,i} a_{r,i}. \quad (1)$$

Due to the fact that there are two hydraulic cylinders working in parallel, the sum and difference forces (f_s , f_d) are analysed to include an additional level of information. They are calculated as

$$f_s = f_{o,1} + f_{o,2} \quad (2)$$

$$f_d = f_{o,1} - f_{o,2}. \quad (3)$$

f_s is the total lifting force and f_d is a measure for the torsion exerted on the boom. In an idealistic case $f_d = 0$ and $f_{o,1} = f_{o,2}$. In Figure 3 those signals are shown for an entire day.

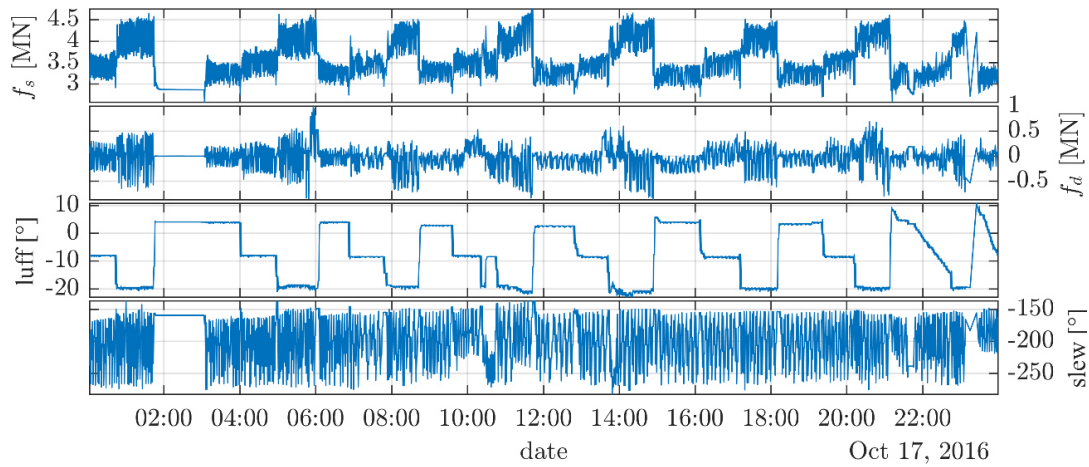


Figure 3 Data for a single day showing the sum and difference forces, the luff angle and the slew angle

As shown in Figure 4, the histogram for f_d reveals a slightly asymmetrical distribution that indicates a higher torsion in one specific direction. The reason is that the bucket-wheel is assembled at a skewed angle on the boom. This behaviour is therefore expected.

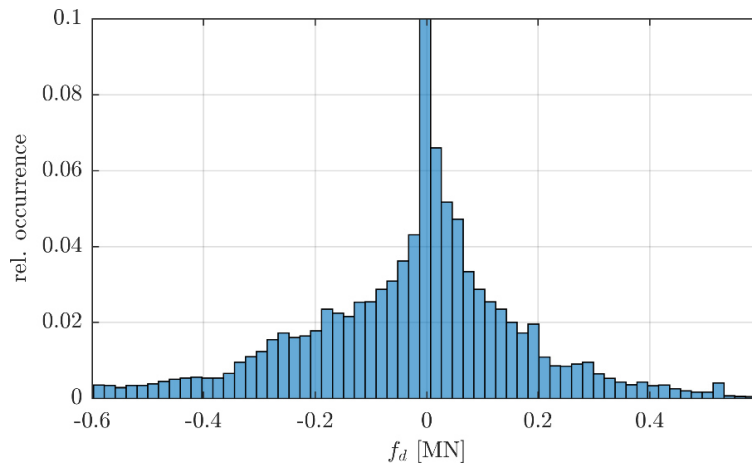


Figure 4 Histogram of f_d in the range of -0.6 to 0.6 MN

Due to the kinematics of the boom and the cylinders, a correlation between the luff angle and the total lifting force is given. This correlation can be quantified by calculating the correlation coefficient of these two signals. To do this for all channels, the correlation coefficient matrix C is derived (Lawson & Hanson, 1987). For the presented data (f_s , f_d , $luff$, $slew$) the correlation coefficient matrix constitutes to

$$C = \begin{bmatrix} 1 & -0.18 & -0.84 & -0.14 \\ -0.18 & 1 & 0.01 & -0.24 \\ -0.84 & 0.01 & 1 & 0.09 \\ -0.14 & -0.24 & 0.09 & 1 \end{bmatrix}.$$

(4)

The largest value at C[1,3] is the above mentioned (negative) correlation between the total lifting force and the luff angle, which is to be expected.

Statistics

The signal channels used in this work were sampled at 1 Hz over a period of 93 days. Thus, $n = 8.035.200$ samples for each channel are available (86.400 samples per day). This large number of samples ensures a well-defined confidence interval in the statistical evaluation of the data. Since there is a contiguous data model provided by the data framework, evaluations can be triggered for different time spans. In this paper, evaluations on a daily basis are presented.

A histogram is computed for each variable for each 24-hour period. The distribution can be characterised by its statistical central moments. The first central moment is the *mean* and the second one is the *variance*. The third moment is the *skewness*: it characterises the asymmetry of the distribution. The fourth central moment is the *kurtosis* and measures the steepness of the distribution (Loether & McTavish, 1980). Those properties can be used to find anomalies and identify time sequences, which demand further investigation.

To quickly get an overview of the entire statistics for a longer time period, the histograms for each day are collected as a column in a matrix, which can be viewed and processed as images. The time-varying histogram is used to visualise the evolution of the system response behaviour to find abrupt changes: these indicate parts that require further investigation. Both, short- and long-term changes in the statistical behaviour of the system become visible quickly (Kollment et al., 2017). This method is a powerful tool for the evaluation of big data sets used in an exploratory phase.

A time-varying histogram for the piston side pressure of cylinder 2 ($p_{p,2}$) is shown in Figure 5. The sudden change indicates a major change in the system behaviour. It was confirmed, that the sensor was flawed and was changed at this exact point in time.

For a more detailed inspection the histograms of single days can be investigated. They are found in the according column of the time-varying histogram matrix.

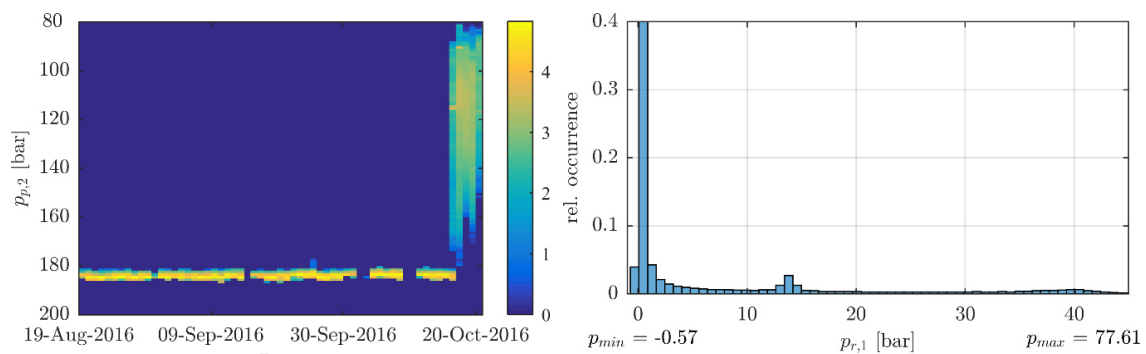


Figure 5 Time-varying histogram of the pressure of cylinder 2 piston side for a period of 93 days

Figure 6 Single-day histogram with a pressure range of -1 to 45 bar of cylinder 1 rod side

In Figure 6, such a histogram is plotted for the rod side pressure of cylinder 1 ($p_{r,1}$). There are 2 modes visible: the mode at $p_{r,1} \approx 30$ bar¹ is the expected pre-tension of the cylinder; the mode at $p_{r,1} \approx 0$ bar indicates a low pre-tension, although the machine is in operation: this may be of interest for domain experts. Rules applied to the value ranges of the pressure signals (p_{max} , p_{min}) can be used to find data sequences where certain limits are exceeded. In Figure 6, the occurrence of negative pressure values is visible. Negative pressures increase the risk of cavitation and can cause dirt and debris to be sucked into the oil cycle through the sealings. Additionally, the stiffness of the luffing system can become more and more unstable: if the load direction of a cylinder changes when the pressure is close to zero in one of its oil chambers, the piston needs to travel a certain distance for the pressure to build up again. Such a scenario would induce or increase superstructure rocking, which can lead to further issues and can have a significant impact on other components as well (Rothschedl, 2016). Hence, close to zero or even negative pressures in the cylinder hydraulics can cause serious problems.

RESULTS AND CONCLUSION

This paper covers the steps of exploratory work on hydraulics data of a mining machine. It was shown that many characteristics of the machine's hydraulic design can be found in the data during and between machine operations. Additionally, a sensor delivering erroneous values was identified. To interpret certain patterns and findings within the data correctly, the involvement of domain experts is required. To integrate this specific domain expertise, data has to be made available to these engineering and process experts in an easy-to-work-with manner. Many of the evaluations provide a technical discussion basis that acts as an interface between the raw data and the engineers.

The presented methods can either be used to analyse historical data sets to provide engineering feedback and to optimise, or can generate reports or notifications in real-time monitoring environments. Notifications triggered by automatic evaluations can facilitate decision-making chains to ensure corrective or preventative actions are taken in a timely manner. Further tools are currently under development with the aim of representing large data sets efficiently in a compact manner, without neglecting important content.

The results presented here have been evaluated by the hydraulics design engineer and the detected defects deemed to be valid results.

¹ The unit *bar* is used for pressure because of common on-site acceptance, although it is not an SI unit.

REFERENCES

- Kollment, W., O'Leary, P., Ritt, R., & Klünsner, T. (2017). Force based tool wear detection using Shannon entropy and phase plane. In *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (pp. 1–6). IEEE. <http://doi.org/10.1109/I2MTC.2017.7969765>
- Lawson, C. L., & Hanson, R. J. (1987). *Solving Least Squares Problems (Classics in Applied Mathematics)*. SIAM classics in applied mathematics (Vol. 15). <http://doi.org/10.1137/1.9781611971217>
- Loether, H. J., & McTavish, D. G. (1980). *Descriptive and Inferential Statistics: An Introduction* (2nd ed.). Allyn and Bacon.
- O'Leary, P., Harker, M., Ritt, R., Habacher, M., Landl, K., & Brandner, M. (2016). Mining Sensor Data in Larger Physical Systems. *IFAC-PapersOnLine*, 49(20), 37–42. <http://doi.org/10.1016/j.ifacol.2016.10.093>
- Rothschedl, C. J. (2016). *Condition Monitoring of Large-Scale Slew Bearings in Bucket-Wheel Boom-Type Reclaimers*. University of Leoben.
- Rothschedl, C. J., Ritt, R., O'Leary, P., Harker, M., Habacher, M., & Brandner, M. (2017). Real-Time-Data Analytics in Raw Materials Handling. In T. van Gerwe & D. Hößelbarth (Eds.), *Proceedings of Real-Time Mining, International Raw Materials Extraction Innovation Conference* (pp. 144–153). Amsterdam: Prof. Dr.-Ing. Jörg Benndorf.

Part V

Discussion and Appendices

18 | Conclusion and Outlook

The work performed during the herein presented research confirmed the thesis proposed at the beginning. Beside the conclusions given in the individual papers, the following statements about future investigations for data science in large cyber physical systems can be deducted:

The importance of including the physics of the system within calculations is stated throughout the entire thesis. Therefore, value and/or derivative data play an important role. Investigating the data in the (pseudo) phase space can yield valuable insights for, e.g. the segmentation of data, distance measures for time series sequences to find similarities or discords, or identifying a change in the system behaviour by investigating the area within a closed loop. Therefore, the concepts of ergodicity and recurrence times are important to consider [72].

A fundamental issue, which needs to be addressed in future when analysing data from physical systems, is the establishment of a generic structure to store more advanced metadata alongside the data. This encompasses metadata dealing with the kinematics of physical systems, interdependencies of data emanating from subsequent assembly-groups and the description of the dynamics of interacting components. Therefore, description languages, e.g. *Automation Markup Language (AutomationML)*¹, need to be considered and investigated.

Along with the handling of metadata, the concept of how hypotheses are formulated for querying data and performing analysis needs to be refined. Symbolic time series analysis as proposed within this work proved to be powerful. In future research, much more ideas from the metaphor of language need to be addressed, e.g. fixed states are seen as *nouns* and the action needed to move from one state to the other are seen as *verbs*. Additionally, the use of *adverbs* and *adjectives* as well as *punctuations* should be addressed. In this case, natural language processing tools can be applied to the emerging machine language to handle and process data analysis queries contextually and abstractly. Additionally, knowledge representation languages, e.g. *Web Ontology Language (OWL)*², need to be considered in this field of research.

To analyse large volumes of data emanating from large physical systems, symbolic time series can be used in hybrid learning approaches in future research. With the

¹<https://www.automationml.org>

²<https://www.w3.org/TR/owl2-overview>

use of symbolic time series the physics of the system is included within the machine learning task. In this context, a consistent distance measure for symbolic time series generated from multi-dimensional time series, i.e. multisyllabic symbolic time series, needs to be established.

A rather philosophical issue in data analysis, which is often neglected but important, is the difference between describing data (ontology) and understanding data (epistemology). Just because an incident can be described, does not automatically imply that it can be predicted as well, since the causal link is not there or not established correctly. As in a lot of cases the underlying *truth* is not known, *intuition* (often based on *belief* and sometimes *justified belief*) is used to formulate hypotheses. Experience, either gained personally or from others, justifies and strengthens the belief, yielding *justified belief*. However, the question followed by this is: *What exactly is the justification in a justified belief?* This question needs to be considered and should be a driver each time data science in cyber physical systems is addressed.

Appendices

A | List of Figures¹

2.1	Digital Twin for Large Physical Systems	12
2.2	CRISP Data Mining Cycle	15
2.3	Data Mining Pyramid as proposed by Mark Embrechts	16
2.4	Fundamental premiss behind data analytics	17
3.1	Data Flow and Ingestion.	20
3.2	Data collection and secure data transfer.	23
3.3	Contiguous Data Model.	24
4.1	Data analytics framework represented in the form of a schematic class diagramm.	27
4.2	Visual representation of the core attributes of the data handling framework.	30
4.3	Graphical visualization of the symbolic representation of a time series. 32	
	(a) Symbolization of a time series based on its slope.	32
	(b) Structure of the symbolic representation of the time series.	32
4.4	Graphical visualization of segments.	33
	(a) Multiple segments of the same type in a time series.	33
	(b) Structure of how segments are handled.	33
4.5	A multi-dimensional time series visualized as a stacked plot with a common time axis. Each channel is plotted as one sub-plot.	36
4.6	Schematic of decimation; Multiple data points within the same pixel are reduced to a single data point, since sub-pixel details cannot be displayed on a monitor. The grey squares indicate the activated pixels. 37	
	(a) Schematic of decimation used in scatter-plots.	37
	(b) Schematic of decimation used in line-plots.	37

¹The figures included in the papers are not within this list.

4.7	Visualization of decorative objects using overlays.	38
9.1	Graphical visualization of the LDO-matrices and the convolutional approach.	96
(a)	Local LDO-matrix L	96
(b)	Partitioned LDO-matrix; green: top part of L ; blue: centre row of L ; red: bottom rows of L	96
(c)	Expanded LDO-matrix L_{full}	96
(d)	Convolving the centre row over the data; top: the flipped centre row acting as kernel; middle: the original data to be operated on; bottom: operated data generated via convolution (with corrected endpoints using the top and bottom part of L).	96

B | List of Tables¹

1.1	Contribution of the Author to the publications collected in the thesis in percent.	6
-----	--	---

C | List of Algorithms²

9.1	Convolutional approach for collocational local linear differential computations.	96
-----	--	----

¹The tables included in the papers are not within this list.

²The algorithms included in the papers are not within this list.

D | List of Author's Publications

- [P1] R. Ritt, M. Harker, and P. O'Leary, "Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials," *arXiv Open Access Journal Article*, Mar. 2019. arXiv: 1903.10810. [Online]. Available: <http://arxiv.org/abs/1903.10810>.
- [P2] ———, "Simultaneous Approximation of Measurement Values and Derivative Data using Discrete Orthogonal Polynomials," in *2019 IEEE Industrial Cyber-Physical Systems (ICPS)*, Taipei: IEEE, 2019.
- [P3] R. Ritt, P. O'Leary, C. J. Rothschedl, A. Almasri, and M. Harker, "Hierarchical Decomposition and Approximation of Sensor Data," in *Proceedings of the 1st International Conference on Numerical Modelling in Engineering*, Lecture Notes in Mechanical Engineering, M. A. Wahab, Ed., vol. NME2018, Springer Singapore, 2019, pp. 351–370. DOI: 10.1007/978-981-13-2273-0_27. [Online]. Available: http://link.springer.com/10.1007/978-981-13-2273-0%7B%5C_%7D27.
- [P4] R. Ritt and P. O'Leary, "Symbolic Analysis of Machine Behaviour and the Emergence of the Machine Language," in *Theory and Practice of Natural Computing*, Springer International Publishing, 2018, pp. 305–316. DOI: 10.1007/978-3-030-04070-3_24. [Online]. Available: http://link.springer.com/10.1007/978-3-030-04070-3%7B%5C_%7D24.
- [P5] R. Ritt, P. O'Leary, C. J. Rothschedl, and M. Harker, "Advanced Symbolic Time Series Analysis In Cyber Physical Systems," in *Proceedings - International work-conference on Time Series, ITISE 2017*, O. Valenzuela, F. Rojas, H. Pomares, and I. Rojas, Eds., vol. 1, Granada: University of Granada, Feb. 2017, pp. 155–160, ISBN: 9788417293017. arXiv: 1802.00617. [Online]. Available: http://arxiv.org/abs/1802.00617%20http://itise.ugr.es/proceedings/ITISE%7B%5C_%7D2017.zip.
- [P6] S. F. Nussdorfer, R. Ritt, and C. J. Rothschedl, "Condition Monitoring of Hydraulics in Heavy Plant and Machinery," in *6th International Congress on Automation in Mining*, V. Babarovich, D. Haro, M. Gajardo, and F. Gómez, Eds., Santiago, Chile: Gecamin, 2018, ch. 6, pp. 417–424, ISBN: 978-956-397-000-5.

- [P7] C. J. Rothschedl, R. Ritt, P. O'Leary, M. Harker, M. Habacher, and M. Brandner, "Real-Time-Data Analytics in Raw Materials Handling," in *Proceedings of Real-Time Mining, International Raw Materials Extraction Innovation Conference*, T. van Gerwe and D. Höfelbarth, Eds., Amsterdam: Prof. Dr.-Ing. Jörg Benndorf, 2017, pp. 144–153, ISBN: 9783938390412. arXiv: 1802.00625. [Online]. Available: <http://arxiv.org/abs/1802.00625>.
- [P8] P. O'Leary, R. Ritt, and M. Harker, "Constrained Polynomial Approximation for Inverse Problems in Engineering," in *Proceedings of the 1st International Conference on Numerical Modelling in Engineering*, Lecture Notes in Mechanical Engineering, M. A. Wahab, Ed., vol. NME2018, Springer Singapore, 2019, pp. 225–244. DOI: 10.1007/978-981-13-2273-0_19. [Online]. Available: http://link.springer.com/10.1007/978-981-13-2273-0%7B%5C_%7D19.
- [P9] P. O'Leary, M. Harker, R. Ritt, M. Habacher, K. Landl, and M. Brandner, "Mining Sensor Data in Larger Physical Systems," *IFAC-PapersOnLine*, vol. 49, no. 20, pp. 37–42, 2016, ISSN: 24058963. DOI: 10.1016/j.ifacol.2016.10.093. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S2405896316316561>.
- [P10] W. Kollment, P. O'Leary, R. Ritt, and T. Klünsner, "Force based tool wear detection using Shannon entropy and phase plane," in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, IEEE, May 2017, pp. 1–6, ISBN: 978-1-5090-3596-0. DOI: 10.1109/I2MTC.2017.7969765. [Online]. Available: <https://ieeexplore.ieee.org/document/7969765/>.
- [P11] R. Schmidt, P. O'Leary, R. Ritt, and M. Harker, "MEMS based inclinometers: Noise characteristics and suitable signal processing," in *I2MTC 2017 - 2017 IEEE International Instrumentation and Measurement Technology Conference, Proceedings*, IEEE, May 2017, pp. 1–6, ISBN: 9781509035960. DOI: 10.1109/I2MTC.2017.7969830. [Online]. Available: <http://ieeexplore.ieee.org/document/7969830/>.

E | References

- [1] K. Weierstrass, *Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen*, 1885. [Online]. Available: <https://www.math.auckland.ac.nz/hat/fpapers/wei4.pdf>.
- [2] R. Schutt and C. O’Neil, *Doing data science*, p. 375, ISBN: 9781449358655.
- [3] D. M. Dedge Parks, “Defining Data Science and Data Scientist,” *ProQuest Dissertations and Theses*, no. October, p. 77, 2017. [Online]. Available: <https://scholarcommons.usf.edu/cgi/viewcontent.cgi?referer=https://www.google.com/%7B%5C%7Dhttpsredir=1%7B%5C%7Darticle=8211%7B%5C%7Dcontext=etd>.
- [4] F. S. Haug, “Bad big data science,” in *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2016, pp. 2863–2871, ISBN: 978-1-4673-9005-7. DOI: 10.1109/BigData.2016.7840935. [Online]. Available: <http://ieeexplore.ieee.org/document/7840935/>.
- [5] *Journal of Data Science*, 2002. [Online]. Available: <http://www.jds-online.com/about%20http://www.sinica.edu.tw/%7B~%7Djds/JDS-7.html> (visited on 03/15/2019).
- [6] V. Dhar, “Data science and prediction,” *Communications of the ACM*, vol. 56, no. 12, pp. 64–73, Dec. 2013, ISSN: 00010782. DOI: 10.1145/2500499. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2534706.2500499>.
- [7] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, “Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2318–2331, Oct. 2017, ISSN: 1041-4347. DOI: 10.1109/TKDE.2017.2720168. arXiv: 1612.08544. [Online]. Available: <http://ieeexplore.ieee.org/document/7959606/>.
- [8] K. Cukier and V. Mayer-Schoenberger, “The Rise of Big Data: How It’s Changing the Way We Think About the World,” *Foreign Affairs*, vol. 92, no. 3, pp. 28–40, 2013. [Online]. Available: <http://www.jstor.org/stable/23526834>.

- [9] NIST and E. Incorporated, “Cyber-Physical Systems: Situation Analysis of Current Trends, Technologies, and Challenges,” *National Institute of Standards and Technology*, p. 67, 2012. [Online]. Available: http://events.energetics.com/NIST-CPSWorkshop/pdfs/CPS%7B%5C_%7DSituation%7B%5C_%7DAnalysis.pdf%7B%5C%7D5Cnhttp://cps-vo.org/node/7139.
- [10] J. Sztipanovits, S. Ying, D. Corman, J. Davis, S. M. Leadership, P. J. Mosterman, V. Prasad, and L. Stormo, “Strategic R&D Opportunities for 21st Century Cyber-Physical Systems,” NIST - Steering Committee for Foundations for Innovation in Cyber-physical Systems, Tech. Rep., 2013.
- [11] D. Spath, S. Gerlach, M. Hämmerle, S. Schlund, and T. Strölin, “Cyber-Physical System for Self-Organised and flexible Labour Utilisation,” *Proceedings of the 22nd International Conference on Production Research*, vol. 50, p. 22, 2013.
- [12] Y. Liu, Y. Peng, B. Wang, S. Yao, and Z. Liu, “Review on cyber-physical systems,” *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 1, pp. 27–40, Jan. 2017, ISSN: 2329-9266. DOI: 10.1109/JAS.2017.7510349. [Online]. Available: <http://ieeexplore.ieee.org/document/7815549/>.
- [13] R. (Rajkumar, I. Lee, L. Sha, and J. Stankovic, “Cyber-physical systems,” in *Proceedings of the 47th Design Automation Conference on - DAC '10*, vol. 156, New York, New York, USA: ACM Press, Apr. 2010, p. 731, ISBN: 9781450300025. DOI: 10.1145/1837274.1837461. [Online]. Available: <https://academic.oup.com/endo/article-lookup/doi/10.1210/en.2014-1673%20http://portal.acm.org/citation.cfm?doid=1837274.1837461>.
- [14] J. Huang and L. Zhang, “Research and challenges of CPS,” *AIP Conference Proceedings*, vol. 1864, no. August, pp. 3–7, 2017, ISSN: 15517616. DOI: 10.1063/1.4992851.
- [15] K.-J. Park, R. Zheng, and X. Liu, “Cyber-physical systems: Milestones and research challenges,” *Computer Communications*, vol. 36, no. 1, pp. 1–7, Dec. 2012, ISSN: 01403664. DOI: 10.1016/j.comcom.2012.09.006. arXiv: arXiv:1011.1669v3. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0140366412003180>.
- [16] Y. Ashibani and Q. H. Mahmoud, “Cyber physical systems security: Analysis, challenges and solutions,” *Computers and Security*, vol. 68, pp. 81–97, 2017, ISSN: 01674048. DOI: 10.1016/j.cose.2017.04.005.
- [17] E. Geisberger and M. Broy, “agendaCPS Integrierte Forschungsagenda Cyber-Physical Systems,” Tech. Rep., 2012. [Online]. Available: www.acatech.de.
- [18] P. Bogdan and R. Marculescu, “Towards a science of cyber-physical systems design,” *Proceedings - 2011 IEEE/ACM 2nd International Conference on Cyber-Physical Systems, ICCPS 2011*, pp. 99–108, 2011, ISSN: 9780769543611 (ISBN). DOI: 10.1109/ICCPS.2011.14.

- [19] E. A. Lee, "Cyber Physical Systems: Design Challenges," in *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, vol. 13, IEEE, May 2008, pp. 363–369, ISBN: 978-0-7695-3132-8. DOI: 10.1109/ISORC.2008.25. [Online]. Available: <http://ieeexplore.ieee.org/document/4519604/>.
- [20] ———, "CPS foundations," in *Proceedings of the 47th Design Automation Conference on - DAC '10*, New York, New York, USA: ACM Press, 2010, p. 737, ISBN: 9781450300025. DOI: 10.1145/1837274.1837462. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1837274.1837462>.
- [21] E. Lee, "The Past, Present and Future of Cyber-Physical Systems: A Focus on Models," *Sensors*, vol. 15, no. 3, pp. 4837–4869, Feb. 2015, ISSN: 1424-8220. DOI: 10.3390/s150304837. [Online]. Available: <http://www.mdpi.com/1424-8220/15/3/4837>.
- [22] H. Kagermann, W. Wahlster, J. Helbig, A. Hellinger, M. A. V. Stumpf, L. Treugut, J. Blasco, and H. Galloway, "Umsetzungsempfehlungen für das Zukunftsproject Industrie 4.0," acatech – Deutsche Akademie der Technikwissenschaften e.V., Tech. Rep. April, 2013. [Online]. Available: https://www.bmbf.de/files/Umsetzungsempfehlungen%7B%5C_%7DIndustrie4%7B%5C_%7D0.pdf.
- [23] S. O. Okolie, S. O. Kuyoro, and O. B. Ohwo, "Emerging Cyber-Physical Systems : An Overview," vol. 3, no. 8, pp. 306–316, 2018.
- [24] D. Spath, O. Ganschar, S. Gerlach, M. Hämmerle, T. Krause, and S. Schlund, *Produktionsarbeit der Zukunft - Industrie 4.0 : [Studie]*, T. Bauernhansl, M. ten Hompel, and B. Vogel-Heuser, Eds. Wiesbaden: Fraunhofer Verlag, 2013, ISBN: 978-3-8396-0570-7. [Online]. Available: <http://link.springer.com/10.1007/978-3-658-04682-8>.
- [25] J. Lee, B. Bagheri, and H.-A. Kao, "Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics," *Int. Conference on Industrial Informatics (INDIN)*, no. November 2015, pp. 1–6, 2014, ISSN: 1935-4576. DOI: 10.13140/2.1.1464.1920. arXiv: arXiv:1011.1669v3.
- [26] P. O'Leary, M. Harker, C. Gugg, P. O. Leary, M. Harker, and C. Gugg, "Sensor-data Analytics in Cyber Physical Systems - From Husserl to Data Mining," in *Proceedings of the 4th International Conference on Sensor Networks*, SCITEPRESS - Science, 2015, pp. 176–181, ISBN: 978-989-758-086-4. DOI: 10.5220/0005328601760181. [Online]. Available: <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005328601760181>.
- [27] R. Baheti and H. Gill, "Cyber-physical systems," *The Impact of Control Technology*, pp. 161–166, 2011.
- [28] A. A. Letichevsky, O. O. Letychevskiy, V. G. Skobelev, and V. A. Volkov, "Cyber-Physical Systems," *Cybernetics and Systems Analysis*, vol. 53, no. 6, pp. 821–834, 2017, ISSN: 15738337. DOI: 10.1007/s10559-017-9984-9.

- [29] Q. Qi, D. Zhao, T. W. Liao, and F. Tao, “Modeling of Cyber-Physical Systems and Digital Twin Based on Edge Computing, Fog Computing and Cloud Computing Towards Smart Manufacturing,” in *Volume 1: Additive Manufacturing; Bio and Sustainable Manufacturing*, ASME, Jun. 2018, V001T05A018, ISBN: 978-0-7918-5135-7. DOI: 10.1115/MSEC2018-6435. [Online]. Available: <http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?doi=10.1115/MSEC2018-6435>.
- [30] C. Gugg, “An Algebraic Framework for the Solution of Inverse Problems in Cyber-Physical Systems,” PhD thesis, Montanuniversitaet Leoben, 2015.
- [31] J. Han, M. Kamber, and J. Pei, *Data mining : concepts and techniques*. Elsevier Science, 2011, p. 744, ISBN: 9780123814791.
- [32] K. J. Cios and L. A. Kurgan, “Trends in Data Mining and Knowledge Discovery,” in *Advanced Techniques in Knowledge Discovery and Data Mining*, Dm, London: Springer London, 2007, pp. 1–26, ISBN: 1846281830. DOI: 10.1007/1-84628-183-0_1. [Online]. Available: http://link.springer.com/10.1007/1-84628-183-0%7B%5C_%7D1.
- [33] K. Cios, A. Teresinska, S. Konieczna, J. Potocka, and S. Sharma, “A knowledge discovery approach to diagnosing myocardial perfusion,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 4, pp. 17–25, 2000, ISSN: 07395175. DOI: 10.1109/51.853478. [Online]. Available: <http://ieeexplore.ieee.org/document/853478/>.
- [34] C. Shaerer, “The CRISP-DM model: the new blueprint for data mining,” *Journal of Data Warehousing*, vol. 5, no. 4, 2000.
- [35] S. Anand, A. Patrick, J. Hughes, and D. Bell, “A Data Mining methodology for cross-sales,” *Knowledge-Based Systems*, vol. 10, no. 7, pp. 449–461, May 1998, ISSN: 09507051. DOI: 10.1016/S0950-7051(98)00035-5. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0950705198000355>.
- [36] S. S. Anand, A. G. Büchner, and Financial Times Management., *Decision support using data mining*. London: Financial Times Management, 1998, p. 168, ISBN: 0273632698.
- [37] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “Advances in knowledge discovery and data mining: an overview,” in *Advances in knowledge discovery and data mining*, AAAI Press, 1996, pp. 1–35, ISBN: 0262560976. [Online]. Available: <https://dl.acm.org/citation.cfm?id=257942>.
- [38] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD process for extracting useful knowledge from volumes of data,” *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, Nov. 1996, ISSN: 00010782. DOI: 10.1145/240455.240464. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=240455.240464>.
- [39] —, “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996. DOI: 10.1609/AIMAG.V17I3.1230. [Online]. Available: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>.

- [40] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “Knowledge Discovery and Data Mining: Towards a Unifying Framework,” *KDD’96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 82–88, 1996, ISSN: 13500872. DOI: 10.1.1.27.363. [Online]. Available: <http://www.aaai.org/Papers/KDD/1996/KDD96-014>.
- [41] L. A. Kurgan and P. Musilek, “A survey of Knowledge Discovery and Data Mining process models,” *The Knowledge Engineering Review*, vol. 21, no. 01, p. 1, Mar. 2006, ISSN: 0269-8889. DOI: 10.1017/S0269888906000737. [Online]. Available: http://www.journals.cambridge.org/abstract%7B%5C_%7DS0269888906000737.
- [42] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISO-DM 1.0: Step-by-step data mining guide*, 2000. [Online]. Available: <https://www.the-modeling-agency.com/crisp-dm.pdf> (visited on 03/19/2019).
- [43] Kenneth Jensen, *CRISP-DM Process Diagram.png - Wikimedia Commons*. [Online]. Available: https://commons.wikimedia.org/wiki/File:CRISP-DM%7B%5C_%7DProcess%7B%5C_%7DDiagram.png (visited on 04/12/2019).
- [44] M. Last, A. Kandel, and H. Bunke, *Data Mining in Time Series Databases*, ser. Series in Machine Perception and Artificial Intelligence. WORLD SCIENTIFIC, Jun. 2004, vol. 57, ISBN: 978-981-238-290-0. DOI: 10.1142/5210. [Online]. Available: <http://books.google.at/books?id=f38wqKjyBm4C%20http://www.worldscientific.com/worldscibooks/10.1142/5210>.
- [45] E. Keogh and S. Kasetty, “On the need for time series data mining benchmarks,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’02*, New York, New York, USA: ACM Press, 2002, p. 102, ISBN: 158113567X. DOI: 10.1145/775047.775062. [Online]. Available: <http://dl.acm.org/citation.cfm?id=775047.775062%20http://portal.acm.org/citation.cfm?doid=775047.775062>.
- [46] E. Fuchs, T. Gruber, H. Pree, and B. Sick, “Temporal data mining using shape space representations of time series,” *Neurocomputing*, vol. 74, no. 1-3, pp. 379–393, Dec. 2010, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2010.03.022. [Online]. Available: <http://dx.doi.org/10.1016/j.neucom.2010.03.022%20http://linkinghub.elsevier.com/retrieve/pii/S0925231210002237>.
- [47] P. Esling and C. Agon, “Time-series data mining,” *ACM Computing Surveys*, vol. 45, no. 1, pp. 1–34, Nov. 2012, ISSN: 03600300. DOI: 10.1145/2379776.2379788. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2379776.2379788>.
- [48] M. Embrechts, B. Szymanski, and K. Sternickel, *Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing*. John Wiley and Sons, New York, 2005, ch. Chapter 10, ISBN: 0471476684.
- [49] R. L. Ackoff, “From Data to Wisdom,” *Journal of Applied Systems Analysis*, vol. 16, pp. 3–9, 1989.

- [50] H. von. Hentig, *Kreativität: Hohe Erwartungen an einen schwachen Begriff*. Beltz, 2007, p. 80, ISBN: 978-3-407-22067-7. [Online]. Available: https://www.amazon.de/Kreativit%7B%5C%22%7Ba%7D%7Dt-Erwartungen-schwachen-Begriff-Taschenbuch/dp/3407220677/ref=pd%7B%5C_%7Dsbs%7B%5C_%7D14%7B%5C_%7D1?%7B%5C_%7Dencoding=UTF8%7B%5C%7Dpsc=1%7B%5C%7DrefRID=AJYPMTSX2CTN4599B16Y.
- [51] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948, ISSN: 00058580. DOI: 10.1002/j.1538-7305.1948.tb01338.x. [Online]. Available: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf%20http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024>.
- [52] W. Mahnke, S.-H. Leitner, and M. Damm, *OPC Unified Architecture*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, ISBN: 978-3-540-68898-3. DOI: 10.1007/978-3-540-68899-0. [Online]. Available: <http://link.springer.com/10.1007/978-3-540-68899-0>.
- [53] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec, “The many faces of publish/subscribe,” *ACM Computing Surveys*, vol. 35, no. 2, pp. 114–131, Jun. 2003, ISSN: 03600300. DOI: 10.1145/857076.857078. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=857076.857078>.
- [54] J. Lin, E. Keogh, L. Wei, and S. Lonardi, “Experiencing SAX: a novel symbolic representation of time series,” *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 107–144, Aug. 2007, ISSN: 1384-5810. DOI: 10.1007/s10618-007-0064-z. [Online]. Available: <http://link.springer.com/10.1007/s10618-007-0064-z>.
- [55] A. Camera, T. Palpanas, J. Shieh, and E. Keogh, “iSAX 2.0: Indexing and mining one billion time series,” *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 58–67, 2010, ISSN: 15504786. DOI: 10.1109/ICDM.2010.124.
- [56] J. Lin, R. Khade, and Y. Li, “Rotation-invariant similarity in time series using bag-of-patterns representation,” *Journal of Intelligent Information Systems*, vol. 39, no. 2, pp. 287–315, Oct. 2012, ISSN: 0925-9902. DOI: 10.1007/s10844-012-0196-5. [Online]. Available: <http://link.springer.com/10.1007/s10844-012-0196-5>.
- [57] C. G. Nevill-Manning and I. H. Witten, “Identifying Hierarchical Structure in Sequences: A linear-time algorithm,” *Journal of Artificial Intelligence Research*, vol. 7, no. 1, pp. 67–82, Aug. 1997, ISSN: 1076-9757. DOI: arXiv:cs/9709102. arXiv: 9709102 [cs]. [Online]. Available: <http://arxiv.org/abs/cs/9709102>.
- [58] L. Ye and E. Keogh, “Time Series Shapelets: A New Primitive for Data Mining,” *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, p. 947, 2009. DOI: 10.1145/1557019.1557122. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1557019.1557122>.

- [59] J. Grabocka, M. Wistuba, and L. Schmidt-Thieme, “Time-series classification through histograms of symbolic polynomials,” *arXiv preprint arXiv:1307.6365*, Jul. 2013. DOI: 10.1109/TKDE.2014.2377746. arXiv: arXiv:1307.6365v4. [Online]. Available: <http://arxiv.org/abs/1307.6365>.
- [60] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein, “GrammarViz 3.0: Interactive Discovery of Variable-Length Time Series Patterns,” *ACM Transactions on Knowledge Discovery from Data*, vol. 12, no. 1, pp. 1–28, Feb. 2018, ISSN: 15564681. DOI: 10.1145/3051126. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=3178542.3051126>.
- [61] E. Keogh, J. Lin, and A. Fu, “HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence,” in *Fifth IEEE International Conference on Data Mining (ICDM’05)*, IEEE, 2005, pp. 226–233, ISBN: 0-7695-2278-5. DOI: 10.1109/ICDM.2005.79. [Online]. Available: <http://ieeexplore.ieee.org/document/1565683/>.
- [62] J. W. Tukey, *Exploratory data analysis*. Addison-Wesley Pub. Co, 1977, p. 688, ISBN: 9780201076165. [Online]. Available: https://www.goodreads.com/book/show/2111739.Exploratory%7B%5C_%7DData%7B%5C_%7DAnalysis.
- [63] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, “Visualizing time-oriented data—A systematic view,” *Computers & Graphics*, vol. 31, no. 3, pp. 401–409, Jun. 2007, ISSN: 0097-8493. DOI: 10.1016/j.cag.2007.01.030. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0097849307000611>.
- [64] W. Javed, B. McDonnell, and N. Elmqvist, “Graphical Perception of Multiple Time Series,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 927–934, Nov. 2010, ISSN: 1077-2626. DOI: 10.1109/TVCG.2010.162. [Online]. Available: <http://ieeexplore.ieee.org/document/5613429/>.
- [65] S. Steinarsson, “Downsampling Time Series for Visual Representation,” M.Sc. thesis, School of Engineering and Natural Sciences University of Iceland, 2013, p. 87. [Online]. Available: https://skemman.is/bitstream/1946/15343/3/SS%7B%5C_%7DMSthesis.pdf.
- [66] W. Shi and C. Cheung, “Performance Evaluation of Line Simplification Algorithms for Vector Generalization,” *The Cartographic Journal*, vol. 43, no. 1, pp. 27–44, Mar. 2006, ISSN: 0008-7041. DOI: 10.1179/000870406X93490. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1179/000870406X93490>.
- [67] D. H. Douglas and T. K. Peucker, “Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature,” in *Classics in Cartography*, 2, vol. 10, Chichester, UK: John Wiley & Sons, Ltd, Mar. 2011, pp. 15–28, ISBN: 9780470669488. DOI: 10.1002/9780470669488.ch2. [Online]. Available: <http://doi.wiley.com/10.1002/9780470669488.ch2>.

- [68] M. Riss, “FTSPlot: Fast Time Series Visualization for Large Datasets,” *PLoS ONE*, vol. 9, no. 4, J. Aerts, Ed., e94694, Apr. 2014, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0094694. [Online]. Available: <http://dx.plos.org/10.1371/journal.pone.0094694>.
- [69] N. Duchonova, “Mutli-resolution Visualisation of Large Time Series,” Master Thesis, Masaryk University, 2014, pp. 1–28. [Online]. Available: <https://is.muni.cz/th/yuh6m/thesis.pdf>.
- [70] L. Berry and T. Munzner, *BinX: Dynamic Exploration of Time Series Datasets Across Aggregation Levels*, 2005. DOI: 10.1109/infvis.2004.11. [Online]. Available: <http://ieeexplore.ieee.org/document/1382913/>.
- [71] P. O’Leary and M. Harker, “Discrete polynomial moments and Savitzky-Golay smoothing,” *World Academy of Science, Engineering and Technology; International Journal of Computer and Information Engineering*, vol. 72, pp. 439–443, 2010, ISSN: 2010376X. [Online]. Available: <https://waset.org/publications/12268/discrete-polynomial-moments-and-savitzky-golay-smoothing>.
- [72] H. Poincaré, “Sur le probleme des trois corps et les équations de la dynamique,” *Acta Mathematica*, vol. 13, no. 1, pp. 5–7, 1890, ISSN: 0001-5962. DOI: 10.1007/BF02392506. [Online]. Available: <http://link.springer.com/10.1007/BF02392506>.