Technische Universität München
Ingenieurfakultät Bau Geo Umwelt
Lehrstuhl für Statik

Combining Physics-Based Models and Machine Learning for an Enhanced Structural Health Monitoring

Mohamed Khalil, MSc.-Hons.

Vollständiger Abdruck der von der Ingenieurfakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender:

Prof. Dr.-Ing. Christian Große

Prüfer der Dissertation:

1. Prof. Dr.-Ing. Kai-Uwe Bletzinger
2. Prof. Dr. ir. Daniel J. Rixen
3. Prof. Riccardo Rossi, Ph.D.

Die Dissertation wurde am 15.04.2020 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 16.03.2021 angenommen.

Schriftenreihe des Lehrstuhl für Statik TU München

Band 46

Mohamed Khalil, MSc.-Hons.

Combining Physics-Based Models and Machine Learning for an Enhanced Structural Health Monitoring

München 2020

To the brain cells, eye balls, and cervical spine vertebrae that propelled beyond their physiological limits during the times of aspiration, desperation, and procrastination.

Acknowledgements

I would like to express my gratitude to everyone who extended their unconditioned support and motivation, without which this work would have had ended up being 200 pages of lorem ipsum and contemporary artwork.

My parents, my father *Prof.-MD. Magdy Khalil*, and my mother *Dr. Azza Nasr*, I owe you endless genuine thanks throughout the journey of life, and not just the years of PhD. Your gleams of belief, encouragement, and consolidation are the reason I keep moving forward, even through the times of despair. My beloved sister, *Noha Khalil*, I will life-long feel indebted to your sincere admiration, reassurance and compassion.

Life during the past couple of years had more than just the scientific content of this work. To my dear friends in Germany, *Abdelrahman Elskhawy*, *Ahmed Khattab*, *Dr. Ahmed Saleh*, *Amr Seleem*, and *Omar Shouman*, it is due to your presence, I felt surrounded by family in Munich.
My school and university friends back in Cairo, *Loay Khalil*, *Mohamed Hamza*, *Noor Elmahallawi*, and *Youssek Elkorma*, you never made it feel like a thousand miles away. My utmost gratitude to you for being by my side when I longed for a brother to vent and a wise man to consult.

Special thanks is forwarded to *Dr. Ahmed Kamal*; the friend and the colleague. I would also like to extend a precious thank-you to all the peers who "voluntarily" reviewed my work. I, here, would like to particularly name *Abdelrahman Elskhawy*, *Ahmed Khattab*, *Omar Shouman*, *Diana Manvelyan*, *Ioannis Kouroudis*, and *Jakob Zietsch*. Without your rounds of constructive feedback, the quality of this dissertation would have never been anywhere close to its current.

I forward my thanks to my advisers and mentors *Dr. Vincent Malik*, *Dr.-Ing. Roland Wüchner* and *Prof. Kai-Uwe Bletzinger* for their undoubted assurance of my work, and their incessant academic and professional advisory and support through the course of my doctoral work.
Further, I would like to thank my former and current managers at Siemens *Dr. Efrossini Tsouchnika* and *Dr. Christoph Heinrich* for putting their trust in me back then in 2017, and granting me the opportunity to conduct my doctoral research at Siemens Corporate Technology.

I feel very grateful to my colleagues *Dr. Arianna Bosco*, *Dr. Christoph Bergs*, *Christoph Ludwig*, *Dr. Dirk Hartmann*, *Ioannis Kourodis*, *Jack Zietsch*, *Dr. Stefan Boschert*, *Stefan Gavranovic*, and *Theo Papadopoulos* for your fruitful collaboration, your guidance through daily corporate business, and, most appreciated, your festive work mood. That latter kept the not-my-best days going and made long evening stays in the office more joyful than they sound.

The final thanks go to the unknown soldiers, who remain in the dark aisles unrecognised enough for their precious roles. I would like to acknowledge the thousands of individuals who have coded for open-source projects and uploaded them to Github for free. It is due to their efforts this scientific work could be managed in such a critical period. I also feel grateful to all Pythoneers on Stack Overflow, and the Latex community on Stack Exchange for their endless contributions.

Abstract

Structural Health Monitoring (SHM) refers to a paradigm enabling maintenance activities to be scheduled based on the forecast of the system degradation. This forecast is primarily derived from the analysis of sensor data. This dissertation presents two novel contributions, which enhance the precision of SHM, through the combination of physics-based simulations and data-driven models.

The first contribution is a robust approach, that finds an optimal configuration of heterogeneous sensors, which maximizes the damage estimation confidence. This approach's novelty lies in three aspects. First, the mathematical definition of the damage estimation process using a Kalman filter is modularly incorporated in the optimization problem's objective function. Second, benefiting from such a definition, the Jacobian of the objective function with respect to the design variables is derived, permitting the usage of a gradient-based method. Third, within the proposed approach, a systematic algorithm for a-priori identifying the optimal number of sensors is derived. The combination of the three aspects in one approach make it customizable according to the subsequent estimation requirements and makes it applicable to complex industrial structures.

The approach's sensitivity to the number of sensors, their types, and the constraint enforcement approaches is rigorously investigated on two structures with ascending physical complexity. The proposed

approach precisely estimated the accumulated damage, and consistently surpassed existing methods in literature. The robustness of the approach to complex cases is evaluated by applying it to two industrial structures under realistic operating conditions.

The second contribution of this thesis is a hybrid model for fatigue damage estimation in fleets of engineering systems. Mainly two novelties are presented in this model. The first novelty is the unprecedented utilization of physical degradation models for fleet estimation. This yields an interpretable damage estimation model, in comparison to a conventional purely-data-driven model. The second novelty is the robust transferability of physics-based degradation models of one engineering system to other unidentical systems. This allows more accurate damage estimation, even where only limited physics-based models and operation data are available.

The hybrid model utilizes the availability of scarce, yet accurate, physics-based degradation models, and combines them to approximate the degradation behaviour of other homogeneous structures in a fleet. The combination is performed by a data-driven weighted-mean filter. Based on the extent of similarity between fleet structures, and the current operating conditions, a data-driven algorithm assigns the suitable weights to each physics-based model.

The model is tested on a fleet of beams, which included a deviant range of damage severities. To assess its robustness, the constituents and parameters of the hybrid model are vigorously varied between nominal and extreme cases. To further evaluate the applicability in an industrial environment, the model is evaluated on a realistic fleet of servomotors. On both use-cases, with only a few physics-based models, and very limited operation data, the hybrid model quantified accurately and precisely the fatigue damage accumulation in other structures of the fleet.

Zusammenfassung

Structural Health Monitoring (SHM) bezieht sich auf ein Konzept, das es ermöglicht, Wartungsaktivitäten auf der Grundlage der Prognose der Systemdegradation zu planen. Diese Prognose wird aus der Analyse von Sensordaten hergeleitet. In dieser Dissertation werden zwei neuartige Beiträge vorgestellt, die durch die Kombination von physikalisch-basierten Simulationsmodellen und datengetriebenen Modellen die Präzision des SHM erhöhen.

Der erste Beitrag ist ein Ansatz, der eine optimale Platzierung heterogener Sensoren findet, die die Zuverlässigkeit der Schadensabschätzung maximiert. Die Neuheit dieses Ansatzes liegt in drei Aspekten. Erstens ist die mathematische Definition eines Kalman-Filter-basierten Schadensabschätzers modular in die Zielfunktion des Optimierungsproblems integriert. Zweitens wird aus einer solchen Definition die Jacobian der Zielfunktion in Bezug auf die Designvariablen abgeleitet, was die Verwendung einer gradientenbasierten Optimierungs-Methode ermöglicht. Drittens wird im Rahmen des Ansatzes ein systematischer Algorithmus zur a-priori-Ermittlung der optimalen Anzahl von Sensoren hergeleitet. Die Kombination der drei Aspekte in einem Ansatz ermöglicht die Anpassung der Formulierung speziell auf komplexe industrielle Strukturen anwendbar.

Die Sensitivität des Ansatzes in Bezug auf die Anzahl der Sensoren, ihre Typen und die Ansätze für Erzwingung der Nebenbedienungen

wird an zwei Strukturen mit aufsteigender physikalischer Komplexität streng untersucht. Darüber hinaus wird die Robustheit des Ansatzes für komplexe Fälle bewertet, indem er auf zwei industrielle Strukturen unter realistischen Betriebsbedingungen angewendet wird. Der Ansatz schätzt die akkumulierten Schäden genau ab und übertrifft die in der Literatur vorhandenen Methoden.

Der zweite Beitrag dieser Arbeit ist ein hybrides Modell zur Abschätzung von Ermüdungsschäden in Flotten von technischen Systemen. Haupsächlich werden zwei Neuheiten bei diesem Modell vorgestellt. Die erste Neuheit liegt bei der neuartigen Nutzung von physikalischen Degradationsmodellen zur Flottenschätzung. Dies führt zu einem interpretierbaren Schadensabschätzungsmodell im Vergleich zu einem konventionellen, rein datengetriebenen Modell. Die zweite Neuheit liegt bei der robusten Übertragbarkeit von physikalisch-basierten Degradationsmodellen eines technischen Systems auf andere unterschiedliche Systeme. Dies ermöglicht eine genauere Schadensabschätzung, selbst wenn nur begrenzte physikalisch-basierte Modelle und Betriebsdaten zur Verfügung stehen.

Das Hybridmodell nutzt die Verfügbarkeit knapper, jedoch genauer, physikalisch-basierter Degradationsmodelle und kombiniert diese, um das Degradationsverhalten anderer ähnlicher Strukturen in einer Flotte zu approximieren. Die Kombination wird durch einen datenbasierten Gewichtsmittelwertfilter durchgeführt. Basierend auf dem Ähnlichkeitsgrad der Flottenstrukturen und den aktuellen Betriebsbedingungen ordnet ein datenbasierter Algorithmus jedem physikalisch-basierten Modell die entsprechenden Gewichte zu.

Das Modell wird exemplarisch an einer Flotte von Balken getestet, die einen abweichenden Bereich von Schweregraden von Schäden enthält. Um seine Robustheit zu beurteilen, werden die Bestandteile und Parameter des Hybridmodells zwischen Nominal- und Extremfällen stark variiert. Um die Anwendbarkeit im industriellen Umfeld zu bewerten, wird das Modell an einer realistischen Flotte von Servomotoren ausgewertet. In beiden Anwendungsfällen ermöglicht das

Hybridmodell die Akkumulation von Ermüdungsschäden in anderen Strukturen der Flotte genau und präzise zu quantifizieren anhand weniger physikalisch-basierten Modellen und sehr beschränkten Betriebsdaten.

Contents

List of Figures

List of Tables

List of Acronyms

General Acronyms

SHM Structural Health Monitoring

CM Condition Monitoring

NDT Non-Destructive Testing

SPC Statistical Process Control

RUL Remaining Useful Life

FEA Finite Element Analysis

FE Finite Elements

BC Boundary Condition

FEM Finite Elements Method

XFEM Extended Finite Elements Model

LENS Laser Engineered Net Shaping

AM Additive Manufacturing

WT Wind Turbine

CBM Condition-Based Maintenance

PdM Predictive Maintenance

PsM Prescriptive Maintenance

PHM Prognosis and Health Management

AI Artificial Intelligence

ML Machine Learning

NN Neural Network

CNN Convolutional Neural Network

RNN Recurrent Neural Network

SVM Support Vector Machine

DOF degree of freedom

PDF Probability Density Function

CDF Cumulative Density Function

KDE Kernel Density Estimate

FFT Fast Fourier Transform

Acquisition and Sensing

OSP Optimal Sensor Placement

EfI Effective Independence

SNR Signal-to-Noise Ratio

MAC Modal Assurance Criterion

POD Proper Orthogonal Decomposition

DMD Dynamic Mode Decomposition

SVD Singular Value Decomposition

ILP Integer Linear Programming

SLSQP Sequential Least Square Quadratic Programming

MMA Method of Moving Asymptotes

S-DOF Single Degree of Freedom

Prognosis and Decisions

SWT Smith-Watson-Topper

GCF Giga-Cycle Fatigue

HCF High-Cycle Fatigue

LCF Low-Cycle Fatigue

CDM Cumulative Damage Model

MoE Mixture of Experts

moSAIc Simulation Models and Artificial Intelligence Combined

RMS Root Mean Square

MSE Mean-Square Error

MARE Mean Absolute Relative Error

ARE Absolute Relative Error

SVR Support Vector Regression

MLP Multi-Layer Perceptron

RF Random Forest

ERTR Extremely Randomized Trees Regressor

List of Symbols

Dimensions and Sizes

n_t	Number of time steps
n_d	Number of system degrees of freedom
n_p	Number of system inputs
n_o	Number of system outputs
n_x	Number of system states
n_m	Number of system modes
n_c	Number of optimization constraints
n_s	Number of sensors
n_s^*	Number of optimal sensors
$n_{\boldsymbol{\sigma}}$	Number of cyclic loads
n_{iter}	Number of optimizer iterations

Fundamental Theory

$\mathbf{M}$	Mass matrix
$\mathbf{V}$	Viscous damping matrix

$\mathbf{K}$	Stiffness matrix
$\mathbf{z}$	Physical/Cartesian displacements
$\mathbf{q}$	Modal displacements
$\mathbf{f}$	Force vector
$\mathbf{x}$	State vector
$\mathbf{u}$	Input vector
$\mathbf{y}$	Output vector
$\mathbf{A}$	State matrix
$\mathbf{B}$	Input matrix
$\mathbf{C}$	Output matrix
$\mathbf{D}$	Transmission matrix
$\mathbf{L}$	Input mapping matrix
$\mathbf{\Phi}$	Displacement modes matrix
$\mathbf{\Psi}$	Strain modes matrix
$\mathbf{\Omega}$	Eigen values matrix
$\mathbf{\Xi}$	Modal damping matrix
d^w	Process noise
$\mathbf{Q}$	Process noise covariance
d^v	Measurement noise
$\mathbf{R}$	Measurement noise covariance
$r_{\boldsymbol{\varepsilon}}$	Measurement variances of strain gauges
$r_{\ddot{z}}$	Measurement variances of accelerometers

$\boldsymbol{\varepsilon}/\boldsymbol{\varepsilon}_{kl}$	Strain tensor / kl^{th} tensor component
$\boldsymbol{\sigma}/\boldsymbol{\sigma}_{ij}$	Stress tensor / ij^{th} component
$\boldsymbol{\zeta}/\boldsymbol{\zeta}_{ijkl}$	Elasticity tensor of the material / $ijkl^{th}$ component
$\lambda,\ \mu$	Lame constants
N_i	Number of load cycles
N_f	Fatigue limit, maximum number of load cycles
D	Instantaneous fatigue damage
D_a	Accumulated fatigue damage
E	Young's Modulus
$\boldsymbol{\sigma}_f$	Material fatigue strength coefficient
b	Material fatigue strength exponent
$\boldsymbol{\varepsilon}_f$	Material fatigue ductility coefficient
c	Material fatigue ductility exponent
$\boldsymbol{\sigma}_a$	Stress amplitude
$\boldsymbol{\sigma}_e$	Effective (mean-corrected) stress
$\mathbf{G}$	Optimal Kalman filter gain matrix
$\mathbf{P}^{x0}$	Covariance matrix of the initial state estimation error
$\mathbf{P}^{x-}$	Covariance matrix of the a-priori state estimation error
$\mathbf{P}^{x+}$	Covariance matrix of the a-posteriori state est. error
$\mathbf{P}^{y}$	Covariance matrix of the output estimation error
$\mathbf{P}^{D}$	Covariance matrix of the damage estimation error

Acquisition and Sensing

$\boldsymbol{\beta}$	Sensor selection matrix - relaxed
$\tilde{\boldsymbol{\beta}}$	Sensor selection matrix- binary
$\boldsymbol{\beta}^*$	Selection matrix of the optimal sensor configuration
$\mathtt{SG}$	Strain gauge
$\mathtt{Accel}$	Accelerometer
$\lambda_{\mathbf{L}}$	Lagrange multiplier
$\lambda_{\mathbf{P}}$	Penalty factor
$f^{\boldsymbol{\beta}}$	Sensor selection variable mapping function
$\boldsymbol{\Psi}$	Cumulative Density Function of a Gaussian distribution
$\boldsymbol{\psi}$	Probability Density Function of a Gaussian distribution
$\bar{\boldsymbol{\beta}}$	Selection variable relaxation threshold
$\mathbf{H}$	Gramian (Gram matrix) of a set of vectors
ϵ_0	SVD reduction tolerance

Prognosis and Decisions

n_a	Number of fleet assets
n_b	Number of basis models
n_D	Number of dynamic features
n_C	Number of static (catalogue) features
n_{D+}	Number of important dynamic features
n_{C+}	Number of important static features
$\mathring{x}$	Fleet asset

$\mathring{\mathcal{X}}$	Set of fleet assets
$\mathring{u}$	Asset input
$\mathring{\mathbf{u}}$	Vector of asset inputs
$\mathring{y}$	Asset output
$\mathring{\mathbf{y}}$	Vector of asset outputs
x	Asset model
$\mathcal{X}$	Set of fleet models
u	Model input
$\mathbf{u}$	Vector of model input
y	Model output
$\mathbf{y}$	Vector of model outputs
c	Model static feature
$\mathbf{c}$	Vector of model static features
x^{P}	Physics-based model
$\mathcal{X}^{\mathsf{P}}$	Set of physics-based models
x^{D}	Data-driven model
$\mathcal{X}^{\mathsf{D}}$	Set of data-driven models
x^{H}	Hybrid model
$\mathcal{X}^{\mathsf{H}}$	Set of hybrid models
x^{M}	*moSAIc* model
$\mathcal{X}^{\mathsf{M}}$	Set of *moSAIc* models
x^{ϕ}	Undefined model

$\mathcal{X}^{\phi}$	Set of undefined models
x^{B}	Basis model
$\mathcal{X}^{\mathsf{B}}$	Set of basis models
$\boldsymbol{\alpha}$	Model weight
$\boldsymbol{\alpha}$	Vector of model weights
$\boldsymbol{\alpha}^{\mathsf{B}}$	Basis model weight
$\boldsymbol{\alpha}^{\mathsf{B}}$	Vector of basis model weight
y^{B}	Basis model output
$\mathbf{y}^{\mathsf{B}}$	Vector of basis model outputs
$\boldsymbol{\alpha}^{\mathsf{M}}$	*moSAIc* weight
$\boldsymbol{\alpha}^{\mathsf{M}}$	Vector of *moSAIc* weights
y^{M}	*moSAIc*'s output
$\mathbf{y}^{\mathsf{M}}$	Vector of *moSAIc*'s outputs
$\mathcal{L}$	Model loss
$\mathcal{L}^{\mathsf{M}}$	*moSAIc* loss
$\mathcal{L}^{\mathsf{G}}$	*moSAIc*'s Multi-Layer Perceptron's loss
$\tilde{y}^{\mathsf{B}}$	Assembled basis model output
$\tilde{\mathbf{y}}^{\mathsf{B}}$	Vector of assembled basis model outputs
$\tilde{\boldsymbol{\alpha}}^{\mathsf{M}}$	Assembled *moSAIc* weight
$\tilde{\boldsymbol{\alpha}}^{\mathsf{M}}$	Vector of assembled *moSAIc* weights
$\tilde{\mathbf{y}}^{\mathsf{M}}$	Vector of assembled *moSAIc*'s outputs
$\mathcal{F}$	Set of features

$\mathcal{F}^{\mathsf{C}}$	Set of static features
$\mathcal{F}^{\mathsf{U}}$	Set of dynamic features
$\mathcal{P}$	Set of parameters
$\mathcal{P}^*$	Set of optimal parameters
$\mathcal{P}_G$	Set of Multi-Layer Perceptron (MLP) parameters
$\mathcal{P}_G^*$	Set of optimal MLP parameters
$\mathcal{T}$	Training set
$n_{\mathcal{T}}$	Number of training set samples
$\mathcal{E}$	Test (Evaluation) set
$n_{\mathcal{E}}$	Number of test set samples
$\mathcal{L}^{\mathsf{G,MLP}}$	Loss of a standard MLP
$\mathcal{L}^{\mathsf{G,M}}$	Loss of *moSAIc*'s MLP
$\mathbf{W}$	MLP's weight matrix
$\mathbf{b}$	MLP's biases vector
f^{act}	MLP's activation function
$a_b^{[l]}$	Output of the b^{th} perceptron in the $[l]^{th}$ layer
$z_a^{[l]}$	Input of the a^{th} perceptron in the $[L]^{th}$ layer
$w_{ab}^{[l]}$	Contribution of the $a_b^{[l]}$ to $z_a^{[l+1]}$
$\Delta w_{ab}^{[l]}$	Update of $w_{ab}^{[l]}$
y^{O}	Target model output
$\boldsymbol{\alpha}^{\mathsf{O}}$	Target *moSAIc* weight
$\boldsymbol{\alpha}^{\mathsf{G}}$	Actual *moSAIc* weight

Part I

Prologue

1 Introduction

Engineering structures are abundant in a plethora of our daily aspects and activities. Many of those are designed to function within a range of environmental and operating conditions, usually pre-defined by the responsible engineers at the design phase. Operating limits dictate influential decisions in the design of structures, such as material choice, manufacturing processes, as well as joint and support mechanisms. In practice, engineers rely on experimental validation and modelling techniques to assess the ruggedness of their designed structures, or prototypes. It is well-known that these two are neither free of approximation nor assumptions about the structures investigated. Typical engineering assumptions could be, for instance, the consistency of applied loading, degradation behaviour of structures, linear response to external environmental excitations, flawless material distribution, and de-coupled interaction between various physical phenomena within the structure.

Many structures have a prevailing complex operation nature, due to the incorporation of multiple coupled physical phenomena in their construction, or due to the indeterministic operation conditions they might undergo; the fact that makes it difficult to accurately model or reproduce their operation condition through experiments. Accordingly, the true *picture* of the structure is only obtained once it is deployed in its operating environment and begins to run. Typically, structures are subject to operation scenarios, unforeseen during the design stage, due to impacts or loads exceeding design limits, as well as due to manufacturing tolerances and deviations. Such *anomalies* could eventually be the root cause of faults in the structure, that could go undetected as their occurrence might be unexpected in

the first place. Undiagnosed, and hence un-remedied, faults could cause severe damages to the structure, leading to unexpected down-time, overwhelming maintenance and overhaul processes, economic devastations, and, in the worst cases, even catastrophic fatalities.

Engineers have been attempting to overcome unfavourable repercussions of unanticipated events during operation in many ways. One classical solution, which is accepted in many engineering domains, is to *over-design* structures, practically by increasing the safety limit upon which material selection and design decision are based. This is often the case in designing residential buildings, for example, where the human safety aspect is decisive. Employing this approach yields a heavy and an expensive structure, which, although more reliable during operation, is still prone to long-term degradation. Additionally, this solution is not robustly applicable to engineering applications, where weight is a critical design parameter, such as aerospace applications and light-weight structures [WDB04].

This has led engineers to design structures with tighter safety factors and rely on maintenance activities during the life cycle of their systems to ensure their performance and integrity as an alternative to over-designing. According to the DIN[1] 13306: 2008 norm, maintenance is defined as

> "*A combination of all technical, administrative and managerial actions during the life cycle of [a system] intended to retain it in, or restore it to, a state in which it can perform the required function.*" ([DIN08])

Before proceeding further with discussion on maintenance, it is essential to clearly define some terms which will be repeatedly used throughout the following paragraphs, and the entire text.
A *system* refers to an intended arrangement of interrelated items or components to achieve a defined function through mapping one

[1] DIN: Deutsches Institut für Normung (translation: German Institute for Standardization)

or more inputs to one or more outputs [Bir13; FB09]. A system is considered in a *healthy condition* as long as it is providing its desired function. For instance, a car is a system, whose function is converting energy resulting from burning fuel to motion.
A system *failure* is the event impeding the system from delivering its desired function. Occurring system failures are caused by faults or defects existing in the system.
A *fault* or a *defect* is a non-inherent deviation of a system's property or parameter from its nominal condition [Ise06]. Unfortunately, failures and faults are erroneously interchangeably used, thus it should be clearly distinguished that a failure is a time-related event occurring on the system, while the fault is an intrinsic system state [Bir13]. In the car example, a failure could be a broken wheel axle, while the fault could be a surface crack which has led to it.
Additionally, *damage* is defined as the temporal event, associated with occurring intentional or unintentional alteration in system properties or parameters, which adversely affects the current or future performance of this system [FW12]. Note that damage refers to an event or a sequence of events that might precede failure. In the car example, while a broken axle is the failure, damage is the growth and propagation of the root-cause surface crack. The scenarios by which a fault or a defect progresses such that it leads to a failure is known as the *failure mechanism*. For instance, the opening modes of a crack subjected to a given stress state are a failure mechanism [Sao17].

That being said, maintenance can pragmatically be perceived as the combination of actions aiming to understand and anticipate the failure mechanisms and act accordingly in order to ensure that the system can perform its desired function [GMZ16].

Historical Overview

2

The concept of *Prognosis and Health Management (PHM)* was first introduced and defined in the medical field as "*the prediction of the future course and outcome of disease processes, which may either concern their natural course or their outcome after treatment*" ([AHL01]). Motivated by the same concept as medical prognostics, several philosophies of maintenance have been adopted and developed over the last decades, see Figure 2.1, until a prognostics paradigm has been reached for engineering applications.

2.1 Evolution of Maintenance

Before 1960s, maintenance was carried out only in a *corrective* fashion. Corrective maintenance is "*carried out after fault recognition and is intended to put an item into a state in which it can perform a required function.*" Such maintenance philosophy could be accepted for applications where failures are neither critical, nor expensive, nor life-threatening. Gradually, this philosophy was complemented and eventually substituted by the *preventive* maintenance philosophy, defined as a "*maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the operation of an item.*" [DIN08].

Two philosophies can be observed under preventive maintenance: *time-based* and *condition-based*. *Time-based maintenance* is when maintenance interventions are carried out at pre-defined intervals. One habitual daily life example of such is the maintenance and inspection schedule for vehicles, performed at mileage- or years-based schedules. It is speculated by system designers, that with such a

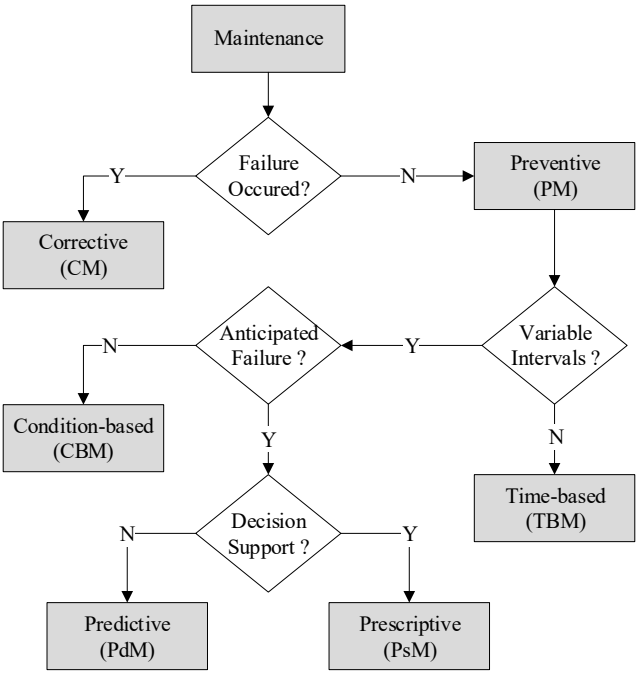


Figure 2.1.: Taxonomy of maintenance philosophies. Adapted from [GMZ16].

routine, critical components of the system would be maintained in a *healthy* condition. The execution of such a maintenance routine doesn't require any monitoring of the system during operation per sé, keeping the cost of running the system as low as possible. Yet it could incur higher maintenance costs due to unnecessary service activities performed on components far from their critical lifetime. In contrast, *condition-based maintenance* is defined as "*a preventive maintenance based on performance and/or parameter monitoring and the subsequent actions*." This maintenance strategy, thus, requires monitoring of the system outputs (for instance vibrations, temperature, etc.) and analysing its responses. It aims to detect existent

anomalies in system outputs, and infer from them the presence of a fault, hence calling for a maintenance intervention based on the *current* state of the system.
What remains not yet addressed is the future projection, to estimate the operation time until the failure event, or what is referred to as the Remaining Useful Life (RUL). This led to the evolution of the *Predictive Maintenance (PdM)* philosophy, defined as "*a condition-based maintenance, carried out following a forecast derived from the analysis and evaluation of the significant parameters of the degradation of the item.*" [DIN08]. Among all maintenance philosophies, this latter one has the highest economic benefit in terms of reduction in maintenance costs, despite the high initial cost required for data acquisition and management systems.

A note on the notion: In many academic and industrial consortia, the term PHM is often used, and for the first sight, it seems like it is often interchanged with Condition-Based Maintenance (CBM) and PdM. As a matter of fact, there is no clear, standard definition proposed for the PHM notion. Nonetheless its primary processes are similar to those of the CBM [GMZ16]. In some context, PHM is distinguished from maintenance as being the discipline concerned with the integrated, system-level health state management, while maintenance is focused on applying similar procedures at a component level. According to the PHM Society, PHM is not limited only to engineering activities, but its research topics extend to Meteorology, Climatology, Decision Policy, Finance, and Economy [Lee+14].
As our work is concerned with the Predictive Maintenance (PdM) activities in engineering systems at component level, we would refer to this maintenance regime and its associated activities as PdM when presenting our ideas. For the sake of consistency in terminology, when presenting ideas of other scientists, such as in Part II, we will refer to all activities associated with health state management as PdM, unless specified explicitly that the higher-level health management discipline is meant; in which case PHM will be used.

2.2 Socio-Economic Aspects

The benefits of adopting a PdM philosophy are motivating many business owners towards enforcing them as acting policies. Unexpected downtimes as well as corrective actions cost companies a fortune! With such policies, reduction in the frequency of system downtimes due to failures could be achieved, and intervals between maintenance activities could be extended, hence maximizing the system availability. This translates in longer periods during which a system is generating revenue rather than draining resources.
Furthermore, the costs of spare parts purchases and stock management would be shredded significantly. According to [FW12], today the annual cost of maintenance of wind turbines, for instance, is estimated to be €20k per Megawatt, and a regular overhaul performed every 15 years costs up to €200k. With adequate maintenance, regular maintenance and overhauls could be spared, unless a failure is prone to occur, leading to a remarkable cut in both figures.
On the other hand, for companies producing high-capital-expenditure products, it is more profitable and appropriate to have a leasing-based revenue under their business model as opposed to an asset-sale-based revenue [Afu04]. Such business models would adhere the company to taking responsibility of the service and maintenance of its equipment. As PdM would allow systems to confidently run in the field up to their utmost load-bearing capacity, a financial advantage could be gained, as well as a more realistic pricing strategy, that could be based on the equipment's lifetime consumption instead of basing it on the leasing period [FW12].
The reliable field performance, guaranteed quality and availability of systems, and lean maintenance plans promise a positive impact on the customer satisfaction and business reputation; two precious intangible assets granting an edge in a competitive market [FMV00].

In addition to the economic value, adequate maintenance functions have significant social and environmental implications as well. Firstly, the failure anticipation of critical system components foresees indus-

trial risks and helps avoiding accidents which can be detrimental to the environment and the humans. In the case of residential and industrial complexes, especially in seismically active regions, with the aid of system monitoring techniques, the readiness and safety of buildings for reoccupation could be assessed with tighter uncertainties. Accordingly, this inflicts satisfactory work conditions and leveraged human safety. Furthermore, monitoring and PdM allow a reliable operation of some structures, e.g. bridges and urban infrastructure, beyond their initial design lifetime, since either their replacement or phase-out inflicts an economic and functional burden [FW12].

2.3 Realizing a Maintenance System

Motivated by these socio-economic implications, many industrialists and engineers have directed their attention towards the development and deployment of PdM policies [MSI08].

In [Int06], a standardized architecture for CBM is proposed, see Figure 2.2. Being a derivative of CBM, the same architecture could be adopted for PdM.

The proposed architecture consists of the following seven functional modules:

1. *Data acquisition*: This module provides the system with digital data acquired from sensors or transducers, as well as meta-data input by operators, e.g. inspection logs.
2. *Data processing*: At this level, the sensor signals are processed in order to extract the important features required for state estimation, fault identification and damage assessment.
3. *Condition assessment*: This module compares the extracted features against predefined or expected benchmarks in order to detect any changes in the system state.

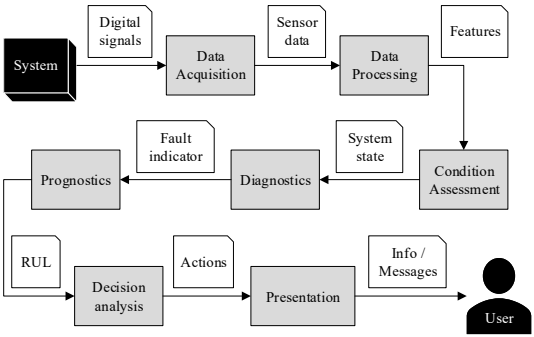


Figure 2.2.: Architecture of CBM according to [Int06]. Grey boxes indicate the seven function levels of the architecture, and the white boxes indicate the data transferred between every two levels.

4. *Diagnostics*: On the basis of the detected state, this module determines whether the system has undergone a damage event and identifies probable causing faults. This module requires thorough knowledge of the system's components, their interactions, and the system's operating conditions.

5. *Prognostics*: This module relies on the accumulated results from the previous modules, to forecast the future state of the system, and its RUL. This prediction generally comes with a certain level of uncertainty as it is based on extrapolated future conditions.

6. *Decision analysis*: Based on the predicted RUL, this module recommends control / maintenance actions to be taken to ensure that the system accomplishes its defined function.

7. *Presentation*: This module collects the outputs and conclusion from all the previous layers for conveying them in a comprehensible fashion to the users through the so-called human-machine interface.

Within industrial domains, as well as academic communities, the terms Structural Health Monitoring (SHM) and Condition Monitoring (CM) are used to refer to the set of tasks comprised by the first five layers. The distinction between the two terms lies in that SHM is rather common for civil and aerospace applications, while CM is used when referring to applications in the domain of rotary and reciprocating machinery [Tid+16]. Both SHM and CM could be applied online, i.e. parallel to the operation of the system without disconnecting it, or offline, i.e. after the system is detached from the operation pipeline.

One way to enhance the accuracy of the monitoring systems is through integrating the operating industrial machinery in the diagnostics loop. This can be achieved through recycling simulation models and upgrading their design-phase features to operation-parallel capabilities, i.e. transferring the classical simulation model created for design and manufacturing purposes into a *digital twin* representing the asset during its operation progression. In the past couple of years, the concept of the *digital twin* has been widely acknowledged in industrial and academic contexts and has been regarded by Gartner among the top strategic technologies of the near future [GV17; Gar19].
Boschert et al. defined the digital twin as

> "*The description of a component or a system, achieved through interlinking its digital artefacts, such as engineering and operational data, via simulation models.*" ([BHR18])

Said simulation models evolve along with the evolution of the system throughout its lifecycle. The vision of the digital twin refers to a comprehensive physical and functional description together with all available operational data of a component, product or system.

In its most generic conception, the digital twin doesn't only hold engineering and functional information of its physical correspondent,

but, additionally, stores its operation history, maintenance activities, its relation to other assets in a larger operating system e.g. the relation between different electric motors in a fabrication plant, as well as information about its subsystems, e.g. information related to the bearings and windings in one electric motor. Herein, this conceptualization of simulation-hardware relationship forms a very profound foundation for seamless SHM and CM. Some holistic perceptions of digital twins in SHM and CM contexts can be found in [BR18; BKR19; Kha+19b]

Research Structure

3

3.1 Problem Statement

When landscaping the scientific work conducted on the five modules of Structural Health Monitoring (SHM) and Condition Monitoring (CM), numerous contributions could be found in the area of signal processing and analysis. Throughout the past five decades, many methods have been proposed to extract the relevant features from both the signal's time and frequency domains to be used in system state identification. These methods have been developed to robustly handle clean as well as noisy, trendy and non-stationary signals.
With respect to diagnostics, a wide range of approaches using output-only-based estimators, linear and non-linear filtering methods, and machine learning algorithms have been developed to identify faults in the system and reconstruct its states and inputs. To the time of writing, new developments are continuously being undisclosed to the academic society addressing the short-comings or enhancing the performance of existing methods.

In contrast, the research on the choice and configuration of sensor networks is not gaining much of attraction. Despite the role optimal sensor placement plays in the accuracy and precision of system states estimation and faults identification, fewer contributions are being published in this area. Browsing the publications in this field released in the last decade, few thousand journal articles and proceedings could be found in the domain of sensor placement, while a ten-fold is published addressing state estimation, state reconstruction and input estimation.

Surveying the sensor placement literature, the following research niches are found:

1. Many contributions address sensor placement for SHM applications, where the underlying system models are relatively simple. For such simpler structures, the scalability of the proposed sensor placement algorithm is not a priority. However, for complex industrial cases with larger system models, the concerns about the algorithmic performance and robustness arise when employing many of the existing methods, e.g. those based on genetic algorithms.
2. Scarce sensor placement strategies consider the downstream condition assessment and diagnostics method employed. Instead, they assume a generic metric representative-enough of the subsequent estimation precision as an objective to find the optimal sensor placement.
3. Only few contributions address the problem of multi-type sensor configurations, whereby sensor are positioned in a configuration that provides accurate prediction of the structure states through fusing signals from different sensor types.

To the time of writing, no known scientific work has been found that has entirely fulfilled the three issues. The shortcomings highlighted within the three niches are addressed in this work, with the goal to

Develop a multi-type sensor configuration approach, robustly applicable to complex structures, that considers the subsequent diagnostics method, and modularly incorporates in it finding the optimal sensor configuration.

Herein, the following steps are employed to achieve this goal:

- Derive the optimization problem based on the assumption that a Kalman filter is to be used to reconstruct the system states and estimate the fatigue damage based on the sensor

data. The objective would then be to minimize the covariance of the filter's error estimate. Such a formulation accounts for the approach employed for the subsequent diagnostics step.

- Formulate the optimal sensor placement problem such that it optimizes the positioning of strain gauges and accelerometers simultaneously, not sequentially. These two sensors are not only chosen because they are common in SHM and CM applications, but also because they measure signals of different amplitudes and encompass variant sensor noises. This would pose a realistic challenge to the robustness of the proposed approach.
- Employ a gradient-based optimization solution method for the sensor placement problem, capable of scaling well with the size of the problem. This method doesn't just outperform the gradient-free genetic algorithms but can also robustly incorporate as many constraints on sensor types, permitted placement domains and budget.

With respect to the work on the prognosis module, two clusters of Remaining Useful Life (RUL) estimation paradigms could be found; approaches estimating qualitative RUL of the structures, e.g. percentage consumed of the system's lifetime, and approaches estimating the physical severity of the underlying failure mechanism, e.g. propagation of the damage-driving crack.
The former relies dominantly on the statistical modelling or machine learning algorithms and necessitates the presence of operational data to build a representative surrogate for the degradation process. It retains the edge of being more easily developed and being extensible to similar systems. In contrast, the latter requires meticulous understanding and elaborate formulation of the principles of material degradation and urges the use of physics-based models of the structure. The presence of such a model allows quantifying the actual degradation state of the system, let it be a crack length or the number of load cycles until failure.

Over the past decades, both paradigms have been and are still being intensively developed. One advantage of the data-driven models is the easier transferability from one system to any other adhering to the same degradation behaviour. This makes these models particularly attractive when monitoring fleets of complex systems – a typical requirement in industrial fields. This comes at the expense of losing, to an extent, the interpretable and quantifiable damage estimate provided by the physics-based model.

Up to the point of writing this document, the exploitation of physics-based models in fleets prognosis has received no scientific contributions except the collaborative work of the author of this thesis. Building upon those early works, in this work, a stride through this deserted path is made with the goal to

> *Develop a prognosis approach, seamlessly applicable to a fleet of systems, that principally bases the damage estimation on the physical description of the degradation process and utilizes operation data only to leverage its robustness.*

Herein, the following steps are employed to achieve this goal:

- Build an elaborate damage accumulation model based on simulating the material fatigue under cyclic loading. The detailed model is limited only to those system retaining the highest similarity to other systems in the fleet, and hence can approximate their degradation behaviour.
- Develop a data-driven algorithm that assesses the similarity between the various fleet systems. Upon this assessment, it decides on the adequate combination of the built physics-based degradation models, such that the most precise approximation of the degradation behaviour is achieved.
- Train the data model on a multitude of operating conditions to enhance its accuracy and robustness over a wide spectrum of

nominal as well as extreme system loading conditions applied to any arbitrary system within the fleet.

Another focus point of this work is to close the gap between the two terminals of the Predictive Maintenance (PdM) pipeline through accounting for the downstream diagnosis and prognosis steps while designing and optimizing the in-feed sensor configuration. Such a closure could be potentiated by employing a physics-based simulation model to forecast the system responses and degradation under prospective operating conditions. This given, the sensor configuration is planned for the ultimate goal of the PdM. The gap is further bridged by employing the very same simulation models during the prognosis step. Since the physics-model encompasses the system's degradation behaviour, the inferred features from the sensor signals, and reconstructed states resulting from the diagnosis step are input to the models to assess the damage state. With this, the simulation model would then be fully utilized throughout the entire PdM pipeline.

3.2 Outline of the Work

Having elaborated the problems under investigation and the research goals of this work, the outline of the document follows in a similar manner.

- *Part II (State of the Art)*: An overview of the main academic and industrial contributions in the domain of PdM is summarized. Following, a detailed survey of the approaches and algorithms published in the area of Optimal Sensor Placement is given. Afterwards, the scientific literature in the field of damage and RUL estimation is outlined.
- *Part III (Fundamental Theory)*: The required knowledge about physics-based damage estimation and state estimation using a Kalman filter is introduced. This serves as a pre-requisite the-

oretical introduction, upon which the consecutive two chapters are built.

- *Part IV (Acquisition and Sensing)*: This chapter addresses the first goal. The role of signal acquisition and sensor configuration in the context of SHM and CM is pointed out. Also, the different motives triggering the deployment of signal acquisition systems are presented and contrasted. This is followed by a proposition of a novel gradient-based algorithms for multi-type sensor configurations. The optimization problem is first formulated, then its constraints are outlined, followed by a discussion on their treatment approaches. Furthermore, numerical issues of the algorithm are addressed. Finally, the constituents of the proposed algorithm are thoroughly investigated on two benchmark examples, followed by a comparison to existing methods.
- *Part V (Prognosis and Damage)*: This chapter addresses the second goal. In this chapter, the use of physics-based models to improve the degradation process modelling of similar systems is presented. First, the architecture of simulation-data-driven hybrid model is presented, wherein the roles of the physics-based model and data-based surrogates are clearly defined. The sequence of simulation model preparation and operation data processing is detailed. A systematic methodology for choosing the physics-models approximating other fleet systems is explained. Finally, the sensitivity of the hybrid model's performance to the choice of these physics-based models and to the data processing is studied on a benchmark example. The chapter is concluded with a comparison between the proposed hybrid model's precision and generalization and purely data-driven approaches.
- *Part VI (Industrial Applications)* : In this chapter, the application of SHM and CM procedures is demonstrated on three industrial use cases. Herein, the strength of utilizing simulations models to implicitly fulfil SHM and CM requirements,

as well as to support data-driven models reaching a precise solution, is emphasized.

- *Part VII (Epilogue)*: A summary of the proposed work and the obtained results are presented along with the conclusion of the main findings of the underlying contribution. The open issues are remarked and prospective future developments on this work are suggested.

Part II

State of the Art

Predictive Maintenance

4

As mentioned in Section 2.1, the terms Prognosis and Health Management (PHM) and Predictive Maintenance (PdM) are sometimes interchangeably used in academic and industrial contexts. In the following paragraphs, we focus on presenting a survey of the main contributions in both fields.

In our opinion, the research in PdM could be investigated in two folds; 1) general contributions addressing PHM holistically as a concept and presenting its frameworks and architectures, and 2) contributions focusing on applications-related realizations of PdM routines and techniques to particular engineering systems.

It is important to note that this section is limited to higher-level state of the art contributions to PdM, while contributions related to the specific technologies involved in the seven elementary modules shown in Figure 2.2 are not in focus here.

4.1 Generic Contributions

Despite the circulation of the PHM concept and term in industry and academic for decades already, and the establishment of the PHM Society[1] in 2009, unfortunately, there has been no systematic framework, or a robust standard for designing and implementing the PHM technology; according to [Lee+14], the developments are mostly system-specific, which are not thoroughly explained or documented. In [FW12], the authors refer as well to the *"dearth"* of publications in Structural Health Monitoring (SHM) and Condition Monitoring

[1] `https://www.phmsociety.org/`

(CM). Nevertheless, when browsing literature, one can come across a couple of preliminary contributions, that could be considered a first seed towards realizing such a standard. These are presented in the following paragraphs.

General PHM

In 2003, members of Los Angeles National Laboratory complemented a literature review produced earlier by their lab in 1996 [Doe+96], with a broader view of statistical pattern recognition paradigm for SHM. The aim of this study was to present techniques and challenges for damage assessment, which in turn is an indicator of Remaining Useful Life (RUL), in the context of SHM. The review starts with proposing a definition of damage, followed by a breakdown of the statistical pattern recognition paradigm, which in their belief is the foundation of damage assessment. The breakdown consisted of four aspects: operational evaluation, data acquisition, feature extraction, and statistical modelling for feature discrimination. Interested readers are referred to the reference for more details about each aspect. The aim of this study was mainly classifying different contributions to the proposed paradigm.
Building upon it, as well as other resources, came the book *Structural health monitoring: a machine learning perspective* published in 2012 [FW12]. The authors, among which were contributors to [Soh+03] too, re-used the definition of damage, and the four-aspects statistical pattern recognition to propose a well-rounded overview for damage detection using novel machine learning supervised and unsupervised learning approaches. The book provides as well basic fundamentals for realising these approaches, lays essential foundations for the damage assessment problem. In addition, the four aspects of the statistical pattern recognition paradigm presented in [Soh+03] are explained in more details and complemented with applications to various use cases: a wind turbine, an aircraft, bridges and buildings.

In 2013, a book composed of a collection of articles [Inm+13] addressing a scope of PHM approaches was presented, *viz.* sensing technology and hardware, monitoring algorithms, damage and prognosis models, and concluded with use cases where SHM and CM proved to be valuable. The book is comprehensive in sense of covering a wide breadth of aspects, yet it fails to connect its elements under one hood. In other words, it doesn't present a comprehensive architecture bundling the flow of data and information across the different modules.

In [Lee+14], Lee et al. were motivated by the fact that no profound framework exists for realising PHM in system-independent way, and by the lack of a proper medium for conveying information among maintenance stakeholders. The authors presented a comprehensive overview of the on-going research efforts in the field, followed by a proposal of system methodology for improving the deployment of PHM to engineering systems. The proposed method was demonstrated on different industrial cases, where PHM procedures *viz.* critical components identification, diagnostics and prognostics algorithms selection, followed by information visualization and decision-making process are explained.

Building upon the work of [Lee+14], in [GMZ16; CMNV17], Gouriveau et al. and Chebel-Morello et al. presented a comprehensive overview and a holistic pipeline for PHM procedures, addressing monitoring, diagnosis, prognostics, data management, reliability, availability, and decision associated to the maintenance activities.
To our knowledge, this reference is by far the most profound and comprehensive source combining a robust, application-independent overview of the PHM procedure. It also provides a deep-dive into the underlying technologies of each of its module with a sufficient level of detail. The authors conclude their work with indicating the main challenges and open questions, which they see vital for the improvement of PHM. These could be summarized into 1) the need for developing a *standard* approach for signal acquisition and raw data

handling, 2) an *objective* methodology for analysis and prognosis, and 3) a *systematic* approach for the validation and verification of the conclusions and decision coming out of the PHM. Additionally, the importance of focusing on an application-independent PHM standard was strongly highlighted in their recommendations.

Further details about different approaches for fault diagnosis and health monitoring techniques can be found in e.g. [GCD15a; GCD15b; Ran11; TDP12]

Frameworks and Architecture

Looking at domain-specific frameworks, the most refined advances in maintenance technology in general could be granted to the aviation industry with their adoption and integration of maintenance and prognostics technologies since the late 1970s. In that regard, the industry has yielded practical frameworks, which have gained a wide consensus as acceptable maintenance strategies by the aerospace industrial community [FW12].

Among those strategies is the Health and Usage Monitoring System (HUMS) [McC05], which is used to diagnose faults in helicopter drive trains, engines, oil systems and the rotor system. The framework has been endorsed by the Federal Aviation Administration (FAA) and the Civil Aviation Authority (CAA) as part of an acceptable maintenance strategy.
Another framework, developed by NASA, is the Space Shuttle Modal Inspection System (SMIS) [Gry94], which was developed to identify fatigue damage in components such as control surfaces, fuselage panels and lifting surfaces using modal testing techniques instead of conventional non-destructive testing. A more detailed discussion on both frameworks, as well as others, could be found in Chapter 2.3 in [FW12].
A recent framework development is being conducted within the context of the European project "REconfiguration of CONtrol in Flight for Integral Global Upset Recovery (RECONFIGURE)", which

aims at automating the handling of abnormal and off-nominal events during flight operation [Gou+14].

In 1999, the US army released a publication presenting the funded project "Prognostics Framework", whose target was providing a holistic architecture to convey PHM information to the personnel involved in the maintenance, logistics and planning activities. Despite being a high-level, comprehensive architecture, the framework is limited to the US army Department of Defence weaponry systems and can't be considered a generalized standard solution. Additionally, the diagnosis and prognosis technologies underlying are based on a knowledge database assembled from experimental conclusions and empirical rules, without much exploitation of more sophisticated analysis of operation information through data-driven and physics-driven models.

In 2002, Hess and Fila proposed a framework for automatic logistics in aircraft maintenance procedures, supported by Condition-Based Maintenance (CBM) activities and PHM technologies. The authors addressed issues such as the architecture of the underlying information system, and technology enablers. Yet this framework is very specific to aircraft systems [HF02].

Roemer et al. proposed a prognostics framework for rotary machinery with an application on bearings of aircraft engines [Roe+13]. The framework utilizes signal analysis and physics-of-failure models to perform online update of material properties, fatigue life estimate and faults severity. The framework provides, as well, online operation feedback and decision-making support based on the in-situ RUL forecast.

A recent contribution in that direction dates to the work in [Kan+17], which proposed a framework using Industry 4.0 capabilities and cloud services for the automation of a plant-wide CBM strategy for rotating machines. In their work, Kande et al. discussed the need for hierarchically-distributed monitoring modules, connected

via communication protocols, which feed information to on-premises analytics units. Said analytics units process the information and supply decision support systems with the result, to close the loop by communicating the taken decisions to the hardware on the plant floor.

Interested readers in more details about PHM frameworks and general contributions can benefit from the discussion presented in [FW12; GMZ16; CMNV17].

4.2 Application-specific Contributions

Application-specific contributions in the field of PdM and PHM are more abundant in the literature. The remarkable work published over the last decade is summarized in the following paragraphs. The paragraphs are arranged according to the applications: rotary machinery, aerospace applications, and electrical applications. The realisation of SHM and CM technology in these fields have been and is still investigated intensively by academic institutions and industrial research centres in attempts to find solutions for relevant daily engineering problems [FW12].

Rotary Machinery

Rotary machinery is one of the fields where maintenance technologies have prospered significantly over the past eight decades, and made a huge leap from being an institutional research topic to an industrial practice [Mit07]. The first research in vibration monitoring in induction motors could be reported way back in 1930s in [Hil30].

As mentioned, the term CM is used to refer to the PHM activities associated with rotary machines. The advancement in CM focused in principle on monitoring and analysing measured signals from rotating machinery using contact and contactless sensing hardware. Among the commonly used signals are acoustic emissions, vibration signals, temperature, and current signature [Cho+18]. Researchers have

been investing huge efforts in attempts to correlate the monitored signatures to specific dynamic responses and faults occurring in the machinery.

In [Ran11], the author presents the basic signal processing techniques utilized for fault identification in both time and, more dominantly, frequency domains. In this book, the author is only concerned with fault detection and diagnosis based on vibration sensors, with an application to different rotary components, e.g. bearings, gears, reciprocating machines, and internal combustion engines. Additionally, the book presents broadly some physics-based and data-driven methods for RUL estimation.
In [Cho+18], the author highlights the advantage of combining vibration signal and current spectrum analyses for CM of induction motors, which share many common operating principles and failure signatures with other rotary machinery. [Zha+17] presents an overview of some methods used for fault detection in rotary machines using Artificial Intelligence (AI) approaches, *viz.* fuzzy logic, Neural Networks (NNs), Support Vector Machines (SVMs), and adaptive neuro-fuzzy inference system.

Rotary machinery are complex systems which are prone to failure due to various possible root causes. The nature of loading on rotary machinery is complex as well, comprising mechanical, electrical, and thermal loads, along with environmental operating conditions. The complexity of the system and its loads results in failures as complex and makes back-tracing their root-causes a challenge for maintenance engineers and researchers in CM field.

Faults in rotary machinery can be, on one side, categorized into internal and external faults, that are further classified into mechanical, electrical and environmental faults. From another perspective, faults may be classified into the rotor, stator, bearing and other mechanical faults based on its location in the equipment, as shown in Figure 4.1. The different faults and their respective root-causes are detailed in [BG12]. The book presents a cross-section documentation of possible

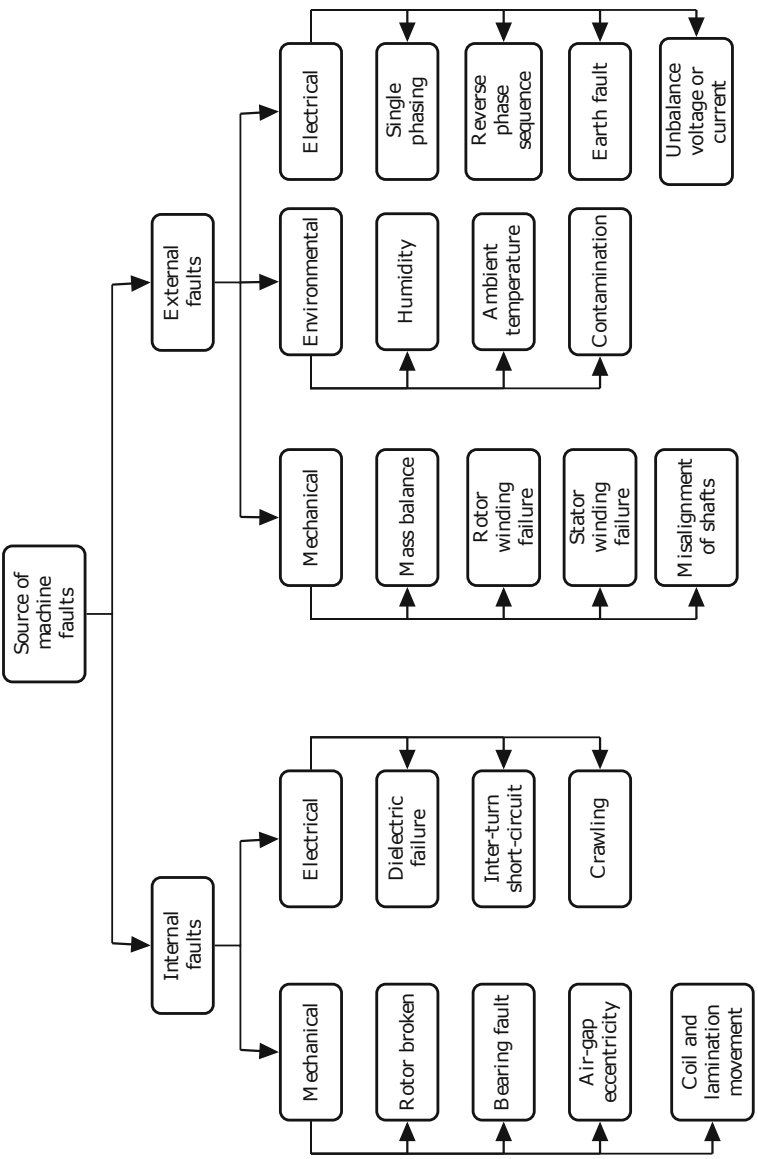


Figure 4.1.: A map of rotary machinery faults [Cho+18]

faults occurring to different subcomponents of a rotary system and illustrates systematic approaches for troubleshooting them based on non-destructive tests, and signals analysis from different sources. The author summarizes for each subcomponent its most prominent root causes of failures and its symptoms that could be detected by monitoring machinery responses.
In [Tol+12], Toliyat et al. present an overview of the common faults in both induction and synchronous motors, including bearings faults, stator faults, and rotor faults. The authors present as well different approaches for building electro-magnetic and structural models for the respective machinery in both healthy and faulty status. With respect to fault diagnosis, techniques based on signal processing, both in time and frequency domain, techniques based on physics models, and techniques based on pattern-recognition in measurements are explained and applications of them on various faults are demonstrated.
According to [O'D+87], the major causes of rotary machinery failure are due to bearings and stator windings, contributing to almost 70 % of failures. This explains the focus of researchers on those two subcomponents in particular, trying to understand their failure modes, identify their root causes, and suggest condition monitoring solutions to mitigate them.

Bearings Faults Bearings faults alone contribute to around 40 % of the root-causes of failures in rotary machinery [O'D+87; Cho+18]. Information about bearing faults could often be related to their fundamental frequencies: inner race frequency, outer race frequency, ball-spin frequency and ball-pass frequency. These frequencies are dictated by both the geometry and the running speed of the bearing [SD+19].

Bearings typically fail due to the development of high contact pressures between its components, i.e. inner and outer race, cage, rolling

elements[2]. High stress concentrations develop due to high operation loads, presence of external faults, e.g. misaligned or bent rotor, unbalance, etc., and/or presence of local or distributed bearing defects, e.g. high surface roughness, debris and inclusions in the lubricant film, etc.
Consequently, local surface cracks and pits start developing on the surfaces of the bearing's components, which as operation goes on evolve into spalls and distributed wear [ETJ14]. A survey conducted by FAG[3], a world-wide bearing provider, presents a wide scope of bearing damages, their causes, inspection procedures and their corrective and preventive remedies [AG03].

Condition Monitoring of bearings requires understanding and modelling the physics of failure, which is a very challenging task, as it inherently contains various interacting physical phenomena and sources of non-linearity. Many dynamic simulation models were and are being developed and researched over the past seven decades, starting with the contribution of [Pal59] in 1959. The complexity of the models developed over the years to include non-linearities and time-variant dynamics, such as speed-related hydrodynamic effects, clearances, surface roughness and localised faults. Various simulation technologies, such as Finite Element Analysis (FEA) and multi-body dynamics were also studied to model bearings operational dynamics and fault characteristics. A comprehensive review of the development of bearings simulation advancements could be found in e.g. [ETJ15].

Besides the development of simulation models, research has been intensively conducted in monitoring and signal analysis of bearings' response signature. Common monitoring techniques for bearing faults include vibration analysis, current signature analysis, acoustic emissions, and oil-debris analysis. Using post-processing techniques

[2]For rolling element bearings, these could be balls, roller, or needles depending on the bearing type. For fluid (journal) bearings, the cage and the rolling elements don't exist.
[3]`https://www.schaeffler.de`

like single-feature extraction, e.g. root-mean-square, skewness, kurtosis, etc., or more complex signal processing techniques, e.g. wavelets, envelopes, Fourier transforms, diagnosis of failure presence, fault location and size, wear evolution and crack sizes could be inferred [Cer+18].

Regarding fault detection in bearings, intensive research is present in literature, where data-driven approaches are exploited for detection of fault-sensitive and damage-related features, e.g. [GCD15a]. Among which are supervised and unsupervised NNs, fuzzy-logic, SVMs, and regression models. Model-based approaches could also be found in literature, but rather less-abundantly, due to the complexity associated with developing a fault model, which is often non-linear and non-deterministic. Examples of model-based approaches for bearing fault diagnosis can be found in [BM97; VP04; ET16].

With respect to the life prognosis in bearings, just like other applications of SHM and CM, physics-based and data-driven models have been developed to predict the RUL. Physics-based models include fatigue-life models, contact stiffness models, damage rules, and crack development/propagation models. More details about physics-based prognosis models are presented in e.g. [San+16; Sao17]. Data-driven models utilize AI and Machine Learning (ML) techniques, such as NNs, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), SVMs, Gaussian processes, Markov chains, etc. Detailed examples of utilizing data-driven RUL methods in life estimation of bearings can be found in [Jia+16; Ahm+18; AWN18; Liu+18b]. [ETJ15; Cer+18; AS18] present a review of different prognosis methods followed by a comparison of their strengths, limitations, and challenges.

Stator Windings Faults Stator faults occur mainly due to inter-turn winding faults caused by insulation breakdown. This in turn results in excessive local heating in the windings, and unbalance in the magnetic field, yielding an additional threat on the bearings and the rotor. In [Sto13], the authors review the different CM approaches

for diagnosing rotor and stator windings faults, both online during operation and offline. The authors stress on the importance of offline tests in revealing some faults that could only be detected through conducting those tests on regular basis. Among the techniques reviewed in [Sto13] is the temperature monitoring of the stator windings using temperature sensors. The sensors indicate the overall temperature of the windings, with few information on the local hot spots, that could occur due to inter-turn faults.

Alternatively, thermal imaging of the stator can provide a heat map of the local hot spots as well as the overall temperature distribution. This technique is nowadays heavily employed to detect faults occurring due to a wide range of root-causes such as shorted strands, local blockages or restrictions in cooling gas flow, extensive contamination of the stator windings and core.
In [Eft+13], the authors provide a review of inter-turn faults, being the most dominant root cause of failures in stator windings, and summarize condition monitoring techniques to diagnose them at early stages of operation. Examples of those techniques include axial leakage flux, current signature analysis, sequence component, vibration monitoring, air-gap torque, temperature monitoring, including both global temperature sensing, and local temperature sensing, e.g. thermal imaging, or estimation using thermal models, e.g. Finite Elements (FE) models, or AI models.

Other faults that occur to rotary machines are faults in rotor cage and windings, faults due to shaft eccentricity and mechanical unbalance, and stator faults occurring in the frame and laminations. Details on these faults and their corresponding CM techniques can be found in e.g. [JLB06; Ran11; BG12; Cho+18].

Aerospace Applications

Research in PdM technology in aerospace applications has been growing over the past couple of decades. The complexity and the criticality associated with aircraft systems, equally military and

commercial, along with the particular safety standards imposed on flights makes the reliability of a PHM framework a crucial and a challenging task. Several PHM frameworks have been deployed and standardized over the past years, such as HUMS [McC05], SMIS [Gry94], and RECONFIGURE [Gou+14]. Such frameworks comprise modules for fault detection, isolation, life prognosis, and decision support.

Fault detection and isolation (FDI) are achieved through model-based approaches [Din08a; Mar+12] attempting to minimize a residual between a real observed state and a simulation-calculated correspondent. Physics-based approaches include Kalman filters, particle filters, input observers, and least-square approaches. A review of physics-based methods can be found in [Doe+96]. FDI could as well be performed using data-driven techniques, sometimes referred to as model-free methods, such as fuzzy logic, and neural networks, although not so common in aerospace applications as in other fields [Fek14]. On the other hand, it is still rather common in aerospace applications to use the traditional Non-Destructive Testing (NDT) techniques, such as ultrasonic scanning, infrared thermography, and X-ray inspection. NDT is an offline maintenance procedure applied after the identification and localisation of faults, from which characteristics and severity of the fault are identified without causing damage to the structure [Shu02].

Aerospace applications, in particular, dictate some weight and safety constraints on primary systems of the aircraft, as well as auxiliary systems, among which are the PdM hardware. In [FW12], Farrar and Worden discuss the impact of such limitations on the choices and decision associated with the configuration of the PdM systems. Further discussion on PdM technologies can be found in more detail in [MK11; Zol+14; Fek14]

Electrical Components

Although not particularly relevant to this work, yet electronics systems play an important role in many applications, and hence have their share of diagnostics and prognosis research. For the sake of completeness, we highlight swiftly some advancements in that field. In [VP07], a prognostics method for estimating damage based on in-situ thermal loads monitoring in electronic components is presented. In [RLT08], the authors presented a plethora of data-driven algorithms for monitoring the health condition and estimating the RUL of batteries. In [SG09], Saha and Goebel utilized particle filtering methods to predict the dynamic behaviour of batteries during discharge cycles and estimate their RUL during the entire life cycle. Interested readers in this subject are recommended to refer to [RBMGSM17] for a more comprehensive overview.

Optimal Sensor Placement

5

The topic of sensor placement has been investigated by many researchers in the Structural Health Monitoring (SHM) and Condition Monitoring (CM) community. Despite its demonstrated importance, it is not as intensively investigated in either fields, in comparison to other research areas such as system parameters identification or life estimation. The following paragraphs will summarize the key relevant contributions in Optimal Sensor Placement (OSP). Interested readers are referred to [Li12; OSM19; YL12] for intensive review of contributions in the art.

The literature concerned with OSP could be classified according to the research question they attempt to address and propose solutions for, as follows:

5.1 Choice of Sensors

The choice of sensors is particularly dictated by the application or system considered, as well as the operating conditions under which it operates. [FW12] has assembled a non-exhaustive list of different prominent SHM applications and their recommended sensor choices. For instance, for rotary machinery and offshore platforms, piezoelectric accelerometers and displacement probes are the most commonly used sensors, due to their ability to monitor vibrations at a broad frequency range. On the other hand, in aerospace applications, strain gauges are more recommended to capture local effects induced by damages and micro-cracks, in addition to their light weight. In critical cases, corrosion and temperature sensors might as well be mounted. For civil structures, accelerometers and strain gauges are

chosen, however, care needs to be taken to ensure they tolerate the environments they operate in.

5.2 Objective Function

The sensor configuration problem is in essence an optimization problem, where a certain objective function needs to be minimized given certain constraints that need to be fulfilled. The design variables, i.e. the independent values of the objective function, are the locations of the sensors, and typical constraints would be the minimal or maximum number of sensors, or infeasible domains (locations) where sensors may not be placed [OSM19].

The earliest effort in that direction was documented in the landmark paper of Kammer [Kam91]. In this work, he proposed the Effective Independence (EfI) technique to optimize uniaxial accelerometers positions for modal identification for large structures. The method was further extended in [Kam05] to accommodate for triaxial accelerometers. The concept underneath attempts to maximize the linear independence of the observed target mode shapes, through selecting positions that minimize the measurement error between the real and estimated modes. The main drawback of this objective function is that it neglects sensor position's contribution to the kinetic energy content of the system, which could result in positions with low Signal-to-Noise Ratio (SNR). To overcome this drawback, OSP objective functions that maximize the kinetic energy content covered by sensors are proposed, see e.g. [Li12].

Another family of objective functions, which are based on information theory, were proposed by [Pap04] and further developed by the author in [PL12]. The proposed objective function provides an estimation uncertainty measure of the system parameters, represented by the inverse of the information entropy. Positioning sensors that minimize this objective in turn leads to maximizing the information and energy content of the sensed positions. For modal analysis, the

Modal Assurance Criterion (MAC) is proposed as an objective for comparing the shapes of the estimated (measured) modes with the expected ones and quantifying the spatial aliasing [Li12].

The observability and controllability Grammians are common indicators on the state estimation uncertainty [Inm17], which are also used as objective functions for sensor placement, cf. [Geo95; SH05]. Typically, matrix functions are used as the objective function, such as determinant, trace, condition number, or maximum/minimum eigenvalues. According to [HM14], all functions hold the same properties, and could be indifferently chosen. [Tam+16] have proposed an observability-based objective function based on the method of Popov-Belevitch-Hautus (PBH) for quantifying systems' observability [GR95]. Due to the computational effort required to calculate the observability Grammian, [SH05; HM14] proposed an empirical approximation of the objective function based on snapshots of simulated response.

For the aforementioned objective functions, systems where single-physics, namely system elasto-dynamics, are considered. In that context, the estimated quantity, the benchmark / real quantity and the measurement belong to the same physical domain; in the case of elasto-dynamics, this is displacement or its temporal or spatial derivatives.
The positioning on a system with multi-physics, e.g. thermo-elastic systems, has been investigated by [HRU17]. In this context, the objective function is chosen to minimize the estimation error or covariance between an unmeasured quantity (e.g. displacement) which is calculated from the coupled measured quantity (e.g. temperature). Due to the complexity of such coupled systems, model order reduction techniques are often used for mathematical modelling [SVR08]. In [Ben+18], the authors compared different model order reduction approaches, to assess the accuracy of the yielded objective function for estimating the displacement from temperature measurements.

In many SHM and CM applications, state reconstruction from sensor data is essential. At this point, it suffices to mention that one of the prominent techniques for system identification is the Kalman filter [Kal60] and its variants. Based on the joint input-state estimation formulation in [Lou+12a], an objective function based on the covariance of the estimated input has been proposed in [ZX16]. The adequateness and limitations of this objective function are heavily investigated in [TJP16; ZAS17]. This work is not only prominent in the sense that a complex objective function has been worked through and thoroughly investigated, but also for generalizing the solution to account for the positioning of multiple sensors, namely displacement probes, accelerometers and strain gauges.
Another formulation of objective, which is in fact a weighted sum of multiple objective functions, has been presented in [LXL18] for structural damage detection. In [HR15], the authors proposed an objective function for joint estimating model parameters and states for thermo-elastic systems. The choice of objective functions for multi-type sensors is still an open research question, which, to our knowledge, faces a lot of issues with respect to stability and complexity, and needs further investigation.

According to [YL12], the objective functions found in literature are all dependent on the dynamics of the structure and are essentially adequate for estimation of displacement and its temporal derivatives (velocity and acceleration). If spatial derivatives, i.e. strains, are to be estimated from vibration monitoring, the metrics may need to be customized similarly for those applications as well.

Strain sensors, such as strain gauges, have been regularly used for damage and fatigue detection due to their characteristic sensitivity to local responses in structures [OSM19]. In the context of damage identification and assessments, many methods infer the presence of damage based on changes in estimated modal shapes from strain gauges readings.
[OSM19] suggests that many OSP formulations involving strain

gauges use an objective function that minimizes the mean-square-error or the mean-absolute-error between the estimate based on the sensor readings and the actual measurement. The drawback of those two objective functions is that, globally, the high objective function value of unfavoured locations tends to fade away when combined with locations with low objective values, hence could be overseen by an on-average acceptable value of the objective function.
[ZW17] proposed an approach for optimal strain gauges placement based on the statistical analysis of the damage event detectability. Based on several simulated damage scenarios, calculated by a finite element model, the strain difference between healthy and damage state are used as the damage-sensitive feature. The proposed objective function is then defined as ratio between the number of damage scenarios detected by the strain gauges to the number of damage scenarios simulated.

5.3 Optimization Algorithms

The OSP problem is inherently an optimization problem, where adapted or adopted minimization algorithms from general optimization literature could be reused. The application, the size of the system, as well as the chosen objective function highly influence the choice of the optimization algorithm. Optimization algorithms in OSP context could be classified as shown in Figure 5.1.

Empirical approaches Empirical approaches are typically standardized engineering or expert guidelines for sensor placement. Normally, these guidelines are closely associated with specific experimental setup, and measurement process. Such approaches can be found in, for instance, the ISO 108 standard “Mechanical vibration, shock and condition monitoring” concerning machine vibrations and vibration diagnostics, and ISO 133 series "Condition Monitoring and Diagnostics of Machines" [KGK13]. Due to the limited scope of those recommendations, which makes them inevitably non-generalizable

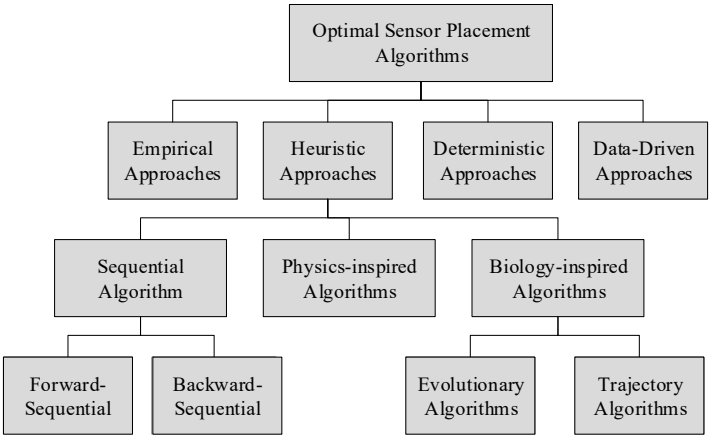


Figure 5.1.: Taxonomy of common algorithms in OSP context.

to real-world engineering applications, they tend to go overseen in many OSP literature.

Heuristic approaches The most prominent family of OSP algorithms are the heuristic approaches. The reason for this is their ease of implementation and the rather small computational overhead required by them to evaluate complex, non-linear objective functions, when compared to the deterministic approaches.
Heuristic methods for OSP could be clustered into two sub-categories: sequential methods [Bis+13] employing greedy approaches, and nature-inspired methods [Bar+17] employing physics- and biology-like routines.

The earliest contribution to sequential methods came along the proposal of the EfI objective function in [Kam91]. A forward sequential sensor placement was proposed, where sensors with minimal contribution to the objective function are sequentially removed one after the other until a constraint (e.g. minimal number of sensors), or a stopping criteria (e.g. relative change in objective function's value)

are met. This algorithm has been later utilized on other objective function as in [Pap+14; Ben+18; ZX16; TJP16].
In [KP08], the same author proposed an additive approach, where basically the sequence of sensor choice is reversed. Instead of having a full pool of sensors, where redundant ones are subtracted, the initial set of sensors is empty, and most influential sensors are sequentially appended to it. This is particularly efficient for system with many degrees of freedom, out of which a much smaller sensor set must be chosen.
The implementation and computational efficiency make them an attractive choice to many researchers. However, due to the greedy nature of those two algorithms, achieving a global optimum solution is not always guaranteed. Also, those methods work optimally only if the objective functions are proven to be decreasing and sub-modular [ZAS17]. Another restriction is their non-suitability to be applied to multi-objective optimization.

On the other hand, nature-inspired approaches are well-known optimization approaches which mimic physical processes, such as simulated annealing [WDB04], and biological patterns, such as species evolution and social behaviours. Biology-inspired algorithms are further classified into *evolutionary algorithms* and *trajectory algorithms*. Evolutionary algorithms are iterative algorithms that follow the Darwinian theory of evolution, where successors surpass their predecessors. Practically, they are initialized by a random set of sensors, and iteratively a new set is generated such that it yields a lower[1] objective function value. The progression from one set of sensors to the following is performed over three stages: selection, crossover, and mutation [Bar+17]. Examples of evolutionary algorithms are the genetic algorithms with their numerous variations and adaptations, see e.g. [LKS08; Liu+08]. Trajectory-based approaches follow the space-search patterns adopted by many creatures in groups. The basic idea evolves about multi-directional exploration of the space,

[1] It is a convention in optimization literature to formulate the objective function to be minimized.

followed by a feedback on the fitness of the explored sub-domain, upon which a proceeding search trajectory is projected. Some famous algorithms in this sub-category are the particle-swarm [RA07], ant-colony[OW04], and frog-leaping [FJ17]. Interested readers are recommended to refer to [Bis+13] and [Bar+17] for an exhaustive compilation of physics- and biology-inspired algorithms, respectively, employed for OSP. The inherent analogy of these algorithms makes them simple to comprehend, and their space-search approaches make them very attractive for multi-objective OSP problems. Nonetheless, they suffer from many disadvantages. Mainly, they require much longer time to converge and like other heuristic algorithms, their converged optimal solution is not guaranteed to be the global optimum. The recent advancement in computational power and parallel computing come in favour of nature-inspired algorithm, since they can elevate their advantages at an alleviated computational effort.

Deterministic approaches Third, deterministic approaches which employ gradient-based methods are as well used to attempt OSP problems. By far, deterministic approaches have been associated with the objective function being one of the measures of the observability and controllability Grammians. The earliest contribution in that direction can be found in [Geo95], where the author proposed the objective function optimized by employing a branch and bound optimization strategy [Wel82] for OSP in linear and non-linear dynamic systems.

Following his work, [Ser+13] exploited a max-det minimization approach to solve the optimization problem using a gradient-based approach. [HM14] proposed an empirical approximation of the Grammian matrix to reduce the complexity associated with objective function and its gradient calculation, the thing which made the gradient-based approach using the Grammian-based objectives an attractive alternative towards obtaining global optima for OSP problems. The PhD work of Hinson [Hin15] presents the most comprehensive and profound work on the optimization problem and objective formulation. Also, Hinson addressed the issues of

the method such as non-guaranteed convexity of the objective and discrete domain of design variables.

Data-driven approaches Finally, a data-driven formulation for sensor placement has been published. [Man18] utilized Proper Orthogonal Decomposition (POD), Dynamic Mode Decomposition (DMD) and empirical interpolation methods to efficiently optimize the sensor locations for estimation, control and field reconstruction.

Hybrid OSP Recently, research has been conducted on hybrid formulation of OSP problem, where a series of optimization algorithms are applied, possibly each on a different objective function. For instance, in [Som+12], a sequential algorithm is used to determine the optimal number of sensors, then followed by a genetic algorithm to find the optimal configuration. Also, a lot of investigation is currently held towards utilizing sensor fusion for leveraged structural monitoring and identification, which necessitates attempting multi-type OSP problems, e.g. [ZX16; LXL18; Kha+18].

5.4 Numerical Issues

Many developments have been carried out on the aforementioned methods to improve their robustness and stability when applied to new applications or larger systems. One challenge faced during solving OSP problems, especially using sequential algorithms, is information uniqueness. In many structures, where candidate positions are spatially close, e.g. due to fine FE discretization, the traditional sequential algorithm yields a cluster of sensors at one hot region on the structure, with essentially redundant information. [FJ17] combined a modal similarity metric with the EfI as an objective function for a frog-leaping algorithm. [GAK16] applied an observability-Grammian-based filter on the result of a forward-sequential minimization of the EfI to reject sensor choices with redundant information to the system's observability. [THVB01]

proposed an entropy-based metric for determining the minimum separation between accelerometers.

Another problem which appears when employing deterministic methods is the non-convexity of the objective function due to the binary representation of the design variables; since a sensor s may either be placed or not, thus represented as $s \in \{0, 1\}$. [Hin15] has presented a convex-relaxation modification to the objective function constraints, where the design variables are represented as a range; $s \in [0, 1]$, and has demonstrated the improved performance of this approach in comparison to mixed integer programming applied to a binary constraint.

An additional problem which appears dominantly in multi-type OSP problems are the ill-conditioned system matrices appearing in the derivation of the objective function due to the combination of heterogeneous measurements, which intrinsically vary in the order of magnitude. [Som+14] proposed to normalize the measured values before combining them in the formulation, [ZX16] proposed pre-multiplying the system's observation matrix by a weighting matrix incorporating information about the sensitivity and accuracy of the sensors used.

5.5 Comparative Literature

Many publications have been concerned with methods comparison, such as [Bak11; Gom+18; LTL04]. In fact, the results of comparative work are only applicable towards similar domain or application field, and are not necessarily transferable, since the formulation and tuning the algorithms' hyper-parameters is very problem-dependent, hence, there is no general solution for OSP [OSM19].

6 Damage Estimation

Damage estimation is an essential goal in any prognosis paradigm. Many contributions can be found in literature attempting to estimate the damage of both systems and components during operation. Damage estimation is also common during the design phase of components and systems; in fact, it is one of the critical design parameters in many applications. The presented discussion in this section is focused on damage prediction for systems deployed in operation, while interested readers in damage evaluation at the design phase are recommended to review the discussion presented in e.g. [Shi11; Hob15; Lot16; Zer+16; Zah19]

Life estimation largely gains the attentions of design and maintenance engineers in applications involving cyclically-loaded structures. Under cyclic loads, which are dominant in engineering domains, structures are subject to gradual degradation, which is caused by the presence or development of defects in the structures. Defects that might lead to fatigue life deterioration could be:

- Material defects, such as pores, shrinkages, and inclusions
- Geometric defects, such as surface roughness, and surface irregularity due to manufacturing inaccuracies
- Surface defects, such as notches, scratches, and pits, resulting mainly from environmental and loading factors.

A detailed discussion on each of these defects, and their corresponding impact on fatigue life is thoroughly presented by Zerbst et al. in a recent three-part review [Zer+19].

The purpose of life estimation models is to provide a prediction of time until the next onset of damage due to the presence of the mentioned defects in the material. To achieve this, the degradation mechanisms leading to defects in materials should have had been investigated. According to [Oko+14], the dominant failure mechanisms in metallic components are:

- Wear defined as the loss of surface's material over the course of structure's operation, mainly caused by mechanical loads, friction and contact,
- Corrosion defined as the loss of material resulting from electrical, chemical or biological reactions occurring between the outer surface of the structures and its surrounding,
- Fracture defined as the separation of material due to cracking (in homogeneous materials) or de-lamination / disintegration (as in composite materials)
- Deformation defined as the mechanism characterized by the change in the geometry of the structure, such as stretching, bending and twisting, resulting in excessive stresses or strains in the structure. Deformation could occur instantaneously (elastic and plastic deformations) or gradually over time (creep).

In contrast, in composites, composed of two or more constituent materials, either stacked as laminates or having a matrix-fiber-reinforced structure, typically encountered failure mechanisms are [SR79; Bat06]:

- Internal matrix cracks, which are micro-cracks localized at a single layer or unit volume of the matrix material,
- Interfacial debonding, which results from coalescing of matrix cracks leading to separation of laminates or separation of the matrix from the fibres,

- Fibre pull-out, which occurs in fiber-reinforced matrials due to excessive interfacial debonding, to the extent that fibres are loosely held inside the matrix,
- Fibre rupture, which occurs both due to sudden or gradual loading being exerted on the fibres.

According to [Kam+15], homogeneous materials and composites materials experience different evolution rates of fatigue damage. The onset of observable fatigue damage in homogeneous material starts after a relatively larger number of loading cycles. Before that, the material encounters cracks at the micro- and meso-scale, boundary dislocations and slippage. Cracks start to develop at the surface of the homogeneous material, and due to cyclic loading, they start to fuse, thus spreading more widely on the surface and propagating deeply into the material. Composites, however, may experience deterioration in their stiffness and integrity after only a few hundred loading cycles, due to the developed internal matrix cracks. Afterwards, the composite undergoes a gradual degradation stage corresponding to the interfacial debonding. Towards the end of life, severe failure mechanisms dominate as the fibres rupture, or as the laminates detach in an unstable manner.

For a more detailed discussion about the different faults and defect evolution mechanisms in different material families, interested readers are referred to cf. [BP13].

Having outlined the different defects that a material could possibly encounter during operation, Remaining Useful Life (RUL) estimation approaches are now introduced, see Figure 6.1. RUL estimation approaches could be categorized into four main classes

6.1 Experience-guided

These are a set of rules formulated and dictated by domain experts correlating the RUL estimate with the current and previous operat-

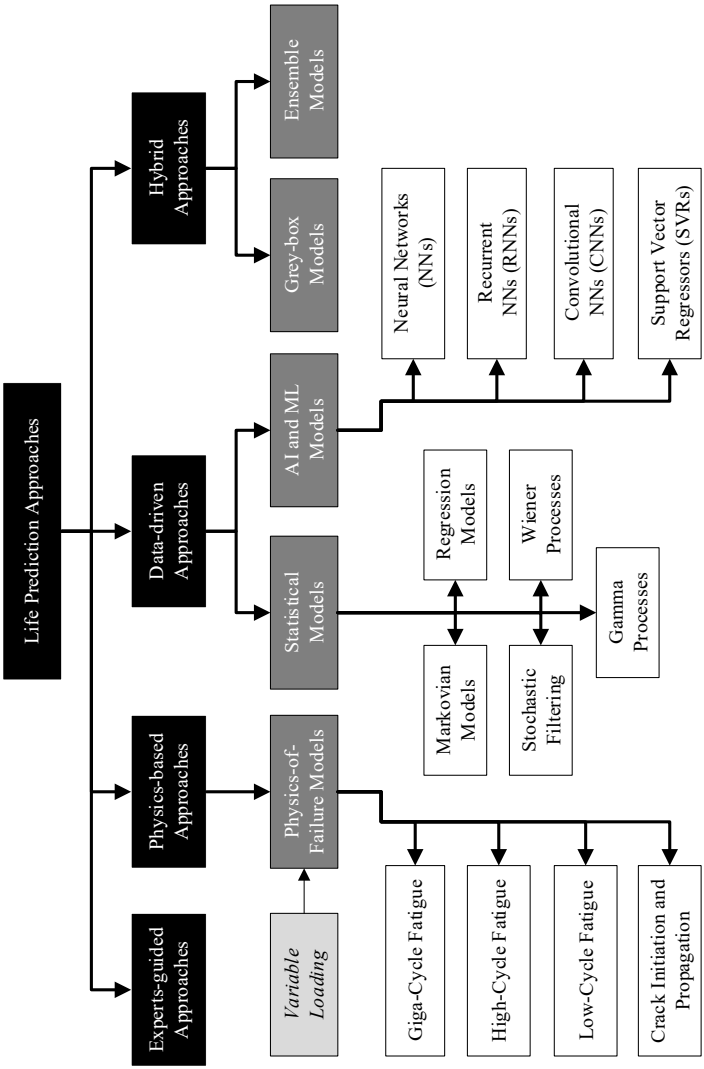


Figure 6.1.: Taxonomy of RUL methodologies in literature.

ing conditions of the structure. This approach is very interpretable as it mimics the sequence followed by experts to solve engineering problems. However, it relies heavily on the specificity of the defined expert rules to be executed, and could quickly suffer from the curse of dimensionality, for systems having a broad scope of operating conditions [LK14]. Some examples of experience-guided, or sometimes called expert systems, could be found in literature, cf. [Lem+89; BS02; BS04; Sci04].

6.2 Physics-based

These are approaches where RUL is predicted based on results of degradation models, built from the first principles and the understanding of the physics of failure mechanisms. Physics-based approaches are sometimes referred to in literature as *model-based* approaches [Luo+03]. Degradation models are parametric mathematical models, whose parameters are initialized based on experimental results and empirical data.
At the deployment phase, the model parameters are typically calibrated using measurements through model update methods, e.g. Kalman filters [Kal60], particle filters [DM96], Bayesian inference [Dow13].

The input to degradation model is the stress distribution on the structure resulting from the solution of a simulation model, e.g. Finite Elements Method [ZZT13], Extended Finite Elements Model [Kho14], or a multi-body dynamics model [RS12], from which the corresponding damage and RUL are estimated.

To domain experts, physics-based approaches provide easily interpretable prognosis measures. Additionally, when available and sufficiently complete, they significantly outperform experience-based and data-driven models. Nonetheless, this is not always the case, as physics-based models are very expensive and resource-intensive to build, and calibrate, making them not always available or sufficiently

complete for complex systems. Moreover, they are usually design for built case-by-case, making them not as robust and generalizable as other models.

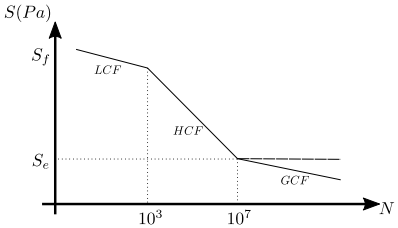


Figure 6.2.: SN curve showing the different CDM domains

Degradation Models Looking deeper into the degradation models over the past decades, one can recognize the following domains, upon which the research has been focused

- **Cumulative Damage Model (CDM)**: Cumulative Damage Models (CDMs) attribute the damage occurring in the material to the experienced load cycles and estimate the total damage severity as the summation of the individual damage contributions of each load cycle [Lee05]. The relationship between the applied load (S) and the fatigue life in number of cycles (N) is given by the *SN-Curve*, see Figure 6.2. CDMs could be sub-categorized into three different domains:
  - **Giga-Cycle Fatigue (GCF)** (also *Very High Cycle Fatigue*): refers to the class of fatigue failure occurring at load cycles $\geq 10^9$ cycles with loads significantly below the yield strength of the material. The remarkable work of Bathias over the last two decades is a key enlightenment in that domain.
    In [Bat99], the author states that fatigue infinite life doesn't exist in different materials. On the contrary, it was found out that the fatigue strength decreases about 50–200 MPa between 10^6 and 10^9 load cycles, depending on the material. Subsequently, many research contributions were

published by the same author and fellow researchers addressing the Giga-Cycle Fatigue (GCF) phenomena in different material, *viz.* ferrous alloys [Wan+99; Wag+10], bearing steels [Ara16], aluminium alloys [Wag+10], titanium alloys [Hua+16], and natural and synthetic rubber [Bay+08].

According to [MA16], under GCF regime, the fatigue life is dominantly spent in crack initiation and not in propagation. Cracks usually initiate in the core of the loaded component, leading eventually to inclusions and large internal grains in the micro-structure of the materials, known as the *Fish-eye*. These act as an internal notch, where regions of stress concentrations and plastic strain localisation are exhibited during operation. More comprehensive discussions about GCF can be found in literature, cf. [BP04; PSB11; MA16; Sak+16; PLJ18].

– **High-Cycle Fatigue (HCF)** This is the class of failures attributed to loads large enough to leave the GCF domain, yet still below the yield strength of the material, applied for $10^3 - 10^6$ cycles.

The relationship between the load and its corresponding fatigue life within the High-Cycle Fatigue (HCF) regime is described by Basquin's exponential law [Kun+08], which relates the number of cycles endured by the material to the applied stress level, as damage occurring in the elastic regime of the material is driven essentially by the material stress.

The simplest method for estimating HCF is Miner's linear degradation rule [Min59], which concludes that the sum of individual damages caused by successive loads constitutes to the total damage in the material. Despite its simplicity, Miner's rule is deficient due to ignoring non-linearity between the number of load cycles and damage, the effects of load sequence, residual stresses, and variable loads.

To overcome its deficiencies, many contributions have been published. For instance, [Mes+05; Ben+15] proposed a non-linear damage model to account for multi-axial variable loads, Lv et al. proposed a modified non-linear damage model which considers load interaction and sequence effects, [Böh+14] employs time history of the trajectory and cumulation of fatigue failures with the help of a coefficient dependent on material memory. Further developments of Miner's rule can be found in [FS11; San+16; Pav18].

– **Low-Cycle Fatigue (LCF)** This is the class of fatigue failure encountered during the first few loading cycles when large stresses (relative to the yield strength of the material) are applied. Typically, faults associated to Low-Cycle Fatigue (LCF) are surface cracks and fractures.
The relationship between the load and its corresponding fatigue life within the LCF regime is described by Coffin-Manson law [BMS92]. This law is analogous to Basquin's law being an exponential relationship between the applied cyclic load and the number of lifetime cycles, except that it attributes the fatigue life degradation to plastic strains around the crack tips.
Different approaches have been used to estimate LCF; In [BS90], the critical plane approach is introduced, where strain components are proposed to the principal plane, at which the material undergoes the maximum damage. The approach is investigated and reviewed on various ductile and brittle materials in [Zhu+18]. In [WB94], the authors have proposed a method to map a history of multi-axial loads to cycles using cycle counting algorithms (e.g. Rainflow counting [ME68]), where for each cycle the critical plane is allowed to change.
Golos and Ellyin proposed a theory based on the total strain energy in the material as a measure of fatigue

[GE89]. The advantage of this method is its robust applicability to both LCF and HCF, as well as combined loading scenarios including mechanical and thermal loads. Based on the work in [GE89], Kreiser et al. proposed a non-linear damage accumulation depending solely on the load history path and considering the stable hysteresis effects. The method has been further investigated on various metal alloys e.g. Ni-based alloys, steels, cast iron, aluminium and titanium [San+16].

Other methods are adopted to estimate the lifetime of the structure due to LCF, e.g. the methods of linear elastic fracture mechanics (LEFM) and elasto-plastic fracture mechanics (EPFM). Interested readers in the underlying assumptions and limitations of those methods are referred to e.g. [And17].

- **Crack Initiation and Propagation** Crack initiation and propagation could be modelled by deterministic as well as stochastic models [San+16]. The earliest contribution to crack initiation models was proposed by [TM81] using the concept of slip plastic flow. Other authors have proposed crack initiation models correlating crack initiation to the geometry of the material specimen examined, the presence of local plastic flow in its micro-structure, and the sizes of short and small cracks, as well as micro-structural defects whose principal dimension is in the order of the micro-structure grains, respectively [San+16].
  Among the earliest contributions to estimate crack growth and propagation using deterministic models is the power law proposed by Paris and Erdogan in [PE63]. The model was initially proposed for macro cracks (cracks having principal dimension larger than the material's micro-structure). Further models have been developed later to account for short and small cracks.
  The growth of fatigue cracks can be also modelled by non-linear stochastic models, with the earliest contributions to this

field dating back to 1985 [LY85].
A summary of developments in crack initiation and growth models can be found cf. [BAA09; Sad+09; Bes10; Sao17; San+16]. Additionally, many efforts have been directed to develop different numerical models to simulate crack initiation and growth, among which are e.g. finite elements method [Bou+00], extended finite elements method [Sut+18], scaled boundary finite elements method [SON18], and virtual elements method [Hus+19].

Variable Loading Many HCF, LCF, and crack initiation and propagation models require information about the sequence, range and mean of the applied loads. Loads in real cases could be random, multi-axial and non-proportional, hence obtaining such information requires a pre-processing step achieved through *cycle counting techniques*.

Among the standardized cycle counting techniques presented in literature are e.g. Level Cross Counting, Peak Counting, and Simple Ranging Counting, cf. [Lee05]. The most robust and widespread cycle counting techniques is the Rainflow Counting [ME68] proposed by Matsuishi and Endo. Two variants of the proposed techniques have been implemented in [DS82] and [Amz+94], and both definitions are included in the ASTM cycle counting standard [FF05]. These technique extracts the load cycles amplitudes and count by processing the load history in time domain.
Recently, Marsh et al. proposed an adaptation to the classical rainflow technique using the endpoint bound sequence (EPBS) to avoid processing the entire load history and spare the computational effort. For each given sub-history, the technique counts the full cycles in the signal, and concatenates the residue of half cycles to the proceeding sub-history, where full cycles are attempted to be counted. In [Hei+19], the authors deployed the technique to an online RUL estimation framework and highlighted its computational advantage over the classical rainflow counting.

On the other hand, many methods were proposed to perform cycle counting in the frequency domain. In fact, these techniques are not explicit cycle counting techniques; they rather attempt to estimate fatigue based on cycle counting from the frequency domain spectra, with an underlying damage accumulation rule [Pap+11]. These techniques rely on the information enclosed in the spectral moments, such as standard deviation, mean frequency, irregularity factor, peaks and upward mean crossing per second, to fit a Probability Density Function (PDF) to the given load sequence. The choice of the PDF is the key element differentiating the different spectral methods [LI15]. A common choice of PDFs include the Rayleigh distribution proposed for narrowband signals in [WL80]. The technique has been extended to accommodate wide-band signals by using empirical correction factors, as in the methods of Dirlik, Wirsching & Lite, Ortiz & Chen, Benasciutti & Tovo [LI15], or by altering the damage formulation as in the single-moment method [BCT13] and the method proposed in [PZ04; KWB19]. The strengths and limitations of each technique, and its applicability to different signals can be found in comparative reviews such as [LI15; QLW16; BT18].

6.3 Data-driven

Data-driven approaches utilize models that rely only on previously observed data to predict the projection of a system's state and infer its RUL without relying on physics or engineering principles. Data-driven models can be classified to statistical models, and artificial intelligence models.

Statistical Models Statistical models are data-driven models that constructs the RUL prediction model by fitting observed data to a probability distribution. Different statistical models have been proposed in literature and interested readers can find more thorough discussions on them in e.g. [Si+11]. Among common statistical models for RUL estimation are:

- **Regression models** Regression models use the observed data from a sample of structures to infer the deterioration of a given derivative of this sample [Ahm+18]. The underlying assumption is that all derivatives follow the same deterioration function. Typically, regression models can provide a PDF of deterioration, either in closed-form or as an approximation. Regression models are, however, limited in a sense that they can't model temporal uncertainty associated with RUL estimates. In addition, regression models are restricted to system exhibiting monotonic deterioration [Si+11].
- **Markovian Models** Markovian models are based on the assumption that the deterioration state possess a Markov property, i.e. 1) the deterioration state can be inferred directly from the observed signals, and 2) the future prediction is solely dependent on the current state (memoryless property). Markovian models can be defined for both continuous and discrete spaces of states, which facilitates their application and understanding. However, its underlying memoryless property limits its robustness. Also, in case of discrete state spaces, the transition between states is often dictated by domain experts, which might be biased, or by large data samples, which might be scarce [Si+11].
- **Wiener models** Wiener models represent RUL by predicting the time, at which the observed signal crosses a certain threshold level, corresponding to the component's end-of-life or a critical deterioration state. This point in time is denoted as the first-passage time (FPT), and its PDF is given by an inverse Gaussian distribution. Wiener models are suitable for non-monotonic degradation processes with temporal variability, unlike regression models. Wiener models have, nonetheless, some restricting assumptions; their application is limited to homogeneous deterioration models, and they ignore the loading history, assuming a Markov property for

the degradation process, which both are not always the case [Wan10b; LS+13].

- **Gamma processes** Gamma processes are models for the degradation processes, where deterioration takes place gradually over time in a sequence of positive increments. The deterioration increments themselves, as well as their cumulative sum, are described by a Gamma distribution [AH75]. The inherent property of the Gamma processes, as well as its possessed physical interpretability, motivated its application to many maintenance-related contexts, e.g. wear processes and crack propagation [Si+11]. Like Wiener models, they withhold the variability in the process uncertainty in their formulation, having an advantage over regression models. Nonetheless, they are, alike Wiener and Markov processes, memoryless processes, that rely on the current deterioration state only for future predictions. A detailed investigation on the use of Wiener processes in the context of maintenance can be found in [VN09].
- **Stochastic Filtering**: Filter-based models in their simplest fashion use historical observations, along with a state-space model of the degradation state evolution, to predict the current and the future degradation state. Examples of stochastic filter models are Kalman filters, and particle filters [SSA15]. A notable advantage of these models is their utilization of historical observations, and their applicability to linear and non-linear degradation behaviour. However, the definition of the degradation state evolution function remains an associated challenge. The main restriction of filtering models is their limitation to single-input-single-output deterioration functions, which makes them inadequate in the cases where deterioration is inferred from various observations. Also, in the context of predictive maintenance, stochastic filters have not been developed to incorporate inter-interval maintenance actions, i.e. assume monotonic deterioration [Si+11].

AI Models Artificial Intelligence (AI) models are data-driven models which use observed data as well statistical techniques, e.g. least squares, to obtain a RUL prediction, but, unlike statistical models, don't provide a PDF of the deterioration. Artificial Intelligence (AI) models are becoming more commonly deployed in many domains, and prognosis is one of them. According to [Oko+14], AI models are more favoured to statistical models in the cases where large amounts of experimental or operation data are available, from which the degradation behaviour could be inferred. A comprehensive review on AI models and their application to Condition Monitoring (CM) can be found in [Zha+16]. Among the common AI models applied in the context of Predictive Maintenance (PdM) are:

- **Neural Network (NN)**: Neural Networks (NNs) are data-driven models in which a network model learns the relationship between a given set of inputs, e.g. vibration signals, and their corresponding output, the increment in RUL. In [MZ05], the author presented a method for predicting RUL under variable loading using a NN and has shown the similarity between the NN's predictions and Miner's rule predictions for different material groups. In [Bez+07; PWB08], the authors presented a RUL prediction model for sandwich composite materials. In [Fre+09], Freitag et al. used accelerated test data to train a NN to predict fatigue life and was applied to a textile-reinforced concrete sample. [MSH10] presented a feed-forward NN to predict the RUL of bearings in rotating machinery. [Tia12] presented a similar model to predict RUL in bearings in pumps, where signals are pre-processed and fitted to a Weibull distribution, which is then used as an input to the NN. The results of his model were compared to CM data collected from pump bearings in the field.
- **Recurrent Neural Network (RNN)**: Recurrent Neural Networks (RNNs) are the most common and intuitive models for time series data, which is a common case in prognosis problems. The first usage of an RNN was reported in 2008,

as Heimes published his results to the IEEE 2008 Prognostics and Health Management conference challenge problem [Hei08]. The challenge involved detecting plant failure events in advance based on time-series signals from plant units with unknown initial wear and manufacturing variation. Following his work, RNNs have been applied in RUL estimation in gears [TZ09], batteries [Liu+10], fuel cells [Liu+18a], bearings [Guo+17; HT18], and turbofan engines [Gug+17; Zhe+17].

- **Convolutional Neural Network (CNN)**: Although Convolutional Neural Networks (CNNs) have been applied on tasks such as computer vision, natural language processing, speech recognition etc., they have only been used lately in RUL estimation in prognostics. To our knowledge, the work in [BZL16] presents the first contribution in that field, where the authors applied a proposed CNN for predicting RUL on two publicly available datasets from NASA. The proposed CNN yielded a consistently lower mean square error in comparison to Multi-Layer Perceptrons (MLPs) and Support Vector Regression (SVR) algorithms. [LDS18] adopted the proposed model in [BZL16] using the proposed time-window interval approach to consider local temporal features as input to the network. The work of [Ren+18] presents the first utilization of CNNs in RUL for bearings. Again, the authors demonstrated a lower mean square error compared to MLPs and SVR.

6.4 Hybrid Models

Hybrid approaches combine the two aforementioned worlds in an attempt to complement the shortcomings of one approach through the strengths of the other. A wide range of hybrid models have been used in many Structural Health Monitoring (SHM) and CM applications in the last decade. In the following paragraphs, we present the most notable contributions.To our knowledge, hybrid

models could be distinguished into two sub-categories: grey-box models, and ensemble models.

Grey-box Models A grey-box model establishes the relationship between its inputs and outputs through an integration of physics-based, experience-based and/or data-driven models [Hau08]. To clarify, take for example a life estimation process that is described by a Paris law (white-box), where its material parameters are not known, thus have to be identified from the sensor data captured from experiment data (black-box).

Grey-box models could comprise any combination of experience-based, physics-based or data-driven models, integrated in different ways. According to [LK14], one way of creating a grey-box model for RUL is through combining rules generated from knowledge bases and domain experts, to determine the faults in the inspected system, and integrate them with data-driven models (either machine learning or statistical filtering models) to refine the RUL estimates. Such models were deployed to assess the health status of industrial gearboxes by combining expert rules with a NNs to discard redundant rules [Gar+01], induction motors bearings by employing a fuzzy logic and NN [SS05], and cutting tools health monitoring by combining NNs and domain knowledge.
With respect to using statistical filtering models, for instance, Swanson estimated crack growth in steel bands using a model comprising a Kalman filter incorporating a crack growth degradation model and fuzzy logic [Swa01]. Similarly, in [BWE04], a fuzzy logic based on a set of predefined rules is combined with a Kalman filter to predict the progression of the damage in aircraft actuator components. In [Nie+15], the authors used a Support Vector Machine (SVM) whose model parameters were calibrated using a particle-swarm optimization algorithm to predict the RUL in an aircraft craft engine. Baptista et al. compared different data-driven models, *viz.* k-nearest neighbours, NN, SVM, and random forests, integrated in a Kalman filter framework to estimate the RUL of aircraft engines,

and concluded with the choice of the k-nearest neighbour method for the data-driven subcomponent of the grey-box model.

Contrarily, another approach towards building grey-box models is through combining different data-driven models. Usually, prognosis in such an architecture is performed in two-folds, with one of the underlying models estimates the system states or RUL-influential features, and another model predicts the RUL. In [Liu+13], a hidden Markov model was used to represent the health state of investigated bearing, which was fed into a least square support vector regression to predict the bearings' health state. Huang et al. used a minimum quantization error to estimate the health state of the system, which was input to a NN to predict RUL [Hua+07].

The three approaches could be combined together as in [BM12], where a dynamic Bayesian network holding information from expert opinions, operational data, experiments, physical models, reliability standards and a particle filter was used to predict the crack length in cantilever beams. In [OSK03], physics-based models of spall initiation and propagation in bearings were used and switched between using a vibration feature-based Dempster-Shafer approach. In [GN11], an analytical parametric physical model for choke valve fluid dynamics was developed, its parameters were identified by filtering the outputs of an ensemble of NNs using a moving average approach, to eventually feed a gamma process describing the degradation of the valves.

The most dominant structure of grey-box models in the context of RUL prognosis is the structure incorporating physics-based and data-driven models [LK14]. Different physics-data interfaces are adopted to realize such grey-box models. In some case, data-driven models are used to replace an unknown or a high-indeterministic sub-component of a physics-model; cf. [Che+10; CVO12], where the authors estimated the crack length in a helicopter's gearbox using a hidden-Markov model (instead of a physics-based approach, e.g. Paris law) within a particle filter framework.

In a second setup, data models are used as a forecast model for the future states based on historical data, while RUL is predicted using a physics-based models. This grey-box architecture was used by Liu et al. to improve the prediction of RUL in lithium-ion batteries, where a particle filter was used to estimate the RUL at the current step, and a NN was used to predict the measurements p-steps ahead in the future based on the historical values [Liu+12].

A third common grey-box architecture is when data-driven models are used to infer a *measurement model*, i.e. the relationship between the measured signal quantities and the unobserved health states, from which RUL is estimated using a physics-based model. In [OV07], the authors predicted the PDF of the fault indicator using a particle filter, which was fed to a Paris law to estimate the PDF of the crack propagation, hence RUL, in planetary carrier plates. [Moh+07] proposed a grey-box model for fatigue crack growth in metallic alloys, where a modified Paris law for crack closure was used to estimate the crack length, while the stress state in the plastic zone around the crack was inferred from a mixed Gaussian process. Similar grey-box models with different physics and data constituents were used by [Pen+12; IEC12] to estimate crack lengths in metallic structures. In the context of RUL estimation in batteries, [Sah+07] used a relevance vector machine (data-driven model) to estimate the initial battery's state of charge to be used in a particle filter-based degradation model. [LK16] used a two data-driven model; one to estimate the unknown internal states in lithium-ion batteries, and one to predict the future states. The results of both models were fed to a physics-based model to estimate the RUL in the battery.

Ensemble Models Ensemble models combine the output of multiple constituent models by a weighted-sum formulation to obtain better prediction than from the constituent models alone [Li+17]. To distinguish it from grey-box models, consider a life estimation process which could be expressed by an LCF model (e.g. Coffin-Manson law), a HCF model (e.g. Basquin law), and a crack model (e.g. Paris

law), an ensemble model assigns a weight to the outputs of each of these three models to improve the accuracy of the life prediction. Unlike grey-box models, no new model is created, but rather the weights assigned to aggregate the existing models are tuned. Just like grey-box models, the constituents of ensemble models could be any combination of experience-based, physics-based or data-driven models.

To our knowledge, the earliest contribution where ensemble models were used in an RUL estimation context was the work of Goebel and Eklund, where the author combined estimates from experience-based data-driven and physics-based models for estimating the length of spall length in bearings [GE07].
In [Pee08], the authors proposed a model composed of an ensemble of MLPs and radial basis functions as constituent models, whose combination weights are guided by a Kalman filter. The model was applied to an aircraft engine life prediction use case. A similar ensemble architecture was used by Baraldi et al., where an ensemble of empirical models was aggregated through a Kalman filter-based algorithm to predict the RUL of turbine blades affected by a developing creep [BMZ12].
In [Jia+10], the authors used an ensemble of MLPs to estimate the lifetime in gas turbines. The author has demonstrated an improvement in prediction accuracy of 35% in comparison to the constituent MLP. Hu et al.combined five data-drive approaches to predict RUL, and demonstrated the power of their model on an aircraft engine lifetime prediction problem, a power transformer ageing problem, and cooling fan life estimation problem [Hu+12].
Xing et al. combined expert-based rules with a regression model to estimate the life of lithium-ion batteries and obtained as well more better prediction accuracy using the ensemble model [Xin+13]. [ZZX13] estimated the RUL of bearings using a weighted combination of two NNs. In [Bar+14; BMZ15], the authors used an ensemble of regression methods to estimate the health state of choke valves in offshore oil platforms.

In [Wu+18], the authors used an aggregate of parallel random forest algorithms to predict the tool wear in machinery. Ensemble models were applied to estimate crack growth in structures [Bar+13; Sba+16], where NN where combined with statistical models to estimate the lifetime of components under fatigue-induced crack. In [Cad+17], Cadini et al. used the results of a committee of NNs to estimate the crack size in a helicopter panel within a particle filter-based probabilistic framework.

There is no single answer to the question on which of the discussed approaches is the best. The choice depends on the nature of the degradation behaviour, the common faults in the material, the failure mechanisms, the knowledge and experience about the underlying physical laws governing degradation, as well as the availability of experimental and operational data. Okoh et al. presented a comparative literature review [Oko+14] mapping the different faults and failure mechanisms, along with other factors influencing the choice of the RUL, to the recommended model. In [AKC15], the authors present a similar comparison, but restricted to AI models and statistical models [1], outlining their associated advantages, restrictions, and challenges [AKC15]. Readers can benefit from the discussion in these two references, as well as [JLB06; KHV06; Hen+09; SHM11] as a guideline for model selection.

[1] Although the authors referred to them as physics-based models, the presented models fit to definition of statistical models used in this work.

Part III

Fundamental Theory

7 Damage Estimation

Damage estimation is one of the essential outcomes for Structural Health Monitoring (SHM) and Condition Monitoring (CM). Many efforts have been invested in finding methods to estimate the fatigue lifetime of materials under different loading conditions, as mentioned in Chapter 6.

Damage occurring due to structural loads is evaluated by post-processing the displacement field, which results from solving the system's equations of motion. To derive the equations of motion, consider a solid structure $\Omega \subset \mathbb{R}^3$, having a weak formulation discretized by a Galerkin approach, e.g. finite element discretization [ZZT13]. The discretization yields a linear system of equations describing the dynamics of the structure, having n_d DOFs and is excited by n_p inputs, and is given in physical coordinates by

$$\mathbf{M}\ddot{\mathbf{z}}(t) + \mathbf{V}\dot{\mathbf{z}}(t) + \mathbf{K}\mathbf{z}(t) = \mathbf{L}\mathbf{u}(t). \tag{7.1}$$

The matrices $\mathbf{M}, \mathbf{V}, \mathbf{K} \in \mathbb{R}^{n_d \times n_d}$ are the system's global mass, damping and stiffness matrices, respectively. $\mathbf{u} \in \mathbb{R}^{n_p}$ and $\mathbf{z} \in \mathbb{R}^{n_d}$ are the inputs and displacements vectors, respectively. $\mathbf{L} \in \mathbb{R}^{n_d \times n_p}$ is the mapping matrix relating the structure's DOFs to the locations of the inputs. The suffix (t) denotes the time dependency of a variable, and the $\dot{(\cdot)}$ and $\ddot{(\cdot)}$ operators denote the first- and second-order temporal derivatives.

Defining the coordinate transformation $\mathbf{z}(t) = \mathbf{\Phi}\mathbf{q}(t)$, where $\mathbf{q} \in \mathbb{R}^{n_m}$, and pre-multiplying Equation (7.1) by $\mathbf{\Phi}^\mathsf{T}$, yields

$$\mathbf{\Phi}^\mathsf{T}\mathbf{M}\mathbf{\Phi}\ddot{\mathbf{q}}(t) + \mathbf{\Phi}^\mathsf{T}\mathbf{V}\mathbf{\Phi}\dot{\mathbf{q}}(t) + \mathbf{\Phi}^\mathsf{T}\mathbf{K}\mathbf{\Phi}\mathbf{q}(t) = \mathbf{\Phi}^\mathsf{T}\mathbf{L}\mathbf{u}(t), \tag{7.2}$$

where $\mathbf{\Phi} \in \mathbb{R}^{n_d \times n_m}$ contains the mass-normalized eigenvectors, obtained by solving the eigenvalue problem of the undamped free-response system, $\mathbf{K\Phi} - \mathbf{M\Phi\Omega} = \mathbf{0}$. Typically, under operating conditions, high frequency modes have negligible contributions to the structural response. Thus, the structural responses could be expressed as a superposition of the first few modal responses. This results in a reduced size of the modal system; $n_m < n_d$ [Inm17].

Let $\mathbf{\Phi}^\mathsf{T}\mathbf{M\Phi} = \mathbf{I}$, $\mathbf{\Phi}^\mathsf{T}\mathbf{V\Phi} = 2\mathbf{\Xi\Omega}$, and $\mathbf{\Phi}^\mathsf{T}\mathbf{K\Phi} = \mathbf{\Omega}^2$, Equation (7.2) can be re-written in modal coordinates as

$$\ddot{\mathbf{q}}(t) + 2\mathbf{\Xi\Omega}\dot{\mathbf{q}}(t) + \mathbf{\Omega}^2\mathbf{q}(t) = \mathbf{\Phi}^\mathsf{T}\mathbf{L}\mathbf{u}(t), \tag{7.3}$$

where $\mathbf{\Xi} \in \mathbb{R}^{n_m \times n_m}$ and $\mathbf{\Omega} \in \mathbb{R}^{n_m \times n_m}$ are the diagonal matrices containing the modal damping ratios ξ_i and the eigen-frequencies ω_i in $\mathrm{rad\,s^{-1}}$, respectively [Inm17].

In a solid mechanics system, the variables $\mathbf{z}(t)$ are the displacements at the discretization nodes. The strains $\boldsymbol{\varepsilon}(\mathbf{z}, t)$ are obtained by calculating the spatial derivatives of $\mathbf{z}(t)$ as follows (the notation ($\mathbf{z}$, t) is dropped for simplicity):

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left(\nabla\mathbf{z} + \nabla\mathbf{z}^\mathsf{T} + \underbrace{\nabla\mathbf{z}\nabla\mathbf{z}^\mathsf{T}}_{\text{non-linear}}\right), \tag{7.4}$$

where ∇ denotes the derivative operator with respect to the material coordinates [Irg08]. For the linear case of small displacements considered in this work, the non-linear term of Equation (7.4) $\nabla\mathbf{z}\nabla\mathbf{z}^\mathsf{T}$ is negligible.

Defining the fourth-order elasticity tensor $\boldsymbol{\zeta}$, where each of its components, ζ_{ijkl}, governs the constitutive relationship between the ij^{th} component of the stress tensor, $\boldsymbol{\sigma}$, and the kl^{th} component of the strain tensor, $\boldsymbol{\varepsilon}$, through the generalized Hooke's law [Irg08],

$$\sigma_{ij} = \zeta_{ijkl}\ \varepsilon_{kl}. \tag{7.5}$$

For a general 3D solid continuum, the elasticity tensor contains 81 components. Without losing generality, considering a linear-elastic, homogeneous, isotropic material, the elasticity tensor is symmetric and reduces to 9 non-zero components, governing the relationship between 6 distinct stress and 6 distinct strain components. The elasticity tensor components are described by the Lamé constants, λ and μ. The Lamé constants are calculated from the material properties; Young's modulus, E, and Poisson's ratio, ν [Irg08]. Based on these assumptions, Equation (7.5) can be re-written as

$$\sigma_{ij} = \lambda \delta_{ij} \varepsilon_{kk} + 2\mu \varepsilon_{ij}. \tag{7.6}$$

Fatigue damage develops in the material when it is subjected to cyclic loading. Consider a material point subjected to n_σ consecutive cyclic loads, each having an amplitude $\sigma_{a,i}$ applied for N_i cycles, where $i \in \{1, \ldots, n_\sigma\}$. The accumulated fatigue damage, D_a, can be estimated by the following rule:

$$D_a = \sum_i^{n_\sigma} D_i = \sum_i^{n_\sigma} \alpha \left(\frac{N_i}{N_{f,i}} \right)^\varsigma. \tag{7.7}$$

where $N_{f,i}$ is the maximum number of load cycles a material can withstand under a given cyclic load, or the fatigue limit. The coefficient α accounts for the stiffness or strength degradation of the material as the damage propagates, and ς is a function to account for non-linearity in the damage accumulation [San+16].

The simplest form of D_i in Equation (7.7) is given by the Miner's rule, where $\alpha = 1$ and $\varsigma = 1$. This is also referred to as the linear damage model, as the damage in the material is a linear relation with the cycle ratio. The accumulated damage in Equation (7.7), hence, simplifies to

$$D_a = \sum_i^{n_\sigma} \frac{N_i}{N_{f,i}} \leq 1. \tag{7.8}$$

Despite its non-realistic simplicity, Miner's rule is predominantly accepted in industrial applications undergoing variable loading. The assumption neglecting the load sequence effects, stiffness degradation of the material and non-linear damage accumulation reduces the independent variables to n_σ and $N_{f,i}$. The first is obtained from the operation, and the latter can be retrieved from the material's SN curve, hence making the rule directly applicable to any problem at almost no additional cost [Lee05].
For traditionally manufactured ferric components, neglecting stiffness degradation can be justified, since the material's Young's modulus at failure has been observed to be at least 95%. In contrast, stiffness degradation in either additive manufactured metallic parts or composites could not be neglected, therefore, a damage accumulation rule accounting for the severe, non-linear time-dependent reduction in material stiffness is necessary [Vas19].

As mentioned, $N_{f,i}$ is obtained from SN curves[1] describing the relationship between the cyclic load and the corresponding fatigue life. A component loaded within the elastic regime of the material, i.e. below its yield strength, its fatigue damage is driven by the stresses, whereas, beyond the elastic domain, strains dominate the fatigue damage development. The SN curve could be applicable to either of them, where the relationship is established between the σ_a and $N_{f,i}$ in the former, and between ε_a and $N_{f,i}$ in the latter, where ε_a is the amplitude of the fatigue-driving strain [San+16].

More generally, to account for both stress- and strain-driven fatigue effects, the Smith-Watson-Topper (SWT) parameter is used to represent the loading bearing capacity of the material with respect to the applied load [SWT70]. The SWT parameter could be expressed as

$$\varepsilon_a \sigma_a = \sigma_f N_f^b \left(\frac{\sigma_f}{E} N_f^b + \varepsilon_f N_f^c \right), \tag{7.9}$$

where ε_a and σ_a are the strain and stress amplitudes at a given material point, σ_f and b are the material's fatigue strength coefficient

[1] In some references, denoted as Wöhler curves.

and exponent, respectively. ε_f and c are the material's fatigue ductility coefficient and exponent, respectively. E is the initial static Young's modulus of the material. E, σ_f, ε_f, b and c are material parameters that can be obtained from material handbooks, e.g. [Han98], or by conducting the appropriate experiments. The fatigue limit for a given load can be obtained by iteratively solving Equation (7.9) for N_f for every applied load.

In Equation (7.9), both ε_a and σ_a are images of the scalar mappings of the strain and stress tensors given by $f^{\varepsilon}(\boldsymbol{\varepsilon}) : \boldsymbol{\varepsilon} \rightarrow \varepsilon_a$ and $f^{\sigma}(\boldsymbol{\sigma}) : \boldsymbol{\sigma} \rightarrow \sigma_a$, respectively. The fundamental idea of the function is to map the components of the strain and stress tensors to the plane at which the maximum damage is prospected to happen; i.e. likely to evolve to the damage plane [Zhu+18].
For the scope of this work, the maximum principal strain plane theory is adopted, and ε_a is chosen to be the maximum principal strain ε_I. In case of strain-driven fatigue, the maximum principal strain is responsible for driving the damage propagation. Otherwise, for stress-driven, the maximum principal stress is responsible for driving the damage, whose plane coincides with the plane on which the maximum principal strain occurs. Hence, the adopted mappings will be applicable to both cases. Other multi-axial loading reduction functions, e.g. the maximum principal stress plane, and maximum shear failure plane, can be reviewed in literature, cf. [Zhu+18].

ε_I is obtained by evaluating the largest eigenvalue of the strain tensor re-written following the Voigt notation in the following symmetric matrix form

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix}, \tag{7.10}$$

and to obtain the corresponding σ_a, the stress components are mapped to the maximum principal strain plane.

Typically, in industrial applications, variable-amplitude loading sequence is experienced, which is not directly applicable to the SWT-curves obtained from constant-amplitude loading experiments. Hence, the load history needs to be processed into constant-amplitude loading cycles through counting algorithms. For an overview on prominent counting algorithms, the reader is directed to [Lee05]. The most common and robust cycle counting method is the *rainflow counting* [Amz+94].

Many modifications have been presented in the literature to account for the shortcoming that the rainflow method doesn't account for the load sequence. Nonetheless, as the Cumulative Damage Model (CDM) used in this work is the Miner's rule, the load sequence effects could be neglected. Hence, the rainflow method is a suitable cycle counting algorithm.

The rainflow counting is herein applied to the strain and stress signals mapped on the critical plane as explained. For each load cycle, the rainflow counting yields the mean load value, as well as the load amplitude. Mean values of the load cycles have an influential effect on the fatigue limit. The mean stress effects can be incorporated in the SWT parameter by substituting the stress amplitude σ_a in Equation (7.9) by an effective value σ_e given by

$$\sigma_e = \sigma_a \sqrt{\frac{2}{1-R}} \ , \tag{7.11}$$

where $R = {}^{\sigma_{min}}/_{\sigma_{max}}$ is the stress ratio, and σ_{min} and σ_{max} are the valley and peak values of the load. For a zero-mean load $R = -1$, and Equation (7.11) becomes $\sigma_e = \sigma_a$.

8 State and Parameter Estimation

In the context of Structural Health Monitoring (SHM) and Condition Monitoring (CM), state estimation is defined as using system models to identify quantities based on sensor data to produce a complete description of the system state (defined in terms of the system variable and its derivatives). State estimation serves a variety of purposes, such as stress prediction in running structures, fatigue estimation and damage identification, either online or offline [Lou+12a].
Various state estimators for structural systems behaving both linearly and non-linearly can be found in literature, including the well-known Kalman filter [Kal60], and the particle filter [DM96], and their advancements [Pas+13].

Damage estimation is an example of state estimation problem. As seen in the previous paragraphs, the starting point of damage estimation is the estimation of the system displacement/state $\mathbf{z}$.

In control and vibration theory, most of the work carried out on state estimation has been developed on the *state-space* representation of the systems to find their equilibrium points, and perform numerical integration [Inm17].

The equation of motion in modal coordinates is considered in the coming paragraphs. To re-write Equation (7.3) in state-space form, the state variable $\mathbf{x} \in \mathbb{R}^{n_x}$ with $n_x = 2n_m$, is defined as

$$\mathbf{x} = \begin{bmatrix} \mathbf{q} \\ \dot{\mathbf{q}} \end{bmatrix}, \quad \mathbf{q}, \dot{\mathbf{q}} \in \mathbb{R}^{n_m} \tag{8.1}$$

where $\mathbf{q}$ and $\dot{\mathbf{q}}$ are the modal displacements and velocities, respectively. Equation (7.3) could be re-written in the continuous state-space form, denoted by the subscript c, as:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}_c\mathbf{x}(t) + \mathbf{B}_c\mathbf{u}(t)\\ \mathbf{y}(t) &= \mathbf{C}_c\mathbf{x}(t) + \mathbf{D}_c\mathbf{u}(t).\end{aligned} \tag{8.2}$$

The first equation expresses the equilibrium of the system. The matrices $\mathbf{A}_c \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{B}_c \in \mathbb{R}^{n_x \times n_p}$ are referred to as the state and input matrix. They hold the right- and left-hand side matrices of Equation (7.3), respectively, and are defined as:

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{\Omega}^2 & -2\mathbf{\Xi\Omega} \end{bmatrix}, \quad \mathbf{B}_c = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Phi}^\mathsf{T}\mathbf{L} \end{bmatrix}. \tag{8.3}$$

The second line of Equation (8.2) expresses the observation (output) model of the system, where $\mathbf{y} \in \mathbb{R}^{n_o}$ is the system output measured by sensors. The matrix $\mathbf{C}_c$ is the observation matrix of the system, relating the outputs to the state variables, and the matrix $\mathbf{D}_c$ is the output feedback matrix of the system relating the outputs to the inputs, given as:

$$\mathbf{y} = \begin{bmatrix} \boldsymbol{\varepsilon} \\ \ddot{\mathbf{z}} \end{bmatrix}, \quad \mathbf{C}_c = \begin{bmatrix} \mathbf{\Psi} & \mathbf{0} \\ -\mathbf{\Phi\Omega}^2 & -2\mathbf{\Phi\Xi\Omega} \end{bmatrix}, \quad \mathbf{D}_c = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Phi\Phi}^\mathsf{T}\mathbf{L} \end{bmatrix}, \tag{8.4}$$

where the considered system outputs are the physical accelerations $\ddot{\mathbf{z}}$ and strains $\boldsymbol{\varepsilon}$, and $\mathbf{\Psi}$ is the matrix containing the strain modes, obtained by differentiating the eigenmodes $\mathbf{\Phi}$ with respect to the material coordinates. The choice of the system outputs in the formulation is dictated by the common sensors used in the context of SHM and CM based on vibration signals. The arrangement

of values in the rows and columns of $\mathbf{C}_c$ and $\mathbf{D}_c$ depends on the discretization approach applied to variational form of the system's governing differential equation. In this work, Finite Elements Method (FEM) is used for spatial and time discretization, and details about the matrices assembly could be found in respective literature, cf. [ZZT13].

Due to the time-discrete form of the sensors' digital signals, the discrete forms of the system matrices are substituted in Equation (8.2). The discrete time step variable

$$t_k = k\Delta t, \quad k = \{1, \ldots, n_t\} \tag{8.5}$$

is introduced, and the discrete states, inputs and outputs at t_k are defined as:

$$\mathbf{x}_k = \mathbf{x}(t_k), \quad \mathbf{y}_k = \mathbf{y}(t_k), \quad \mathbf{u}_k = \mathbf{u}(t_k), \tag{8.6}$$

The discrete state space form of Equation (8.2), denoted by the subscript d, is written as follows:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_d \mathbf{x}_{k-1} + \mathbf{B}_d \mathbf{u}_{k-1}, \\ \mathbf{y}_k &= \mathbf{C}_d \mathbf{x}_k + \mathbf{D}_d \mathbf{u}_k, \end{aligned} \tag{8.7}$$

where the discrete forms of the system matrices in Equations (8.3) and (8.4) are defined as:

$$\begin{aligned}
\mathbf{A}_d &= e^{\mathbf{A}_c \Delta t} \in \mathbb{R}^{n_x \times n_x}, \\
\mathbf{B}_d &= (\mathbf{A}_d - \mathbf{I})(\mathbf{A}_c^{-1}\mathbf{B}_c + \bar{\mathbf{B}}) \ \in \mathbb{R}^{n_x \times n_p} \\
\bar{\mathbf{B}} &= (\mathbf{A}_d - \mathbf{A}_c \Delta t - \mathbf{I}) \left(\frac{\mathbf{A}_c^2 \ \mathbf{B}_c}{\Delta t} \right) \\
\mathbf{C}_d &= \mathbf{C}_c \qquad \mathbf{D}_d = \mathbf{D}_c
\end{aligned} \tag{8.8}$$

In the coming paragraphs, the subscript d would be dropped out for the ease of notation. To incorporate the modelling errors in the equilibrium equation, and the measurement noise in the measurement model, process noise $d^w \in \mathbb{R}^{n_x}$, and measurement noise, $d^v \in \mathbb{R}^{n_o}$ are introduced, respectively. Both variables are assumed to be time-invariant, and following a normal probability distribution with zero mean and covariances $\mathbf{Q} \geq \mathbf{0} \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{R} > \mathbf{0} \in \mathbb{R}^{n_o \times n_o}$, respectively. Also, both variables are assumed to be independent, hence both covariance matrices are diagonal [Lou+12a]. The process noise is assumed constant, while the measurement noise is assumed constant for measurements from the same sensor type, but not necessarily constant for different sensors. This results from the different measurement variances associated with strain and acceleration measurements, due to e.g. internal sensor noise, wiring, or environment noise [ZX16]. The measurement noise covariance holds the corresponding measurement variance associated to each sensor.
Let r_ε and $r_{\ddot{z}}$ be the measurement variances associated to the strain and acceleration measurement, respectively, the covariance matrix $\mathbf{R}$ is assembled as follows:

$$\mathbf{R} = \begin{bmatrix} r_\varepsilon \mathbf{I} & \mathbf{0} \\ \mathbf{0} & r_{\ddot{z}} \mathbf{I} \end{bmatrix}. \tag{8.9}$$

In the scope of this work, a linear state estimation problem is attempted. Therefore, a Kalman filter is used to estimate the state vector $\mathbf{x}_k^+$ based on the measurements vector $\mathbf{y}_k$ at each evaluation step k. In brief, the estimation is achieved through the filter over two steps; 1) prediction, and 2) update [Kal60]. Non-linear state estimation applications are beyond the scope of this work, and interested readers are directed to e.g. [Lou+12a; Lou+12b].

After introducing the process and measurement noise, Equation (8.7) is written as:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + d^w{}_{k-1}, \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + d^v{}_k \end{aligned} \tag{8.10}$$

It is assumed that the initial state x_0 and an initial state-estimate error covariance matrix $\mathbf{P}^{\times 0}$ are known. In the prediction step, an a-priori estimate of the states $\mathbf{x}_k^-$ and their respective estimation error covariance $\mathbf{P}_k^{\times -}$ are evaluated through:

$$\mathbf{x}_k^- = \mathbf{A}\mathbf{x}_{k-1}^+ + \mathbf{B}\mathbf{u}_k, \tag{8.11}$$

$$\mathbf{P}_k^{\times -} = \mathbf{A}\mathbf{P}_{k-1}^{\times +}\mathbf{A}^\mathsf{T} + \mathbf{Q}. \tag{8.12}$$

The superscripts $(\cdot)^+$ and $(\cdot)^-$ denote a-posteriori and a-priori evaluations. For $k = 0$, the initial values of the state estimate and its error covariance are used; $\mathbf{x}_{k-1}^+ = x_0$ and $\mathbf{P}_{k-1}^{\times +} = \mathbf{P}^{\times 0}$. The a-posteriori values are evaluated during the update step by considering the measurements $\mathbf{y}_k$, as follows:

$$\mathbf{x}_k^+ = \mathbf{x}_k^- + \mathbf{G}_k \left(\mathbf{y}_k - \mathbf{C}\mathbf{x}_k^- - \mathbf{D}\mathbf{u}_k\right), \tag{8.13}$$

$$\mathbf{P}_k^{\times +} = \left(\mathbf{I} - \mathbf{G}_k\mathbf{C}\right)\mathbf{P}_k^{\times -}, \tag{8.14}$$

where $\mathbf{G}_k$ is the optimal Kalman filter gain matrix evaluated at the k^{th} time increment, given by [HJS09]:

$$\mathbf{G}_k = \mathbf{P}_k^{\times -}\mathbf{C}^\mathsf{T}\left(\mathbf{R} + \mathbf{C}\mathbf{P}_k^{\times -}\mathbf{C}^\mathsf{T}\right)^{-1} \quad \in \mathbb{R}^{n_x \times n_o}. \tag{8.15}$$

Typically, only a limited number of the system's outputs are measured. The estimate of the non-measured outputs $\hat{\mathbf{y}}$ at any given time step k could be evaluated from the calculated a-posteriori state estimates $\mathbf{x}_k^+$ by:

$$\hat{\mathbf{y}}_k = \mathbf{C}\mathbf{x}_k^+ + \mathbf{D}\mathbf{u}_k + {d^v}_k. \tag{8.16}$$

Defining an output estimate error ϵ_y as:

$$\begin{aligned}\epsilon_y &= \mathbf{y}_k - \hat{\mathbf{y}}_k \\ &= \Big(\big(\mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + {d^v}_k\big)\Big) - \Big(\big(\mathbf{C}\mathbf{x}_k^+ + \mathbf{D}\mathbf{u}_k + {d^v}_k\big)\Big) \\ &= \mathbf{C}\big(\mathbf{x}_k - \mathbf{x}_k^+\big),\end{aligned} \tag{8.17}$$

then, the error covariance of the estimated outputs is given by:

$$\begin{aligned}\mathbf{P}_k^{\mathsf{y}} &= \mathbb{E}\left[\epsilon_y\, \epsilon_y^{\mathsf{T}}\right] = \mathbb{E}\left[\Big(\mathbf{C}\left(\mathbf{x}_k - \mathbf{x}_k^+\right)\Big)\Big(\mathbf{C}\left(\mathbf{x}_k - \mathbf{x}_k^+\right)\Big)^{\mathsf{T}}\right] \\ &= \mathbb{E}\left[\mathbf{C}\left(\left(\mathbf{x}_k - \mathbf{x}_k^+\right)\left(\mathbf{x}_k - \mathbf{x}_k^+\right)^{\mathsf{T}}\right)\mathbf{C}^{\mathsf{T}}\right]\end{aligned} \tag{8.18}$$

where $\mathbb{E}$ is the expected value operator. Given that the observation matrix $\mathbf{C}$ is time- and state-invariant, the error covariance of the estimated output can be re-written as [ZX16]:

$$\mathbf{P}_k^{\mathsf{y}} = \mathbf{C}\,\mathbb{E}\left[\left(\mathbf{x}_k - \mathbf{x}_k^+\right)\left(\mathbf{x}_k - \mathbf{x}_k^+\right)^{\mathsf{T}}\right]\mathbf{C}^{\mathsf{T}} = \mathbf{C}\mathbf{P}_k^{\mathsf{x}+}\mathbf{C}^{\mathsf{T}}. \tag{8.19}$$

To calculate the covariance in the damage estimation error, $\mathbf{P}^{\mathsf{D}}$, first, a mapping function

$$f^{\hat{\mathsf{D}}} : \hat{y} \rightarrow \hat{D} \tag{8.20}$$

is defined to map the output estimates to a damage estimate.

Mapping strain estimates to damage is straightforward and is obtained by following the derivation presented in Equations (7.6) to (7.8). Mapping accelerations estimates, however, requires the double temporal integration of the estimated values to obtain a displacement estimate. The displacement estimate is then substituted in Equation (7.4) to calculate a strain estimate, which is mapped in the same way as the strain estimates calculated directly from the Kalman filter.

In both procedures, the mapping is independent of the filtering process detailed above, and doesn't introduce any uncertainty to the output estimate of the filter. It is, hence, concluded that a reduction in the covariance of the estimated output error results in a reduced covariance of the damage estimate error, and a higher certainty in the damage estimate. Therefore, and based on Equation (8.19), the covariance in the damage estimation error can be expressed to be:

$$\mathbf{P}_k^{\mathsf{D}} \propto \mathbf{P}_k^{\mathsf{y}} \propto \mathbf{P}_{\kappa}^{\mathsf{x+}}. \tag{8.21}$$

Part IV

Acquisition and Sensing

9 Motivation

In Structural Health Monitoring (SHM) and Condition Monitoring (CM) applications, sensors represent the gateway through which operators could grasp an insight about their structures. In brief, SHM and CM techniques exploit sampled dynamic responses obtained from sensors to extract features sensitive to the system's damage. Extracted features afterwards are input to analysis procedures to infer or estimate the current state of the monitored system. Accordingly, the overall performance of any SHM and CM technique is essentially sensitive to the quality of the collected information. Besides the functional advantage of having an optimally-configured sensor network, the economic impact on the lifecycle costs of systems is also significant. With optimal deployment of sensors, all costs associated with instrumentation, wiring, data acquisition, management, storage and processing could be spared. This includes both hardware and labor costs. Additionally, in many applications, e.g. aerospace, the structure's weight is critical, hence, the optimal setup of acquisition systems could lead to reduction in weight, and consequently, operating costs. Optimal sensor placement mitigates the risk of false-positive alarms of damage, and reduces unnecessary system shutdowns and maintenance activities [OSM19]. That being said, finding an optimal choice of sensing devices, and an optimal configuration for mounting and deploying them on a running system becomes an inevitable demand and a huge incentive.

There is a wide spectrum of sensors, with an as-wide underlying operating principles, e.g. current generating sensors, charge generating sensors, resistive sensors, inductive sensors, capacitive sensors. In this text, the investigation of different sensor technologies is out

of scope, therefore, interested readers are recommended to browse [Asc11]. In general, the configuration of a sensor network incorporates aspects such as performance (precision, sensitivity, etc.), cost (e.g. initial, operation and maintenance costs), dimensions and space constraints, operating environment (e.g. extreme humidity, high temperature, strong electromagnetic field), fixation type (magnetic, screwing, adhesive, etc.), connectivity (wired or wireless), sensor family (vibration, strain, temperature, etc.), number, and positioning. [GMZ16; FW12] provide a more holistic overview regarding optimizing sensor networks with respect to other aspects, as well as respective literature. The focus of this chapter is limitedly directed towards optimizing sensor networks with respect to the choice of sensor family, number and positions, which will be referred to later on as Optimal Sensor Placement (OSP)

Farrar and Worden in [FW12] have summarized three main applications for which sensors are deployed; 1) detection and tracking system's responses, 2) model validation, update, and uncertainty quantification, and 3) system control. Both SHM and CM are implicitly concerned with detection and tracking in operating systems. According to [FW12], two paradigms can be distinguished as strategies for Optimal Sensor Placement (OSP) in the context of SHM and CM:

- *Experience-guided Strategy* which involves a sparse arrangement of sensors on the structure, typically guided by previous experience of system developers and operators. This strategy is rather observed in earlier attempts of SHM and CM studies. The physical quantities are measured without any a-priori definition of system parameter to be identified, e.g. damage or load magnitude, thus parameter sensitivity to measurements is not guaranteed. In fact, this strategy assumes that systems are subject to the same nominal excitation, regardless the value of the parameter of interest. With the help of large data sets of archived measurements, from which parameter-sensitive

features are extracted, statistical inference and outliers' detection approaches could be employed to detect such parameter changes. A crucial drawback of this strategy is the inability to classify parameter outliers induced by operational and environmental changes from those outliers induced by parameter change. Also, the lack of a-priori parametrized definition of the system reduces the certainty of the estimated parameter value, and the reliability of the sensing network to detect its entire range of values.

- *Parameter-guided Strategy* which is driven by an a-priori parametrized definition of the monitored system. Based on the results of experimental work or numerical simulations, a model of the system is built, which can simulate the system dynamic response under different values of the parameter of interest. Herein, the positions and types of sensors are selected based on the developed model. Associated optimization procedures indicate that changes in the system parameters yield statistically significant effects that can be observed in the measurements. It is obvious that such strategy guarantees a more precise parameter estimate since it is inherently based on the definition of the parameter in the developed model. Additionally, this strategy could yield a *leaner* sensor network, with less redundant or irrelevant information. This has a positive impact on operation costs of the entire SHM and CM process on the long run. Additionally, it alleviates the uncertainty associated with human decision regarding sensor configuration.

Having briefly shown the benefits of basing sensing configuration on a parameter-guided strategy, a choice of whether to create the parametrized system definition based either on experimental results or simulation models has to be taken. In fact, there is not a clear-cut answer regarding which is a better way to build a model. Experiments will always retain the edge of providing responses governed by the actual system's dynamics, without any assumptions or simplifications.

On the other hand, simulation models are mathematical description of the underlying physical laws governing the system and incorporate some assumptions and simplifications. With the aid of a validated simulation model, the entire spectrum of the system's operating conditions could be simulated, even going through extreme cases which could have been costly or non-feasible through experimental work. Building a simulation model and an experimental setup are both labour intensive tasks, however, extracting results from a simulation model is relatively less time and effort consuming than experiments, as the data acquisition task could be automated.

In the scope of this work, simulation models are chosen as a basis for parametrized description of the system for many reasons. First, a profound verified mathematical description is available in literature for the benchmark examples, as well as the industrial use cases presented. Contrarily, the lack of broad experimental data available for the industrial use cases was feared to jeopardize the accuracy of a model generated based solely on experimental results. Second, the robustness granted by simulation models in investigating the systems' behaviour under extreme operating conditions is deemed to come at a lower cost than the cost of experimental work. This cost comprises the hardware, as well as the labour cost involved. Finally, non-conventional, complex operating scenarios could be set up and investigated at remarkably less effort.

It should be made clear to the reader that solving the OSP problem is considered an application of utilizing simulation models during systems' operation. It can be argued that this step is typically done offline and prior to deploying the systems in their operating environments. Nonetheless, it is a fact that should be regarded, that the quality of the downstream SHM and CM activities is as high as the data collected by sensors. Additionally, the choice of a sensor configuration takes into account the operating conditions, the intended analysis, the system parameters of interest, and the

responses to be analysed during the system's operation. Hence, it is inherently a function of the system's operation.

The following chapters are organized as follows:
In Chapter 10, a procedure for finding an optimal sensor configuration is derived. The procedure's target is to reduce the degree of uncertainty in the state and accordingly damage estimate in structures based on a Kalman filter formulation. The sensor configuration involves the placement of a multitude of sensors types, namely accelerometers and strain gauges, as they are the most practical choice of sensors in SHM and CM, as indicated in Chapter 5. The objective function is first derived, followed by a formulation of an optimization problem, and a justification of the suggested optimization algorithm. Later, special issues related to the objective derivation, the problem formulation, and the numerical implementation are addressed. In Chapter 11, a numerical investigation is carried out on a Single Degree of Freedom (S-DOF) system and a 2D truss structure. The investigation involves examining the optimization parameters, assessing the solution quality, and comparing the results of the proposed OSP approach to the sequential approach, being the most commonly used in literature.

Methodology 10

In industrial real-world applications, an *injective* mapping, in a mathematical fashion, between the measured signals and the damage features, whether local or global, is not always guaranteed. In fact, signals are affected by system parameters, such as the structure's materials, or operating conditions, as well as environmental conditions, such as background noise and sensor quality, which jeopardize the uniqueness of the signal for a given state. Mohanty suggests the utilization of sensor fusion to overcome this shortcoming [Moh18]. Sensor fusion has had many definitions and classifications in literature, cf. [Cas13]. Here, the discussion is focused on *redundant* sensor fusion, where multiple input sources (sensors) provide information about the same observed target, resulting in an incremented estimate confidence. Besides, in the context of Structural Health Monitoring (SHM) and Condition Monitoring (CM), sensor fusion guarantees higher reliability of the monitoring process, better signal-to-noise ratio (as noisy signals get compensated by fusing information from different sources), a higher-resolution and wider insight on the system's state, as well as increased hypothesis discrimination [FW12].

10.1 Optimization Problem Statement

As explained in Chapter 8, in order to reduce the uncertainty in the damage estimate, the a-posteriori error covariance of the state estimate, $\mathbf{P}_k^{x+}$ in Equation (8.14), is to be minimized. For the scope of this work, a generalized damage accumulation rule is selected, cf. Equation (7.7), in association with the rainflow cycle counting algorithm for variable loading [Lee05]. The relationship between the load and the fatigue life is established through the Smith-Watson-

Topper (SWT) parameter. These choices constitute to the definition of the mapping function in Equation (8.20), hence the damage estimate. Nonetheless, these choices are arbitrary as long as they are accompanied by the consistent definition of the mapping function, and should not have an influence on the optimal sensor placement problem formulation presented in the following paragraphs.

To begin with, let the Boolean variable $\beta \in \{0, 1\}$ be defined to indicate whether the system output is measured, i.e. a sensor is placed at a given position. Hence, for the whole system, the sensor selection matrix $\boldsymbol{\beta}$ is defined as:

$$\boldsymbol{\beta} = \text{diag}\left(\beta_1, \ldots, \beta_{n_o}\right) \in \{0, 1\}^{n_o \times n_o}, \tag{10.1}$$

where e.g. $\beta_1 = 1$ indicates a measured system output at position 1, and vice-versa for $\beta_1 = 0$. Re-writing Equation (8.14) in terms of the selection matrix yields:

$$\mathbf{P}_k^{x+}(\boldsymbol{\beta}) = \left(\mathbf{I} - \mathbf{G}_k(\boldsymbol{\beta})\ \mathbf{C}(\boldsymbol{\beta}) \right) \mathbf{P}_k^{x-}(\boldsymbol{\beta}), \tag{10.2}$$

where $\mathbf{C}(\boldsymbol{\beta}) = \boldsymbol{\beta}\ \mathbf{C}_0$, and $\mathbf{C}_0$ refers to the observation matrix with all system outputs being measured.

Effectively, the selection of an optimal configuration of sensors, $\boldsymbol{\beta}^*$, which minimizes $\mathbf{P}_k^{x+}$, is obtained by solving the following optimization problem

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \varphi\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right), \tag{10.3}$$

where φ is a function mapping the matrix $\mathbf{P}_k^{x+} \in \mathbb{R}^{n_x \times n_x}$ to a real number, and is evaluated at a given k-th Kalman filter step. Herein, Equation (10.3) presents an unconstrained optimization problem, where the objective is to minimize the $\mathbf{P}_k^{x+}$ through minimizing the function $\varphi\left(\mathbf{P}_k^{x+}\right)$.

It is important for the function φ to preserve the inherent information in $\mathbf{P}_k^{x+}$ with respect to the uncertainty of the estimation error [Hin15].

[Hin15] presented the most common functions used in this kind of problems, along with an interpretation of their relation to the estimation error covariance. These include

- *Condition number*, which is the ratio of the maximum to minimum singular values of the error covariance matrix, indicating the shape of the estimation uncertainty ellipsoid,
- *Determinant*, which is the product of the singular values, indicating the volume of the estimation uncertainty ellipsoid, and
- *Trace*, which is the summation of the singular values, indicating the average estimation uncertainty ellipsoid.

Since different functions implicitly interpret the estimation uncertainty of the system states, the choice of the function used is arbitrary. In the course of this work, the *trace* is selected, as it is relatively easier to be analytically differentiated. This shall become a requirement as presented in a Section 10.4. Herein, the optimization problem in Equation (10.3) becomes

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \ \mathrm{Tr}\left(\mathbf{P}_k^{\times+}(\boldsymbol{\beta})\right). \tag{10.4}$$

10.2 Choice of Optimization Approach

In recent works, e.g. [Pap04; KP08; Li12; ZX16], the authors resorted to a forward-sequential greedy algorithm for attempting the problem formulated in Equation (10.3) due to its undeniable computational efficiency and implementation simplicity. An inherent property of the system is assumed for a greedy algorithm to be adequate: the objective function to be minimized must be *non-increasing* and *sub-modular* [ZAS17].

To understand the criticality of the requirement, consider a linear dynamic system, cf. Equation (8.7). For any arbitrary number of sensors n_s, a state estimator, e.g. a Kalman filter, provides the optimal estimate of the state based on the sensor measurements by minimizing the mean square estimation error. Assuming the pair $(\mathbf{A}, \mathbf{C})$ is detectable, and the pair $(\mathbf{A}, \mathbf{Q}^{\frac{1}{2}})$ is stabilizable, their exists a unique lower limit, $\varphi^* = \inf(\varphi(\mathbf{P}_k^{x+}))$, to which the a-posteriori estimation error covariance will converge [ZAS17]. The solution of the sensor placement optimization problem attempts to find the sensor configuration (number of sensors and their positions) which achieves this limit. In a forward-sequential greedy algorithm, the following steps are followed:

1. A set of candidate sensors is initialized
2. The function φ is evaluated for all candidate sensors
3. The sensor which contributes most to the minimization of φ is appended to the set of optimal sensors, and dropped out of the set of the candidate sensors,
4. Steps 2 and 3 are iteratively repeated until a stopping criteria is reached, e.g. a maximum number of sensors, or a threshold value of the function φ or the change in it.

Herein, to guarantee that no consecutive sensor additions would lead to a reduction in estimation uncertainty, regardless of their sequence of choice, the function φ should follow that 1) its value decreases or stays constant for each additional sensor, i.e. *non-increasing* function, and 2) that the ratio of decrease for a given sensor is less than or equal to the ratio of the decrease from the previous sensor addition, i.e. *sub-modular* function, see Figure 10.1.

For limited cases, the mapping function φ is a *non-increasing*, *sub-modular* function, however, this could not be generalized to all systems [ZAS17]. Hence, it couldn't be taken for granted that the sequential addition of sensor locations unconditionally has a positive impact on the certainty of the state estimate.

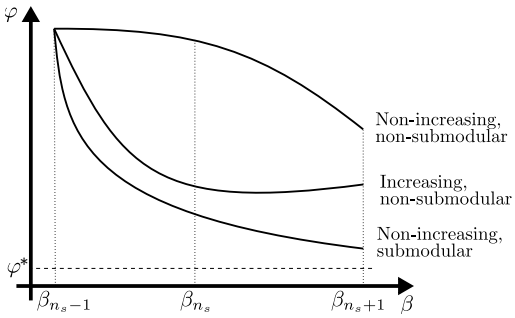


Figure 10.1.: Given $\|\beta_{n_s-1}\|_1 < \|\beta_{n_s}\|_1 < \|\beta_{n_s+1}\|_1$, the function $\varphi(\boldsymbol{\beta})$ is non-increasing if and only if $\varphi(\beta_{n_s+1}) \leq \varphi(\beta_{n_s}) \leq \varphi(\beta_{n_s-1})$, and sub-modular if and only if $\varphi(\beta_{n_s}) - \varphi(\beta_{n_s+1}) \leq \varphi(\beta_{n_s-1}) - \varphi(\beta_{n_s})$

On the other hand, exhaustive search algorithms, see e.g. [Kha+18], fail definitely for large systems, as the possible number of sensor combinations explodes; the size of the search would be $n_o!/(n_s!(n_o - n_s)!)$. For the 5-nodes Single Degree of Freedom (S-DOF) spring-mass system, cf. Figure 11.1, there exists 1023 possible sensor combinations, and adding one more node increase the number of combinations to 4095. Similarly, evolutionary algorithms suffer from the curse of dimensionality as they require large number of iterations, compared to sequential methods, to reach an optimal solution [Pap04].

Given the discrete definition of the design variable, $\beta \in \{0, 1\}$, it is rather common to consider mixed integer programming as a solution algorithm. However, the scalability of mixed integer programming are notorious. Because mixed integer programming doesn't scale to large problems, a relaxation of the binary variable to the entire range $[0, 1]$ is often resorted to. In this case, the relaxed problem provides an approximate fast solution to the original mixed integer problem [HM14]. A discussion about the design variable relaxation follows in the upcoming paragraphs.

[Hin15] proposed using a gradient-based approach to minimize the empirical observability Grammian of non-linear dynamic systems, relating the measures of the Grammian to the estimation uncertainty of the system states. The gradient-based approach is an attractive alternative to sequential approaches, as its applicability is not limited to non-increasing, sub-modular objective functions. Additionally, benefiting from the gradient information, the optimal solution could be reached faster compared to the search and evolutionary algorithms, as well as mixed integer programming [OSM19].

To get an impression about the complexity of the objective function space, Figure 10.2 shows a contour plot of the objective function $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)$ evaluated for the different combinations of 1 and 2 sensors (`Accel` and `SG`) for the 5-node S-DOF system shown in Figure 11.1. The diagonal solid line refers to the objective function value for 1 sensor, projected in Figure 10.3. As shown in both figures, the objective function is neither convex with respect to the position of the sensors nor to the type of sensors. Additionally, for 2 sensors, there seems to be no distinct global optimum for the function; the local optima are indicated by the dots in Figure 10.2.

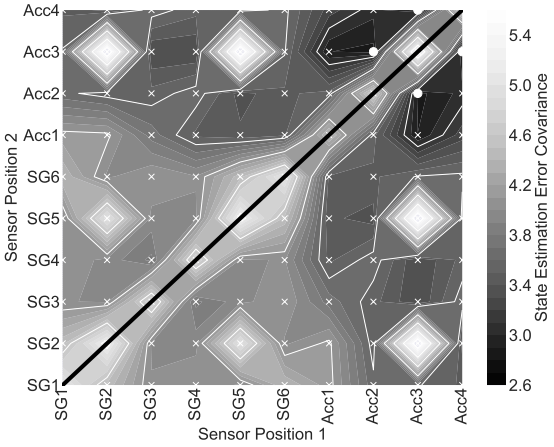


Figure 10.2.: Example Contour of $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)$ for 2 sensors, $\|\boldsymbol{\beta}\|_1 = 2$.

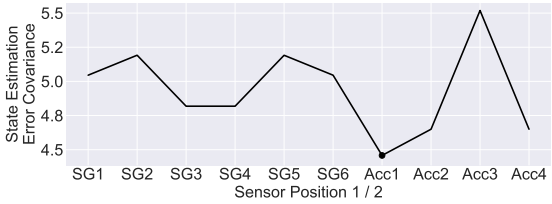


Figure 10.3.: Example of $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)$ for 1 sensors, $\|\boldsymbol{\beta}\|_1 = 1$.

The complexity of the objective function contours is expected to increase with larger number of sensors to be positioned. Motivated by this, it is decided, within the context of this work, to solve Equation (10.3) using a gradient-based approach.

10.3 Formulation Approaches

Equation (10.4) presents an unconstrained optimization problem formulation. Such a formulation might result in a dense sensor configuration (holding a large number of sensors), since this would promote a lower state estimation covariance [Hin15].
In the following paragraphs, two alternative formulations are presented to overcome this challenge.

10.3.1 Regularized Formulation

In Optimal Sensor Placement (OSP), a sparse solution that could maintain an acceptable certainty in the state estimate is desired. The sparsity of the solution could be promoted using the l_1-norm regularization technique (Reg) [BV04].
The optimization problem re-writes to:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} f_{obj} = \arg\min_{\boldsymbol{\beta}} \left\{ \mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right) + \lambda_{\mathrm{P}} \ \|\boldsymbol{\beta}\|_1 \right\}. \qquad (10.5)$$

where $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)$ represents the measure of state estimation uncertainty to be minimized by choosing an optimal sensor configuration $\boldsymbol{\beta}^*$. $\|\boldsymbol{\beta}\|_1$ is the regularization function, representing the number of

sensors, given by the l_1-norm of $\boldsymbol{\beta}$, and $\lambda_{\mathsf{P}} > 0$ is a user-defined constant. Herein, no constraints are imposed on the resulting number of sensors n_s. Instead, n_s is a free variable in the optimization problem determined by $\boldsymbol{\beta}^*$.

The l_1-norm regularization functions as such: λ_{P} penalises the obtained number of sensors, thus increasing the optimization cost for denser sensor configurations. This, in turn, introduces a trade-off between the desired sparse solution and a solution with low $\mathrm{Tr}\left(\mathbf{P}_k^{\mathsf{x}+}(\boldsymbol{\beta})\right)$. The trade-off between the solution sparsity and the state estimation certainty is a user decision, dictated by the choice of λ_{P}. Hence, the choice of a suitable λ_{P} poses a concern about this problem formulation.

Another concern about this formulation is the risk of obtaining a solution with no sensors, especially if a large λ_{P} is chosen. This concern can be alleviated by examining Equation (10.2). For an empty vector of sensors, the observation matrix $\mathbf{C}(\boldsymbol{\beta})$, hence $\mathbf{P}_k^{\mathsf{x}+}(\boldsymbol{\beta})$, would be undefined. Accordingly, a solution with no sensors is not a plausible solution.
Yet another perspective; $\mathrm{Tr}\left(\mathbf{P}_k^{\mathsf{x}+}(\boldsymbol{\beta})\right)$ is the measure of state estimation uncertainty of the Kalman filter given sensor measurements. Should no sensor measurements be fed to the filter, the state estimation uncertainty would increase as the filter progresses, eventually reaching ∞. Thus, for an empty configuration, the first term of the regularized cost function would be very large, making this solution far from the optimal.
Contrarily, [HM14] presents several studies in favour of the formulation's performance. The authors conclude that l_1-norm regularization in OSP often results in configurations near the global optimal solution found by the mixed integer program.

Using a gradient-based optimization method, the regularized objective function should be differentiated with respect to $\boldsymbol{\beta}$ as follows:

$$\frac{\partial f_{obj}}{\partial \boldsymbol{\beta}} = \frac{\partial \ \mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} + \lambda_P \frac{\partial \|\boldsymbol{\beta}\|_1}{\partial \boldsymbol{\beta}}. \tag{10.6}$$

10.3.2 Constrained Formulation

A typical constraint for OSP problems is an a-priori determined number of sensors, n_s. Herein, the optimization problem in Equation (10.4) becomes

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \ \mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right), \quad \text{s.t.} \quad \sum_j^{n_o} \beta_j = n_s. \tag{10.7}$$

Assuming the existence of a unique number of sensors n_s^* for which a minimum value φ^* of the objective function is reached, Equation (10.7) could be solved by employing the *Fixed-point approach* (FP) proposed in [Kha+19a]. Herein, the optimization problem is divided into two consequent sub-problems:

1. Finding the optimal number of sensors n_s^* where

$$\varphi^*(\boldsymbol{\beta} \mid \|\boldsymbol{\beta}\|_1 = n_s^*) \ \leq \ \varphi^*(\boldsymbol{\beta} \mid \|\boldsymbol{\beta}\|_1 = n_s) \ \forall \ n_s \neq n_s^*,$$

2. Finding the optimal configuration of n_s^* sensors, where

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \ \mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right), \quad \text{s.t.} \ \|\boldsymbol{\beta}\|_1 = n_s^*. \tag{10.8}$$

The motivation for a fixed-point approach stems from the fact that the objective function is a function in its own constraint. This effectively yields a complex optimization process if both sub-problems are attempted simultaneously [Kha+19a].

Moreover, the a-priori choice of n_s^* implicitly defines a feasibility domain for the optimization function and narrows down the solution space to the combinations satisfying the constraint $\|\boldsymbol{\beta}\|_1 = n_s^*$. This allows faster convergence and reduces the risk of being trapped into less favourable local optima.

The procedure for finding n_s^* was introduced in [Kha+19a] and is detailed in Appendix A. Equation (10.8) could be reformulated into an unconstrained problem by using constraint enforcement methods. In this work, Lagrange multipliers and penalty methods are used.

Lagrange Multipliers The Lagrange multiplier (LM) method introduces an additional design variable λ_{L} to the optimization problem. As a result, the number of the design variables of the system increases from n_o (number of outputs) to $n_o + n_c$, where $n_c = 1$ is the number of constraints. Hence, the resulting optimization problem becomes

$$\boldsymbol{\beta}^* = \underset{\lambda_{\mathsf{L}},\boldsymbol{\beta}}{\arg\min}\ f_{obj} = \underset{\lambda_{\mathsf{L}},\boldsymbol{\beta}}{\arg\min}\ \left\{\mathrm{Tr}\left(\mathbf{P}_k^{\mathsf{x}+}(\boldsymbol{\beta})\right) + \lambda_{\mathsf{L}}\ g\left(\boldsymbol{\beta}\right)\right\}, \tag{10.9}$$

where $g\left(\boldsymbol{\beta}\right)$ is the constraint function, defined as:

$$g\left(\boldsymbol{\beta}\right) = \|\boldsymbol{\beta}\|_1 - n_s^*. \tag{10.10}$$

Examining the constraint qualification of the chosen $g(\boldsymbol{\beta})$, it is found that $\nabla_{\boldsymbol{\beta}}\ g(\boldsymbol{\beta}) \neq 0$ for all *feasible* solutions. The qualification is only violated when $\boldsymbol{\beta} = \mathbf{0}$; i.e. a solution with no sensors. As mentioned earlier, an empty-vector solution is non-feasible as it would result in an undefined $\mathbf{C}(\boldsymbol{\beta})$, hence an undefined $\mathbf{P}^{\mathsf{x}+}(\boldsymbol{\beta})$. Therefore, this solution is intuitively not the optimal, thus, the constraint non-qualification at which shouldn't introduce numerical issues.

Using LM, the application of gradient-based approach requires differentiating f_{obj} with respect to both the vector of selection variables $\boldsymbol{\beta}$ as well as the Lagrange multiplier λ_{L}, as follows:

$$\begin{aligned}\frac{\partial f_{obj}}{\partial \boldsymbol{\beta}} &= \frac{\partial\ \mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} + \lambda_{\mathsf{L}} \frac{\partial \|\boldsymbol{\beta}\|_1}{\partial \boldsymbol{\beta}}, \\ \frac{\partial f_{obj}}{\partial \lambda_{\mathsf{L}}} &= g\left(\boldsymbol{\beta}\right) = \|\boldsymbol{\beta}\|_1 - n_s^*.\end{aligned} \tag{10.11}$$

Penalty Method An alternative method to enforce the equality constraint is the penalty method. The penalty method (Pen) enforces the constraint by scaling the constraint violations by the penalty factor λ_{P}. Unlike LM, λ_{P} is not a design variable, but a constant user-defined parameter in the optimization problem. Thus, the problem size remains n_o (number of outputs).

The optimization problem using a penalty method is written as:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}}\ f_{obj} = \arg\min_{\boldsymbol{\beta}}\ \left\{\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right) + \lambda_{\mathsf{P}}\ p\left(\boldsymbol{\beta}\right)\right\}, \tag{10.12}$$

where $p\left(\boldsymbol{\beta}\right)$ is the *penalty* function, defined in terms of the constraint $\|\boldsymbol{\beta}\|_1 = n_s^*$.
The quadratic penalty is a common function, defined as:

$$p\left(\boldsymbol{\beta}\right) = \left(\max\ \{0,\ \|\boldsymbol{\beta}\|_1 - n_s^*\}\right)^2.$$

When employing a gradient-based approach, the *max* function has the shortcoming of being non-differentiable. Additionally, the maximum function is an ideal penalty function for inequality or bounded constraints, where the domain in which the $p(\beta)$ equates to 0 is permissible, and otherwise a violation should be penalised. However, in this work, since an equality constraint is employed, the following penalty function is, alternatively, introduced

$$p\left(\boldsymbol{\beta}\right) = \left(\|\boldsymbol{\beta}\|_1 - n_s^*\right)^2. \tag{10.13}$$

This penalty function is C1-continuous. Also, it equates to the quadratic penalty function for solutions with a number of sensors

$\geq n_s^*$. However, unlike the quadratic function, solutions with number of sensors $< n_s^*$ are as well penalised. The penalty function equates to zero only when the number of sensors in the solution reaches n_s^*. Using gradient-based approaches with the penalty method, f_{obj} should be differentiated only with respect to $\boldsymbol{\beta}$ as follows:

$$\frac{\partial f_{obj}}{\partial \boldsymbol{\beta}} = \frac{\partial \ \mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta})\right)}{\partial \boldsymbol{\beta}} + 2\lambda_P \left(\|\boldsymbol{\beta}\|_1 - n_s^*\right) \frac{\partial \|\boldsymbol{\beta}\|_1}{\partial \boldsymbol{\beta}}. \tag{10.14}$$

10.3.3 Discrete Variable Treatment

Earlier in the discussion, it has been mentioned that mixed integer programming doesn't scale to large problems. On the other hand, for a gradient-based approach, the binary definition of β poses as a challenge for calculating the sensitivity of the objective functions as shown in Equations (10.6), (10.11) and (10.14).

In order to employ a gradient-based approach, it becomes, thus, necessary to relax the binary selection variable β to the interval $[0, 1]$, as shown in Figure 10.4. A condition needs to be kept in mind, however; in order for the solution of the relaxed convex problem to serve as a solution for the original mixed integer problem, it needs to be mapped back to the original domain $\{0, 1\}$.

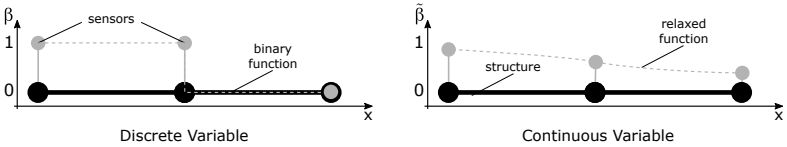


Figure 10.4.: Discrete vs. Continuous Sensor Placement Variables.

Let $\beta \in [0, 1]$ be the relaxed selection variable. A mapping function $f^{\beta} : \beta \rightarrow \tilde{\beta}$ is defined to map back the relaxed variable to a binary value $\tilde{\beta} \in \{0, 1\}$. Hence, the mapped selection matrix is defined as:

$$\tilde{\boldsymbol{\beta}} = \mathrm{diag}\left(\tilde{\beta}_1, \ldots, \tilde{\beta}_{n_o}\right) = \mathrm{diag}\left(f^{\beta}\left(\beta_1\right), \ldots, f^{\beta}\left(\beta_{n_o}\right)\right) \tag{10.15}$$

The mapped variable is now used during the optimization procedure, such that the objective function and its Jacobian are evaluated with respect to the relaxed variables. The term $\frac{\partial \|\boldsymbol{\beta}\|_1}{\partial \boldsymbol{\beta}}$ in Equations (10.6), (10.11) and (10.14) is hence substituted by:

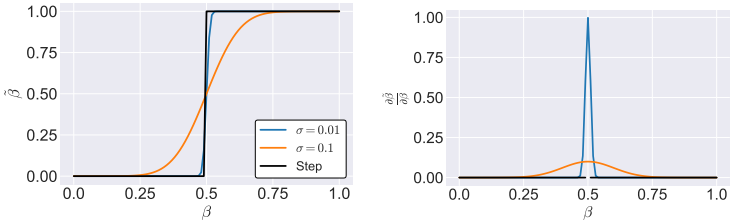
$$\frac{\partial \|\tilde{\boldsymbol{\beta}}\|_1}{\partial \boldsymbol{\beta}} = \frac{\partial \|\tilde{\boldsymbol{\beta}}\|_1}{\partial \tilde{\boldsymbol{\beta}}} \frac{\partial \tilde{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}.$$

A simple definition of the mapping function f^{β} is a piecewise (step) function, as follows:

$$\tilde{\beta}_j = f^{\beta}\left(\beta_j\right) = \begin{cases} 0 & \text{if } \beta_j < \bar{\beta} \\ 1 & \text{if } \beta_j \geq \bar{\beta} \end{cases}, \quad \frac{\partial \tilde{\beta}_j}{\partial \beta_i} = \delta_{ij} \begin{cases} 0 & \text{if } \beta_j \neq \bar{\beta} \\ \phi & \text{if } \beta_j = \bar{\beta} \end{cases} \tag{10.16}$$

with $\bar{\beta}$ being a user-defined relaxation threshold, above which β is rounded up to 1, and below to 0. It can be immediately noticed that this function is discontinuous at $\beta = \bar{\beta}$, thus having an undefined derivative at this point (as denoted by ϕ). This would cause algorithmic instabilities if incorporated into a gradient-based optimization procedure.

Alternatively, a continuous function exhibiting a similar step-function profile has to be chosen. An example of such a function is the Cumulative Density Function (CDF) of the Gaussian distribution, having its mean value $\mu = \bar{\beta}$, and a very small variance $\sigma_{\beta}^2 \ll 1$. Figure 10.5 shows that the Cumulative Density Function (CDF) tends to the piece-wise step function in Equation (10.16) as $\sigma_{\beta}^2 \to 0$. In contrast to the step function, the CDF of the Gaussian distribution, denoted by Ψ, is continuous and differentiable with respect to its random variable $\forall\, \beta \in [0, 1]$. The first derivative of the CDF of the Gaussian distribution is given by its Probability Density Function



(a) Function (b) Derivative

Figure 10.5.: Influence of the choice of the Gaussian distribution variable on the shape of the function and its derivative. $\bar{\beta} = 0.5$.

(PDF), denoted by ψ. Hence, the mapping function and its derivative are defined as follows:

$$\tilde{\beta}_j = \Psi\left(\beta_j \mid \bar{\beta}, \sigma_\beta^2\right), \qquad \frac{\partial \tilde{\beta}_j}{\partial \beta_i} = \delta_{ij}\psi\left(\beta_i \mid \bar{\beta}, \sigma_\beta^2\right) \tag{10.17}$$

10.4 Solution Strategy

The optimization problem of interest falls under the family of non-linear optimization problems as shown in Section 10.1. Iterative approaches are employed to solve this kind of optimization problems. Iterative methods are known to be rather practical for problems involving large number of DOFs, for which the evaluation of the objective function as well as the gradients could be computationally demanding [NW06].

10.4.1 Optimization Method

One of the iterative approaches, having a wide consensus in structural optimization problems, is the Method of Moving Asymptotes (MMA) [Sva87]. In structural optimization problems, the evaluation of the objective and Jacobian functions involves intensive finite-element calculations at the spatially-discrete nodal locations to find the solution regions. The formulation of the OSP problem shows a great resemblance to topology optimization, in a sense that the objective function and the Jacobian are evaluated at discrete nodal positions, resulting in discrete design variables. In this case, the said variables are the selection variables β. Exploiting the similarity between the problems, MMA is used to solve the presented OSP problem.

The execution of MMA runs in two nested iteration loops, with the superscript (k) denoting the outer loop counter. At the beginning, the design variables are initialized. In the scope of this work, the elements of $\boldsymbol{\beta}^{(0)}$ are randomly sampled from a uniform distributions $\mathcal{U}(0,1)$ and $\lambda^{(0)}$ is user-initialized[1]. The outer loop starts with a value of the design variables $\boldsymbol{\beta}^{(k)}$ and $\lambda^{(k)}$ and increments to $\boldsymbol{\beta}^{(k+1)}$ and $\lambda^{(k+1)}$ once the inner loop is terminated. Before entering the inner loop, the gradient of the objective function is evaluated at $\left(\boldsymbol{\beta}^{(k)}, \lambda^{(k)}\right)$.

[1] Only when a LM approach is employed. In penalty approach, λ is not a design variable.

In the inner loop, the original optimization problem is approximated by an explicitly convex sub-problem about the point $\left(\boldsymbol{\beta}^{(k)}, \lambda^{(k)}\right)$. The approximating function is calculated using the evaluated gradient and the method parameters, referred to as the *moving asymptotes*. The convex sub-problem is iteratively solved using until a convergent solution of the sub-problem is reached. This solution becomes the next iteration point $\left(\boldsymbol{\beta}^{(k+1)}, \lambda^{(k+1)}\right)$.

The outer loop keeps iterating until a defined stopping criteria is reached. Within the scope of this work, three stopping criteria are set to control the termination of the algorithm.

1. l_2-norm of the relative change in design variables $\underline{\Delta} x^{(k)} \leq \epsilon_{tol}^{x}$

$$\underline{\Delta} x^{(k)} = \left\| \begin{bmatrix} \underline{\Delta}\lambda^{(k)} & \underline{\Delta}\boldsymbol{\beta}^{(k)} \end{bmatrix}^{\mathsf{T}} \right\|_2$$

$$\text{where} \quad \underline{\Delta}\lambda^{(k)} = \frac{\lambda^{(k)} - \lambda^{(k-1)}}{\lambda^{(k)}}, \tag{10.18}$$

$$\underline{\Delta}\boldsymbol{\beta}^{(k)} = \left(\frac{\beta_j^{(k)} - \beta_j^{(k-1)}}{\beta_j^{(k)}} \right)_{j=1}^{n_o}$$

   Here, $\underline{\Delta}\lambda^{(k)}$ corresponds to the Lagrange multiplier, when applicable, and $\underline{\Delta}\boldsymbol{\beta}^{(k)}$ corresponds to the sensor selection variables.

2. Absolute change in objective function $\Delta f^{(k)} \leq \epsilon_{tol}^{f}$

$$\Delta f^{(k)} = f_{obj}^{(k)} - f_{obj}^{(k-1)} \tag{10.19}$$

3. Number of iterations $k = n_{max}^{iter}$

The details of defining the sub-problem and performing the inner iterations are given in [Sva87] and a more stable version of the algorithm is presented in [Sva02]. For the discussion here, it suffices to mention that the approximate sub-problem requires the first-order derivative

of the objective function with respect to the design variables to be derived and evaluated.
The Jacobian expression of f_{obj} is given in Equations (10.6), (10.11) and (10.14), depending on the problem formulation. The detailed derivation of the Jacobian of Tr $\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)$ is presented in Appendix B.

10.4.2 Solution Pipeline

To summarize, the following steps are required to solve the optimal solution problem using the procedure presented in this work. In Algorithm 10.1, a pseudo-code is given, showing the details of the following steps.

1. Assembly of system matrices
2. Building state-space representation
3. Assumption on process and measurement noise
4. System simulation to retrieve prospective output at candidate output locations
5. Estimation of n_s^*
6. Formulation of optimization problem, Equation (10.5) or Equation (10.8)
7. Initialization of optimization variables, $\boldsymbol{\beta}^{(0)}$ and $\lambda^{(0)}$
8. Optimization problem solution

Algorithm 10.1 Algorithm to find an optimal sensor configuration

procedure SYSTEM DEFINITION
    Assemble system matrices $\mathbf{K}$, $\mathbf{V}$, and $\mathbf{M}$
    Assemble input vector $\mathbf{u}$ and input mapping matrix $\mathbf{L}$.
    Assume process noise d^w and measurement noise d^v
    Build state space representation of equations of motion Equation (8.7).
    Solve Equation (8.7) to simulate system output, $\mathbf{y}$.
end procedure
procedure ESTIMATE n_s^*
    Evaluate Grammian: $\mathbf{G} = \int_{t_0}^{t_f} \mathbf{Y}^T \mathbf{Y} dt$
    SVD decomposition: $\mathbf{G} = \mathbf{U}_f \, \Sigma_f \, \mathbf{Z}_f^{\mathsf{T}}$ (cf. Equation (A.2))
    $n_s^* \leftarrow r$ (find r by evaluating Equation (A.4))
end procedure
procedure SOLVE OPTIMIZATION PROBLEM
    Set optimization stopping criteria: ϵ_{tol}^x, ϵ_{tol}^f and n_{max}^{iter}
    Initialize MMA outer loop counter: $k \leftarrow 0$
    Initialize $\boldsymbol{\beta}^{(0)}$ and $\lambda^{(0)}$
    while $k < n_{iter}^{max}$ **do**
        Map selection variable $\tilde{\boldsymbol{\beta}}^{(k)}$ (cf. Equation (10.17))
        Evaluate $f_{obj}^{(k)}$
        Evaluate Jacobian $\dfrac{\partial f_{obj}^{(k)}}{\partial \boldsymbol{\beta}}$ and $\dfrac{\partial f_{obj}^{(k)}}{\partial \lambda}$
        Evaluate stopping criteria $\underline{\Delta} x^{(k)}$ and $\Delta f^{(k)}$
        if $\underline{\Delta} x^{(k)} > \epsilon_{tol}^x$ or $\Delta f^{(k)} > \epsilon_{tol}^f$ **then**
            Update design variables $\boldsymbol{\beta}^{(k+1)}$ and $\lambda^{(k+1)}$
            Increment outer loop counter: $(k) \leftarrow (k) + 1$
        else
            Optimal sensor configuration found: $\boldsymbol{\beta}^* \leftarrow \boldsymbol{\beta}^{(k)}$
            Terminate
        end if
    end while
end procedure

11 Numerical Investigation

In this chapter, a numerical investigation of the proposed approach is carried out. The goals of this investigation are

1. Examine the solution space, and observe the objective function behaviour in terms of sub-modularity and non-increasingness,
   (a) Variation of the objective function with respect to the number and types of sensors
   (b) Solution spaces of the estimation error and estimation covariance as functions of the sensor configuration
   (c) Distribution of the resultant sensors from the different formulations in the solution spaces
2. Study the sensitivity of the resultant sensor configuration and estimation quality to the problem formulation,
   (a) Formulation approach: *fixed-point* vs. *regularized*,
   (b) Constraint enforcement: *penalty* vs. *Lagrange multipliers*,
3. Verify the advantage of using a multi-sensor configuration over a single-sensor one
4. Compare the performance of the proposed Optimal Sensor Placement (OSP) approach to the forward-sequential approach applied on the same objective function.

Within the context of this investigation, the following quantities are chosen to be used as evaluation metrics:

1. n_s^*: The number of sensors the approach decides for, being implicitly part of the optimization's cost,
2. n_{iter}: The speed of convergence, expressed in terms of the number of iterations required by the optimization algorithm to reach a solution,
3. f_{obj}^*: The value of the unconstrained objective function at the optimal (converged) solution,
4. $\text{Tr}(\mathbf{P}^{x+})$: The average estimation error covariance, when the optimal solution is used as output locations to feed a Kalman filter solving for the hidden states,
5. ϵ_x: The mean absolute relative error of the estimated states calculated across the Kalman filter evaluation horizon over all states

11.1 Problem Description

The aforementioned investigation is conducted on two systems; a Single Degree of Freedom (S-DOF) system, Figure 11.1, and a truss structure, Figure 11.2, both modelled using Finite Elements.
The decision for these two specific systems is due to their relative simplicity in comparison to more complex industrial engineering systems. Hence, they pose as minimal case problems, upon which extreme scenarios could be applied, and a broad range of interesting parameters could be tested. Additionally, both problems have closed form mathematical models, thus minimizing the errors introduced by the modelling process, and cornering the error sources to those due to the estimation procedure.
Despite their non-complex mathematical formulations, the two cases pose two levels of complexity. The Single Degree of Freedom (S-DOF) has fewer finite elements, as well as holding only a single DOF per node, thus resulting in significantly fewer possible sensor configurations, and a simpler solution space. On the other hand, the truss structure is a 2D structure, with more elements, and 2 DOFs

per node, thus allowing acceleration sensors to be oriented in either directions. This, in turn, appends an additional dimension to the solution space as will be explained in the coming paragraphs.

The following points summarize the main geometric properties and dynamic characteristics of the two use cases.

Test Case 1: S-DOF System

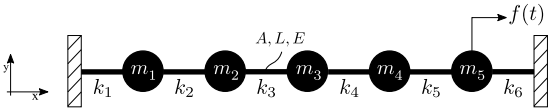


Figure 11.1.: S-DOF system.

- **Geometry:**
  - 6 springs, 7 nodes, with the two end nodes fixed
  - lengths, 0.3 m
  - circular cross-sections of a 7 mm diameter
- **Material:**
  - AISI-1045: $E = 210\,\text{GPa}$, $\nu = 0.3$, $\rho = 7850\,\text{kg}\,\text{m}^{-3}$
  - fatigue properties: $\sigma_f = 948\,\text{MPa}$, $b = -0.092$, $\varepsilon_f = 0.26$, and $c = -0.445$
  - linear, elastic material model
- **Load $f(t)$:**
  - location: Node 5,
  - characteristics: periodic, sinusoidal, frequency = 10 Hz, amplitude = 1 kN,
  - superposed by a random Gaussian process following $\mathcal{N}(0, 1)\,\text{kN}$
- **Model:**
  - 1D constant-stiffness spring elements connecting mass elements located at the nodes,
  - nodal lumped masses: 21, 15, 12, 15 and 21 kg,
  - system damping: 1 % of all system natural frequencies,
- **Sensors:**
  - Uni-axial strain gauges on the spring elements, oriented along the elements axis,
  - Uni-axial accelerometers on all non-boundary nodes except 5 (load node).

Test Case 2: Truss Structure

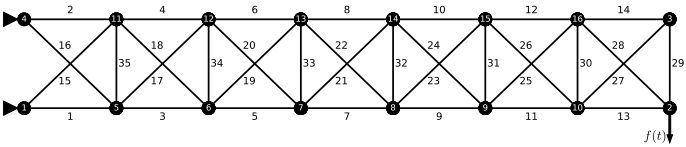


Figure 11.2.: Truss structure, adapted from [Pap+11].

- **Geometry:**
  - cantilever structure, composed of 35 truss members
  - horizontal and vertical elements have equal lengths 1 m.
  - circular cross-sections of a 7 mm diameter
  - identical mass $200\rho AL$ kg, where A and L are the cross-sectional area and length.
- **Material:**
  - AISI-1045: $E = 210\,\text{GPa}$, $\nu = 0.3$, $\rho = 7850\,\text{kg m}^{-3}$
  - fatigue properties: $\sigma_f = 948\,\text{MPa}$, $b = -0.092$, $\varepsilon_f = 0.26$, and $c = -0.445$
  - linear, elastic material model
- **Load $f(t)$:**
  - location: Node 2,
  - characteristics: periodic, sinusoidal, frequency = 100 Hz, amplitude = 100 N,
  - superposed by a random Gaussian process following $\mathcal{N}(0, 1)$ N
- **Model:**
  - truss finite element formulation,
  - system damping: 5 % of all system natural frequencies,
- **Sensors:**
  - `SG`: uni-axial strain gauges on the truss members, oriented along the elements axis - total: 35,
  - `AccX`, `AccY`: uni-axial accelerometers on all nodes except 2 (load node) and 1 and 4 (boundary nodes), oriented either in the x- or y-directions - total: $13 \times 2 = 26$.

To evaluate the precision of the hidden state estimation and the associated estimate confidence, uncertainty is introduced to the both problems by perturbing the masses of the elements by a random $\Delta m \sim \mathcal{N}(0, 0.05)\,\mathrm{kg}$. The perturbed system is depicted as a real system with unknown deviations from the mathematical model used for the sensor placement. Outputs at the chosen sensor positions, $\mathbf{y}$, are fed into a Kalman filter to estimate the hidden states $\hat{\mathbf{x}}^{\mathsf{h}} \in \mathbb{R}^{n_h}$, while the actual system states , $\mathbf{x}^{\mathsf{h}}$, are calculated from the simulated outputs of the perturbed system. The mean absolute relative error between both quantities is given by:

$$\epsilon_x = \frac{1}{n_t n_h} \sum_{k=1}^{n_t} \sum_{l=1}^{n_h} \left| \frac{x_{kl}^{\mathsf{h}} - \hat{x}_{kl}^{\mathsf{h}}}{x_{kl}^{\mathsf{h}}} \right|. \tag{11.1}$$

Output reconstruction refers to estimating the unmeasured outputs (i.e. outputs associated to DOFs where no sensors are mounted), $\hat{\mathbf{y}}^{\mathsf{r}} \in \mathbb{R}^{n_r}$, based on the estimated states. The actual values of the reconstructed outputs $\mathbf{y}^{\mathsf{r}}$ are obtained from the perturbed system simulation, and the corresponding Mean Absolute Relative Error (MARE) of reconstruction can be calculated by:

$$\epsilon_y = \frac{1}{n_t n_r} \sum_{k=1}^{n_t} \sum_{l=1}^{n_r} \left| \frac{y_{kl}^{\mathsf{r}} - \hat{y}_{kl}^{\mathsf{r}}}{y_{kl}^{\mathsf{r}}} \right|. \tag{11.2}$$

The process noise, d^w, is assumed to be $10^{-2}\,\mathrm{m}$ for displacement states and $10^{-2}\,\mathrm{m\,s^{-1}}$ for velocity states. All sensors from one category are assumed to be identical sensors, having independent and identical Gaussian noise. The measurement noise, d^v, is assumed 10^{-1} for strain gauges, and $10^{-3}\,\mathrm{m\,s^{-2}}$ for accelerometers. That's because accelerometers are assumed to output a signal with less noise compared to strain gauges.

It is worth noting that the discussion in this section is focused intentionally on the state estimates and their corresponding estimation

error covariance, and not on damage estimate, since these quantities are immediately related to the estimation process, while the damage estimation is a post-processing procedure that follows. Only at the latest stage of the investigation is the damage estimation examined and compared against the actual values.

11.2 Solution Spaces

A first step towards benchmarking is setting the ground truth against which the model results are assessed. Herein, the solution spaces of both problems are investigated. Due to the large number of sensor configurations, especially for the truss problem – going beyond 10^{17} combinations, exhaustive solution space evaluation is almost impossible. Instead, the combinations are sampled randomly from sub-spaces generated by full factorial design of experiments applied to the *number* of sensors.

For instance, taking the S-DOF problem, the total number of sensors are 6 `SG` and 4 `Accel`, which makes 34 full-factorial experiments with pairs of sensor counts (0, 1), (0, 2), ..., (6, 4). Here, for each experiment, sensor positions from each category are randomly drawn according to the corresponding count. For this work, 30 unique sensor configuration are chosen for each experiment, except for the last experiment with (6, 4) sensors, since there is only one possible combination associated with it; i.e. a fully covered system.

The exact procedure is repeated for the truss problem, with 35 `SG` and 28 `Accel`, with the x- and y-accelerometers being drawn from the same pool of sensors, and not separated. This yields full-factorial experiments with pairs of sensor counts instead of triplets.

For each experiment, $\mathrm{Tr}(\mathbf{P}^{x+})$ and ϵ_x are evaluated by feeding the sensor selection to the Kalman filter, as explained earlier. Additionally, the respective value of the objective function is evaluated (as if such a configuration has resulted from the optimization procedure), by considering $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)$ at $k = 2$. This is referred to as the *Raw* f_{obj}, to distinguish it from the constrained objective functions,

using either the fixed point or the regularization formulations, cf. Equations (10.5) and (10.8). The latter are as well evaluated for the configurations of each experiment's configuration, with $\|\tilde{\boldsymbol{\beta}}\|_1$ taken as the sum of the sensor counts, and n_s^* calculated using pre-defined values for $\epsilon_0 \in \{10^{-2}, 10^{-4}\}$.

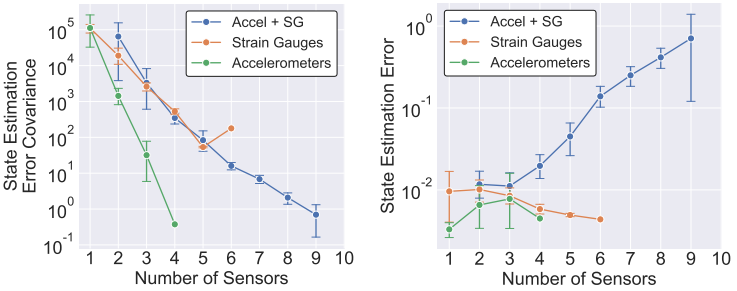
Besides exploring the solutions space, the results of 30 optimization runs using each constraint formulation approach are shown. This provides a visualization of the locations of the optimization algorithm procedures in the space representing the solution space.

S-DOF System

Figure 11.3 shows the relationships between the estimation error and estimation covariance on one side, and the number of sensors on the other side, taking into consideration whether only a single sensor type is mounted, or a multi-type configuration, denoted by *"Accel + SG"*.

Examining the state estimation error, for a single sensor configuration, mounting more than 2 sensors results in a monotonic reduction in the average estimation error as well as the uncertainty associated with it, indicated by the error bars. When deploying a multi-type sensor configuration, the average estimation error decreases to a minimum for $n_s = 3$, then gradually increases with the number of sensors . The sample deviation of the error doesn't change notably, except for $n_s = 9$.
On the other hand, the covariance of the estimation error is non-increasing through the whole domain for single-type accelerometers and multi-type configurations. In contrast, only for strain gauges when $n_s \leq 6$, the function is non-increasing; i.e. a system fully covered with strain gauges alone doesn't guarantee a more confident estimate.

For this case, one might rashly conclude that sequential sensor placement is an adequate approach. This could even be supported



(a) $\mathrm{Tr}(\mathbf{P}^{x+})$ (b) ϵ_x

Figure 11.3.: Variation of the estimation error and covariance with respect to the number and type of sensors - S-DOF system

by the raw objective function contours in Figure 11.4 (a).
The contours in Figures 11.4 and 11.5, as well as Figures 11.7 and 11.8 are generated by considering the minimum value of the corresponding quantity for the respective sensor configuration. For some cases, and due to the large range of the plotted variables, the contour values are capped to ease visualization, and to highlight the gradual change of the variable's values.

The contours in Figure 11.4 (a) give the impression that the more the accelerometers the lower the objective function, while strain gauges' presence is of no effect. Therefore, favouring the least number of sensors, an optimal configuration becomes 4 accelerometers and no strain gauges. Nevertheless, as indicated by the colour scale, the contours of the raw objective function are almost identical, which falsely indicates that the estimation uncertainty is independent of sensor configurations. This impression is proven rash when comparing the contours of the raw objective function of the optimization with the contours of the actual estimation error covariance in Figure 11.5 (a). Herein, the constrained objective functions are appreciated when considering their contours, whether using the regularization formulation, Figure 11.4 (b), or the fixed-point, Figure 11.4 (c) and Figure 11.4 (d). In a sense, the constrained objective functions form

a front of optimal sensor configurations similar to the front of low covariance.

Through visual inspection of the distribution of sensors for each constraint formulation approach in Figure 11.5, it can be seen that all approaches yield sensor configurations within the bounds of the low uncertainty contours in the solution space. Additionally, with respect to the estimation error, the results seem to be formulation independent, with all chosen sensor configurations yielding an average error less than 2 %.

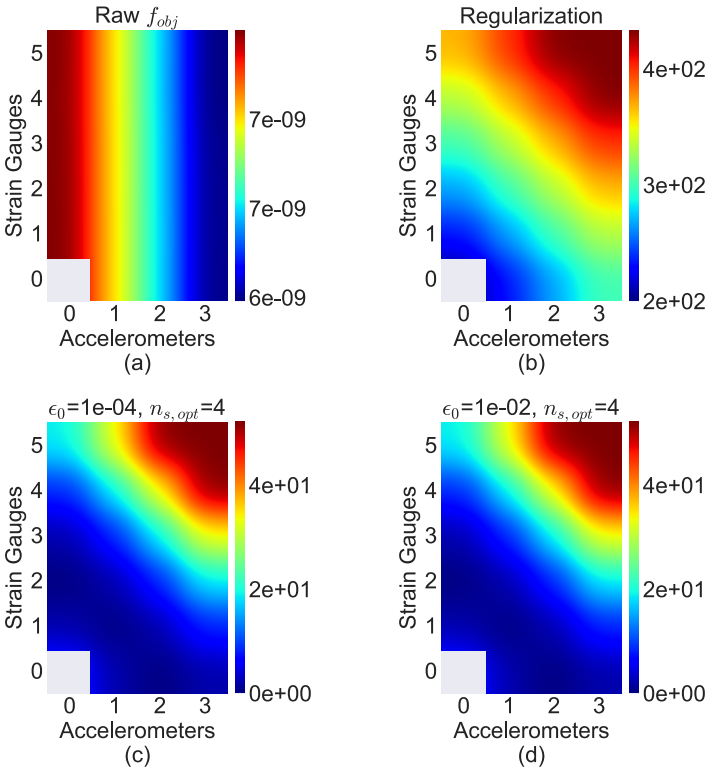


Figure 11.4.: Contours of unconstrained (raw) and constrained objective function using FP and Reg - S-DOF system

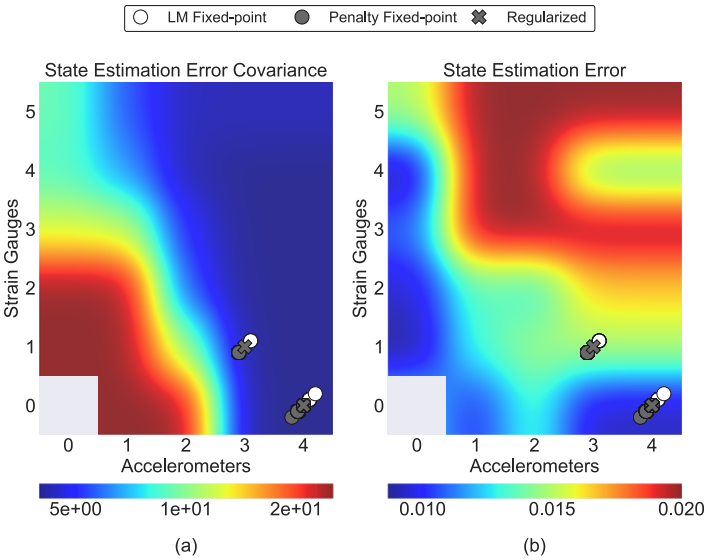
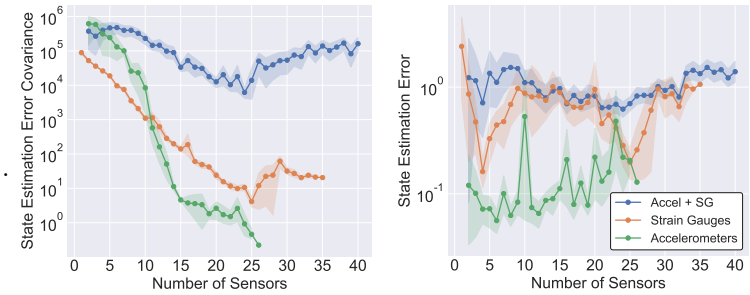


Figure 11.5.: Contours of estimation error and covariance, with the OSP solution using different formulations - S-DOF system

Truss Structure

The truss structure is a more complex system in comparison to the S-DOF system. It has a larger number of hidden states to be estimated, more candidate positions to place sensors, as well as the additional dimension introduced by the sensors orientation of accelerometers.

Examining the progression of the estimation error and the corresponding estimation error covariance in Figure 11.6, it can be seen that the non-increasing behaviour observed in the S-DOF no longer holds for both single- and multi-type configurations. In this case, for instance, it could be asserted that the trace of the covariance is not always a non-increasing, sub-modular function, according to the findings of [ZAS17]. Herein, the choice of a gradient-based OSP could be further motivated and justified.



(a) $\mathrm{Tr}(\mathbf{P}^{x+})$ (b) ϵ_x

Figure 11.6.: Variation of the estimation error and covariance with respect to the number and type of sensors - Truss structure

Similar to the S-DOF system, the contours of the minimum objective function for each sampled sensor configuration are plotted, and are shown in Figure 11.7, where four main regions could be distinguished;

1. A region with few accelerometers and few strain gauges (lower left corner).
   In this region, the values are the highest in the objective function space.

2. A region with few accelerometers, but > 15 strain gauges (upper left corner of the plot).
   In this region, the raw objective function values start to decrease, and could seem promising for the optimizer to traverse. The region corresponding to this one in the state estimation covariance space, Figure 11.8(a), yields a close range of high covariance values. This makes it as unfavoured as the previous region. Like the S-DOF, the constrained objective functions, but not the raw, capture this behaviour, and result in a qualitatively equivalent solution space.

3. The subspace with > 20 accelerometers, and >5 strain gauges (right section of the plot).
   This is deemed as a region of instability for the solver. This

can as well be indicated by the high values in the estimation error space, Figure 11.8(a). To ease the visualization, some of these values are capped, while their true values are actually higher than the scale limits.

4. The mid section of the objective function space.
   This region bears a remarkable resemblance of its correspondent in the state estimation error covariance space. It poses a favourable sub-space for the optimizer to traverse and search for an optimal configuration.

In this problem, the optimal sensor configurations form a cluster with the number of strain gauges lying between 0 and 2, and the number of accelerometers (X or Y) lying between 14 and 16. This cluster is located well inside the domain of low state estimation error covariance, as well as the state estimation error solution spaces (Region 4). Furthermore, the choice of the constraint formulation doesn't seem to hinder the optimization procedure from reaching a solution located at a low-level contour of the solution space.

Since the optimal configurations generated are dominated by accelerometers, the accelerometers' solution space is thoroughly investigated. Figure 11.10 shows the raw objective function, alongside the fixed-point and regularization objective functions. Figure 11.11 shows the corresponding distribution of the state estimation error and covariance in relationship to the sensor configurations.

Comparing the objective function contours to the contours of the state estimation error covariance, the constrained functions, irrespective of the constraint formulation, show a better resemblance than the unconstrained one. Although all solutions lie within a low covariance and a low error domain, it can be noticed that the configurations from regularization formulation are more widely spread within the solution space, and approach the fronts of the low-high estimation error and covariance domains. On the other hand, the fixed-point results are located at stably low regions of state estimation error and

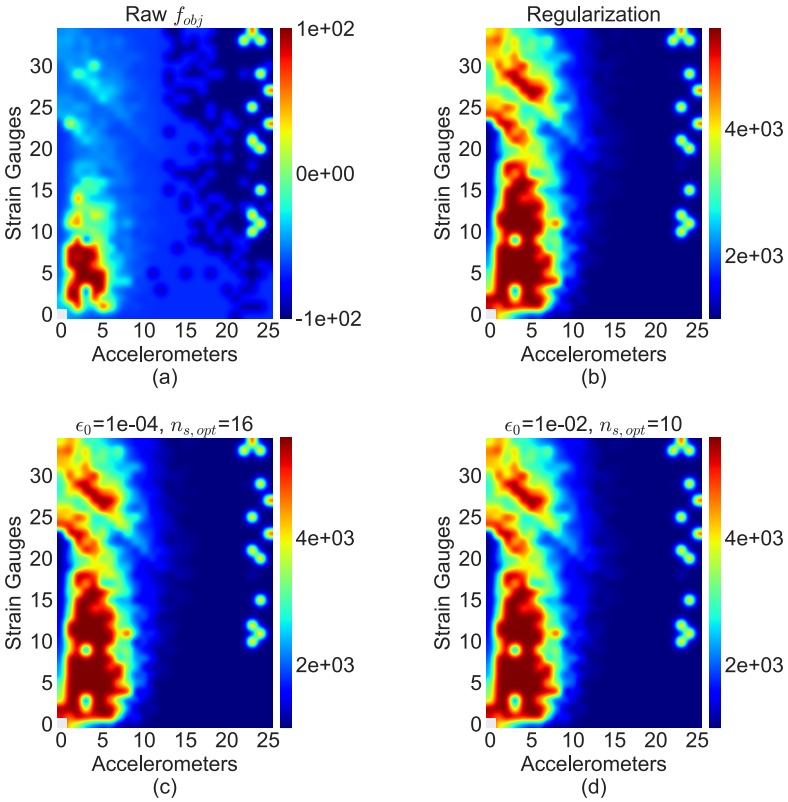


Figure 11.7.: Contours of unconstrained (raw) and constrained objective function using FP and Reg - Truss structure

covariances. A more detailed quantitative investigation is triggered by this observation.

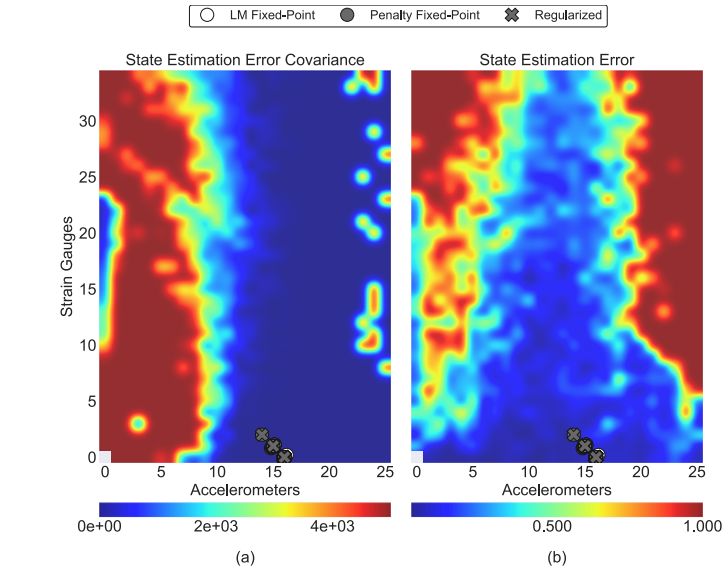


Figure 11.8.: Contours of estimation error and covariance, with the OSP solution using different formulations - Truss structure

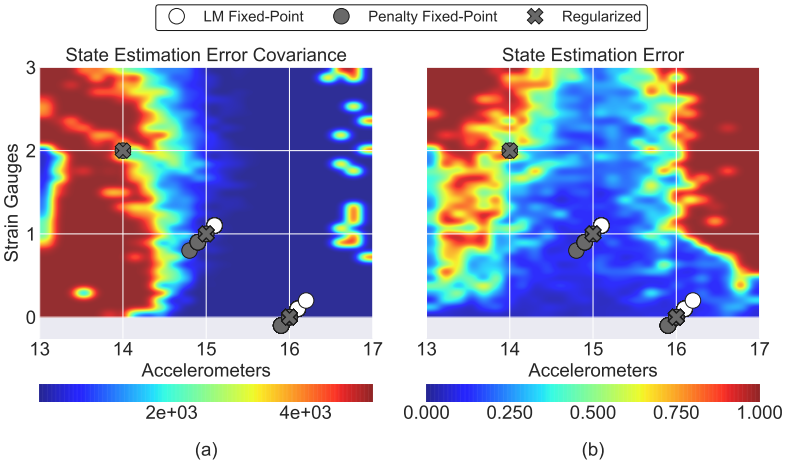


Figure 11.9.: Contours of estimation error and covariance, with the OSP solution using different formulations - Truss structure

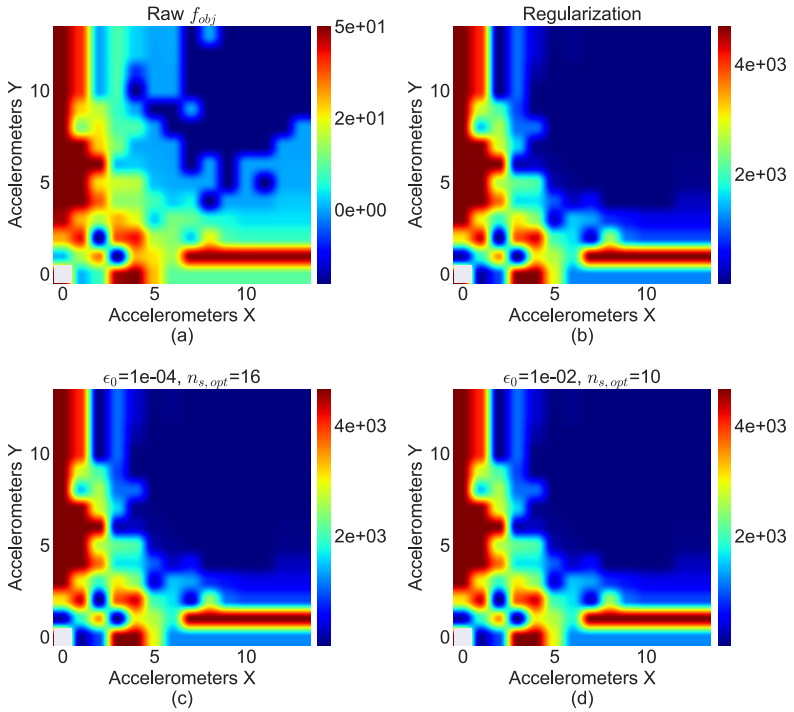


Figure 11.10.: Contours of unconstrained (raw) and constrained objective function using FP and Reg - Truss structure

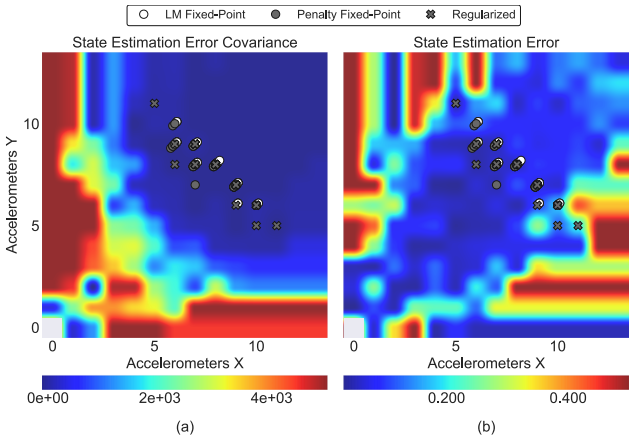


Figure 11.11.: Contours of estimation error and covariance, with the OSP solution using different formulations - Truss structure

At this point, it is worth mentioning that the choice of the reduction threshold ϵ_0 in the fixed-point formulation seems to have a small effect on the constrained objective function. In Figure 11.7(c) and Figure 11.7(d), the constrained objective functions using a fixed-point formulation with $\epsilon_0 = 10^{-4}$ and 10^{-2}, respectively, are plotted. For these values of ϵ_0, $n_s^* = 16$ and 10 are yielded, respectively. Although it could be expected that the value of ϵ_0 would possibly affect the results of the optimization problem, both plots show similar distribution of the objective function values with respect to the sensor configuration. For the coming investigations, $\epsilon_0 = 10^{-4}$ is chosen.

11.3 Problem Formulation and Constraint Enforcement

To illustrate the influence of the constraint functions choice and the constraint enforcement approaches on the result on the optimization problem, the problems in Figures 11.1 to 11.2 are evaluated using the following four settings:

1. LM, Fixed-point,
2. Penalty, Fixed-point,
3. Regularization.

To account for the possible variations occurring during the optimization procedure, the optimization problem is solved 30 times for each setting, each with a different random initialization of $\boldsymbol{\beta}$.

In some investigations, it would be necessary to examine the validity of a hypothesized assumption or observation for statistical significance. Herein, the following statistical inference tests [MGB74] are used:

- Two-tailed T- test: To test whether two samples (a and b) come from populations having identical means. For this test, the null and alternative hypotheses are written as:

$$\mathcal{H}_0 : \mu^{\mathrm{a}} = \mu^{\mathrm{b}}, \qquad \mathcal{H}_1 : \mu^{\mathrm{a}} \neq \mu^{\mathrm{b}}. \tag{11.3}$$

- One-tailed T-test: To test whether the mean population A is significantly less than the of population B based on the corresponding samples a and b. For this test, the null and alternative hypotheses are written as:

$$\mathcal{H}_0 : \mu^{\mathrm{a}} \leq \mu^{\mathrm{b}}, \qquad \mathcal{H}_1 : \mu^{\mathrm{a}} > \mu^{\mathrm{b}}. \tag{11.4}$$

Similarly, one-tailed and two-tailed Mood's test are also applied. These have a hypothesis formulation similar to the equivalent T-tests, except that the sample's median is used as a test statistic instead of the sample's mean.

For any of the above tests, the null hypothesis is rejected when the *p-value* is smaller than the significance value, α. In this work, $\alpha = 5 \times 10^{-3}$; the test inferences regarding the consistency of the

data with the null and alternative hypothesis are drawn with a confidence of 99.5%.

S-DOF System

In this study, $\epsilon_0 = 10^{-4}$ is used, and a fixed-point $n_s^* = 4$ is yielded.

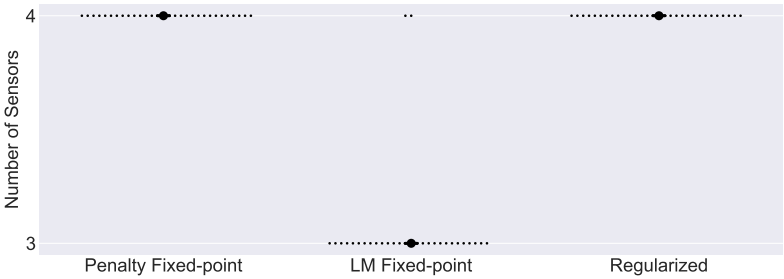


Figure 11.12.: Distributions of $\|\boldsymbol{\beta}^*\|_1$ for the different constraint treatment approaches - S-DOF system

As shown in Figure 11.12, enforcing the constraint using the penalty methods yields the target number of sensors for all the 30 experiments conducted, either using a regularized or a fixed-point formulation. On the other hand, using LM tends to yield one sensor less than n_s^*, with 28 out of 30 experiments yielding 3 sensors and only 2 yielding the target 4 sensors. Comparing the initialization of $\boldsymbol{\beta}$ for the experiments with 3 and 4 sensors using LM enforcement, it can be observed that the initial set of selection variables are different, and only certain initializations yield the desired n_s. On the other hand, the initial sets that yield 3 sensors using LM are all found to yield 4 sensors when using penalty.

This concludes that the LM enforcement is somehow sensitive to the initialization of the optimization variable, while this is not observed in the penalty approach, while the formulation of the constraint has no impact on the yielded number of sensors.

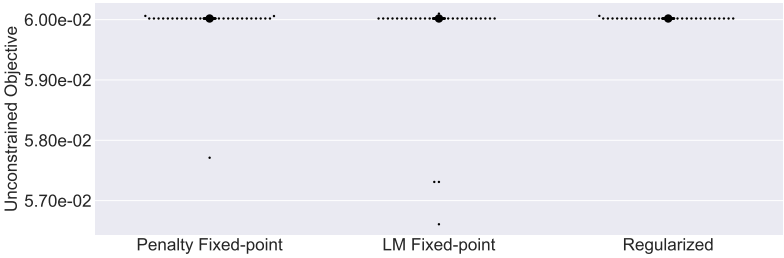


Figure 11.13.: Distribution of $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)$ for the different constraint treatment approaches - S-DOF system

The distribution of the objective function reached at the end of the optimization procedure is shown in Figure 11.13. This is the yielded value of the *unconstrained* term of the objective function $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\boldsymbol{\beta}^*)\right)$.

It can be observed that, on average, all algorithms yield the same value. The objective function from all experiments ranges between 5.66×10^{-2} and 6.00×10^{-2}, which is a rather insignificant deviation.

The regularization formulation results in the most consistent optimization results. For this formulation, 29 experiments out of 30 converged to the same state estimation covariance, despite the different initializations of $\boldsymbol{\beta}$.
Meanwhile, using a LM results in a slightly larger variance among the result set compared to the using penalty, as indicated by the wider scatter in Figure 11.13.

Figure 11.14 shows the distribution of the number of optimization iterations required to reach convergence for each formulation, n_{iter}. From the distributions, no significant difference in terms of speed of convergence could be observed, with almost all methods converging in 2 iterations.

Figure 11.14.: Distribution of n_{iter} for the different constraint treatment approaches - S-DOF system

As an additional step, using each of the three formulations, the resulting optimal sensors are used to estimate the hidden states of the system using a Kalman filter, as explained in Chapter 8. The error in the state estimation ϵ_x as well as the trace of the error covariance matrix $\mathrm{Tr}(\mathbf{P}_k^{x+})$ are evaluated after 400 Kalman filter steps. The distributions of both variables are shown in Figures 11.15 and 11.16.

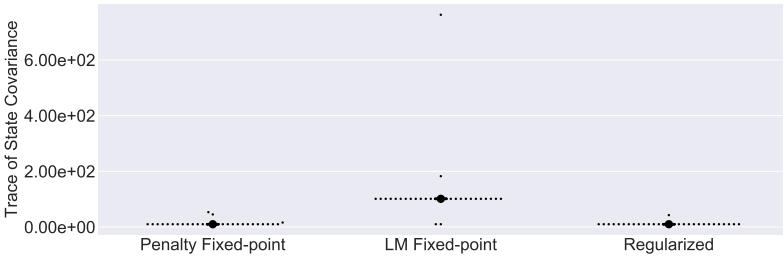


Figure 11.15.: Distribution of $\mathrm{Tr}(\mathbf{P}^{x+})$ for the different constraint treatment approaches - S-DOF system

To recall, the objective of the proposed sensor placement method is alleviate the uncertainty in the state estimate by reducing $\mathrm{Tr}(\mathbf{P}_k^{x+})$. As shown in Figure 11.15 using a constrained formulation with a penalty method yields a lower mean $\mathrm{Tr}(\mathbf{P}_k^{x+})$ compared to the LM and the regularized formulation. The same observation holds with respect to outliers, as hinted in the discussion about the objective function.

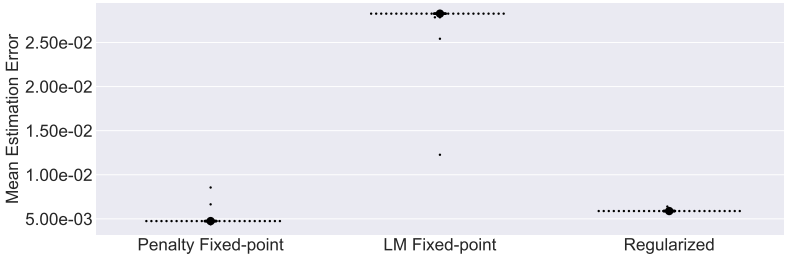


Figure 11.16.: Distribution of ϵ_x for the different constraint treatment approaches - S-DOF system

Given certain initial sensor configurations, the optimization yields outliers when employing LM, which is less likely when employing penalty.

Likewise for the state estimation error, the penalty method outperforms the LM method. However, since all experiments yield an error well below 5 %, the deviation between the algorithms could be neglected.

To sum up, for the S-DOF system, both the penalty method and the regularized formulations exhibit a more reliable constraint enforcement, yields lower state estimation error covariance, with no significant increase in the computational cost of optimization.

Truss Structure

As for the S-DOF system, $\epsilon_0 = 10^{-4}$ is used, and a fixed-point $n_s^* = 16$ is yielded. The difference between the two benchmarks is the extent of complexity, with this system having a larger space of sensor configurations, and `AccY` as an additional sensor type.

Examining the l_1-norm of $\boldsymbol{\beta}^*$ in Figure 11.12, the penalty method formulation and the regularized formulation meet the target n_s^* for all the 30 experiments. The results of the LM-constrained formulation evenly fluctuate between 15 and 16.

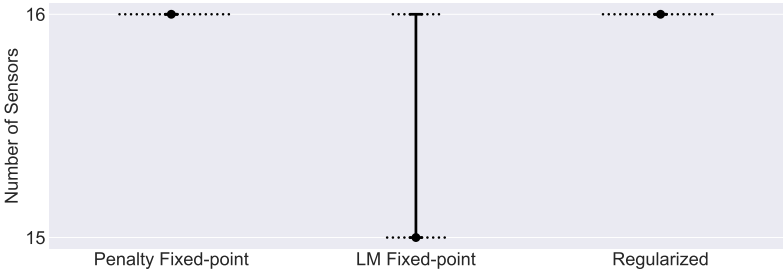


Figure 11.17.: Distributions of $\|\boldsymbol{\beta}^*\|_1$ for the different constraint treatment approaches - Truss structure

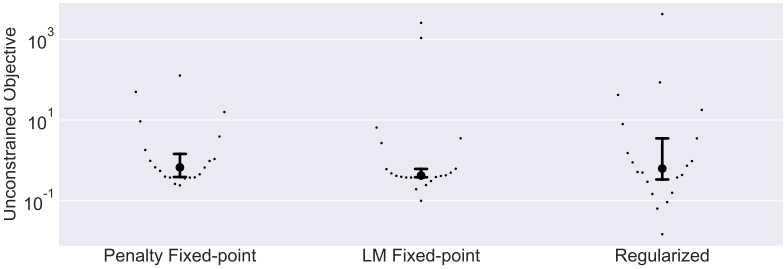


Figure 11.18.: Distribution of Tr $\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}}^*)\right)$ for the different constraint treatment approaches - Truss structure

Looking at Figure 11.18, it can be observed that the median values of the samples using the different algorithms are almost identical; 0.42 for LM Fixed-point, 0.67 for penalty fixed-point, and 0.63 for the regularized formulation. It needs to be remarked that the median in this case is investigated to alleviate the influence of the outliers, which can be observed in all samples.

This observation is further assured by a two-tailed Mood's test, having the hypotheses $\mathcal{H}_0 : \text{med}_{LM} = \text{med}_{Pen}$, $\mathcal{H}_0 : \text{med}_{LM} = \text{med}_{Reg}$, and $\mathcal{H}_0 : \text{med}_{Pen} = \text{med}_{Reg}$. The null hypothesis fails to be rejected in all cases, with p-values between 0.17 and 0.5. Based on the evidence from the samples, it can concluded that the approaches are unlikely to yield non-identical objective functions when repeated.

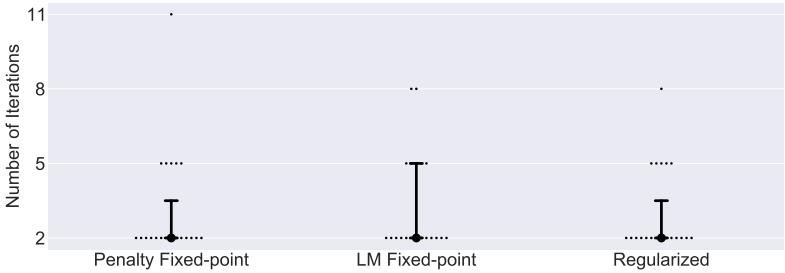


Figure 11.19.: Distribution of n_{iter} for the different constraint treatment approaches - Truss structure

A very efficient performance of all algorithms is demonstrated in Figure 11.17, where all approaches converge in two optimization iterations.

Speaking of the quality of the state estimate, when neglecting the outliers, both formulations seem to perform equally as shown in Figures 11.20 and 11.21. The estimation error for all cases lies below 2 %, including outlier sample points, whereas, the median values of the errors for all approaches' samples lie well below 1 %.

Visually, it seems that the LM method yields smaller estimation error medians than the penalty method. When conducting a one-tailed Mood's test to verify the null hypothesis

$$\mathcal{H}_0 : \operatorname{med}\left(\epsilon_x\right)^{\mathrm{Pen}} \leq \operatorname{med}\left(\epsilon_x\right)^{\mathrm{LM}},$$

both tests failed to reject $\mathcal{H}_0$. Similarly, when statistically comparing the LM-constrained formulation to the regularized formulation, $\mathcal{H}_0$ fails to be rejected. This concludes that the observed superiority of the LM formulation is statistically insignificant. Hence, the choice of the enforcement method are insignificant to the resultant estimation error.

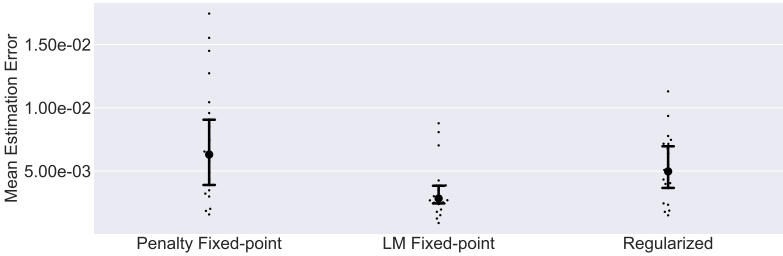


Figure 11.20.: Distribution of ϵ_x for the different constraint treatment approaches - Truss structure

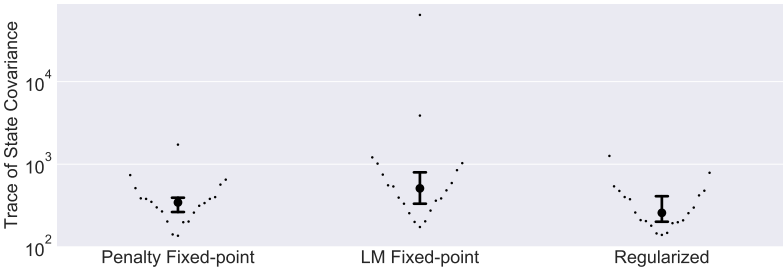


Figure 11.21.: Distribution of $\mathrm{Tr}(\mathbf{P}^{x+})$ for the different constraint treatment approaches - Truss structure

The last comparison metric is trace of the estimation error covariance, shown in Figure 11.21. The three approaches yield very close values of $\mathrm{Tr}(\mathbf{P}^{x+})$. Using the LM formulation yields a relatively higher average value of the trace, which could be justified by performing the estimation with one-less sensor on average, as shown in Figure 11.17. Although this deficiency in the number of sensors didn't have a notable influence on the precision of the state estimate, its impact on the estimation confidence, manifested by $\mathrm{Tr}(\mathbf{P}^{x+})$, is obvious.

Summing up the conclusions reached in this investigation, it has been found that

- With respect to the number of sensors, the penalty approach is more likely to yield the target sensors, while the LM ap-

proach has shown some evidence of sensitivity towards the initialization of the optimization variable.

- With respect to the estimation error, and the optimization procedure convergence, no approach is concluded to be favoured over the others.
- The concern that the regularized formulation might favour sparse sensor configurations is alleviated.
- All algorithms reach nearly the same objective function value at the end of the optimization algorithm, with the penalty fixed-point method showing a slightly higher robustness indicated by the tighter scatter around the average.
- The constrained formulation using the penalty method yields smaller state estimation covariance, hence lower estimation uncertainty.

Accordingly, the penalty method is favoured to the LM method, and for the following investigations, the penalty fixed-point is chosen as the gradient-based OSP method.

11.4 Multi-Sensor Configuration

In the following paragraphs, the influence of limiting candidate sensors to a single type versus choosing from both sensor types is investigated. Moreover, the choice of the sensor type (accelerometers vs. strain gauges) in case of single sensor type is studied.

The samples for this study are generated as follows; for each of the two considered systems, the domain of candidate sensors is limited to either strain gauges only, accelerometers only (for the truss structure, this includes x- and y-accelerometers), or both (denoted by "Accel + SG"). The former two are referred to as single-type configurations, and the latter as multi-type. For each set of candidate sensors, 30 unique random initializations of the selection variables β_i are made, and the optimization problem is solved. The chosen optimizer setting

throughout this investigation is *fixed-point* formulation for defining the constraint, and the penalty method is applied to enforce it.

For this investigation, the following two claims are tested for validity based on the evidence from the samples:

- *Claim 1:* Using single-type sensor configuration, accelerometers yield lower estimation errors, as well as lower estimation error covariance than strain gauges. The corresponding null and alternative hypotheses are

$$\mathcal{H}_0 : \mu(\cdot)^{\mathrm{SG}} \leq \mu(\cdot)^{\mathrm{Acc}}, \qquad \mathcal{H}_1 : \mu(\cdot)^{\mathrm{SG}} > \mu(\cdot)^{\mathrm{Acc}}. \tag{11.5}$$

- *Claim 2:* Using multi sensor-type configuration yields lower estimation errors, as well as lower estimation error covariance than single sensor-type configuration. The corresponding null and alternative hypotheses are

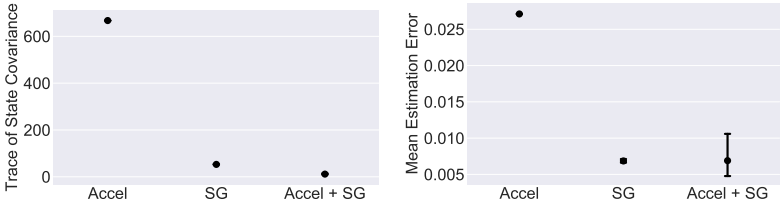
$$\mathcal{H}_0 : \mu(\cdot)^{\mathrm{single}} \leq \mu(\cdot)^{\mathrm{multi}}, \qquad \mathcal{H}_1 : \mu(\cdot)^{\mathrm{single}} > \mu(\cdot)^{\mathrm{multi}}. \tag{11.6}$$

Both claims are tested using a one-tailed T-test with a significance value $\alpha = 5 \times 10^{-3}$.

S-DOF System

Figure 11.22 shows the distribution of the state estimation error and state estimation error covariance when using different sensor configurations.

Clearly, a multi-type sensor configuration results in 20-folds lower estimation error covariance, as shown in Figure 11.22(a). Also, a multi-type configuration results in a smaller uncertainty in the state estimates more consistently, as indicated by the smaller sample variance. With respect to the choice of the sensor type in case of single-type configurations, Figure 11.22(b) indicates that strain gauges yield lower covariances.



(a) $\mathrm{Tr}(\mathbf{P}^{x+})$ (b) ϵ_x

Figure 11.22.: Variation of the estimation error and covariance with respect to the configuration and type of sensors - S-DOF system

Additionally, using a multi-type configuration results for state estimation yields a smaller estimation covariance than any single-type configuration. This is visually obvious when comparing multi-type to accelerometers in Figure 11.22. To confirm the validity of this for multi-type vs. strain gauges, a one-tailed T-test is conducted, with the null hypothesis being $\mathcal{H}_0 : \mu^{\mathrm{SG}} \leq \mu^{\mathrm{multi}}$. This is, in fact, the statistical significance test of claim 2 stated above, with the single-type sample being exclusive to the strain-gauges. The resultant p-value of that test is $10^{-27} \approx 0$, thus the null hypothesis is confidently rejected. Hence, it is concluded that the multi-type configuration yields a smaller estimation covariance.

Looking at the state estimation error distributions, using strain gauges alone yields a significantly lower state estimation error than using accelerometers alone; almost 10-folds smaller on average. Astonishingly, the precision of the estimate when combining accelerometers and strain gauges didn't surpass that of the strain gauges alone, which is seconded by the one-tailed T-test. The p-value calculated from the strain-gauges' experiments and the multi-type experiment is $0.490 > \alpha$, hence providing no sufficient evidence to reject $\mathcal{H}_0$.

From the 30 experiments conducted for each of the candidate sensor configurations, the number of times each sensor position is chosen as optimal is counted, and the corresponding tally is shown in

Figure 11.23(a). Using a reduction tolerance, $\epsilon_0 = 10^{-4}$, the fixed points for accelerometers only, strain gauges only, and multi-type configurations are 3, 3, and 4, respectively. In light of this, the most frequently selected positions for each configuration are chosen, and shown on the structure in Figures 11.23(b) to 11.23(d).

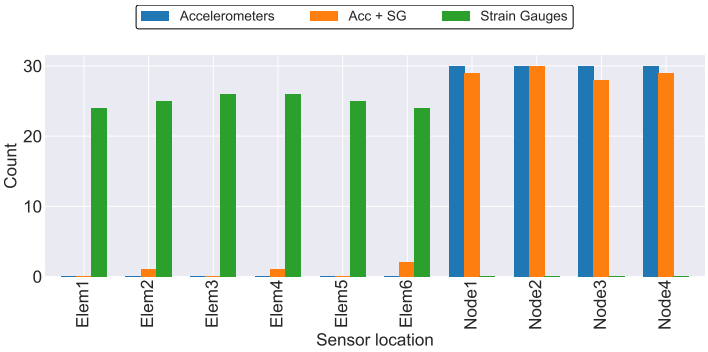
Here, it can be seen that the multi-type optimal sensor configuration is not simply the union of the single-type optimal sensor positions. In fact, the most selected sensors in a multi-type configuration are accelerometers. This could possibly be explained by the lower measurement noise attributed to accelerometers, compared to strain gauges.

Truss Structure

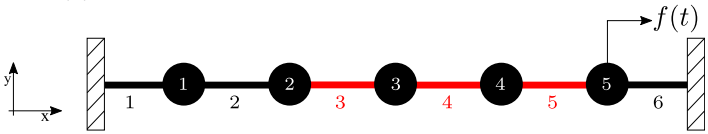
Looking at the second system, Figure 11.24 shows the distribution of the state estimation error and state estimation error covariance when using different sensor configurations.

Alike the S-DOF system, a multi-type sensor configuration results in lower estimation error covariance, as shown in Figure 11.24(a), and the strain gauges outperform accelerometers with respect to reducing state estimation uncertainty. This accords with the hypothesis test results in Table 11.1. It is shown that no evidence is presented to support the claim that accelerometers result in lower estimation covariance. On the other hand, the claim that multi-type configurations surpass single-type configurations (whether accelerometers or strain gauges) is assured by rejecting the null hypothesis.

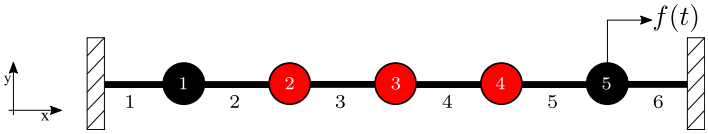
As previously observed in the S-DOF system, strain gauges state estimation precision outperforms accelerometers, as shown in Figure 11.24(b). On the other hand, using a multi-type configuration yields an even smaller estimation error compared to any of the single-type sensor configurations. Likewise, the one-tailed T-test results in Table 11.1 affirm this, with all the claim 2 null hypotheses being rejected.



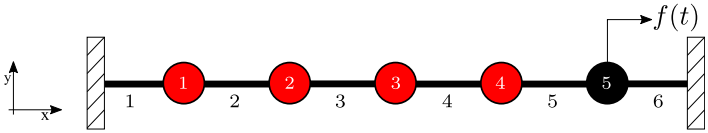
(a) Count of sensor positions being in the optimal solution



(b) Strain Gauges Only



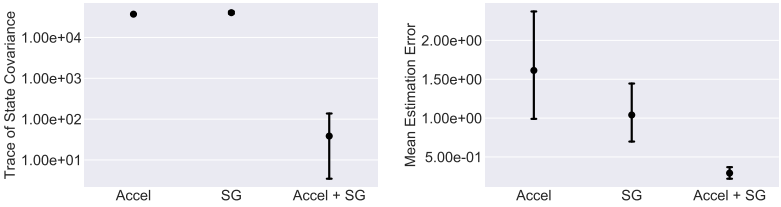
(c) Accelerometers Only



(d) Multi-type

Figure 11.23.: Most selected sensor positions - S-DOF system

The individual count of sensor positions in the samples is shown in Figure 11.25(a). It may be noticed that the multi-type configuration is dominated by accelerometers, with very few experiments involving strain gauges choice. Figures 11.25(b) to 11.25(d) show the positions of the most frequently chosen 3 strain gauges, 3 accelerometers, and 16 multi-type sensors, as dictated by the fixed-points calculated with $\epsilon_0 = 10^{-4}$.



(a) $\mathrm{Tr}(\mathbf{P}^{x+})$ (b) ϵ_x

Figure 11.24.: Variation of the estimation error and covariance with respect to the configuration and type of sensors - Truss structure

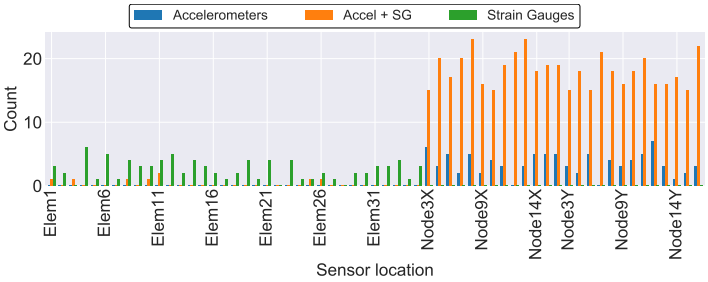
Table 11.1.: Results of the statistical significance test - Truss structure

	$\mu(\cdot)^{\mathrm{SG}} \leq \mu(\cdot)^{\mathrm{Acc}}$		$\mu(\cdot)^{\mathrm{SG}} \leq \mu(\cdot)^{\mathrm{multi}}$	
Variable	p-value	$\mathcal{H}_0$	p-value	$\mathcal{H}_0$
$\mathrm{Tr}(\mathbf{P}^{x+})$	9.204×10^{-3}	No Evidence	0.00	Reject
ϵ_x	8.278×10^{-2}	No Evidence	1.049×10^{-3}	Reject

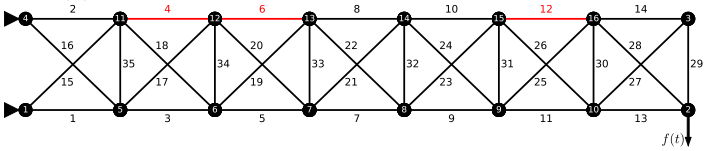
	$\mu(\cdot)^{\mathrm{Acc}} \leq \mu(\cdot)^{\mathrm{multi}}$	
Variable	p-value	$\mathcal{H}_0$
$\mathrm{Tr}(\mathbf{P}^{x+})$	0.00	Reject
ϵ_x	7.574×10^{-4}	Reject

The distribution of the multi-type configuration sensors on the structure is, at the first glance, confusing. No accelerometers are placed at node 3. This would have been a first educated guess, especially from those with high confidence in their engineering intuition, given a vertically oriented load. Recalling that the goal of the sensor configuration is to reduce the estimation error covariance – in other words, reduce estimation uncertainty – it becomes more intuitive to think of the optimal sensor configuration as the one that observes a broader spatial spectrum of the structure, even near its supports.

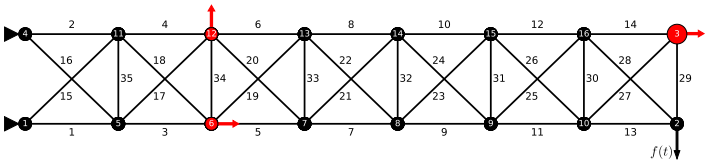
This explains the scatter of the sensor over the upper and lower *fibres* of the truss structure.



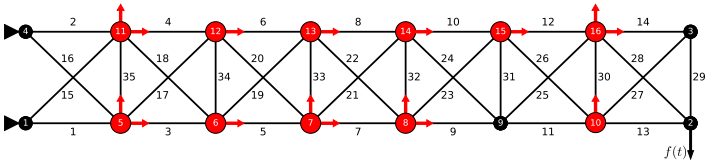
(a) Count of sensor positions being in the optimal solution



(b) Strain Gauges Only



(c) Accelerometers Only



(d) Multi-type

Figure 11.25.: Most selected sensor positions - Truss structure

11.5 Comparison to Sequential Approach

In the following paragraphs, the proposed gradient-based OSP approach is compared with the greedy forward sequential OSP approach. The sequential approach for OSP is the most abundant method in lit-

erature being applied to many objective functions used by authors to measure the quality of sensor configurations, e.g. [Kam91], [Kam05], [KP08], [Pap+11], [ZX16], and [Tam+16].

The sequential approach is compared with a penalty fixed-point gradient-based approach, since it was shown to outperform the other constraint enforcement approaches in the previous paragraphs. Accordingly, the stopping criteria of the additive sequential OSP procedure is terminated when reaching the *fixed-point*, n_s^*.

Herein, two results are compared; the state estimation error and the error estimation covariance, shown in Figure 11.26. Moreover, the respective mean values of 30 experiments applying gradient-based OSP and the result of the sequential OSP are summarized in Tables 11.2 and 11.3.

Table 11.2.: Comparison between gradient-based and sequential sensors - S-DOF system

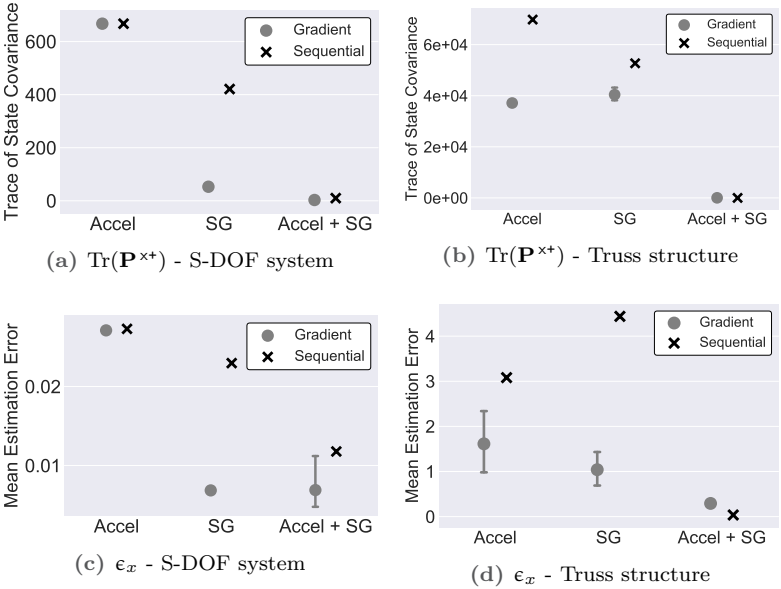
	Method	$\mathrm{Tr}(\mathbf{P}^{x+})$	ϵ_x
Accelerometers	Gradient	6.67×10^{2}	2.70×10^{-2}
	Sequential	6.67×10^{2}	2.70×10^{-2}
Strain Gauges	Gradient	5.44×10^{1}	7.00×10^{-3}
	Sequential	4.21×10^{2}	2.30×10^{-2}
Multi-type	Gradient	3.75×10^{-1}	5.00×10^{-3}
	Sequential	1.00×10^{1}	1.20×10^{-2}

Table 11.3.: Comparison between gradient-based and sequential sensors - Truss system

	Method	$\mathrm{Tr}(\mathbf{P}^{x+})$	ϵ_x
Accelerometers	Gradient	3.74×10^4	1.50×10^{-2}
	Sequential	6.98×10^4	5.80×10^{-2}
Strain Gauges	Gradient	3.77×10^4	1.40×10^{-2}
	Sequential	5.27×10^4	8.40×10^{-2}
Multi-type	Gradient	2.66	5.00×10^{-3}
	Sequential	4.96	1.00×10^{-3}

With respect to $\mathrm{Tr}(\mathbf{P}^{x+})$, the gradient-based OSP results consistently in lower state estimation error covariance, which reflects an improvement in the state estimation confidence. Applying the proposed method on the S-DOF system results in a relative reduction in the state estimation uncertainty by 87 % and 96 % for the strain-gauges and multi-type configurations, respectively. For the accelerometers configuration, both approaches yielded identical configurations, hence equal values of $\mathrm{Tr}(\mathbf{P}^{x+})$. For the truss structure, the improvement achieved by the gradient-based method is 46 % and 23 % for the single-type and multi-type configurations, respectively.

Both algorithms yield very precise state estimates, with respective errors not exceeding 10 % in any experiment. When comparing the approaches, the gradient-based approach demonstrates steadily lower estimation errors than the sequential approach. For the S-DOF system, the error drops by 70 % for strain gauges and by 58 % for multi-type configurations, whereas for the truss structure, the error for the accelerometers and strain gauges configurations drops to 74 % and 83 %. For a multi-type configuration, the gradient-based approach yielded a higher average estimation error than the sequential. Nonetheless, both errors retain equivalent order of magnitude and are both below 1 %, making the estimate error in an acceptable range.



(a) $\mathrm{Tr}(\mathbf{P}^{x+})$ - S-DOF system

(b) $\mathrm{Tr}(\mathbf{P}^{x+})$ - Truss structure

(c) ϵ_x - S-DOF system

(d) ϵ_x - Truss structure

Figure 11.26.: Variation of the estimation error and covariance with respect to the configuration and type of sensors - Truss structure

11.6 Damage Estimation

Damage estimation based on the captured sensor signals is the goal of SHM. That a procedure for configuring an optimal sensors constellation has been investigated, the captured sensor data are used to estimate the damage the structure is prone to experiencing under prospective loading conditions.

Herein, damage introduced in the structure can be estimated as explained in Chapter 7. To recall, the damage can be inferred based on the strain signals, by correlating strains to stresses using the constitutive law, or based on the acceleration signals, by applying a more complex function to correlate both quantities. The function involves applying a double integral on the acceleration signal to

infer the nodal displacements, from which the elements' strains are computed, as per Equation (7.4). Clearly, using acceleration signals to infer damage could involve higher process error, which could downgrade the quality of the *damage* estimate, even if the associated measurement error is lower.

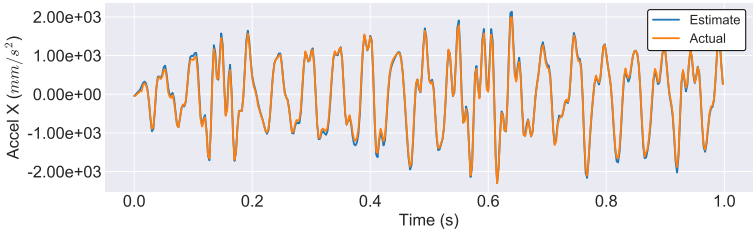
As shown in the previous paragraphs, the optimal sensor configuration is mostly dominated by accelerometers. Hence, an attempt to reconstruct the strain field from the chosen acceleration measurements is undergone. Without being exhaustive, this following discussion will consider the nodes associated to the highest and lowest strain amplitudes, as these present the limits of the foreseen damage spectrum. Also, these present the extrema of the signal-to-noise ratios attainable for this problem, since all sensors of the same type have identical and independent measurement noise.

S-DOF System

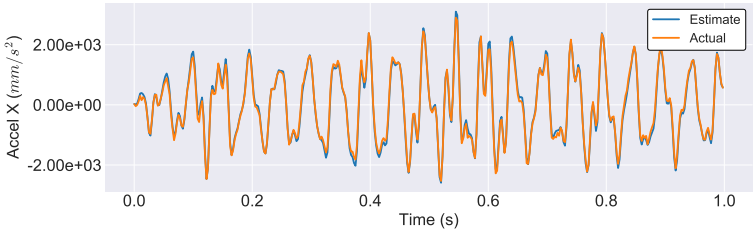
From the simulation results of the perturbed system, element 4 experiences the smallest strain, in the order of 10^{-4}, while element 6 experiences the largest strain, in the order of 10^{-3}. Element 4 connects nodes 3 and 4, while element 6 connects node 5 and the fixed support. The acceleration signals of the three nodes are shown in Figure 11.27.

In each of the three plots in Figure 11.27, the actual acceleration signal is plotted. Additionally, the mean and the one-standard-deviation (1-STD) band are plotted based on a data set of 30 experiments using a penalty fixed-point approach. All reconstructed acceleration signals exhibit a remarkable overlap with the actual signals, with a mean relative error of 7.74 %.

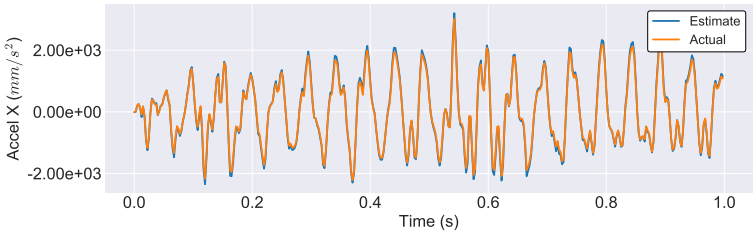
Figures 11.28(a) and 11.28(c) show the strain signals for elements 4 and 6, respectively. The actual signal is obtained from the perturbed system simulation, whereas the estimated signals are reconstructed from the nodal accelerations in Figure 11.27. The frequency spectra



(a) Acceleration X, node 3 - S-DOF system



(b) Acceleration X, node 4 - S-DOF system

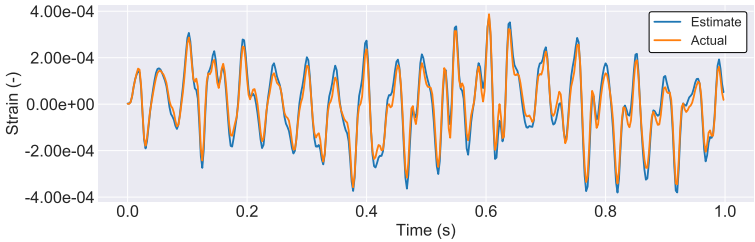


(c) Acceleration X, node 5 - S-DOF system

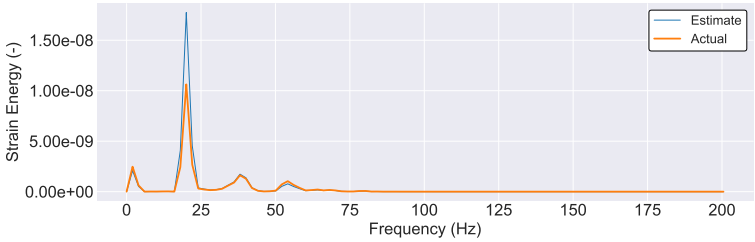
Figure 11.27.: Accelerations of the nodes associated to the elements with lowest and highest strains - S-DOF system

of both strain signals is obtained by applying a Fast Fourier Transform (FFT), and the resulting spectra are shown in Figures 11.28(b) and 11.28(d), respectively.

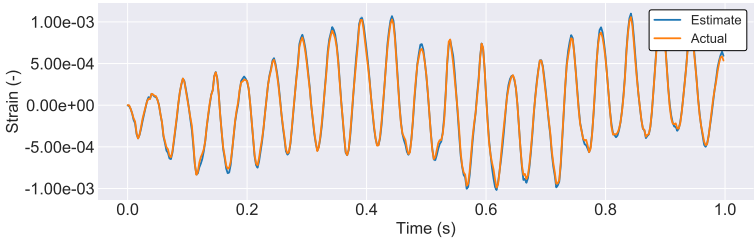
As shown in the Figure 11.28, using the proposed gradient-based approach, the difference between the actual and the reconstructed strain is rather insignificant. Based on the simulation results, given the described load, elements 4 and 6 experience damage of 0.13 %



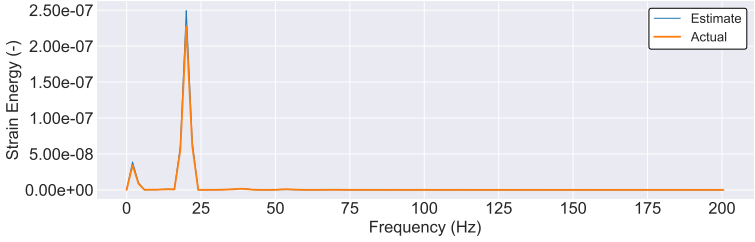
(a) ε_4 - S-DOF system



(b) Frequency spectrum ε_4 - S-DOF system



(c) ε_6 - S-DOF system



(d) Frequency spectrum ε_6 - S-DOF system

Figure 11.28.: Reconstructed strain signals and corresponding frequency spectra for elements with lowest and highest strains - S-DOF system

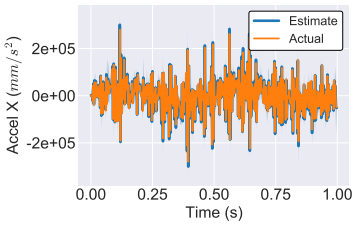
and 0.71 %, which, using the reconstructed signals, are estimated to be 0.11 % and 0.67 %, respectively.

Truss Structure

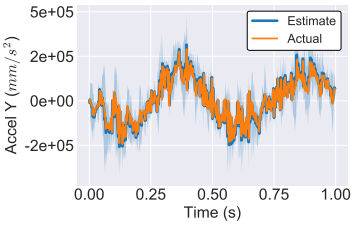
From the simulation results of the perturbed system, it can be determined that element 16 experiences the smallest strain, in the order of 10^{-3}, while element 27 experiences the largest strain, in the order of 10^{-1}. The nodes associated with these two elements are (1, 11) and (10, 3), respectively. Since node 1 is fixed in all directions, the acceleration signals in x and y directions of nodes 3, 10, and 11 are shown in Figure 11.29. In each of the six plots, the actual acceleration signal, along with the reconstructed signal from the sensors obtained using the gradient-based OSP algorithm are plotted. For the OSP results, the mean and the 1-STD band are plotted based on a data set of 30 experiments. All reconstructed acceleration signals exhibit a remarkable overlap with the actual signals, even capturing the high-frequency vibrations, with relative errors 9.21 % and 7.95 % for the x- and y-accelerations, respectively.

Figures 11.30(a) and 11.30(c) show the strain signals for elements 16 and 27, respectively. The actual signal is obtained from the perturbed system simulation, whereas the other two signals are reconstructed from the nodal accelerations in Figure 11.29. The frequency spectra of both strain signals is obtained by applying a FFT, and the resulting spectra are shown in Figures 11.30(b) and 11.30(d), respectively.

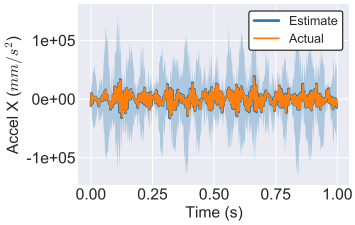
As shown in Figure 11.30, the difference between the actual and the reconstructed strain almost vanishes in the time-domain, with the actual signal lying within the bounds of the 1-STD band of the reconstructed signals. In the frequency domain, insignificant deviation between both spectra is also observed. From the simulation results, elements 16 and 27 experience damages of 3.87 % and 51.82 %, respectively. Estimating the damage from the reconstructed signals, the obtained damage estimates evaluate to 4.58 % and 52.85 %.



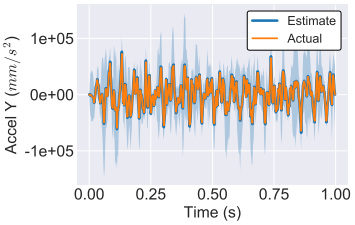
(a) Acceleration X, node 3 - Truss structure



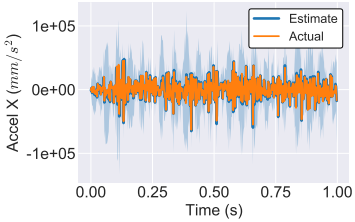
(b) Acceleration Y, node 3 - Truss structure



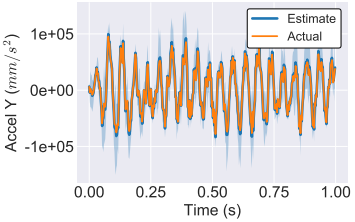
(c) Acceleration X, node 10 - Truss structure



(d) Acceleration Y, node 10 - Truss structure

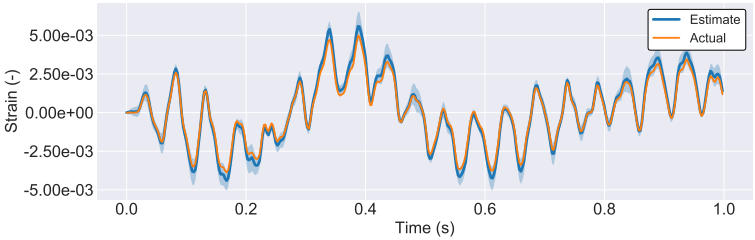


(e) Acceleration X, node 11 - Truss structure

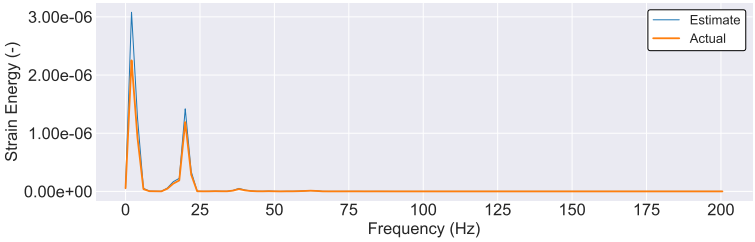


(f) Acceleration Y, node 11 - Truss structure

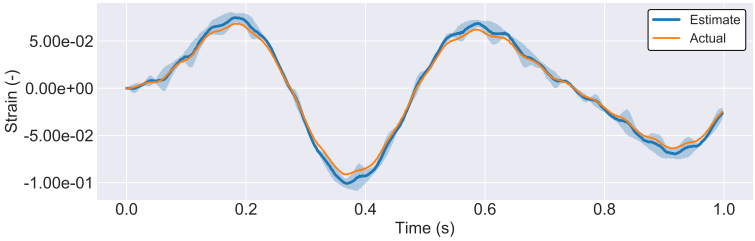
Figure 11.29.: Accelerations of the nodes associated to the elements with lowest and highest strains - Truss structure



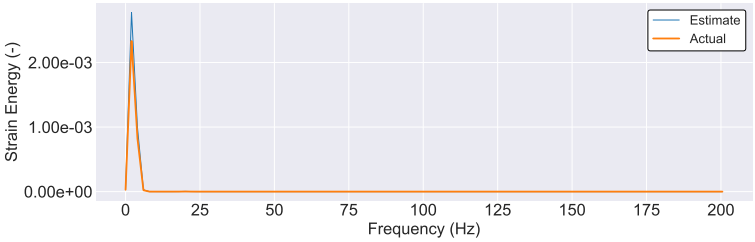
(a) ε_{16} - Truss structure



(b) Frequency spectrum ε_{16} - Truss structure



(c) ε_{27} - Truss structure



(d) Frequency spectrum ε_{27} - Truss structure

Figure 11.30.: Reconstructed strain signals and corresponding frequency spectra for elements with lowest and highest strains - Truss structure

Part V

Prognosis and Damage

12 Motivation

In Section 2.3, it was introduced that prognosis is one of the execution modules as well as goals of a Predictive Maintenance (PdM) routine. In the context of the upcoming discussion, we refer to prognosis as a functional block receiving the predicted system states as input, and yielding an estimate of the accumulated damage as an output. This estimate is fed as an input to the proceeding functional blocks in the PdM routine to take or suggest maintenance-related decisions such as operating conditions adjustments, failure and anomalies corrections, rescheduling maintenance or overhaul, etc.

Having such an importance in the context of PdM, many models have been proposed and developed over the previous decades for predicting damage in operating systems. Prognosis methods can be categorized into three main classes:

- *Physics-based models*: These are models which require the description of the system's behaviour based on the physical laws governing its dynamics and degradation. Among these laws are the laws representing the evolution of damage in the system, which could be both deterministic or stochastic [CNM11; Wan10a]. These models offer more precise and interpretable results as they, in fact, do mimic the system's behaviour, and their parameters are related to physical properties of the system. This facilitates the process of estimating damage based on a current system state and investigating its root cause, in comparison to the other two models. However, the construction of such models retains a great deal of assumptions and uncertainties, e.g. within material consecutive laws, or defining components' connections and interfaces. Typically,

such non-deterministic properties are modelled by stochastic random variables, which are calibrated based on measurements from the real system, presumed that the model represents its corresponding state [Din08b; Ise05]. Even though, for real systems, it is challenging to obtain an analytical description of degradation processes due to their complexity. Another limitation lies in the restriction to the transferability of a model specifically developed for one system to another.

- *Data-driven models*: These models exploit the abundance of monitoring data and pattern-recognition models to learn a surrogate relationship between the system inputs and outputs, and derive characteristics depicting the system's behaviour and degradation. Data-driven models are particularly advantageous when limited knowledge about the system's physical laws compared to input-output data are available. These models are comparably simpler and require less effort to build. However, they demand intricate pre-processing of monitoring data to extract the relevant features [Din+11; JQ03]. Also, they lack interpretability in most of the cases, especially in engineering applications[1]
- *Hybrid models*: These models offer an integration of a physical model and a data-driven approach aiming to combine the benefits of both. The integration could be achieved in either of two manners; a *series* manner, where data driven models are used to estimate an uncertain and unobserved parameter in the physical model, thus complementing its definition, or a *fusion* manner, where the outputs of both models are combined to reconstruct a global output. Either ways, hybrid models are gaining a lot of attention due to their exhibited performance and precision. Nonetheless, they could be computationally costly to deploy [RPK19].

[1] Currently, intensive research is being conducted to improve the interpretability of data-driven models, yet it is still more prominent in natural language processing and computer vision domains [Sam19].

It is rather typical in industrial applications to have systems composed of an array of different assets, which need to be monitored and maintained. When considering the choice of a damage estimation method, the degree of variability of system assets needs to be accounted for, while not overseeing the extent of homogeneity among them.

A single-asset system typically holds one or more identical assets. An example on that could be an array of bearings in machines, all having the same designation; hence, identical dimensions, identical build-up, identical operating principles, identical range of load bearing capacities (reflected in stress and heat distribution in the bearing), as well as identical set of potential failure mechanisms. In this case, it suffices having a single damage model (whether physics-based, data-driven, or hybrid) that could be instantiated and executed for each asset. The different instances of this model would differ in the input system states and hence output damage estimates as the operating conditions and anomalies differ from one to the other.

In the context of this work, a multi-asset system refers to a system comprising assets with variant principles of operations, build-ups, or purposes. With respect to a multi-asset system, a system composed of heterogeneous assets is referred to as a *portfolio*, while a system composed of homogeneous assets is referred to as a *fleet* [Pet18].
Taking bearings as a clarification example. In terms of variant *operation principles*, a system composed of roller bearings and fluid bearings is a multi-asset system, since the mode of operation is distinct between both families of bearings, and has a direct impact on failure modes and degradation. In terms of variant *build-up*, a system composed of roller bearings and needle bearings is considered a multi-asset system. In this case, both bearings belong to the rolling element bearing family (i.e. same operation principle), yet they differ in structural build-up. Needle bearings have slenderer races and rolling elements resulting in reduced stress peaks under axial load compared to roller bearings. This in turn reflects on the

failure mechanisms; needle bearings typically experience failure due to brinelling, while roller bearings tend to experience spalling and pitting. The two previous cases are examples of *portfolios*.
On the other hand, a system composed of different roller bearings is considered a *fleet*, even if the bearings vary in dimensions or run under different operating parameters. All the assets have the same operating principles and build-up, thus resulting in similar loading profiles. The variation in dimensions and operating conditions reflects mainly in the magnitude of loading and stress but not in their distribution profiles. Hence, for the different roller bearings, similar degradation and failure mechanisms could be expected [AG03].

When deriving a damage model for single-asset systems, physics-based models is relatively less arduous since only one asset with its respective degradation model is under consideration. On the other hand, for data-driven models, the diversity of the experimental and operational data required by the model is limited to the diversity of the operating conditions only; i.e. diversity of the model inputs. This means that historical data of different assets can be accumulated for the model development.
In multi-asset systems, the diversity of the assets must be covered in the damage models. Considering the example of the portfolio of roller and fluid bearings, two physical models would need to be developed. With respect to data-driven model, either two separate models would need to be developed, or alternatively a more sophisticated model which can distinguish between the operation principles of the two bearings. The latter is more computationally intensive and both alternatives would require a larger dataset to develop the model.

Prognosis models for fleets lie midway. With fleets, the similarity in the structural build-up and operation principles allows, to an extent, for the re-usage of models from other assets. It should be made clear that damage estimates from those models are not precise due to the dimensionality variance.

Take an example of a fleet of two roller bearings with one having shorter roller length than the other, yet both are undergoing the same cyclic loads. Both bearings will have equivalent contact stress distribution profiles, yet the one with shorter rollers will experience higher contact pressure, thus higher stress peaks on the races. If the applied load is such that the stress in one bearing cross the yield strength of the material, while the stress in the other is still below, different material fatigue laws would need to be applied to estimate the damage. For the one with stress crossing the yield limit, High-Cycle Fatigue (HCF) will not be valid in this range of loading. Alternatively, data-driven models can be used. The degradation in the bearing with stress below the yield strength is stress-driven, while in the other the degradation is strain- and crack-driven [ETJ15]. Plasticity associated with strain-driven degradation induces different dynamics to the system's output signals, while cracks developing in the races induce cyclic impulses in the output signals. The signal profile observed for a system experiencing stress-driven degradation is different. Hence, using one bearing's set of signals to train the model for the other bearing's model means that the training set would differ from the deployment set, leading to large errors in the damage estimates. The same problem would arise if a fleet is composed of two roller bearings with identical dimensions, but one is loaded more than the other, to the extent that it experiences plasticity and crack initiation, while the other does not.

From this discussion, it can be concluded that, for fleets, models could be transferred among the assets, however, only within a limited range of operation.

For the scope of this work, the focus is dedicated to damage estimation for fleets, and not portfolios. Motivated by the discussion in the previous paragraphs, the approach Simulation Models and Artificial Intelligence Combined (*moSAIc*) is proposed as a methodology for damage estimation for fleet assets based on hybrid modelling. The hybrid model consists of an ensemble of physics-based models, whose

damage estimates are combined using a weighted average to yield a damage estimate of an asset that has no damage estimation model. The weights assigned to each simulation model are calculated by a data-driven model, which is trained to generate weights such that the deviation between the estimated damage value and the corresponding ground-truth value is minimized.

The following chapters are organized as follows:
In Chapter 13, ensemble and Mixture of Experts (MoE) models are first introduced as a foundation for *moSAIc*, followed by a detailed formulation of the proposed methodology, and a description of the architecture of the hybrid model. Afterwards, the specific model preparation steps as well as the model pipeline are elaborated. Finally, in Chapter 14, a benchmarking example is presented to investigate the robustness and sensitivity of the model to its constituents, parameters, and assumptions. In addition, the model's performance is compared to that of purely data-driven models applicable to the problem of damage estimation in fleets.

13 Methodology

As pointed out earlier, the fundamental concept of data-driven approaches is trying to find a black-box relationship between inputs and outputs by adjusting the weights and parameters of the underlying model. In the context of supervised Machine Learning (ML), the hypothesis space is searched for an appropriate mapping between the inputs and the outputs, such that it satisfies, as much as possible, the prescribed performance metric, e.g. accuracy, precision, etc. In many cases, different methods compete in terms of performance on the considered task. In fact, in most settings, there is not a single method that consistently outperforms the others, rather, different methods achieve different performance levels in different input domains. Hence, it can be inadequate to restrict the choice of the method to a single model, and one could instead favour combining the results of different base models together under a committee or *ensemble*, in a process referred to as *ensemble learning*. Different forms of ensemble learning include the combination of different model results through Bayesian averaging, Random Forests, Bootstrap Aggregation (Bagging), and Boosting [ZM12].

13.1 Mixture of Experts

One form of ensemble learning is the Mixture of Experts (MoE) models. In Mixture of Expertss (MoEs), it is assumed or proven that no stand-alone model can attempt the underlying problem. Hence, the problem space is divided into different subspaces, each is allocated to an expert base model, hypothesized to yield optimal performance on such a subspace. The results of different base models are combined to create the relationship between the input-output

sets of the problem. The combination is achieved by associating each base model with a weight, which is typically decided by a gating network. The weights in that sense resemble the competency of the expert base model towards the whole problem. Building an MoE model occurs in two folds: 1) selecting and building base models, and 2) training the gating network to yield optimal weights, such that the output of the MoE satisfies the prescribed performance metrics [ME14].

13.2 Architecture of *moSAIc*

Having given an overview of the MoE in the previous paragraphs, in the following paragraphs, the details of the proposed model, Simulation Models and Artificial Intelligence Combined (*moSAIc*), is explained. The fundamental idea of *moSAIc* is to exploit the existing similarity between the assets' geometric features, material laws, and operation principles, which implicitly influence the failure mechanisms, such that the existing damage models of assets can be re-used by other assets.

The first works on *moSAIc* were first introduced in the collaborative work [Hil+19], where the concept of using physics-based degradation models as experts, and combining them linearly using a trained Multi-Layer Perceptron (MLP) was proposed and demonstrated on a synthetic example. To the time of writing this work, a MoE model with physics-based expert models is a novelty, and [Hil+19] is considered the first contribution in that direction. The model however didn't incorporate either extraction of time-dependent features from signals, and didn't provide a criteria for selecting the *expert* degradation models. Later, in [Ber+19], the same authors highlighted the impact of expert models selection on the hybrid model's performance, albeit the absence of proposing a systematic approach for this selection. These two aspects are explicitly attempted rigorously in the scope of the current work.

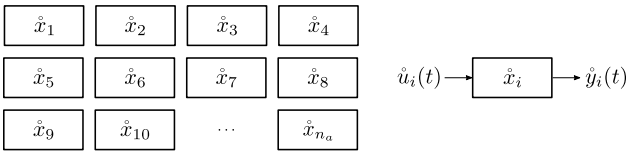


Figure 13.1.: A fleet of assets

Consider a fleet of n_a distinct assets $\mathring{\mathcal{X}} = \{\mathring{x}_1, \ldots, \mathring{x}_{n_a}\}$. Each asset $\mathring{x}_i$ is represented by the states of an engineering system whose inputs and outputs are the tensors $\mathring{\mathbf{u}}_i$ and $\mathring{\mathbf{y}}_i$, respectively, see Figure 13.1. In the context of a damage calculation problem, $\mathring{\mathbf{u}}$ is an array of time-variant operating conditions of the system, and $\mathring{\mathbf{y}}$ is the accumulated damage of the asset.

The estimate of the damage of $\mathring{x}_i$ could be evaluated using a dedicated degradation model, x_i, defined as $x_i : \mathbf{u}_i \mapsto y_i$. Herein, $\mathbf{u}_i$ is a vector of inputs of the i^{th} model, which is the discrete form of the i^{th} asset's time-variant operating parameters, and y_i is the corresponding damage estimate. Due to modelling and measurement errors, ${d^w}_i$ and ${d^v}_i$, respectively, deviations between the real asset's damage value $\mathring{y}$ and the corresponding model's estimate y usually exist.

The set of time-variant inputs of the i^{th} model, $\mathring{x}_i$, is referred to as the dynamic features, and denoted by ${\mathcal{F}^{\mathsf{U}}}_i$, and defined as:

$$ {\mathcal{F}^{\mathsf{U}}}_i = \left\{ u_i^{(1)}(t),\ \ldots\ , u_i^{(\mathsf{n_D})}(t) \right\} \tag{13.1} $$

where $u_i^{(\mathsf{j})}$ is the j^{th} dynamic feature of the i^{th} asset.

Moreover, each i^{th} asset $\mathring{x}_i$ is characterized by a unique set of static, time-invariant parameters, such as geometric dimensions, material properties, etc. known as the static features, and denoted by ${\mathcal{F}^{\mathsf{C}}}_i$. The set of static features of $\mathring{x}_i$ is defined as:

$$ {\mathcal{F}^{\mathsf{C}}}_i = \left\{ c_i^{(1)},\ \ldots\ , c_i^{(\mathsf{n_C})} \right\}, \tag{13.2} $$

where $c_i^{(j)}$ is the j^{th} static feature of the i^{th} asset. The static features are used as parameters in the damage models; i.e. $x_i = x_i \left[\mathcal{F}^{C}{}_i\right]$, with [.] denoting the parametrisation of x_i by $\mathcal{F}^{C}{}_i$. The parametrization notation will be dropped out throughout the discussion for the sake of notation simplicity.

Herein, the set of dynamic and static features for the whole fleet are given as:

$$\begin{aligned} \mathcal{F}^{U} &= \left\{\mathcal{F}^{U}{}_1, \ \dots \ , \mathcal{F}^{U}{}_{n_a}\right\}, \\ \mathcal{F}^{C} &= \left\{\mathcal{F}^{C}{}_1, \ \dots \ , \mathcal{F}^{C}{}_{n_a}\right\}. \end{aligned} \tag{13.3}$$

The sets of static features of different assets are defined to be non-identical; i.e. $\left(\mathcal{F}^{C}{}_i \cap \mathcal{F}^{C}{}_j\right) \subset \mathcal{F}^{C}{}_i, \ i \neq j$. Since $\mathring{x}_i$ and $\mathring{x}_j$ are non-identical engineering systems, hence for any identical set of dynamic features $\mathcal{F}^{U}{}_{\cdot} = \mathcal{F}^{U}{}_i = \mathcal{F}^{U}{}_j$ applied to both systems, the damage estimates of both systems would be non-identical, i.e. $\mathring{y}_i \neq \mathring{y}_j$.

As mentioned earlier, there exists different classes of damage models; physics-based, data-driven, and hybrid models, denoted by x^{P}, x^{D}, and x^{H}, respectively. For any asset, there could exist a damage model of either of these classes. In some cases, some assets could have no model at all, either due to the inability to derive a mathematical formulation of the damage behaviour (for physics-based models), or unavailability of a data set of correlated operation parameters $\mathring{u}$ and the damage values $\mathring{y}$ (for data-driven models). In this case, an undefined damage estimation model x^{ϕ} is associated to the asset, with the aim to define it.

Let $\mathcal{X} = \{x_1, \dots, x_{n_a}\} = \mathcal{X}^{P} \cup \mathcal{X}^{D} \cup \mathcal{X}^{H} \cup \mathcal{X}^{\phi}$, where $\mathcal{X}^{P}$, $\mathcal{X}^{D}$, $\mathcal{X}^{H}$, and $\mathcal{X}^{\phi}$ are the disjoint sets of physics-based, data-driven, hybrid, and undefined damage models, respectively, Figure 13.2.

In this work, *moSAIc* is proposed as a hybrid damage model that combines an ensemble of existing physics-based models using a data-driven fusion method. *moSAIc* can be categorized as an MoE model

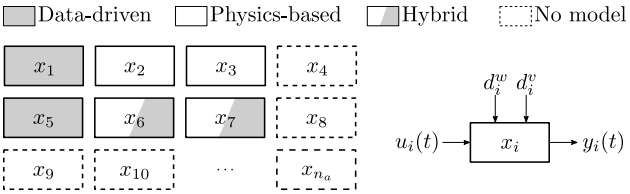


Figure 13.2.: A fleet of damage models

for the damage estimates for fleet assets with unknown damage models. For an MoE model, a set of *expert* models needs to be defined. Here, we refer to this set as the set of n_b *basis models* $\mathcal{X}^{\mathsf{B}}$. Details about the proper selection of basis models are detailed later in this chapter.

The vector $\mathbf{u}_i$ is defined to be the vector of inputs to the i^{th} model as

$$\mathbf{u}_i = \left(u_i^{(\mathrm{j})}\right)_{j=1}^{n_D}, \tag{13.4}$$

and $\mathbf{c}_i$ to be the vector of static features of the i^{th} asset as

$$\mathbf{c}_i = \left(c_i^{(\mathrm{j})}\right)_{j=1}^{n_C}. \tag{13.5}$$

For an arbitrary vector of operating conditions $\mathbf{u}$, each basis model $x^{\mathsf{B}}{}_i$ yields a corresponding damage estimate $y^{\mathsf{B}}{}_i$. The damage estimates are fused together by assigning a weight $\alpha^{\mathsf{B}}{}_i \in [0,1]$ to each. Hence, the weighted-sum of the damage estimates is given by:

$$\hat{y} = \sum_{i}^{n_b} \alpha^{\mathsf{B}}{}_i y^{\mathsf{B}}{}_i, \quad \sum_{i}^{n_b} \alpha^{\mathsf{B}}{}_i = 1. \tag{13.6}$$

In MoE models, the weights are decided by a gating network to account for the non-linear relationship between the inputs and outputs.

In *moSAIc*, the gating network is an MLP, parametrized by the set of parameters $\mathcal{P}_G$, and is defined as a mapping function such that

$$f^{\mathsf{G}}[\mathcal{P}_G] : (\mathbf{c}, \mathbf{u}) \mapsto \boldsymbol{\alpha}^{\mathsf{B}}, \tag{13.7}$$

where $\mathbf{u}$ is the vector of model inputs, $\mathbf{c}$ is the vector of static features, and $\boldsymbol{\alpha}^{\mathsf{B}}$ is the vector of weights assigned to the basis models. f^{G} is the function relating the model inputs and static features vectors to the weights vector. $\mathcal{P}_G$ is the set of MLP's hyper-parameters, e.g. the number of layers, number of perceptrons in each layer, perceptrons' activation functions, learning rate, the weight assigned to each perceptron's output, the bias in each layer, etc.

To this end, the only element in Equation (13.7) that remains undefined is the set of function parameters $\mathcal{P}_G$. Within *moSAIc*, the goal is to estimate a weights vector $\boldsymbol{\alpha}^{\mathsf{B}}$ for a given $(\mathbf{c}, \mathbf{u})$, which combines the basis models' damage estimates to approximate the target damage of the unknown-model asset as precisely as possible. Herein, the objective function to be optimized is the difference between *moSAIc*'s damage estimate and the target damage.

The optimization problem is defined as follows:
Let y^{O} be the target damage and its corresponding estimate $\hat{y}$. Following the ML jargon, the difference between the target damage and the estimated damage values is referred to as the loss function, and is denoted by $\mathcal{L}^{\mathsf{M}}$. Denoting $\mathcal{P}_G^*$ to be the set of optimal MLP parameters, that minimize $\mathcal{L}^{\mathsf{M}}$, $\mathcal{P}_G^*$ can be obtained by solving the optimization problem

$$\mathcal{P}_G^* = \underset{\mathcal{P}_G}{\arg\min}\ \mathcal{L}^{\mathsf{M}}(\mathcal{P}_G)\,. \tag{13.8}$$

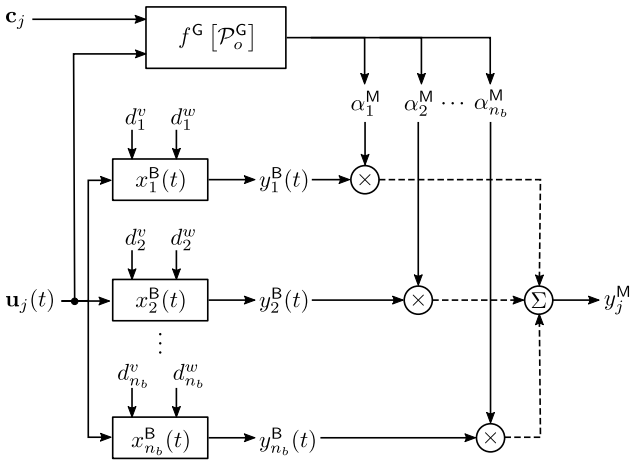


Figure 13.3.: Architecture of *moSAIc*

Defining $\boldsymbol{\alpha}^{\mathsf{M}} = f^{\mathsf{G}}\left[\mathcal{P}_G^*\right](\mathbf{c}, \mathbf{u})$, obtained by substituting $\mathcal{P}_G^*$ in Equation (13.7), *moSAIc*'s damage estimate can be re-written by substituting $\boldsymbol{\alpha}^{\mathsf{M}}$ into Equation (13.6)

$$y^{\mathsf{M}} = \sum_{i}^{n_b} \alpha^{\mathsf{M}}{}_i y^{\mathsf{B}}{}_i =< \boldsymbol{\alpha}^{\mathsf{M}}, \mathbf{y}^{\mathsf{B}} > . \tag{13.9}$$

The architecture of *moSAIc* is shown in Figure 13.3. For complex problems, there is no closed form solution to the optimization problem in Equation (13.8). Alternatively, machine learning is applied to learn a surrogate relationship between $(\mathbf{c}, \mathbf{u})$ and $\boldsymbol{\alpha}^{\mathsf{M}}$ through a gradient-descent optimization procedure.

The elements of $\mathcal{P}_G$ are classified into two groups: 1) the matrices of weights and biases associated to the individual perceptrons of the MLP, $\mathbf{W}$ and $\mathbf{b}$, respectively, and 2) the set of other MLP hyperparameters, such as the number of layers, the size of each layer, the learning rate, the activation function of each perceptron etc. In practice, only $\mathbf{W}$ and $\mathbf{b}$ are solved for by applying gradient-descent

algorithms, while the rest of the parameters could be chosen using hyper-parameter tuning methods, e.g. random search or grid search. Further details about the theoretical principles of MLPs can be found in [Kro16].

In a supervised learning setting, given a training set of $n_{\mathcal{T}}$ examples, defined as

$$\mathcal{T} = \left\{ \left\{ \left(c^{\{\mathsf{q}\}}, u^{\{\mathsf{q}\}} \right), y^{\{\mathsf{q}\}} \right\} \mid q \in \{1, \ldots, n_{\mathcal{T}}\} \right\}, \tag{13.10}$$

the gradient descent procedure attempts to find the appropriate weights and biases of the MLP's perceptrons. To avoid over-fitting to the training set, elastic-net regularization is employed to penalize large weights. The loss function $\mathcal{L}^{\mathsf{M}}(\mathcal{P}_G)$ is, herein, re-written as

$$\mathcal{L}^{\mathsf{M}} = \frac{1}{n_{\mathcal{T}}} \sum_{q=1}^{n_{\mathcal{T}}} \left| 1 - \frac{y^{\mathsf{M}\{q\}}}{y^{\mathsf{O}\{q\}}} \right| + \lambda_1 \|\mathbf{W}\|_1 + \lambda_2 \|\mathbf{W}\|_F, \tag{13.11}$$

where $\lambda_1, \lambda_2 \in \mathcal{P}_G$ are the lasso and ridge regularization parameters, respectively. $\|\mathbf{W}\|_1$ is the l_1-norm of the weights matrix and $\|\mathbf{W}\|_F$ is the Frobenius norm of the weights matrix.

At the initialization of the training process, the weights and biases are randomly initialized, and an estimate of damage is computed, from which the loss is evaluated using Equation (13.11). This estimate is used to feed-in the back-propagation process to iteratively update the individual weights of the MLP by

$$w_{ab}^{[l]}|_{m+1} = w_{ab}^{[l]}|_m - \ \Delta w_{ab}^{[l]}|_m, \tag{13.12}$$

where $w_{ab}^{[l]}|_m$ is the contribution of the b^{th} perceptron's output to the value of the a^{th} perceptron in the l^{th} layer at the m^{th} epoch, $\Delta w_{ab}^{[l]}|_m$ is the update increment of the said weight.

13.2.1 The general case of *moSAIc*

In the previous paragraphs, the simplest form of the proposed damage estimation model has been introduced. The underlying assumptions for the proposed architecture and formulations are:

1. For every asset $\mathring{x}_i$, there exists one and only one degradation model x_i, i.e. the degradation of the asset is attributed to a single degradation mode, e.g. mechanical degradation due to Low-Cycle Fatigue (LCF).
2. Every model x_i outputs only a single damage estimate y_i, i.e. reducing the damage of the system to a scalar metric.

In reality, engineering systems are rather complex, and their degradation is typically caused by a multitude of failure mechanisms, hence dictating the need for as-many degradation models describing each. The significance of each failure mechanism's contribution to the overall degradation of the asset is assigned using a decision mechanism, which is not the focus of this discussion at the moment. Additionally, each degradation model doesn't necessarily reduce to a scalar damage, but typically to a vector of damage estimates for the same asset undergoing the same inputs $\mathbf{u}_i$.
For example, let x_i be a LCF model for a given roller bearing, possible model outputs may include the Remaining Useful Lifes (RULs) of each roller, the inner and outer races, and the cage. This general architecture including models having multiple as well as single outputs is shown in Figure 13.4.

Recalling the definition of the model parametrized by the set of static features introduced earlier, $x_{id}\left[\mathbf{c}_i\right]$ denotes the model describing the d^{th} degradation phenomena of the i^{th} asset parametrized by the static features $\mathbf{c}_i$. Herein, it should be noted that not all assets should necessarily have a thoroughly complete set of degradation models. For instance, for a fleet of roller bearings, a subset might have both structural LCF and degradation models attributed to grease

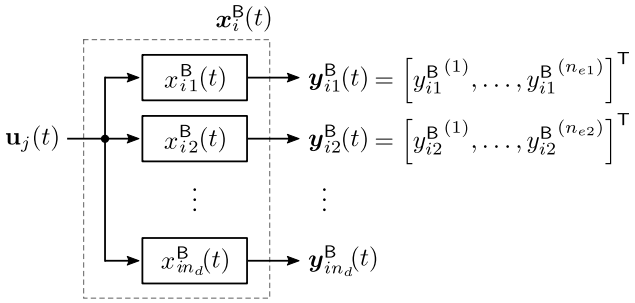


Figure 13.4.: General configuration of a damage model set-up and output

contamination, while the rest have only corresponding contamination-driven degradation models.

In this work, for simplicity, it is assumed that the different damage models are independent. Hence, deriving the damage model x^{ϕ} attributed to the d^{th} degradation phenomena requires only the existence of basis models attributed to the same phenomena. Herein, the general form of the definition of x^{B} can be re-written as

$$\mathcal{X}^{\mathsf{B}} = \left\{ x^{\mathsf{P}}{}_{id} \in \mathcal{X}^{\mathsf{P}}{}_{d} \mid i \in \{1, \ldots, n_b\} \right\}, \tag{13.13}$$

where $\mathcal{X}^{\mathsf{P}}{}_{d}$ is the subset of physics-driven models having damage models corresponding to the d^{th} degradation phenomena. The vectors of damage estimates of the i^{th} basis model and its associated vector of *moSAIc* weights are given by:

$$\begin{aligned} \mathbf{y}^{\mathsf{B}}{}_{i} &= x^{\mathsf{B}}{}_{i}\left[\mathbf{c}_i\right](\mathbf{u}) && \in \mathbb{R}^{n_{e,i}}, \\ \boldsymbol{\alpha}^{\mathsf{M}}{}_{i} &= \left[f^{\mathsf{G}}\left[\mathcal{P}_G^*\right](\mathbf{c}, \mathbf{u})\right]_i && \in \mathbb{R}^{n_{e,i}}, \end{aligned} \tag{13.14}$$

where $n_{e,i}$ is the number of damage estimates from the i^{th} basis model. For generality, the basis models don't yield an equal number of estimate; i.e. $n_{e,i} \neq n_{e,j}$, and don't yield correspondent damage estimates; e.g. taking two basis models of roller bearings, the first yields

damage for the rollers only, while the second yield damage estimates for the rollers and the races. Hence, $\{1, \ldots, n_{e,i}\} \neq \{1, \ldots, n_{ej}\}$. Following this assumption, defining *moSAIc*'s damage estimate as $\mathbf{y}^{\mathsf{M}} = \sum_{i}^{n_b} < \boldsymbol{\alpha}^{\mathsf{M}}{}_i, \mathbf{y}^{\mathsf{B}}{}_i >$ results in mathematical inconsistency in dimensions and indexing.

To resolve this inconsistency, the dimensions of the individual basis damage and weights vectors should be adjusted, as well as the indexing of their constituent damage estimates, such that all corresponding constituents are positioned at the same index, thus allowing vectorized operations to be applied. This procedure is referred to as the assembly of consistent vectors [1], and its assembled outcomes are scribed as $\tilde{(\cdot)}$.

To define the assembled vectors, we first define the size $\tilde{n_e} = \left| \bigcup_{i=1}^{n_b} \{1, \ldots, n_{e,i}\} \right|$, calculated by evaluating the cardinality of the union of sets of indexes of all basis models. Additionally, the function $f^{\tilde{\mathsf{a}}} : \mathbb{R}^{n_{e,i}} \rightarrow \mathbb{R}^{\tilde{n_e}}$ is defined, which maps the indexes of the individual basis output / weights vectors, e, to their corresponding index in the consistently assembled vectors, $\tilde{\mathsf{e}}$. Therefore, the $\tilde{e}^{th}$ element of the assembled vector of basis damage estimates and the vector of their associated weights are defined as:

$$(\tilde{y}^{\mathsf{B}}{}_i)^{\tilde{\mathsf{e}}} = \begin{cases} (y^{\mathsf{B}}{}_i)^{\mathsf{e}}, & \text{if } (y^{\mathsf{B}}{}_i)^{\mathsf{e}} \in \mathbf{y}^{\mathsf{B}}{}_i \\ 0 & \text{else} \end{cases}, \tag{13.15}$$

$$(\tilde{\alpha}^{\mathsf{M}}{}_i)^{\tilde{\mathsf{e}}} = \begin{cases} (\alpha^{\mathsf{M}}{}_i)^{\mathsf{e}}, & \text{if } (\alpha^{\mathsf{M}}{}_i)^{\mathsf{e}} \in \boldsymbol{\alpha}^{\mathsf{M}}{}_i \\ 0 & \text{else} \end{cases} \tag{13.16}$$

[1] Naming adopted from the Finite Elements Method (FEM) assembly procedure of matrices and vectors.

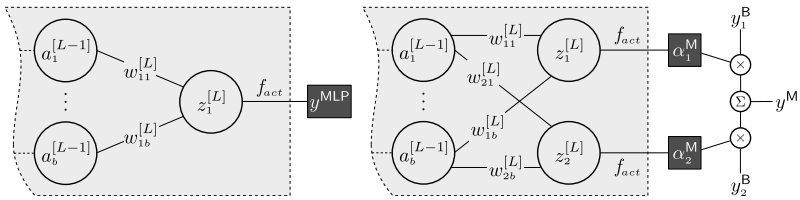
Herein, the vector of *moSAIc*'s damage estimates for any arbitrary asset can be written as:

$$\tilde{\mathbf{y}}^{\mathsf{M}} = \sum_{i}^{n_b} < \tilde{\boldsymbol{\alpha}}^{\mathsf{M}}{}_i, \tilde{\mathbf{y}}^{\mathsf{B}}{}_i > \quad \in \mathbb{R}^{\tilde{n_e}}. \tag{13.17}$$

In Appendix C, the assembly procedure is elaborated on a minimal example.

13.2.2 Back-propagation in *moSAIc*

In *moSAIc*, the calculation of the update increment of the weights of the hidden layers follows the standard calculation procedure of back-propagation in MLPs, see e.g. [Kro16] for details. Only the update of the MLP's output layer weights is influenced by the architecture of *moSAIc*, and differs from the standard MLP's output layer's back-propagation procedure. This is explained in the following paragraphs.



(a) Standard MLP **(b)** *moSAIc*

Figure 13.5.: Output terminal of *moSAIc*'s MLP vs. a standard MLP. The MLP is enclosed by the light grey dotted polygon, and its outputs(s) are contained in the dark grey boxes

Figure 13.5 shows the terminal layers in a standard MLP deployed as a data-driven damage estimation model, and the MLP deployed in *moSAIc* (n_b= 2) for estimating the weights associated with the basis models. The architectures are shown for models estimating a single damage, and could be generalized by applying the relevant

linear algebra operations to the tensors introduced in the previous paragraphs.

Three main differences can be remarked between the two architectures:

1. The output of *moSAIc*'s MLP is weights vector $\boldsymbol{\alpha}^{\mathsf{M}}$ which is afterwards used to predict the damage, while the output of the standard MLP is the damage prediction y directly.
2. The output of *moSAIc*'s MLP is a vector of size n_b (or a tensor of size $n_b \times \tilde{n_e}$ for the general case), while the output of the standard MLP is a scalar; i.e. size 1 (or a vector of size $\tilde{n_e}$ for the general case).
3. The loss function of *moSAIc* is not calculated directly from the MLP's outputs, but rather on the proceeding value y^{M}, cf. Equation (13.11). Contrarily, in a standard MLP, the model's loss function is evaluated directly on the MLP's output.

Let the superscript $[L]$ denote quantities associated to the last layer of the MLP, $z_b^{[L-1]}$ denotes the output of the b^{th} perceptron in the second-to-last layer, $w_{ab}^{[L]}$ be its associated weight, and let $z_a^{[L]} = \sum_b w_{1b}^{[L]} a_b^{[L-1]}$ be the a^{th} perceptrons' input to the last layer, cf. Figure 13.5.

For a classical MLP, the non-regularized loss function given by the Mean Absolute Relative Error (MARE) between the target damage and the MLP's damage is expressed as follows:

$$\mathcal{L}^{\mathsf{G,MLP}} = \left| 1 - \frac{y^{\mathsf{MLP}}}{y^{\mathsf{O}}} \right|, \tag{13.18}$$

where $y^{\mathsf{MLP}} = f^{\mathsf{act}}\left(z_1^{[L]}\right)$, f^{act} is the activation function, and y^{O} is the target model output.

Herein, the update increment of the weights of the last layer, $\Delta w_{ab}^{[L]}$, can be obtained by applying the chain rule, as follows:

$$\Delta w_{ab}^{[L]} = \eta \frac{\partial \mathcal{L}^{\mathsf{G,MLP}}}{\partial w_{ab}^{[L]}} = \eta \underbrace{\left(\frac{-1}{y^{\mathsf{O}}}\right)}_{\frac{\partial \mathcal{L}^{\mathsf{G,MLP}}}{\partial y^{\mathsf{MLP}}}} \underbrace{\delta f^{\mathsf{act}}\left(z_a^{[L]}\right)}_{\frac{\partial y^{\mathsf{MLP}}}{\partial z_1^{[L]}}} \underbrace{a_b^{[L-1]}}_{\frac{\partial z_1^{[L]}}{\partial w_{ab}^{[L]}}}, \tag{13.19}$$

where δf^{act} is the derivative of the activation function, whose closed-form can be looked up in machine learning literature for standard activation functions. $\eta \in \mathcal{P}_G$ is the learning rate, a hyper-parameter of the MLP, which controls the gradient descent's speed.

On the other hand, in *moSAIc*, the output of the MLP is the weights vector, $\boldsymbol{\alpha}^{\mathsf{M}}$, and not the damage. The loss function of the MLP[2] is given by the mean square error between the *target* weights vector $\alpha_i^{\mathsf{M,O}}$ and the MLP's *output* weights vector $\alpha_i^{\mathsf{M,G}}$. The MLP's loss is herein expressed as:

$$\mathcal{L}^{\mathsf{G,M}} = \frac{1}{2} \sum_{i}^{n_b} \left(\alpha_i^{\mathsf{M,O}} - \alpha_i^{\mathsf{M,G}}\right)^2, \tag{13.20}$$

where $\alpha_i^{\mathsf{M,G}} = f^{\mathsf{act}}\left(z_i^{[L]}\right)$.

In the case of *moSAIc*, although $\boldsymbol{\alpha}^{\mathsf{M,O}}$ is the target weights vector, it is not provided as a label for training set examples. Instead, it has to be inferred from the given target damage y^{O}. The target damage

[2]Not to be confused with the loss function of the whole model in Equation (13.11).

can be expressed in terms of the target weights and the basis models damage estimates as

$$y^{\mathsf{O}} = \sum_{i=1}^{n_b} \alpha_i^{\mathsf{M,O}} y^{\mathsf{B}}{}_i = \alpha_a^{\mathsf{M,O}} y^{\mathsf{B}}{}_a + \sum_{\substack{i=1 \\ i \neq a}}^{n_b} \alpha_i^{\mathsf{M,O}} y^{\mathsf{B}}{}_i, \tag{13.21}$$

where the last term represents the target contributions of the basis damage estimates, excluding the a^{th} estimate. As mentioned earlier, the target weights are not provided as training labels, thus Equation (13.21) represents an under-determined system. This problem could be overcome by approximating the summation in the last term using the corresponding MLP's output weights, by introducing

$$y_{\neq a}^{\mathsf{O}} = \sum_{\substack{i=1 \\ i \neq a}}^{n_b} \alpha_i^{\mathsf{M,O}} y^{\mathsf{B}}{}_i \approx \sum_{\substack{i=1 \\ i \neq a}}^{n_b} \alpha_i^{\mathsf{M,G}} y^{\mathsf{B}}{}_i \tag{13.22}$$

Substituting Equation (13.22) in Equation (13.21) and arranging the terms, the $\alpha_a^{\mathsf{M,O}}$ could be expressed as

$$\alpha_a^{\mathsf{M,O}} = \frac{y^{\mathsf{O}} - y_{\neq a}^{\mathsf{O}}}{y^{\mathsf{B}}{}_a}. \tag{13.23}$$

Substituting Equation (13.23) into Equation (13.20), the update increment of the weights of the last layer, $\Delta w_{ab}^{[L]}$, can be obtained by applying the chain rule, as follows:

$$\begin{aligned} \Delta w_{ab}^{[L]} &= \eta \frac{\partial \mathcal{L}^{\mathsf{G,M}}}{\partial w_{ab}^{[L]}} \\ &= \eta \underbrace{\left(\frac{y^{\mathsf{O}} - y_{\neq a}^{\mathsf{O}}}{y^{\mathsf{B}}{}_a} - \alpha_a^{\mathsf{M,G}} \right)}_{\frac{\partial \mathcal{L}^{\mathsf{G,M}}}{\partial \alpha_a^{\mathsf{M,G}}}} \underbrace{\delta f^{\,\mathsf{act}}\left(z_a^{[L]} \right)}_{\frac{\partial \alpha_a^{\mathsf{M,G}}}{\partial z_1^{[L]}}} \underbrace{a_b^{[L-1]}}_{\frac{\partial z_1^{[L]}}{\partial w_{ab}^{[L]}}}, \end{aligned} \tag{13.24}$$

13.3 Hybrid Model Preparation

In this section, four necessary model preparation steps are introduced, and their application in the context of *moSAIc* is elaborated; 1) classification of damage severity prior to interpolating among basis models, 2) dynamic features extraction from sensor signals, 3) the selection of static and dynamic features, $\mathbf{c}$ and $\mathbf{u}$, respectively, and 4) choice of the set of basis models $\mathcal{X}^{\mathsf{B}}$. Conventional data pre-processing practices, e.g. hyper-parameter tuning, input standardization, etc. are not addressed in this text.

13.3.1 Damage pre-classification

Most degradation models in literature describe accumulated damage D_a in Equation (7.7) as a value between 0 and 1. Hence, under a given operation scenario, the damage accumulation severity could be categorized into: negligibly damaged when $D_a \approx 0$, fully damaged $D_a \approx 1$, or partially damaged $0 < D_a < 1$.

The dominance of the extreme values, i.e. 0 and 1, in a dataset could pose a challenge to the model trying to fit one regression model to samples from three different distributions. To overcome this and ensure a model fits a regression kernel capable of distinguishing the three behaviours, larger training sets, as well as computation effort for training is needed. Typically, obtaining a training set generous with samples having full or negligible damage is a challenge.

Alternatively, this problem could be broken down into a preliminary classification step, whereby, given a set of inputs $(\mathbf{c}, \mathbf{u})$, a classifier algorithm, e.g. Random Forests (RFs) [Bre01], predicts the damage severity. Having categorized the damage severity class, only the cases with partial damage are passed to the hybrid model to quantify their damage severity *numerically*.

In this work, the classifier is trained on the same training set as the hybrid model. Since the decision of the classifier is passed

downstream to the hybrid model, it is essential to make sure that the predicted severity matches the actual, and not misclassified as another severity. Hence, the classifier parameters are to be tuned such that it favours maximizing precision over recall, while not neglecting the latter.

Nonetheless, not even a perfect classifier can be 100 % precise, and misclassified information can inevitably pass to the hybrid model. To address this issue, during the training phase, not only the samples labelled as partially damage are let through to the hybrid model, but a portion of the samples labelled as fully or negligibly damaged. Under this setting, the regression step performed by the hybrid model gets trained on extreme cases by incorporating the entire spectrum of accumulated damage values. Given that, the probability of a misclassified case getting corrected by the regression step is higher, hence yielding a more precise result.

13.3.2 Dynamic features extraction

As mentioned earlier, dynamic features are time-variant features, provided as input to the degradation model. In many data-driven applications, time-series data is handled using Recurrent Neural Networks (RNNs) or its more advanced variants, e.g. LSTMs. Such models come with additional computational overhead. On the other hand, another approach to handle time-series data is through extracting descriptive statistical quantities, distinguishably representing the signal characteristics.

In the context of this work, the focus is directed to estimating damage for structures undergoing oscillatory loading, where cyclic signals are commonly recorded by sensors. Hence, the following discussion focuses on the extraction of features relevant to them. Other applications having different sensor data signatures might require different treatment to extract representative features.

For cyclic signals, [GMZ16] suggests extracting features from the time domain (temporal features), frequency domain (spectral features), as well as time-frequency-related features, especially for non-stationary signals.

Temporal features represent the statistical properties of a signal. Some common temporal features are

1. Mean: filters the signal, or windows of it, to an average value
2. Median: filters the signal to an average value, while being less sensitive to outliers
3. Root Mean Square (RMS): represents the energy content of the signal
4. Peak: measures the strength of the vibrations, and reflects the extent of damage
5. Skewness: indicates the presence of rough surfaces between contacting components (e.g. rolling elements in bearings)
6. Kurtosis: characterizes the impulsive and noisy behaviour of the signal

Spectral features are extracted by performing a Fourier transformation on the signal, from which fundamental frequencies of the signal and their corresponding signal power content can be retrieved. In many rotating machinery applications, for instance, the fundamental frequencies depend on the frequency of the system loads, and its multiples. In other applications, e.g. bearings, the geometry of the components contribute to the fundamental frequencies. Due to the periodic nature of operation, damage-related features are associated to such fundamental frequencies. Therefore, their corresponding power spectral density functions are used as spectral features.

Finally, temporal-spectral features are features connecting both time and frequency domains. These are particularly relevant for non-

stationary signals, where signal trends and mean variations over time are as relevant to the damage. Temporal-spectral features can be obtained from short-term Fourier transformation, which performs a sliding-window Fourier transformation on the signal, yielding a time-variant power spectral density functions. Further elaboration on the temporal-spectral features is not relevant to this work, and interested readers are directed to [GMZ16]

13.3.3 Feature selection

Feature selection is a very common preparatory step for statistical models. Feature selection doesn't only bring an advantage to the models' performance, but also the failure to select the relevant features from a larger set could impair the accuracy and generalization capability of the model.

In this work, features are classified into static and dynamic features. Among either sets, some features are more correlated to damage than the others, in addition to some features being dependent on others features. Accordingly, the feature selection process should exclude features dependent on others, and extract those that exhibit the highest contribution to the damage estimation precision.

To asses the contribution of features to the damage estimate, an Extremely Randomized Trees Regressor (ERTR) is employed. ERTR is an evolutionary development of the Random Forest (RF), known for its accuracy, robustness and computational efficiency [GEW06]. An importance threshold is set on the feature importances, below which features are deemed insignificant to the damage estimate.

The correlation among the important features selected by the ERTR is assessed to rule out the correlated features. Since the features investigated are quantitative, Pearson's and Spearman's correlation coefficients pose as alternatives. In the course of this work, Spearman's correlation coefficient is preferred since it can measure the correlation between ranked values, thus highlighting the monotonic-

ity between the variables, rather than the linearity. Additionally, Spearman's correlation is less affected by outliers in the data, making it more attractive, especially for industrial use-cases, where outliers are common [CSP02].

The Spearman's correlation coefficient scales between -1 and 1, with 1 indicating negative and positive monotonic correlations, respectively, while 0 is a sign of no correlation at all. To decide on dependent features, a correlation threshold is set on the absolute values of the coefficient. Features with correlations coefficients higher than the threshold are excluded from the set, leaving behind only the feature with the highest correlation to the damage variable.

The outcome of both the feature importance step and the correlation identification steps are governed by the importance and correlation thresholds. Both thresholds are considered hyper-parameters of the hybrid model, which ought to be identified either by search methods, or by recommendations from experts based on their know-how about the assets degradation.

13.3.4 Basis models choice

Ideally, the selected basis models should be chosen such that they span the entire range of static features describing the physics-based models. As well, the basis should span the entire space of dynamic features, describing the operation scenarios. In this work, it is assumed that all physics-based models can estimate the damage equally-precise for the entire domain of dynamic features. Therefore, the basis selection procedure should yield basis models that scatter uniformly within the space of the static features, and are independent from one another; i.e. the information added by one basis model should not be already represented by another.

A first step for identifying the basis models is to select the damage-relevant features, so as not to overwhelm the hybrid model with redundant basis models. As explained in the previous paragraphs, the

feature importance is obtained from an ERTR model. Depending on the chosen importance threshold, unimportant features are dropped from $\mathbf{c}$, thus reducing its size to n_{C+}. For the upcoming discussion, $\mathbf{c}$ is taken to hold only the important features.

A first question to care for is the number of basis models required to adequately represent the degradation behaviour of assets in the fleet. A linear n-dimensional feature space can be spanned and represented by $n+1$ points. Hence, $n_b = 1 + n_{C+}$ models are assumed sufficient to represent the fleet.

The choice of the basis models is performed sequentially. As a prerequisite, the function f^{dist} which evaluates the distance between every i^{th} static features vector and any other l^{th} vector in $\mathcal{F}^{\mathsf{C}}$ is defined as

$$f^{\mathsf{dist}}(\mathbf{c}_i, \mathbf{c}_l) = \|\mathbf{c}_i - \mathbf{c}_l\|_2 . \tag{13.25}$$

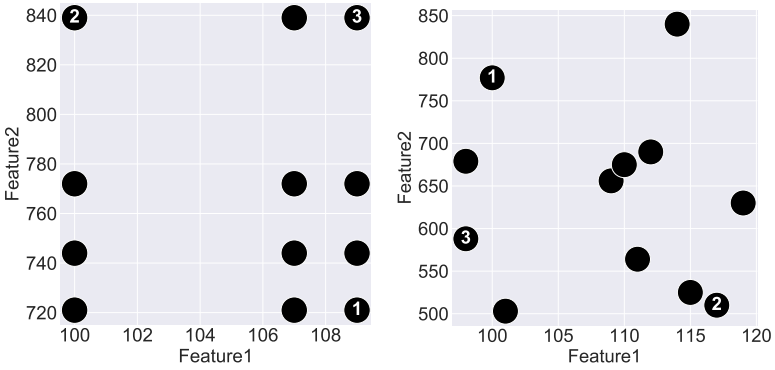
It is important to note that all the vectors in $\mathcal{F}^{\mathsf{C}}$ used through the discussion should be normalized.

The set of basis models $\mathcal{X}^{\mathsf{B}} \subseteq \mathcal{X}^{\mathsf{P}}$ is hereafter defined as

$$\mathcal{X}^{\mathsf{B}} = \left\{ x^{\mathsf{B}}{}_k \right\}_{k=1}^{n_b} = \left\{ x^{\mathsf{P}}{}_b \mid b \in \left\{ 1, \ldots, \left| \mathcal{X}^{\mathsf{P}} \right| \right\} \right\}, \tag{13.26}$$

where

$$b = \arg\min_i \begin{cases} \left(\left\| \left(f^{\mathsf{dist}}(\mathbf{c}_i, \mathbf{c}_l) \right)_{l=1}^{n_a} \right\|_2 \right)_{i=1}^{n_a} & \text{if } k = 1, \\ \left(\dfrac{\left\| \left(f^{\mathsf{dist}}(\mathbf{c}_i, \mathbf{c}_l) \right)_{l=1}^{n_a} \right\|_2}{\left\| \left(f^{\mathsf{dist}}(\mathbf{c}_i, \mathbf{c}_l) \right)_{l=1}^{k-1} \right\|_2} \right)_{i=1}^{n_a} & \text{if } k > 1, \end{cases} \tag{13.27}$$



(a) Structured (b) Unstructured

Figure 13.6.: Example for basis selection in different feature spaces

Figure 13.6 shows a minimal example demonstrating the basis selection process, for a fleet with two features. The models' positions in the features are indicated by the black circles. The selected models and their order of selection are indicated by the number inside.

In a structured feature space, the basis models land automatically to the corner models, where they can fully span the entire feature space. On the other hand, for the unstructured space, the basis models get pushed towards the models lying in the *outskirts* of the scatter. In that case, the fact that the algorithm considers the already chosen basis while choosing the consequent ones can be observed. Basis 2 is chosen as far as possible from basis 1, whereas basis 3 maintains a distance from the two previously selected basis.

13.3.5 Preprocessing pipeline

To summarize, Figure 13.7 illustrates the model preparation pipeline. The following sequence is undergone:

1. Extract the temporal, spectral, and tempo-spectral dynamic features from the raw sensor signals. Simultaneously, the relevant static features are inferred from the assets' catalogues
2. Consistently assign the damage labels to the corresponding extracted features
3. Split the features and labels into train and test sets
4. Assess the feature importance on the training set and exclude less-important and correlated features
5. Perform the basis selection procedure, and extract a subset of basis models from the training set. Hence, reduce the training set to the unselected models only
6. Train a classifier to categorize the damage severity
7. Perform cross-validation on K-folds of the training set to tune the model hyper-parameters
8. Train the model on the entire training features and labels, and the basis sets labels
9. Evaluate the trained model's performance on the test set

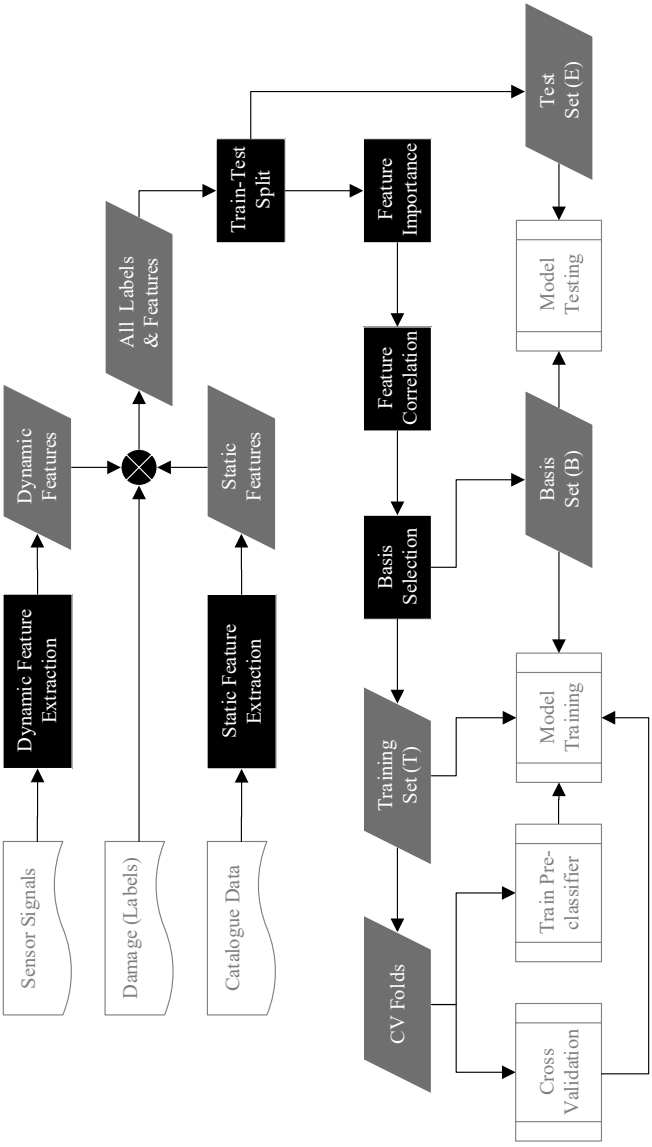


Figure 13.7.: Model preparation pipeline

Numerical Investigation 14

In this section, numerical investigations are conducted on the proposed hybrid approach for damage estimation. The role the physics-based simulation models play in enhancing the accuracy and precision of the estimate is highlighted. Furthermore, the edge of this hybrid model over traditional data-driven models in terms of robustness and intuitiveness of the results is assessed.
Throughout this study, the models' sensitivity to the following aspects is investigated:

- Impact of pre-classifying the damage severity
- Size and selection method of the set of basis models $\mathcal{X}^{\mathsf{B}}$
- Importance and correlation thresholds of static and dynamic features extraction
- Performance compared to simulation-free models

To allow for an exhaustive investigation of Simulation Models and Artificial Intelligence Combined (*moSAIc*)'s sensitivity to its constituents, inputs, and hyper-parameters, the following problem is introduced.

14.1 Problem Description

The problem shown in Figure 14.1 consists of a cantilever structure of total length l, and is discretized using 10 equisized Euler-Bernoulli beam finite elements, each of length dl. All elements have the same isotropic material properties: Young's modulus E, density ρ, and Poisson's ratio ν, and their material behaviour is governed by a linear-elastic material law. Two solid alloys are used in this study; Stainless

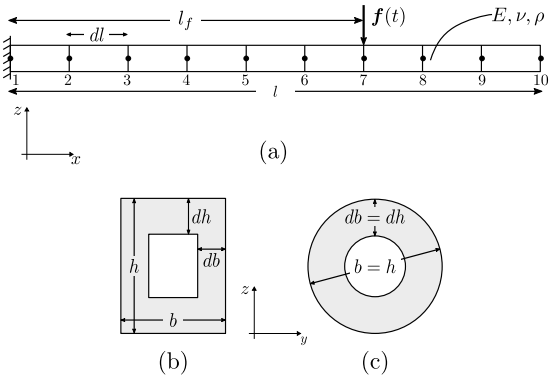


Figure 14.1.: Beam benchmark example. (a) Beam elements discretization and boundary conditions, (b) Box cross-section, and (c) Tube cross-section.

steel alloy SS316, and Aluminium alloy Al5086. Their corresponding nominal material properties, denoted by $(\cdot)_o$ are shown in Table 14.1. Additionally, the fatigue strength S_f and fatigue coefficient b, which will be required to evaluate damage are shown in the same table.

Table 14.1.: Nominal material properties of the beam structure

Material	E_o (GPa)	ρ_o (kg/m^3)	ν_o	S_f (MPa)	b
SS316	210	7809	0.30	1762	-0.209
Al5086	70	2710	0.27	491	-0.081

Fleet assets

The beams have a constant cross-section profile throughout their lengths, shown in Figure 14.1 (b) and (c). Two profiles are considered in this study; 1) box profile, defined by the breadth b, height h, horizontal thickness db, and vertical thickness dh, 2) a tube profile, defined by the radius $b = h$, and the radial thickness $db = dh$.

Herein, the different cross-section profiles, cross-section dimensions, and material properties can be combined to create the set of static features of fleet assets. The respective domains of each static feature are summarized in Table 14.2

Table 14.2.: Domains of asset variant parameters

Parameter	**Domain**
Profile, P	$\in \{\text{Tube}, \text{Box}\}$
Material, M	$\in \{\text{SS316}, \text{Al5086}\}$
b (m)	$\in \{0.05, 0.01\}$
h (m)	$\in \{0.05, 0.01\}$
$^{db}/_{b}$	$\in \{0.2, 0.5\}$
$^{dh}/_{h}$	$\in \{0.2, 0.5\}$

The reader is reminded at this point that the values in Table 14.2 are not essentially static features of the model. Rather, these values are used to create a fleet of beam assets, while the definition of the set of static feature follows in the subsequent paragraphs. Examining the possible combinations of parameters, it is found that 32 are possible for *Box* profile, only 8 are possible for the *Tube* profile, since $b = h$ and $^{db}/_{b} = {}^{dh}/_{h}$. This yields a fleet of 40 unique assets.

For the definition of the fleet assets, variances in material properties are accounted for, i.e. not all beams from the same material have the same material properties. Instead, E, ν, and ρ are respectively sampled from the following Gaussian distributions,

$$E \sim \mathcal{N}\left(E_o, \sigma_E^2\right) \text{GPa},\ \nu \sim \mathcal{N}\left(\nu_o, \sigma_\nu^2\right),\ \text{and}\ \rho \sim \mathcal{N}\left(\rho_o, \sigma_\rho^2\right) \text{kg m}^{-3},$$

where $\sigma_E = 5\,\text{GPa}$, $\sigma_\nu = 0.01$, and $\sigma_\rho = 15\,\text{kg m}^{-3}$ denote the standard deviations assumed in this study for the three material properties. The sampled material properties are used solely in the simulation of the different dynamic inputs, as will be shown later, and will not be exposed as inputs to the damage model.

Boundary conditions

The beam is excited by the sinusoidal force vector

$$\mathbf{f}(t) = [f_x(t),\ f_y(t),\ f_z(t)]^\mathsf{T},$$

where $f_x(t) = 0$, and the y- and z-components are expressed by:

$$f_{y/z}(t) = f_{o,\ y/z}\ sin(\omega_{f,y/z}t) + d^f_{y/z}, \tag{14.1}$$

where $d^f \sim \mathcal{N}\left(0, 6.25 \times 10^{-2}\right)$ N is a Gaussian noise super-positioned on the sinusoidal force signal, f_o is the amplitude of the force signal, and ω_f is its frequency. The uncertainty in the loading conditions is demonstrated by sampling f_o and ω_f as well from the Gaussian distributions, where $\omega_f \sim \mathcal{N}\left(10, 2\right)$ Hz. Limiting this study to High-Cycle Fatigue (HCF) damage, it is necessary to remain in the linear elastic domain of the material. Hence, the means and the standard deviations of the force amplitude are varied with the material choice. Accordingly, $f_o \sim \mathcal{N}\left(6.5 \times 10^2, 25\right)$ N for Al5086 beams, and $f_o \sim \mathcal{N}\left(10^4, 2.5 \times 10^3\right)$ N for SS316.

The force vector is applied at an arbitrary node at distance l_f from the support, as shown in Figure 14.1(a). The distance l_f is given by $l_f = dl \cdot \mathcal{U}(3, 10)$, where $dl = 0.1\,\text{m}$ is the element's length, and $\mathcal{U}(3, 10)$ denotes sampling a node from a discrete uniform distribution ranging from 3 to 10; i.e. the load is neither allowed on node 1 (support node) nor node 2 (the node next to the support), cf. Figure 14.1. For each asset, a total of 800 scenarios of unique combinations of load amplitudes, frequencies, and positions are generated.

The responses of the structure to the load scenarios are simulated by applying them as boundary conditions on a finite element model for each fleet variant. A linear transient solver with a $10^{-3}\,\text{s}$ time step is used. The critical time step is inferred from the maximum eigenfrequency of the system to be $10^{-2}\,\text{s}$.

To capture the cantilever's response, three sensors are placed at node 2; a uni-axial stain gauge oriented in the x-axis, and two uni-axial accelerometers oriented in the y- and z-axes. The position of the sensors is intentionally chosen neither to be at node 1 (where damage is evaluated), nor at nodes 3 to 10 (force application domain). Also, it is intentional to fix the sensor position for all assets and loading scenarios. Therefore, node 2 is chosen.

Static features

Static features can be inferred from Table 14.2. Bearing in mind that the loading scheme induces damage caused by the normal stresses at the support, the physics of failure can be exploited to derive additional features. The normal stress σ_{xx} at the support of a cantilever beam is inversely proportional to the second moment of area $I_{y/z}$, depending on whether the transverse load is applied in the y/z directions, respectively. Therefore, the second moments of area $I_y = \frac{bh^3}{12}$, and $I_z = \frac{hb^3}{12}$ are added to the set of static features.

With respect to the material parameters, besides the nominal values of the Young's modulus E_0, Poisson's ratio ν_0, and density ρ_0, which describe the dynamics of the structure, the yield and ultimate tensile strengths σ_Y and σ_U, reflecting its mechanical strength, as well as the fatigue strength σ_f and fatigue exponent b, dictating the endurance of the material, are appended to the set of static features.

In total, an initial set of 20 static features is used, out of which the unimportant and correlated will be filtered out before training the model.

Dynamic features

Dynamic features, $\mathbf{u}$, are to be generated by processing the 3 signals, both in time and frequency domains. Exploiting the cyclic nature of the load, hence the response, from the time domain, the mean, peak, Root Mean Square (RMS), skewness, and kurtosis of the signal

are extracted. From the frequency spectrum of signal, ideally, the power spectrum density value at the load frequency is assumed to be the most valuable information. Nevertheless, the information about the load is to be unknown during the damage estimation procedure. Hence, considering this as a feature to train the model is physically an invalid assumption, and from a machine learning point of view introduces data leakage.

Alternatively, a convolution operation with a square kernel function is applied to the power spectrum to evaluate the mean energy content in subsequent frequency bands. In this problem, the domain of the frequency spectrum obtained from a one-sided Fast Fourier Transform (FFT) extends to 500 Hz, and is divided into 10 equisized frequency bands, each is 50 Hz wide. Hence, the convolution operation yields 10 mean band-energy values. In the upcoming text, these features are denoted by *PSDmean*.

The dynamic features naming convention follows the template

"acc<Node ID><direction>_<statistic>",

e.g. the feature "acc2Y_mean" refers to the mean of the accelerometer's Y-direction signal at node 2.

Having 5 temporal features, and 10 spectral features for each of the 3 sensor signals, in total, 45 initial dynamic features are used.

Damage estimation

From the problem definition, the loading of the cantilever is dominated by bending, and the most stressed point in the cantilever is the outermost fibre of the support (at node 1). Therefore, for each simulated boundary condition, the stress tensor is evaluated on node 1 at $z = {}^{h}/_{2}$ to estimate the damage, cf. Chapter 7.

For this problem, since HCF damage is dominant, Miner's linear damage accumulation rule is used. Since the stresses in the material are below its yield strength, the stress-strain relationship remains linear, and thus the relationship between the load and fatigue limit could be simplified from the Smith-Watson-Topper (SWT) rule to the Basquin's rule, given by

$$\sigma_{e,i} = \sigma_f N_{f,i}^{b}, \tag{14.2}$$

and $\sigma_{e,i}$ is obtained from Equation (7.11). A rainflow algorithm is applied on the variable loading sequence to extract the mean stress and stress amplitudes.

The damage for each beam is estimated after applying the load prescribed in Equation (14.1) for 10^8 load cycles. Figure 14.2 shows the distribution of accumulated damage obtained for each asset.

Two families of damage distributions could be observed in Figure 14.2.

The first family contains beams 1 to 4 and beam 33, and has a broad distribution of damages, inducing a first impression of high sensitivity to the loading conditions.
Beams 1 and 33 share the same dimensions and material properties, except that asset 1 is a box, while asset 33 is a tube. Both assets have vertical and horizontal wall thickness of 0.2 m, which is the thinnest wall in the fleet. This explains their large median damage value.
On the other hand, beams 2, 3 and 4 have box cross-sections, just like

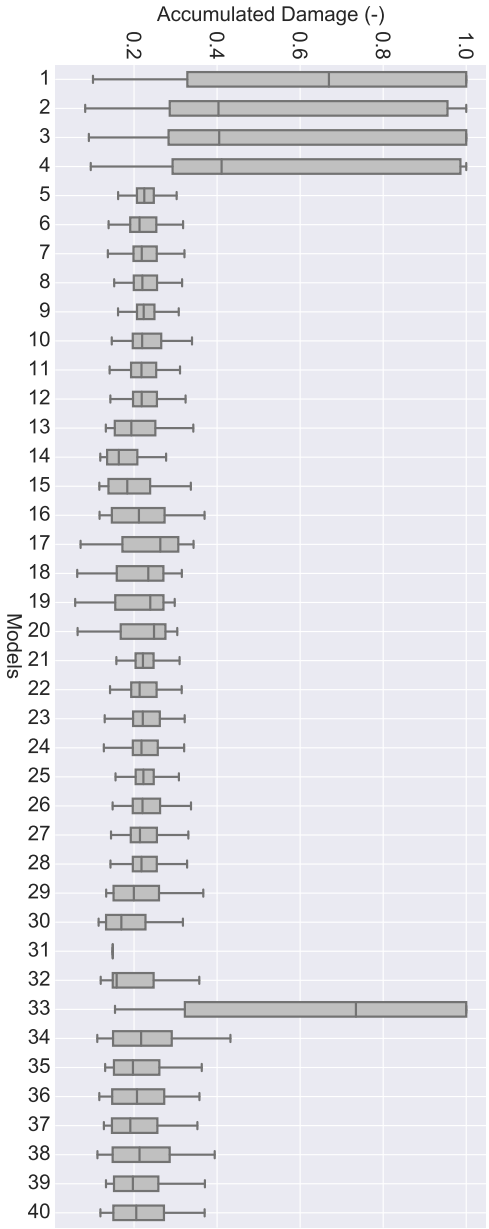


Figure 14.2.: Distribution of damage for each fleet asset

model 1, except that their vertical and/or horizontal wall thickness is 0.5 m instead of 0.2 m, hence making their smaller median damage value justifiable.

Contrarily, the remaining 36 exhibit both a smaller average value and variance in the damage accumulation with respect to the loading conditions. Their average damage values scattered between 0.15 to 0.26, and their standard deviations between 10^{-2} to 10^{-1}, compared to a standard deviation of ≈ 0.3 for the first family.

MLP configuration

To perform the training and evaluation of the Multi-Layer Perceptron (MLP), the fleet of beams is divided into

1. Basis models set $\mathcal{X}^{\mathsf{B}}$, whose damage estimates are to be combined linearly using the output weights $\boldsymbol{\alpha}$ from the MLP,
2. Training set $\mathcal{T}$, whose inputs $\mathbf{u}$ and $\mathbf{c}$, as well as corresponding damages $\mathbf{y}$, are provided to the MLP to infer their relationship, hence the appropriate $\boldsymbol{\alpha}^{\mathsf{M}}$,
3. Test set $\mathcal{E}$, whose $\mathbf{u}$ and $\mathbf{c}$ are input to the MLP to evaluate whether it can suggest $\boldsymbol{\alpha}^{\mathsf{M}}$ correctly, and hence yield a correct prediction of damage.

An important remark at this point is that the MLP during the training phase doesn't learn the relationship between the inputs and the outputs of models in $\mathcal{X}^{\mathsf{B}}$. Instead it only receives the *resulting outputs* from the basis models as input. Accordingly, when evaluating the model's performance during the testing phase, the model is tested both on the test set and the set of basis models. For the upcoming discussion, when referring to the *test set* without explicit exclusion of the basis models, the combined set of test models and basis models is hereby referred to.

14.2 Impact of Pre-classification

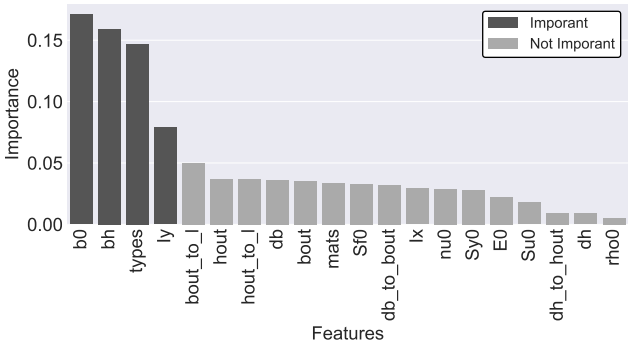
The first aspect investigated is the impact of pre-classifying damage severity based on the input static and dynamic features. Herein, two models are created:

1. The first model feeds all the data samples to the MLP, irrespective of their labelled damage severities. Thus, during training, the MLP is trained on the entire training set, and during testing, the model evaluates all test samples.
2. The second model incorporates a *Random Forest Classifier*, which is trained to classify the severity of damage. During training, the MLP is trained dominantly on the partially damaged samples, and fewer samples from both extremes. During testing, only if the sample is classified "partially damaged", its damage severity gets quantified using *moSAIc*. Else, its damage severity is mapped to the corresponding 0 and 1 damages.

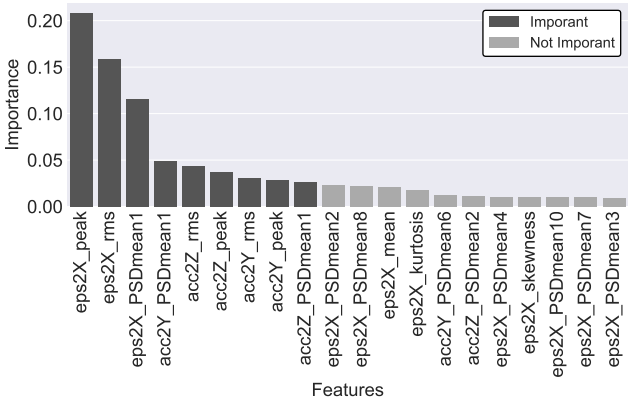
The hypothesis to be tested within this study is that the incorporation of a pre-classifier prior to the model's MLP yields a smaller than or equal Mean Absolute Relative Error (MARE) as compared to model without a pre-classifier.

For this investigation, an Extremely Randomized Trees Regressor (ERTR) is applied to assess the importance of static and dynamic features to the damage. The top 20 %of the features are chosen, then those among them having a Spearman correlation coefficient $\geq$ 0.9 are omitted. The importance ranking of the static and dynamic features are shown in Figure 14.3
Out of the static features, the nominal fatigue exponent b_0, the cross-section bounding area bh, the cross-section profile *type*, and the second moment of area around the y-axis I_y are considered important. Out of these, the second moment of area exhibits a 0.97 correlation coefficient with the bounding area, hence dropped, and only the first 3 features are considered.

From the dynamic features, 9 are deemed important, with the first three showing surpassing importance, see Figure 14.3(b). The first three features are all extracted from the strain signal, being its peak, RMS and average energy content in the first frequency band. A comprehensible selection, in fact, it is; the peak and RMS values relate to the amplitude of the cyclic load, whereas the energy in the first band encloses the signal energy content at the cyclic load's frequency. All 9 features exhibit a correlation coefficient ≥ 0.9, thus funnelling down to the peak value of the strain signal.



(a) Static



(b) Dynamic

Figure 14.3.: Feature importance ranking

The basis models are selected according to the procedure described in Section 13.2, hence, $n_b = 4$. The selected basis models are 13, 25, 33 and 37, and their corresponding selected static features are listed in Table 14.3. It can be seen that the reduced static features space is fully bounded by the chosen basis models, and that all variants of the reduced static features are covered within.

Table 14.3.: Static features of basis models

Model	b_0	bh	**Profile**
13	-0.209	1.0×10^{-2}	Box
25	-0.081	5.0×10^{-3}	Box
33	-0.209	2.5×10^{-3}	Tube
37	-0.081	2.5×10^{-3}	Tube

Both models undergo a 5-fold cross validation procedure to tune their hyper-parameters to attain the best performance. The two models are thereafter trained for 20 epochs, then evaluated on their combined test sets.

For this fleet, the beams are either fully or partially damaged, and none of the beams under any loading condition is negligibly damaged. This reduces the number of severities predicted by the pre-classifier to 2. For the model with a pre-classifier, the precision of the fully damaged and partially damaged severities are 0.78 and 0.97, respectively. In such a set-up, the precision is the most valued classification metric, since it is more important to have the severities correctly classified before being passed to *moSAIc*'s MLP. Of greater importance is the precision of the partially damaged samples, since a false prediction of the extreme severities could be corrected by the regression step downstream, whereas an over/under-estimation of the partial damage sticks to it.

Figure 14.4 shows the respective MARE of each set using both models. Generally, the model with a pre-classifier out-performs or at least yields equal MARE as the model without. Recall from Figure 14.2 that assets 1, 2, 3, 4 and 33 exhibited high sensitivity to

the operating conditions, and a non-trivial subset of their samples reached full damage after 10^8 load cycles. Particularly on these assets, the model with the pre-classifier demonstrates a satisfactory performance, with the MARE dropping by 18 to 28 %.

14.3 Selection of Features

The second point under investigation is the influence of the importance and correlation thresholds on the number of features yielded, and hence the implications this has on the train and test MARE. In this study, both importance and correlation thresholds are assigned a value from {0.2, 0.4, 0.6, 0.8, 1.0}.
With respect to the feature importance, this threshold presents the fraction of ranked feature importances, whose features are deemed important; i.e. a threshold of 0.4 means the features corresponding to the largest 40 % of the ranked importances. Whereas for feature correlation, the threshold is the minimum Spearman's coefficient of a feature pair, larger than which the features are assumed correlated. An importance/correlation threshold of 1.0 means that all features are assumed important/uncorrelated, respectively, hence none of them is filtered out.

Figure 14.5 shows a map of the resultant total number of features for each combination of the thresholds. In total, this problem had 20 static features, and 45 dynamic features.

Generally, the impact of the importance threshold on the resultant number of features is less pronounced than that of the correlation threshold. This is clearly indicated in the domain [0.4, 0.8], where the number of features remains constant with respect to the feature importance threshold. This indicates that even if the important features are over-selected, the correlation filter would still omit the superfluous features out.

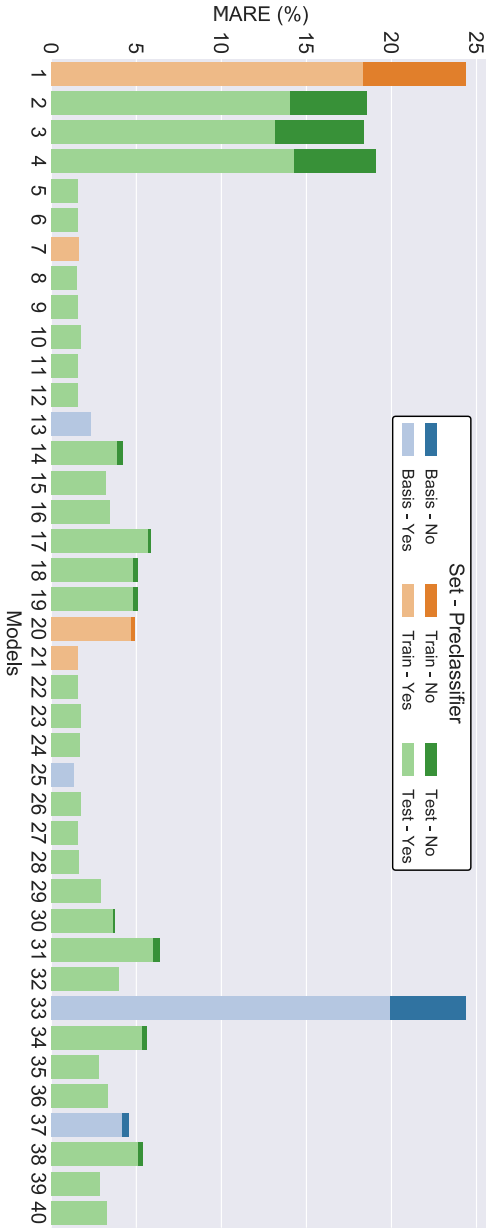


Figure 14.4.: MARE using a model with and without pre-classification on basis, train, and test sets

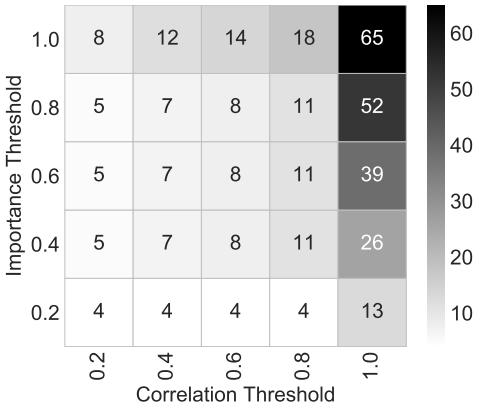
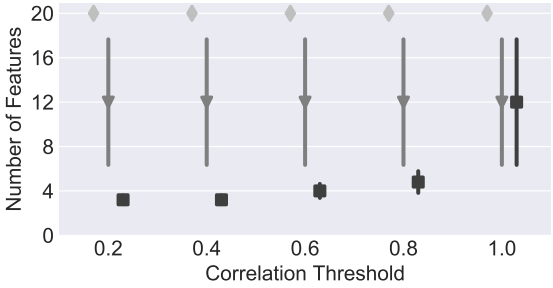


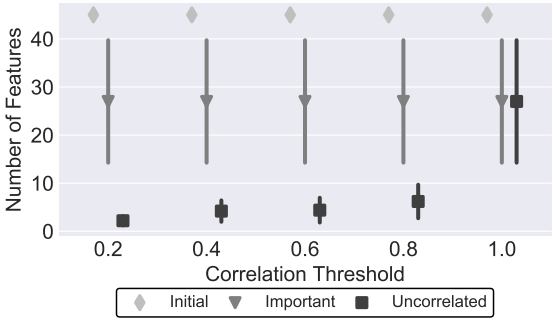
Figure 14.5.: Number of model features for different importance and correlation thresholds

Having concluded the stronger impact of the correlation threshold, the outcome of the feature selection steps for static as well as dynamic features is detailed in Figure 14.6. The different markers indicate the number of initial features, the important features, and the important, uncorrelated features. The error bars correspond to the variances in the number of features due to the different importance thresholds. As shown, when a correlation filter is applied, the variances in both static and dynamic features are significantly small.

The influence of the correlation threshold on the MARE is shown in Figure 14.7. The variance in the MARE over the range of thresholds considered is barely noticeable. Although the choice of the threshold significantly affected the number of features, the accuracy of the model seems to remain consistent after sufficient optimizer epochs, in this case 20.



(a) Static



(b) Dynamic

Figure 14.6.: Effect of correlation and importance thresholds on the number of features

14.4 Choice of Basis Models

The third aspect to be studied is the influence of the basis models selection criteria on the model's performance, evaluated on the train and test sets. In this study, two selection criteria are considered:

1. Optimal criteria: the basis models are chosen from the fleet assets based on the distance metric described in Equations (13.25) to (13.27).
2. Random selection: the basis models are drawn from the set of fleet assets without replacement. In this study, 3 draws are performed.

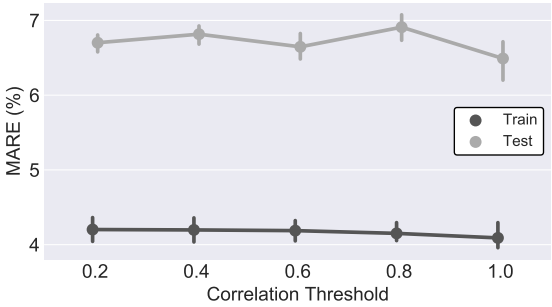


Figure 14.7.: Effect of correlation and importance thresholds on MARE

Additionally, the number of basis models is regarded as a variable, whose effect on the MARE to be investigated within this study. To this end, the condition that $n_b = n_{C+} + 1$ is ignored. Accordingly, two hypothesis can be formulated here:

1. Using the proposed criteria for optimal basis selection should yield smaller or equal MARE as compared to randomly selecting the basis models
2. An increase in the number of basis models yields smaller or equal MARE

For both hypothesis, the MARE is evaluated on the train and test sets. In this specific study, since the basis test changes for each test group, when evaluating the model's performance, the basis set is not included in the test set. However, for all groups, the test set $\mathcal{E} = \{1, 7, 20, 21\}$ is kept fixed. Moreover, the influence of hyper-parameter tuning is muted for this study, and the same set of hyper-parameters is used for all test cases.
In deployment, this is definitely not ideal, since it results in a poor model quality. However, as the focus of this study is the selection and size of the basis set, all other degrees of freedom in the model are deliberately fixed.

The variation of the train and test sets' MARE with the number of basis models is shown in Figure 14.8. Additionally, the 95 % confidence interval is evaluated for the 3 random selections and is plotted as well.

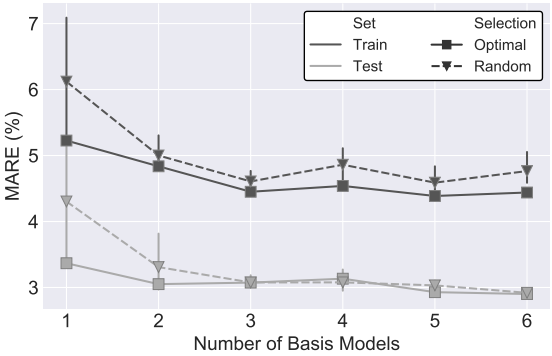


Figure 14.8.: Effect of basis choice on MARE

As shown, for the training set, the choice of the basis models using the proposed procedure yields a MARE smaller than or equal to the lower bound of the 95 % confidence interval of the MARE yielded from the random selection. With respect to the test set, the same observation holds for $n_b \leq 2$, however for $n_b > 2$ the selection criteria of the basis is less significantly impacting the performance.

The consistently smaller error observed for the train set is enough of a motivation to consider an optimal selection criteria for the basis. Although the impact on the test set errors is mild, the simplicity of the problem should be borne in mind. The static feature space is Cartesian within which the assets are uniformly scattered, thus interpolating between the assets degradations is rather straightforward. Therefore, a random selection of sufficient models would cover the feature space. In contrast, for a feature space with randomly scattered assets, like the one shown in Figure 13.6(b), the random choice of assets to be used as basis models is less prone to completely enclose the feature space.

14.5 Performance Assessment

The final stage of this numerical investigation is to assess the performance of the proposed hybrid model against other algorithms, typically applied to this class of problems. As hinted earlier, *moSAIc* is a Mixture of Experts model, whereby the damage estimates from an ensemble of physics-based degradation models are averaged by the weights suggeted from an MLP.

To compare its performance, the following three machine learning models are proposed and justified:

1. *Elastic-Net*: Linear regression combined with L1 and L2 regularization, attempting to minimize Equation (13.11).
2. *SVR*: A Support Vector Regression (SVR) with a radial basis function. Support Vector Regressions (SVRs) are commonly used in literature for this class of problems.
3. *Ensemble*: An ensemble of 50 SVRs with radial basis functions. As *moSAIc* is an ensemble of physics-based degradation models, it is relevant to have it compared to an ensemble of purely data-driven models.
4. *MLP*: A Multi-Layer Perceptron predicting the accumulated damage directly, unlike the MLP of *moSAIc*, which predicts the weights vector.

The four models above fit the set of features $(\mathbf{c}, \mathbf{u})$ to the accumulated damage y. The hyper-parameters of the four models are tuned after conducting a 5-fold cross-validation procedure. The four models, as well as *moSAIc* are evaluated on the combined test set, and the resulting mean and standard deviation of the MARE for each asset are shown in Figure 14.9.

Compared to other models, *moSAIc* yields a smaller MARE except on assets 1 and 33. These two assets exhibited full damage in some of their samples, as well as a large variance in the damage, indicating

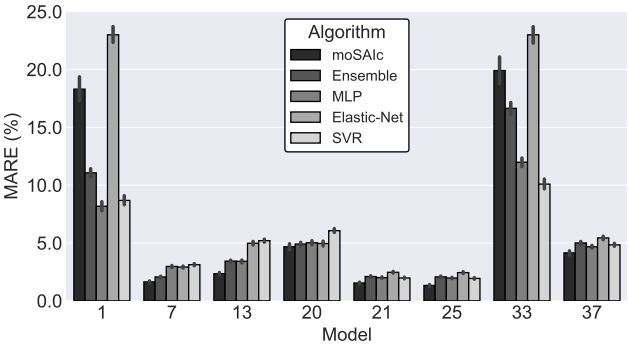


Figure 14.9.: Test set MARE using *moSAIc* and different data-driven models

higher sensitivity to the loading conditions. In these two cases, the regression capability of data-driven regression model surpasses the interpolation accuracy of *moSAIc*'s MLP in correctly fitting to the extreme damage cases, even with the presence of a pre-classifier.

Part VI

Industrial Applications

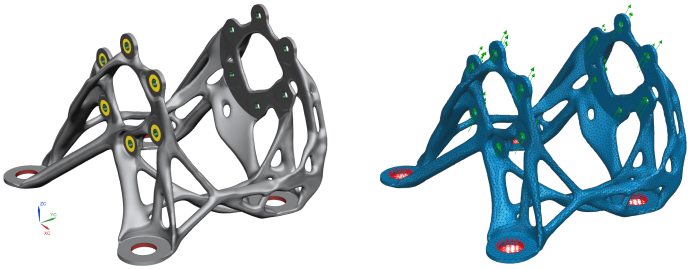
Gearbox Mount

15

The first presented industrial application is a gearbox mount for a two-seats vehicle. The mount was originally fabricated using traditional manufacturing processes, but with the rise of the unprecedented capabilities of Additive Manufacturing (AM), new optimal designs are now under investigation. One of the proposed optimal designs is shown in Figure 15.1(a).

15.1 Description

The mount is designed to withstand a gearbox with an average mass of 25 kg, fastened to the mount using 12 M8 bolts, six on each side, whose holes are indicated by the green surfaces in Figure 15.1(a). The mount is fastened to the chassis underneath with four M10 bolts, whose holes are shown as red surfaces.



(a) Geometry and regions (b) Mesh and boundary conditions

Figure 15.1.: Geometry of the gearbox mount, optimized for 3D printing.

15.1.1 Material and fabrication

As mentioned, AM techniques are used to fabricate the new design of the mount. Particularly, direct metal laser sintering (DMLS) [Fra14] process is used to print the component using Ti-6Al-4V grade 23. Further details about the parameters of the printing parameters and the mechanical tests conducted to infer the fatigue parameters of the material can be found in [Lam+19]. Various material specimens have been investigated with different print orientation and surface treatment settings.
For this work, material properties are drawn from the specimen with 0° print orientation, without any surface treatment after printing (designated *as-built*). The corresponding material properties of the said specimen are summarized in Table 15.1.

Table 15.1.: Material properties of the gearbox bracket

Material property	Value
E	75 GPa
ν	0.27
σ_f	6652 MPa
b	−0.319
ε_f	0.084 mm/mm
c	−0.319
d	7.0

15.1.2 Damage accumulation law

The specimens manufactured from Ti-6Al-4V using 3D printing commonly exhibit non-linear fatigue degradation behaviour. Thus, applying the linear Miner's degradation rule in Equation (7.7) yields inaccurate estimates. Instead, a modification is introduced, and the damage accumulation rule is re-written as

$$D_a = \left(\sum_{i}^{n_\sigma} \frac{N_i}{N_{f,i}} \right)^d , \tag{15.1}$$

where d is an experimentally-determined material parameter, describing the rate of Young's modulus degradation in relation to the life fraction $N_i/N_{f,i}$ [EP11a; EP11b].

15.1.3 FE modelling

Finite elements are used to discretize the geometry using linear 4-noded tetrahedral elements, with an average element size of 5 mm is applied. A damping ratio of 5% of the first system natural frequencies is assumed to calculate the system damping. The covariance of the process noise is assumed identical and independent for all system states, and equal to 5×10^{-2}. The material model is assumed linear and elastic, since only High-Cycle Fatigue (HCF) is under investigation, and no stresses beyond the yield strength are expected. For simplicity, the distribution of porosities in the print is assumed uniform throughout the component. Hence, an isotropic material model is used and material properties are assumed constant at all material points.

Both the bolts and the gearbox assembly are assumed significantly stiffer than the mount, hence replaced by boundary conditions in this analysis. Full supports are used to fix the mount at the location where it is fastened to the chassis (red surfaces), while the forces resulting from the acceleration of the gearbox assembly are applied directly on the inner surfaces of the 12 M8 holes.

15.1.4 Load Scenarios

Different road conditions are assumed for the vehicle, which inflict different cyclic loads on the gearbox. The road profile considered varies between *smooth* and *rough*, whose imposed loads are described

as gravitational forces (g-forces) on the bracket. The load properties of both road conditions are summarized in Table 15.2.

Table 15.2.: Load parameters for different road conditions

Parameter	Smooth	Rough
Duration	1200 s	2400 s
Frequency	2 Hz	2 Hz
Amplitude	[2, 2, 5] g	[5, 4, 10] g
Static Offset	[0, 0, 0] g	[0, 2, 4] g
Noise	$\mathcal{N}(0, 0.5)$ g	$\mathcal{N}(0, 2)$ g

Both load signals are sinusoidal functions having 2 Hz frequency. The vehicle is assumed to drive 1200 s on a smooth road, and twice as much on a rough road. On a smooth road, zero-mean cyclic load is imposed with magnitudes 2 g, 2 g and 5 g in the X-, Y- and Z-directions, respectively. On the other hand, static load is superposed on the cyclic load in the rough road section, resulting in shifted-mean cyclic loads. This will effectively introduce mean effects to the stress and strain cycles, and will have a pronounced effect on the components fatigue behaviour. For both load profiles, corresponding Gaussian noise is superposed, where their parameters are indicated in Table 15.2. Figure 15.2 shows a close-up on the transition between the smooth and rough road at 1200 ± 2 s.

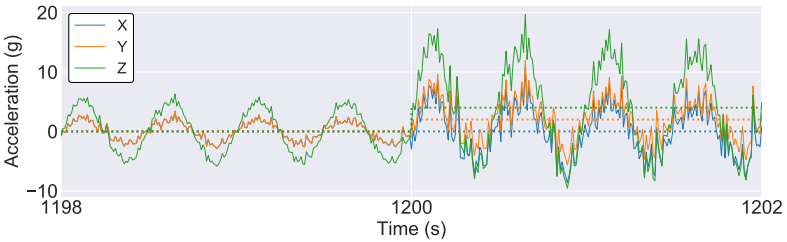


Figure 15.2.: Applied load profile - load signal (solid), static offset(dotted)

The two load conditions are applied periodically for a total operation time of 10^5 hours, where the vehicle is supposed to be running in consequent laps, each lap lasting one hour. To reduce the computational cost of the problem, only a small portion of this operation profile is estimated using the Kalman filter procedure. Particularly, the 4 s window shown in Figure 15.2, during which the mount experiences the transition from load condition to the other is of interest. Elsewhere, the quality of the filter is assumed to remain constant throughout the operation due to the periodicity of the load.

15.1.5 Candidate sensors

Due to the curved, non-uniform topology of the optimized mount's geometry, the regions where sensors could be installed are rather limited. Permitted surfaces for sensor placement are indicated by the yellow faces in Figure 15.1(a), and opposing faces on the other side of the mount. The permissible regions are essentially the periphery of the holes where the gearbox is fastened, which were intentionally constrained during the optimization process to remain flat to ease the gearbox attachment.

For this problem, the sensor configurations are composed of uni-axial accelerometers and strain gauges, with the measurement noises assumed identical and independent among sensors of the same type. The associated noises are assumed $10^{-3}\,\mathrm{mm\,s^{-2}}$ and 10^{-1} for accelerometers and strain gauges, respectively [Pap+11; ZX16].

Typically, uni-axial strain gauges are assembled into strain gauge rosettes, with $0-45-90$ or $0-60-120$ arrangements. This constraint is, however, not taken into consideration, and all sensors are assumed uni-axial and can be mounted independent from one another, to avoid imposing additional complexity in the formulation of the output matrix in Equation (8.7). Thus, either sensors can be oriented in the global Cartesian X-, Y- or Z-directions.

15.2 Results

Due to the limited space allocated for sensor placement, the number of sensors attached to this bracket is to be kept to minimum. Hence, a threshold $\epsilon_0 = 10^{-2}$ is used to determine n_s^*. For this problem, this yields $n_s^* = 2$ when mounting a multi-type configuration.

Five experiments are conducted for this use case. Figure 15.3 shows the sensor configuration yielding the smallest $\mathrm{Tr}\,(\mathbf{P}^{x+})$, with a value of 0.934, and a mean state estimation error $\bar{\epsilon}_x$ of 4 %. The configuration is composed of one accelerometer and one strain gauge both oriented in the X-direction, with each mounted on one side of the structure as indicated by the two views in Figure 15.3.

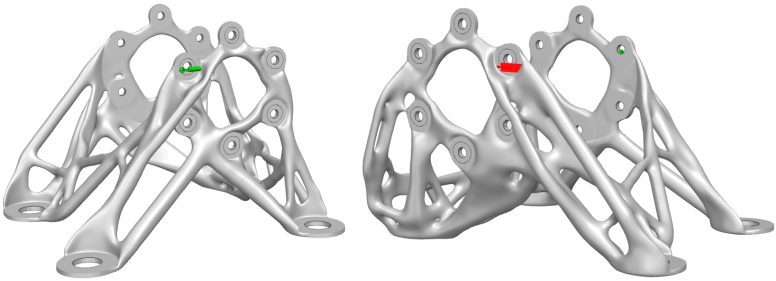
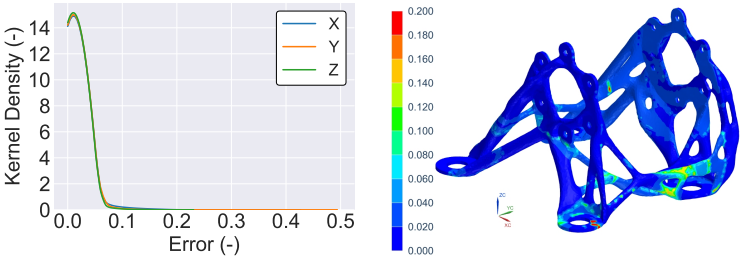


Figure 15.3.: Optimal sensor configuration – accelerometers (green), strain gauges (red).

The relative error of each state estimate is calculated using Equation (11.1), and is averaged over the 4 s duration of the Kalman filter. The errors of the displacement and velocity state estimates, respectively $\tilde{\mathbf{z}}$ and $\dot{\tilde{\mathbf{z}}}$, are examined separately. Figures 15.4(a) and 15.5(a) show the Kernel Density Estimate (KDE) of the X-, Y-, and Z-displacement and velocity estimation errors, respectively.

Impressively small errors can be observed for displacement estimates, with a mean value less than 1 %. The error variable is nearly following an exponential distribution, and its KDE function diminishes

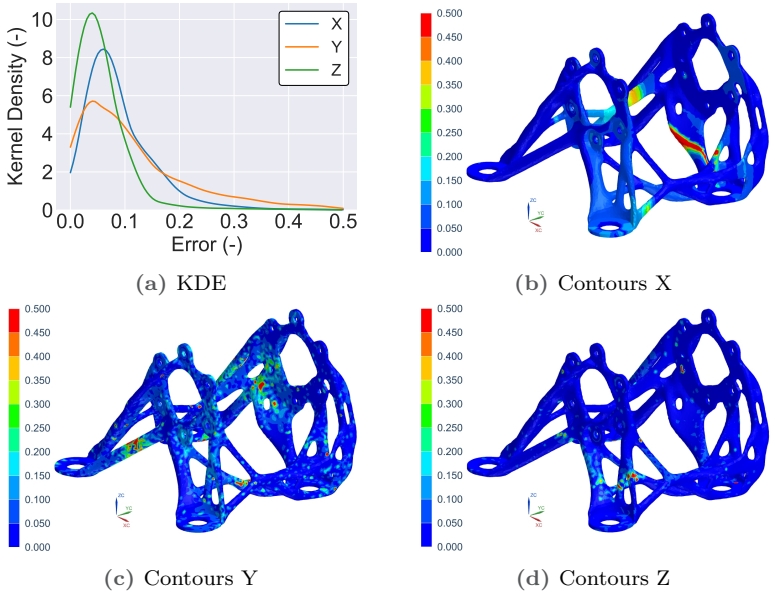
significantly for error values $\geq 10\,\%$.
Figure 15.4(b) shows the contours of the mean X-, Y-, and Z-errors at each node. The colour scale is capped at $20\,\%$. The contours are clearly dominated by error values below $5\,\%$, and only $3\,\%$ of the nodes have associated errors between $5\,\%$ and $20\,\%$. Such outliers are not scattered in one region, but rather scattered over the geometry, therefore, they can be attributed to numerical inaccuracies in the model, e.g. due to coarse mesh, or poor element quality.



(a) KDE (b) Contours

Figure 15.4.: KDE and contours of the displacements estimation mean relative error

The estimate of the velocity states are not as accurate as the displacements', as indicated by the KDEs in Figure 15.5(a). Although the mean values of the X-, Y- and Z-velocity estimation errors evaluate to $10\,\%$, $13\,\%$ and $6\,\%$, which are acceptable, the scale of the KDE, especially of the X- and Y- errors indicates that a non-ignorable number of states have larger errors. This can be better visualized in Figures 15.5(b) to 15.5(d), where the contours of each error are shown on the mount's geometry. With respect to the contours of the X- component, two regions seem to have spikes of relative errors, whereas for the Y-components, the nodes with high errors are scattered over the structure. Nonetheless, the three error components are dominated by values below the $10\,\%$ mark, which is very acceptable for the velocity field, since it doesn't contribute directly to the reconstruction of the strain field used in damage estimation.



(a) KDE

(b) Contours X

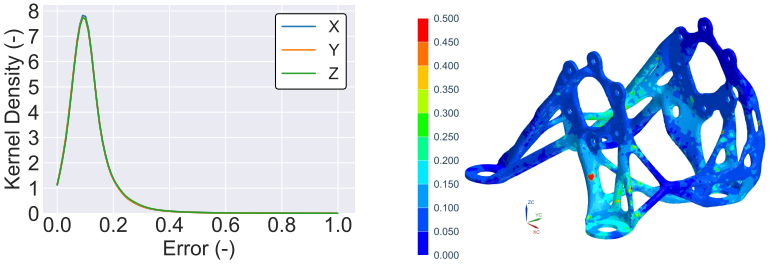
(c) Contours Y

(d) Contours Z

Figure 15.5.: KDE and contours of velocities estimation mean relative error

From the estimated states, particularly displacement, the strain field is reconstructed by applying Equation (7.4). The reconstructed strains are to be used in damage prediction, following the procedure and assumptions explained in Chapter 7. The distributions of the reconstruction errors over the system outputs are shown in Figure 15.6(a). Figure 15.6(b) shows the distribution of the mean errors at each element overlaid on the geometry. The strain reconstruction errors follow coherently the distribution of the displacement estimation errors, due to the physical relationship between the two variables, as expressed by Equation (7.4). Albeit its mean being about 11 %, the distribution of the strain reconstruction error has a relatively wider scale. Such a spread manifests to the accumulated error from the state estimation, and is suspected to impact the damage estimates. Nevertheless, for the rather restricted instrumentation

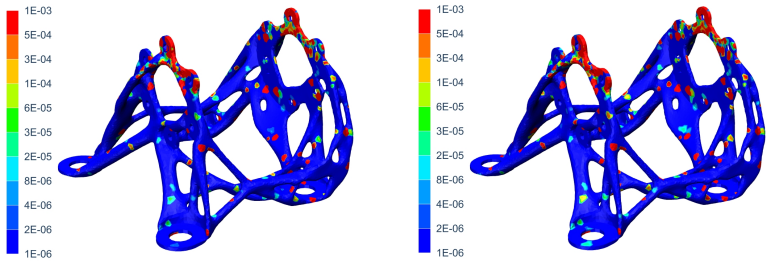
and non-trivial system, this error is regarded within tolerance, and the reconstructed field is proceeded with to the damage estimation procedure.



(a) KDE (b) Contours

Figure 15.6.: KDE and contours of relative output reconstruction error

The accumulated damage after the 10^5 hours of operation is evaluated, and its distribution is displayed on the geometry as shown in Figure 15.7. Although the bracket is optimized for low weight-to-strength ratio, its design is still rugged and can withstand a typical load scenario of a conventional vehicle. The accumulated damage on the structure ranges between 10^{-6} to 10^{-3}, indicating at the first glance a healthy condition of the component.



(a) Estimate (b) Actual

Figure 15.7.: Contours of the accumulated damage

Nonetheless, it is worth mentioning that the non-linearity of the damage accumulation rule incurs a delusional interpretation on the component's reliability. It could be misinterpreted that the component is still far away from its end-of-life, only because its accumulated damage is not close to 1. This correlation doesn't hold for non-linearly degrading materials. Looking at Figure 15.8, the material having a non-linear damage accumulation behaviour ($d = 7$) consumes almost 90 % of its useful lifetime with the corresponding accumulated damage just turning 0.5, whereas the last 10 % of the lifetime corresponds to as much of the damage accumulation. Such abrupt material degradation towards the end of life is rather common in fabricated composites and printed materials [Kam+15].

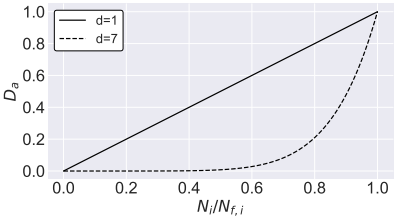
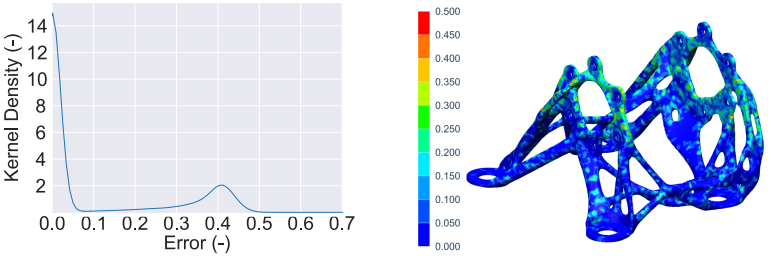


Figure 15.8.: Linear and non-linear damage accumulation behaviour

Examining the distribution of the relative error in the accumulated damage value calculated after 10^5 hours, Figure 15.9(a) shows a bi-modal KDE, with the first and dominant mode at 1.42 % relative error, and the second mode at 38.6 %. Although more than 75 % of the elements have associated errors below 10 %, the error in the remaining elements is alerting, especially for a material that exhibits rapid degradation towards its end-of-life expectancy.

Comparing the accumulated damage distribution in Figure 15.7(b) with the error distribution, it can be observed that some regions with high actual damage values have a discrepancy in their estimates. The errors in those regions cannot be claimed to be sporadic, as they form a continuous gradient with the values around them.



(a) KDE (b) Contours

Figure 15.9.: KDE and contours of relative accumulated damage estimation error

Although the strain reconstruction seems accurate, small variations in the signals could yield errors during the cycle counting procedure, which get augmented by the exponential damage rule.
The restricted domain for placing sensors is suspected to impede further improvements in the accuracy of the solution. It would be ideal to consider the original geometry of the mount while performing optimal sensor placement, and to constrain the surfaces where the sensors are to be placed during the topology optimization.

16 Bearing Support

The second presented application is a bracket acting as a bearing housing, found in aircrafts. The bracket was originally designed by a manufacturer of fastening systems and rings for aerospace and industrial applications to be fabricated using conventional manufacturing processes. The bracket design was published in 2016 through a design competition[1], aiming at optimizing the old design to suit Additive Manufacturing (AM), while retaining its mechanical performance. Figure 16.1 shows one of the selected optimal designs.

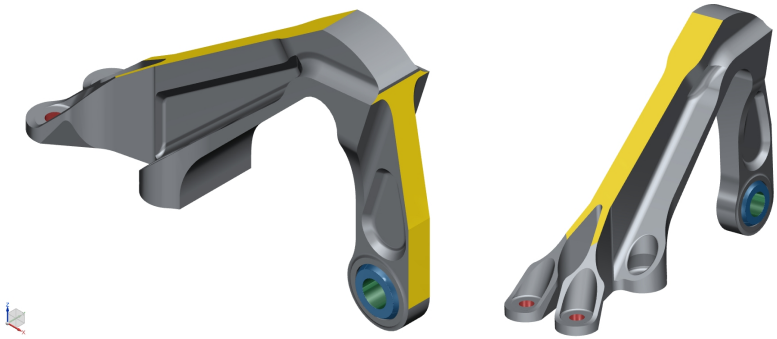


Figure 16.1.: Bearing support geometry

16.1 Description

The bracket is fastened with four 10-32 high strength tension rated bolts, mounted in the red-coloured holes at the rear end of the bracket. A custom-design spherical bearing (the blue component) is

[1] For more details:
`https://grabcad.com/challenges/airplane-bearing-bracket-challenge`

mounted in the special grooving at the front end. External forces from the assembled component are applied on the inner surface (green face) of the bearing. As specified by the manufacturer, the bracket is to hold a dynamic load of 2.5 kN in the horizontal direction, and 5.5 kN in the vertical direction.

16.1.1 Material and fabrication

No detailed specifications about the fatigue properties of the material used in AM were provided, thus the material specifications presented in [Ste+15] are assumed. The material properties are for Ti-6Al-4V alloy fabricated using Laser Engineered Net Shaping (LENS) 3D-printing process, proceeded by hot isostatic pressing (HIP) treatment. Titanium alloys are highly attractive materials to be fabricated using AM processes, due to their easy manufacturability, as well as high strength-to-weight ratios. In aerospace applications, this results in lighter components, hence significant cost reduction.

16.1.2 Damage accumulation law

In this use case, to estimate fatigue damage in the component, the Smith-Watson-Topper (SWT) parameter in Equation (7.9) is applied to calculate the fatigue limit, as described in Chapter 7. It is common to observe a highly non-linear damage accumulation in AM components, thus the non-linear damage accumulation rule in Equation (15.1) is applied.

16.1.3 FE modelling

A linear elastic material model is assumed, with the mechanical and fatigue properties summarized in Table 16.1. A damping ratio of 5% of the first system natural frequencies is assumed to calculate the system damping.

Linear tetrahedral finite elements (4-noded), with average edge size of 3 mm are used to mesh the geometry, as shown in Figure 16.2. The

Table 16.1.: Material properties of the bearing support

Material property	Value
E	106 GPa
ν	0.29
σ_f	2310 MPa
b	−0.136
ε_f	0.256 mm/mm
c	−0.797
d	7.0

boundary conditions are, as well, indicated on the same figure. The covariance of the process noise is assumed identical and independent for all system states, and equal to 5×10^{-2}. The support is assumed to be fastened to a rigid foundation, using bolts with significantly higher stiffness. Hence, the DOFs of fasteners' holes and the base are fully constrained. The stiffness of the spherical bearing is assumed very high, thus the bearing is dropped off the Finite Elements (FE) model, and the external load is applied directly on the bearing housing as a distributed surface force.

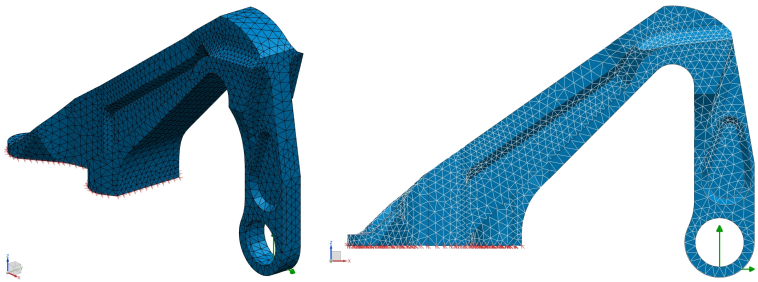


Figure 16.2.: Bearing support mesh, fixed-constraints (red), and loads (green)

16.1.4 Load scenarios

A 10 Hz sinusoidal load of magnitudes 2.5 kN and 5.5 kN in the horizontal and vertical directions is applied. To introduce noise, a Gaussian process, following $\mathcal{N}(\mu_n, 1)$, where μ_n is a 100 Hz sine wave is superposed on the load signal. A sample of the applied load profile is shown in Figure 16.3

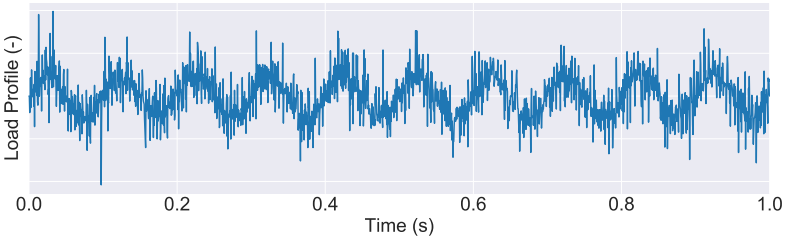


Figure 16.3.: Applied load profile

16.1.5 Candidate sensors

Due to the curved surfaces of the optimized bracket, limited positions are available to mount the sensors. The yellow surfaces, shown in Figure 16.1, are the only candidate surfaces to mount either strain gauges or accelerometers. The associated measurement noise covariances are independent and identical for all sensors of the same type, and are assumed $10^{-3}\,\mathrm{mm\,s^{-2}}$ and 10^{-1} for accelerometers and strain gauges, respectively, [Pap+11; ZX16]. For this structure, either sensor type can be oriented in the global X-, Y- or Z-directions, and all sensors are assumed uni-axial.

Since weight is a major concern in aerospace applications, the number of sensors attached to this bracket is to be kept to a minimum. Hence, a threshold $\epsilon_0 = 10^{-2}$ is used to determine n_s^*. For this problem, this yields $n_s^* = 4$ when mounting a multi-type configuration.

16.2 Results

Five experiments are conducted for this use case. Figure 16.4 shows a resulting sensor configuration, consisting of one accelerometer in the Z-direction (green), and three strain gauges (red), each oriented in one of the three spatial directions. The shown configuration yields the smallest $\mathrm{Tr}\,(\mathbf{P}^{x+})$, with a value of 0.820, and a mean state estimation error $\bar{\epsilon}_x$ of 9.58 % with a standard deviation of 5.12 %. Over the five experiments, the resulting mean and standard deviation of $\mathrm{Tr}\,(\mathbf{P}^{x+})$ are 0.828 and 0.01, respectively, with $\bar{\epsilon}_x$ ranging between 3.58 to 17.0 %. Despite the discrepancy of the resulting sensor configurations, the yielded estimates are rather accurate, and their associated uncertainty is significantly low.

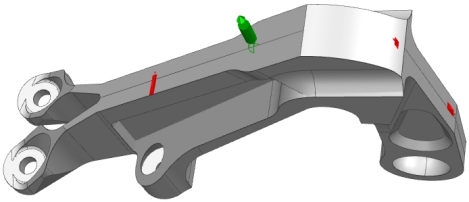
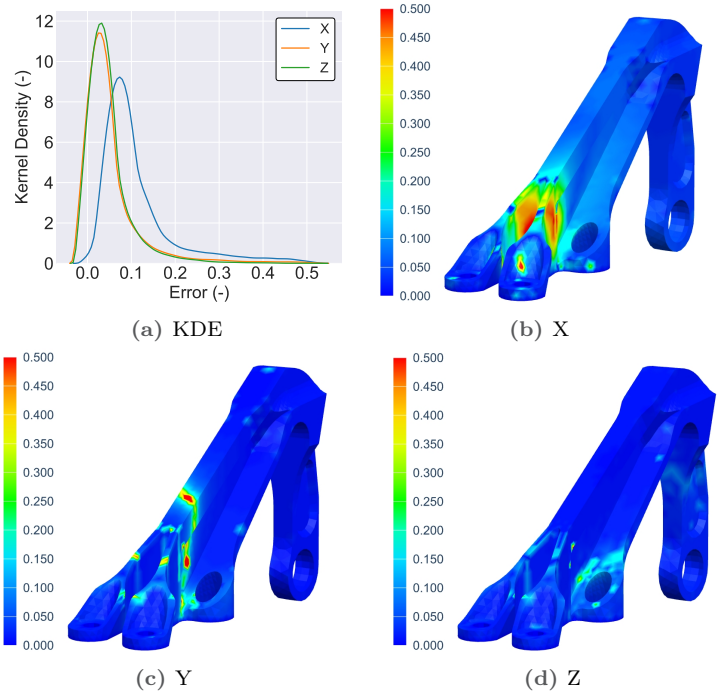


Figure 16.4.: Optimal sensor configuration – accelerometers (green), strain gauges (red).

The mean relative error of the state estimates over the 1 s window is calculated using Equation (11.1). Figures 16.5(a) and 16.6(a) show the Kernel Density Estimate (KDE) of the error for X-, Y-, and Z-displacement and velocity states, respectively. The density functions are clearly left-skewed, with medians errors 8.37 %, 2.72 % and 2.83 % for X-, Y-, and Z-displacement estimates, respectively, and 7.79 %, 4.70 % and 4.99 % for X-, Y-, and Z-velocity estimates, respectively.

Figures 16.5(b) to 16.5(d) show the contours of the mean estimation relative error of each displacement component. The error is calculated

over a 1 s interval after the Kalman filter's convergence; i.e. after the state estimation covariance plateaus at 0.828. As shown, for the Y- and Z-displacement fields, the time-averaged error remains below 10 %, with the error going sporadically higher at some points. The region around the bracket's support exhibits larger errors in the X-displacement field.

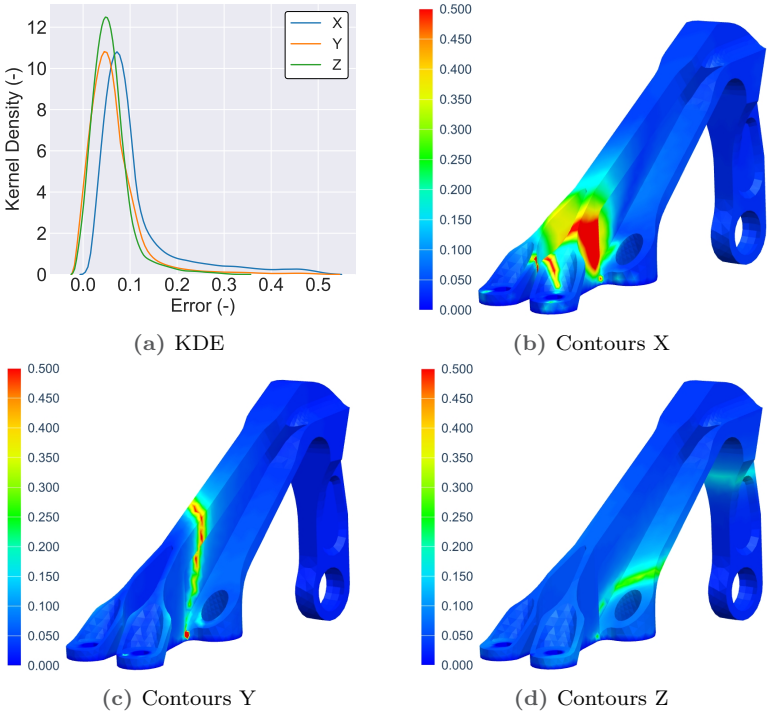


(a) KDE (b) X

(c) Y (d) Z

Figure 16.5.: KDE and contours of the mean estimation relative error of displacement

The mean estimation relative error of the X-, Y- and Z-velocities are shown in Figures 16.6(b) to 16.6(d), respectively. Likewise, the three components exhibit estimation errors well below 10 % on average, with the X-component showing a larger deviation from the true values around the supports. Such discrepancy in the X-velocity correlates strongly to the discrepancy of the X-displacement estimates.



(a) KDE (b) Contours X

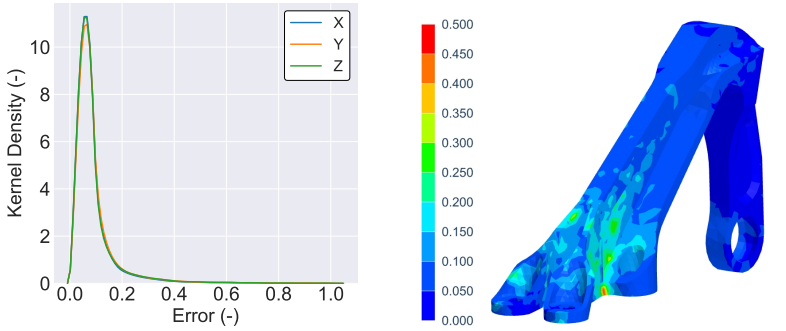
(c) Contours Y (d) Contours Z

Figure 16.6.: KDE and contours of velocity estimation relative error

From the estimated states, the strain field is reconstructed to be used in damage prediction. The distribution of the respective mean strain reconstruction error over system outputs is shown in Figure 16.7(a). Error patterns similar to those of the displacement estimates can be observed due to the physical relationship between the two variables. The errors from estimating the three normal strain fields are also left skewed, and all have mean values below 10 %.

Similar to the state estimation error, the contours of the reconstruction error averaged over a one-second window after the Kalman filter's convergence are calculated, and shown in Figure 16.7(b). Ignoring the sporadic outlier values, the temporally-averaged strain

field reconstruction error remains mostly below 15 %. The regions exhibiting larger reconstruction errors are non-critical from an engineering perspective, them being in the vicinity of the support. Herein, the shown contours present satisfactory estimates, upon which the damage in the structure under the assumed loading profile can be estimated.

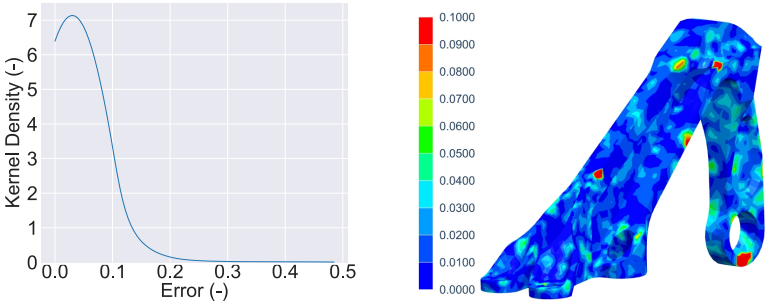


(a) KDE (b) Contours

Figure 16.7.: KDE and contours of relative strain reconstruction error

The one-second load profile shown in Figure 16.3 is repeatedly applied 10^7 times on the structure, and based on the reconstructed strain field, Equation (15.1) is used to estimate the accumulated damage. The true values of the damage estimate are evaluated based on the outputs of the simulated perturbed system. The damage estimation error covariance matrix is also evaluated at the end of the simulation, and the resulting trace of the covariance of the damage estimation error $\mathrm{Tr}\left(\mathbf{P}^{\mathrm{D}}\right)$ is 0.236.

Figure 16.8(b) shows the contours of the mean relative damage estimation error. As indicated by the colour scale, the error falls dominantly below 10 %, with an average and standard deviation of the 3.69 % and 6.65 %. The distribution of the error is skewed towards the lower error values, with 95 % of the elements having relative errors $\leq$10 %.



(a) KDE (b) Contours

Figure 16.8.: KDE and contours of relative accumulated damage estimation error

Nonetheless, scatters of error values greater than 10 % can be observed, despite the high accuracy and precision of the state estimation and output reconstruction. Those exceptionally larger errors could be attributed to the high non-linearity in the damage accumulation rule.

First, the formulation in Equation (15.1) is very sensitive to variations in the cycle ratio, ${}^{N_i}/_{N_{f,i}}$, due to the exponent d. Second, the accumulated damage is a function of the cumulative sum of cycle ratios; unlike Miner's rule which is a function of the cycle ratios themselves. The combination of these two factors lead to the propagation and augmentation of small errors that arise in the cycle ratio. Errors in the cycle ratio could stem either from the cycle counting algorithm, introducing errors in N_i , or from the variations in the stress and/or strain values used to calculate the SWT parameter, thus leading to errors in $N_{f,i}$. The latter seems to be the source of discrepancy in this problem, due to the slight variations in the reconstructed strains.

17 Servomotors

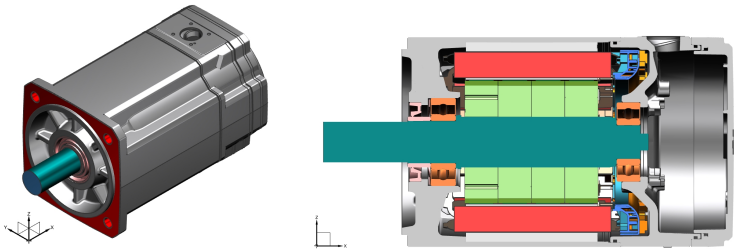
Servomotors are abundantly used in a broad domain of industrial applications, both as rotary and linear motion driver. Typical engineering systems, in which servomotors are used, are e.g. robotics, CNC machines, and automated manufacturing. Servomotors come with several advantages, such as broader range of operating speeds, and precise and smooth torques, even at high speeds.
Despite their ruggedness and reliability, servomotors can't afford being operated at their limits for long periods. Running at maximum torque for extended durations makes them susceptible to failure.

17.1 Description

In this chapter, the estimation of the accumulated damage in servomotors exposed to a range of typical and excessive operating conditions is investigated. Herein, Simulation Models and Artificial Intelligence Combined (*moSAIc*) is tested on a fleet of servomotors adopted from Siemens *SIMOTICS* series with power rating from 0.05 kW to 0.9 kW.

For modelling simplicity, the power electronics, encoder, and control circuits are omitted and only the mechanical sub-components are considered. Figure 17.1 shows a typical isometric and a cross-section view of a SIMOTICS motor, excluding the unconsidered units.

Based on the shown cross-section, 18 servomotor variants are generated following the dimensions and specifications in the product catalogues. Despite the strong similarity between the SIMOTICS motors and the models shown here, it is crucial to note that the assets considered are not the actual Siemens products, but rather synthesized replicas intended solely for the sake of this study.



(a) Isometric (b) Cross-section

Figure 17.1.: Typical CAD of SIMOTICS servomotor

17.1.1 Model features

The information retrieved from the product catalogues are considered to define the set of static features $\mathcal{F}^{\mathsf{C}}$. In practice, catalogue information are provided by manufacturers to motor operators, hence it is fair to assume the availability of such information for both motors with and without degradation models.

The set of static features can be sub-categorized into:

- Operation features, represented by the parameters *nrated, Prated, M0, Mrated, Irated, n_poles, J, m* denoting the rated speed, rated power, static torque, rated torque, rated current, number of poles, polar moment of inertia, and mass, respectively.
- Driving end/non-driving end bearings features, represented by the parameters *BDi, BDo, BB* denoting the bearings' inner, outer, and ball diameters, respectively, and the parameters *Cd, C0* denoting the dynamic and static load ratings, respectively.
- Geometric features, represented by the parameters *SH, P, N, LA, M, AB, T, I2, S, D, E, GA, F, LB, O1, O2*, and denoting dimensions of the housing, shaft, and connection flange.

Material properties are not included in the static features for this application since the sub-components of all considered fleet variants are manufactured of the same material.

A total of 29 static features are considered for each fleet variant. Each servomotor is not necessarily described by a fully-unique set of static features. For instance, the fleet covers only three unique rated speeds, 2000, 3000 and 6000 RPM, associated to motors 1, 5 and 12. This would in turn have an influence on the choice of the set of basis motors, which should cover the space of static features. Additionally, this irregularity in parameters distribution among the fleet variants could inevitably introduce class unbalances between the train, test and basis sets.

With respect to the dynamic features set $\mathcal{F}^{\mathsf{U}}$, the statistical quantities listed in Section 13.2 are extracted from acceleration signals of triaxial sensors mounted at 10 different positions on the motor's housing, as indicated by the green cylinders in Figure 17.2.

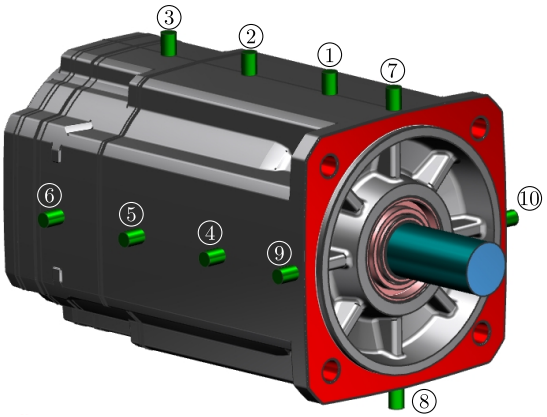


Figure 17.2.: Positions of accelerometer sensors (green), fixation surface (red), and Boundary Conditions (BCs) (teal and light blue)

For each acceleration signal, the mean, peak, Root Mean Square (RMS), skewness, and kurtosis values are evaluated from the time series. In addition, the power spectrum densities at the first three harmonics are extracted from the signal frequency spectrum. No temporal-spectral features are used in this problem. This yields 8 dynamic features per signal, and a total of 240 dynamic features for each motor's operation scenario.

The dynamic features naming convention follows the template

"acc<ID><direction>_<statistic>".

For instance, the feature "acc5Y_mean" refers to the mean of the 5^{th} accelerometer's Y-direction signal.

17.1.2 Data synthesis

The acceleration signals are obtained for different operation scenarios designed based on the operation specifications and load capacity of each fleet variant. Each scenario is described by three Boundary Conditions (BCs) applied to the driving end of the shaft

1. Radial preload force, $F_r \sim \mathcal{N}\left(\alpha_F F_{r0}, 10^{-4}\right)$,
2. Axial moment, $M_x \sim \mathcal{N}\left(\alpha_M M_{x0}, 10^{-2}\right)$,
3. Parallel misalignment, $\Delta_p = \alpha_\Delta \Delta_{p0}$.

$\mathrm{F_{r0}}$ (kN), $\mathrm{M_{x0}}$ (Nm), and $\mathrm{\Delta_{p0}}$ (mm) are nominal radial force, axial moment, and shaft misalignments specific to each motor, and assumed based on its bearing's static load capacity, nominal torque and shaft's diameter, respectively. The nominal values for each fleet variant are summarized in Table 17.1. $\alpha_F, \alpha_M, \alpha_\Delta \sim \mathcal{U}(1, 25)$ are multipliers applied to each BC to vary their magnitude with respect to the nominal values.

Variation in the loads imposed on the motor from the driven appliance is accounted for by sampling the radial force and axial moment from a

Gaussian distribution, whereas the parallel misalignment is assumed to remain constant. A set of 11 random values are sampled from the uniform distribution $\mathcal{U}(1, 25)$, and are assigned to all the multipliers. This yields a total of 1331 combinations of multipliers, hence, 1331 operation scenarios for each motor.

Table 17.1.: Nominal boundary conditions of different motors

Motor	Δ_{p0} (mm)	F_{r0} (kN)	M_{x0} (Nm)
1	2.0	1.550	0.08
2	2.0	1.550	0.08
3	2.0	1.550	0.16
4	2.0	1.550	0.16
5	2.8	1.875	0.60
6	2.8	1.875	0.60
7	3.6	2.375	1.00
8	3.6	2.375	0.80
9	3.6	2.375	1.20
10	3.6	2.375	0.90
11	3.6	2.375	1.45
12	3.6	2.375	1.00
13	4.8	3.900	1.10
14	4.8	3.900	2.80
15	4.8	3.900	2.60
16	4.8	3.900	2.60
17	4.8	3.900	1.50
18	4.8	3.900	1.50

The fleet motors operations are numerically simulated using Finite Element Analysis (FEA). Linear, tetrahedral (4-noded) finite elements are used to mesh the geometry, and a time step of 1.25×10^{-3} s is applied for temporal discretization. The critical time step is inferred from the maximum eigenfrequency of the system to be 2×10^{-3} s. For all the subcomponents, a linear-elastic, isotropic material definition is used.

The torque and axial preload are applied on the curved surface of the exposed section of the shaft (teal surface in Figure 17.2), while the misalignment is applied on its flat surface (light blue in Figure 17.2). The motor is assembled such that the driven component is vertically underneath, thus an additional gravity load is applied in the x-direction on the entire assembly. The four holes on the driving end are used to fasten the motor housing, whereby the flange becomes fully in contact with the motor mount. Hence, the red surfaces in Figure 17.2 are fully supported.

17.1.3 Damage estimation

The root of the shaft; the section at which it enters the motor housing, is highly prone to failure due to extended operation times at high torques, especially in the presence of misalignments. Typically, bearings would fail as well, however in the scope of this work, bearing faults are not in focus.

Damage is estimated based on the stress distribution evaluated at the nodes located at the shaft's root. Because the motors are initially designed with a high safety factor, the applied loads, even when exceeding the nominal operating conditions, don't lead to material plasticity. Herein, the Smith-Watson-Topper (SWT)-N relationship reduces to the SN relationship, where the number of cycles to fatigue failure are expressed as a function of the cyclic stresses only.

Since fatigue failure is stress-drive, High-Cycle Fatigue (HCF) rules are applied, where the maximum principal stress represents the stress tensor. Basquin's equation is used to estimate the fatigue life, with the fatigue strength and exponent for the shaft equal to 440 MPa and -0.088, respectively.

For each load scenario, the simulation results are replicated to generate 3×10^4 hours of operation. The accumulated damage is estimated at the end of the operation using Miner's rule. The application of Miner's rule to this case is justified by the HCF damage, where the

material is loaded in the elastic regime, hence load sequence effects can be ignored. In the case of motor failure before completing 30 000 hours, its accumulated damage is capped at 1.

Figure 17.3 shows the distribution of the accumulated damage for each motor across the different operation scenarios. Motors 5, 6, 11 and 12 seem to be lightly damaged, despite of the overload conditions. On the other hand, motors 7, 8, 13 and 18 have all failed before completing 30 000 hours of operation. The other fleet motors experience a wider spread of damage accumulation, depending on the severity of the operation condition. The estimated accumulated damages are used as labels to train and test the hybrid model.

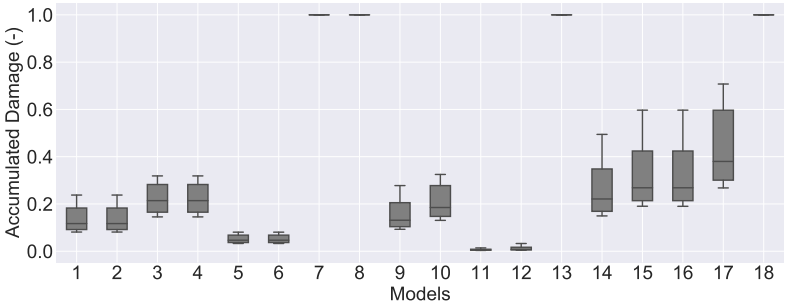


Figure 17.3.: Distribution of accumulated damage

17.2 Results

Data pre-processing

The very first step is to isolate a test (evaluation) set $\mathcal{E}$ from the pre-processing and training processes. Having 18 motors, each with 1331 operation scenarios, yields a dataset of 23 958 samples.
A conventional train-test split on the entire dataset would be inappropriate. In such a case, the Multi-Layer Perceptron (MLP) would be exposed to samples relating the operating conditions features $\mathbf{u}$ and catalogue data feature $\mathbf{c}$ to the degradation $\mathbf{y}$ for motors with physical degradation models $\mathcal{X}^{\mathsf{P}}$ as well as models with unknown degradation models $\mathcal{X}^{\phi}$. This would definitely lead to data leakage, since in practice, the degradation models in set $\mathcal{X}^{\phi}$ are not yet defined.

Alternatively, the train-test split is performed among the fleet motors using a *Grouped Split*, whereby each motor with its corresponding operation scenarios are treated as a *group*, and are splits as such. The train and test subset of motors represent 90 % and 10 % of the entire fleet, resulting in 16 models for training and 2 models for testing.

Having split the data, the feature extraction and basis selection procedures could be executed on the training set.

As explained in Section 13.2, an Extremely Randomized Trees Regressor (ERTR) is used to rank the feature importance with respect to the accumulated damage. This process is conducted on the sets of static and dynamic features separately.

For both feature sets, an ERTR with 100 randomized decision trees, each with a maximum depth of 50 is used. The features corresponding to the top 20 % importances are deemed *important*.

Figure 17.4 shows the static features importances. Five features out of 29 are deemed important; $LB, O1, J, m$ and *nrated*. $O1$ and LB are dimensions, representing the housing length and the distance from the driving end to the power electronics compartment. Typically, LB is about 65 to 75 % of $O1$, hence, the two features are obviously correlated. J and m are the polar moment of inertia and the mass, which are as well correlated, and *nrated* is the rated speed of the motor.

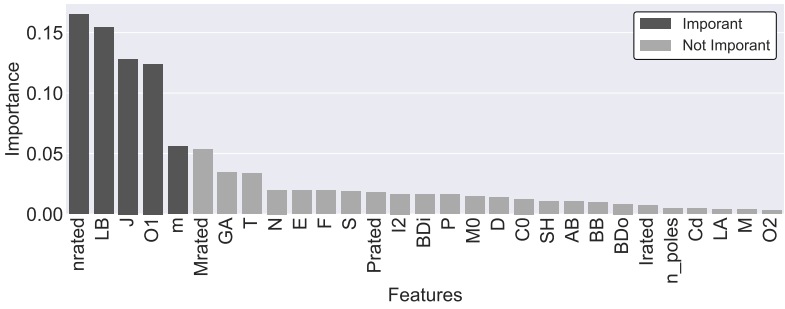


Figure 17.4.: Static features importance

Figure 17.5 shows the dynamic features importances. Due to the large number of features, only the highest 50 are shown. Out of 240 features, 48 are selected. It is clearly shown that the top important features are dominated by the signal *kurtosis* and *mean* values from different sensors in different directions. Spectral features are less significant in this case compared to temporal ones.

Looking more thoroughly at the important dynamic features, for each feature, the corresponding sensor position, measurement direction and signal statistic are identified. The importances corresponding to each sensor position, each measurement direction, and each signal statistic are summed up as shown in Figure 17.6.

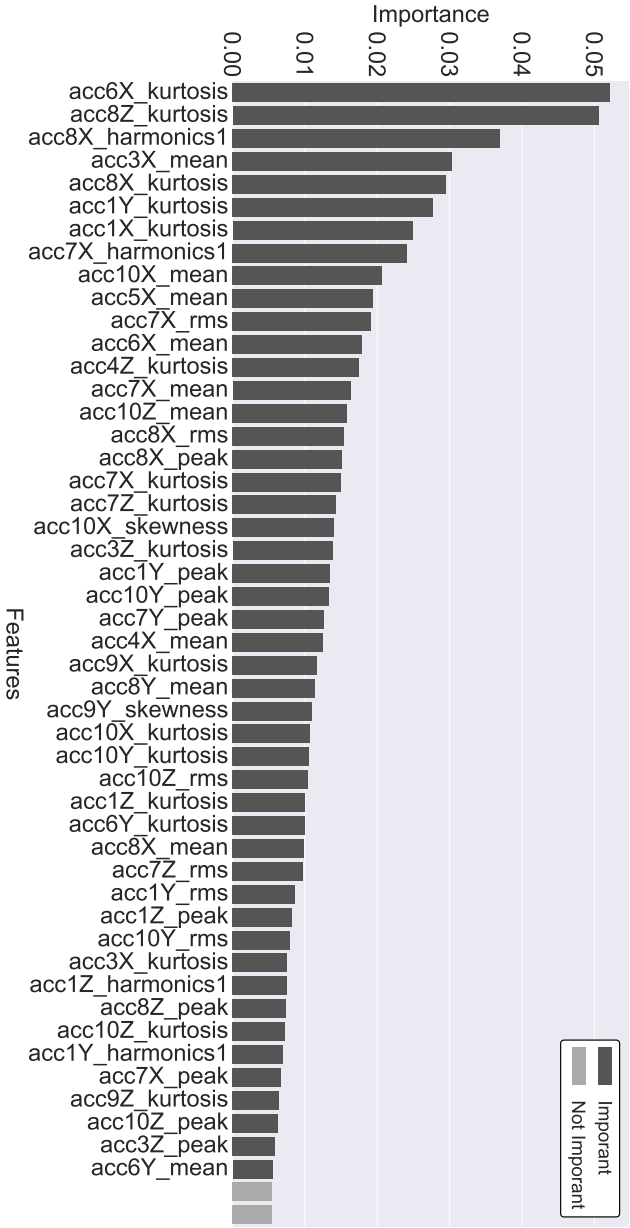


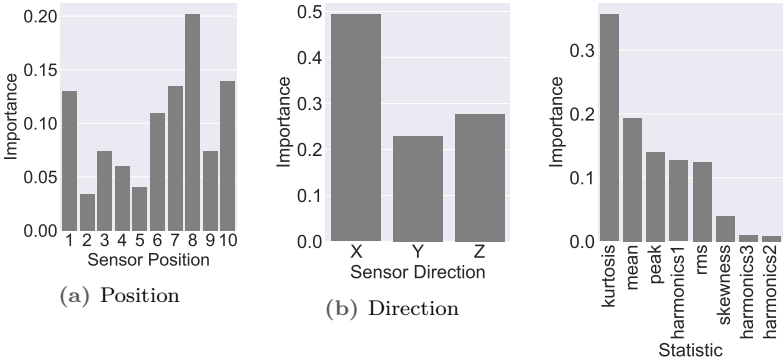
Figure 17.5.: Dynamic features importance

Looking at the sensor positions, the most frequently appearing positions among the top 50 features is *position* 8, located along the driving end flanges. Positions 1, 7 and 10 are ranked second, with the importance of position 6 being comprehensible. Position 1's importance is justified in similar manner as position 8, with position 8 being more important due to its proximity to the driving end, hence location of damage.

The damage introduced to the shaft is mainly dominated by stresses due to bending, hence normal stresses in the x-direction. This explains the dominance of features extracted from signals in the *x-direction*. Although torque results shear stresses in the material, its impact could be less significant on the damage, it being an operation condition a motor is designed for. On the other hand, the presence of misalignments is considered an operation anomaly, a load condition to which the motor is less tolerant.

The *kurtosis* represents the impulsive characteristic of the signal, being a measure of outliers. Impulses in the acceleration signal result from the superposition of the shaft's driving end's eccentricity (due to misalignment) and offset (due to the pre-load), leading to a cam-shaped profile motion of the driving end. On the other hand, the importance of the *mean* stems from being an indicator of the pre-load, which introduces mean effects to the cyclic stress signals.

The features with the highest importance are the kurtosis of the x acceleration signal at position 6 and of the z acceleration signal at position 8. This is justified by the vertical orientation of the motor. Position 6 corresponds to a section with relatively lower stiffness, due to the un-modelled electronic and control components. Hence, sensor 6 experiences the highest vibration impulses (hence signal kurtosis) due to axial gravitational loads. On the other hand, the proximity of position 8 to the damaged component, the radial nature of the cyclic and static loads (i.e. in the z-direction), along with the impulsive or eccentric load profile explains the importance of the kurtosis of acceleration z at position 8.



(a) Position (b) Direction (c) Statistic

Figure 17.6.: Dynamic feature importances breakdown

Having identified the important features, some features are suspected to be correlated. Hence their presence bears a computation cost and threatens the generalization of the model. A Spearman correlation is calculated for every two features, and features with correlation coefficients greater than 0.9 are dropped out, except for the feature among them with the largest importance (i.e. the feature most significant to the damage estimation).

Out of the 5 important static features, only *nrated* and *LB* are considered, as the other features are found correlated to either of them, as shown by the correlation matrix in Figure 17.7. Consequently, with respect to motor specifications, the damage is mainly influenced by the nominal rotation speed and the housing length.

Figure 17.8 shows the correlation matrix of important dynamic features, with the darker entries corresponding to high correlations. Several blocks of correlated features can be identified. After dropping out the correlated dynamic features, the number of dynamic features drops from 48 to 22.

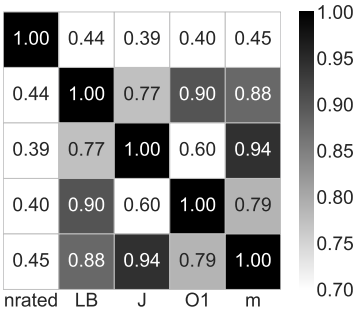


Figure 17.7.: Spearman correlation matrix of important static features

As shown in Figure 17.9, the breakdown of signal direction and signal static remains almost unchanged after dropping out correlated features, which comes in line with the justification of feature importance from a physical point of view. However, the distribution of *position* significance looks different. First, all features extracted from the signals at position 10 are dropped, and position 8 lost its significance dramatically, being now among the position with the fewest associated important feature. Since the signals at the flange are strongly correlated, both positions have lost their leading significance to position 7, which became the leading position with associated important features.

Coming back to the basis selection, the 2 uncorrelated static features deduced are used to calculate the relative distance function in Equation (13.25), upon which the basis models are chosen. The basis models are chosen and excluded from the training set resulting from the aforementioned split. Herein, the basis set holds 3 models, the training set size reduces to 13, and the evaluation set remains at 2 as is. This results in a typical 80-20 train-test split.

Figure 17.10 shows the scatter of the fleet assets in the space of extracted static features, indicating the basis, training and test sets.

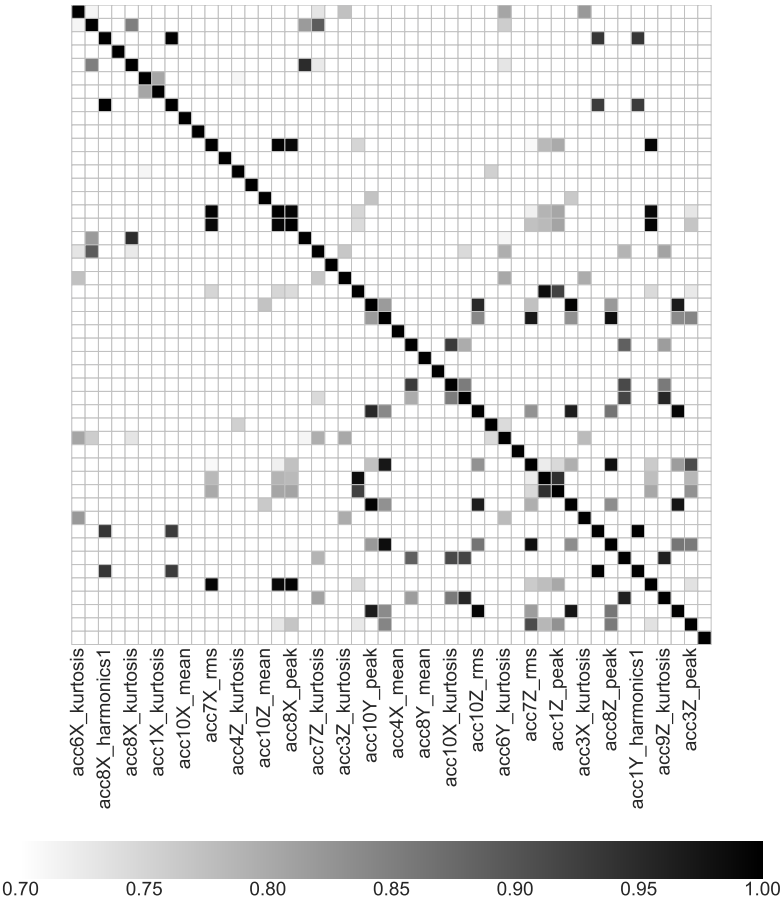
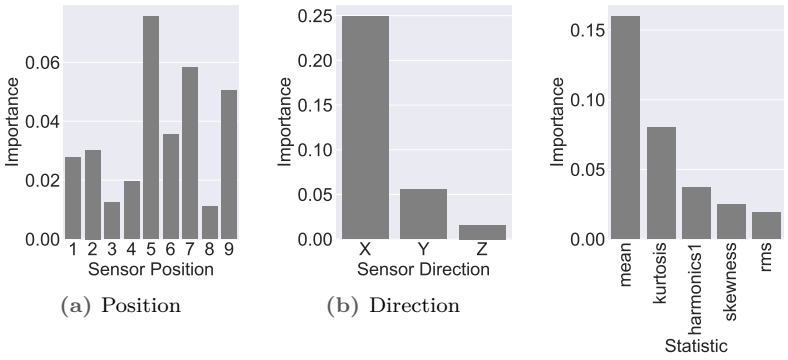


Figure 17.8.: Spearman correlation matrix of important dynamic features



(a) Position (b) Direction (c) Statistic

Figure 17.9.: Uncorrelated dynamic feature importances breakdown

As shown, the scatter of the basis models covers the entire space of reduced static features.

Figure 17.10.: Distribution of models in the static feature space

Pre-classification

An ensemble of Random Forests (RFs) with 50 estimators and a maximum depth of 100 splits is trained to classify the damage severity before predicting its value using *moSAIc*.

Figure 17.11 shows the confusion matrix of the classification process when performed on the test set, with the value normalized to the

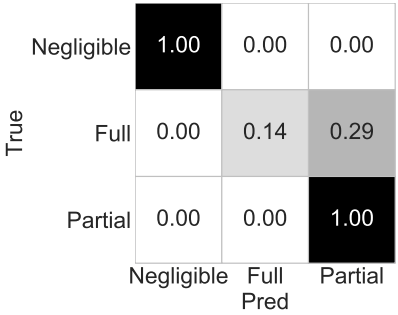


Figure 17.11.: Confusion matrix of the RF classifier

true count of each class. Both the fully and negligibly damaged samples were predicted with 100 % precision. On the other hand, only 78 % of the samples predicted partially damaged are in fact partially damaged, while the remaining 22 % are *falsely* predicted as fully damaged.

To this end, this result is satisfactory, since all the partially damaged samples could be 100 % *recalled*, whereas the presence of some false predictions in the training of the downstream MLP could be corrected by the hybrid model's regression task.

MLP setup

The hyper-parameters of the MLP are tuned using random search. Ideally, the training set is split into a training and a validation set to perform the parameter tuning. However, due to the small number of samples, such a split could impact the generalization of the trained model. This limitation is overcome by conducting a cross-validation procedure using 5 training-validation folds.

To accelerate the cross validation loops, only 10 epochs are performed by the optimizer. The mean and the 95 % confidence interval of the Mean Absolute Relative Error (MARE) evaluated on all epochs after 10 folds are shown in Figure 17.12.

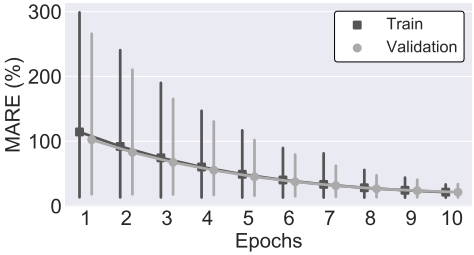


Figure 17.12.: Average loss with a 5-fold cross validation

To this end, with the tuned parameters, the model converges to the same mean MARE and standard deviation. The details of the hyper-parameter tuning using random search are omitted due to irrelevance, and only the concluded optimal hyper-parameters are summarized in Table 17.2.

Table 17.2.: Set of optimal MLP parameters, $\mathcal{P}_G^*$

Hidden Units	λ_1	λ_2	Activation	Learning rate
(40, 30, 20, 10)	5×10^{-3}	5	tanh	10^{-4}

The MLP is trained using the hyper-parameters in Table 17.2, and evaluated on the test set. In this case, the basis models are regarded as part of the test set, as the MLP has not learned the relationship between their inputs and weights during the training phase. Instead, they have only been regarded as expert models, to which the learned weights are assigned.

The training process of the model is shown Figure 17.13. After 20 epochs, the MAREs calculated on the training and test set evaluate to 14.9 % and 14.6 %. As shown, both sets converge to nearly the same value, showing confidence about the concluded hyper-parameters.

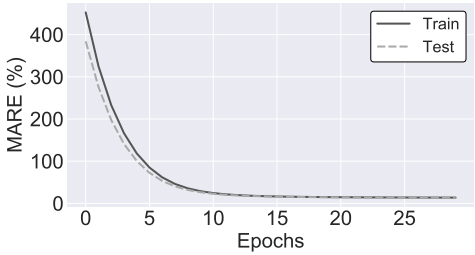


Figure 17.13.: MARE evolution over the training and test sets

The inclusion of the basis set doesn't only increase the size of the test set, but also increases the robustness of the model, in a sense that the model gets tested on motors with different degradation behaviour.

Figure 17.15(a) shows the distribution of the accumulated damage in each motor after 30 000 hours of operation. If the basis models are to be excluded, the model would have been evaluated only on models 10 and 12, which is a very unrepresentative sample of the degradation behaviour. For instance, the degradation of models 15 and 16 exhibiting partial damage with a large variance, or the models 7 and 8 exemplifying the extreme case of full damage are not in the test set. Hence, the performance of the model on these behaviour is not assessed. With the inclusion of the basis models, these behaviours are covered and the model is assessed on them.

Figure 17.15(b) shows the distribution of the Absolute Relative Error (ARE) of each motor model, evaluated only on the samples passed through the MLP. Half the models, whether in train, test or basis set, have mean error below 10 %. These very models, nonetheless, share in common that they are partially damaged. Contrarily, those negligibly damaged exhibit errors between 10 and 20 %, while the full damaged have the highest error spikes with errors going up to 40 %.

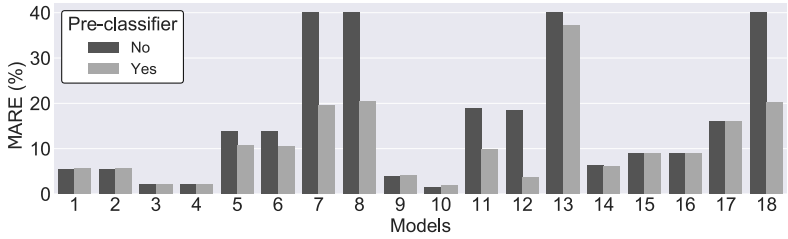
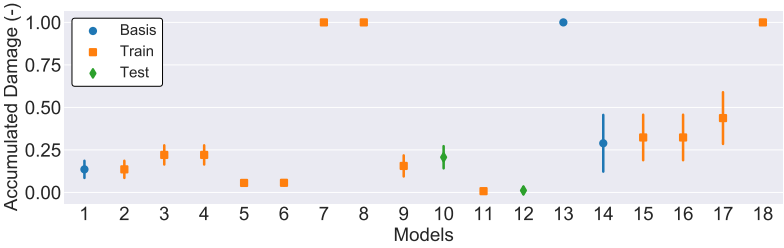


Figure 17.14.: Error comparison between *moSAIc* with and without pre-classifier

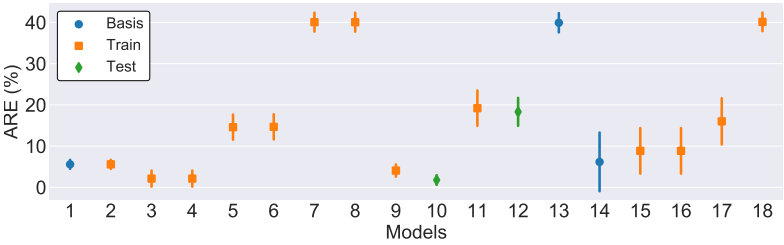
Accounting for the pre-classifications, and mapping the pre-classified severity to 0 and 1 for negligible and full damage, respectively, the distribution of the error changes to that shown in Figure 17.15(c). While the partially damaged models don't show much change in the error distribution, the error of the models with extreme damage severity drops to almost the half. This highlights the added benefit of the pre-classifier.

A final remark with respect to the large errors and variances associated with models 7, 8, 13 and 18: due to the high pre-classifier's precision, many of these models' samples get excluded from the training of the MLP, hence leading to insufficient data to learn from during the training phase. This reflects in the observed large errors. This problem, however, is not addressed in this work, and is recommended for further investigation.

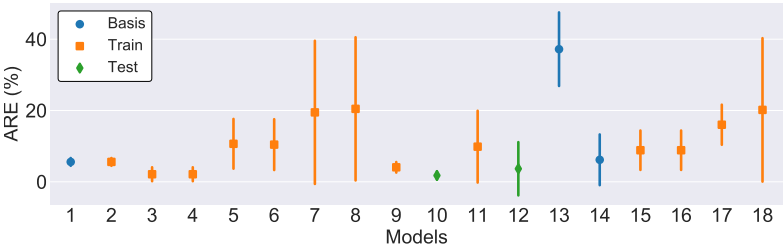
To highlight further the added value of the pre-classifier, a *moSAIc* model without the pre-classification step is trained on samples with all damage severities. The MARE associated to each motor model is shown in Figure 17.14, for both *moSAIc* configuration. As expected, the precision of the damage estimate for the extreme cases increases dramatically when inferred using a classifier rather than from a regression model.



(a) Accumulated Damage



(b) ARE - excluding pre-classification



(c) ARE - including pre-classification

Figure 17.15.: Distribution of ARE and accumulated damage for different models

Part VII

Epilogue

18 Summary of Contributions

In this dissertation, two state-of-the-art contributions were presented in the context of combining physics-based simulation models with machine learning in the pipeline of Predictive Maintenance (PdM), particularly the two modules: sensing and signal acquisition, and prognosis and damage estimation.

Acquisition and Sensing

The first contribution was an Optimal Sensor Placement (OSP) approach for configuring a network of multi-type sensors to improve the confidence in the damage estimates in complex industrial structures. The novelty herein lay in three aspects:

1. First, the formulation of the OSP approach incorporated the objectives of the downstream state and damage estimation within its objective function. The optimization problem formulation was based on a Kalman filter framework, that integrated the mathematical description of the structure, whose state were to be reconstructed from a limited number of sensors.
2. Second, the formulation of the problem constraints permitted optimizing a configuration of heterogeneous sensor types. In addition, the optimal sensor budget was not defined as a user input, unlike prior scientific work. On the contrary, a systematic method, based on simulating the system responses under the prospective operating conditions, was derived. From the simulation results, the minimum number of sensors required for a confident damage estimate could be a-priori inferred.

3. Third, the proposed approach is based on a convex optimization problem formulation. The gradient-based Method of Moving Asymptotes benefited from the physical description of the system and from the analytical form of the Jacobian to solve for the optimal sensor configuration. This guaranteed a scalable algorithm, that is ruggedly applicable to large and complex industrial structures.

The sensitivity of the approach towards varying its ingredients and constituents was rigorously investigated on two simple systems: a mass-spring system and a 2D truss structure. The proposed method outperformed consistently the sequential approach when applied on the same problem. It yielded a smaller damage estimation error as well as a lower covariance of the estimation error.

Furthermore, on two real-life industrial cases, the proposed approach showed very satisfactory precision. On both cases, the resulting state reconstruction and damage estimation errors were dominantly below 10 %, despite the restricted domain of candidate sensor positions.

Prognosis and Damage

The second contribution was a physics-data-based hybrid approach for degradation modelling and damage estimation within a fleet of industrial structures. The approach exploited the presence of elaborate physics-based degradation models of limited structures in the fleet to approximate the fatigue damage in other fleet structures. The novelty in this contribution lay in the following two aspects

1. First, this contribution is the first-of-a-kind in the PdM scientific work that enables the transferability of physics-based degradation models among fleet structures. This was achieved through weighted-mean filtering the physics-based damage models' outputs. The weights were assigned based on the similarity physics-based models express to the structure in question.

2. Second, the measure of similarity, and hence the weights of the physics-based damage models, were determined from a Multi-Layer Perceptron (MLP). The MLP suggested the weights based on the characteristics features of the structure, e.g. its material properties and geometric dimensions, as well as on its endurance sensitivity towards a broad set of operation conditions. This, in turn, extended the applicability of the hybrid model to nominal and extreme operating conditions.

This approach allowed precise and quantitative damage estimation in structures, for which a physics-based degradation model would have had been laborious or expensive to formulate. The hypothesized performance of the approach was corroborated by the rigorous study conducted on a fleet of cantilever beams. In this study, the sensitivity of the approach is assessed towards choosing the physics-based models, which were used to approximate other beams' degradation. Also, the performance of the model was evaluated with respect to the choice of the fleet structures' characteristics features and the choice of the operation conditions, upon which the weights of the physics-based models are assigned.
The precision of the approach was compared against other Machine Learning (ML) models, namely a MLP, an elastic-net regressor, a Support Vector Regression (SVR), and an ensemble of SVRs. In most of the cases, it demonstrated a surpassing performance, with some limitations remarked at the extreme cases of damage.

Additionally, the approach was applied to a fleet of 18 synthetic servomotors, with different geometric dimensions and operation ranges. In that use case, with the help of the 3 physics-based degradation models motors in the fleet, the damage in the rest of the fleet was calculated. The damage was predicted under an exhaustive spectrum of nominal and extensive operation conditions, and for the most part, the damage estimation error remained below $<20\,\%$.

19 Open Issues

The research presented in this thesis paves the road for many intriguing avenues of future work.

Acquisition and Sensing

Among the open issues to be investigated is the applicability of the proposed Optimal Sensor Placement (OSP) approach to non-linear systems. In this work, the discussion was limited to linear systems, where cyclic fatigue damage was to be estimated using standard Kalman filters. Generally, state estimation problems with unknown inputs involve using non-linear filters and estimators, e.g. extended Kalman filters, augmented Kalman filters, or particle filters. Additionally, other forms of damage, such as crack initiation and propagation result in non-linear system matrices, that must solved within filtering framework. The current formulation of the optimization problem doesn't account for the case of non-linear estimation and has to be adapted when considering such systems.

Further, filtering methods are not the only diagnostics methods. In fact, hybrid models, such as multi-fidelity models, are heavily applied at this stage of the Predictive Maintenance (PdM) pipeline. Multi-fidelity models combine limited, high-fidelity sensor data with abundant, low-fidelity simulation results for a better state reconstruction. Typically, the performance of these models is a function of the locations of sensors capturing the high-fidelity data. Given a physics-based simulation model is available, it could be utilized, with an adapted optimization problem formulation, to find the optimal configuration for such a hybrid diagnosis approach.

Another potential extension of this work is optimal sensor placement for multi-physics problems. [Ben+18] and [HR15] investigated sequential OSP for thermo-elastic problems. The robustness of the proposed method in this work could be applied to similar problems, whereby the estimated quantity is not only different from the measured one, but also doesn't belong to the same domain. Other multi-physics Structural Health Monitoring applications, to which the proposed approach is applicable, are fluid-structure interaction problems, e.g. in wind turbines, gas turbines, high rise buildings, dams and railway foundations. Herein, wind velocity, temperatures or water pressure are the quantities recorded by the sensors, which are used to estimate the presence of cracks or fatigue damage in the structure.

Prognosis and Damage

Regarding the damage estimation of fleet assets using physics-data-based hybrid modelling, one potential would be to investigate applying more rigorous gating networks instead of the Multi-Layer Perceptron (MLP) to estimate the weights assigned to the basis physics-based models. Classical machine learning regressors has shown outstanding performance compared to the proposed approach when applied to extreme cases. This sparks a flame of curiosity to investigate whether they could enhance the robustness of *moSAIc* if they substitute the MLP.

Additionally, for the discussion in this work, only cyclic fatigue damage models were considered when investigating Simulation Models and Artificial Intelligence Combined (*moSAIc*). Cyclic fatigue damage is not the only failure mechanism. Cracks, wear, and creep are as prevailing in industrial applications. When broadening the scope of research beyond metallic materials, a multitude of more sophisticated failure mechanisms associated with composites show up on the table.

Moreover, it was assumed that the appropriate degradation model of the structure is known a-priori. In the daily life of a maintenance engineer inspecting an engineering system, this doesn't go beyond an assumption associated with a level of confidence. Another alternative for *moSAIc* is to use it as an ensemble of candidate physics-based degradation models for the same structure, where the voting algorithm decides how they could be combined to obtain the most accurate damage model. Herein, the variability is not among fleet assets, but among different failure mechanisms of the same structure. Some industrial applications where such an approach could be reasonably applicable are gas turbines prone to simultaneous thermal- and mechanical-driven degradation. Also, this could be applicable to rotating machinery, where an interaction between faults in the electrical and the mechanical sub-systems are root causes for damage accumulation.

Part VIII

Appendices

A-Priori Identification of Optimal Number of Sensors

A

The paragraphs in this appendix summarize the methodology presented in [Kha+19a] for a-priori estimating the optimal number of sensors required for an accurate state estimation. The estimated number of sensors is used as a constraint for the optimization problem, presented in Equation (10.8)

Let $\mathbf{Y} = (\mathbf{y}_i)_{i=1}^{n_t} \in \mathbb{R}^{n_t \times n_o}$ be the matrix of output snapshots captured at n_t discrete time steps, obtained either through experimental measurements or a simulation of the system. From linear algebra, given a set of real-valued functions $\{\mathbf{y}_i | i \in \{1, \dots n_t\}\}$ on the interval $[t_0, t_f]$, the Gram matrix is given by the inner product on functions. One application of the Gramian is to evaluate the linear independence of the set of functions $\mathbf{y}_i$. The vectors $\mathbf{y}_i$ are linearly independent if the Gramian is non-singular. The Gramian of the matrix of outputs $\mathbf{Y}$ is given by

$$\mathbf{H} = \mathbf{Y}^\mathsf{T}\mathbf{Y} \in \mathbb{R}^{n_o \times n_o}. \tag{A.1}$$

Defining d to be the rank of $\mathbf{H}$, the column space spanning $\mathbf{H}$ can be represented in terms of d linearly-independent columns contained in $\mathbf{U}_f \in \mathbb{R}^{n_o \times d}$ [KV10]. Singular Value Decomposition (SVD) guarantees the existence of real, positive eigenvalues $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d \overset{!}{\geq} 0$, and two sets of orthogonal basis vectors $\mathbf{U}_f \in \mathbb{R}^{n_o \times d}$ and $\mathbf{Z}_f \in \mathbb{R}^{n_o \times d}$ such that

$$\mathbf{H} = \mathbf{U}_f \, \Sigma_f \, \mathbf{Z}_f^\mathsf{T}; \quad \Sigma_f = \operatorname{diag}(\sigma_1, \dots, \sigma_d) \in \mathbb{R}^{d \times d} \tag{A.2}$$

According to [KV10], for a matrix $\mathbf{H}$, having orthogonal basis vectors $\mathbf{U}_f$ and $\mathbf{Z}_f$, there exists a reduced sub-space represented by fewer basis vectors, $\mathbf{U}_r$ and $\mathbf{Z}_r$, that can express the information contained in $\mathbf{H}$. The components of the reduced sub-space are given by

$$\begin{aligned} \mathbf{U}_{r,\,ij} &= \mathbf{U}_{f,\,ij} \quad \text{for } i \in \{1,\ldots,n_o\}\ ,\ j \in \{1,\ldots,r\}, \\ \mathbf{Z}_{r,\,ij} &= \mathbf{Z}_{f,\,ij} \quad \text{for } i \in \{1,\ldots,n_o\}\ ,\ j \in \{1,\ldots,r\}, \end{aligned} \tag{A.3}$$

where $1 \leq r \leq d$ is defined such that

$$1 - \frac{\sum_{i=1}^{r} \sigma_i}{\sum_{i=1}^{d} \sigma_i} \leq \epsilon_0, \tag{A.4}$$

and ϵ_0 is a pre-defined tolerance for the reduction.

In relation to the optimal sensor placement problem, the number of sensors to capture the dynamics of the system represented by r linearly-independent basis vectors is given by $n_s = r$. The inner product $\mathbf{Y}^{\mathsf{T}}\mathbf{Y}$ used to calculate $\mathbf{H}$ is proven by [PK96] to be equivalent to the Fisher Information Matrix, which represents the estimation covariance of the set of output responses. Accordingly, the Gram matrix $\mathbf{H}$, as well as its reduced-order representation, inherently reflect the estimation error covariance. This concludes that the optimal number of sensors n_s^* required to minimize the estimation error covariance $\mathbf{P}_k^{x+}$ in Equation (10.2) equals to the number of reduced basis vectors r according to Equation (A.4).
It needs to be noted that the choice of r depends dramatically on the choice of the reduction tolerance ϵ_0. In the literature of model-order reduction, e.g. [KV10], it is common to choose the value of ϵ_0 according to the application in hand. To our knowledge, there exists no clear-cut rule of thumb according to which this tolerance value is set.

Jacobian of Trace of A-Posteriori Estimation Covariance Matrix

B

The paragraphs in this appendix present the detailed derivation of the first term of the Jacobian of the objective function, $Tr\left(\mathbf{P}_k^{+}(\tilde{\boldsymbol{\beta}})\right)$. .

The chain rule could be used to find the derivative of the trace of the state estimation error covariance, $\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)$, with respect to the *relaxed* selection matrix $\boldsymbol{\beta}$, as follows:

$$\frac{\partial \,\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)}{\partial \boldsymbol{\beta}} = \left(\frac{\partial \,\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)}{\partial \beta_j}\right)_{j=1}^{n_o} \in \mathbb{R}^{n_o} \tag{B.1}$$

where the j^{th} component of the Jacobian vector is

$$\frac{\partial \,\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)}{\partial \beta_j} = \underbrace{\frac{\partial \,\mathrm{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)}{\partial \tilde{\beta}_i}}_{\text{Term I}} \underbrace{\frac{\partial \tilde{\beta}_i}{\partial \beta_j}}_{\text{Term II}} . \tag{B.2}$$

The evaluation of Term II is rather simple. From Equation (10.17), the mapping function of the relaxed variable β_j to the binary variable $\tilde{\beta}_i$ is written as

$$\tilde{\beta}_i(\beta_j) = \delta_{ij}\Psi\left(\beta_j \mid \bar{\beta}, \sigma^2\right) \implies \frac{\partial \tilde{\beta}_i(\beta_j)}{\partial \tilde{\beta}_j} = \delta_{ij}\psi\left(\beta_j \mid \bar{\beta}, \sigma^2\right), \tag{B.3}$$

where Ψ and ψ are the Cumulative Density Function (CDF) and Probability Density Function (PDF) of the Gaussian distribution $\mathcal{N}(\bar{\beta},\ \sigma^2)$, respectively, and δ_{ij} is the Kronecker delta.

Term I requires a more elaborate step-by-step derivation.

First, exploiting the linear property of the trace operator, which allows commuting it with the partial derivative operator, yields

$$\frac{\partial \operatorname{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)}{\partial \tilde{\beta}_i} = \operatorname{Tr}\left(\frac{\partial \mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\beta}_i}\right) \tag{B.4}$$

Starting from Equation (10.2), the error covariance matrix is formulated as a function of the argument variable $\tilde{\beta}_i$ as

$$\mathbf{P}_k^{x+}(\tilde{\beta}_i) = \left(\ \mathbf{I} - \mathbf{G}_k(\tilde{\beta}_i)\ \mathbf{C}(\tilde{\beta}_i)\ \right)\mathbf{P}_k^{x-}(\tilde{\beta}_i), \tag{B.5}$$

where the matrices $\mathbf{G}_k(\tilde{\beta}_i),\ \mathbf{C}(\tilde{\beta}_i),\ \mathbf{P}_k^{x-}(\tilde{\beta}_i)$ are given by:

$$\mathbf{G}_k(\tilde{\beta}_i) = \mathbf{P}_k^{x-}(\tilde{\beta}_i)\mathbf{C}(\tilde{\beta}_i)^{\mathsf{T}}\left(\underbrace{\mathbf{R}(\tilde{\beta}_i) + \mathbf{C}(\tilde{\beta}_i)\mathbf{P}_k^{x-}(\tilde{\beta}_i)\mathbf{C}(\tilde{\beta}_i)^{\mathsf{T}}}_{\tilde{\mathbf{R}}(\tilde{\beta}_i)}\right)^{-1}$$

$$\mathbf{P}_k^{x-}(\tilde{\beta}_i) = \mathbf{A}\mathbf{P}_{k-1}^{x+}(\tilde{\beta}_i)\mathbf{A}^{\mathsf{T}} + \mathbf{Q}, \tag{B.6}$$

$$\mathbf{C}(\tilde{\beta}_i) = \tilde{\beta}_i \mathbf{C}_{0,\ i*},$$

$$\mathbf{R}(\tilde{\beta}_i) = \tilde{\beta}_i^2 \mathbf{R}_{0,\ i*}.$$

The matrices $\mathbf{C}_0$ and $\mathbf{R}_0$ refer to the original observation and observation covariance matrices with all n_o system outputs being observed, as defined in Equations (8.8) and (8.9). The subscripts $(\cdot)_{i*}$ / $(\cdot)_{*i}$ are used to refer to the i^{th} row / column of a matrix, respectively.

For the rest of the derivation, the argument $(\tilde{\beta}_i)$ is dropped out for the simplicity of the notation.

Applying the product differentiation rule to Equation (B.5), the derivative of the estimation error covariance is given by:

$$\begin{aligned}\frac{\partial \mathbf{P}_k^{\times+}}{\partial \tilde{\beta}_i} &= (\ \mathbf{I} - \mathbf{G}_k\ \mathbf{C}\ ) \frac{\partial \mathbf{P}_k^{\times-}}{\partial \tilde{\beta}_i} \ + \ \frac{\partial (\ \mathbf{I} - \mathbf{G}_k\ \mathbf{C}\ )}{\partial \tilde{\beta}_i} \mathbf{P}_k^{\times-} \\ &= (\ \mathbf{I} - \mathbf{G}_k\ \mathbf{C}\ ) \frac{\partial \mathbf{P}_k^{\times-}}{\partial \tilde{\beta}_i} \ - \ \frac{\partial (\mathbf{G}_k\ \mathbf{C}\ )}{\partial \tilde{\beta}_i} \mathbf{P}_k^{\times-} \\ &= (\ \mathbf{I} - \mathbf{G}_k\ \mathbf{C}\ ) \frac{\partial \mathbf{P}_k^{\times-}}{\partial \tilde{\beta}_i} \ - \ \left(\frac{\partial \mathbf{G}_k}{\partial \tilde{\beta}_i}\ \mathbf{C} + \mathbf{G}_k\ \frac{\partial \mathbf{C}}{\partial \tilde{\beta}_i} \right) \mathbf{P}_k^{\times-}.\end{aligned} \tag{B.7}$$

Looking at the individual matrices separately, the derivative of the a-priori estimation error covariance, $\mathbf{P}_k^{\times-}$, could be written as:

$$\begin{aligned}\frac{\partial \mathbf{P}_k^{\times-}}{\partial \tilde{\beta}_i} &= \frac{\partial \left(\mathbf{A} \mathbf{P}_{k-1}^{\times+} \mathbf{A}^\mathsf{T} + \mathbf{Q} \right)}{\partial \tilde{\beta}_i} = \mathbf{A} \frac{\partial \mathbf{P}_{k-1}^{\times+}}{\partial \tilde{\beta}_i} \mathbf{A}^\mathsf{T} \\ &= \begin{cases} \mathbf{A} \dfrac{\partial \mathbf{P}^{\times 0}}{\partial \tilde{\beta}_i} \mathbf{A}^\mathsf{T} = \mathbf{0}, & \text{if } k = 1 \\ \mathbf{A} \dfrac{\partial \mathbf{P}_{k-1}^{\times+}}{\partial \tilde{\beta}_i} \mathbf{A}^\mathsf{T}, & \text{if } k > 1 \end{cases}.\end{aligned} \tag{B.8}$$

Clearly, $\dfrac{\partial \mathbf{P}_k^{\times-}}{\partial \tilde{\beta}_i}$ is dependent on $\dfrac{\partial \mathbf{P}_{k-1}^{\times+}}{\partial \tilde{\beta}_i}$ evaluated at the previous Kalman filter step. For $k = 1$, $\mathbf{P}_{k-1}^{\times+} = \mathbf{P}^{\times 0}$, which is the initial state covariance of the problem, and is independent of $\tilde{\beta}_i$. Hence, the derivative vanishes for the first step. For the following steps, the derivative could be calculated. This could be comprehended by the fact that the Kalman filter intrinsically attempts to minimize the covariance in the estimated joint probability of the states based on the array of measurements updates. Accordingly, the a-posteriori covariance is effectively a function of the measurements array, and thus their positions in the system. For the first step, however, the covariance is effectively an assumption on the degree of uncertainty

in the initial state estimate, $\mathbf{x}_0$, thus remains uninfluenced by the choice of system output positions.

Due to recursive procedure of the filter, which results in the dependence of the current state estimation covariance on the previous one, the derivative of the covariance matrices propagates in a similar recursive fashion; i.e.

$$\frac{\partial \mathbf{P}_k^{x-}}{\partial \tilde{\beta}_i} = f^{\partial \mathbf{P}^{x-}}\left(\frac{\partial \mathbf{P}_{k-1}^{x+}}{\partial \tilde{\beta}_i}\right) = f^{\overrightarrow{\partial \mathbf{P}^{x-}}}\left(\frac{\partial \mathbf{P}_2^{x+}}{\partial \tilde{\beta}_i}\right) \tag{B.9}$$

where $f^{\partial \mathbf{P}^{x-}}$ is a function performing the steps shown in Equation (B.8), and $f^{\overrightarrow{\partial \mathbf{P}^{x-}}}$ indicates the recursive call of the function.

Nonetheless, since the filter estimate at step k requires only knowledge from the current step's measurements and the previous steps covariance, so does the derivatives of the covariance matrix. In other words, the derivative $\dfrac{\partial \mathbf{P}^{x+}}{\partial \tilde{\beta}_i}$ should only be stored for the $(k-1)^{th}$ step, and recursively updated as the filter progresses.

The derivatives of the observation and observation covariance matrices $\mathbf{C}$ and $\mathbf{R}$ with respect to $\tilde{\beta}_i$ are given as

$$\frac{\partial \mathbf{C}}{\partial \tilde{\beta}_i} = \begin{bmatrix} \mathbf{0} \\ \mathbf{C}_{0,\, i*} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n_o \times n_x}, \qquad \frac{\partial \mathbf{R}}{\partial \tilde{\beta}_i} = 2\tilde{\beta}_i \begin{bmatrix} \mathbf{0} \\ \mathbf{R}_{0,\, i*} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n_o \times n_o}, \tag{B.10}$$

respectively, where the derivatives are matrices holding the elements of the i^{th} row of the original matrix and having zeros elsewhere.

The derivative of the Kalman filter gain, $\mathbf{G}$, is given by applying the product rule, as follows:

$$\begin{aligned}\frac{\partial \mathbf{G}_k}{\partial \tilde{\beta}_i} &= \frac{\partial \left(\mathbf{P}_k^{\text{x-}} \mathbf{C}^\mathsf{T} \tilde{\mathbf{R}}^{-1}\right)}{\partial \tilde{\beta}_i} \\ &= \underbrace{\frac{\partial \mathbf{P}_k^{\text{x-}}}{\partial \tilde{\beta}_i}}_{Eq.\ (B.8)} \mathbf{C}^\mathsf{T} \tilde{\mathbf{R}}^{-1} + \mathbf{P}_k^{\text{x-}} \underbrace{\left(\frac{\partial \mathbf{C}}{\partial \tilde{\beta}_i}\right)^\mathsf{T}}_{Eq.\ (B.10)} \tilde{\mathbf{R}}^{-1} + \mathbf{P}_k^{\text{x-}} \mathbf{C}^\mathsf{T} \frac{\partial \tilde{\mathbf{R}}^{-1}}{\partial \tilde{\beta}_i}\end{aligned} \tag{B.11}$$

The term $\tilde{\mathbf{R}}$, introduced in Equation (B.6), is differentiated by applying the chain rule on the inverse operator. Recalling the definition of the derivative of an inverse operator, the term $\frac{\partial \tilde{\mathbf{R}}^{-1}}{\partial \tilde{\beta}_i}$ is given by

$$\frac{\partial \tilde{\mathbf{R}}^{-1}}{\partial \tilde{\beta}_i} = -\tilde{\mathbf{R}}^{-1} \frac{\partial \tilde{\mathbf{R}}}{\partial \tilde{\beta}_i} \tilde{\mathbf{R}}^{-1}. \tag{B.12}$$

The derivative of the matrix, $\tilde{\mathbf{R}}$, is given by recalling the definition in Equation (B.6), as follows

$$\begin{aligned}\frac{\partial \tilde{\mathbf{R}}}{\partial \tilde{\beta}_i} &= \frac{\partial \left(\mathbf{R} + \mathbf{C} \mathbf{P}_k^{\text{x-}} \mathbf{C}^\mathsf{T}\right)}{\partial \tilde{\beta}_i} \\ &= \underbrace{\frac{\partial \mathbf{R}}{\partial \tilde{\beta}_i}}_{Eq.\ (B.10)} \\ &+ \left(\underbrace{\frac{\partial \mathbf{C}}{\partial \tilde{\beta}_i}}_{Eq.\ (B.10)} \mathbf{P}_k^{\text{x-}} \mathbf{C}^\mathsf{T} + \mathbf{C} \underbrace{\frac{\partial \mathbf{P}_k^{\text{x-}}}{\partial \tilde{\beta}_i}}_{Eq.\ (B.8)} \mathbf{C}^\mathsf{T} + \mathbf{C} \mathbf{P}_k^{\text{x-}} \underbrace{\left(\frac{\partial \mathbf{C}}{\partial \tilde{\beta}_i}\right)^\mathsf{T}}_{Eq.\ (B.10)} \right)\end{aligned} \tag{B.13}$$

Substituting Equation (B.13) in Equation (B.12) to obtain $\frac{\partial \tilde{\mathbf{R}}^{-1}}{\partial \tilde{\beta}_i}$, then substituting Equation (B.12) back into Equation (B.11) to obtain $\frac{\partial \mathbf{G}_k}{\partial \tilde{\beta}_i}$, and finally substituting Equations (B.8), (B.10) and (B.11) into Equation (B.7), the derivative matrix $\frac{\partial \mathbf{P}_k^{x+}}{\partial \tilde{\beta}_i}$ could be evaluated. The trace of $\frac{\partial \mathbf{P}_k^{x+}}{\partial \tilde{\beta}_i}$ is evaluated to obtain the derivative of the objective function with respect to β_i as per Equation (B.4).

To calculate the Jacobian vector, $\frac{\partial \operatorname{Tr}\left(\mathbf{P}_k^{x+}(\tilde{\boldsymbol{\beta}})\right)}{\partial \boldsymbol{\beta}}$, Equation (B.2) is then evaluated for all $\beta_j \in \boldsymbol{\beta}$.

Assembly Procedure for the General Case of *moSAIc*

C

In the paragraphs of this appendix, the assembly procedure of the general case of moSAIc explained in Section 13.2 is elaborated on a minimal example attempting to estimate the damage in a roller bearing.

Consider a fleet of roller bearings, where Simulation Models and Artificial Intelligence Combined (*moSAIc*) is applied to estimate the damage of an unknown asset. For simplicity, we consider an architecture with $\mathcal{X}^{\mathsf{B}} = \left\{x^{\mathsf{B}}{}_1, x^{\mathsf{B}}{}_2\right\}$, and given that each model may yield up to four damage estimates, $(y^{\mathsf{B}}{}_1)^{(\mathsf{r})}, (y^{\mathsf{B}}{}_1)^{(\mathsf{i})}, (y^{\mathsf{B}}{}_1)^{(\mathsf{o})}$,and $(y^{\mathsf{B}}{}_1)^{(\mathsf{c})}$, corresponding to the most damaged roller, the inner race, the outer race, and the cage, respectively. The alphabetic characters are used in the example for the sake of clarity, but typically, numeric indexing is used.

Let

$$\mathbf{y}^{\mathsf{B}}{}_1 = \left[(y^{\mathsf{B}}{}_1)^{(\mathsf{r})}, (y^{\mathsf{B}}{}_1)^{(\mathsf{o})}\right]^{\mathsf{T}}, \quad \mathbf{y}^{\mathsf{B}}{}_2 = \left[(y^{\mathsf{B}}{}_2)^{(\mathsf{r})}, (y^{\mathsf{B}}{}_2)^{(\mathsf{i})}\right]^{\mathsf{T}}.$$

Correspondingly,

$$\boldsymbol{\alpha}^{\mathsf{M}}{}_1 = \left[(\alpha^{\mathsf{M}}{}_1)^{(\mathsf{r})}, (\alpha^{\mathsf{M}}{}_1)^{(\mathsf{o})}\right]^{\mathsf{T}}, \quad \boldsymbol{\alpha}^{\mathsf{M}}{}_2 = \left[(\alpha^{\mathsf{M}}{}_2)^{(\mathsf{r})}, (\alpha^{\mathsf{M}}{}_2)^{(\mathsf{i})}\right]^{\mathsf{T}}$$

Here, the union of indexes is given by

$$\{(r), (o)\} \cup \{(r), (i)\} = \{(r), (i), (o)\},$$

and has a cardinality of 3. Therefore, $\tilde{n_e} = 3$, which is the total number of distinct damage estimates yielded by the basis models.

The mapping function is defined as $f^{\tilde{a}} : \{(r), (i), (o)\} \rightarrow \{1, 2, 3\}$. Applying Equation (13.15) to the first basis model's outputs and weights, we observe that for both $\tilde{e} = 1$ and $\tilde{e} = 3$, the correspondents $(y^{B_1})^{(r)}, (y^{B_1})^{(o)} \in \mathbf{y}^{B_1}$, in contrast to $(y^{B_1})^{(i)} \notin \mathbf{y}^{B_1}$. Similarly, $(\alpha^{M_1})^{(r)}, (\alpha^{M_1})^{(o)} \in \boldsymbol{\alpha}^{M_1}$, while $(\alpha^{M_1})^{(i)} \notin \boldsymbol{\alpha}^{M_1}$.

Hence, the assembled vectors corresponding to x^{B_1} are written as:

$$\tilde{\mathbf{y}}^{B_1} = \begin{bmatrix} (y^{B_1})^{(r)} \\ 0 \\ (y^{B_1})^{(o)} \end{bmatrix}, \quad \tilde{\boldsymbol{\alpha}}^{M_1} = \begin{bmatrix} (\alpha^{M_1})^{(r)} \\ 0 \\ (\alpha^{M_1})^{(o)} \end{bmatrix}.$$

Equivalently, the mapping function is applied to vectors of x^{B_2}, yielding:

$$\tilde{\mathbf{y}}^{B_2} = \begin{bmatrix} (y^{B_2})^{(r)} \\ (y^{B_2})^{(i)} \\ 0 \end{bmatrix}, \quad \tilde{\boldsymbol{\alpha}}^{M_2} = \begin{bmatrix} (\alpha^{M_2})^{(r)} \\ (\alpha^{M_2})^{(i)} \\ 0 \end{bmatrix}.$$

Finally, substituting the yielded consistent vectors into Equation (13.17), yields

$$\begin{aligned} \tilde{\mathbf{y}}^{M} &= < \tilde{\boldsymbol{\alpha}}^{M_1}, \tilde{\mathbf{y}}^{B_1} > + < \tilde{\boldsymbol{\alpha}}^{M_2}, \tilde{\mathbf{y}}^{B_2} > \\ &= \begin{bmatrix} (\alpha^{M_1})^{(r)}(y^{B_1})^{(r)} + (\alpha^{M_2})^{(r)}(y^{B_2})^{(r)} \\ 0 + (\alpha^{M_2})^{(i)}(y^{B_2})^{(i)} \\ (\alpha^{M_1})^{(o)}(y^{B_1})^{(o)} + 0 \end{bmatrix}. \end{aligned}$$

Effectively, the hybrid model's estimate of the roller element's damage is the combined damage of x^{B_1} and x^{B_2}, weighted by the vectors $\tilde{\boldsymbol{\alpha}}^{M_1}$ and $\tilde{\boldsymbol{\alpha}}^{M_2}$. Due to the non-comprehensiveness of the basis models in estimating the damage of the inner and outer rings, the estimate of both quantities is attributed solely to x^{B_1} and x^{B_2}, respectively, without a contribution from the other basis model.

Part IX

Bibliography

List of References

[Afu04] Allan Afuah. *Business models: A strategic management approach*. McGraw-Hill/Irwin, 2004 (cit. on p. 10).

[AG03] FAG Kugelfischer AG. "Rolling Bearing Damage - Recognition of damage and bearing inspection". In: *Publ. No. WL* 82.102 (2003), p. 2 (cit. on pp. 34, 156).

[AH75] Mohamed Abdel-Hameed. "A gamma wear process". In: *IEEE transactions on Reliability* 24.2 (1975), pp. 152–153 (cit. on p. 61).

[AHL01] Ameen Abu-Hanna and Peter JF Lucas. "Prognostic models in medicine". In: *Methods of information in medicine* 40.01 (2001), pp. 1–5 (cit. on p. 7).

[Ahm+18] Wasim Ahmad, Sheraz Ali Khan, MM Manjurul Islam, and Jong-Myon Kim. "A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models". In: *Reliability Engineering & System Safety* (2018) (cit. on pp. 35, 60).

[AKC15] Dawn An, Nam H Kim, and Joo-Ho Choi. "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews". In: *Reliability Engineering & System Safety* 133 (2015), pp. 223–236 (cit. on p. 68).

[Amz+94] C Amzallag, JP Gerey, JL Robert, and J Bahuaud. "Standardization of the rainflow counting method for fatigue analysis". In: *International journal of fatigue* 16.4 (1994), pp. 287–293 (cit. on pp. 58, 76).

[And17] Ted L Anderson. *Fracture mechanics: fundamentals and applications*. CRC press, 2017 (cit. on p. 57).

[Ara16] Nagaraj K Arakere. “Gigacycle rolling contact fatigue of bearing steels: A review”. In: *International Journal of Fatigue* 93 (2016), pp. 238–249 (cit. on p. 55).

[AS18] Isaac Animah and Mahmood Shafiee. “Condition assessment, remaining useful life prediction and life extension decision making for offshore oil and gas assets”. In: *Journal of Loss Prevention in the Process Industries* 53 (2018), pp. 17–28 (cit. on p. 35).

[Asc11] Georges Asch. *Acquisition de données-3e éd.: Du capteur à l'ordinateur*. Dunod, 2011 (cit. on p. 88).

[AWN18] HOA Ahmed, MLD Wong, and AK Nandi. “Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features”. In: *Mechanical Systems and Signal Processing* 99 (2018), pp. 459–477 (cit. on p. 35).

[BAA09] SM Beden, Shahrum Abdullah, and AK Ariffin. “Review of fatigue crack propagation models for metallic components”. In: *European Journal of Scientific Research* 28.3 (2009), pp. 364–397 (cit. on p. 58).

[Bak11] Pelin Gündeş Bakir. “Evaluation of optimal sensor placement techniques for parameter identification in buildings”. In: *Mathematical and Computational Applications* 16.2 (2011), pp. 456–466 (cit. on p. 48).

[Bap+18] Marcia Baptista, Elsa MP Henriques, Ivo P de Medeiros, et al. “Remaining useful life estimation in aeronautics: Combining data-driven and Kalman filtering”. In: *Reliability Engineering & System Safety* (2018) (cit. on p. 64).

[Bar+13] Piero Baraldi, Michele Compare, Sergio Sauco, and Enrico Zio. “Ensemble neural network-based particle filtering for prognostics”. In: *Mechanical Systems and Signal Processing* 41.1-2 (2013), pp. 288–300 (cit. on p. 68).

[Bar+14] Piero Baraldi, Enrico Zio, Francesca Mangili, Giulio Gola, and Bent H Nystad. “Ensemble of kernel regression models for assessing the health state of choke valves in offshore oil platforms”. In: *International Journal of Computational Intelligence Systems* 7.2 (2014), pp. 225–241 (cit. on p. 67).

[Bar+17] Alberto Barontini, Maria-Giovanna Masciotta, Luís F Ramos, Paulo Amado-Mendes, and Paulo B Lourenço. "An overview on nature-inspired optimization algorithms for Structural Health Monitoring of historical buildings". In: *Procedia engineering* 199 (2017), pp. 3320–3325 (cit. on pp. 44–46).

[Bat06] Claude Bathias. "An engineering point of view about fatigue of polymer matrix composite materials". In: *International journal of fatigue* 28.10 (2006), pp. 1094–1099 (cit. on p. 50).

[Bat99] Claude Bathias. "There is no infinite fatigue life in metallic materials". In: *Fatigue & fracture of engineering materials & structures* 22.7 (1999), pp. 559–565 (cit. on p. 54).

[Bay+08] E Bayraktar, N Isac, K Bessri, and C Bathias. "Damage mechanisms in natural (NR) and synthetic rubber (SBR): nucleation, growth and instability of the cavitation". In: *Fatigue & Fracture of Engineering Materials & Structures* 31.2 (2008), pp. 184–196 (cit. on p. 55).

[BCT13] D Benasciutti, A Cristofori, and R Tovo. "Analogies between spectral methods and multiaxial criteria in fatigue damage evaluation". In: *Probabilistic Engineering Mechanics* 31 (2013), pp. 39–45 (cit. on p. 59).

[Ben+15] S Benkabouche, H Guechichi, A Amrouche, and M Benkhettab. "A modified nonlinear fatigue damage accumulation model under multiaxial variable amplitude loading". In: *International Journal of Mechanical Sciences* 100 (2015), pp. 180–194 (cit. on p. 56).

[Ben+18] Peter Benner, Roland Herzog, Norman Lang, Ilka Riedel, and Jens Saak. "Comparison of model order reduction methods for optimal sensor placement for thermo-elastic models". In: *Engineering Optimization* (2018), pp. 1–19 (cit. on pp. 41, 45, 254).

[Ber+19] Christoph Bergs, Mohamed Khalil, Marcel Hildebrandt, et al. "Health indication of electric motors using a hybrid modeling approach". In: *Teschnische Messen* (2019) (cit. on p. 160).

[Bes10] Jacques Besson. "Continuum models of ductile fracture: a review". In: *International Journal of Damage Mechanics* 19.1 (2010), pp. 3–52 (cit. on p. 58).

[Bez+07] Abderrezak Bezazi, S Gareth Pierce, Keith Worden, and El Hadi Harkati. “Fatigue life prediction of sandwich composite materials under flexural tests using a Bayesian trained artificial neural network”. In: *International Journal of Fatigue* 29.4 (2007), pp. 738–747 (cit. on p. 62).

[BG12] Heinz P Bloch and Fred K Geitner. *Machinery Failure Analysis and Troubleshooting: Practical Machinery Management for Process Plants*. Butterworth-Heinemann, 2012 (cit. on pp. 31, 36).

[BHR18] Stefan Boschert, Christoph Heinrich, and Roland Rosen. “Next generation digital Twin”. In: *Proceedings of TMCE, Las Palmas de Gran Canaria, Spain Edited by: Horvath I., Suarez Rivero JP and Hernandez Castellano PM* (2018) (cit. on p. 13).

[Bir13] Alessandro Birolini. *Reliability engineering: theory and practice*. Springer Science & Business Media, 2013 (cit. on p. 5).

[Bis+13] Anupam Biswas, KK Mishra, Shailesh Tiwari, and AK Misra. “Physics-inspired optimization algorithms: a survey”. In: *Journal of Optimization* 2013 (2013) (cit. on pp. 44, 46).

[BKR19] Stefan Boschert, Mohamed Khalil, and Roland Rosen. “Digital Twin Based Diagnosis and Condition Prediction”. In: *8th IFAC Symposium on Mechatronic Systems*. 2019 (cit. on p. 14).

[BM12] Gregory Bartram and Sankaran Mahadevan. “Prognostics and health monitoring in the presence of heterogeneous information”. In: *Proc. Annual. Conference. Prognostic. Health Management Society*. Vol. 3. 126. 2012 (cit. on p. 65).

[BM97] DC Baillie and J Mathew. “Nonlinear model-based fault diagnosis of bearings”. In: *NDT and e International* 5.30 (1997), p. 328 (cit. on p. 35).

[BMS92] Y Brechet, T Magnin, and D Sornette. “The Coffin-Manson law as a consequence of the statistical nature of the LCF surface damage”. In: *Acta metallurgica et materialia* 40.9 (1992), pp. 2281–2287 (cit. on p. 56).

[BMZ12] Piero Baraldi, Francesca Mangili, and Enrico Zio. “A Kalman filter-based ensemble approach with application to turbine creep prognostics”. In: *IEEE Transactions on Reliability* 61.4 (2012), pp. 966–977 (cit. on p. 67).

[BMZ15] Piero Baraldi, Francesca Mangili, and Enrico Zio. “A belief function theory based approach to combining different representation of uncertainty in prognostics”. In: *Information Sciences* 303 (2015), pp. 134–149 (cit. on p. 67).

[Böh+14] Ewelina Böhm, Marta Kurek, Grzegorz Junak, Marek Cieśla, and Tadeusz **L**agoda. “Accumulation of fatigue damage using memory of the material”. In: *Procedia materials science* 3 (2014), pp. 2–7 (cit. on p. 56).

[Bou+00] Pierre-Olivier Bouchard, François Bay, Yvan Chastel, and Isabelle Tovena. “Crack propagation modelling using an advanced remeshing technique”. In: *Computer methods in applied mechanics and engineering* 189.3 (2000), pp. 723–742 (cit. on p. 58).

[BP04] Claude Bathias and Paul C Paris. *Gigacycle fatigue in mechanical practice*. CRC Press, 2004 (cit. on p. 55).

[BP13] Claude Bathias and André Pineau. *Fatigue of materials and structures: Application to damage and design*. Vol. 2. John Wiley & Sons, 2013 (cit. on p. 51).

[BR18] Stefan Boschert and Roland Rosen. “Digital Twin: a Second Life for Engineering Models”. In: *ERCIM NEWS* 115 (2018), pp. 8–9 (cit. on p. 14).

[Bre01] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32 (cit. on p. 174).

[BS02] T Biagetti and Enrico Sciubba. “A first step towards unmanned intelligent process management: A procedure for the diagnostics and prognostics of energy conversion plants”. In: *Int. J. Appl. Thermodyn* 5 (2002), pp. 85–99 (cit. on p. 53).

[BS04] Tatiana Biagetti and Enrico Sciubba. “Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems”. In: *Energy* 29.12-15 (2004), pp. 2553–2572 (cit. on p. 53).

[BS90] Julie Ann Bannantine and DF Socie. “A variable amplitude multiaxial fatigue life prediction methods”. In: *ICBMFF3*. 1990 (cit. on p. 56).

[BT18] D Benasciutti and R Tovo. “Frequency-based analysis of random fatigue loads: Models, hypotheses, reality: Frequenzbasierte Analyse zufälliger Ermüdungsbelastungen: Modelle, Hypothesen, Praxis”. In: *Materialwissenschaft und Werkstofftechnik* 49.3 (2018), pp. 345–367 (cit. on p. 59).

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004 (cit. on p. 99).

[BWE04] Carl S Byington, Matthew Watson, and Doug Edwards. “Data-driven neural network methodology to remaining life predictions for aircraft actuator components”. In: *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)*. Vol. 6. IEEE. 2004, pp. 3581–3589 (cit. on p. 64).

[BZL16] Giduthuri Sateesh Babu, Peilin Zhao, and Xiao-Li Li. “Deep convolutional neural network based regression approach for estimation of remaining useful life”. In: *International conference on database systems for advanced applications*. Springer. 2016, pp. 214–228 (cit. on p. 63).

[Cad+17] Francesco Cadini, Claudio Sbarufatti, Matteo Corbetta, and Marco Giglio. “A particle filter-based model selection algorithm for fatigue damage identification on aeronautical structures”. In: *Structural Control and Health Monitoring* 24.11 (2017), e2002 (cit. on p. 68).

[Cas13] Federico Castanedo. “A review of data fusion techniques”. In: *The Scientific World Journal* 2013 (2013) (cit. on p. 93).

[Cer+18] Mariela Cerrada, René-Vinicio Sánchez, Chuan Li, et al. “A review on data-driven fault severity assessment in rolling bearings”. In: *Mechanical Systems and Signal Processing* 99 (2018), pp. 169–196 (cit. on p. 35).

[Che+10] Chaochao Chen, Bin Zhang, George Vachtsevanos, and Marcos Orchard. “Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering”. In: *IEEE Transactions on Industrial Electronics* 58.9 (2010), pp. 4353–4364 (cit. on p. 65).

[Cho+18] Anurag Choudhary, Deepam Goyal, Sudha Letha Shimi, and Aparna Akula. "Condition Monitoring and Fault Diagnosis of Induction Motors: A Review". In: *Archives of Computational Methods in Engineering* (2018), pp. 1–18 (cit. on pp. 30–33, 36).

[CMNV17] Brigitte Chebel-Morello, Jean-Marc Nicod, and Christophe Varnier. *From Prognostics and Health Systems Management to Predictive Maintenance 2: Knowledge, Reliability and Decision*. John Wiley & Sons, 2017 (cit. on pp. 27, 30).

[CNM11] Mohamed Chookah, Mohammad Nuhi, and Mohammad Modarres. "A probabilistic physics-of-failure model for prognostic health management of structures subject to pitting and corrosion-fatigue". In: *Reliability Engineering & System Safety* 96.12 (2011), pp. 1601–1610 (cit. on p. 153).

[CSP02] Peter Y Chen, Michael Smithson, and Paula M Popovich. *Correlation: Parametric and nonparametric measures*. 139. Sage, 2002 (cit. on p. 178).

[CVO12] Chaochao Chen, George Vachtsevanos, and Marcos E Orchard. "Machine remaining useful life prediction: An integrated adaptive neuro-fuzzy and high-order particle filtering approach". In: *Mechanical Systems and Signal Processing* 28 (2012), pp. 597–607 (cit. on p. 65).

[DIN08] EN DIN. "13306: Maintenance–Maintenance terminology". In: *DIN Deutsches Institut für Normung eV* (2008) (cit. on pp. 4, 7, 9).

[Din08a] Steven X Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008 (cit. on p. 37).

[Din08b] Steven X Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008 (cit. on p. 154).

[Din+11] Steven X Ding, Ping Zhang, Torsten Jeinsch, et al. "A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis". In: *IFAC Proceedings Volumes* 44.1 (2011), pp. 12380–12388 (cit. on p. 154).

[DM96] Pierre Del Moral. “Non-linear filtering: interacting particle resolution”. In: *Markov processes and related fields* 2.4 (1996), pp. 555–581 (cit. on pp. 53, 77).

[Doe+96] Scott W Doebling, Charles R Farrar, Michael B Prime, and Daniel W Shevitz. *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review*. Tech. rep. Los Alamos National Lab., NM (United States), 1996 (cit. on pp. 26, 37).

[Dow13] Allen Downey. *Think Bayes: Bayesian statistics in python*. " O'Reilly Media, Inc.", 2013 (cit. on p. 53).

[DS82] Stephen D Downing and DF Socie. “Simple rainflow counting algorithms”. In: *International journal of fatigue* 4.1 (1982), pp. 31–40 (cit. on p. 58).

[Eft+13] M Eftekhari, M Moallem, S Sadri, and A Shojaei. “Review of induction motor testing and monitoring methods for inter-turn stator winding faults”. In: *2013 21st Iranian Conference on Electrical Engineering (ICEE)*. IEEE. 2013, pp. 1–6 (cit. on p. 36).

[EP11a] Elias N. Eliopoulos and Theodore P. Philippidis. “A progressive damage simulation algorithm for GFRP composites under cyclic loading. Part I: Material constitutive model”. In: *Composites science and technology* 71.5 (2011), pp. 742–749 (cit. on p. 207).

[EP11b] Elias N. Eliopoulos and Theodore P. Philippidis. “A progressive damage simulation algorithm for GFRP composites under cyclic loading. Part II: FE implementation and model validation”. In: *Composites Science and Technology* 71.5 (2011), pp. 750–757 (cit. on p. 207).

[ET16] Idriss El-Thalji. “Dynamic modelling and fault analysis of wear evolution in rolling bearings”. In: *VTT Science* 127 (2016) (cit. on p. 35).

[ETJ14] Idriss El-Thalji and Erkki Jantunen. “A descriptive model of wear evolution in rolling bearings”. In: *Engineering failure analysis* 45 (2014), pp. 204–224 (cit. on p. 34).

[ETJ15] Idriss El-Thalji and Erkki Jantunen. “A summary of fault modelling and predictive health monitoring of rolling element bearings”. In: *Mechanical Systems and Signal Processing* 60 (2015), pp. 252–272 (cit. on pp. 34, 35, 157).

[FB09] Thomas Frey and Martin Bossert. *Signal- und Systemtheorie*. Vol. 2. Springer, 2009 (cit. on p. 5).

[Fek14] Afef Fekih. "Fault diagnosis and fault tolerant control design for aerospace systems: A bibliographical review". In: *2014 American Control Conference*. IEEE. 2014, pp. 1286–1291 (cit. on p. 37).

[FF05] ASTM Committee E-8 on Fatigue and Fracture. *Standard practices for cycle counting in fatigue analysis*. ASTM International, 2005 (cit. on p. 58).

[FJ17] Shuo Feng and Jinqing Jia. "Acceleration sensor placement technique for vibration test in structural health monitoring using microhabitat frog-leaping algorithm". In: *Structural Health Monitoring* (2017), p. 1475921716688372 (cit. on pp. 46, 47).

[FMV00] Esteban Fernández, José M Montes, and Camilo J Vázquez. "Typology and strategic analysis of intangible resources: A resource-based approach". In: *Technovation* 20.2 (2000), pp. 81–92 (cit. on p. 10).

[Fra14] William E. Frazier. "Metal additive manufacturing: a review". In: *Journal of Materials Engineering and Performance* 23.6 (2014), pp. 1917–1928 (cit. on p. 206).

[Fre+09] S Freitag, M Beer, W Graf, and M Kaliske. "Lifetime prediction using accelerated test data and neural networks". In: *Computers & Structures* 87.19-20 (2009), pp. 1187–1194 (cit. on p. 62).

[FS11] Ali Fatemi and Nima Shamsaei. "Multiaxial fatigue: An overview and some approximation models for life estimation". In: *International Journal of Fatigue* 33.8 (2011), pp. 948–958 (cit. on p. 56).

[FW12] Charles R Farrar and Keith Worden. *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, 2012 (cit. on pp. 5, 10, 11, 25, 26, 28, 30, 37, 39, 88, 93).

[GAK16] Mladen Gibanica, Thomas JS Abrahamsson, and Daniel C Kammer. "Redundant information rejection in sensor localisation using system gramians". In: *Topics in Modal Analysis & Testing, Volume 10, Proceedings of the 34th IMAC, A Conference and Exposition on Structural Dynamics*. Springer, 2016, pp. 325–333 (cit. on p. 47).

[Gar+01] Amulya K Garga, Katherine T McClintic, Robert L Campbell, et al. “Hybrid reasoning for prognostic learning in CBM systems”. In: *2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542)*. Vol. 6. IEEE. 2001, pp. 2957–2969 (cit. on p. 64).

[Gar19] Inc. Gartner. “Top 10 Strategic Technology Trends for 2019”. recieved from https://www.gartner.com/ smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019/. 2019 (cit. on p. 13).

[GCD15a] Zhiwei Gao, Carlo Cecati, and Steven X Ding. “A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches”. In: *IEEE Transactions on Industrial Electronics* 62.6 (2015), pp. 3757–3767 (cit. on pp. 28, 35).

[GCD15b] Zhiwei Gao, Carlo Cecati, and Steven X Ding. “A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches”. In: *IEEE Transactions on Industrial Electronics* 6.62 (2015), pp. 3768–3774 (cit. on p. 28).

[GE07] Kai Goebel and Neil Eklund. “Prognostic fusion for uncertainty reduction”. In: *AIAA Infotech@ Aerospace 2007 Conference and Exhibit*. 2007, p. 2843 (cit. on p. 67).

[GE89] K Golos and F Ellyin. “Total strain energy density as a fatigue damage parameter”. In: *Advances in Fatigue Science and Technology*. Springer, 1989, pp. 849–858 (cit. on pp. 56, 57).

[Geo95] Didier Georges. “The use of observability and controllability gramians or functions for optimal sensor and actuator location in finite-dimensional systems”. In: *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*. Vol. 4. IEEE. 1995, pp. 3319–3324 (cit. on pp. 41, 46).

[GEW06] Pierre Geurts, Damien Ernst, and Louis Wehenkel. “Extremely randomized trees”. In: *Machine learning* 63.1 (2006), pp. 3–42 (cit. on p. 177).

[GMZ16] Rafael Gouriveau, Kamal Medjaher, and Noureddine Zerhouni. *From Prognostics and Health Systems Management to Predictive Maintenance 1: Monitoring and Prognostics*. John Wiley & Sons, 2016 (cit. on pp. 5, 8, 9, 27, 30, 88, 176, 177).

[GN11] Giulio Gola and Bent Helge Nystad. “From measurement collection to remaining useful life estimation: defining a diagnostic-prognostic frame for optimal maintenance scheduling of choke valves undergoing erosion”. In: *Annual Conference of the Prognostics and Health Management Society*. 2011, pp. 26–29 (cit. on p. 65).

[Gom+18] Guilherme Ferreira Gomes, Sebastiao Simões da Cunha, Patricia da Silva Lopes Alexandrino, Bruno Silva de Sousa, and Antonio Carlos Ancelotti. “Sensor placement optimization applied to laminated composite plates under vibration”. In: *Structural and Multidisciplinary Optimization* 58.5 (2018), pp. 2099–2118 (cit. on p. 48).

[Gou+14] Philippe Goupil, Josep Boada-Bauxell, Andres Marcos, et al. “AIRBUS efforts towards advanced real-time fault diagnosis and fault tolerant control”. In: *IFAC Proceedings Volumes* 47.3 (2014), pp. 3471–3476 (cit. on pp. 29, 37).

[GR95] Bijoy K Ghosh and Joachim Rosenthal. “A generalized Popov-Belevitch-Hautus test of observability”. In: *IEEE transactions on automatic control* 40.1 (1995), pp. 176–180 (cit. on p. 41).

[Gry94] Micheal S Grygler. “Modal Test Technology as Non-Destructive Evaluation of Space Shuttle Structures”. In: (1994) (cit. on pp. 28, 37).

[Gug+17] Narendhar Gugulothu, Vishnu TV, Pankaj Malhotra, et al. “Predicting remaining useful life using time series embeddings based on recurrent neural networks”. In: *arXiv preprint arXiv:1709.01073* (2017) (cit. on p. 63).

[Guo+17] Liang Guo, Naipeng Li, Feng Jia, Yaguo Lei, and Jing Lin. “A recurrent neural network based health indicator for remaining useful life prediction of bearings”. In: *Neurocomputing* 240 (2017), pp. 98–109 (cit. on p. 63).

[GV17] Michael Grieves and John Vickers. “Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems”. In: *Transdisciplinary perspectives on complex systems*. Springer, 2017, pp. 85–113 (cit. on p. 13).

[Han98] Military Handbook. “MIL-HDBK-5H: Metallic Materials and Elements for Aerospace Vehicle Structures”. In: *US Department of Defense* (1998), pp. 3–244 (cit. on p. 75).

[Hau08] Jan Hauth. “Grey-Box Modelling for Nonlinear Systems”. PhD thesis. Technische Universität Kaiserslautern, 2008 (cit. on p. 64).

[Hei08] Felix O Heimes. “Recurrent neural networks for remaining useful life estimation”. In: *2008 international conference on prognostics and health management*. IEEE. 2008, pp. 1–6 (cit. on pp. 62, 63).

[Hei+19] Christoph Heinrich, Mohamed Khalil, Kirill Martynov, and Utz Wever. “Online remaining lifetime estimation for structures”. In: *Mechanical Systems and Signal Processing* 119 (2019), pp. 312–327 (cit. on p. 58).

[Hen+09] Aiwina Heng, Sheng Zhang, Andy CC Tan, and Joseph Mathew. “Rotating machinery prognostics: State of the art, challenges and opportunities”. In: *Mechanical systems and signal processing* 23.3 (2009), pp. 724–739 (cit. on p. 68).

[HF02] Andrew Hess and Leo Fila. “The joint strike fighter (JSF) PHM concept: potential impact on aging aircraft problems”. In: *Proceedings, IEEE Aerospace Conference*. Vol. 6. IEEE. 2002, pp. 6–6 (cit. on p. 29).

[Hil+19] Marcel Hildebrandt, Mohamed Khalil, Christoph Bergs, et al. “Remaining useful life estimation for unknown fleet derivatives using a hybrid modeling approach”. In: *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)* (2019) (cit. on p. 160).

[Hil30] LE Hildebrand. “Quiet induction motors”. In: *Transactions of the American Institute of Electrical Engineers* 49.3 (1930), pp. 848–852 (cit. on p. 30).

[Hin15] Brian Thomas Hinson. “Observability-based guidance and sensor placement”. PhD thesis. University of Washington, 2015 (cit. on pp. 46, 48, 94, 95, 98, 99).

[HJS09] Elbert Hendricks, Ole Jannerup, and Paul Haase Sørensen. “Optimal observers: Kalman filters”. In: *Linear Systems Control*. Springer, 2009, pp. 431–491 (cit. on p. 81).

[HM14] Brian T Hinson and Kristi A Morgansen. “Observability-based optimal sensor placement for flapping airfoil wake estimation”. In: *Journal of Guidance, Control, and Dynamics* 37.5 (2014), pp. 1477–1486 (cit. on pp. 41, 46, 97, 100).

[Hob15] Adolf Hobbacher. *Recommendations for fatigue design of welded joints and components*. Springer, 2015 (cit. on p. 49).

[HR15] Roland Herzog and Ilka Riedel. "Sequentially optimal sensor placement in thermoelastic models for real time applications". In: *Optimization and Engineering* 16.4 (2015), pp. 737–766 (cit. on pp. 42, 254).

[HRU17] R. Herzog, I. Riedel, and D. Ucinski. "Optimal Sensor Placement For Joint Parameter And State Estimation Problems In Large-Scale Dynamical Systems With Applications To Thermo-Mechanics". TU Chemnitz. 2017 (cit. on p. 41).

[HT18] Ahmed Zakariae Hinchi and Mohamed Tkiouat. "Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network". In: *Procedia Computer Science* 127 (2018), pp. 123–132 (cit. on p. 63).

[Hu+12] Chao Hu, Byeng D Youn, Pingfeng Wang, and Joung Taek Yoon. "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life". In: *Reliability Engineering & System Safety* 103 (2012), pp. 120–135 (cit. on p. 67).

[Hua+07] Runqing Huang, Lifeng Xi, Xinglin Li, et al. "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods". In: *Mechanical systems and signal processing* 21.1 (2007), pp. 193–207 (cit. on p. 65).

[Hua+16] Zhi Yong Huang, Han Qing Liu, Hao Min Wang, et al. "Effect of stress ratio on VHCF behavior for a compressor blade titanium alloy". In: *International Journal of Fatigue* 93 (2016), pp. 232–237 (cit. on p. 55).

[HUL93] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. "Abstract duality for practitioners". In: *Convex Analysis and Minimization Algorithms II*. Springer, 1993, pp. 137–193.

[Hus+19] Ali Hussein, Fadi Aldakheel, Blaz Hudobivnik, et al. "A Computational framework for brittle crack propagation based on an efficient virtual element method". In: *Finite Elements in Analysis and Design, submitted* (2019) (cit. on p. 58).

[IEC12] P Irving, O Eker, and F Camci. "Prognostics control of cracking in structures and components operating in hydrogen environments". In: *Proc. Int. Hydrogen Conf.* 2012 (cit. on p. 66).

[Inm+13] Daniel J. Inman, Charles R. Farrar, Vicente Lopes Junior, and Valder Steffen Junior. *Damage Prognosis for Aerospace, Civil and Mechanical Systems*. Ed. by Daniel J. Inman, Charles R. Farrar, Vicente Lopes Junior, and Valder Steffen Junior. John Wiley & Sons, 2013 (cit. on p. 27).

[Inm17] Daniel J Inman. *Vibration with control*. John Wiley & Sons, 2017 (cit. on pp. 41, 72, 77).

[Int06] International Standard Organization (ISO). *ISO 13374 - Condition Monitoring and Diagnostics of Machines – Data Processing, Communication and Presentation – Part 2: Data Processing*. 2006 (cit. on pp. 11, 12).

[Irg08] Fridtjov Irgens. *Continuum mechanics*. Springer Science & Business Media, 2008 (cit. on pp. 72, 73).

[Ise05] Rolf Isermann. "Model-based fault-detection and diagnosis–status and applications". In: *Annual Reviews in control* 29.1 (2005), pp. 71–85 (cit. on p. 154).

[Ise06] Rolf Isermann. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media, 2006 (cit. on p. 5).

[Jia+10] Sun Jianzhong, Zuo Hongfu, Yang Haibin, and Michael Pecht. "Study of ensemble learning-based fusion prognostics". In: *2010 Prognostics and System Health Management Conference*. IEEE. 2010, pp. 1–7 (cit. on p. 67).

[Jia+16] Feng Jia, Yaguo Lei, Jing Lin, Xin Zhou, and Na Lu. "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data". In: *Mechanical Systems and Signal Processing* 72 (2016), pp. 303–315 (cit. on p. 35).

[JLB06] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. "A review on machinery diagnostics and prognostics implementing condition-based maintenance". In: *Mechanical systems and signal processing* 20.7 (2006), pp. 1483–1510 (cit. on pp. 36, 68).

[JQ03] S Joe Qin. “Statistical process monitoring: basics and beyond”. In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 17.8-9 (2003), pp. 480–502 (cit. on p. 154).

[Kal60] Rudolph Emil Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D (1960), pp. 35–45 (cit. on pp. 42, 53, 77, 81).

[Kam05] Daniel C Kammer. “Sensor set expansion for modal vibration testing”. In: *Mechanical systems and signal processing* 19.4 (2005), pp. 700–713 (cit. on pp. 40, 142).

[Kam+15] M Kaminski, F Laurin, JF Maire, C Rakotoarisoa, and E Hémon. “Fatigue damage modeling of composite structures: the onera viewpoint”. In: *AerospaceLab* 9 (2015), p–1 (cit. on pp. 51, 214).

[Kam91] Daniel C Kammer. “Sensor placement for on-orbit modal identification and correlation of large space structures”. In: *Journal of Guidance, Control, and Dynamics* 14.2 (1991), pp. 251–259 (cit. on pp. 40, 44, 142).

[Kan+17] Mallikarjun Kande, Alf Isaksson, Rajeev Thottappillil, and Nathaniel Taylor. “Rotating electrical machine condition monitoring automation - A review”. In: *Machines* 5.4 (2017), p. 24 (cit. on p. 29).

[KGK13] Horst Kuemmlee, Thomas Gross, and Josef Kolerus. “Machine vibrations and diagnostics the world of ISO”. In: *Industry Applications Society 60th Annual Petroleum and Chemical Industry Conference*. IEEE. 2013, pp. 1–13 (cit. on p. 43).

[Kha+18] Mohamed Khalil, Christoph Heinrich, Roland Wüchner, and Kai-Uwe Bletzinger. “Robust algorithm for optimal configuration of a multi-type network of sensors”. In: *Proceedings of ISMA 2018*. 2018 (cit. on pp. 47, 97).

[Kha+19a] Mohamed Khalil, Ioannis Kouroudis, Roland Wüchner, and Kai-Uwe Bletzinger. “Optimal sensor configuration for fatigue life prediction in structural applications”. In: *Proceedings of ASME 2019 Dynamic Systems and Control Conference* (2019) (cit. on pp. 101, 102, 259).

[Kha+19b] Mohamed Khalil, Theodoros Papadopoulos, Christoph Bergs, et al. “IIoT-based Fatigue Life Indicationusing Augmented Reality”. In: *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)* (2019) (cit. on p. 14).

[Kho14] Amir R Khoei. *Extended finite element method: theory and applications*. John Wiley & Sons, 2014 (cit. on p. 53).

[KHV06] Ranganath Kothamasu, Samuel H Huang, and William H VerDuin. “System health monitoring and prognostics—a review of current paradigms and practices”. In: *The International Journal of Advanced Manufacturing Technology* 28.9-10 (2006), pp. 1012–1024 (cit. on p. 68).

[KP08] Daniel C Kammer and Jeffrey A Peck. “Mass-weighting methods for sensor placement using sensor set expansion techniques”. In: *Mechanical Systems and Signal Processing* 22.7 (2008), pp. 1515–1525 (cit. on pp. 45, 95, 142).

[Kre+07] Dean Kreiser, Sheng Xiang Jia, Jian Jun Han, and Manicka Dhanasekar. “A nonlinear damage accumulation model for shakedown failure”. In: *International Journal of Fatigue* 29.8 (2007), pp. 1523–1530 (cit. on p. 57).

[Kro16] Andreas Kroll. *Computational intelligence: probleme, methoden und technische anwendungen*. Walter de Gruyter GmbH & Co KG, 2016 (cit. on pp. 166, 170).

[Kun+08] Ferenc Kun, HA Carmona, José S Andrade Jr, and Hans Jürgen Herrmann. “Universality behind Basquin’s law of fatigue”. In: *Physical review letters* 100.9 (2008), p. 094301 (cit. on p. 55).

[KV10] Karl Kunisch and Stefan Volkwein. “Optimal snapshot location for computing POD basis functions”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 44.3 (2010), pp. 509–529 (cit. on pp. 259, 260).

[KWB19] Mohamed Khalil, Roland Wüchner, and Kai-Uwe Bletzinger. “Generalization of spectral methods for high-cycle fatigue analysis to accommodate non-stationary random processes”. In: *Proceedings of ASME 2019 Dynamic Systems and Control Conference* (2019) (cit. on p. 59).

[Lam+19] Nicolas Lammens, Hunor Erdelyi, Tom Craeghs, Brecht Van Hooreweder, and Wim Van Paepegem. "Effect of Process Induced Artefacts on the Fatigue Life Estimation of Additive Manufactured Metal Components". In: *Proceedings of European Conference on Structural Integrity of Additively Manufactured Materials (ESIAM19)*. 2019 (cit. on p. 206).

[LDS18] Xiang Li, Qian Ding, and Jian-Qiao Sun. "Remaining useful life estimation in prognostics using deep convolution neural networks". In: *Reliability Engineering & System Safety* 172 (2018), pp. 1–11 (cit. on p. 63).

[Lee05] Yung-Li Lee. *Fatigue testing and analysis: theory and practice*. Vol. 13. Butterworth-Heinemann, 2005 (cit. on pp. 54, 58, 74, 76, 93).

[Lee+14] Jay Lee, Fangji Wu, Wenyu Zhao, et al. "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications". In: *Mechanical systems and signal processing* 42.1-2 (2014), pp. 314–334 (cit. on pp. 9, 25, 27).

[Lem01] Claude Lemaréchal. "Lagrangian relaxation". In: *Computational combinatorial optimization*. Springer, 2001, pp. 112–156.

[Lem+89] E Lembessis, G Antonopoulos, RE King, C Halatsis, and J Torres. "CASSANDRA: an on-line expert system for fault prognosis". In: *Proc. the 5th CIM Europe Conference on Computer Integrated Manufacturing*. Vol. 371377. 1989 (cit. on p. 53).

[Li12] Dongsheng Li. "Sensor placement methods and evaluation criteria in structural health monitoring". PhD thesis. Universitat Siegen, 2012 (cit. on pp. 39–41, 95).

[LI15] Curtis E Larsen and Tom Irvine. "A review of spectral methods for variable amplitude fatigue prediction and new results". In: *Procedia Engineering* 101 (2015), pp. 243–250 (cit. on p. 59).

[Li+17] Zhixiong Li, Dazhong Wu, Chao Hu, and Janis Terpenny. "An Ensemble Learning-based Prognostic Approach with Degradation-Dependent Weights for Remaining Useful Life Prediction". In: *Reliability Engineering & System Safety* (2017) (cit. on p. 66).

[Liu+08] Wei Liu, Wei-cheng Gao, Yi Sun, and Min-jian Xu. "Optimal sensor placement for spatial lattice structure based on genetic algorithms". In: *Journal of Sound and Vibration* 317.1-2 (2008), pp. 175–189 (cit. on p. 45).

[Liu+10] Jie Liu, Abhinav Saxena, Kai Goebel, Bhaskar Saha, and Wilson Wang. *An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries*. Tech. rep. NATIONAL AERONAUTICS and SPACE ADMINISTRATION MOFFETT FIELD CA AMES RESEARCH . . ., 2010 (cit. on p. 63).

[Liu+12] J Liu, W Wang, F Ma, YB Yang, and CS Yang. "A data-model-fusion prognostic framework for dynamic system state forecasting". In: *Engineering Applications of Artificial Intelligence* 25.4 (2012), pp. 814–823 (cit. on p. 66).

[Liu+13] Zhijuan Liu, Qing Li, Xianhui Liu, and Chundi Mu. "A hybrid LSSVR/HMM-based prognostic approach". In: *Sensors* 13.5 (2013), pp. 5542–5560 (cit. on p. 65).

[Liu+18a] Jiawei Liu, Qi Li, Weirong Chen, et al. "Remaining useful life prediction of PEMFC based on long short-term memory recurrent neural networks". In: *Int. J. Hydrogen Energy* (2018) (cit. on p. 63).

[Liu+18b] Ruonan Liu, Boyuan Yang, Enrico Zio, and Xuefeng Chen. "Artificial intelligence for fault diagnosis of rotating machinery: A review". In: *Mechanical Systems and Signal Processing* 108 (2018), pp. 33–47 (cit. on p. 35).

[LK14] Linxia Liao and Felix Köttig. "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction". In: *IEEE Transactions on Reliability* 63.1 (2014), pp. 191–207 (cit. on pp. 53, 64, 65).

[LK16] Linxia Liao and Felix Köttig. "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction". In: *Applied Soft Computing* 44 (2016), pp. 191–199 (cit. on p. 66).

[LKS08] Jing Li, Rakesh K Kapania, and William B Spillman. "Placement optimization of distributed-sensing fiber optic sensors using genetic algorithms". In: *AIAA journal* 46.4 (2008), pp. 824–836 (cit. on p. 45).

[Lot16] Inge Lotsberg. *Fatigue design of marine structures*. Cambridge University Press, 2016 (cit. on p. 49).

[Lou+12a] E Lourens, Costas Papadimitriou, Steven Gillijns, et al. "Joint input-response estimation for structural systems based on reduced-order models and vibration data from a limited number of sensors". In: *Mechanical Systems and Signal Processing* 29 (2012), pp. 310–327 (cit. on pp. 42, 77, 80, 81).

[Lou+12b] E Lourens, Edwin Reynders, Guido De Roeck, Geert Degrande, and Geert Lombaert. "An augmented Kalman filter for force identification in structural dynamics". In: *Mechanical Systems and Signal Processing* 27 (2012), pp. 446–460 (cit. on p. 81).

[LS+13] Khanh Le Son, Mitra Fouladirad, Anne Barros, Eric Levrat, and Benoît Iung. "Remaining useful life estimation based on stochastic deterioration models: A comparative study". In: *Reliability Engineering & System Safety* 112 (2013), pp. 165–175 (cit. on p. 61).

[LTL04] ZN Li, J Tang, and QS Li. "Optimal sensor locations for structural vibration measurements". In: *Applied Acoustics* 65.8 (2004), pp. 807–818 (cit. on p. 48).

[Luo+03] Jianhui Luo, Madhavi Namburu, Krishna Pattipati, et al. "Model-based prognostic techniques [maintenance applications]". In: *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference.* IEEE. 2003, pp. 330–340 (cit. on p. 53).

[Lv+15] Zhiqiang Lv, Hong-Zhong Huang, Shun-Peng Zhu, Huiying Gao, and Fangjun Zuo. "A modified nonlinear fatigue damage accumulation model". In: *International Journal of Damage Mechanics* 24.2 (2015), pp. 168–181 (cit. on p. 56).

[LXL18] Jian-Fu Lin, You-Lin Xu, and Siu-Seong Law. "Structural damage detection-oriented multi-type sensor placement with multi-objective optimization". In: *Journal of Sound and Vibration* 422 (2018), pp. 568–589 (cit. on pp. 42, 47).

[LY85] YK Lin and JN Yang. "A stochastic theory of fatigue crack propagation". In: *AIAA journal* 23.1 (1985), pp. 117–124 (cit. on p. 58).

[MA16] Haël Mughrabi and Stephen D Antolovich. "A tribute to Claude Bathias–Highlights of his pioneering work in Gigacycle Fatigue". In: *International Journal of Fatigue* 93 (2016), pp. 217–223 (cit. on p. 55).

[Man18] Krithika Manohar. "Data-Driven Sensor Placement Methods". PhD thesis. University of Washington, 2018 (cit. on p. 47).

[Mar+12] Julien Marzat, Hélène Piet-Lahanier, Frédéric Damongeot, and Eric Walter. "Model-based fault diagnosis for aerospace systems: a survey". In: *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of aerospace engineering* 226.10 (2012), pp. 1329–1360 (cit. on p. 37).

[Mar+16] Gabriel Marsh, Colin Wignall, Philipp R Thies, et al. "Review and application of Rainflow residue processing techniques for accurate fatigue damage estimation". In: *International Journal of Fatigue* 82 (2016), pp. 757–765 (cit. on p. 58).

[McC05] J McColl. "HUMS in the era of CAA, JAA, EASA and ICAO". In: *AIAC Conference, Melbourne, Paper*. Vol. 112. 2005 (cit. on pp. 28, 37).

[ME14] Saeed Masoudnia and Reza Ebrahimpour. "Mixture of experts: a literature survey". In: *Artificial Intelligence Review* 42.2 (2014), pp. 275–293 (cit. on p. 160).

[ME68] Masanori Matsuishi and Tatsuo Endo. "Fatigue of metals subjected to varying stress". In: *Japan Society of Mechanical Engineers, Fukuoka, Japan* 68.2 (1968), pp. 37–40 (cit. on pp. 56, 58).

[Mes+05] Gérard Mesmacque, S Garcia, Abdelwaheb Amrouche, and C Rubio-Gonzalez. "Sequential law in multiaxial fatigue, a new damage indicator". In: *International Journal of Fatigue* 27.4 (2005), pp. 461–467 (cit. on p. 56).

[MGB74] Alexander M Mood, Franklin A Graybill, and D Boes. "Introduction to the theory of statistics, McGraw-Hill". In: *DaCosta, CJ and Baenziger, JE et al (2003). A rapid method for assessing lipid: protein and detergent: protein ratios in membrane–protein crystallization* 59 (1974), pp. 77–83 (cit. on p. 127).

[Min59] Milton A Miner. "Estimation of fatigue life with particular emphasis on cumulative damage". In: *Metal Fatigue* (1959), pp. 278–89 (cit. on p. 55).

[Mit07] John S Mitchell. "From vibration measurements to condition-based maintenance". In: *Sound and Vibration* 41.1 (2007), p. 62 (cit. on p. 30).

[MK11] Nader Meskin and Khashayar Khorasani. *Fault detection and isolation: Multi-vehicle unmanned systems*. Springer Science & Business Media, 2011 (cit. on p. 37).

[Moh+07] Subhasish Mohanty, Rikki Teale, Aditi Chattopadhyay, Pedro Peralta, and Christina Willhauck. "Mixed Gaussian process and state-space approach for fatigue crack growth prediction". In: *International workshop on structural heath monitoring*. Vol. 2. Citeseer. 2007, pp. 1108–1115 (cit. on p. 66).

[Moh18] Amiya Ranjan Mohanty. *Machinery condition monitoring: Principles and practices*. CRC Press, 2018 (cit. on p. 93).

[MSH10] Abd Kadir Mahamad, Sharifah Saon, and Takashi Hiyama. "Predicting remaining useful life of rotating machinery based artificial neural network". In: *Computers & Mathematics with Applications* 60.4 (2010), pp. 1078–1087 (cit. on p. 62).

[MSI08] Alexandre Muller, Marie-Christine Suhner, and Benoît Iung. "Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system". In: *Reliability Engineering & System Safety* 93.2 (2008), pp. 234–253 (cit. on p. 11).

[MZ05] C Marquardt and H Zenner. "Lifetime calculation under variable amplitude loading with the application of artificial neural networks". In: *International journal of fatigue* 27.8 (2005), pp. 920–927 (cit. on p. 62).

[Nie+15] PJ García Nieto, E Garcia-Gonzalo, F Sánchez Lasheras, and Francisco Javier de Cos Juez. "Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability". In: *Reliability Engineering & System Safety* 138 (2015), pp. 219–231 (cit. on p. 64).

[NW06] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006 (cit. on p. 107).

[O'D+87] Pat O'DONNELL, C Heising, C Singh, and SJ Wells. "Report of large motor reliability survey of industrial and commercial installations. I, II, III". In: *IEEE Transactions on Industry Applications* 23.1 (1987), pp. 153–158 (cit. on p. 33).

[Oko+14] Caxton Okoh, Rajkumar Roy, Jorn Mehnen, and L Redding. "Overview of remaining useful life prediction techniques in through-life engineering services". In: *Procedia CIRP* 16 (2014), pp. 158–163 (cit. on pp. 50, 62, 68).

[OSK03] Rolf F Orsagh, Jeremy Sheldon, and Christopher J Klenke. "Prognostics/diagnostics for gas turbine engine bearings". In: *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*. Vol. 7. IEEE. 2003, pp. 3095–3103 (cit. on p. 65).

[OSM19] Wieslaw Ostachowicz, Rohan Soman, and Pawel Malinowski. "Optimization of sensor placement for structural health monitoring: a review". In: *Structural Health Monitoring* (2019), p. 1475921719825601 (cit. on pp. 39, 40, 42, 48, 87, 98).

[OV07] Marcos E Orchard and George J Vachtsevanos. "A particle filtering approach for on-line failure prognosis in a planetary carrier plate". In: *International Journal of Fuzzy Logic and Intelligent Systems* 7.4 (2007), pp. 221–227 (cit. on p. 66).

[OW04] G Overton and K Worden. "Sensor optimisation using an ant colony metaphor". In: *Strain* 40.2 (2004), pp. 59–65 (cit. on p. 46).

[Pal59] Arvid Palmgren. "Ball and roller bearing engineering". In: *Philadelphia: SKF Industries Inc., 1959* (1959) (cit. on p. 34).

[Pap04] Costas Papadimitriou. "Optimal sensor placement methodology for parametric identification of structural systems". In: *Journal of sound and vibration* 278.4 (2004), pp. 923–947 (cit. on pp. 40, 95, 97).

[Pap+11] Costas Papadimitriou, Claus-Peter Fritzen, Peter Kraemer, and Evangelos Ntotsios. "Fatigue predictions in entire body of metallic structures from a limited number of vibration sensors using Kalman filtering". In: *Structural Control and Health Monitoring* 18.5 (2011), pp. 554–573 (cit. on pp. 59, 114, 142, 209, 220).

[Pap+14] Maria Papadopoulou, Benny Raphael, Ian FC Smith, and Chandra Sekhar. "Hierarchical sensor placement using joint entropy and the effect of modeling error". In: *Entropy* 16.9 (2014), pp. 5078–5101 (cit. on p. 45).

[Pas+13] Roland Pastorino, Dario Richiedei, Javier Cuadrado, and Alberto Trevisani. "State estimation using multibody models and non-linear Kalman filters". In: *International Journal of Non-Linear Mechanics* 53 (2013), pp. 83–90 (cit. on p. 77).

[Pav18] Dimitrios G Pavlou. "The theory of the SN fatigue damage envelope: generalization of linear, double-linear, and non-linear fatigue damage models". In: *International Journal of Fatigue* 110 (2018), pp. 204–214 (cit. on p. 56).

[PE63] Pe Paris and Fazil Erdogan. "A critical analysis of crack propagation laws". In: *Journal of basic engineering* 85.4 (1963), pp. 528–533 (cit. on p. 57).

[Pee08] Leto Peel. "Data driven prognostics using a Kalman filter ensemble of neural network models". In: *2008 International Conference on Prognostics and Health Management*. IEEE. 2008, pp. 1–6 (cit. on p. 67).

[Pen+12] Tishun Peng, Jingjing He, Yongming Liu, et al. *Integrated fatigue damage diagnosis and prognosis under uncertainties*. Tech. rep. NATIONAL AERONAUTICS and SPACE ADMINISTRATION MOFFETT FIELD CA AMES RESEARCH ..., 2012 (cit. on p. 66).

[Pet18] Sanyapong Petchrompo. "A review of asset management literature on multi-asset systems". In: *Reliability Engineering & System Safety* (2018) (cit. on p. 155).

[PK96] Youn-Sik Park and Hong-Bae Kim. "Sensor placement guide for model comparison and improvement". In: *PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING*. SPIE INTERNATIONAL SOCIETY FOR OPTICAL. 1996, pp. 404–409 (cit. on p. 260).

[PL12] Costas Papadimitriou and G. Lombaert. “The effect of prediction error correlation on optimal sensor placement in structural dynamics”. In: *Mechanical Systems and Signal Processing* 28.4 (2012), pp. 105–127 (cit. on p. 40).

[PLJ18] Thierry Palin-Luc and Dalenda Jeddi. “The gigacycle fatigue strength of steels: a review of structural and operating factors”. In: *Procedia Structural Integrity* 13 (2018), pp. 1545–1553 (cit. on p. 55).

[PSB11] B Pyttel, D Schwerdt, and C Berger. “Very high cycle fatigue–is there a fatigue limit?” In: *International Journal of fatigue* 33.1 (2011), pp. 49–58 (cit. on p. 55).

[PWB08] S Gareth Pierce, Keith Worden, and Abderrezak Bezazi. “Uncertainty analysis of a neural network used for fatigue lifetime prediction”. In: *Mechanical Systems and Signal Processing* 22.6 (2008), pp. 1395–1411 (cit. on p. 62).

[PZ04] G Petrucci and B Zuccarello. “Fatigue life prediction under wide band random loading”. In: *Fatigue & Fracture of Engineering Materials & Structures* 27.12 (2004), pp. 1183–1195 (cit. on p. 59).

[QLW16] John P Quigley, Yung-Li Lee, and Liang Wang. “Review and assessment of frequency-based fatigue damage models”. In: *SAE International Journal of Materials and Manufacturing* 9.3 (2016), pp. 565–577 (cit. on p. 59).

[RA07] A Rama Mohan Rao and Ganesh Anandakumar. “Optimal placement of sensors for structural system identification and health monitoring using a hybrid swarm intelligence technique”. In: *Smart materials and Structures* 16.6 (2007), p. 2658 (cit. on p. 46).

[Ran11] Robert Bond Randall. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011 (cit. on pp. 28, 31, 36).

[RBMGSM17] Juan Rivera-Barrera, Nicolás Muñoz-Galeano, and Henry Sarmiento-Maldonado. “SoC estimation for lithium-ion batteries: review and future challenges”. In: *Electronics* 6.4 (2017), p. 102 (cit. on p. 38).

[Ren+18] Lei Ren, Yaqiang Sun, Hao Wang, and Lin Zhang. “Prediction of bearing remaining useful life with deep convolution neural network”. In: *IEEE Access* 6 (2018), pp. 13041–13049 (cit. on p. 63).

[RLT08] Freeman Rufus, Seungkoo Lee, and Ash Thakker. “Health monitoring algorithms for space application batteries”. In: *2008 International Conference on Prognostics and Health Management*. IEEE. 2008, pp. 1–8 (cit. on p. 38).

[Roe+13] M. Roemer, G. Kacpryznski, R. Orsagh, and B. Marshall. “Prognosis of Rotating Machinery Components”. In: ed. by Daniel J. Inman, Charles R. Farrar, Vicente Lopes Junior, and Valder Steffen Junior. John Wiley & Sons, 2013. Chap. 19, pp. 385 –420 (cit. on p. 29).

[RPK19] M Raissi, P Perdikaris, and GE Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *Journal of Computational Physics* 378 (2019), pp. 686–707 (cit. on p. 154).

[RS12] Robert E Roberson and Richard Schwertassek. *Dynamics of multibody systems*. Springer Science & Business Media, 2012 (cit. on p. 53).

[Sad+09] Farshid Sadeghi, Behrooz Jalalahmadi, Trevor S Slack, Nihar Raje, and Nagaraj K Arakere. “A review of rolling contact fatigue”. In: *Journal of tribology* 131.4 (2009), p. 041403 (cit. on p. 58).

[Sah+07] Bhaskar Saha, Scott Poll, Kai Goebel, and Jon Christophersen. “An integrated approach to battery health monitoring using Bayesian regression and state estimation”. In: *2007 IEEE Autotestcon*. Ieee. 2007, pp. 646–653 (cit. on p. 66).

[Sak+16] Tatsuo Sakai, Akiyoshi Nakagawa, Noriyasu Oguma, et al. “A review on fatigue fracture modes of structural metallic materials in very high cycle regime”. In: *International Journal of Fatigue* 93 (2016), pp. 339–351 (cit. on p. 55).

[Sam19] Wojciech Samek. *Explainable AI: interpreting, explaining and visualizing deep learning*. Vol. 11700. Springer Nature, 2019 (cit. on p. 154).

[San+16] E Santecchia, AMS Hamouda, F Musharavati, et al. “A review on fatigue life prediction methods for metals”. In: *Advances in Materials Science and Engineering* 2016 (2016) (cit. on pp. 35, 56–58, 73, 74).

[Sao17] Victor E. Saouma. *Lecture Notes in Fracture Mechancis*. Virgina Tech. Jan. 2017 (cit. on pp. 5, 35, 58).

[Sba+16] Claudio Sbarufatti, Matteo Corbetta, Andrea Manes, and Marco Giglio. “Sequential Monte-Carlo sampling based on a committee of artificial neural networks for posterior state estimation and residual lifetime prediction”. In: *International Journal of Fatigue* 83 (2016), pp. 10–23 (cit. on p. 68).

[Sci04] Enrico Sciubba. “Hybrid semi-quantitative monitoring and diagnostics of energy conversion processes”. In: *International Journal of Thermodynamics* 7 (2004), pp. 95–106 (cit. on p. 53).

[SD+19] JJ Saucedo-Dorantes, M Delgado-Prieto, RA Osornio-Rios, and RJ Romero-Troncoso. “Spectral Analysis of Nonlinear Vibration Effects Produced by Worn Gears and Damaged Bearing in Electromechanical Systems: A Condition Monitoring Approach”. In: *Nonlinear Structural Dynamics and Damping*. Springer, 2019, pp. 293–320 (cit. on p. 33).

[Ser+13] Mitch Serpas, Gabriel Hackebeil, Carl Laird, and Juergen Hahn. “Sensor location for nonlinear dynamic systems via observability analysis and MAX-DET optimization”. In: *Computers & Chemical Engineering* 48 (2013), pp. 105–112 (cit. on p. 46).

[SG09] Bhaskar Saha and Kai Goebel. “Modeling Li-ion battery capacity depletion in a particle filtering framework”. In: *Proceedings of the annual conference of the prognostics and health management society*. 2009, pp. 2909–2924 (cit. on p. 38).

[SH05] Abhay K Singh and Juergen Hahn. “On the use of empirical gramians for controllability and observability analysis”. In: *American Control Conference, 2005. Proceedings of the 2005*. IEEE. 2005, pp. 140–141 (cit. on p. 41).

[Shi11] Joseph Edward Shigley. *Mechanical Engineering Design*. Tata McGraw-Hill Education, 2011 (cit. on p. 49).

[SHM11] JZ Sikorska, Melinda Hodkiewicz, and Lin Ma. “Prognostic modelling options for remaining useful life estimation by industry”. In: *Mechanical systems and signal processing* 25.5 (2011), pp. 1803–1836 (cit. on p. 68).

[Shu02] Peter J Shull. “Nondestructive Evaluation: Theory”. In: *Techniques, and Applications, Marcel Ekker Inc* (2002), pp. 17–18 (cit. on p. 37).

[Si+11] Xiao-Sheng Si, Wenbin Wang, Chang-Hua Hu, and Dong-Hua Zhou. "Remaining useful life estimation–a review on the statistical data driven approaches". In: *European journal of operational research* 213.1 (2011), pp. 1–14 (cit. on pp. 59–61).

[Sny05] Jan A Snyman. *Practical mathematical optimization*. Springer, 2005.

[Soh+03] Hoon Sohn, Charles R Farrar, Francois M Hemez, et al. "A review of structural health monitoring literature: 1996–2001". In: *Los Alamos National Laboratory, USA* (2003) (cit. on p. 26).

[Som+12] R Soman, T Onoufriou, R Votsis, CZ Chrysostomou, and MA Kyriakides. "Optimization of wireless sensor locations for SHM based on application demands and networking limitations". In: *Bridge Maintenance, Safety, Management, Resilience and Sustainability: Proceedings of the Sixth International IABMAS Conference, Stresa, Lake Maggiore, Italy, 8-12 July 2012*. CRC Press. 2012, p. 313 (cit. on p. 47).

[Som+14] Rohan N Soman, Toula Onoufrioua, Marios A Kyriakidesb, Renos A Votsisc, and Christis Z Chrysostomou. "Multi-type, multi-sensor placement optimization for structural health monitoring of long span bridges". In: *Smart Structures and Systems* 14.1 (2014), pp. 55–70 (cit. on p. 48).

[SON18] Chongmin Song, Ean Tat Ooi, and Sundararajan Natarajan. "A review of the scaled boundary finite element method for two-dimensional linear elastic fracture mechanics". In: *Engineering Fracture Mechanics* 187 (2018), pp. 45–73 (cit. on p. 58).

[SR79] WW Stinchcomb and KL Reifsnider. "Fatigue damage mechanisms in composite materials: a review". In: *Fatigue mechanisms*. ASTM International, 1979 (cit. on p. 50).

[SS05] B Satish and NDR Sarma. "A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors". In: *IEEE Power Engineering Society General Meeting, 2005*. IEEE. 2005, pp. 2291–2294 (cit. on p. 64).

[SSA15] Rodney K Singleton, Elias G Strangas, and Selin Aviyente. "Extended Kalman filtering for remaining-useful-life estimation of bearings". In: *IEEE Transactions on Industrial Electronics* 62.3 (2015), pp. 1781–1790 (cit. on p. 61).

[Ste+15] Amanda Sterling, Nima Shamsaei, Brian Torries, and Scott M Thompson. "Fatigue Behaviour of Additively Manufactured Ti-6Al-4 V". In: *Procedia Engineering* 133 (2015), pp. 576–589 (cit. on p. 218).

[Sto13] GC Stone. "Condition monitoring and diagnostics of motor and stator windings–A review". In: *IEEE Transactions on Dielectrics and Electrical Insulation* 20.6 (2013), pp. 2073–2080 (cit. on pp. 35, 36).

[Su+99] LP Su, M Nolan, Greg DeMare, and David R Carey. "Prognostics framework [for weapon systems health monitoring]". In: *1999 IEEE AUTOTESTCON Proceedings (Cat. No. 99CH36323)*. IEEE. 1999, pp. 661–672 (cit. on p. 29).

[Sut+18] Danas Sutula, Pierre Kerfriden, Tonie Van Dam, and Stéphane PA Bordas. "Minimum energy multiple crack propagation. Part-II: Discrete solution with XFEM". In: *Engineering Fracture Mechanics* 191 (2018), pp. 225–256 (cit. on p. 58).

[Sva02] Krister Svanberg. "A class of globally convergent optimization methods based on conservative convex separable approximations". In: *SIAM journal on optimization* 12.2 (2002), pp. 555–573 (cit. on p. 108).

[Sva87] Krister Svanberg. "The method of moving asymptotes—a new method for structural optimization". In: *International journal for numerical methods in engineering* 24.2 (1987), pp. 359–373 (cit. on pp. 107, 108).

[SVR08] Wilhelmus HA Schilders, Henk A Van der Vorst, and Joost Rommes. *Model order reduction: theory, research aspects and applications*. Vol. 13. Springer, 2008 (cit. on p. 41).

[Swa01] David C Swanson. "A general prognostic tracking algorithm for predictive maintenance". In: *2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542)*. Vol. 6. IEEE. 2001, pp. 2971–2977 (cit. on p. 64).

[SWT70] K. N. Smith, P. Watson, and T. H. Topper. "A Stress-Strain Function for the Fatigue of Metals". In: *Journal of Materials., JMLSA* 5 (1970), p. 767 (cit. on p. 74).

[Tam+16] Tommaso Tamarozzi, Enrico Risaliti, Ward Rottiers, Wim Desmet, et al. “Noise, ill-conditioning and sensor placement analysis for force estimation through virtual sensing”. In: *In International Conference on Noise and Vibration Engineering (ISMA2016),* 2016, pp. 1741–1756 (cit. on pp. 41, 142).

[TDP12] Dhirendra Nath Thatoi, Harish Ch Das, and Dayal R Parhi. “Review of techniques for fault diagnosis in damaged structure and engineering system”. In: *Advances in Mechanical Engineering* 4 (2012), p. 327569 (cit. on p. 28).

[THVB01] Irina Trendafilova, Ward Heylen, and Hendrik Van Brussel. “Measurement point selection in damage detection using the mutual information concept”. In: *Smart materials and structures* 10.3 (2001), p. 528 (cit. on p. 47).

[Tia12] Zhigang Tian. “An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring”. In: *Journal of Intelligent Manufacturing* 23.2 (2012), pp. 227–237 (cit. on p. 62).

[Tid+16] Khaoula Tidriri, Nizar Chatti, Sylvain Verron, and Téodor Tiplica. “Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges”. In: *Annual Reviews in Control* 42 (2016), pp. 63–81 (cit. on p. 13).

[TJP16] Vasileios Tzoumas, Ali Jadbabaie, and George J Pappas. “Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms”. In: *American Control Conference (ACC), 2016*. IEEE. 2016, pp. 191–196 (cit. on pp. 42, 45).

[TM81] K Tanaka and T Mura. “A dislocation model for fatigue crack initiation”. In: *Journal of Applied Mechanics* 48.1 (1981), pp. 97–103 (cit. on p. 57).

[Tol+12] Hamid A Toliyat, Subhasis Nandi, Seungdeog Choi, and Homayoun Meshgin-Kelk. *Electric machines: modeling, condition monitoring, and fault diagnosis*. CRC press, 2012 (cit. on p. 33).

[TZ09] Zhigang Tian and Ming J Zuo. “Health condition prognostics of gears using a recurrent neural network approach”. In: *2009 Annual Reliability and Maintainability Symposium*. IEEE. 2009, pp. 460–465 (cit. on p. 63).

[Vas19] Anastasios P Vassilopoulos. *Fatigue life prediction of composites and composite structures*. Woodhead publishing, 2019 (cit. on p. 74).

[VN09] JM Van Noortwijk. "A survey of the application of gamma processes in maintenance". In: *Reliability Engineering & System Safety* 94.1 (2009), pp. 2–21 (cit. on p. 61).

[VP04] Andrea Vania and Paolo Pennacchi. "Experimental and theoretical application of fault identification measures of accuracy in rotating machine diagnostics". In: *Mechanical Systems and Signal Processing* 18.2 (2004), pp. 329–352 (cit. on p. 35).

[VP07] Nikhil Vichare and Michael Pecht. "Enabling electronic prognostics using thermal data". In: *arXiv preprint arXiv:0709.1813* (2007) (cit. on p. 38).

[Wag+10] D Wagner, N Ranc, C Bathias, and PC Paris. "Fatigue crack initiation detection by an infrared thermography method". In: *Fatigue & Fracture of Engineering Materials & Structures* 33.1 (2010), pp. 12–21 (cit. on p. 55).

[Wan10a] Tianyi Wang. "Trajectory similarity based prediction for remaining useful life estimation". PhD thesis. University of Cincinnati, 2010 (cit. on p. 153).

[Wan10b] Xiao Wang. "Wiener processes with random effects for degradation data". In: *Journal of Multivariate Analysis* 101.2 (2010), pp. 340–351 (cit. on p. 61).

[Wan+99] QY Wang, JY Berard, 1 A Dubarre, et al. "Gigacycle fatigue of ferrous alloys". In: *Fatigue & Fracture of Engineering Materials & Structures* 22.8 (1999), pp. 667–672 (cit. on p. 55).

[WB94] CH Wang and MW Brown. "A study of the deformation behaviour under multiaxial loading". In: *European journal of mechanics. A. Solids* 13.2 (1994), pp. 175–188 (cit. on p. 56).

[WDB04] Keith Worden and Janice M Dulieu-Barton. "An overview of intelligent fault detection in systems and structures". In: *Structural Health Monitoring* 3.1 (2004), pp. 85–98 (cit. on pp. 4, 45).

[Wel82] William J Welch. “Branch-and-bound search for experimental designs based on D optimality and other criteria”. In: *Technometrics* 24.1 (1982), pp. 41–48 (cit. on p. 46).

[WL80] Paul H Wirsching and Mark C Light. “Fatigue under wide band random stresses”. In: *Journal of the Structural Division* 106.7 (1980), pp. 1593–1607 (cit. on p. 59).

[Wu+18] Dazhong Wu, Connor Jennings, Janis Terpenny, Soundar Kumara, and Robert X Gao. “Cloud-based parallel machine learning for tool wear prediction”. In: *Journal of Manufacturing Science and Engineering* 140.4 (2018), p. 041005 (cit. on p. 68).

[Xin+13] Yinjiao Xing, Eden WM Ma, Kwok-Leung Tsui, and Michael Pecht. “An ensemble model for predicting the remaining useful performance of lithium-ion batteries”. In: *Microelectronics Reliability* 53.6 (2013), pp. 811–820 (cit. on p. 67).

[YL12] Ting-Hua Yi and Hong-Nan Li. “Methodology developments in sensor placement for health monitoring of civil infrastructures”. In: *International Journal of Distributed Sensor Networks* 8.8 (2012), p. 612726 (cit. on pp. 39, 42).

[Zah19] Eliahu Zahavi. *Fatigue design: life expectancy of machine parts*. CRC press, 2019 (cit. on p. 49).

[ZAS17] Haotian Zhang, Raid Ayoub, and Shreyas Sundaram. “Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms”. In: *Automatica* 78 (2017), pp. 202–210 (cit. on pp. 42, 45, 95, 96, 120).

[Zer+16] Uwe Zerbst, Michael Vormwald, Reinhard Pippan, et al. “About the fatigue crack propagation threshold of metals as a design criterion–a review”. In: *Engineering Fracture Mechanics* 153 (2016), pp. 190–243 (cit. on p. 49).

[Zer+19] Uwe Zerbst, Mauro Madia, Christian Klinger, Dirk Bettge, and Y Murakami. “Defects as a root cause of fatigue failure of metallic components”. In: *Engineering Failure Analysis* (2019) (cit. on p. 49).

[Zha+16] Rui Zhao, Ruqiang Yan, Zhenghua Chen, et al. “Deep learning and its applications to machine health monitoring: A survey”. In: *arXiv preprint arXiv:1612.07640* (2016) (cit. on p. 62).

[Zha+17] Wan Zhang, Min-Ping Jia, Lin Zhu, and Xiao-An Yan. "Comprehensive overview on computational intelligence techniques for machinery condition monitoring and fault diagnosis". In: *Chinese Journal of Mechanical Engineering* 30.4 (2017), pp. 782–795 (cit. on p. 31).

[Zhe+17] Shuai Zheng, Kosta Ristovski, Ahmed Farahat, and Chetan Gupta. "Long short-term memory network for remaining useful life estimation". In: *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE. 2017, pp. 88–95 (cit. on p. 63).

[Zhu+18] Shun-Peng Zhu, Zheng-Yong Yu, José Correia, Abílio De Jesus, and Filippo Berto. "Evaluation and comparison of critical plane criteria for multiaxial fatigue analysis of ductile and brittle materials". In: *International Journal of Fatigue* 112 (2018), pp. 279–288 (cit. on pp. 56, 75).

[ZM12] Cha Zhang and Yunqian Ma. *Ensemble machine learning: methods and applications*. Springer, 2012 (cit. on p. 159).

[Zol+14] Ali Zolghadri, David Henry, Jérôme Cieslak, Denis Efimov, and Philippe Goupil. *Fault diagnosis and fault-tolerant control and guidance for aerospace vehicles*. Springer, 2014 (cit. on p. 37).

[ZW17] K Zhou and ZY Wu. "Strain gauge placement optimization for structural performance assessment". In: *Engineering Structures* 141 (2017), pp. 184–197 (cit. on p. 43).

[ZX16] CD Zhang and YL Xu. "Optimal multi-type sensor placement for response and excitation reconstruction". In: *Journal of Sound and Vibration* 360 (2016), pp. 112–128 (cit. on pp. 42, 45, 47, 48, 80, 82, 95, 142, 209, 220).

[ZZT13] O. C. Zienkiewicz, J.Z. Zhu, and Robert L. Taylor. *The Finite Element Method: its Basis and Fundamentals*. 7th ed. Butterworth-Heinemann, 2013 (cit. on pp. 53, 71, 79).

[ZZX13] Bin Zhanga, Lijun Zhang, and Jinwu Xub. "Remaining useful life prediction for rolling element bearing based on ensemble learning". In: *CHEMICAL ENGINEERING* 33 (2013) (cit. on p. 67).

List of Publications

[Kha+18] Mohamed Khalil, Christoph Heinrich, Roland Wüchner, and Kai-Uwe Bletzinger. "Robust algorithm for optimal configuration of a multi-type network of sensors". In: *Proceedings of ISMA 2018*. 2018 (cit. on pp. 47, 97).

[Kha+19a] Mohamed Khalil, Ioannis Kouroudis, Roland Wüchner, and Kai-Uwe Bletzinger. "Optimal sensor configuration for fatigue life prediction in structural applications". In: *Proceedings of ASME 2019 Dynamic Systems and Control Conference* (2019) (cit. on pp. 101, 102, 259).

[Kha+19b] Mohamed Khalil, Theodoros Papadopoulos, Christoph Bergs, et al. "IIoT-based Fatigue Life Indicationusing Augmented Reality". In: *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)* (2019) (cit. on p. 14).

[KWB19] Mohamed Khalil, Roland Wüchner, and Kai-Uwe Bletzinger. "Generalization of spectral methods for high-cycle fatigue analysis to accommodate non-stationary random processes". In: *Proceedings of ASME 2019 Dynamic Systems and Control Conference* (2019) (cit. on p. 59).