



**HAL**  
open science

# Data-driven prognostics for maintenance decision making: Facing industrial small datasets using Data Augmentation with prior knowledge integration

Antonin Gay

► **To cite this version:**

Antonin Gay. Data-driven prognostics for maintenance decision making: Facing industrial small datasets using Data Augmentation with prior knowledge integration. Automatic. Université de Lorraine, 2023. English. NNT: 2023LORR0059 . tel-04205398

**HAL Id: tel-04205398**

**<https://hal.univ-lorraine.fr/tel-04205398>**

Submitted on 12 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**UNIVERSITÉ  
DE LORRAINE**

**BIBLIOTHÈQUES  
UNIVERSITAIRES**

## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)  
*(Cette adresse ne permet pas de contacter les auteurs)*

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

# Pronostic de défaillance basé sur les données pour la prise de décision en maintenance :

Exploitation du principe d'augmentation de données avec intégration de connaissances à priori pour faire face aux problématiques du small data set.

## THÈSE

présentée et soutenue publiquement le 06 Juin 2023

pour l'obtention du

**Doctorat de l'Université de Lorraine**

**(Automatique)**

par

**Antonin GAY**

### Composition du jury

*Président :* Christophe BERENGUER (Université Grenoble Alpes)

*Rapporteurs :* François PERES (Université de Toulouse)  
Birgit VOGEL-HEUSER (Technische Universität München)

*Examineurs :* Olga FINK (EPFL Lausanne)  
Rémi BONIDAL (ArcelorMittal), co-encadrant industriel  
Ahmed KHELASSI (ArcelorMittal), co-encadrant industriel  
Alexandre VOISIN (Université de Lorraine), co-directeur de thèse  
Benoit IUNG (Université de Lorraine), directeur de thèse

Mis en page avec la classe thesul.

**Data-driven prognostics for  
maintenance decision making:**  
Facing industrial small datasets using  
Data Augmentation with prior knowledge  
integration.



## Remerciements

Paradoxically, the acknowledgements of the PhD are the last part left to write before definitively sending the manuscript. Not that it is difficult to determine who is to be thanked, I've been well accompanied in this adventure. But rather for the difficulty to find the right words to share all my gratitude to all the people who shared this slice of life with me.

First and foremost, I would like to thank my five PhD supervisors, Benoit Iung, Alexandre Voisin, Phuc Do, Rémi Bonidal and Ahmed Khelassi, despite their number, the thesis would have been completely different if only one of them had not been there. Despite the conditions in which this PhD started (in this particular month of Mars 2020) and was conducted, they have always been present and made all their possible for the PhD to move forward in the best conditions. Thank you to Alexandre and Rémi for all the technical talks around a coffee, where small ideas have emerged that ended up filling tens of pages in this manuscript. Thank you to Benoit for reminding me of the “Big Picture” when I mostly wanted to go back to Python coding. Thank you to Phuc for following me in some strange ideas for decision-making, it ended up as a full chapter so in the end it was probably worth it. And thank you to Ahmed, for always finding time in your busy schedule to discuss on everything.

Finally, in this time of writing, thank to you all for your efficiency in reviewing this manuscript. We would have not respected the schedule if not for your dedication.

I would like to address my acknowledgement to my PhD committee members, Mr. François Peres, Ms. Birgit Vogel-Heuser, Mr. Christophe Berenguer and Ms. Olga Fink for their evaluation of my PhD work, for their appreciation and interesting insight to complete and enrich the PhD contributions. Likewise, I would like to thank the members of my *comité de suivi* Mr. Christophe Berenguer and Mr. Romain Postoyan for their insights all along the PhD, they helped more than you can imagine.

On a less PhD-oriented note, I have to thank the MC department of ArcelorMittal Maizières-lès-Metz for their kindness. Those last four years have been intense but it was always a pleasure going to work thank to you all. Thank you to Armelle for all our talks while sharing the road. Thank you to Loann for the coffee breaks. Thank you to all the others for those great times Sophie, Morgan, Pauline, Vincent, Grace, Gwenaël, Miranda, Guillaume, Emmanuelle, Jean-Pierre, Pierre-Jean, Benoit, Morvan. And those who were there for a shorter time Patricia, Paul, Léa. A special thank to Gabriel, without whom none of this would have happened.

Similarly, a big thank to all the PhD's student of the CRAN who shared this journey with me. A sincere thought to those who are already done Loïc, Guilain, Paul. A big tap in the back to those who are still in it Gauthier, Alaaeddine. A cheerful shout out to those who are finishing along with me Thaï, Ayoub. A particular thank to Alaaeddine for sharing its code and model with me, as you can see it is the base of all further work. And last but not least, a great thank to Simon, for hosting me and all those times shared together!

Finally, I would like to warmly thank my friends and family.

Valentin, Pierre-Jean and Simon for still being there every week after all those years.

My sister, Valentine, for all that we share.

My mother, for still taking care of me after all those years.

Margaux, my love, for those wonderful years spend together and all those to come.

Celeste, my daughter. I can't wait to see you grow.

And my father, for everything.



*À mes deux étoiles,  
Celle dans le ciel et celle sur Terre.*



# Contents

<b>List of Tables</b>	<b>xvii</b>
<b>Résumé en français (French summary)</b>	<b>1</b>
1 Portée industrielle de la thèse . . . . .	1
1.1 Contexte industriel . . . . .	1
1.2 Objectifs industriels . . . . .	2
2 Approches scientifiques de la maintenance . . . . .	3
3 Questions scientifiques . . . . .	4
4 Contributions aux questions scientifiques . . . . .	6
4.1 Contributions aux pronostics . . . . .	6
4.2 Contributions à la prise de décision . . . . .	8
<b>Introduction</b>	<b>11</b>
1 Industrial scope of the PhD . . . . .	11
1.1 Industrial context . . . . .	11
1.2 Industrial objectives . . . . .	12
2 Scientific approaches of maintenance . . . . .	13
3 Scientific issues . . . . .	13
4 Contributions to the scientific issues . . . . .	15
4.1 Contributions to prognostics . . . . .	15
4.2 Contributions to decision making . . . . .	16
5 Organisation of the manuscript . . . . .	16
<b>1 Maintenance decision making optimization at ArcelorMittal: A planification challenge</b>	<b>21</b>
1.1 Introduction . . . . .	21
1.2 Industrial context - From steel-making descaling process towards a generalization as industrial objectives . . . . .	21
1.2.1 Hot Strip Mill (HSM) process . . . . .	22
1.2.2 Descaling use-case description and generalization . . . . .	25

Contents

1.2.3	ArcelorMittal maintenance constraints . . . . .	26
1.2.4	Maintenance formalization - toward industrial objectives for ArcelorMittal downscaling . . . . .	27
1.2.5	Conclusion - Industrial objectives . . . . .	29
1.3	Scientific context – Prognostics and Health Management (PHM) . . . . .	29
1.3.1	PHM paradigm definitions . . . . .	29
1.3.2	Prognostics . . . . .	33
1.3.3	Small datasets among data-driven prognostic . . . . .	34
1.3.4	Data Augmentation as an opportunity for prognostics . . . . .	35
1.3.5	Expert Knowledge . . . . .	41
1.3.6	Decision Making . . . . .	44
1.4	Conclusion: industrial and scientific issues list . . . . .	47
<b>2</b>	<b>Agnostic data augmentation for data-driven prognostics on <i>small datasets</i></b>	<b>49</b>
2.1	Introduction . . . . .	49
2.2	<i>Small datasets</i> test methodology: full datasets sub-sampling. . . . .	50
2.2.1	Sources of variance in <i>small datasets</i> evaluation . . . . .	51
2.2.2	Facing variance with multi-picks and multi-trainings . . . . .	52
2.3	Dataset and prognostics models . . . . .	53
2.3.1	Dataset requirements . . . . .	53
2.3.2	C-MAPSS turbofan public dataset . . . . .	54
2.3.3	LSTM-based model . . . . .	57
2.3.4	ROCKET and MiniROCKET prognostics models . . . . .	59
2.4	Prognostics performance loss on <i>small datasets</i> . . . . .	60
2.4.1	Determination of number of picks for further evaluations . . . . .	61
2.4.2	Prediction of performance loss due to <i>small dataset</i> . . . . .	64
2.5	Agnostic data augmentation for Time Series - Literature Survey . . . . .	65
2.5.1	Data augmentation for time series in the literature . . . . .	66
2.5.2	Random transformations family . . . . .	69
2.5.3	Pattern mixing family . . . . .	71
2.5.4	Decomposition family . . . . .	72
2.6	Single-transformation data augmentation . . . . .	74
2.6.1	Single-transformation data augmentation methodology . . . . .	74
2.6.2	Hyper-parameters optimization . . . . .	74
2.6.3	Single-transformation data augmentation results . . . . .	75
2.6.4	Analysis of hyper-parameters optimization results. . . . .	79
2.7	Multi-transformation data augmentation . . . . .	79
2.7.1	Multi-transformation methodology . . . . .	82

2.7.2	Multi-transformation evaluation perspective . . . . .	83
2.8	Conclusion . . . . .	83
<b>3</b>	<b>Knowledge-based data augmentation for prognostics</b>	<b>87</b>
3.1	Introduction . . . . .	87
3.2	Description and formalization of available knowledge on dataset . . . . .	88
3.2.1	Introduction to expert-knowledge for data augmentation . . . . .	88
3.2.2	“Expert” knowledge on turbofan systems for the C-MAPSS dataset . . . . .	89
3.2.3	Statistical knowledge on turbofan systems for the C-MAPSS dataset . . . . .	91
3.2.4	Knowledge formalization for data augmentation on the C-MAPSS dataset . . . . .	95
3.3	Implementation and evaluation of knowledge-based data augmentation for prognostics . . . . .	96
3.3.1	Knowledge-based data augmentation through signal exclusion and grouping . . . . .	96
3.3.2	Knowledge-based data augmentation through correlation matrices distance . . . . .	98
3.4	Conclusion . . . . .	101
<b>4</b>	<b>Impact of prognostics performances on maintenance decision-making</b>	<b>103</b>
4.1	Introduction . . . . .	103
4.2	ArcelorMittal decision-making constraints . . . . .	104
4.3	Prognostics-based maintenance decision-making cost . . . . .	105
4.3.1	Introduction to maintenance cost modeling . . . . .	105
4.3.2	Notations . . . . .	105
4.3.3	Preventive maintenance cost modeling . . . . .	106
4.3.4	Maintenance cost model based on prognostics . . . . .	108
4.4	Numerical application for ArcelorMittal descaling valve use-case . . . . .	114
4.4.1	Weibull distribution for component failure rate . . . . .	114
4.4.2	Assumptions and used data . . . . .	115
4.4.3	Simulation model . . . . .	116
4.4.4	Maintenance cost - Scheduled (preventive) maintenance . . . . .	117
4.4.5	Maintenance cost - Predictive maintenance . . . . .	118
4.4.6	Maintenance cost - Evaluation of the gains with RMSE optimization . . . . .	119
4.5	Current limits and perspectives . . . . .	122
4.6	Conclusion . . . . .	124
<b>5</b>	<b>Conclusion and perspectives</b>	<b>127</b>
5.1	Perspectives . . . . .	129
	<b>Annexes</b>	<b>131</b>
1	Student’s t-test values . . . . .	131
	<b>Bibliography</b>	<b>133</b>

*Contents*

# List of Figures

1.1	Illustration of the rolling process of the hot strip mill. On the diagram, rolls are represented in blue and the steel slab in grey. . . . .	23
1.2	Schematics of the hot rolling process. It begins on the left with the two reheating furnaces, heating slabs up to 1300°C. Then heating slabs go through stands to reduce their thickness from 10cm to 1.8mm. Finally, they are cooled down by the cooling table before being wound in the winders. . . . .	24
1.3	Illustration of the distributed conveying surface studied by Skima et al. 2019. The whole system is parallel in the sense that if one “tile” have failed, the function can still be ensured by using the other tiles. On the Figure, one example path is represented. If one of its tile were to fail, another equivalent path could be used. . . . .	24
1.4	Cross-sectional image of the scale on the surface of a steel slab. Steel is in white, the air is in black and the scale is in gray. . . . .	25
1.5	(a) shows a schema of the descaling of a slab. (b) is a picture of an ongoing descaling inside a descaling stand. Low picture quality is due to the constrained environment. It can be seen the nozzles and projected water. . . . .	26
1.6	Illustration of the PHM processes. Source of the image Atamuradov et al. 2017. . . . .	30
1.7	Illustration of the prognostics step of PHM and its inherent uncertainty. The RUL prediction is made at the time of detection of the degradation. The uncertainty on the prediction ends up with the resulting distribution of possible RUL, illustrated with the blue distribution. Source: Atamuradov et al. 2017. . . . .	32
1.8	Example of data augmentation transformations in the case of image processing. From one original sample of a butterfly picture, six new ones are generated by altering the pixels of the image without altering its label “butterfly”. Source Ahmad, Muhammad, and Baik 2017. . . . .	36
1.9	Taxonomy of time series data augmentation from Iwana and Uchida 2021. . . . .	38
1.10	Graphical examples of random magnitude transformations. Graph (a) shows an example of jittering while (b) shows an example of magnitude scaling. Original signals are in blue, augmented ones in red. . . . .	39
1.11	Graphical examples of random time transformations. Graph (a) shows an example of slicing while (b) shows an example of time warping. Original signals are in blue, augmented ones in red. . . . .	40
1.12	Interpolation (a) and extrapolation (b) between handwritten characters. Character (0,i) is interpolated/extrapolated with character (j,0) to form character (i,j), where i is the row number and j is the column number. Original characters are shown in bold. Image and text from DeVries and Taylor 2017a. . . . .	40

List of Figures

1.13	Graphical examples pattern mixing transformations. Graph (a) shows an example of interpolation while (b) shows an example of extrapolation. Original signals are in blue and green, augmented ones in red. . . . .	41
1.14	Suggestion matrix proposed by Bozkurt, Orak, and Tunçkaya 2021 as an expert suggestion system for temperature control decisions in steel-making blast furnace process. . . . .	44
2.1	Methodology flowchart. . . . .	50
2.2	Histogram of prognostics performances on a sub-samplings of a full dataset. The dataset is the C-MAPSS FD001, the model used is MiniROCKET, both described in 2.3. Sub-samplings size is 10%. . . . .	51
2.3	Histogram of prognostics performances on 50 sub-samplings of a full dataset. The dataset is the C-MAPSS FD001, the model used is MiniROCKET, both described in 2.3. Sub-samplings size is 25%. For each sub-sampling, the RMSE value is averaged over 5 model trainings. . . . .	52
2.4	Methodology flowchart with sources of variance identified. . . . .	52
2.5	Final flowchart representation of the methodology, including the multi picks and trainings. . . . .	53
2.6	Example of a run-to-failure trajectory from the C-MAPSS dataset (FD001). . . . .	55
2.7	Example of the RUL to predict on one trajectory. The “gold RUL” of 130 is seen at the begining of the trajectory. . . . .	56
2.8	Architecture of the LSTM-based model proposed and shared by Chaoub, Voisin, et al. 2021. This picture comes from the paper. . . . .	58
2.9	Synthetic schema of ROCKET structure, proposed by Dempster, Petitjean, and Webb 2020. The first step consists in convolution by a large amount of random convolution kernels. The second step is pooling of the convoluted signals with two specific functions (see main text). Last step is the prediction model done through Ridge Classifier. Image shared by <i>ROCKET</i> 2023. . . . .	59
2.10	Prognostics performances on the C-MAPSS dataset with (a) the LSTM-based model and (b) ROCKET for various sizes of sub-samplings (small datasets). Performances are measured with RMSE on the test set. . . . .	60
2.11	Distributions of RMSE on the test set for various percentage values for MiniROCKET on C-MAPSS FD001. For each percentage values, 20 to 60 independent sub-samplings are picked, and the model is trained multiple times. Each point corresponds to one training. Those distributions are used to evaluate normality (see Table 2.4). . . . .	63
2.12	Average RMSE of the LSTM-based model on the C-MAPSS dataset along the fitted curve for predicting performances. The last two points (75% and 100%) are not used for curve fitting. So they can be used for validation. . . . .	65
2.13	Example of data augmentation transformations in the case of image processing. From one original sample of a butterfly picture, six new ones are generated by altering the pixels of the image without altering its label “butterfly”. Source Ahmad, Muhammad, and Baik 2017. . . . .	66
2.14	Taxonomy of time series data augmentation from Iwana and Uchida 2021. The highlighted methods are the one selected for the study. The other ones are ignored as they are not adapted to prognostics. . . . .	68

2.15	Graphical examples of the random magnitude transformations selected for the study. Graph (a) shows an example of jittering with a high value of $p$ while (b) show an example of magnitude scaling. Original signals are in blue, augmented ones in red. Signals are from the C-MAPSS dataset. . . . .	69
2.16	Examples of cubic splines curves used by Um et al. 2017 to generate magnitude warping augmented trajectories by multiplying them with the original signal. The image was generated by Um et al. 2017 and is available here (GitHub). . . . .	70
2.17	Graphical examples of the random time transformations selected for the study. Graph (a) shows an example of slicing while (b) shows an example of time warping. Original signals are in blue, augmented ones in red. Signals are from the C-MAPSS dataset. . . . .	70
2.18	Graphical examples of the pattern mixing transformations selected for the study. Graph (a) shows an example of interpolation while (b) shows an example of extrapolation. Original signals are in blue and green, augmented ones in red. Signals are from the C-MAPSS dataset. . . . .	72
2.19	Interpolation (a) and extrapolation (b) between handwritten characters. Character (0,i) is interpolated/extrapolated with character (j,0) to form character (i,j), where i is the row number and j is the column number. Original characters are shown in bold. Image and text from DeVries and Taylor 2017a. . . . .	73
2.20	Schema of the data augmentation test process. Step 1 is the reduction of the size of the dataset into sub-samplings, repeated $n_{picks}$ times. Step 2 is the augmentation of each sub-samplings with one augmentation at a time (out of $m$ transformations), repeated $n_{augm}$ times per DA transformation. Step 3 is training the prognostics model with the $n_{picks} * n_{augm} * m$ obtained augmented subsets and measure their performances on the test set. . . . .	75
2.21	Performances of LSTM model trained on C-MAPSS dataset augmented with (a) interpolation and (b) time slicing, with various values of their hyper-parameters $p$ , given $n = 65$ . Those results were not obtained with cross-validation, and therefore should be considered with care as of their absolute value. . . . .	77
2.22	Cross validation performance gains obtained with data augmentation applied to 25% of C-MAPSS FD001 for ROCKET training. The evaluation is done on 10 independent sub-samplings of the dataset, optimization being done for each of them through cross-validation. (a) shows average performances obtained for each pick for each optimised DA transformation. (b) shows performance gain obtained for each pick for each optimised DA transformation. . . . .	78
2.23	Test performance gains obtained with data augmentation applied to 25% of C-MAPSS FD001 for ROCKET training. The evaluation is done on 10 independent sub-samplings of the dataset, using optimized hyper-parameters. (a) shows average performances obtained for each pick for each optimised DA transformation. (b) shows performance gain obtained for each pick for each optimised DA transformation. . . . .	80
2.24	Distribution of <b>interpolation</b> performances among sub-samplings, depending on parameters values, extrapolated from optimizer results. Each graph represents a sub-sampling from the C-MAPSS dataset. Best performances are brighter, colors are relative to each graph, the darkest and brightest blue are for the worst and the best performance for the given sub-sampling. Black points corresponds to evaluated values, contours are extrapolated around them. This Figure illustrates the variability of optimal values from a sub-sampling to another. . . . .	81

List of Figures

2.25	Schema of the multiple-data augmentation process, indicating in which order are applied the transformations. Each transformation uses as original trajectories to augment the trajectories supplied by the previous step in the chart. In the case of parallel transformations, generated trajectories are merged after application. . . .	83
2.26	Best results obtained on a <i>small dataset</i> for various data augmentation transformations. The <i>small dataset</i> is obtained as a 25% subset of the C-MAPSS dataset. Prognostics performances are measures as a RMSE (Root Mean Square Error, to be minimized). Prognostics is done with the ROCKET model. The box in red corresponds to the performances on the un-augmented dataset. . . . .	84
3.1	Schematic diagram showing the components and station numbers of the C-MAPSS engine. LPC: low pressure compressor, HPC: high pressure compressor, HPT: high pressure turbine, LPT: low pressure turbine. Source Frederick, DeCastro, and Litt 2007 . . . . .	89
3.2	Fitted exponential function on one signal of the C-MAPSS FD001 dataset. The fitted equation coefficient for this trajectory is $y(t) = -0.12 * \exp(.026 * t) + 8133$ . If this fitting would have been used for dimension reduction, the coordinates in the projected space would have been $(a, b, c) = (-0.12, 0.26, 8133)$ . . . . .	90
3.3	Fitted ARIMA estimator to one signal of the C-MAPSS FD001 dataset. . . . .	91
3.4	All the signals from the C-MAPSS dataset (FD001) plotted for eleven run-to-failure trajectories. Scales are not the same accross signals, but time axis is similar. . . .	92
3.5	Correlation matrix for C-MAPSS FD001 dataset. Values correspond to the absolute values of the correlations between signals. On the Figure (b) they are grouped by similarity. The empty lines and rows in Figure (a) correspond to signals always constant. The correlation is computed on the integrality of the 100 train trajectories of the dataset. . . . .	93
3.6	Hierarchical clustering applied to the C-MAPSS FD001 signals. Signals can be grouped by cutting the branches of the dendrogram at a selected height. . . . .	94
3.7	Plot of twelve trajectories of the two signals of the first group, sig_9 and sig_14 from C-MAPSS dataset. These two signals are highly correlated together and less with the rest. The signals on the plots have been scaled (using Standard Scaler) to set them to the same scale. . . . .	94
3.8	Plot of twelve trajectories of the two signals of the first group, sig_8 and sig_13 from C-MAPSS dataset. These two signals are highly correlated together and less with the rest. The signals on the plots have been scaled (using Standard Scaler) to set them to the same scale. . . . .	95
3.9	Plot of twelve trajectories of the three signals of the last group from C-MAPSS dataset. These three signals have no correlations with either other signals, due to their very noisy shape. The signals on the plots have been scaled (using Standard Scaler) to set them to the same scale. . . . .	95
3.10	Distribution of performances of data augmentation methods (a) through cross-validation and (b) on the test set of the C-MAPSS dataset with hyper-parameters optimized. Those results are obtained with the MiniROCKET model. Methods presented here are the random magnitude transformations (in red) and their knowledge-based counterparts (in blue). . . . .	97

3.11	Distributions of Correlation Matrix Distances (CMD) among 25% sub-samplings of the C-MAPSS dataset. Each box represents the distribution of CMDs computed among the considered sub-samplings and the 20 others. It is seen that distances are low, as could be expected. From these data it is seen that 90% of distances are below 0.0065 and 95% below 0.0085, with maximum value at $d_{max} = 0.020$ . . . .	99
3.12	Graphs of sample augmented trajectories through knowledge-based jittering from C-MAPSS dataset. Red signals are original (un-augmented) signals. Blue signals are augmented signals. Augmented signals on the figure (a) are rejected by the correlation matrices comparison. Their Correlation Matrix Distance (CMD) is $d \approx 0.030 > 0.02$ . Augmented signals on the figure (b) were accepted with the same method. Their Correlation Matrix Distance (CMD) is $d \approx 0.003 < 0.02$ . . .	99
3.13	Performances of CMD-based jittering on C-MAPSS dataset with MiniROCKET for various values of its $p$ hyper-parameter, compared to knowledge-based jittering. Knowledge-based jittering implements the knowledge-based signal selection by not augmenting constant and discrete signals. CMD-based jittering implements both signal selection and CMD-based Algorithm 1 of augmentation rejection if generated signals are too different of original ones. . . . .	100
3.14	Correlation Matrix Distance between augmented and original datasets with EK jittering for various values of $p$ . Values obtained on the C-MAPSS dataset with MiniROCKET and $n = 25$ . . . . .	100
4.1	Schema of the decision tree implied by Equation (4.2). . . . .	108
4.2	Maintenance decision tree integrating prognostics in decision-making for a single component. $Recall(T_p)$ is the recall indicator of the prognostics model. It corresponds to the true positive rate. $Specificity(T_p)$ is the specificity indicator of the prognostics model. It corresponds to the true negative rate. $f$ is the factor corresponding to the lost product remaining useful life when preventive maintenance is done "too soon" due to False Positive from the prognostics model. It is included in the equation by the impact of the Specificity on the average $T$ . . . . .	109
4.3	Schema of the basic decision tree linked to the application of prognostics to decision making. $T_i$ is the time of the decision, during a maintenance slot. $T_{step}$ is the maintenance slots period. . . . .	111
4.4	Cumulative distribution function (CDF) of the standard normal distribution, named $\Phi$ , for a law $\mathcal{N}(0, 1)$ . . . . .	112
4.5	Weibull distribution probability density function (PDF) depending on the two parameters $k$ (a) and $\lambda$ (b). Source: wikimedia.org . . . . .	114
4.6	Probability density function (PDF) and cumulative distribution function (CDF) of the Weibull distribution used to model the failure of the considered component (a valve). Vertical red dash lines correspond to the available maintenance stops. . .	116
4.7	Evolution of the preventive maintenance cost defined in Equation (4.2). Units are daily cost accordingly to arbitrary units defined earlier ( $C_p = 1,000$ and $C_c = 20,000$ ). Red lines correspond to the available maintenance stops. Red points correspond to the value that the preventive period can take. . . . .	118
4.8	Average daily cost of maintenance as a function of the RMSE. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ , $C_c = 20,000$ , $T_{step} = 21days$ , $T_{opt} = 4 \cdot 21days$ and a Weibull distribution with $k = 4$ , $lambda = 221$ ). . . . .	120

List of Figures

4.9 Relative improvement of average daily maintenance cost for a 10% or 20% improvement of the RMSE, as a function of the original RMSE. The graph can be read as follows: given an initial RMSE  $rmse = 13$ , if it was improved by 10% (getting to  $rmse = 11.7$ ), the average daily maintenance cost would improve by 0.6%. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21$  days,  $T_{opt} = 4 \cdot 21$  days and a Weibull distribution with  $k = 4$ ,  $lambda = 221$ ). . . . . 121

4.10 Average daily cost of maintenance as a function of the preventive period T. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21$ days,  $T_{opt} = 4 \cdot 21$ days and a Weibull distribution with  $k = 4$ ,  $lambda = 119$ ). . . . . 121

4.11 Evolution of the maintenance cost as a function of  $T$  and the  $RMSE$ . The white dash lines corresponds to available values for  $T$ , *i.e.* maintenance slots. The color corresponds to the maintenance cost value, purple corresponding to smaller cost values. . . . . 122

4.12 Average daily cost of maintenance as a function of the RMSE, with and without updated  $T_{opt}$ . Updated  $T_{opt}$  means that the optimal value of preventive maintenance period  $T$  is computed based on the cost considering prognostics, see Equation (4.10). The blue points corresponds to the optimal value of  $T$  associated to the  $rmse$  value.  $T$  can only take values proportionnal of the maintenance slots period  $T_{step}$ . This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21$ days,  $T_{opt} = 4 \cdot 21$ days and a Weibull distribution with  $k = 4$ ,  $lambda = 119$ ). . . . . 123

4.13 Average daily cost of maintenance as a function of the RMSE. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21$ days,  $T_{opt} = 4 \cdot 21$ days and a Weibull distribution with  $k = 4$ ,  $lambda = 119$ ). . . . . 124

# List of Tables

1.1	Classification of the papers reviewed in Gay, Benoit Iung, et al. 2021 among the defined categories. . . . .	43
2.1	C-MAPSS sub-datasets description. Based on Chaoub, Voisin, et al. 2021 . . . .	57
2.2	Values of the hyper-parameters of the LSTM-based model, as optimized by Chaoub, Voisin, et al. 2021 . . . . .	58
2.3	Summary of performances loss due to small datasets for all datasets and models. Performances loss are evaluated on the average <i>rmse</i> . Loss are exprimed as percentage of RMSE reduction compared to 100% of the dataset. . . . .	61
2.4	p-values of the D'Agostino and Pearson's normality tests for LSTM and ROCKET on various sub-samplings sizes of the C-MAPSS FD001 dataset. It is reminded that the null hypothesis is that a sample comes from a normal distribution. Therefore, a p-value greater than 0.05 means there are more than 95% chances that the samples follow a normal distribution. . . . .	63
2.5	Minimum number of picks needed to differentiate a performances improvement down to 10% among experiments . . . . .	64
2.6	Fitted coefficient of the model in equations (2.5) for the various models. . . . .	65
2.7	List of references for each DA transformations . . . . .	73
3.1	List of T-test p-values obtained between knowledge-based and standard version of augmentation methods, for various methods and evaluation sets. Those results were obtained on 25% sub-samplings of the C-MAPSS dataset with the MiniROCKET model. . . . .	96
4.1	Numerical application of decision-making maintenance cost when prognostics is applied, for two values of <i>rmse</i> . Those two values correspond to LSTM performances on the C-MAPSS dataset for various dataset sizes. . . . .	119

*List of Tables*

# Résumé en français (French summary)

## 1 Portée industrielle de la thèse

Cette thèse CIFRE est un projet commun entre ArcelorMittal et le laboratoire CRAN. ArcelorMittal est le leader mondial de la production d'acier, implanté dans des centaines de pays. Le laboratoire CRAN est un acteur majeur de la recherche scientifique en France et à l'étranger dans les domaines de l'automatisation, du pronostic et de la gestion de la santé. Cette thèse est le résultat d'une réflexion commune sur les problématiques de maintenance industrielle, auxquelles les deux acteurs sont confrontés, d'un point de vue industriel et scientifique.

### 1.1 Contexte industriel

La décarbonation, la réduction de la consommation d'énergie et la révolution de l'*Industrie 4.0* sont des sujets clés des tendances futures de l'industrie. Leur importance s'est encore accrue ces dernières années dans le contexte du réchauffement climatique et de l'augmentation des prix de l'énergie due à la crise géopolitique.

L'amélioration de l'efficacité des processus est une solution clé pour aborder la décarbonation et la réduction de l'énergie dans l'industrie manufacturière. En effet, des processus plus efficaces consomment moins d'énergie, de matériaux, de temps et d'outils pour produire le même produit. L'amélioration de l'efficacité des processus conduit à une réduction des coûts et donc à des gains de compétitivité, elle a donc toujours été l'objectif principal de l'industrie (Porzio et al. 2013).

La tendance industrielle actuelle de l'*Industrie 4.0* est le levier clé actuel pour aborder l'optimisation des processus. Les installations industrielles sont en transition vers l'*Industrie 4.0* dans le but d'exploiter les données disponibles pour de nouvelles avancées techniques. La disponibilité croissante des données et les développements récents d'algorithmes puissants pour les exploiter justifient l'exploration de ce paradigme pour relever les défis industriels (Sahli, Evans, and Manohar 2021; Manan et al. 2017).

L'amélioration de l'efficacité des processus est soutenue par deux leviers principaux, d'abord l'amélioration du temps d'exécution des processus, ensuite le maintien des systèmes en conditions opérationnelles. Cette deuxième étape consiste principalement à réduire les temps d'arrêt des systèmes. Ce dernier point est particulièrement important pour la fabrication de l'acier, où certaines machines (comme les fours de réchauffage ou même les hauts fourneaux) ne sont pas arrêtées pendant les temps d'arrêt, ce qui entraîne une forte consommation d'énergie et des émissions de carbone pendant les périodes improductives (Backman, Kyllönen, and Helaakoski 2019; Merkert et al. 2015). La réduction des temps d'arrêt dans les aciéries est donc un objectif industriel majeur pour ArcelorMittal, car ils ont un impact particulier sur la consommation

d'énergie, la réduction des coûts et les émissions de carbone.

Pourtant, jusqu'à 50% des temps d'arrêt actuels de certains processus dans les aciéries sont encore des arrêts non planifiés dus à des défaillances inattendues entraînant une maintenance curative des systèmes. Ces temps d'arrêt non souhaités pourraient être réduits grâce à une maintenance efficace. L'amélioration de la maintenance est donc l'un des **principaux objectifs industriels** pour la recherche future de l'industrie de la fabrication de l'acier (R. X. Gao et al. 2020).

En outre, l'amélioration de la maintenance peut contribuer à réduire l'état de dégradation des systèmes, qui peut entraîner une dégradation de la qualité du produit. Des pronostics de maintenance plus précis permettent également de mieux anticiper les temps d'arrêt nécessaires et donc d'adapter la planification de la production pour les inclure.

**L'objectif industriel principal** de ce travail de doctorat est l'optimisation de la prise de décision en matière de maintenance industrielle par l'exploitation des sources d'information disponibles, *i.e.* les données industrielles et les connaissances d'experts, sous les contraintes industrielles présentées par le contexte de la sidérurgie. Cet objectif est détaillé et divisé en sous-objectifs dans la section suivante.

## 1.2 Objectifs industriels

La stratégie actuelle de maintenance des lignes d'acier est basée sur une maintenance préventive régulière. Elle consiste en des opérations de maintenance systématique régulières sur les systèmes en fonction de leur durée moyenne avant défaillance. En raison de la caractéristique statique de la stratégie de maintenance préventive systématique, certaines défaillances se produisent encore, ce qui donne lieu à une maintenance curative d'urgence. Comme les lignes d'aciérie sont principalement des systèmes en série, cela nécessite l'arrêt de toute la ligne de production et est donc extrêmement coûteux.

Les opérations de maintenance systématique tiennent compte de l'évolution statistique de l'état du système. Cela conduit à l'apparition de défaillances précoces. Pour éviter ces défaillances précoces, une stratégie plus personnalisée doit être envisagée. En effet, les décisions de maintenance doivent être optimisées en tenant compte, au moment de la décision, de l'état actuel et prévu du système. En outre, une approche dynamique de la maintenance pourrait être envisagée, y compris le regroupement des opérations ou les opérations de maintenance opportunistes (Do 2015).

**Le premier objectif industriel** de cette thèse est la proposition d'une prise de décision dynamique en matière de maintenance afin de réduire les occurrences de défaillances inattendues et donc les opérations de maintenance curative.

L'évolution de la maintenance préventive vers une stratégie dynamique se fait par le biais de la maintenance prédictive. La maintenance prédictive vise à optimiser les décisions de maintenance en se basant sur les états de santé actuels et futurs des systèmes plutôt que sur des règles statistiques (Gouriveau, Medjaher, and Zerhouni 2016). Par conséquent, la prise de décision dynamique est basée sur l'état de santé du système.

Dans le cadre du paradigme *Industrie 4.0* susmentionné, la maintenance prédictive tire parti des données disponibles sur les systèmes. Les données sont largement disponibles dans les aciéries sous la forme d'enregistrements de capteurs, principalement exploités pour le contrôle des outils ou l'évaluation de la qualité des produits. La maintenance prédictive nécessite la détermination de l'état de santé actuel des systèmes ainsi que la prédiction des états futurs. Ces tâches tirent parti de l'industrie 4.0 pour sa capacité à exploiter les informations subsidiaires contenues dans les enregistrements de données. L'exploitation des données dans les pronostics industriels est un sujet de recherche en cours (Leukel, González, and Riekert 2021).

**Le deuxième objectif industriel** que ce doctorat vise à aborder est l'exploitation des

données de processus industriels pour mettre en œuvre la maintenance prédictive dynamique sur les systèmes industriels.

Les approches basées sur les données (comme l'apprentissage automatique) invoquées par l'*Industrie 4.0* nécessitent de grandes quantités de données pour atteindre leur plein potentiel. Grâce à la troisième révolution industrielle et encore plus à la quatrième, les installations industrielles fournissent de grandes quantités de données historiques sous la forme d'enregistrements de capteurs à long terme. Ces grands ensembles de données définissent le contexte *Big Data* des applications axées sur les données (Jiao, K. Peng, and J. Dong 2021). Ils garantissent l'efficacité des algorithmes axés sur les données à long terme.

Toutefois, parmi ces enregistrements, les algorithmes guidés par les données nécessitent des exemples étiquetés pour apprendre, que les installations industrielles ne parviennent pas à fournir en grande quantité dans le cas de la maintenance prédictive, ce qui se traduit par des petits ensembles de données (connus sous le nom de *small datasets*). En effet, dans le cas de la maintenance prédictive guidée par les données, les exemples étiquetés (d'apprentissage) consistent en des enregistrements de "pannes" des systèmes. De tels événements sont rares sur les systèmes pour lesquels des stratégies de maintenance préventive sont déjà mises en œuvre. Le manque d'exemples étiquetés concerne également les systèmes récemment installés, pour lesquels les enregistrements historiques ne sont pas encore disponibles. Les performances de l'apprentissage automatique sont liées à la taille des ensembles de données et les ensembles de données plus petits réduisent les performances de prédiction (Catal and Diri 2009). Néanmoins, il reste important de fournir des solutions de maintenance prédictive à ces systèmes dès que possible.

**Le troisième objectif industriel** que ce doctorat vise à aborder est le problème de la mise en œuvre sur de nouveaux systèmes auquel est confrontée la maintenance prédictive industrielle. Ce problème est étendu aux systèmes récemment instrumentés.

## 2 Approches scientifiques de la maintenance

La maintenance prédictive industrielle est abordée dans la littérature scientifique par la communauté Prognostics and Health Management, à laquelle le laboratoire CRAN contribue largement.

La maintenance prédictive a été formalisée dans le paradigme de la *Prognostics and Health Management* (PHM) sous la forme d'un processus en sept étapes. Les sept étapes proposées sont les suivantes : acquisition des données, traitement, évaluation de l'état, diagnostic, pronostic, prise de décision et interface homme-machine. Chaque étape fournit des résultats pour la suivante et une mise en œuvre complète de la maintenance prédictive nécessite la réalisation de toutes ces étapes (Atamuradov et al. 2017).

Parmi ces étapes de la gestion de la santé prédictive, les travaux de ce doctorat se concentrent sur le *pronostic* et la *prise de décision*. Le pronostic de l'état de santé consiste à prédire l'évolution future de l'état de santé du système. La prédiction est basée sur l'état de santé actuel fourni par l'étape de *diagnostic*, sur les données disponibles et sur l'utilisation future prévue du système. Ce doctorat se concentre sur le pronostic en tant que prédiction de la durée de vie utile restante (RUL), *i.e.* temps restant avant que le système ne soit pas en mesure d'assurer ses fonctions correctement (Gouriveau, Medjaher, and Zerhouni 2016). Cette définition spécifie le pronostic comme un problème de régression.

Trois approches principales sont disponibles pour la prédiction de la durée de vie utile : le pronostic basé sur un modèle, le pronostic basé sur des données et le pronostic hybride. La première nécessite des modèles physiques du système pour extrapoler l'état de santé à partir de KPI fondés sur la physique. Cette approche est efficace lorsque des modèles physiques avancés sont disponibles sur les systèmes. Lorsque de tels modèles ne sont pas disponibles, des approches

basées sur les données sont utilisées. Le modèle de dégradation est alors appris à partir des enregistrements historiques du comportement du système (Ellefsen 2019). Les approches hybrides combinent des modèles et des méthodes basées sur les données.

Les informations RUL sont ensuite fournies à l'étape de *prise de décision*, dont l'efficacité des décisions est liée à la précision du pronostic. La prise de décision en matière de maintenance dynamique dépend de la RUL prévue et de la planification de la production à venir. Les décisions sont prises pour minimiser le coût global de la maintenance (Bougacha, Varnier, and Zerhouni 2020). De nombreux leviers peuvent être utilisés pour optimiser la prise de décision, notamment le regroupement des opérations (Do 2013) et la maintenance opportuniste (Benoit Iung, Do, et al. 2016; Derigent et al. 2009).

Le travail de cette thèse se concentre sur le pronostic et la prise de décision en raison du positionnement des objectifs industriels susmentionnés. La section suivante présente les questions scientifiques découlant des objectifs industriels parmi les travaux des communautés scientifiques.

### 3 Questions scientifiques

Les objectifs industriels sont placés dans le contexte scientifique de la PHM. Le **premier** objectif industriel se concentre sur l'optimisation de la prise de décision en matière de maintenance sur la base des résultats des pronostics. Les objectifs industriels **deuxième** et **troisième** se concentrent sur l'exploitation des données industrielles pour la mise en œuvre et l'optimisation des pronostics de maintenance. Les questions scientifiques abordées par ce doctorat concernent donc ces deux étapes du processus PHM. Suivant l'ordre du processus PHM, les premières questions scientifiques portent sur le pronostic et les suivantes sur la prise de décision. Les objectifs industriels ont été déroulés de l'échelle du système à l'échelle de la maintenance, ce qui donne un ordre opposé, commençant par les considérations de prise de décision et finissant par les pronostics. Ces deux ordres sont opposés et, par conséquent, la première question scientifique traite du dernier objectif industriel, et ainsi de suite.

Le **troisième** objectif industriel met l'accent sur les *small datasets* dans le domaine de la maintenance industrielle. Il concerne la mise en œuvre de pronostics guidés par les données dans le cadre des travaux de PHM. Le pronostic basé sur les données appris sur de petits ensembles de données souffrent de performances inférieures à celles attendues (Zhang et al. 2019).

Les méthodes de pronostic basé sur les données sont bien étudiées dans la littérature. De nombreux articles proposent des modèles axés sur les données pour les pronostics (Ellefsen 2019). Du simple k-means dans Diez-Olivan 2017. Jusqu'aux modèles avancés tels que les auto-encodeurs dans Yu, I. Y. Kim, and Mechefske 2019, ou les LSTM dans Chaoub, Voisin, et al. 2021 ou même les modèles basés sur la convolution dans Z. Wu and K. Wu 2022.

Peu d'études se concentrent sur l'impact de la taille des ensembles de données sur les performances des pronostics. Par exemple Vapnik, Levin, and Cun 1994 a proposé une formule pour lier la performance du modèle à la taille de l'ensemble de données à des fins de classification.

Ainsi, la **première question scientifique** que ce doctorat vise à aborder est de formuler le lien entre la taille de l'ensemble de données et les performances du pronostic pour prédire l'amélioration des performances apportée par des données supplémentaires.

Le **second** objectif industriel met l'accent sur l'exploitation des données industrielles pour mettre en œuvre la maintenance prédictive. Il concerne également la mise en œuvre de pronostics fondés sur des données dans le cadre des travaux de PHM.

Malgré des performances moindres sur les petits ensembles de données, la mise en œuvre de la PHM reste pertinente sur les systèmes industriels et les performances des pronostics doivent être optimisées dans toutes les situations pour soutenir une prise de décision efficace. En effet, la mise

en œuvre de PHM peut améliorer l'efficacité de la maintenance même lorsque les performances des pronostics sont faibles (Skima et al. 2019). En outre, les *small dataset* dans le contexte industriel peuvent également correspondre à une situation temporaire, comme dans les cas d'utilisation des "nouveaux systèmes" mentionnés précédemment. Les performances des pronostics doivent être améliorées autant que possible afin de fournir les meilleures prédictions possibles pour la prise de décision.

Les performances des méthodes basées sur les données peuvent être améliorées par l'application de l'augmentation des données (Iwana and Uchida 2021). Elle est largement utilisée dans le traitement des images pour améliorer les performances et la résilience des algorithmes d'apprentissage automatique, voir par exemple Shorten and Khoshgoftaar 2019.

Toutefois, l'augmentation des données dans le cas des pronostics sur petits ensembles de données n'a pas encore fait l'objet d'une étude approfondie. Elle a été étudiée en profondeur pour les séries temporelles (comme disponible pour le pronostic) dans Iwana and Uchida 2021. Mais ces études se concentrent sur les applications de classification et peu sur l'impact des petits ensembles de données.

Ainsi, la **deuxième question scientifique** que ce doctorat vise à aborder est l'exploitation et l'optimisation de l'augmentation des données sur les données de séries temporelles pour améliorer les performances de pronostic lorsqu'elles sont appliquées à des *petits ensembles de données* recueillis à partir d'applications industrielles.

Une autre approche prometteuse pour améliorer les performances des pronostics sur les systèmes industriels affectés par des petits ensembles de données est l'exploitation de la connaissance experte. Les connaissances expertes peuvent compenser les informations manquantes dans les données en raison de la dimension réduite des ensembles de données (Chao et al. 2020). Le contexte industriel de cette étude doctorale fournit des connaissances sur les systèmes de la part d'experts industriels (opérateurs, ingénieurs, chercheurs de R&D, ...).

L'intégration des connaissances dans les algorithmes guidés par les données peut améliorer les performances, en particulier sur les petits ensembles de données où l'information est manquante dans les ensembles de données. Nous avons proposé un bref examen de l'intégration des connaissances pour les pronostics dans l'article Gay, Benoit Iung, et al. 2021.

L'intégration des connaissances pour les pronostics a été étudiée par le biais de modèles basés sur la physique, mais pas encore par l'adaptation de l'augmentation des données. Par exemple, Xiahou, Zeng, and Liu 2021 utilise les connaissances d'experts pour les pronostics avec un modèle hybride. L'augmentation des données a été utilisée avec des connaissances d'experts par Pan et al. 2022 pour une application de classification.

Ainsi, la **troisième question scientifique** que ce doctorat vise à aborder est l'incorporation de connaissances expertes dans les méthodologies d'augmentation des données afin d'améliorer encore les performances et la fiabilité des pronostics pour les applications industrielles.

Enfin, le **premier** objectif industriel se concentre sur l'optimisation de la prise de décision en matière de maintenance sur la base des pronostics fournis. Il concerne le processus de prise de décision post-prognostique parmi les travaux de PHM (Wesendrup and Hellingrath 2020).

L'efficacité de la prise de décision est liée à la performance des pronostics et est influencée à la fois par sa dégradation due aux petits ensembles de données et par son amélioration potentielle due aux deux objectifs scientifiques précédents. La dégradation des performances des pronostics induite par les petits ensembles de données dégrade le processus de prise de décision, qui repose sur la prédiction de la RUL. Si les performances des pronostics deviennent trop faibles, la prise de décision dynamique pourrait être moins performante que des stratégies préventives plus simples (K. Nguyen and Do 2019). Inversement, l'amélioration des performances du pronostic améliore la prise de décision.

Enfin, les performances des pronostics seront commentées à travers l'impact qu'elles ont sur

la prise de décision en matière de maintenance. Les coûts de maintenance sont calculés dans la littérature sur la base de la maintenance préventive.

Ainsi, la **quatrième question scientifique** que cette thèse vise à aborder est l'évaluation de l'impact des performances des pronostics sur la prise de décision dans le contexte industriel fourni par les cas d'utilisation d'ArcelorMittal, afin d'évaluer les gains de décision obtenus par l'amélioration des pronostics grâce à l'augmentation des données. Cette dernière question est examinée à la lumière de l'amélioration du pronostic visée par les deux premiers objectifs.

## 4 Contributions aux questions scientifiques

Les contributions de cette thèse sont directement proposées comme réponses aux questions scientifiques susmentionnées. Elles sont classées en deux catégories, contribution au pronostic et contribution à la prise de décision. Toutefois, il convient de noter que les contributions aux pronostics tendent à améliorer les performances des pronostics et sont donc destinées à améliorer la prise de décision. Par conséquent, la séparation entre ces deux catégories n'est pas aussi forte que ce qui suit pourrait le laisser entendre.

### 4.1 Contributions aux pronostics

Les deux premières contributions de la thèse consistent à évaluer l'impact du *petits ensembles de données* sur les performances des pronostics et à étudier la capacité de l'augmentation agnostique des données comme solution pour les améliorer sans utiliser la connaissance du système. Elles sont adressées dans le chapitre 2. Elles sont évaluées sur le dataset C-MAPSS avec deux modèles de pronostic, un modèle basé sur LSTM agissant comme l'état de l'art sur le dataset et ROCKET, un modèle prometteur considéré pour accélérer le calcul des résultats. À cette fin, des méthodologies de test approfondies sont proposées pour simuler les *petits ensembles de données* et évaluer l'augmentation de données avec un bon niveau de confiance dans les résultats.

Tout d'abord, les *petits ensembles de données* sont obtenus en tant que sous-échantillonnages aléatoires indépendants du dataset complet. Le sous-échantillonnage étant une source de variance des résultats, il est répété plusieurs fois pour garantir la représentativité des résultats. De même, l'initialisation du modèle étant une source de variance, en particulier pour les petits ensembles de données, l'initialisation et l'apprentissage du modèle sont également répétés plusieurs fois pour chaque sous-échantillon.

Deuxièmement, les méthodes d'augmentation de données sont optimisées par validation croisée pour correspondre à un usage industriel. En effet, les méthodes d'augmentation des données considérées sont paramétrées par deux hyper-paramètres. Leur impact sur les performances du pronostic étant confirmé, ils sont optimisés pour chaque sous-échantillon. Cette optimisation se fait par une approche de validation croisée, l'ensemble de test ne devant pas être utilisé pour une telle tâche. Un ensemble de test est finalement utilisé pour l'évaluation des performances. Cette optimisation est réalisée pour les méthodes appliquées individuellement.

Troisièmement, un ordre est défini pour les méthodes d'augmentation de données multiples appliquées simultanément. En effet, l'ordre d'application de plusieurs méthodes a un impact sur leurs performances finales. Un ordre fixe basé sur l'interaction entre les méthodes est défini afin de réduire le risque de dégradation des performances.

Une première contribution relie avec succès les performances des pronostics à la taille des ensembles de données par le biais d'une équation. L'équation proposée, inspirée par Vapnik, Levin, and Cun 1994, est écrite dans l'équation (2.5). Elle peut être utilisée pour prédire le gain de performance attendu de données supplémentaires lorsque seuls *petits ensembles de données* sont disponibles.

Une deuxième contribution confirme l'amélioration des performances des pronostics par l'application d'une augmentation individuelle des données agnostiques jusqu'à 10%. Toutes les méthodes sélectionnées ne permettent pas d'obtenir des améliorations de performances similaires, certaines d'entre elles entraînant même une baisse des performances. De grandes améliorations sont systématiquement obtenues sur la validation croisée, mais elles ne sont pas systématiquement observées sur l'ensemble de test (qui sert de données futures pour les systèmes industriels). Enfin, une ouverture est proposée sur l'application simultanée de plusieurs méthodes d'augmentation de données. Tous ces résultats sont résumés dans la Figure 2.26.

Enfin, ce chapitre met en évidence la prudence nécessaire lors de l'application de l'augmentation des données aux systèmes industriels, qui peut être abordée avec l'exploitation des connaissances d'experts sur les systèmes. Ces résultats soulignent l'importance et la difficulté de la sélection et de l'optimisation des méthodes d'augmentation de données lorsqu'elles sont appliquées à un petit ensemble de données. Des propositions visant à résoudre ce problème sont mentionnées dans Fink et al. 2020. L'une d'entre elles est l'exploitation de connaissances supplémentaires grâce aux experts lorsqu'elles sont disponibles dans les systèmes industriels. Par exemple, les deux méthodes les plus efficaces identifiées sont la gigue et l'interpolation, qui sont cohérentes avec la physique du système. En effet, la gigue ajoute du bruit à des données déjà bruyantes, tandis que l'interpolation génère de nouvelles trajectoires en mélangeant les trajectoires existantes, ce qui augmente leur potentiel de correspondance avec les données réelles.

Par conséquent, la troisième contribution de la thèse pour le pronostic est l'intégration des connaissances d'experts dans l'augmentation des données. Deux approches sont proposées. Tout d'abord par le biais de connaissances expertes telles que les modèles physiques, l'analyse fonctionnelle ou l'analyse des dysfonctionnements. Deuxièmement, les connaissances statistiques extraites des données. Cette contribution est adressée dans le chapitre 3.

Ce chapitre porte sur l'intégration des connaissances expertes dans les méthodes d'augmentation de données afin d'améliorer leurs performances sur de petits ensembles de données pour les systèmes industriels. Cette approche est soutenue par la littérature. Elle présente un potentiel élevé car les performances sur les petits ensembles de données dépendent des nouvelles informations apportées par les données supplémentaires dans le processus d'apprentissage. Par conséquent, l'optimisation des méthodes d'augmentation des données à l'aide de connaissances d'experts peut accroître les informations ajoutées. Dans ce chapitre, les connaissances sont exploitées à partir de deux sources : les connaissances d'experts et les connaissances statistiques.

Les connaissances sur l'ensemble de données C-MAPSS utilisé sont rassemblées à partir de sources expertes et d'analyses statistiques afin de proposer des améliorations des méthodes d'augmentation des données. On trouve peu de connaissances d'experts dans la littérature concernant cet ensemble de données basé sur la simulation.

Les connaissances statistiques sont donc exploitées pour proposer trois améliorations des méthodes d'augmentation de données basées sur la magnitude ("jittering" et "magnitude scaling"). Premièrement, les signaux constants et discrets sont exclus des méthodes d'augmentation de données sur la base des résultats de l'analyse des données. Deuxièmement, les signaux sont regroupés sur la base de leurs corrélations afin d'appliquer des facteurs de mise à l'échelle de l'amplitude par groupes plutôt que par signaux, dans un souci de cohérence. Troisièmement, les trajectoires augmentées sont validées en exploitant la distance entre la matrice de corrélation avant et après l'augmentation pour rejeter les augmentations aberrantes sur la base d'un seuil.

Les trois améliorations proposées donnent des résultats mitigés, sans amélioration ni dégradation notables des performances, bien qu'elles aient amélioré la résilience de l'augmentation de données. Les approches sont testées à l'aide de la méthodologie d'essai proposée dans le précédent chapitre 2. L'exclusion de signaux ne modifie pas les performances dans les conditions testées (MiniROCKET appliqué à 25% de C-MAPSS). Le regroupement des signaux n'a pas non plus

d'impact sur les performances. Cependant, la validation de l'augmentation des données basée sur la distance améliore la résistance de l'augmentation des données à la réduction des performances (notamment en raison des valeurs des hyperparamètres).

On peut espérer qu'en améliorant la résistance des modèles de pronostic à l'entraînement, l'exploitation des connaissances les rendra plus aptes à soutenir la prise de décision en matière de maintenance. En effet, comme indiqué dans l'introduction générale, la prise de décision nécessite des prévisions de pronostic précises ainsi qu'une réduction de l'incertitude sur le pronostic. La réduction des risques d'ajustement inefficace des modèles et de surajustement des ensembles d'entraînement garantit à la prise de décision que la documentation des données ne peut pas dégrader les performances des pronostics. Par conséquent, la prise de décision peut être plus fiable.

Cette tentative infructueuse d'exploitation des connaissances des experts a conduit à la formulation de perspectives pour une exploitation plus poussée des connaissances des systèmes en vue d'une augmentation des données. L'analyse fonctionnelle des systèmes pourrait être utilisée pour définir le lien entre les signaux et guider ainsi l'augmentation des données en générant de nouveaux signaux plus cohérents avec le comportement réel du système. De la même manière que le regroupement des signaux effectué après la connaissance statistique, mais avec l'appui de la connaissance et des modèles de système. La connaissance pourrait également être extraite de l'analyse des dysfonctionnements, comme HAZOP, pour guider la génération de méthodes d'augmentation des données. L'article récent Han et al. 2022 propose par exemple d'utiliser des graphes de causalité pour l'intégration des connaissances d'experts, qui pourraient être utilisés pour représenter les informations contenues dans l'analyse (dis)fonctionnelle. Cela serait particulièrement efficace pour faire face à de nouveaux modes de dégradation, qui ne sont pas ou peu représentés dans les ensembles de formation. Cette question n'a pas été prise en compte dans ce travail, qui s'est concentré sur des ensembles de données concernant un seul mode de dégradation.

D'autres améliorations pourraient également être mises en œuvre dans les méthodes fondées sur la connaissance proposées. Les matrices de corrélation ont été utilisées pour le regroupement des signaux et l'évaluation de la divergence des trajectoires nouvellement générées. Les variations des matrices de corrélation pourraient être utilisées pour identifier les trajectoires qui devraient varier le plus afin de concentrer l'augmentation des données sur celles-ci. Cette approche devrait permettre d'appliquer l'augmentation des données à une échelle "microscopique" plutôt que "macroscopique" comme le font les transformations agnostiques.

## 4.2 Contributions à la prise de décision

Les contributions à la prise de décision consistent en l'intégration des résultats du pronostic dans la modélisation des coûts de maintenance et en l'évaluation de l'impact du pronostic sur les coûts de la décision de maintenance. Cette contribution est adressée dans le chapitre 4. Ce chapitre vise à évaluer l'impact des performances du pronostic sur le coût de maintenance dans le contexte inspiré des usines d'ArcelorMittal et de ses contraintes spécifiques.

Ce chapitre propose une formalisation du pronostic développé précédemment par le biais d'une modélisation statistique. Le pronostic  $\widehat{rul}$  est modélisé comme une variable stochastique obéissant à une distribution normale de moyenne  $\overline{rul}$  la RUL "réelle" et de variance la  $rmse$  du modèle de pronostic. La défaillance du composant est ensuite modélisée comme une variable stochastique obéissant à une loi de Weibull.

Ces formalisations du pronostic et du contexte sont utilisées pour définir le coût moyen de maintenance sur la base de considérations statistiques. Cette formulation aboutit à l'expression du coût dans l'équation (4.6). Cette définition est basée sur l'expression du Recall et de la Specificity

du pronostic. Dans le contexte considéré, et sur la base d'un pronostic formalisé sous la forme d'une régression, ces deux éléments sont développés en profondeur. Elles sont définies dans les équations (4.7) et (4.8).

En parallèle, un modèle de simulation a été développé pour vérifier/comparer les résultats théoriques et assurer la pertinence des formulations. Il est développé en Python à l'aide d'une méthode incrémentale et est exécuté des dizaines de milliers de fois pour approximer les coûts de maintenance. Il valide avec succès les valeurs prédites par les modèles pour des valeurs considérées de *rmse* pour lesquelles le modèle de pronostic aurait été rejeté. Les résultats de l'équation ne commencent à diverger des valeurs de simulation que pour de grandes valeurs de *rmse*  $> 60$ , qui correspondent à des valeurs non réalistes de *rmse*. Par conséquent, la formulation est considérée comme valide.

Enfin, la formulation du coût de maintenance basée sur les caractéristiques du pronostic (y compris *rmse*) est utilisée pour calculer le gain en coût de maintenance (i) en appliquant une maintenance prédictive basée sur le pronostic plutôt qu'une maintenance préventive et (ii) en améliorant les performances du pronostic comme proposé dans les chapitres 2 et 3 (*i.e* en améliorant sa *rmse*). Les expériences sont paramétrées avec des valeurs numériques basées sur une vanne du système de détartrage d'ArcelorMittal, décrit en détail dans le chapitre 1. Il est démontré que la mise en œuvre de la maintenance prédictive réduit les coûts de maintenance jusqu'à 18–20%. Enfin, il est démontré que l'amélioration des performances de pronostic de 10% réduit les coûts de maintenance jusqu'à 1%. L'amélioration des performances de pronostic de 20% réduit les coûts de maintenance de 2%. Dans le chapitre 2, les performances de pronostic ont été améliorées de 10% grâce à l'augmentation des données.

Enfin, l'impact de la mise à jour de la période de maintenance préventive sur l'efficacité du pronostic est également pris en compte. En effet, la période de maintenance préventive était initialement fixée pour optimiser le coût de la maintenance préventive. Suite à l'implémentation du pronostic, la période de maintenance préventive peut être mise à jour pour être effectuée moins souvent et profiter de la capacité du pronostic à éviter les défaillances. Cette optimisation dynamique de la période permet de réduire encore davantage les coûts de maintenance.

*Résumé en français (French summary)*

# Introduction

## 1 Industrial scope of the PhD

This CIFRE PhD is a joint project between ArcelorMittal and the CRAN laboratory. ArcelorMittal is the worldwide leader of steel manufacturing, located in hundreds of countries. The CRAN laboratory is a main actor of the scientific research in France and abroad in the topics of automation as well as Prognostics and Health Management. This PhD is the result of a joint consideration for industrial maintenance issues, which both actors face, from an industrial and scientific point of view.

### 1.1 Industrial context

Manufacturing decarbonization, energy reduction and the *Industry 4.0* revolution are key topics of the industry future trends. Their importance increased even more those last few years within the context of global warming and energy price increase due to geopolitics crisis.

The improvement of processes efficiency is a key solution to address both decarbonization and energy reduction for manufacturing. Indeed, more efficient processes consumes less energy, materials, time and tools to produce the same product. Process efficiency improvement leads to cost reduction and therefore concurrency gains, it has therefore always been the main goal of industry (Porzio et al. 2013).

The current industrial trend of *Industry 4.0* is the current key lever to tackle process optimization. Industrial plants are transitioning toward *Industry 4.0* with the purpose to exploit available data for new breakthrough applications. The increasing availability of data and recent developments of powerful algorithms to exploit them justifies the exploration of this paradigm to face industrial challenges (Sahli, Evans, and Manohar 2021; Manan et al. 2017).

Process efficiency improvement is supported by two main levers, firstly process runtime improvement, secondly maintaining systems in operational conditions. This second step mostly consists in the reduction of systems downtime. The last is particularly important for steel manufacturing, where some machines (like reheating furnaces or even blast-off furnaces) are not stopped during downtimes, resulting in high energy consumption and carbon emissions for unproductive times (Backman, Kyllönen, and Helaakoski 2019; Merkert et al. 2015). Downtime reduction in steel plants is therefore a major industrial objective for ArcelorMittal as they are particularly impactful on both energy consumption, cost reduction and carbon emissions.

Yet, up to 50% of current downtime on some processes in steel plants are still un-planned stops due unexpected failure leading to curative maintenance of the systems. These un-wanted stop times could be reduced with efficient maintenance. Maintenance improvement is thus one of the **main industrial objectives** for the future research of the steel manufacturing industry (R. X. Gao et al. 2020).

Moreover, maintenance improvement can help reduce the degradation state of systems, which can lead to product quality degradation. More precise maintenance prognostics also allow for a better anticipation of required downtimes and thus help to adapt production planing to include them.

**The main industrial objective** of this PhD's work is the industrial maintenance decision-making optimization through the exploitation of the available sources of information, *i.e.* industrial data and expert knowledge, under the industrial constraints presented by the steel-making context. This objective is further detailed and split in sub-objectives in the following section.

## 1.2 Industrial objectives

Current maintenance strategy on steel lines is based on regular preventive maintenance. It consists in regular systematic maintenance operations on systems based on their average time before failure. Due to the static characteristic of preventive systematic maintenance strategy, some failures are still occurring, resulting in emergency curative maintenance. As steel plant lines are mainly series system, it requires the stops of the whole producing line and is therefore extremely costly.

Systematic maintenance operations consider the statistical health evolution of the system. This leads to the occurrence of early failures. To avoid those early failures, more customized strategy must be considered. Indeed, maintenance decisions shall be optimized by considering, at the time of decision, the current and predicted system health. Moreover, dynamic approach of maintenance could be considered including operation grouping or opportunistic maintenance operations (Do 2015).

**The first industrial objective** this PhD is the proposition of dynamic maintenance decision-making to reduce unexpected failures occurrences and therefore curative maintenance operations.

Evolution of preventive maintenance towards a dynamic strategy is done through predictive maintenance. Predictive maintenance aims at optimizing maintenance decisions based on current and future health states of the systems rather than statistical rules (Gouriveau, Medjaher, and Zerhouni 2016). Therefore, dynamic decision-making is based on system's health state.

Within the aforementioned *Industry 4.0* paradigm, predictive maintenance takes advantage of the available data of the systems. Data are widely available on steel plants in the form of sensors recordings, mostly exploited for tools control or product quality assessment. Predictive maintenance requires the determination of current health state of systems as well as prediction of future ones. Such tasks takes advantage of *Industry 4.0* for its ability to exploit subsidiary information contained in data recordings. Data exploitation in industrial prognostics is an ongoing research topic (Leukel, González, and Riekert 2021).

**The second industrial objective** this PhD aims to tackle is the exploitation of industrial process data to implement dynamic predictive maintenance on industrial systems.

Data-driven approaches (like machine learning) summoned by *Industry 4.0* require large amounts of data to achieve their full potential. Thanks to the 3rd industrial revolution and even more with the 4th one, industrial plants supply large amount of historical data under the form of long-term sensor recording. Those large datasets defines the *Big Data* context for data-driven applications (Jiao, K. Peng, and J. Dong 2021). They ensure the efficiency of data-driven algorithms in the long-run.

However, among those recordings, data-driven algorithms require labeled examples to learn, which industrial plants fail to supply in large amount in the case of predictive maintenance, resulting in *small datasets*. Indeed, in the case of data-driven predictive maintenance, labeled (training) examples consist in recorded "runs to failure" of the systems. Such events are rare on systems when preventive maintenance strategies are already implemented. Lack of labeled

examples also happens for recently installed systems, for which historical recordings are not yet available. Machine learning performances are linked to dataset size and smaller datasets reduce prediction performances (Catal and Diri 2009). Nevertheless, it remains important to supply predictive maintenance solutions to those systems as soon as possible.

**The third industrial objective** this PhD aims to tackle is the implementation on new systems issue faced by industrial predictive maintenance. This issue is extended to systems recently instrumented.

## 2 Scientific approaches of maintenance

Industrial predictive maintenance is addressed in the scientific literature by the Prognostics and Health Management community, in which the CRAN laboratory largely contributes.

Predictive maintenance has been formalized within the Prognostics and Health Management (PHM) paradigm as a seven steps process. The seven steps proposed are *Data Acquisition, Processing, Condition Assessment, Diagnostics, Prognostics, Decision-Making and Human Machine interface*. Each step provides outputs for the next one and a complete predictive maintenance implementation requires the fulfilling of all of them (Atamuradov et al. 2017).

Among these PHM steps, this PhD's work focuses on *prognostics* and *decision-making*. Prognostics of the health state consists in the prediction of the future evolution of the system health state. The prediction is based on the current health state provided by the *diagnostic* step, on available data and on planned future use of the system. This PhD focuses on prognostics as a prediction of the Remaining Useful Lifetime (RUL), *i.e.* time left before the system is not able to ensure its functions properly (Gouriveau, Medjaher, and Zerhouni 2016). This definition specifies prognostics as a regression problem.

Three main approaches are available for RUL prediction: model-based, data-driven and hybrid prognostics. The first one requires physics models of the system to extrapolate the health state from physics-based KPI. Such approach is efficient when advanced physics models are available on the systems. When such models are not available, data-driven approaches are used. Degradation model are then learned from the historical recordings of the system behavior (Ellefsen 2019). Hybrid approaches combine models and data-driven methods.

The RUL information is then supplied to the *decision-making* step, whose decisions efficiency are related to prognostics' precision. Dynamic maintenance decision-making depends on the predicted RUL and the upcoming production planning. Decisions are made to minimise the global cost of maintenance (Bougacha, Varnier, and Zerhouni 2020). Many levers can be used to optimize decision-making, including operation grouping (Do 2013) and opportunistic maintenance (Benoit Iung, Do, et al. 2016; Derigent et al. 2009).

This PhD's work focuses on the prognostics and decision-making due to the positioning of the industrial objectives aforementioned. The next section derives scientific issues from the industrial objectives among the scientific communities works.

## 3 Scientific issues

Industrial objectives are positioned within the PHM scientific context. The **first** industrial objective focuses on the optimization of maintenance decision-making on the basis of prognostics outputs. While the **second** and **third** industrial objectives focus on industrial data exploitation for maintenance prognostics implementation and optimization. The scientific issues tackled by this PhD therefore address these two steps of the PHM process. Following the order of the PHM

## Introduction

process, the first scientific issues address prognostics and the following ones address decision-making. Industrial objectives were unrolled from system-wide to maintenance-wide, resulting in opposite order, beginning with decision-making considerations and ending with prognostics. These two orders are opposite and therefore, the first scientific issue addresses the last industrial objective, and so on.

The **third** industrial objective puts emphasis on the *small dataset* issue in industrial maintenance topic. It relates to the implementation of data-driven prognostics among PHM works. Data-driven prognostics trained on small datasets suffers from lower performances compared to what's expected (Zhang et al. 2019).

Data-driven prognostics methods are well studied in the literature. Many papers propose data-driven models for prognostics (Ellefsen 2019). From simple k-means in Diez-Olivan 2017. To advanced models such as auto-encoders in Yu, I. Y. Kim, and Mechefske 2019, or LSTMs in Chaoub, Voisin, et al. 2021 or even convolution-based models in Z. Wu and K. Wu 2022.

There are few studies focusing on dataset size impact on prognostics performances. For instance Vapnik, Levin, and Cun 1994 proposed a formula to link model performance to the size of dataset for classification purpose.

So, the **first scientific issue** this PhD aims to address is to formulate the link between dataset size and prognostics performances to predict the performances improvement brought by additional data.

The **second** industrial objective puts emphasis on the exploitation of industrial data to implement predictive maintenance. It also relates to the implementation of data-driven prognostics among PHM works.

Despite lower performances on *small datasets*, PHM implementation yet remains pertinent on industrial systems and prognostics performances are to be optimized in any situation to support efficient decision-making. Indeed, PHM implementation can improve maintenance efficiency even when facing low prognostics performances (Skima et al. 2019). Moreover, *small dataset* in industrial context can also be a temporary situation, like in “new systems” use-cases mentioned before. Prognostics performances should be improved as much as possible to provide the best predictions to decision-making as possible.

Performances of data-driven methods can be improved with the application of data augmentation (Iwana and Uchida 2021). It is widely used in image processing to improve performances and resilience of machine learning algorithms, see for instance Shorten and Khoshgoftaar 2019.

However, data augmentation in the case of prognostics on *small datasets* has not yet been thoroughly studied. It has been thoroughly studied for Time Series (like available for prognostics) in Iwana and Uchida 2021. But such studies focus on classification applications and not much on the impact for *small datasets*.

So, the **second scientific issue** this PhD aims to address is the exploitation and optimization of data augmentation on Time Series data to improve prognostics performances when applied to *small datasets* gathered from industrial applications.

Another promising approach to improve prognostics performances on industrial systems impacted by *small datasets* is the exploitation of expert knowledge. Expert knowledge can compensate the information missing from the data due to the reduced dimension of the datasets (Chao et al. 2020). The industrial context of this PhD's study provides knowledge on the systems from industrial experts (operators, engineers, R&D researchers, ...).

Integration of knowledge into data-driven algorithm can improve performances, especially on small dataset where information is lacking from the datasets. We have proposed a short review of knowledge integration for prognostics in the paper Gay, Benoit Jung, et al. 2021.

Knowledge integration onto prognostics has been studied through physics-based models but not yet through the adaptation of data augmentation. For instance, Xiahou, Zeng, and Liu 2021

uses expert knowledge for prognostics with a hybrid model. Data augmentation has been used with expert knowledge by Pan et al. 2022 for a classification application.

So, the **third scientific issue** this PhD aims to address is the incorporation of expert knowledge into data augmentation methodologies to improve even further the performances and reliability of prognostics for industrial applications.

Finally, the **first** industrial objective focus on the optimization of maintenance decision-making on the basis of the supplied prognostics. It relates to the post-prognostics decision-making process among PHM works (Wesendrup and Hellingrath 2020).

Decision-making efficiency is linked to prognostics performance and is impacted by both its degradation from *small dataset* and its potential improvement from the two previous scientific objectives. The degraded prognostics performances induced from *small dataset* degrades the decision-making process, which relies on the RUL prediction. If prognostics performances get too low, dynamic decision-making could perform worse than simpler preventive strategies (K. Nguyen and Do 2019). Oppositely, prognostics performances improvement improves decision-making.

Finally, prognostics performances shall be commented through the impact they bring to maintenance decision-making. Maintenance cost is computed in the literature based on the preventive maintenance.

So, the **fourth scientific issue** this PhD aims to address is the evaluation of prognostics performance impact on decision-making under the industrial context provided by ArcelorMittal use-cases, in order to evaluate the decision gains achieved by prognostics improvement through data augmentation. This last issue is considered in the light of prognostics improvement pursued by the two first objectives.

## 4 Contributions to the scientific issues

This PhD's contributions are directly proposed as answers to the aforementioned scientific issues. They are below classified in two categories, contribution to prognostics and contribution to decision-making. However, it must be noted that contribution to prognostics tend to improve prognostics performances and are therefore intended to improve *in fine* decision-making. Therefore, the separation between those two categories is not as strong as the following might insinuate.

### 4.1 Contributions to prognostics

The first contribution of the PhD consists in proposing an equation to link prognostics performances to the number of available training samples. The purpose of this equation is to be fitted on the current dataset to predict the performance gain that can be expected with additional data. The proposed equation is adapted from Vapnik, Levin, and Cun 1994 and successfully validated on a public dataset.

The second contribution of the PhD focuses on evaluating and analyzing the performance of data augmentation when applied to prognostics on *small datasets*. It defines a precise test methodology to ensure the validity of results, especially as samples distribution among *small datasets* might not be representative of the global data distribution and as such great differences among different instances of *small datasets* may exist. The study focuses on data augmentation by altering existing samples thanks to data transformation functions. The considered functions are obtained from the thorough review of Iwana and Uchida 2021, with a selection done to avoid methods unfitted to prognostics.

Finally, the third contribution of the PhD for prognostics is the integration of expert knowledge into data augmentation. Two approaches are proposed. Firstly through expert knowledge such

as physics models, functional analysis or dysfunctional analysis. Secondly through statistical knowledge extracted from the data.

Available industrial datasets being too small, expert knowledge could only be sketched, however statistical knowledge proved efficient to avoid performance degradation caused by data augmentation under some conditions.

## 4.2 Contributions to decision making

The contributions to decision-making consist in the integration of prognostics results in maintenance cost modeling and the evaluation of prognostics impact on maintenance decision cost. Prognostics is integrated in maintenance cost through an alteration of the preventive maintenance cost by considering the prognostics contribution to maintenance in a probabilist way. Prognostics impact on maintenance decision cost is then evaluated by the application of this evolution to an ArcelorMittal industrial use-case. It first demonstrates the cost reduction obtained by upgrading preventive maintenance to prognostics-based predictive maintenance. It then demonstrates the positive impact of prognostics improvement brought by data augmentation previously studied.

## 5 Organisation of the manuscript

In order to present and defend these contributions, the manuscript is organized as follows.

Firstly, Chapter 1 formalizes the context within which this PhD works is positioned, through the definition of industrial constraints and a thorough literature review of related works. This chapter firstly begins by describing precisely an ArcelorMittal industrial use-case used as a baseline for industrial context. This system constraints are then generalized and placed in their scientific context, to achieve a precise positioning of the PhD's work inside scientific communities and previous works. The chapter then reviews scientific state-of-the-art surrounding the considered research subjects, including Prognostics and Health Maintenance (PHM) through prognostics and decision-making, data-driven prognostics, machine learning and data augmentation for Time Series. Finally, the chapter defines precisely the industrial and scientific issues addressed by the PhD.

Chapter 2 then evaluates the capability of data augmentation to improve prognostics performances on *small datasets*. The chapter starts by precisely defining its test methodology, to ensure validity and representativity of the obtained results. Then, the chapter evaluates the performance loss due to *small datasets*. These *small datasets* are augmented using "agnostic" data augmentation methods, *i.e.* methods not using knowledge specific to the data origin. This work is done on a public dataset described in the chapter.

Chapter 3 deepens data augmentation study with knowledge-based augmentation methods, using data specificities to guide data augmentation, in order to get more relevant augmented trajectories. Firstly it lists available knowledge on the previous dataset. Then proposes specific data augmentation methods adaptation to integrate this knowledge. Finally, it tests those approaches on the dataset and compares their results with "agnostic" data augmentation transformations.

Chapter 4 goes back to maintenance decision-making by evaluating the improvement in decision consecutively to prognostics improvement. This chapter concludes the manuscript by measuring the cost reduction of predictive maintenance thanks to data augmentation on prognostics. It first proposes a formulation of the prognostics incorporation into maintenance. Then it parametrizes the ArcelorMittal use-case. And finally uses it to measure the cost evolution linked to (1) prognostics integration and (2) prognostics improvement.

Finally, Chapter 5 concludes the manuscript, going back to the original objectives of maintenance optimization and recalling the proposed answers to the addressed industrial and scientific

## 5. *Organisation of the manuscript*

issues. The chapter also present the remaining perspectives able to serve as a base for future works on the subject.

*Introduction*

## Résumé

Cette thèse CIFRE est un projet commun entre ArcelorMittal et le laboratoire CRAN, dont l'objectif est d'optimiser la prise de décision en maintenance industrielle par l'exploitation des sources d'information disponibles : les données et des connaissances expertes, dans le cadre des contraintes industrielles présentées par le contexte sidérurgique. La stratégie actuelle de maintenance des lignes sidérurgiques est basée sur une maintenance préventive planifiée. L'évolution de la maintenance préventive vers une stratégie dynamique se fait par le biais de la maintenance prédictive. La maintenance prédictive a été formalisée au sein du paradigme Prognostics and Health Management (PHM) sous la forme d'un processus en sept étapes. Parmi ces étapes de la PHM, ce doctorat se concentre sur la prise de décision et le pronostic. En regard de cette maintenance prédictive, le contexte de l'Industrie 4.0 met l'accent sur les approches basées sur les données, qui nécessitent une grande quantité de données que les systèmes industriels ne peuvent pas fournir systématiquement. La première contribution de la thèse consiste donc à proposer une équation permettant de lier les performances du pronostic au nombre d'échantillons d'entraînement disponibles. Cette contribution permet de prédire quelles performances le pronostic pourraient atteindre avec des données supplémentaires, dans le contexte de petits jeux de données (*small datasets*). La deuxième contribution de la thèse porte sur l'évaluation et l'analyse des performances de l'augmentation de données appliquée au pronostic sur des petits jeux de données. L'augmentation de données conduit à une amélioration de la performance du pronostic jusqu'à 10%. La troisième contribution de la thèse est l'intégration de connaissances expertes au sein de l'augmentation de données. L'intégration de connaissances statistiques s'avère efficace pour éviter la dégradation des performances causée par l'augmentation de données sous certaines conditions défavorables. Enfin, la quatrième contribution consiste en l'intégration des résultats du pronostic dans la modélisation des coûts de la prise de décision en maintenance et en l'évaluation de l'impact du pronostic sur ce coût. Elle démontre que (i) la mise en œuvre de la maintenance prédictive réduit les coûts de maintenance jusqu'à 20% et (ii) l'amélioration de 10% du pronostic peut réduire les coûts de maintenance de 1% supplémentaire.

**Mots-clés:** Maintenance, Prognostics and Health Management, Augmentation de données, Petits jeux de données, Prise de décision en maintenance, Connaissances expertes

## Abstract

This CIFRE PhD is a joint project between ArcelorMittal and the CRAN laboratory, with the aim to optimize industrial maintenance decision-making through the exploitation of the available sources of information, *i.e.* industrial data and expert knowledge, under the industrial constraints presented by the steel-making context. Current maintenance strategy on steel lines is based on scheduled preventive maintenance. Evolution of preventive maintenance towards a dynamic strategy is done through predictive maintenance. Predictive maintenance has been formalized within the Prognostics and Health Management (PHM) paradigm as a seven steps process. Among these PHM steps, this PhD's work focuses on *decision-making* and *prognostics*. The *Industry 4.0* context puts emphasis on data-driven approaches, which require large amount of data that industrial systems cannot systematically supply. The first contribution of the PhD consists in proposing an equation to link prognostics performances to the number of available training samples. This contribution allows to predict prognostics performances that could be obtained with additional data when dealing with *small datasets*. The second contribution of the PhD focuses on evaluating and analyzing the performance of data augmentation when applied to prognostics on *small datasets*. Data augmentation leads to an improvement of prognostics performance up to 10%. The third contribution of the PhD consists in the integration of expert knowledge into data augmentation. Statistical knowledge integration proved efficient to avoid performance degradation caused by data augmentation under some unfavorable conditions. Finally, the fourth contribution consists in the integration of prognostics in maintenance decision-making cost modeling and the evaluation of prognostics impact on maintenance decision cost. It demonstrates that (i) the implementation of predictive maintenance reduces maintenance cost up to 20% and (ii) the 10% prognostics improvement can reduce maintenance cost by an additional 1%.

**Keywords:** Maintenance, Prognostics and Health Management, Data augmentation, Small datasets, Decision-making, Expert Knowledge

# Chapter 1

## Maintenance decision making optimization at ArcelorMittal: A plannification challenge

### 1.1 Introduction

This chapter defines industrial challenges faced by ArcelorMittal which this PhD ought to tackle, positions them within the scientific context and highlights the scientific issues induced by the industrial objectives.

To do so, the current chapter begins with the presentation of the industrial context in which the PhD's work takes place. It is directly defined through the constraints defined by ArcelorMittal steel-making activities. These constraints are linked to the two current main objectives of steel-making, decarbonization and process optimization. They are also linked to the strong stress to ensure product quality. Among those constraints, the optimization of assets maintenance is primordial. Industrial maintenance is constrained by the characteristics of steel-making plants: their linearity, the impact of assets degradation on product quality, the high cost of un-planned stops of the line and the possibility to operate maintenance only during fixed regular time slots.

The industrial context leads to three main industrial objectives to be tackled in the current ArcelorMittal state, which are defined in section 1.2. These objectives concern the upgrade of maintenance from preventive to predictive while facing the issue of *small datasets* due to the ongoing implementation of Industry 4.0, through the proposal of advanced data-driven prognostics methods exploiting expert knowledge.

The industrial objectives echo the industrial works proposed within the PHM scientific community, which are defined in section 1.3. They are connected to the prognostics and decision-making process of the PHM paradigm.

### 1.2 Industrial context - From steel-making descaling process towards a generalization as industrial objectives

The purpose of this section is to provide a comprehensive overview of the constraints and objectives that are associated with the steel-making process at ArcelorMittal, with specific focus on the Hot Strip Mill (HSM) use-case, representative of these constraints. This will be accomplished

by first detailing the steel-making process, with a particular focus on the Hot Strip Mill (HSM) in subsection 1.2.1. Following this, subsection 1.2.2 will provide a more detailed examination of one specific system, descaling, which is representative of the maintenance constraints. These descriptions will be used to formulate the industrial constraints relevant to ArcelorMittal steel-making process. Finally, by reviewing available literature, we will derive generalized industrial objectives that are aligned with these constraints.

ArcelorMittal is a multinational corporation that specializes in steel manufacturing. The company was created in 2006 when Mittal Steel, an Indian-owned enterprise, acquired and merged with Arcelor. As of 2022, ArcelorMittal is the world's second-largest producer of steel, with an annual crude steel production of 88 million metric tonnes. The company directly and indirectly employs approximately 200,000 people. ArcelorMittal is engaged in various activities, including research and development, mining, and steel production. The company is involved in every stage of the steel-making process, from mining to the final product. A particular emphasis for ArcelorMittal is the quality of the final product, which is a primary focus of their steel-making efforts.

Steel-making involves several steps including iron-making, steel-making, casting, rolling, and finishing (Béranger et al. 1996). In iron-making, iron ore, coke, and limestone are heated together to produce pig iron. In steel-making, impurities in pig iron are removed by blowing oxygen into the molten pig iron in a basic oxygen furnace. The molten steel is then usually cast at the continuous casting, cooled and solidified, and formed into sheets or other shapes through rolling. Finally, the steel is finished by cutting, stamping, drilling, or shaping it to the final product. This part is usually done by the final user of the steel sheet. The process requires precise control of various parameters to ensure that the final product quality meets the required specifications (such as tensile strength or ductability).

The following part of this section will concentrate on the hot rolling stage of the steel-making process. This particular step embodies the intricacies of steel-making and highlights the influence of the systems involved on the final product's quality.

### 1.2.1 Hot Strip Mill (HSM) process

Hot rolling is a crucial step in steel-making that involves reducing the thickness of hot steel slabs by passing them through successive sets of rolls. The initial slabs have an average thickness of 235mm and length of 10m (coming from the continuous casting), and are transformed into plates with an average thickness of 1.8mm through the rolling process. In order to do so, the plates are rolled between pairs of rolls, known as stands. The gap between the rolls becomes increasingly smaller until the desired thickness is achieved. At the end of the rolling process, once the coil is cooled according to a given pattern, the product is wound into a coil. This process is depicted in Figure 1.1.

The primary focus of steel industry is on the quality of the final product, for both economic and ecologic reasons, which the HSM process illustrates. Although steel is a 100% recyclable product, it requires a large amount of energy and thus emits a large amount of CO<sub>2</sub> as well, as documented in de Beer, Worrell, and Blok 1998, with energy consumption estimated at 24 GJ/tonne. The HSM process requires furnaces to heat up plates to 1300°C, which represents a significant amount of energy, and is often done through gas burning. Many research works are being conducted to reduce the energy consumption and CO<sub>2</sub> emissions associated with steel-making, as discussed in Xu and Cang 2010. When products do not meet the required quality standards, they are sent for recycling, as they cannot be sold to the clients. As a result, any substandard product represents a net loss of energy, money, and CO<sub>2</sub>. The HSM process is important contributor to the final quality of the product, which is of the first priority for final clients, therefore it is critical

## 1.2. Industrial context - From steel-making descaling process towards a generalization as industrial objectives

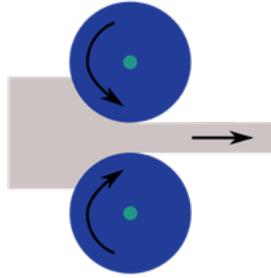


Figure 1.1: Illustration of the rolling process of the hot strip mill. On the diagram, rolls are represented in blue and the steel slab in grey.

for final product quality. This highlights the importance of considering product quality as the primary constraint of the steel industry. Thus, the first industrial constraint identified in this PhD work is the focus on product quality. It is noted as industrial constraint 1.

### Industrial Constraint 1: Product Quality

Product quality must be satisfied at all time.

The following is a description of the hot rolling process, depicted in Figure 1.2. The hot rolling process begins with reheating the slabs in a furnace for about 3 hours at a temperature up to 1300°C, to prepare them for the rolling process. Usually the slabs then enter a reversible roughing mill, where their thickness is quickly reduced from 235mm to 60mm by going back and forth through a stand. The finishing mill, consisting of 7 stands (in the case of Florange) reduces the thickness further from 60mm to about 1.8mm. At the exit of these stands, the strip reaches a speed of 14m/s. After passing through the stands, the strip is cooled on a table (named Roll-Out Table) with water flowing over it to reach a temperature of 700°C according to a predetermined colling pattern. The strip is eventually coiled in the winder and transported to storage. Any defects on the strip will be detected and either repaired on a separate line or sent back for recycling. The entire process from furnace to winder takes less than 5 minutes.

The hot rolling mill is a linear process in which all systems are arranged in series, meaning that systems cannot be run in parallel. If a system fails, the entire line becomes non-operational until the system is repaired. However, many industrial applications are asserted by the same constraint, such as continuous chemical processes, assembly lines, or conveyor belt systems. These systems also require sequential processing and can be disrupted by failures or malfunctions. Therefore, redundancy and contingency plans are often implemented to minimize downtime and ensure continuity of production. Other industries can avoid this constraint by using parallel systems, like studied by Skima et al. 2019, where a distributed conveying surface with alternate routes is available in case individual “tiles” fail (see Figure 1.3). The linearity of the process constitutes the industrial constraint 2.

### Industrial Constraint 2: Linear process

The stopping of one subprocess requires the stop of the whole line due to the linearity of the steel-making process.

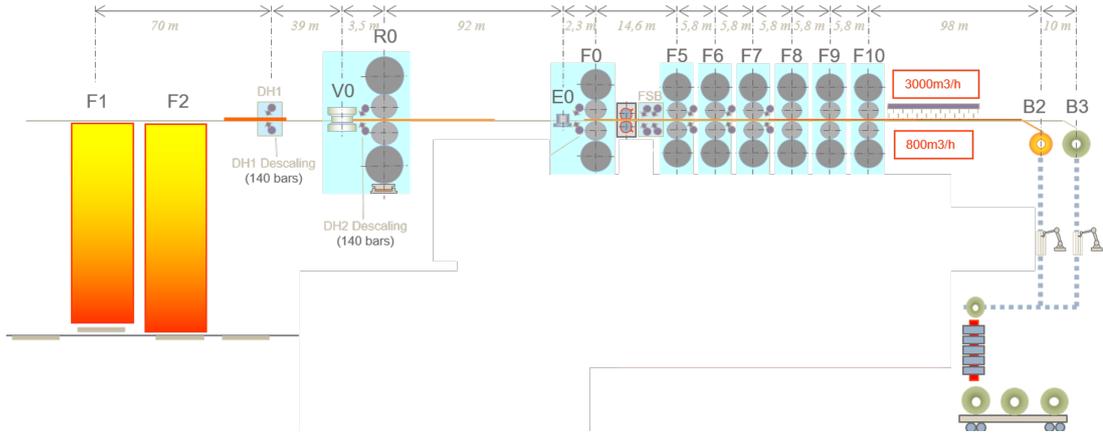


Figure 1.2: Schematics of the hot rolling process. It begins on the left with the two reheating furnaces, heating slabs up to 1300°C. Then heating slabs go through stands to reduce their thickness from 10cm to 1.8mm. Finally, they are cooled down by the cooling table before being wound in the winders.

When steel is exposed to air at high temperatures, it undergoes a chemical reaction with oxygen, resulting in the formation of a surface oxide layer called *scale*. This phenomenon occurs on the hot rolling lines as soon as the steel slab exits the furnace at 1300°C, and it affects the entire line. Figure 1.4 provides an example of a steel surface covered with scale. Scale can significantly compromise product quality during the rolling process, as it can get crushed by the rolls and become embedded in the steel. Scale is very hard and when rolled along the strip, it will penetrate into the first layers of the product creating surface defects. Additionally, scale damages the rolls when they come into contact with the product. For these reasons, it is removed from the surface of the product at several points of the hot rolling process, in an operation called **descaling**.

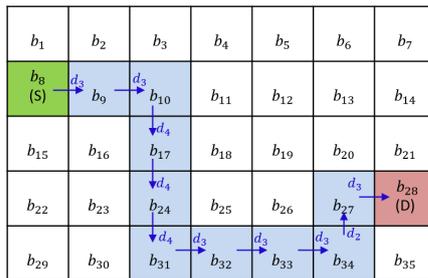


Figure 1.3: Illustration of the distributed conveying surface studied by Skima et al. 2019. The whole system is parallel in the sense that if one “tile” have failed, the function can still be ensured by using the other tiles. On the Figure, one example path is represented. If one of its tile were to fail, another equivalent path could be used.

## 1.2. Industrial context - From steel-making descaling process towards a generalization as industrial objectives

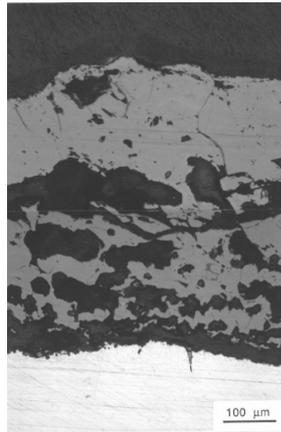


Figure 1.4: Cross-sectional image of the scale on the surface of a steel slab. Steel is in white, the air is in black and the scale is in gray.

### 1.2.2 Descaling use-case description and generalization

To remove scale from the surface of steel slabs prior to rolling, cold water under high pressure is projected on the product's surface, as illustrated in Figure 1.5.a. Descaling relies on two forces to remove scale from the surface of the steel: the impact of high-pressure water sprays on the scale and the thermal shock of cold water on the surface (Kermanpur et al. 2008).

The descaling operation is achieved through descaling stands that are either directly integrated into the stands or separate elements. Water is only projected in the presence of slabs, the water circuit being closed in the absence of slabs. The descaling system is composed of pumps to send high pressure water to all the stands, independent valves in each stand to open and close the circuit and nozzles to project the water on the slabs. Additionally, rollers are located at the entrance and exit of the stand to prevent excessive cooling of the slab due to water escaping the stand. Figure 1.5.b provides a visual example of the descaling process.

The descaling process is essential in ensuring the quality of the final product during hot rolling, as failure to properly remove scale leads to significant quality issues. Consequently, even minor problems in the descaling circuit have a direct impact on the final product quality, leading to a complete shutdown of the entire line until the issue is resolved. Common issues that can occur are nozzle blockage due to poor water quality, resulting in a failure to descale a portion of the steel surface; and pressure loss, which corresponds to a degradation of the system functioning and have many possible causes. In such cases, immediate corrective maintenance action is required to prevent any negative impact on the product quality. This leads to the industrial constraint 3.

#### **Industrial Constraint 3: Descaling system degradation impact**

The degradation of system health impacts product quality. Therefore, any system failure requires immediate corrective maintenance operation.

This highlights the necessity of efficient maintenance on steel plants to ensure product quality. The next subsection addresses the question of industrial maintenance in ArcelorMittal industrial plants.

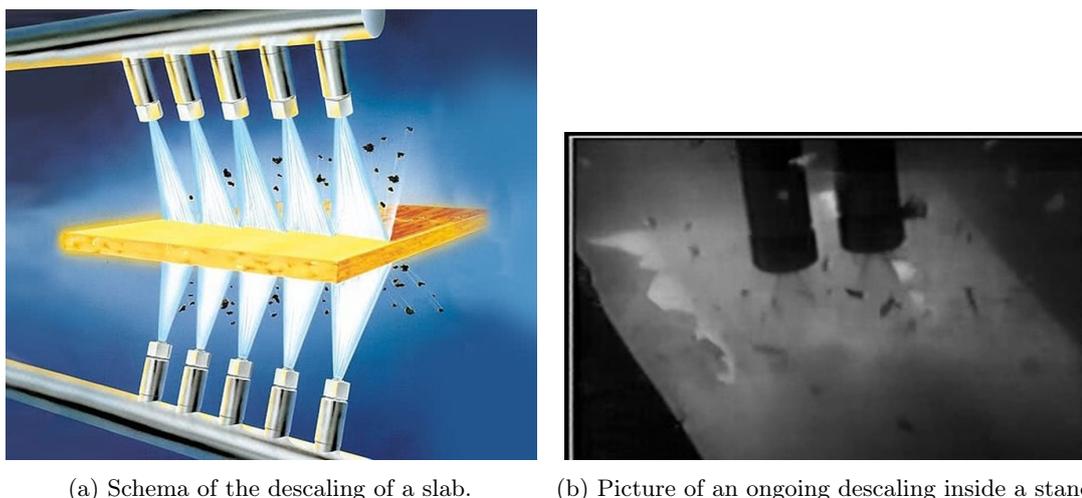


Figure 1.5: (a) shows a schema of the descaling of a slab. (b) is a picture of an ongoing descaling inside a descaling stand. Low picture quality is due to the constrained environment. It can be seen the nozzles and projected water.

### 1.2.3 ArcelorMittal maintenance constraints

The HSM process maintenance undergoes two main constraints directly implied by the physical constraints of the systems.

The reheating furnaces undergo extended heating and cooling phases, primarily attributable to their high temperature capabilities. The products, in turn, require several hours of reheating within these furnaces. As a consequence, it is impractical to bring the furnace to a halt for brief periods of corrective maintenance. The associated costs of unplanned stops is thus high. This is identified as industrial constraint 4.

#### Industrial Constraint 4: High stop cost

Un-planned corrective stops are financially and environmentally expensive. It is considered up to several thousands of euros due to reheating furnace of the HSM not being shut down and the corresponding CO<sub>2</sub> consumption.

The constraint of process continuity of service also affects the organization of maintenance in a steel plant, resulting in the planning of only a few maintenance slots. Due to the long shutdown and heating time required for reheating furnaces, maintenance times are postponed and consolidated as much as possible. For example, on the hot strip mill of Florange, maintenance slots are scheduled for two consecutive days every three weeks. This is a common approach for most HSM of ArcelorMittal group. This scheduling constraint is referred to as industrial constraint 5.

#### Industrial Constraint 5: Fixed regular maintenance slots

Maintenance can only be done during fixed regular slots available as two days every three weeks.

## 1.2. Industrial context - From steel-making descaling process towards a generalization as industrial objectives

Considering these five constraints, it is ought to upgrade current maintenance at ArcelorMittal to dynamic maintenance. Current scheduled preventive maintenance fails to prevent expensive un-planned stops (constraint 4) due to the rigidity of the maintenance organization (constraint 5). Consistent prediction of the upcoming failures requires an advanced monitoring of all systems (constraints 2 and 3). Such approach is achieved by dynamic predictive maintenance.

### 1.2.4 Maintenance formalization - toward industrial objectives for ArcelorMittal descaling

Maintenance of industrial assets can be roughly categorized as corrective or preventive. Corrective maintenance is the most basic maintenance, where maintenance is only applied on occurring failure and is the first maintenance to be implemented on industrial asset, as it is a naive strategy. It is worth replacing it with preventive maintenance when failure cost become too expensive, *i.e.* downtime cost with respect to maintenance cost.

Preventive maintenance aims to do maintenance before the failure happens and is categorized in three types. *Scheduled maintenance*, also called “predetermined”, with periodic planned operations based on statistical properties of the maintained systems. Systematic maintenance can lead to “over care” ( Gouriveau, Medjaher, and Zerhouni 2016) and therefore higher costs than required as it is based on average failure times. *Condition based maintenance (CBM)*, which ought to maintain equipment based on an assessment of its condition (sensor or inspection based) (Kalgren et al. 2006). *Predictive Maintenance (PdM)* considers both current health state and future solicitations of the asset to predict when failures may happen and take decision accordingly (Shukla, Fan, and Jennions 2020).

Current maintenance strategy on ArcelorMittal production lines is *systematic maintenance*. It consists in regular maintenance operations on systems (constraint 5) based on their average time before failure. Due to the static characteristic of preventive maintenance, remaining failures cannot be avoided, resulting in emergency corrective maintenance of the failed asset and stop of the whole line (constraint 2). To avoid those expensive corrective maintenance (constraint 4), maintenance decisions undertaken shall be optimized by considering system’s health and its evolution. This consists in the evolution of maintenance towards predictive maintenance. This dynamic approach of maintenance requires the optimization of decision based on predicted system health and maintenance cost. It can integrate operation grouping or opportunistic maintenance operations (Do 2015). Evolution to predictive maintenance is necessary to consider the environmental constraints and to achieve a high level of flexibility for the industrial equipment. Those objectives are included in the scope of industry 4.0 (Vogel-Heuser and Hess 2016; Sahli, Evans, and Manohar 2021). It results in industrial objective 1 for ArcelorMittal.

#### Industrial Objective 1: Dynamic maintenance decision-making

The development of dynamic maintenance strategies based on current and future system’s health state to reduce corrective maintenance operations occurrences.

Predictive maintenance requires abundant data on the steel-making systems, available in the context of Industry 4.0. Condition-based and predictive maintenance strategies rely on real-time assessment of the current and future health state of assets. The advanced data analytics techniques made possible by Industry 4.0 facilitate the extraction of relevant information from the sensor data. For instance, Bazart et al. 2014 propose an adaptation of diagnostics tools for detecting anomalies in continuous casting, one of the early steps of steel-making. Within the descaling process, degradation takes the form of continuous wear of the components. This is

observed in other assets of the steel-making line, line in Souto et al. 2022, to which the author of this manuscript participated and which seek to predict the wear of work rolls used in the hot rolling process. Wear measurement is only available at the end of eight-hours campaigns, and advanced decision-making requires a precise estimation during the campaign to estimate its impact on product quality and therefore anticipate quality problems. Industry 4.0 is backed by data-driven approaches, such as machine learning, which necessitate copious amounts of data to fully exploit their potential. Steel plants provide large quantities of historical data recorded by sensors, mostly installed for the system automation conducted within the Industry 3.0 evolution. This establishes a Big Data context for data-driven applications, as noted in Jiao, K. Peng, and J. Dong 2021. These large datasets are essential for ensuring the long-term efficacy of data-driven algorithms.

Data exploitation in industrial prognostics is an ongoing research topic (Leukel, González, and Riekert 2021). It results in Industrial objective 2.

### Industrial Objective 2: Prognostics based on industrial process data

The exploitation of industrial process data to implement dynamic predictive maintenance on industrial systems, through the development of robust health prognostics algorithm.

Despite the large amount of available data history, as describe above, data-driven models also require the availability of labeled examples, which are often limited in the case of predictive maintenance due to the scarcity of recorded failure events in industrial plants, resulting in *small datasets*. Indeed, in the case of data-driven predictive maintenance, labeled (training) examples consist in recorded failures of the systems. Such events are rare on systems when preventive maintenance strategies are already implemented. The lack of labeled data is particularly acute in the case of steel manufacturing, where system failures can be prohibitively expensive and preventive maintenance strategies are already in place. Additionally, data in steel-making are primarily gathered by sensors deployed for the implementation of Industry 3.0 (automation) and are commonly used to control tools and assess product quality. Thus, those sensors are not necessarily adapted to predictive maintenance applications, which may require new sensors whose data history would not be available. Similarly, recently installed systems may not have sufficient historical data for training. This is the case of ArcelorMittal descaling, for which new sensors have recently been installed to gather wider datasets, therefore data collection is still ongoing and cannot supply long data history. *Small datasets* can negatively impact the performance of machine learning algorithms, which are typically dependent on the size of the dataset (Catal and Diri 2009). Despite this limitation, it is imperative to develop and apply predictive maintenance solutions to these systems as soon as possible. This results in Industrial objective 3.

### Industrial Objective 3: Small datasets

Facing the *small dataset* issue that industrial predictive maintenance is eager to encounter through the proposal of advanced data-driven prognostics methods.

The application of condition-based or predictive maintenance to industrial assets is an active topic of research, which revolves around the **Prognostic and Health Management (PHM)** paradigm (Atamuradov et al. 2017). This paradigm defines a set of rules and models that aid in determining the health state of a system. It is addressed in further section 1.3.

### 1.2.5 Conclusion - Industrial objectives

The industrial context around steel-making descaling process maintenance concluded in a set of constraints developed in the section and summarized as:

1. *Product quality incentive*: product quality is critical due to its manufacturing cost.
2. *Linear process*: descaling is a mandatory step of HSM, any unavailability requires a stop of the whole line.
3. *System degradation impact on product quality*: descaling dysfunction lead to un-scaled product and thus unacceptable surface defects.
4. *High cost of un-planned stops*: the HSM downtime is expensive due to reheating furnaces not being shut down.
5. *Fixed regular maintenance slots*: maintenance on the HSM (including descaling) has to be made during regular dedicated time slots.

From these constraints, and in the desire to upgrade current maintenance from scheduled preventive to predictive maintenance, this PhD's ought to tackle the three industrial objectives, relevant for improving maintenance at ArcelorMittal:

1. The development of dynamic maintenance strategies based on current and future system's health state to reduce corrective maintenance operations occurrences.
2. The exploitation of industrial process data to implement dynamic predictive maintenance on industrial systems, through the development of robust health prognostics algorithm.
3. Facing the *small dataset* issue that industrial predictive maintenance is eager to encounter through the proposal of advanced data-driven prognostics methods.

The next section presents the scientific positioning of predictive maintenance in the light of these industrial objectives. It develops and formulates scientific issues to tackle to face them.

## 1.3 Scientific context – Prognostics and Health Management (PHM)

Prognostics and Health Management (PHM) is a field of study that focuses on predicting the future health and performance of systems by monitoring and analyzing data in real-time. In recent years, PHM has gained increasing attention in the research community due to its potential to improve the reliability and safety of complex systems in various domains, including aerospace, automotive, energy, and healthcare.

### 1.3.1 PHM paradigm definitions

Prognostics and Health Management is a global methodology to monitor, control and optimize systems health. It ought to assess the current health of a system as a diagnostics step and predict its future health as a prognostics step (Laloix 2018). It is done through the exploitation of past, present and future information in relation with its environmental, operational and usage condition (Kalgren et al. 2006). Finally, “the PHM paradigm provides a robust environment to optimize maintenance and logistics”( Kalgren et al. 2006).



Figure 1.6: Illustration of the PHM processes. Source of the image Atamuradov et al. 2017.

The PHM approach addresses the maintenance problematics in a variety of scales, from single-component to multi-systems maintenance strategies. The PHM paradigm is for instance applied to isolated component in Satish and Nuthalapati 2005 where the health state of bearings inside motors are assessed and predicted. Other works consider multi-component systems, such as Dinh, Do, and Benoit Iung 2020b which considers both the impact of replacing one component on the degradation of others and the structural dependence of disassembly to group components maintenance and therefore reduce maintenance cost. Similarly, K.-A. Nguyen, Do, and Antoine Grall 2017 considers joint prognostics and maintenance decision-making for multi-components, to improve their global reliability and optimize maintenance decision-making through the scope of inventory management. Finally, additional works adopt multi-systems approaches such as Voisin, Medina-Oliva, et al. 2013 which propose data-driven prognostics learned on fleets of similar systems. A large diversity of tools and approaches co-exist among the PHM paradigm, which some papers like Lee et al. 2014 ought to formalize into a systematic approaches to designing a PHM system and selecting proper tools.

The PHM paradigm defines seven consecutive processes for the implementation of predictive maintenance, through the determination of current and future health states. These processes are data acquisition, data processing, condition assessment, diagnostic, prognostics, decision-making and presentation / Human Machine Interface (HMI). Only few papers focus on this last one, as mentioned in Ellefsen 2019. They are based on the MIMOSA group standard OSA/CBM, and are represented in Figure 1.6. Therefore, the application of predictive maintenance to a system is a stepwise process that follows the PHM processes. For each of them, papers are available in the literature presenting adapted algorithms, models and solutions adapted to various use cases (Laloix 2018). Finally, Vogl, Weiss, and Donmez 2014 propose a guide of best practices for PHM system development on the basis of discussion about diagnostics and prognostics activities.

### Data Acquisition

The first step, *data acquisition*, consists in identifying the relevant parameters and signals to be monitored for the PHM application. They can be identified using systems functional analysis like proposed by Matsokis et al. 2010, combined with ontologies. Relevant parameters can be raw sensor data as well as logs of all the occurring events, such as faults, maintenance operations, etc.

(Ellefsen 2019).

### Data processing

The second step, *data processing*, consists in constructing health indicators from acquired data to be used for following condition assessment. It consists of two essential steps, namely data cleaning and data analysis (Atamuradov et al. 2017). A popular approach for errors and noise removal is the use of wavelet transform, as reviewed for condition monitoring by Z. Peng and Chu 2004. When the number of features is large, one can perform an approach based on genetic algorithm as done by Ooi, Lim, and Leong 2019 with genetic algorithms to select adequate features.

### Condition Assessment

The third step, *condition assessment*, “aims at facilitating the PHM process by constructing features that, alone, reveal the health state of a component and its evolution” (Gouriveau, Medjaher, and Zerhouni 2016). It is an intermediate step transforming data processing features into features representing the health state of the system. If the health state of the monitored system is available, supervised approaches are used such as in Chawathe 2019, who proposes a classification-based approach for determining those health-related features, learned from a dataset annotated with current system condition. Otherwise, unsupervised approaches are used, like in Schneider, Helwig, and Schütze 2018, where the authors combine multiple unsupervised automated feature extraction methods.

### Diagnostics

The fourth step, *diagnostics*, consists in assessing the current health state of the system and is composed of three tasks: fault detection, fault isolation and fault identification (Jardine, Lin, and Banjevic 2006). Fault detection relates to the anomaly detection field, which represents a developed research topic for predictive maintenance, as presented in Carrasco et al. 2021. Fault isolation consists in locating the fault, either among the components in the case of multi-components systems, or among the component, as proposed in civil engineering structure by Frigui et al. 2018 or by Ben Hassen et al. 2015 by sensor fusion based on clustering. Fault identification consists in identifying the type of fault or failure mode concerned among possible failure mode for the considered system.

### Prognostics

The fifth step, *prognostics*, consists in predicting the evolution of the health state of the system. This step is responsible for predicting failures before they occur (Jardine, Lin, and Banjevic 2006). Prognostics can either take the form of forecast of remaining useful lifetime (RUL) or upcoming health conditions, or take the form of an estimator of the probability of the system reliability for upcoming operations (Heng et al. 2009; Voisin, Medina-Oliva, et al. 2013). The RUL is defined as the time at which a component will no longer perform a particular function ( Atamuradov et al. 2017).

Prognostics prediction is based on the features generated by the condition assessment step and the output of the diagnostics step, while its outputs are designed to be used by the following decision-making. Therefore, prognostics results shall be supplied along with their inherent uncertainty, such as mentioned in the review Atamuradov et al. 2017. This topic is further exposed in Sankararaman and Kai Goebel 2015, in which the authors propose approaches to face and quantify uncertainty in prognostics. This includes the identification of the source of uncertainty and the considerations of the propagation of uncertainty along the prognostics process. According

to Javed, Gouriveau, and Zerhouni 2017, the four main sources of uncertainty in prognostics are: input uncertainty from system, measurement uncertainty from sensors, operational environment uncertainty from usage conditions, and modeling uncertainty from degradation model. An illustration of prognostics uncertainty propagation in the prediction is shown in Figure 1.7.

Following subsection 1.3.2 gets more into details on prognostics implementation, especially for data-driven approaches.

### Decision Making

The sixth step, *decision-making*, considers the process of taking maintenance decisions based on the outputs of prognostics and previous PHM steps. It can either take the form of complete autonomous decision-making (decision automation) in which the machine decides by itself of the maintenance operations to conduct. Or a decision support, in which the algorithm provides advices or insights to a human decision maker on the decisions to make. Decision-making can be extended to operational decisions such as production planning, like proposed in Bougacha, Varnier, Zerhouni, and Hajri-Gabouj 2018 where the authors propose to iteratively integrate RUL values along production to update the global tasks schedule including maintenance interventions. This formulation of maintenance and PHM as a feedback loop is also presented in Huynh, A. Grall, and Bérenguer 2017 where diagnostics and prognostics are regularly updated to support decision-making. In Bole et al. 2011, the authors propose to adapt the load of the systems accordingly to their current and predicted health. They illustrate the joint optimization of prognostics and decision-making, where prognostics is required to adapt to proposed decisions to update the predicted RUL of systems. Decision-making is extended to inventory strategy by K.-A. Nguyen, Do, and Antoine Grall 2017 for multi-components system where replacement parts can be shared among components.

Decision-making can be extended to design-oriented decisions. Design based outcomes could be adding and/or replacing sensors observability and redesign and/or components placement( Atamuradov et al. 2017).

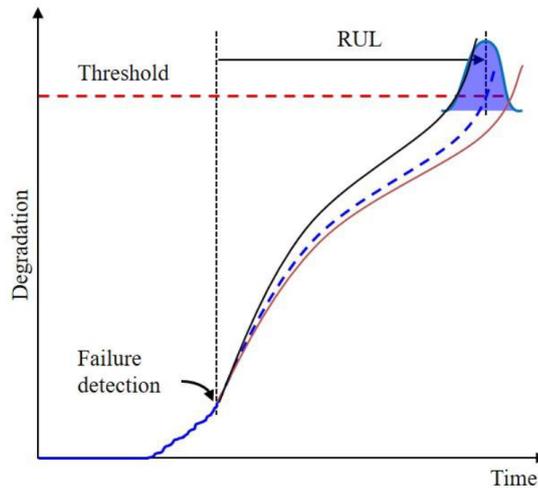


Figure 1.7: Illustration of the prognostics step of PHM and its inherent uncertainty. The RUL prediction is made at the time of detection of the degradation. The uncertainty on the prediction ends up with the resulting distribution of possible RUL, illustrated with the blue distribution. Source: Atamuradov et al. 2017.

In either case, decision-making is formalized as an optimization problem, which requires a thorough definition of a maintenance cost to minimize. Following subsection 1.3.6 gets more into details on decision-making cost evaluation.

### 1.3.2 Prognostics

The field of prognostics can be broadly classified into three main categories, namely physics-based models, data-driven models, and hybrid models, as highlighted in the literature (Y. Peng, M. Dong, and M. J. Zuo 2010; Guo, Z. Li, and M. Li 2020). Physics-based models use physic laws to predict the behavior and future evolution of the system. Data-driven methods use historical data to learn the behavior of the system to predict the RUL. Hybrid methods combine both approaches. Data-driven approaches are promoted by the introduction of the Industry 4.0 paradigm (R. X. Gao et al. 2020).

Leveraging data from industrial plants is in line with the Industry 4.0 paradigm, which emphasizes the use of digital technologies to improve industrial processes and performance. This aligns with the Industrial objective 2 of developing robust prognostics algorithms for predictive maintenance. Therefore, the study will focus on **data-driven prognostics** methods.

Prognostics applies to single-component as well as multi-components systems. Single components approach are often considered for model development, optimization and testing. Though they are less representative of real-life use-cases, where systems are composed of multiple components interacting among each other. Multi-components system requires to consider interactions between components, as proposed in Assaf et al. 2018. They propose an incremental degradation model where the degradation of one component  $X_t$  is expressed as  $X_{t+1} = X_t + [\Delta O_t + \Delta X_t + \sum_{other\ comp} \Delta X_t^{ij}]$  with  $\Delta O_t$  the degradation due to the operation,  $\Delta X_t$  the intrinsic degradation of the component and  $\Delta X_t^{ij}$  the induced degradation of component j over the current component i. Methods exist to aggregate component-level prognostics into system-wide, like proposed by Laloix et al. 2019 through the use of Choquet integral operator. The Choquet integral was also exploited by Rizzolo et al. 2011 to generate a unique aggregated indicator for maintenance decision-making from multiple ones.

Prognostics in ArcelorMittal case is still at the early stage of development and implementation. In this case, single-component prognostics is sufficient to fulfil Industrial objective 2, concerning the development of robust health prognostics algorithms. It is indeed ought to realize a proof of concept (POC) of the implementation of prognostics on one component before extending to further multi-component approaches. Therefore, the rest of the study considers **single-component prognostics**.

Prognostics is mostly found under two forms, either remaining useful lifetime (RUL) prediction or prediction of future failure states, like presented in Gouriveau, Medjaher, and Zerhouni 2016. The first approach is the most found in data-driven prognostics literature, it is assimilated to a regression problem, in which the RUL is to be predicted for each time sample as the time left before the end of a run-to-failure trajectory. The second approach is often modeled as hidden Markov chain, where the considered systems and components are linked to hidden states, which diagnostics ought to evaluate and prognostics seek the evolution. Transitions between states are thus modeled through probabilistic approaches such as Monte Carlo. This system modeling can be more adapted to decision-making when various failure modes have different impacts on the product and are therefore exposed to varying maintenance criticality.

In the rest of the study, prognostics is considered as **Remaining useful lifetime prediction**. Firstly because it is the most common for data-driven approaches. Secondly because steel-making degradations fit to regressions, with regular wear leading to failures (see for instance Souto et al. 2022).

To summarize, prognostics in the following is considered as a prediction of the Remaining Useful Lifetime (RUL) of single-component system, by exploiting available data through data-driven methods.

### Data-driven prognostics

Data-driven prognostics is related to the use of machine learning methods, which has already been largely applied to system health prognostics in the literature, with promising performances. The authors of Leukel, González, and Riekert 2021 propose a systematic review on prognostic on industrial systems with machine learning. Advanced learning models such as long short term memory neural networks (LSTM) have been applied to many applications. In the case of steel-making, Jiao, K. Peng, and J. Dong 2021 exploited LSTM to generate health indicators for rolls in the hot strip mill process. Those health indicators are then used for prognostics using a double exponential model (Pascanu, Mikolov, and Bengio 2013). In the aero-spatial field, LSTM have been used by Chaoub, Voisin, et al. 2021 to predict the RUL of turbofan engines. In steel-making, the prediction of HSM rolls wear relates to predictive maintenance as wear is an indicator of the health of a critical component of the system, which moreover directly impacts product quality. Ensemble methods are successfully applied in G. Li et al. 2021 to predict the wear profile or rolls, *i.e.* the wear estimation all along the width of the roll. Similarly, Vasilyeva, Fedorova, and Kolesnikov 2021 exploits gradient boosting to develop a predictive model of rolling mill roll wear based on a large array of operational control data. In their paper, X. Chen 2021 use machine learning models to predict the wear of galvanization rolls.

Data-driven prognostics is improved through the consideration of fleet of systems to optimize prognostics among multiple systems at once Fleet-based approaches supply information (data) in lower quality but greater quantity. It is proposed in Voisin, Léger, et al. 2014, where the fleet is composed of similar systems rather than identical ones, with a study of data-driven prognostic methods adapted to this fleet context. The model used is based on Relevant Vector Machine, and a global health index is defined to tackle the fleet heterogeneity. This heterogeneity is tackled in Medina-Oliva et al. 2014 by exploiting ontologies to formalize systems functioning.

Despite its great performances, data-driven prognostics faces several challenges (Fink et al. 2020) among which data scarcity, also called *small datasets*, is of the most importance. Similarly, *small datasets* is listed in Atamuradov et al. 2017 as a challenge for PHM.

### 1.3.3 Small datasets among data-driven prognostic

Industrial applications regularly fail to supply large amount of learning data due to systematic complexity, leading to *small datasets* problematics. Industrial causes of *small datasets* includes the following, as enumerated in Zio 2022. They apply to ArcelorMittal descaling as well:

- Poor quality of collected data: noise, missing data or faulty sensors
- Low availability of training examples for prognostics. Indeed, they consist in recorded failures of systems and are prevented by preventive maintenance strategies
- The variety of signals collected
- The variety of working conditions of systems.

Small sample size is known to reduce data-driven models performances. In their paper, Zhang et al. 2019 precise that the number of samples for training has a direct impact on the upper bound of a deep neural network's accuracy. It is shown that smaller datasets may reduce the

prognostics performance (Catal and Diri 2009), but the impact of dataset size on the prognostics performance is still an open scientific issue which should be addressed. The paper by Raudys and Jain 1991 proposes a set of recommendations for facing small sample size effects in statistical pattern recognition. They first identify models operating better on small datasets than others, for instance Fisher linear discriminant function (for pattern recognition). They then put emphasis on the selection of test set to validate results on *small datasets*, as the limited amount of available samples induce bias and variance in the test results.

Nevertheless, none of those references propose a direct link between prognostics performances and dataset size, which would formalize the impact of *small datasets* on prognostics. Such approach also allows to predict the performance improvement that can be expected from additional data. Such results is useful for industrials to evaluate the value of gathering additional data on industrial assets. It echoes the industrial objective 3 identified for descaling. Therefore, as a *first scientific issue*, this PhD seeks to formulate the link between dataset size and prognostics, then to assess what performance gain can be obtained by adding data to a small prognostics dataset. It is summoned as scientific issue 1

#### Scientific Issue 1: Small datasets impact on prognostics

Is it possible to anticipate for improvements of data-driven prognostics achieved with additional data in an industrial *small datasets* context?

#### 1.3.4 Data Augmentation as an opportunity for prognostics

The *small datasets* issue relates in the literature to the Few Shot Learning (FSL) domain. As related in Y. Wang et al. 2020, “FSL is a type of machine learning problems where [the learning database] contains only a limited number of examples with supervised information for the target”. This problem is often met in industrial context and has been listed as PHM challenges previously, see Fink et al. 2020. The survey lists three types of approaches to face FSL: model adaptation, model training adaptation and data augmentation.

Model adaptation can be done through multi-task learning, consisting for the model to predict multiple targets, including the one of interest and other related targets for which larger sets of labeled data are available. This is done in health prognostics by T. S. Kim and Sohn 2021 by training a CNN to predict both the RUL and the health status of the system, which achieved better performances than baseline models. Additionally, unsupervised learning can be used for pre-training like in Ellefsen 2019. An implementation is presented with details in Erhan et al. 2009. Their approach consists in training neural networks firstly without labels in an AutoEncoder-like architecture to initialize weights. Then, the network is trained traditionally with the few available labels. This approach increases performance of deep networks with few labelled data.

Model training adaptation consists in optimizing the machine learning model optimizer to converge faster and with less training examples to optimal solutions. This is proposed as a perspective for prognostics by Chaoub, Cerisara, et al. 2022 and is proposed in Ding, Jia, and Zhao 2021 along with transfer learning to improve prognostics precision on FSL, they improve performances on a rotary-machine datasets use-cases. Similarly, Nikolić 2017 explains how fixed neural networks architectures are to be equipped with mechanisms dynamically altering the architectures.

Finally, Zhang et al. 2019 propose to use data augmentation techniques to obtain more training samples to improve the performance of PHM applications. The use of data augmentation is promising, as it has already demonstrated its efficiency in other machine learning fields. Data

augmentation is referenced in papers as a perspective for PHM, such as Fink et al. 2020. The rest of this subsection focuses on this method for facing *small datasets* in prognostics.

Data augmentation consists in generating new samples from existing ones. It is widely used in image processing to improve performances and resilience of machine learning algorithms, see for instance Shorten and Khoshgoftaar 2019. The Figure 1.8 shows an example of data augmentation applied to an image in the case of image categorisation. In this example, the new generated images are obtained by altering the pixels of the original image, but the label is not changed, therefore increasing the dataset size. Data augmentation is also used in speech recognition, for instance in Jaitly and Hinton 2013, who use data augmentation first in the model training phase to increase the number of learning example. Secondly they use it in the prediction phase to improve predictions by expanding each test sample into multiple ones to obtain the final prediction as the average over all predictions.

Data augmentation compensates for the lack of historical data available for learning and improves models performances. It can also improves models robustness to variance, which is used by Fields, Hsieh, and Chenou 2019 to face data drift, *i.e.* the change in data over time.

A diversity of data augmentation transformations are defined in the literature, with some surveys proposing to group them in families. The paper Iwana and Uchida 2021 proposes a comprehensive survey of data augmentation for time series, and summarize them in an exhaustive list of transformations, available in Figure 1.9. Data augmentation transformations are classified in four families by both Iwana and Uchida 2021; Shorten and Khoshgoftaar 2019, which are covered in more details in the next subsections:

- Random transformations
- Pattern mixing
- Generative models
- Decompositions

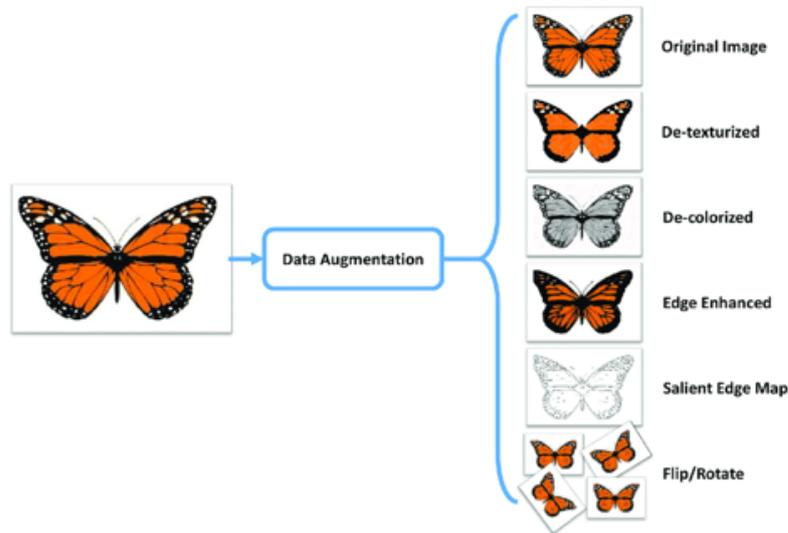


Figure 1.8: Example of data augmentation transformations in the case of image processing. From one original sample of a butterfly picture, six new ones are generated by altering the pixels of the image without altering its label “butterfly”. Source Ahmad, Muhammad, and Baik 2017.

The study of those transformation methods is conducted with more details in further Chapter 2, including a selection of methods based on their pertinence for prognostics.

Not all data augmentation transformations should be applied to any type of data at the risk of degrading performances, which include prognostics-oriented data. Data augmentation impact (both positive and negative) has not yet been studied thoroughly for *small datasets* nor prognostics application. Therefore, this PhD intends to extend the data augmentation benchmarks to industrial prognostics applied to *small datasets*. This is identified as the scientific issue 2.

### Scientific Issue 2: Data augmentation for prognostics

Can data augmentation applied and optimized on Time Series data improve prognostics performances when applied to *small datasets* gathered from industrial applications?

Performances of data augmentation depend strongly on transformation methods' hyper-parameters value, as shown in Rashid and Louis 2019. Parameters optimization is an important issue to implement machine learning algorithms, as hyper-parameters optimization in any machine learning algorithm. Their optimal values may depend on the considered dataset and need to be optimized on the available training data. However, these issues have not yet been addressed thoroughly in the literature when facing *small datasets*, which reduces the possibility to optimize performances. Therefore, this PhD ought to address hyper-parameter optimization on *small datasets* as a part of the scientific issue 2.

The next subsections introduce the data augmentation transformations family as defined in Iwana and Uchida 2021. A more thorough description is supplied in Chapter 2.

#### DA family - Random Transformations

Random transformations consist in altering time samples of the time series based on factors randomly picked. The most common method of this family is jittering, consisting in adding noise to the signals, which can with few effort improve model resilience to noise. It was applied to Time Series in Um et al. 2017 combined with a CNN but failed to improve performances with their application. In this example, the randomness comes from the noise being drawn from a normal distribution. Another classical magnitude-based random transformation is scaling, which consists in multiplying signals by a random factor. It was successfully applied to Time Series by Um et al. 2017, improving predictions by 2.6%. Examples of random transformations are shown in Figure 1.10.

Random transformations are also applied to the time-axis of signals. Classic time-transformation methods are time permutation, time masking, time slicing or time warping. The randomness in those transformations concerns the width of time windows of the value of dilation factor for time warping. Time masking can be efficient when used with Neural Networks for which input-masking can be an efficient method, but is less pertinent with other types of models such as the ones we use (DeVries and Taylor 2017b). For the record, time masking achieved gains up to 100% when combined with other transformations among the SpecAugment framework proposed by Park et al. 2019. Time slicing is successfully applied in Le Guennec, Malinowski, and Tavenard 2016, where the best results were obtained when applying time slicing, though the exact gain is hard to determine from their experiments. Time warping is successfully applied to time series by both Rashid and Louis 2019; Um et al. 2017, the first one achieving gain from 10% to 100% (depending on the model and metric considered), the latter achieving a gain of 6%. Examples of random time-based transformations are shown in Figure 1.11.

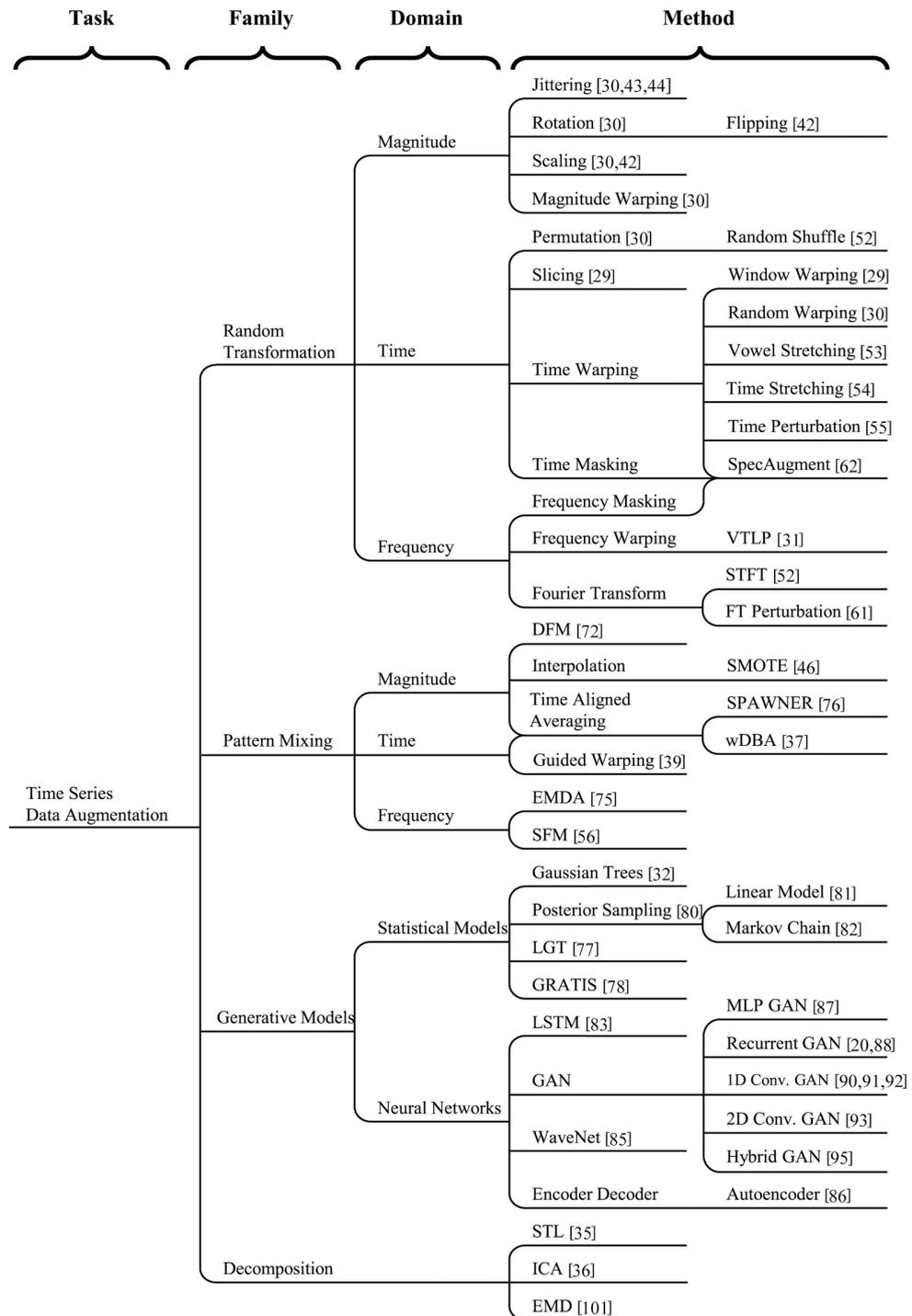


Figure 1.9: Taxonomy of time series data augmentation from Iwana and Uchida 2021.

### 1.3. Scientific context – Prognostics and Health Management (PHM)

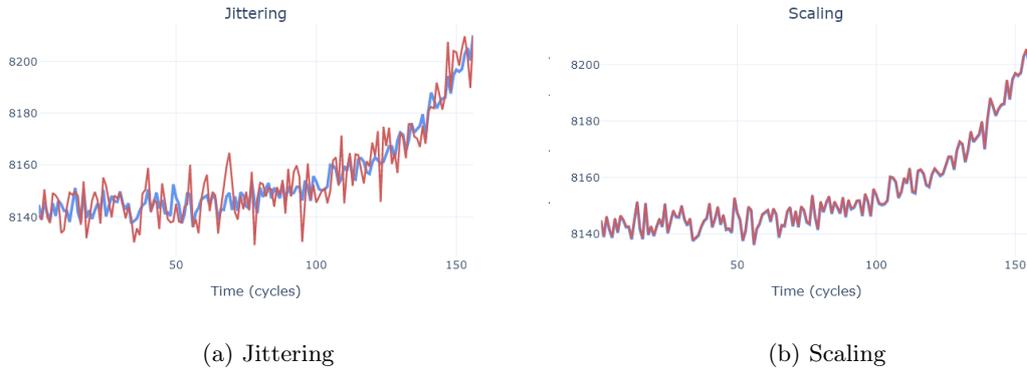


Figure 1.10: Graphical examples of random magnitude transformations. Graph (a) shows an example of jittering while (b) shows an example of magnitude scaling. Original signals are in blue, augmented ones in red.

Random transformations can also be applied in the frequency space. Those achieve great results, for instance J. Gao et al. 2021 achieves a 5% improvement with magnitude and phase augmentation on Fourier Transform. They are often similar to previous transformations, but applied to the Fourier transform of Time series, with additional care to respect the constraints of frequency data.

All these methods are efficient as they reproduce the natural variation of physical data, they however require care in their application. Firstly by ensuring they are pertinent given the physics information represented by the measurements. For instance, rotation like proposed in Um et al. 2017 consists either in exchanging the values of two signals or in inverting the sign of signals variation. It can be pertinent to simulate an inversion in sensors or an upside-down placement of a sensor, like mentioned in Rashid and Louis 2019. It is usable and pertinent for some type of data, such as position of velocity, achieving improvements of 6.6% in Um et al. 2017, which is the best performances among the tested transformations when evaluated individually. Though it has no physics sense for pressure data and may therefore reduce performances if applied to such data. Secondly by ensuring that the modification remains within a physically acceptable range.

#### DA family - Pattern mixing

Pattern mixing transformations consist in generating new trajectories as combinations of multiple original ones. The most common one is interpolation, which consists in generating new trajectories as a weighted average of two existing trajectories. A popular implementation of interpolation data augmentation is SMOTE, described and evaluated in Chawla et al. 2002, which obtained performances improvement. SMOTE is also applied on sensor signals in Arslan et al. 2019 conjointly with jittering with great results. A second transformation is extrapolation, which is an extension of interpolation allowing weights to be negatives, such that generated trajectories can go further than the space between the two original trajectories. Pattern mixing examples on an image dataset are shown on the Figure 1.12 from the paper of DeVries and Taylor 2017a. Examples on time series are shown in Figure 1.13.

Similarly to random transformations, the pattern mixing technics can be applied to the frequency domain if pertinent for the application.

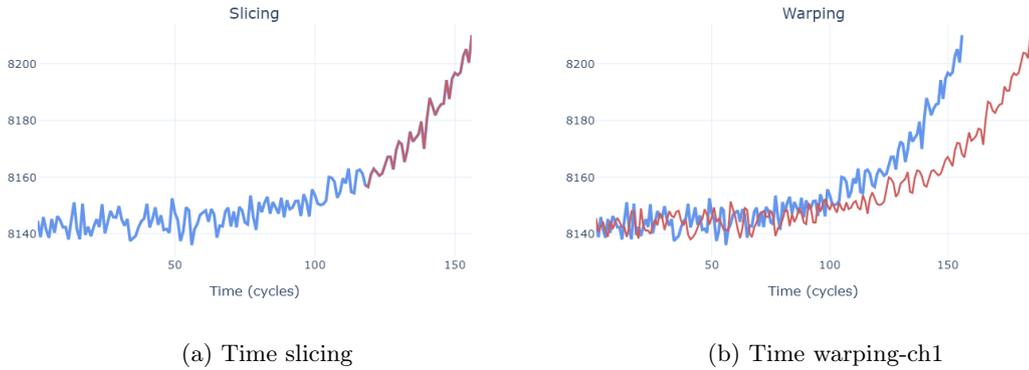


Figure 1.11: Graphical examples of random time transformations. Graph (a) shows an example of slicing while (b) shows an example of time warping. Original signals are in blue, augmented ones in red.

### DA family - Generative models

Generative models for data augmentation consist in training deep neural networks to generate new samples. A trending approach is Generative Adversarial Neural Networks (GAN), which can be extremely efficient once trained to generate new samples. GAN are used to generate human faces in Karras et al. 2020 when trained on a dataset of human faces. GAN can therefore be used to generate new learning samples and thus perform data augmentation.

The review Iwana and Uchida 2021 defines two types of generative models: statistical and neural networks. Examples of neural network-based GAN for Time Series are for instance proposed in Yoon, Jarrett, and van der Schaar 2019. Neural networks require a large amount of training data, so they are not compatible with *small datasets*. Statistical models, such as Gaussian Trees proposed in Cao, Tan, and Pang 2014 can be used with less training data but they still require

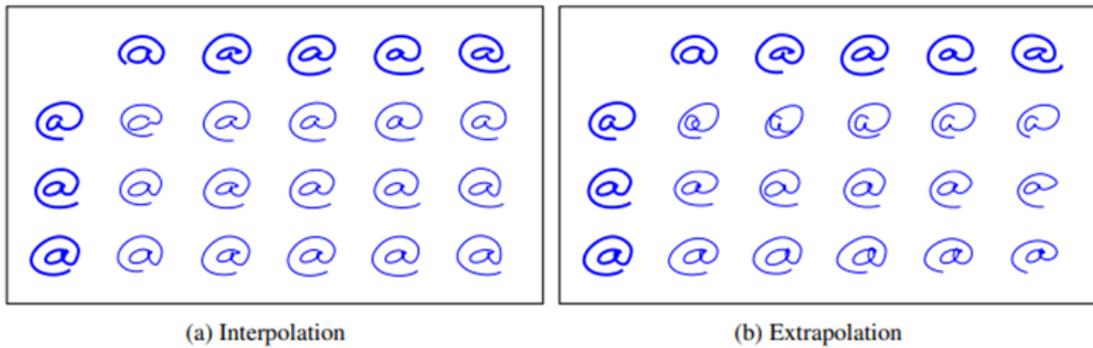


Figure 1.12: Interpolation (a) and extrapolation (b) between handwritten characters. Character  $(0,i)$  is interpolated/extrapolated with character  $(j,0)$  to form character  $(i,j)$ , where  $i$  is the row number and  $j$  is the column number. Original characters are shown in bold. Image and text from DeVries and Taylor 2017a.

### 1.3. Scientific context – Prognostics and Health Management (PHM)

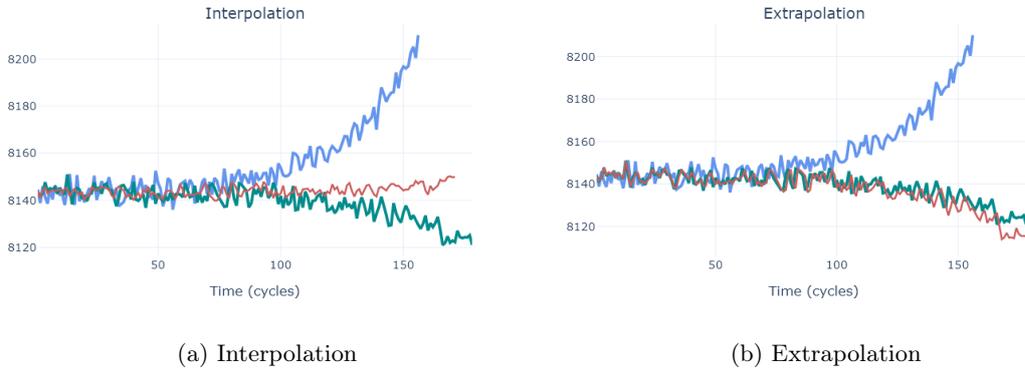


Figure 1.13: Graphical examples pattern mixing transformations. Graph (a) shows an example of interpolation while (b) shows an example of extrapolation. Original signals are in blue and green, augmented ones in red.

too much data example to be confidently used in a *small dataset* context.

#### DA family - Decompositions

The last family is decomposition, consisting in decomposing the signals in sub-signals and applying data transformations to one or more of the decomposed signals. The purpose is to keep the support of the information to ensure the pertinence of the generated trajectories. A common decomposition method is STL, which consists in splitting the signal in three components representing the seasonality (S) of the signal, *i.e.* its periodic part, then its trend (T), *i.e.* global change of the average value, and its remainder (R), *i.e.* the rest of the signal not explained in the first two components. Data augmentation transformations are then applied on the remainder to modify the random noise, like in Bergmeir, Hyndman, and Benítez 2016, or on the trend to modify the degradation profile. The seasonality signal can also be modified with frequency-based transformations. Other decomposition exists such as ICA (independent component analysis), which decomposes the signals in independent components, on which DA transformations can be applied, before reconstructing new signals. Transformations such as interpolation can be especially efficient as it can fill the space between samples in the latent space of the decomposition. It was successfully applied in Eltoft 2002.

Data augmentation as presented in the surveys is applied with no prior knowledge on the dataset. “Agnostic” data augmentation transformations consist in mathematical alterations of the signals, independent of their origin. This approach can be completed and extended by proposing alteration of data based on the physics properties inherent to signal gathered by sensors. Such approach is known in the literature as expert knowledge.

#### 1.3.5 Expert Knowledge

The majority of data-driven methods used in prognostics require high-quality data for achieving good performance. This includes a low level of noise, sufficient examples of labeled failures and the representation of each functioning modes and working conditions. However, real-world industrial data, among the *small datasets* issue, only partially fulfills these requirements.

To overcome *small datasets* and their poor quality, *expert knowledge* is a crucial factor. *Expert knowledge* encompasses all the system-related information that cannot be directly measured by sensors. It is a significant information source concerning systems. It is argued that *expert knowledge* can provide additional information to data-driven approaches to compensate for the lack of information in available data, and improve data-driven prognostic performance (Xiahou, Zeng, and Liu 2021). In their review, Sikorska, Hodkiewicz, and Ma 2011 mention that the successful practical implementation of prognostics model selection requires both a “mathematical understanding of each model type” and “an appreciation of how a particular business intends to utilize a model and its outputs”. The latter is obtained through the appreciation of system experts.

Nonetheless, expert knowledge comes in various forms, and many integration methods have been proposed in the literature. Further research is still necessary to explore the most effective ways of integrating expert knowledge into data-driven prognostics. This would enable the development of more accurate and reliable prognostic models that can enhance system performance and reduce maintenance costs.

We previously presented in Gay, Benoit Iung, et al. 2021 a classification of Expert Knowledge that can assist in determining which types of prognostics methods are suitable for a particular application based on the available knowledge. By providing a comprehensive classification of Expert Knowledge, this study sought to enhance the selection of appropriate prognostic methods and improve their performance in real-world industrial applications. It defined a multidimensional space to discriminate and give proper insight about the types of knowledge used in existing works, based on Almuiet and Salim 2014; Bonissone and Naresh Iyer 2008. Three classification dimensions were identified for the available Expert Knowledge:

- **Nature** of the knowledge: Is the Knowledge Quantitative or Qualitative?
- **Scope** of the knowledge: Does the Knowledge concern the System/Component or the Context / Environment?
- **Domain** of the knowledge: Does the Knowledge concern Health Management or not?

Finally, a selection of papers using Expert Knowledge was identified and classified among these dimensions. It resulted in the Table 1.1.

### **Steel-making expert knowledge**

Steel making data is a source of expert knowledge in the topic of data interpretation. For instance, raw thermochemical information of steel products are available in datasets. However, they are rarely usable as is, and require preprocessing. The development and implementation of this preprocessing relies on the knowledge of the product. For instance, H.-T. Nguyen et al. 2021 cluster thermochemical information using k-means before passing them as inputs for Neural Nets.

Expert knowledge is also available to optimize the decision-making step of PHM. For instance, Bozkurt, Orak, and Tunçkaya 2021 propose an “expert suggestion system” to propose blast furnace temperature control operations based on measurements and predictions. The expert suggestion system consists in a two-axis table with suggestions based on the last measured temperature and the temperature direction, it is illustrated on Figure 1.14.

### **Expert knowledge integration methods for prognostics**

Expert knowledge is exploited in the literature to improve various steps of prognostics models.

It is used in feature extraction by generating new features as combination of existing ones based on physics-based equation. For instance, Chao et al. 2020 add new unobservable variables

### 1.3. Scientific context – Prognostics and Health Management (PHM)

Table 1.1: Classification of the papers reviewed in Gay, Benoit Iung, et al. 2021 among the defined categories.

Nature	Scope	Domain	References
Qualitative	System	Health M.	Garga et al. 2001; Byrne et al. 2021
Qualitative	System	General K.	Y. Li and C. Yang 2021; Diez-Olivan 2017; Satish and Nuthalapati 2005; Helwig, Klein, and Schütze 2015; Nuñez and Borsato 2017; Matsokis et al. 2010; Emmanouilidis et al. 2010; Behera 2021
Qualitative	Context	Health M.	McDonnell 2018
Qualitative	Context	General K.	Ruiz-Tagle Palazuelos and Droguett 2021
Quantitative	System	Health M.	Lyu et al. 2020; Roemer and Kacprzyński 2001; Hua 2021; Behera 2021
Quantitative	System	Health M.	Fink et al. 2020; Łomowski and Hummel 2020; Z.-Q. Wang et al. 2018; Chao et al. 2020; Karpatne et al. 2018; Ruiz-Tagle Palazuelos and Droguett 2021
Quantitative	Context	Health M.	Johansson 2014
Quantitative	Context	General K.	She et al. 2020

**Health M.** = Health Management | **General K.** = General Knowledge.

to input data by combining available measurements. As a result, the proposed hybrid framework outperforms purely data-driven approaches by extending the prediction horizon by nearly 127%. Additionally, Karpatne et al. 2018 present physical model used to directly predict the RUL. Then, a data-driven model (using Neural Network) takes the predicted RUL as an input to compute an enhanced-RUL, obtaining better precision.

Another use of *expert knowledge* is found in model design, where the knowledge of the topology of data for the concerned application guides towards a specific type of models, adapted to it. For instance, Long Short-Term Memory (LSTM) are used in R. Wang et al. 2020 to adapt to chaotic Time Series, as they are recurrent neural networks suited for time-dependent input data. Meanwhile, Y. Li and C. Yang 2021 customizes a Genetic Algorithm (GA) used for feature extractions by defining original evolution operators for the Genetic Algorithms based system characteristics, leading to a performance improvement of 25%. To this extent, ontologies are used in various applications to formalize knowledge about systems into representations suitable for models. For instance, Medina-Oliva et al. 2014 uses ontologies to predict energy consumption. The ontologies are used to get a general representation of both systems and main concepts. Ontologies give a representation of the systems in a topological space, which is also achieved with system design using SysML, like proposed in Vogel-Heuser, Schütz, et al. 2014. Trends are then identified in the data from the exploitation of ontologies.

Finally, *expert knowledge* is used to optimize data-driven model learning phase. Learning can

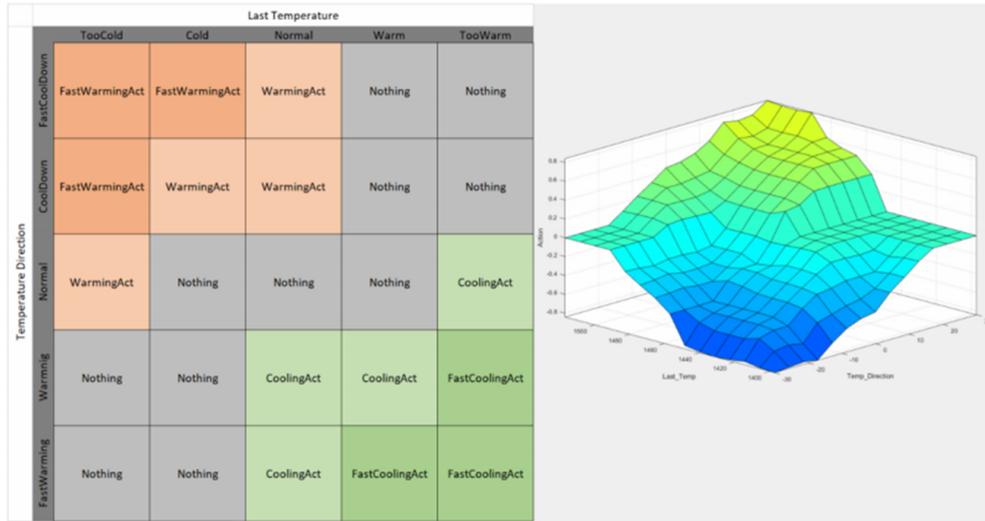


Fig. 8. Suggestion matrix design and fuzzy inference system output surface.

Figure 1.14: Suggestion matrix proposed by Bozkurt, Orak, and Tunçkaya 2021 as an expert suggestion system for temperature control decisions in steel-making blast furnace process.

be optimized with knowledge by integrating it into the loss function. For instance, Karpatne et al. 2018 uses knowledge in the form of equations between the target and input variables. They include these relations in the loss function as a penalization when the relations are violated. The authors of Byrne et al. 2021 use a loss-tuning approaches during a fine-tuning phase of a Neural Networks. They optimize the CNN weights with respect to a loss function including knowledge-based rules concerning the expected shape of objects in the pictures.

Those approaches improve prognostics, and similar knowledge-integration could be used to improve data augmentation methods by specific tuning. However, this proposal is not much exploited in the literature, and not in the case of prognostics. Therefore, this PhD ought to address the scientific issue of knowledge-integration in data augmentation for prognostics. It is formalized as scientific issue 3.

**Scientific Issue 3: Expert knowledge integration in data augmentation**

Is it possible to incorporate system knowledge into data augmentation methodologies to improve performances and reliability of prognostics for industrial applications? *Knowledge is here considered under all available forms, both expert knowledge and statistical knowledge.*

**1.3.6 Decision Making**

Decision-making within the PHM paradigm is known as “post-prognostics decision-making” (Wesendrup and Hellingrath 2020). Decision-making implementation requires the selection of adapted algorithms based on the instantiation of the problem, as mentioned in Wesendrup and Hellingrath 2020, defining the so-called per-instance algorithm selection problem, as presented in Kerschke et al. 2019. For example, in the case of steel industry, Lian, Zheng, and X. Gao 2022 mention the possibility to reschedule steel operations to face oxygen supply changes and potential

failures.

Popular decision-making strategies are based on the comparison of the prognostics output with a given threshold. Naive strategies use static thresholds, to which are compared the predicted degradation level or RUL to decide whether maintenance operations are to be done immediately or postponed. Advanced strategies propose dynamic thresholds. For instance, Vieira, Lietti, and Sanz-bobi 2012 propose a new variable health threshold that helps to re-schedule and to optimize the maintenance plan of the assets of a wind farm. Further, Barros, C. Berenguer, and A. Grall 2003 propose an evolution of the threshold-based decision to integrate the uncertainty of the RUL as an imperfectly monitored two-unit parallel system.

Decision-making efficiency depends on the quality of the prognostics, therefore decision-making optimization can be done jointly with prognostics optimization to improve their joint performances. Joint optimization of prognostics and decision-making, like proposed in K. Nguyen and Do 2019 improves global maintenance performances by considering prognostics uncertainty in the decision-making. Joint optimization is at the core of the modeling of imperfect maintenance operations, like proposed in Do et al. 2015. It is considered that a maintenance operation can be imperfect and not restore the degradation of the system to brand new but to an already-degraded state. Imperfect maintenance source can either be intrinsic or the consequence of independent external factors like human behaviours. Imperfect maintenance operations are modeled as less expensive than perfect ones, opening the way to advanced maintenance policies. They consist in a joint optimization of decision-making and prognostics, the latter being adapted to the expected post-operation degradation. They can be linked with imperfect inspections, impacting prognostics uncertainty as it is based on uncertain inputs, like in Hao, J. Yang, and Christophe Berenguer 2020

Current research topic in decision-making tends toward the consideration of multi-component systems dependencies, opening the way to advanced strategies based on dynamic maintenance grouping, opportunistic maintenances or continuity of service through load redistribution. Maintenance grouping is presented in Do 2015 as an opportunity to reduce cost through the reduction of downtime. The authors present dynamic grouping as an NP-complete problem and propose the MULTIFIT algorithm to tackle the optimization, which consists in three steps 1. optimizing maintenance at component-level 2. predicting the timing to conduct individual maintenances 3. optimizing maintenance grouping at system level. Meanwhile, Skima et al. 2019 propose maintenance strategies for parallel conveyor belt to uniformize wear and ensure continuity of service. Similarly, Herr, Nicod, and Varnier 2014 address the problem of parallel tools for continuity of service. The paper N. Iyer, K. Goebel, and Bonissone 2006 proposes a framework for post-prognostics decision-making. They define it in multiple objects interacting together.

Multi-component systems are also considered for the mutual impacts among component degradations. For instance, Dinh, Do, and Benoit Iung 2020a consider the impact of components disassembly on other component degradation. Prognostics of components therefore depends on the operation maintenance on other components. It is then considered in the decision-making as proposed in the further Dinh, Do, and Benoit Iung 2020b to group maintenance operations based on disassembly links between components.

In this study, it is not seek to develop a new decision-making algorithm but rather to evaluate the impact of prognostics implementation on decision. It is ought to do so in the industrial context of steel-making along with specific constraints like *small datasets*. One way to evaluate this impact is to consider the maintenance cost.

## Maintenance cost modeling

Decision-making is classically considered as a maintenance cost optimization problem. The usual criteria on optimization of maintenance policies are based on maintenance cost measures only: expected maintenance costs per unit of time, total discounted costs, gain, etc. (Hongzhou and P. Hoang 2006). Hence, the optimal maintenance policies are the ones that minimize (maximize) a given cost (gain) criterion (Jensen 1996). In most cases, the considered cost is financial, or a derivative like proposed by Pérès and Noyes 2003 with the consideration of the rate of repair of a system. Otherwise, A. Hoang 2016 proposes to add the consideration of sustainability to the PHM process. They create indicators based on the system energy consumption and efficiency, which are estimated during diagnostic, predicted during prognostic and considered in decision-making. Similarly, Xi et al. 2021 does energy scheduling decision-making for steel plants based on an ecology-based loss function. They use Machine Learning for decision-making to optimize a loss function defined to minimize CO2 emission and consumed energy. The optimization is done through “heuristic search algorithm of particle swarm optimization applied to find the low-carbon and economical scheduling”.

The definition of the cost to optimize can be enlarged to take into account the final product quality. Such approach is proposed by Benoit Iung, Laloix, and Voisin 2018 by estimating product quality from the current degradation of the machine. This model is then used to predict product quality evolution during prognostics.

In the literature, maintenance cost in the case of preventive maintenance considers both operation cost and the cost of the operation duration. For instance, Hou-bo, Jian-min, and Chang-an 2011 formulate the cost of scheduled operations as  $c = c1 + c2$  with “ $c1$  the cost of repairing the component, including man-hour, spare parts and other related costs” and “ $c2$  the cost of the interruption by the scheduled maintenance”. Additionally, Feldman, Jazouli, and P. A. Sandborn 2009 considers both maintenance operation cost and out-of-service cost of the considered component for different types of operations. Their maintenance cost is formulated as  $c = fC_{comp} + (1 - f)C_{comprepair} + fT_{replace}V + (1 - f)T_{repair}V$ . Where  $C_{comp}$  is the cost of procuring a new component while  $C_{comprepair}$  is the cost of repairing it,  $f$  is the fraction of operations requiring a replacement rather than reparation and  $T_{replace}$  and  $T_{repair}$  are the duration of both operations and  $V$  the value of time out of service.

However, considering system out-of-service cost as an additional cost do not account that a corrective (un-scheduled) maintenance operation also postpone future scheduled maintenances. This can be accounted by modeling the maintenance cost as an average cost over time. Therefore, maintenance operation cost are divided by the asset runtime, like proposed in H. Wang and Pham 2006.

Finally, the formulation of the cost to optimize must take into consideration the maintenance formalization. In ArcelorMittal descaling case, it consists in the fixed maintenance slots and functioning criticality aforementioned. The work of H. Wang and Pham 1996 can be considered similar with the consideration of cost per unit of time requiring an average over working campaigns. Though they do not consider maintenance strategy based on prognostics. Prognostics integration in maintenance strategy is addressed for instance by Lei et al. 2015 with maintenance operations planned based on RUL prognostics. The authors evaluate the cost of maintenance using a simulation of the wind farm. However, they do not consider regular maintenance slots and the necessity to stop all operations under failure happening.

This PhD ought to address the scientific issue of integration of prognostics in maintenance decision-making under the aforementioned ArcelorMittal descaling constraints. The ArcelorMittal case aforementioned, with regular stops and high corrective cost has few equivalent in the literature. This is summed up in scientific issue 4.

#### Scientific Issue 4: Prognostics impact on decision-making

How much is decision-making impacted by prognostics integration under the industrial context provided by ArcelorMittal use-case? How can the impact on decision-making of prognostics improvement through data augmentation be evaluated?

*This last issue is considered in the light of prognostics improvement pursued by the two first objectives.*

## 1.4 Conclusion: industrial and scientific issues list

This PhD work aims to tackle industrial maintenance in steel-making context through ArcelorMittal specificities. To do so, this chapter identified four industrial objectives:

- The development of dynamic maintenance strategies based on current and future system's health state to reduce corrective maintenance operations occurrences.
- The exploitation of industrial process data to implement dynamic predictive maintenance on industrial systems, through the development of robust health prognostics algorithm.
- Facing the *small dataset* issue that industrial predictive maintenance is eager to encounter through the proposal of advanced data-driven prognostics methods.
- The development of dynamic maintenance strategies based on current and future system's health state to reduce corrective maintenance operations occurrences.

The scientific literature addresses these subjects through the Prognostics and Health Management (PHM) paradigm (Atamuradov et al. 2017). The industrial objectives are related to the *decision-making* and *health prognostics* parts of it. To this extent, the steel-making industrial context highlights four scientific issues in the literature, that this PhD ought to tackle.

**The first scientific issue** this PhD aims to tackle is: Is it possible to anticipate for prognostics performances improvement obtained from additional data gathered in an industrial context of *small datasets*?

**The second scientific issue** this PhD aims to tackle is: Can data augmentation applied and optimized on Time Series data improve prognostics performances when applied to *small datasets* gathered from industrial applications?

**The third scientific issue** this PhD aims to tackle is: Is it possible to incorporate system knowledge into data augmentation methodologies to improve performances and reliability of prognostics for industrial applications? *Knowledge is here considered under all available forms, both expert knowledge and statistical knowledge.*

**The fourth scientific issue** this PhD aims to tackle is: How much is decision-making impacted by prognostics integration under the industrial context provided by ArcelorMittal use-case? How can the impact on decision-making of prognostics improvement through data augmentation be evaluated? *This last issue is considered in the light of prognostics improvement pursued by the two first objectives.*

Chapter 2 addresses the two first scientific issues, through *small datasets* performance evaluation and agnostic data augmentation implementation. Chapter 3 addresses the third scientific issue by proposing and evaluating upgrades of data augmentation transformations by integrating knowledge. Finally, Chapter 4 addresses the fourth scientific issue through the integration of prognostics prediction into decision-making cost according to formalized steel-making constraints.



## Chapter 2

# Agnostic data augmentation for data-driven prognostics on *small datasets*

### 2.1 Introduction

The second Chapter is a first step to tackle data-driven prognostics on *small datasets* with data augmentation. It addresses the **first scientific issue**, *i.e.* the formulation of the link between dataset size and prognostics performances to allow for anticipation of performances expected from additionally gathered data; and the **second scientific issue**, *i.e.* exploitation and optimization of data augmentation on Time Series to improve prognostics performances when applied to *small datasets* gathered from industrial applications.

Prognostics is the fifth step of the PHM methodology (Figure 1.6), consisting in predicting system health evolution. In this study, it is considered as the estimation of the Remaining Useful Lifetime (RUL), *i.e.* remaining time before the failure of the system.

Industrial applications regularly fail to supply large amount of learning data due to systematic complexity, leading to *small datasets* problematics. Industrial causes of *small datasets* include, as enumerated in Zio 2022:

- Poor quality of collected data: noise, missing data or faulty sensors
- Low availability of training examples for prognostics. Indeed, they consist in recorded failures of systems and are prevented by preventive maintenance strategies
- The variety of signals collected
- The variety of working conditions of systems.

This PhD's work proposes to face this challenge through the use of:

- Data augmentation to compensate the low availability of learning examples, which will be tackled in the current chapter
- Expert knowledge to face the issue of signals variety, tackled in the next chapter

This chapter addresses the data augmentation approach, while expert knowledge is addressed in the following chapter 3.

To this end, this chapter proposes two contributions consisting in (i) the evaluation of data-driven prognostics performances loss when facing *small datasets* (ii) the evaluation of data augmentation (DA) transformations capability to improve prognostics on *small datasets*. The latter is considered without exploiting system-specific or database-specific knowledge, which is left to the next chapter.

In that way, section 2.2 defines the test methodology implemented to simulate *small datasets* on prognostics while obtaining generalized results. Subsequently, Section 2.3 describes the dataset (C-MAPSS) and prognostics models (LSTM and ROCKET) used for experiments.

The following section 2.4 applies the methodology to the dataset to measure prognostics performance loss with respect to dataset size reduction, *i.e.* with *small dataset*. A general mathematical function is fitted between performances and dataset size in order to predict performances that can be expected from additional training data.

Once performance loss stated and measured, section 2.5 reviews common approaches for data augmentation on *small datasets* to improve performances.

Then section 2.6 evaluates prognostics performance improvement supplied by data augmentation when transformations are applied individually. It first defines the test methodology to evaluate data augmentation in an industrial-like context in subsection 2.6.1, in the continuity of the previous methodology. It is then applied to the dataset, models and DA transformation methods selected. Results are analysed in regards to various parameters to assert the complexity of application of this methodology on real industrial systems.

Finally, section 2.7 proposes a similar approach with multiple transformation methods applied simultaneously.

## 2.2 *Small datasets* test methodology: full datasets sub-sampling.

The first objective of the study is to evaluate prognostics performance loss on *small datasets*, and then predict what they would be on the same dataset with more available samples.

To this extent, *small datasets* of various size are obtained by sub-sampling *full datasets*. Indeed, to evaluate prognostics performances on *small datasets*, some instances are required to train models and measure their performances. The proposed methodology ought to do the opposite: a *large dataset* is selected, then its size is reduced to match the characteristics of a *small dataset*. The methodology concept is summarized in a first flowchart on Figure 2.1.

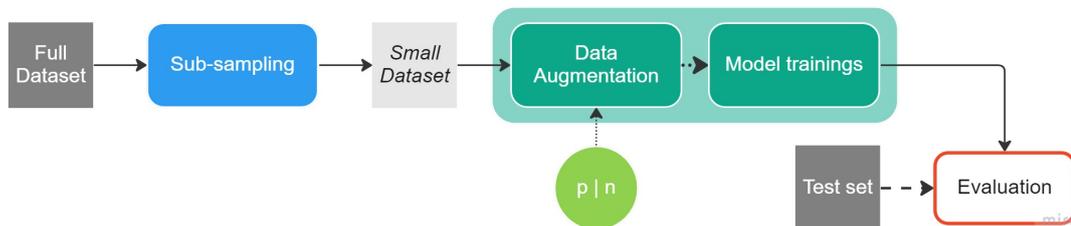


Figure 2.1: Methodology flowchart.

This methodology (dataset size reduction) will also be used in the rest of the manuscript to evaluate performances gain of additional generated data (data augmentation).

## 2.2. Small datasets test methodology: full datasets sub-sampling.

The next subsections describe the sources of variance and define a test methodology to face them, in order to ensure results consistency. The defined test methodology is then used for the rest of the PhD's work.

### 2.2.1 Sources of variance in *small datasets* evaluation

Three sources of variance have been identified: model training, datasets reduction and data augmentation generation.

Model training on a sub-sampling is a source of variance due to model initialisation. Indeed, machine learning models weights' initialization impacts final predictions and thus results (Narkhede, Bartakke, and Sutaone 2022). To face it, the model has to be trained multiple times for each sub-sampling. An example of the dispersion of prognostics performances among model trainings is illustrated on Figure 2.2 (the dataset and model used for this Figure will be introduced later in section 2.3).



Figure 2.2: Histogram of prognostics performances on a sub-samplings of a full dataset. The dataset is the C-MAPSS FD001, the model used is MiniROCKET, both described in 2.3. Sub-samplings size is 10%.

*Small datasets* are simulated as sub-samplings of full datasets. Sub-sampling consists in picking samples among the original *full dataset* to create a new smaller dataset. Therefore, the obtained *small dataset* represents few data samples collected on the system. Sub-sampling is done randomly as is usually done in the literature (see Pedregosa et al. 2011).

Sub-sampling is a source of variance among experiments. Indeed, two independent sub-samplings can show great disparities in data distribution and representativity between. This is even more true for smaller sub-samplings, because the smaller the sub-sampling, the greater the variability between sub-samplings. This diversity of distribution among sub-samplings can impact the result of prognostics performances. For instance, one pick could have a similar distribution to the test set and a model trained on it would achieve great test results. While another pick may be differently distributed than the test set, and a trained model would then achieve worst test results. An example of the dispersion of prognostics performances among sub-samplings is illustrated on Figure 2.3.

Data augmentation is a source of variance due to its random generation. Data augmentation used in the latter are based on random transformations of original trajectories. Therefore, new trajectories generation is subject to randomness, which can impact the performances of models trained on the augmented dataset. To face it, data augmentation and corresponding model training have to be done multiple times for each sub-sampling.

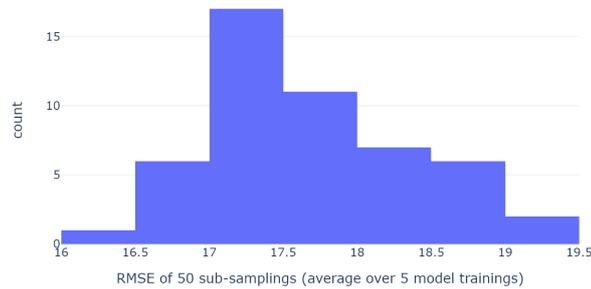


Figure 2.3: Histogram of prognostics performances on 50 sub-samplings of a full dataset. The dataset is the C-MAPSS FD001, the model used is MiniROCKET, both described in 2.3. Sub-samplings size is 25%. For each sub-sampling, the RMSE value is averaged over 5 model trainings.

Finally, the origin of variance inside of the methodology is summarized on the flowchart in Figure 2.4.

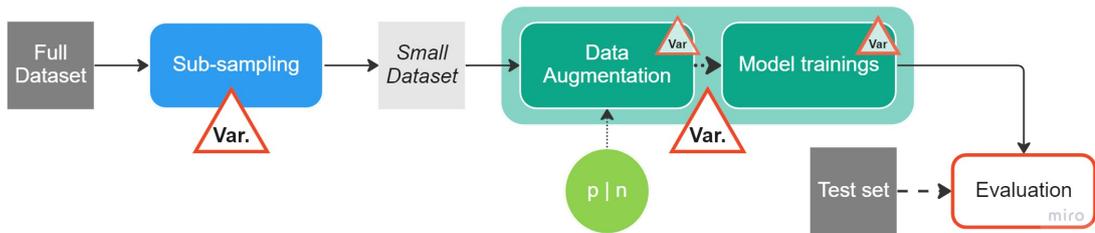


Figure 2.4: Methodology flowchart with sources of variance identified.

## 2.2.2 Facing variance with multi-picks and multi-trainings

To face the variances aforementioned and thus ensure results representativity, both sub-sampling picking and model training (*i.e.* model initialization) are conducted multiple times. This methodology is summarized on the flowchart in Figure 2.5. The required number of picks  $n_{picks}$  and number of trainings  $n_{train}$  to use for each sub-sampling size is therefore to be determined to ensure the representativity of the results with a given confidence. The next subsection focuses on this topic.

After each model training on a sub-sampling, a common independent test set is used to measure prognostics performances. The test set is never seen during the training phase. This is done accordingly to what is done in data competition (Abhinav Saxena et al. 2008) to ensure results representativity. This methodology is used for all the experiments.

Finally, in all experiments conducted in this manuscript, a fixed seed is used for random sub-samplings. Fixing the random seed before picking ensures that one random pick is always the same among various experiments. This means that pick number  $i$  will always contain the same trajectories, while being not the same of pick number  $j$  with  $j \neq i$ . Moreover, the sub-sampling is implemented such that bigger sub-samplings  $j$  would include smaller sub-sampling  $j$ ,

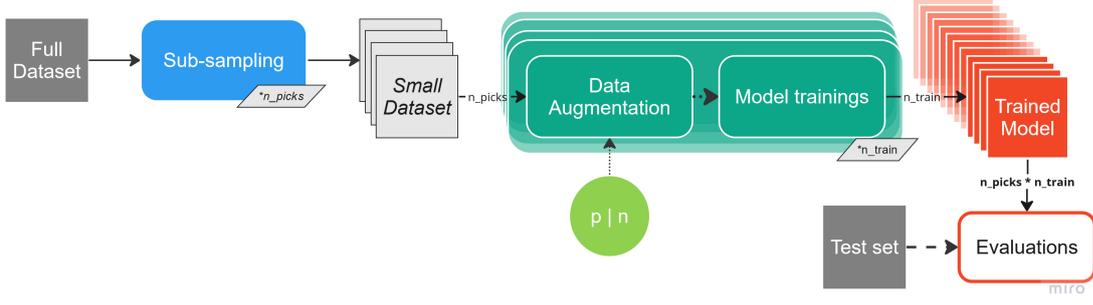


Figure 2.5: Final flowchart representation of the methodology, including the multi picks and trainings.

*i.e.*  $D_{i_1\%}^{(j)} \subset D_{i_2\%}^{(j)}, \forall i_1 < i_2, \forall j$  with  $D_{i\%}^{(j)}$  the  $j$ -th pick of the subset of size  $i$ . This both ensures repeatability of experiments and eases posterior analysis of the results.

The test methodology applied in the following is finally split in three consecutive steps:

1. Evaluation of *small dataset* performance loss with sub-samplings of the *full dataset*. This is done in section 2.4.
2. Determination of the optimal number of sub-samplings  $n_{picks}$  to evaluate in further experiments. Indeed,  $n_{picks}$  must not be too small to ensure results representativity. While a large value of  $n_{picks}$  would require too much computational power. This study is conducted in section 2.4.1.
3. Evaluation of data augmentation performances on *small datasets*. This is done in sections 2.6.

## 2.3 Dataset and prognostics models

This section describes the dataset and prognostics models selected to run the experiments. The dataset has to comply with some requirements. These requirements are a consequence of the dataset subsampling methodology described above and the industrial context of the PhD.

### 2.3.1 Dataset requirements

The first requirement is to be able to compare performances on sub-sampled *small datasets* with a full-size dataset. *Small datasets* are expected to have about 20 training trajectories available. Therefore, full datasets should have at least 50 available training examples. The second requirement is that the dataset must supply an independent test set to be used for final performance benchmark.

It was initially expected to use data from ArcelorMittal industrial plants. Unfortunately, this could not be done for the following reasons. Firstly, too few failure examples were available to fit the first requirement, with only about ten failures recorded on the descaling system. Secondly, for the same reason of lack of recorded failures, it was not possible to get enough examples for an independent test set. Thirdly, the recorded failures presented a high diversity of causes, making very difficult to develop a common prognostic model for all of them with data-driven approaches. Finally, the trajectories were not always running up to system failure, some of them being identified beforehand and production stopped to fix them. Therefore, “ground truth” RUL

was not available precisely on them. For those reasons, available industrial dataset from the Arcelor Mittal use-case was not adapted to the study’s requirements. Though, its exploitation is foreseen for future works once the dataset will be consolidated.

Therefore, it was decided to use a public dataset for prognostics shared by the scientific community. Various prognostics datasets were identified in the literature matching the above-mentioned requirements. Using a public dataset also gives the advantages to compare obtained performances with previous works from the literature, and therefore locating the proposed methods among the contributions of the scientific communities. Moreover, it simplifies model development as it is possible to use prognostics models already proposed by previous researches, either directly when sources are shared, or by reproducing the models described in scientific publications.

Based on all those criteria, the C-MAPSS turbofan dataset has been selected. It is described in the following section.

**In the rest of the document, all experiments are conducted on the public C-MAPSS turbofan dataset.** The limits on the industrial dataset discarded it for any experimental use.

### 2.3.2 C-MAPSS turbofan public dataset

The C-MAPSS public dataset is proposed by A. Saxena and K. Goebel 2008 and is widely used in the PHM community. It may be the most used in the prognostics literature, with 570 citations at the time of writing. Additionally to RUL prognostics, other uses in the literature consist in the prediction of the degradation trajectory. For instance, a synthetic Health Index (HI) is build thanks to sensor aggregation and the model aims at predicting its variation. Other approaches try to determine the fault state, like Moghaddam, Q. Chen, and Deshmukh 2020.

The data consists of run-to-failure trajectories from simulated turbofan engines. The input data are Time Series of sensor measurements and control signals. One sample is recorded for each working cycle of the turbofan engines. They are run-to-failure trajectories, *i.e.* the end of each trajectory corresponds to the time of failure of the engine, hence trajectories length is variable. There are 21 sensors inputs recorded, called *sig\_1* to *sig\_21*, representing physical indicators of the systems. There are 3 control signals recorded, called *set\_1* to *set\_3*. All of them are called input features. This dataset is similar to what is available on the descaling system, with various sensors (pressure, temperature, etc.) and control signals.

Figure 2.6 shows superposed samples of run-to-failure trajectories. It shows that trajectories have different lengths as well as variability in the sensor deviation. Trajectories often exhibit a breaking point where signal trend changes. In addition, as expected for real systems, the final value of signals, *i.e.* the failure point, shows variability. For instance *sig\_15* increases most of the time near the end, though its final offset go from  $-30$  to  $+100$ . Oppositely, some signals such as *sig\_14* are constant on all the dataset samples. More thorough data exploration on the dataset is done in the Chapter 3 to apply knowledge-based data augmentation. Indeed the first part of the work ought to apply data augmentation with (almost) no prior knowledge on the system and a thorough data analysis would have brought additional knowledge and could have biased the results.

Prognostics on C-MAPSS dataset takes the form of RUL prediction. The RUL is defined as the number of working cycles (the time unit of the dataset) remaining before the end of the trajectory. The *ground truth* RUL is capped with a “gold RUL” similarly to Chaoub, Voisin, et al. 2021. Every RUL value greater than 130 is transformed to 130, as presented in Figure 2.7.

The test set is composed of truncated run-to-failure trajectories, for which only the last sample’s RUL has to be predicted. These predicted RUL are then compared to the “ground truth” RUL of the test trajectories using a Root Mean Square Error (RMSE), which constitute the performance indicator of models and is reminded in Equation(2.1). The original paper also

### 2.3. Dataset and prognostics models

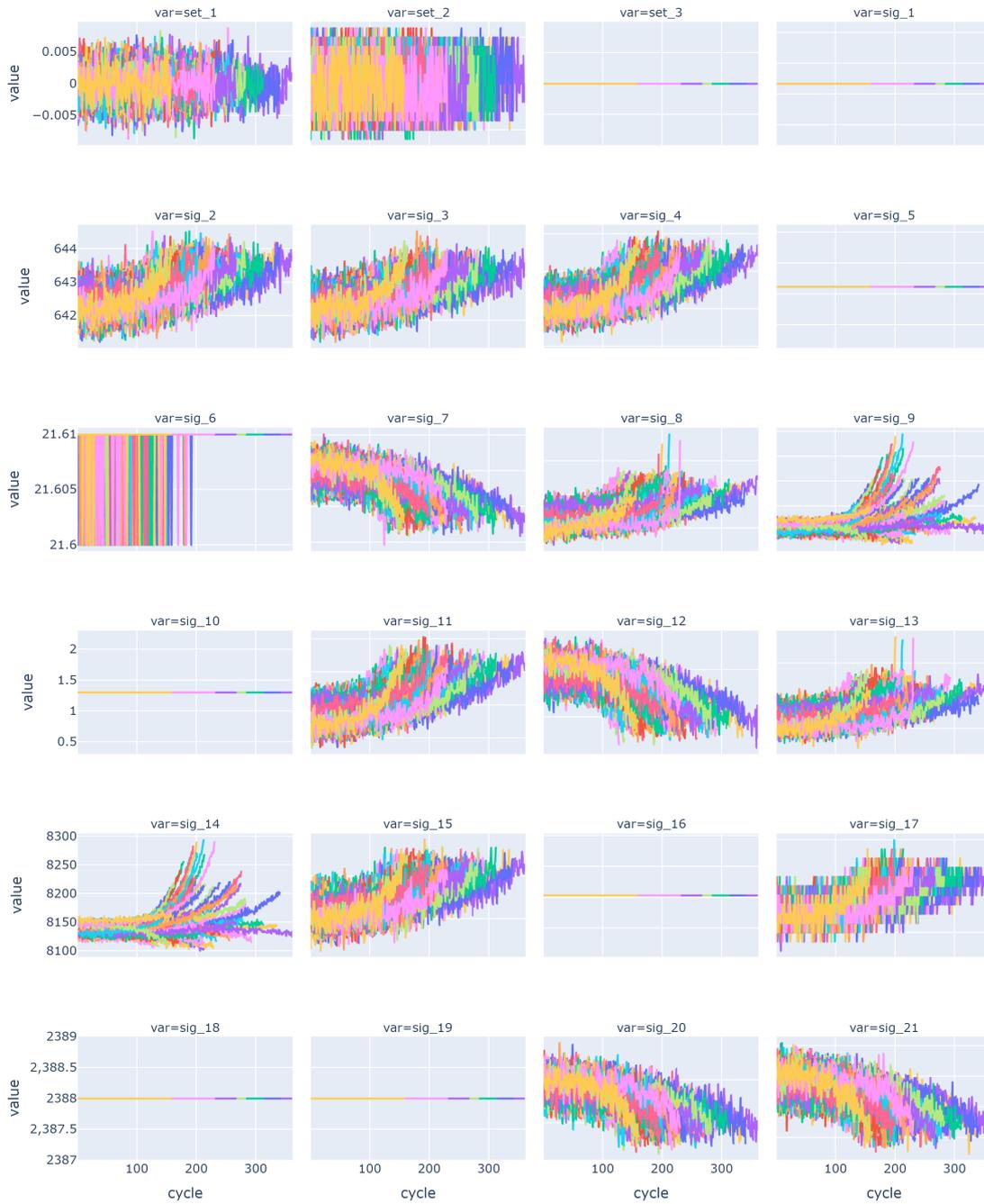


Figure 2.6: Example of a run-to-failure trajectory from the C-MAPSS dataset (FD001).



Figure 2.7: Example of the RUL to predict on one trajectory. The “gold RUL” of 130 is seen at the beginning of the trajectory.

proposes an asymmetric custom score shown in Equation (2.2) to consider the fact that “early prediction is preferred over late predictions”. Both metrics have to be minimized. In the following, only the RMSE will be considered for matter of simplicity.

$$rmse = \sqrt{\sum_{i=1}^n (rul_i - \widehat{rul}_i)^2} \quad (2.1)$$

$$s = \begin{cases} \sum_{i=1}^n e^{-\frac{d}{a_1}} - 1 & \text{for } d < 0 \\ \sum_{i=1}^n e^{-\frac{d}{a_2}} - 1 & \text{for } d \geq 0 \end{cases} \quad (2.2)$$

where

$s$  is the computed score

$$d = \widehat{t_{RUL}} - t_{RUL} \text{ (Estimated RUL - True RUL)}$$

$$a_1 = 10 \text{ and } a_2 = 13$$

The C-MAPSS dataset is composed of four sub-datasets named FD001 to FD004 which differ by the number of operating conditions and numbers failure modes. Table 2.1 summarizes the information on these sub-datasets. Each sub-dataset provides an independent test set which can be used to evaluate the model performance.

For the PhD’s work, only the first sub-dataset FD001 is used. It has only one failure mode and one operating condition, which relates to ArcelorMittal descaling, whose operating conditions are similar for all products. Additionally, it simplifies the analysis of the results. Meaning, the same condition, *i.e.* with respect to condition and failure mode, applies for all the trajectories. As such the variability of the results could not come from these parameters. It shall be noted that data augmentation requires care when applied to multiple failure modes, so generated trajectories correspond to the same failure mode as the trajectory it originated. The case of multi-degradation modes is similar to multi-class categorization considered by Iwana and Uchida 2021. They show

Table 2.1: C-MAPSS sub-datasets description. Based on Chaoub, Voisin, et al. 2021

DATASET		FD001	FD002	FD003	FD004
Number of Train trajectories		100	260	100	249
Number of Test trajectories		100	259	100	249
Number of Operating conditions		1	6	1	6
Number of Fault models		1	1	2	2
Trajectories length distribution	Max	360	378	525	543
	Mean	206	206	247	245
	Min	128	128	145	128

that data augmentation improvement is greater when the intra-class variance is higher, *i.e.* time series among failure modes diverge. FD001 is one of the smallest sub-datasets, however this is not a problem in our application as we seek for small datasets and will reduce its size for the study. Furthermore, it will enable us to test if we can improve the result by augmenting the full dataset

FD001 is composed of 100 run-to-failure time series and supplies a test set composed of an additional 100 time series (see Table 2.1). In this study, we generated “small datasets” by randomly sub-sampling the initial dataset as describe earlier in section 2.6.1. This methodology as been used in Gay, Voisin, et al. 2022. Experiments were conducted firstly on 25% then 50% and 75% of the dataset, 25% being the closest to industrial *small dataset* use-case. These dataset sizes will be used as “references”. In the following, the results for each of the two considered models, are presented with respect to these sizes.

### 2.3.3 LSTM-based model

The first prognostics model is an LSTM-based Neural Network proposed by Chaoub, Voisin, et al. 2021 and whose sources were shared to us. It was the state-of-the art model on the C-MAPSS dataset at the time of starting the tests. The model has been built in an end-to-end manner *i.e.* it is directly applied on all the time series and requires no feature engineering and almost no data pre-processing, except for RUL clipping, (see Figure 2.7) and signals standardization by a Standard Scaler <sup>1</sup>.

Therefore, it is particularly adapted to the evaluation of “agnostic” data augmentation performances, because it is directly applied to raw signals and will directly be “fed” with the new augmented trajectories.

The model is based on a MLP followed by the LSTM and then a final MLP for prognostics. As explained by the author: “after normalization of all data, inputs are fed directly to an MLP layers for feature learning, then to an LSTM layer to capture temporal dependencies, and finally to other MLP layers for RUL prognostic”. Long Short Term Memory (LSTM) Neural Networks are a structure of Recurrent Neural Networks including recall systems to get in “memory” its internal states when applied to previous samples. They are especially adapted to ordered data such as Time Series.

This allowed the LSTM-based model of Chaoub, Voisin, et al. 2021 to get the best RMSE on the C-MAPSS dataset (considering all sub-datasets). Figure 2.8 shows the model architecture with only one layer drawn for the MLPs to simplify the diagram, while Table 2.2 shows the optimized hyper parameters. The proposed model is composed of two Multi Layer Perceptron (MLP) with three layers each surrounding an LSTM.

---

<sup>1</sup>sklearn StandardScaler

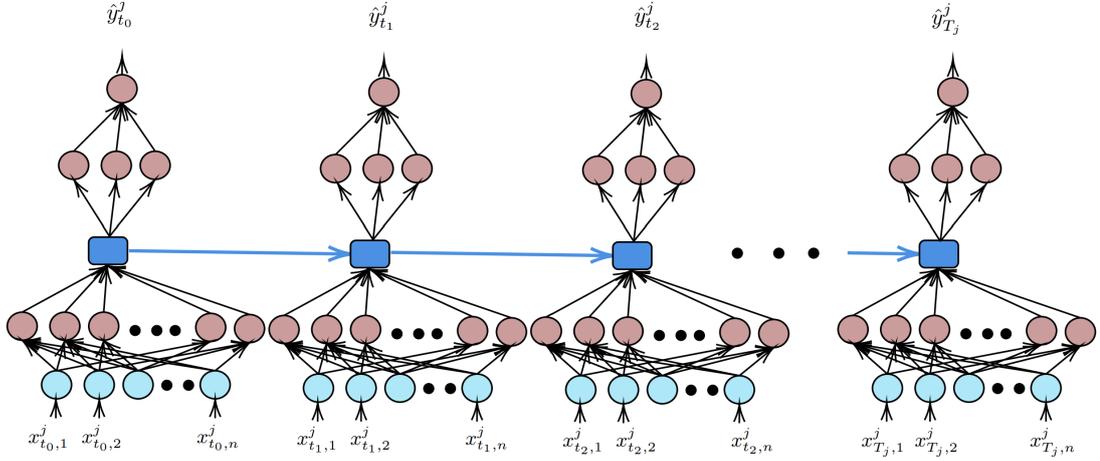


Figure 2.8: Architecture of the LSTM-based model proposed and shared by Chaoub, Voisin, et al. 2021. This picture comes from the paper.

Table 2.2: Values of the hyper-parameters of the LSTM-based model, as optimized by Chaoub, Voisin, et al. 2021

Hyper-parameter	Value
Learning Rate	0.0001
Number of MLP layers before LSTM	3
Number of neurons in MLP layers	100/50/50
Number of LSTM layers	1
Number of LSTM cells	60
Number of MLP layers after LSTM	3
Number of neurons in MLP layers	60/30/1
Activation function for MLP layers	Tanh()
Batch size	5
Dropout percentage	0%

Nevertheless, this model is long to train, as it is based on Neural Network with lot of parameters to be optimized. Indeed, each model training takes few minutes on the GPU used, which repeated hundreds of times represents hours for each data augmentation evaluation. Therefore, a second model, ROCKET, is used to evaluate data augmentation performances faster, which is necessary for hyper-parameters optimization, which requires hundreds of runs.

### 2.3.4 ROCKET and MiniROCKET prognostics models

ROCKET is a fast and efficient model based on convolutions kernels proposed by Dempster, Petitjean, and Webb 2020. It was developed in order to have a model based on convolution faster to train than Convolution Neural Networks (CNN). It was initially used for classification, while our problem is a regression. But it can be repurposed for regression by replacing the ridge classifier in the last layer by a ridge regression.

ROCKET is composed of a convolution layer based on random kernels, followed by a ridge regression. ROCKET structure is represented in Figure 2.9. The first layer transforms the input time series by applying a large number of convolutional kernels and pooling on them. The originality of ROCKET is that the kernels of the first layer are not learned, like in a CNN, but randomly generated. The random generation follows few constraints to ensure the relevance and variety of the kernels. Kernels are defined by their length, weights, bias, dilation and padding, each respecting the following constraints:

- Length is selected randomly from 7, 9, 11 with equal probability
- Weights are sampled from a normal distribution,  $\forall w \in W, w \sim \mathcal{N}(0, 1)$ , and are mean centered after being set,  $\omega = W - \bar{W}$
- Bias is sampled from a uniform distribution,  $b \sim \mathcal{U}(-1, 1)$
- Dilation is sampled on an exponential scale  $d = \lfloor 2^x \rfloor, x \sim \mathcal{U}(0, A)$  with  $A$  defined such that the effective length of the kernel is up to the length of the input time series.

This is the main cause of the fast learning speed of ROCKET, as it has not to learn the kernels. The pooling step extracts from each transformed time series the max value and the

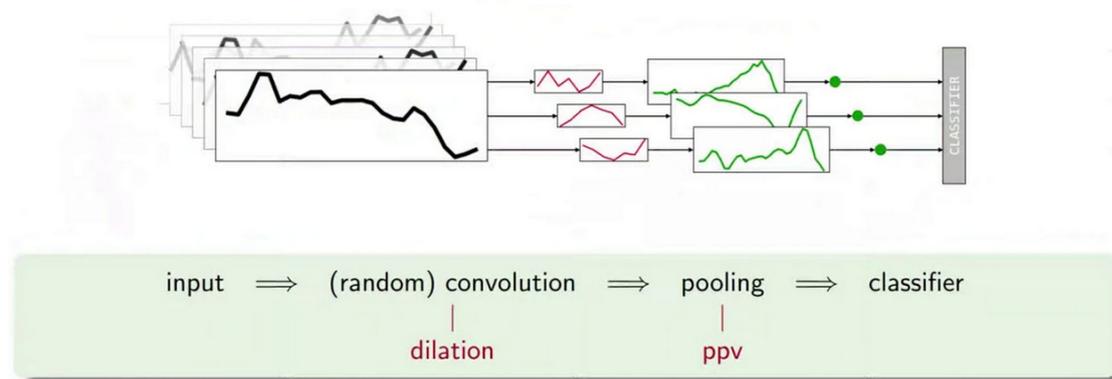


Figure 2.9: Synthetic schema of ROCKET structure, proposed by Dempster, Petitjean, and Webb 2020. The first step consists in convolution by a large amount of random convolution kernels. The second step is pooling of the convoluted signals with two specific functions (see main text). Last step is the prediction model done through Ridge Classifier. Image shared by *ROCKET* 2023.

proportion of positive values. Thus, two features are extracted per kernel per signal. The authors recommend a value of  $n_{kernels} = 10000$ . In the case of the C-MAPSS dataset with its  $21 + 3$  signals, it corresponds to  $24 * 2 * n_{kernels} = 48,000$  features in the new latent space. Eventually, those features are given as input to a regularized linear regression, specifically a Ridge regression, which outputs the RUL. Finally, ROCKET is preceded by a standardisation of the signals, for which we used a standard scaler.

ROCKET was previously used on the C-MAPSS dataset for classification by Z. Wu and K. Wu 2022. The paper defined health state categories from the signal recordings and successfully used ROCKET to predict them. This approach corresponds to diagnostics rather than prognostics but is encouraging concerning the potential of ROCKET on the C-MAPSS dataset.

Meanwhile, MiniROCKET is a reformulation of ROCKET by its authors proposed in Dempster, Schmidt, and Webb 2021. It consists in a modification of ROCKET removing randomness in the kernels generation. It allows MiniROCKET to be almost deterministic and to run even faster than ROCKET. It was therefore used to fasten results evaluation. As it was published during our tests, some of the results have also been executed with ROCKET.

## 2.4 Prognostics performance loss on *small datasets*

In this section, prognostics models performances are evaluated on the dataset accordingly to the test methodology defined in Section 2.2. Sub-samplings are used to evaluate the degradation of performances due to the *small datasets* with respect to their size. Finally, it is determined the required number of picks for further evaluations.

The two selected models (LSTM and ROCKET) are evaluated on the whole dataset and then on 20 randomly picked sub-samplings. It is reminded that the sub-samplings are the same for the two models and for the future experiments thanks to fixed random seed.

Starting with the LSTM model, the performance degradation is shown in Figure 2.10a. These results show both a decrease in the average precision and an increase in the variability between trainings as the number of samples decreases. Indeed, average performances range from  $rmse_{100\%} = 13.9$  (same value as in the original paper of Chaoub, Voisin, et al. 2021) to

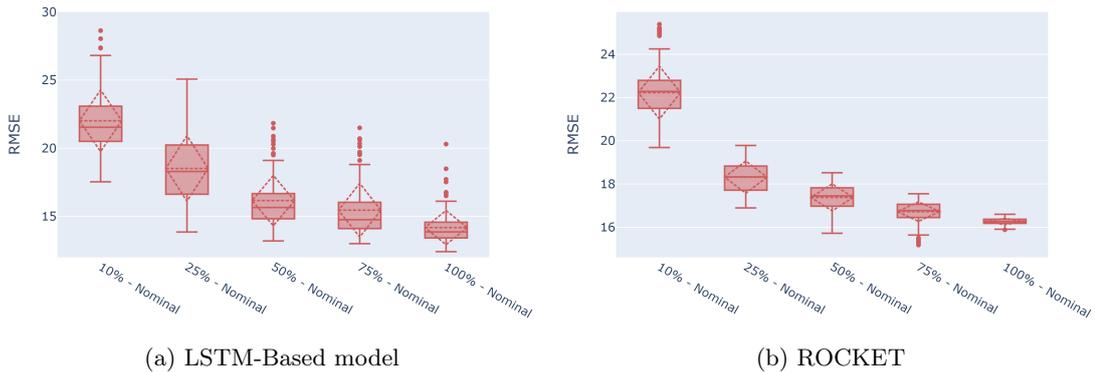


Figure 2.10: Prognostics performances on the C-MAPSS dataset with (a) the LSTM-based model and (b) ROCKET for various sizes of sub-samplings (small datasets). Performances are measured with RMSE on the test set.

## 2.4. Prognostics performance loss on small datasets

$rmse_{75\%} = 14.8$ ,  $rmse_{50\%} = 15.6$ ,  $rmse_{25\%} = 18.3$  and  $rmse_{10\%} = 21.5$ . It is worth mentioning that the increase in the dispersion may be caused by the model’s hyper-parameters being optimized for the full dataset. All the results are summarized in Table 2.3.

Considering the ROCKET model, the results are shown in Figure 2.10b. Results are consistent with the LSTM, although ROCKET does not achieve as good results as the LSTM. Average performances range from  $rmse_{100\%} = 16.2$  to  $rmse_{75\%} = 16.7$ ,  $rmse_{50\%} = 17.4$ ,  $rmse_{25\%} = 18.3$  and  $rmse_{10\%} = 22.2$ .

Prognostics performances drop on the C-MAPSS dataset when training on 25% of the dataset is high enough to degrade maintenance decision-making. For the LSTM models, performances drop from  $rmse_{100\%} = 13.9$  to  $rmse_{25\%} = 18.3$ , corresponding to a 32% performance decrease. For ROCKET, performances drop from  $rmse_{100\%} = 16.2$  to  $rmse_{25\%} = 18.3$ , corresponding to a 13% performance decrease. This loss impact on decision making is evaluated in the latter Chapter 4 of the manuscript and is confirmed to increase maintenance cost. Cost is increased even more when considering the LSTM, as performance drop is proportionally higher. An explanation for the higher performance drop on the LSTM than ROCKET is the number of tunable parameters. The LSTM model has many hyper parameters to tune, see 2.2, while the ROCKET model has only 1 hyper parameter, *i.e.* the number of kernels<sup>2</sup>. As such the ROCKET model is less sensitive to a change in the dataset.

Table 2.3: Summary of performances loss due to small datasets for all datasets and models. Performances loss are evaluated on the average  $rmse$ . Loss are exprimed as percentage of RMSE reduction compared to 100% of the dataset.

Dataset	Model	RMSE 100%	Loss @75%	@50%	@25%	@10%
C-MAPSS	LSTM	13.6	6%	12%	32%	55%
FD001	ROCKET	16.2	3%	7%	13%	37%

### 2.4.1 Determination of number of picks for further evaluations

This subsection aims at finding the smallest number of picks  $n_{picks}$  required such that obtained results are representative of the distribution which could be expected.

It is ought to determine  $n_{picks}$  such as performance variations above 10% is confidently identified. The value of 10% is chosen accordingly to results of chapter 4, stating that an improvement in decision-making can be expected for prognostics improvements starting at 10%. “Confidently” is here understood as a confidence index over 95%.

#### Theoretical work

Student’s independent two-sample t-test is considered to compare two results distributions  $X_1$  and  $X_2$ . The test null hypothesis is that “the means of the populations from which the two samples were taken are equal”. The Student t-test is applied considering two samples  $X_1$  and  $X_2$  with same sizes  $n_{picks}$  drawn from distributions whose means are different of 10%:  $|\bar{X}_2 - \bar{X}_1| = 0.1 * \bar{X}_2$ . To use the Student t-test under such conditions, two hypothesis are considered:

- Prediction results for a given configuration follow a normal distribution.
- Compared distributions have similar variances, *i.e.*  $s_{X_1}^2 = s_{X_2}^2$ .

<sup>2</sup>Which is considered to have low impact on performances (Dempster, Petitjean, and Webb 2020)

These hypothesis are tested in the next subsection on the model trained without data augmentation.

Considering the notations and hypothesis, the value of the  $t$  statistic of the Student t-test is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{2}{n_{picks}}}}$$

where

$$s_p = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}} = \sqrt{\frac{2 * s_{X_1}^2}{2}} = \sqrt{s_{X_1}^2} = |s_{X_1}|$$

and

$$\bar{X}_1 - \bar{X}_2 = \bar{X}_1 - 0.9 * \bar{X}_1 = 0.1 * \bar{X}_1$$

therefore

$$t = \frac{0.1 * \bar{X}_1}{|s_{X_1}| \sqrt{\frac{2}{n_{picks}}}}$$

Considering  $\bar{X}_2 \neq \bar{X}_1$ , the null hypothesis is to be rejected. Therefore, the t-statistic should be superior to the threshold value that can be found in tables such as found in Annex 5.1. The threshold  $t_{thres}$  corresponding to a 95% confidence is found in the table considering for a degree of freedom  $\nu = 2n_{picks} - 2$ . It is noted  $t_{thres}(n_{picks}) = t_{thres}^{(\nu=2n_{picks}-2, 0.1)}$  in the following. Therefore,  $n_{picks}$  can be determined by solving:

$$\begin{aligned} t &\geq t_{thres}^{(\nu, 0.1)} \\ \frac{0.1 * \bar{X}_1}{|s_{X_1}| \sqrt{\frac{2}{n_{picks}}}} &\geq t_{thres}^{(\nu, 0.1)} \\ \sqrt{n_{picks}} * \frac{0.1 * \bar{X}_1}{|s_{X_1}| \sqrt{2}} &\geq t_{thres}^{(\nu, 0.1)} \\ \sqrt{n_{picks}} &\geq t_{thres}^{(\nu, 0.1)} * \frac{|s_{X_1}| \sqrt{2}}{0.1 * \bar{X}_1} \end{aligned} \quad (2.3)$$

The determination of  $n_{picks}$  is defined as the minimal value verifying equation (2.3). Therefore, the minimal value of  $n_{picks}$  is defined in equation (2.4). This value can be determined using the table provided in Annex 5.1.

$$n_{picks} = \min \left\{ n \mid n \geq 2 * 100 * t_{thres}(n)^2 * \left( \frac{|s_{X_1}|}{\bar{X}_1} \right)^2 \right\} \quad (2.4)$$

The determination of  $n_{picks}$  is done further for each model and dataset size by applying the following steps:

1. Verification of the normality of the distribution.
2. Determination of the minimal  $n_{picks}$  using Equation (2.4).

The determined values are then used in the next experiments. The hypothesis of similar variance shall be re-evaluated on further experiments and the number of picks could be updated in the case the difference is significant.

### Validation of the normality of the distribution

The obtained results are finally used to assert the normality of the RMSE distribution. The Figure 2.11 shows the distributions of RMSE for all evaluated dataset sizes and both models. D’Agostino and Pearson’s test is then used to test the normality of the distributions. The Python’s Scipy test is used, calculated p-values are compiled in Table 2.4 The hypothesis that RMSEs among picks follow normal distributions is verified for most cases. The cases where it is not verified are not considered in the rest of the study, which focuses on 100% and 25% sizes.

Table 2.4: p-values of the D’Agostino and Pearson’s normality tests for LSTM and ROCKET on various sub-samplings sizes of the C-MAPSS FD001 dataset. It is reminded that the null hypothesis is that a sample comes from a normal distribution. Therefore, a p-value greater than 0.05 means there are more than 95% chances that the samples follow a normal distribution.

C-MAPSS percentage	10%	25%	50%	75%	100%
LSTM	0.01	0.31	0.30	0.00	0.38
ROCKET	0.13	0.87	1.00	0.02	0.96
MiniROCKET	0.13	0.26	0.68	0.32	0.19

### Determination of minimal $n_{picks}$ for each sub-sampling size and models

Given the previous results for all datasets and models, we computed this value, and compiled them in table 2.5. For instance, for C-MAPSS FD001 with LSTM average is  $\bar{X}_1 = 18.3$  and

LSTM on C-MAPSS FD001



Figure 2.11: Distributions of RMSE on the test set for various percentage values for MiniROCKET on C-MAPSS FD001. For each percentage values, 20 to 60 independent sub-samplings are picked, and the model is trained multiple times. Each point corresponds to one training. Those distributions are used to evaluate normality (see Table 2.4).

Table 2.5: Minimum number of picks needed to differentiate a performances improvement down to 10% among experiments

Dataset	Model	100%	75%	50%	25%	10%
C-MAPSS	LSTM	8	15	13	15	10
FD001	ROCKET	2	3	3	3	4

variance is  $s_{X_1} = 2.4$ , therefore we are seeking for  $n_{picks} = \min \{n \mid n \geq 13.5 * t_{thres}(n)^2\} = 15$ .

The values obtained in the previous table 2.5 are used in the following experiments to ensure the representativity of results. Those results are based on the hypothesis that the variance of additional results are the same as the one obtained with no augmentation. As it could not be verified, most of the experiments will be repeated more than the minimal number identified in the table.

### 2.4.2 Prediction of performance loss due to *small dataset*

The previous results demonstrated the reduction of prognostics performances with the reduction of the dataset size, confirming the impact of small datasets on prognostics. However, in real application, the opposite is sought, *i.e.* one would like to predict the performance gain that can be expected from additional data. We initiated a work in this direction in our article Gay, Voisin, et al. 2022 and previous work like Catal and Dirı 2009 studied size impact conjointly with metrics and feature selection techniques.

It is found relevant to fit a function  $RMSE = f(n_{samples})$  to the datasets. To do so, the model proposed by Vapnik, Levin, and Cun 1994 is adapted to our data. This model is proposed to predict the performances of classifier models based on the dataset size. It is adapted to small sizes of dataset and offers great performances in the classification case. It is therefore pertinent to use in this application. The model is depicted eq. (2.5) where  $n \in [0\%, 100\%]$  is the dataset size and the average RMSE is used:

$$RMSE(n) = RMSE^* + c_1 * \frac{\log(c_2 * n) + 1}{\frac{1}{2} * c_2 * n} \tag{2.5}$$

The model is fitted on the average RMSE values obtained for the dataset sizes tested earlier and summarized in Table 2.3. To correspond to a realistic use-case and validate the model precision, the coefficient are fitted with the RMSE vales of dataset sized from 10% to 50% and the RMSE prediction error was measured on the RMSE for 75% and 100% of the dataset. Coefficients fitting is done with Scipy least square optimizer<sup>3</sup> in Python with a L2 regularization term of  $\lambda = 0.1$ .

For instance, applied on the C-MAPSS FD001 with the LSTM-based model, the coefficient are fitted on the RMSE from 10% to 50%, which gives  $c_1 = 5.7$ ,  $c_2 = 0.12$  and  $RMSE^* = 10.2$ .  $RMSE^*$  corresponds to the limit of the function for  $n \rightarrow \inf$ , *i.e.* the best RMSE value that can be expected with this model. It can be considered as the intrinsic error of the problem. These values are then used to predict performances expected for 75% and 100%, which gives  $\widehat{rmse}_{75\%} = 14.3$  and  $\widehat{rmse}_{100\%} = 13.6$ , which corresponds to errors of respectively 3.2% and 2.4%. Those prediction errors are small enough to validate the model. The Figure 2.12 shows the curve along with the average performances. These results open the perspectives to apply data augmentation to the full dataset with hope to yet improve model performances towards the identified limit of  $RMSE^* = 10.2$ .

<sup>3</sup>[https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least\\_squares.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html)

## 2.5. Agnostic data augmentation for Time Series - Literature Survey

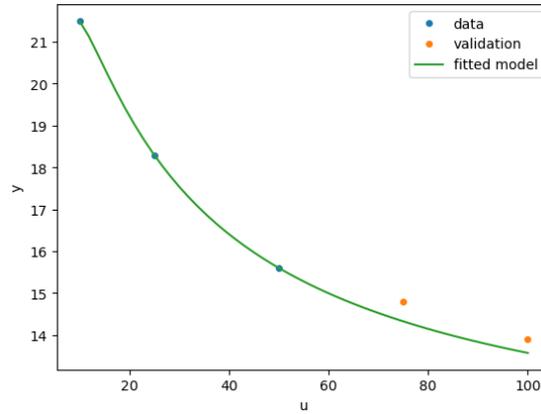


Figure 2.12: Average RMSE of the LSTM-based model on the C-MAPSS dataset along the fitted curve for predicting performances. The last two points (75% and 100%) are not used for curve fitting. So they can be used for validation.

This model is therefore considered efficient in predicting prognostics performances with additional data. The same study is applied to ROCKET, with similar results to the LSTM. Table 2.6 summarizes the fitted values of the coefficient for the various datasets and models.

Those results can be kept in mind to compare the improvements obtained from data augmentation in the following sections with expected improvements from additional data.

The presented approach of reducing a dataset size to simulate a *small dataset* is opposed to the real-case situation. In real situation, only *small datasets* are available. In such cases, it is ought to know how much performances can be improved from additional data and therefore the interest of gathering additional data. To face this question the proposed formula demonstrated its relevance to predict the performances on the full dataset when fitted on performances for smaller sizes.

## 2.5 Agnostic data augmentation for Time Series - Literature Survey

The rest of the Chapter focuses on data augmentation to face *small datasets*. The current section 2.5 begins by reviewing what is done in the scientific literature. Then section 2.6 evaluates the efficiency of data augmentation transformations when applied individually. Finally, section 2.7 does the same for transformations applied simultaneously in the hope to improve even more performances. These last two sections first defines their test methodologies in subsections 2.6.1 and 2.7.1, similarly to what was done in Section 2.2.

Table 2.6: Fitted coefficient of the model in equations (2.5) for the various models.

Dataset	Model	$RMSE^*$	$c_1$	$c_2$	Error @75%	Error @100%
C-MAPSS	LSTM	10.2	5.7	0.12	3.2%	2,4%
FD001	ROCKET	15.6	173	37	0.17%	1.7%

### 2.5.1 Data augmentation for time series in the literature

*This section is a summarised literature review. A full literature review can be found in Chapter 1.*

Data augmentation is a way to compensate for the lack of historical training data and provide better prognostics performances. Data augmentation consists in generating new samples from existing ones. It is widely used in image processing to improve performances and resilience of machine learning algorithms, see for instance Shorten and Khoshgoftaar 2019. The Figure 2.13 shows an example of data augmentation applied to an image in the case of image categorisation. In this example, the new generated images are obtained by altering the pixels of the original image, but the label is not changed, therefore increasing the dataset size.

Several approaches for data augmentation exist, they can be classified in four families like Iwana and Uchida 2021; Shorten and Khoshgoftaar 2019 propose:

- Random transformations
- Pattern Mixing
- Generative models
- Decomposition

Generative models for data augmentation consist in training deep neural networks to generate new samples. A trending approach to do so is Generative Adversarial Neural Networks (GAN), which can be extremely efficient once trained to generate new samples. GAN were used to generate human faces in Karras et al. 2020 when trained on a dataset of human faces<sup>4</sup>. GAN can therefore be used to generate new learning samples and thus perform data augmentation.

The Iwana and Uchida 2021 review defines two types of generative models: statistical and neural networks. Examples of neural network-based GAN for Time Series are for instance proposed

<sup>4</sup>See <https://thispersondoesnotexist.com/> for an impressive demo.

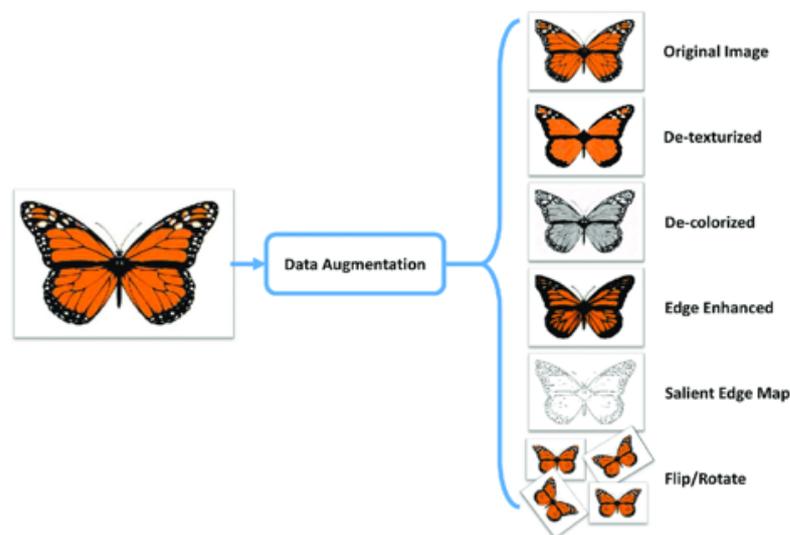


Figure 2.13: Example of data augmentation transformations in the case of image processing. From one original sample of a butterfly picture, six new ones are generated by altering the pixels of the image without altering its label “butterfly”. Source Ahmad, Muhammad, and Baik 2017.

in Yoon, Jarrett, and van der Schaar 2019. Neural networks require a large amount of training data, so they are not compatible with *small datasets*. Statistical models, such as Gaussian Trees proposed in Cao, Tan, and Pang 2014 can be used with less training data but the exploited dataset still require too much data example to be confidently used in a *small dataset* context.

GAN require for a large amount of training data to be efficient, which make them not adapted for data augmentation with *small datasets*. Therefore, in this work, we focus on data augmentation by altering existing samples to increase the data set thanks to a transformations, *i.e.* mathematical function. Iwana and Uchida 2021 proposed a review of data transformations for time series by testing data augmentation transformations on public datasets. However, they did not focus on *small dataset* nor data from industrial context nor regression task, as their study focused on classification datasets. Thus, a thorough study of the potential of data transformation for industrial prognostics has still to be completed (Fink et al. 2020).

It is shown in Rashid and Louis 2019 that the performances of data transformation depend strongly on hyper-parameters tuning (the hyper-parameters of the data transformation). Indeed, parameters optimization is an important issue to implement machine learning algorithms, as hyper-parameters optimization in any machine learning algorithm. Their optimal values may depend on the considered dataset and need to be optimized on the available training data. However, these issues have not yet been well addressed in the literature.

In the following we first precise in section 2.6.1 a methodology to evaluate the performances of various DA transformations on *small datasets* and a discussion of the results.

In a second step, we ought to address the tuning hyper-parameter optimization in an industrial context with *small dataset* for data-driven prognostics. This work was already initiated in an accepted conference article for the IFAC WC 2023.

The first step before evaluating data augmentation on the dataset is to select DA transformations to test. To do so, the quite exhaustive transformations list proposed by Iwana and Uchida 2021 is used as a base. Nevertheless, all the transformations proposed there were not adapted to prognostics or *small datasets*. The complete list and classification is available in the Figure 2.14.

Transformations are grouped into four families, which will be covered in the next subsection to identify which ones have to be tested. The Generative Models family has been ignored as discussed before. The transformations are presented hereafter with their mathematical formulations. In the following, we note a time series from the original data set  $\underline{x} = \{x_i\}_{i \in [0, T]}$  with  $T$  the time series length, *i.e.* the time of the failure.

The parametrization has been uniformized among transformations with two parameters defining the amount of new trajectories generated and their variety. Firstly,  $n \in \mathbb{N}^{+*}$  is the number of new trajectories added to the dataset. Secondly,  $p \in ]0, 1]$  defines the width of the distribution among which new trajectories are sampled. The definition of  $p$  varies among transformations and is specified in the following equations.

Finally, data augmentation on Time Series requires care on the alteration of the signal label. In the case of time series categorization, Iwana and Uchida 2021 exposed that “the data augmentation methods which transform the patterns so much that the classes are overlapped [...] reflected significant losses in accuracy”. In the case of prognostics, the label of the time series is the RUL, predicted at each time sample. In the considered C-MAPSS dataset, the RUL is accounted in number of remaining working cycles, *i.e.* number of time samples before the end of the trajectory, as shown on Figure 2.7. Some methods in the following required an alteration of the RUL signal, in which case it will be specified (mostly when the time-axis of the time series is not altered). Otherwise, the RUL signal is not modified by the data augmentation method.

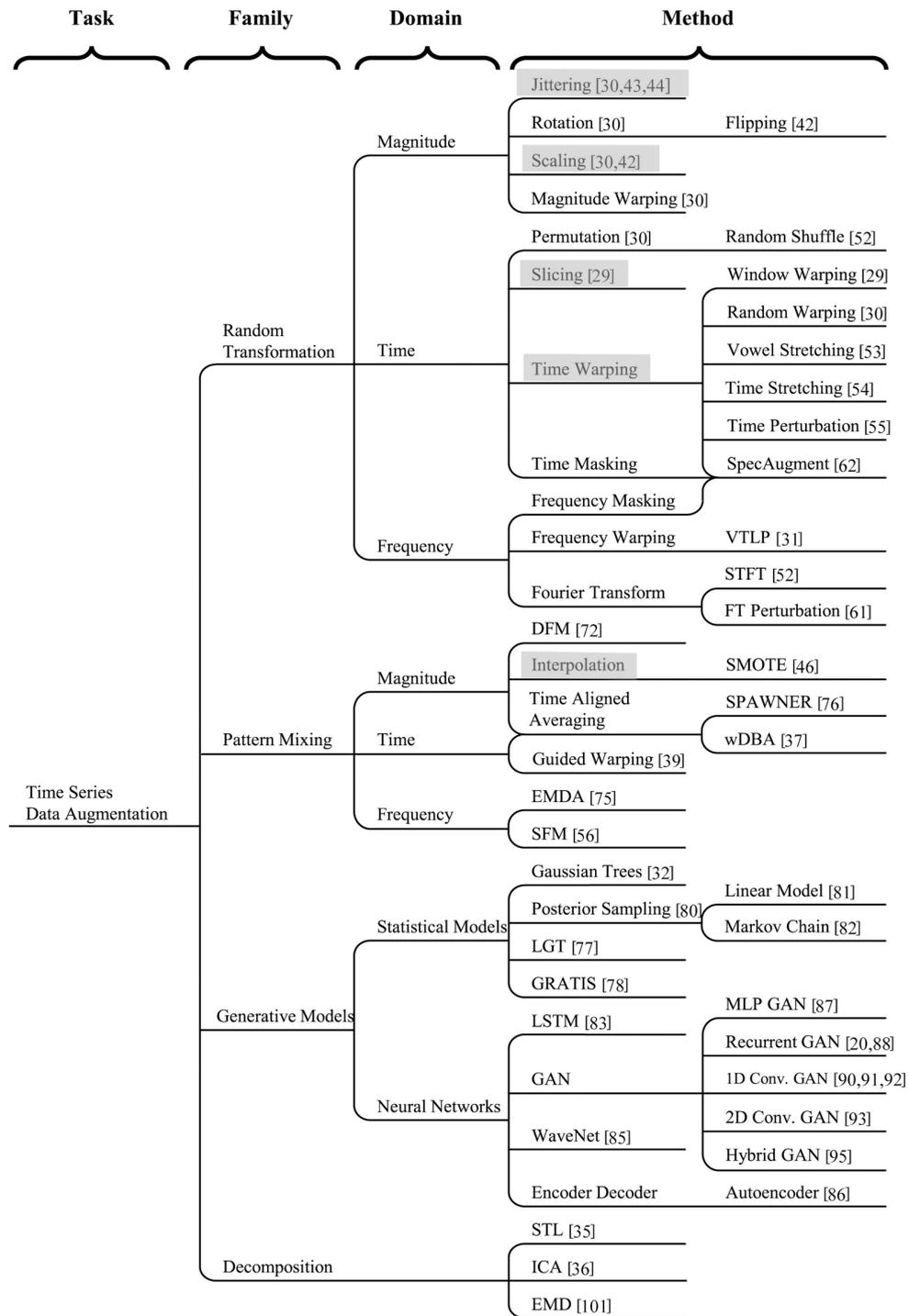


Figure 2.14: Taxonomy of time series data augmentation from Iwana and Uchida 2021. The highlighted methods are the one selected for the study. The other ones are ignored as they are not adapted to prognostics.

## 2.5.2 Random transformations family

Random transformations consist in altering time samples of the time series based on factors randomly picked.

### Random Transformations - Magnitude

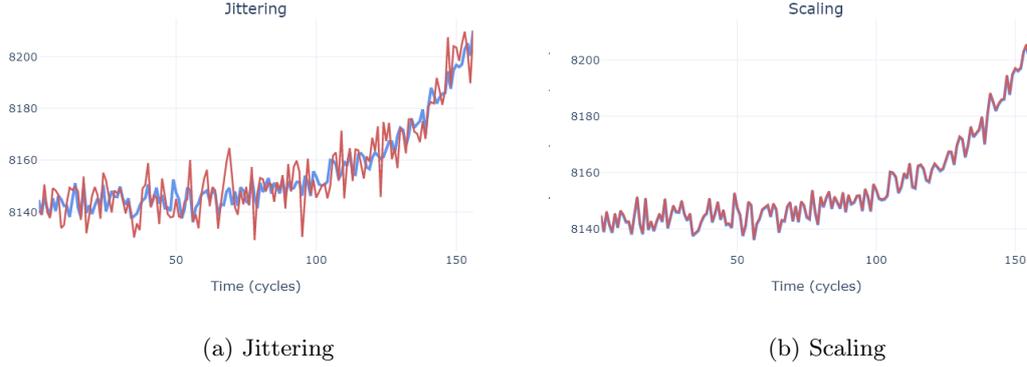


Figure 2.15: Graphical examples of the random magnitude transformations selected for the study. Graph (a) shows an example of jittering with a high value of  $p$  while (b) show an example of magnitude scaling. Original signals are in blue, augmented ones in red. Signals are from the C-MAPSS dataset.

**Jittering** consists in adding noise to the signals, it is illustrated on the C-MAPSS data on Figure 2.15a. It is one of the most common DA transformations, as it can with few effort improve model resilience to noise. It was applied to Time Series in Um et al. 2017 combined with a CNN but failed to improve performances in their application. Jittering was applied by An 1996 both on input data, output data and neural network weights, which achieved performances improvements up to 61%, at the condition to keep low values for the added noise. Random noise is drawn from a centered normal distribution of standard deviation  $p$ .

$$x_i^{jittering} = x_i + \varepsilon_i * \sigma_x \quad \text{with} \quad \varepsilon_i \sim \mathcal{N}(0, p)$$

**Rotation** like proposed in Um et al. 2017 consists either in exchanging the values of two signals to simulate an inversion in sensors (if pertinent) or in inverting the sign of signals variation, to simulate for instance an upside-down placement of a sensor, like mentioned in Rashid and Louis 2019. Rotation achieved improvements of 6.6% in Um et al. 2017, achieving the best performances among the tested transformations when evaluated individually. Nevertheless, it did not seem appropriate to the considered dataset, on which no knowledge about sensors was considered, therefore it was not used in the study.

**Scaling** consists in multiplying signals by a random factor, it is illustrated on the C-MAPSS data on Figure 2.15b. It can be efficient under the condition that the magnitude modification remains within a physically acceptable range, therefore  $p$  optimization is of huge importance. It was successfully applied to Time Series by Um et al. 2017, the first improving its prediction of 2.6%. The scaling factor  $\varepsilon$  can either be equal for all signals (*uniform scaling*) or be independently picked for every signal (*per-signal scaling*). This factor is drawn from a centered uniform distribution of width  $2 * p$ .

$$\begin{aligned}
 (\underline{x}^{(j)})_i^{uniform\ scaling} &= (\underline{x}^{(j)} * (1 + \varepsilon))_i \quad \text{with } \varepsilon \sim \mathcal{U}(-p, p) \\
 (\underline{x}^{(j)})_i^{per-signal\ scaling} &= (\underline{x}^{(j)} * (1 + \varepsilon_j))_i \quad \text{with } \varepsilon_j \sim \mathcal{U}(-p, p), \forall j
 \end{aligned}$$

**Magnitude warping** is similar to magnitude scaling excepts the multiplicative factor is time-dependent. For instance, Um et al. 2017 multiplies the signals with cubic splines<sup>5</sup> centered around one, such as shown in Figure 2.16. They achieve an improvement of performance of 2.3% with this transformation. This transformation was not applied in this chapter as the generation of efficient warping curves would require knowledge on the signals and system.

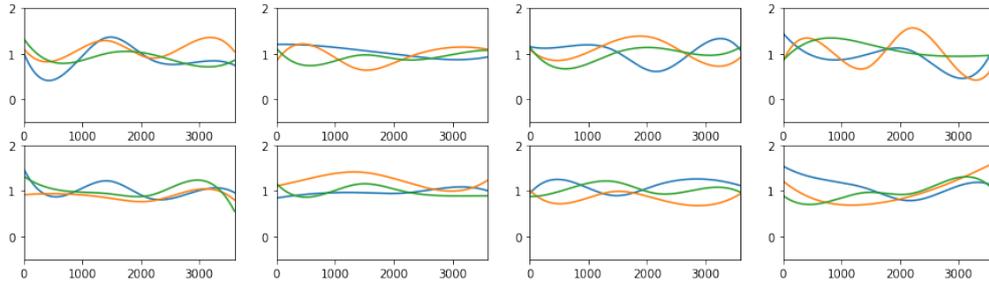


Figure 2.16: Examples of cubic splines curves used by Um et al. 2017 to generate magnitude warping augmented trajectories by multiplying them with the original signal. The image was generated by Um et al. 2017 and is available here (GitHub).

### Random Transformations - Time

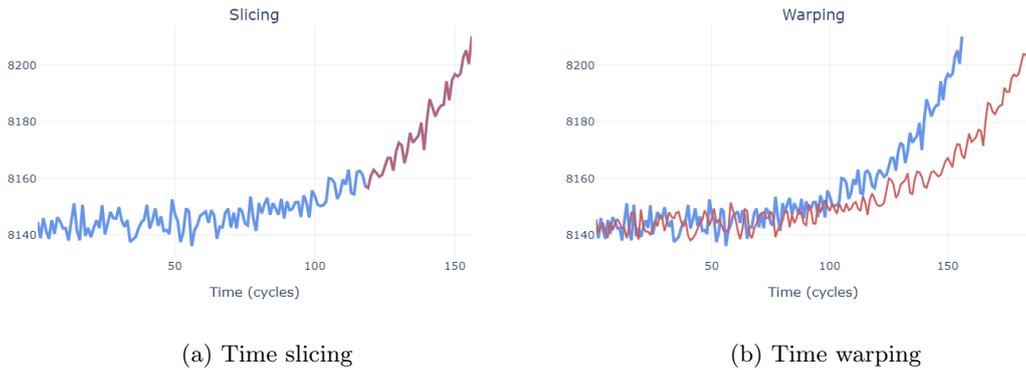


Figure 2.17: Graphical examples of the random time transformations selected for the study. Graph (a) shows an example of slicing while (b) shows an example of time warping. Original signals are in blue, augmented ones in red. Signals are from the C-MAPSS dataset.

<sup>5</sup>Scipy CubicSpline

## 2.5. Agnostic data augmentation for Time Series - Literature Survey

Among the random time transformation proposed in Iwana and Uchida 2021, *time permutation* and *time masking* were excluded of the study because they were not adapted to prognostics application. Indeed, the prognostics input data are run-to-failure trajectories, meaning time-ordered data where the order of the sample is key information, thus time permutation is not pertinent. Time masking can be efficient when used with Neural Networks for which input-masking can be an efficient method, but is less pertinent with other types of models such as the ones we use (DeVries and Taylor 2017b). For the record, time masking achieved gains up to 100% when combined with other transformations among the SpecAugment framework proposed by Park et al. 2019.

**Time slicing** consists in extracting segments of the original signal to generate new samples, it is illustrated on the C-MAPSS data on Figure 2.17a. It was successfully applied in Le Guennec, Malinowski, and Tavenard 2016, where the best results were obtained when applying time slicing, though the exact gain is hard to determine from their experiments. Extracted segments have a length of  $p * T$ .

$$\underline{x}^{t. \text{ slicing}} = \{x_j\}_{j \in [t, \min(t+s, T)]} \quad \text{with} \quad \begin{cases} s = p * T \\ t \sim \mathcal{U}(0, T) \end{cases} \quad (2.6)$$

**Time warping** consists in compressing or dilating the time axis of a time series, it is illustrated on the C-MAPSS data on Figure 2.17b. It can be done with a dilatation factor varying among time, but the proposed implementation applies a linear scaling to time dimension. It was successfully applied to time series by both Rashid and Louis 2019; Um et al. 2017, the first one achieving gain from 10% to 100% (depending on the model and metric considered), the latter achieving a gain of 6%. The linear coefficient is drawn from a centered uniform distribution of width  $2 * p$ .

$$\underline{x}^{t. \text{ warping}} = \{x_{\tau(j)}\}_{j \in [0, T]} \quad \text{with} \quad \begin{cases} \tau(x) = \alpha * x \\ \alpha \sim \mathcal{N}(1, p) \end{cases}$$

In the case of time warping, the RUL signal has to be modified. Indeed, the number of remaining working cycles for each time sample before and after the transformation is altered. Therefore, the RUL signal is recreated for the new generated signal, as the number of time samples before the end of the trajectory, as shown on Figure 2.7.

### Random Transformations - Frequency

The review also mentions methods for random transformation in the frequency space. Those can achieve great results, for instance J. Gao et al. 2021 achieves a 5% improvement with magnitude and phase augmentation on Fourier Transform. However, they were not evaluated in this work as the considered datasets do not rely on frequencies.

### 2.5.3 Pattern mixing family

**Interpolation** consists in generating new trajectories as a weighted average of two existing trajectories, it is illustrated on the C-MAPSS data on Figure 2.18a. A classical implementation of interpolation data augmentation is SMOTE, described and evaluated in Chawla et al. 2002, which obtained the best performances in 44 out of the 48 experiments conducted by the authors. SMOTE was also applied on sensor signals in Arslan et al. 2019 conjointly with jittering with great results. In this work, it was implemented as a weighted average of two time series with positive weights (whose sum is equal to one). Weights are sampled from a uniform distribution centered on 0.5 and of width  $p$ .

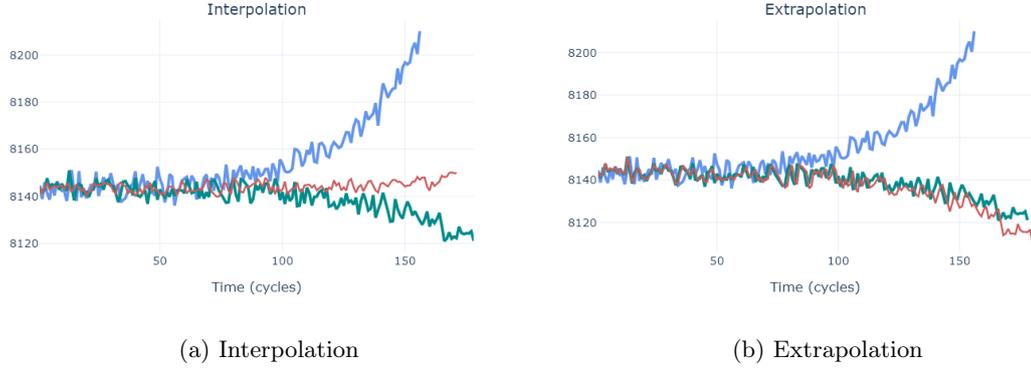


Figure 2.18: Graphical examples of the pattern mixing transformations selected for the study. Graph (a) shows an example of interpolation while (b) shows an example of extrapolation. Original signals are in blue and green, augmented ones in red. Signals are from the C-MAPSS dataset.

$$\underline{x}^{interp.} = \varepsilon * \underline{x}_a + (1 - \varepsilon) * \underline{x}_b \quad \text{with} \quad \varepsilon \sim \mathcal{U}\left(0.5 - \frac{p}{2}, 0.5 + \frac{p}{2}\right)$$

In the case of interpolation, the RUL signal has to be modified. Indeed, the number of remaining working cycles for each time sample before and after the transformation is altered, as the length of the new trajectory is the weighted average of the duration of the original ones, as illustrated on Figure 2.18a. Therefore, the RUL signal is recreated for the new generated signal, as the number of time samples before the end of the trajectory, as shown on Figure 2.7.

Additionally, **extrapolation** is an extension of interpolation allowing weights to be negatives, such that generated trajectories can go further than the space between the two original trajectories, it is illustrated on the C-MAPSS data on Figure 2.18b. It is an extension proposed in this work, which was not mentioned in the above-mentioned review. First weight is therefore sampled from a uniform distribution centered on 0 and of width  $p$ .

$$\underline{x}^{extrap.} = \varepsilon * \underline{x}_a + (1 - \varepsilon) * \underline{x}_b \quad \text{with} \quad \varepsilon \sim \mathcal{U}\left(-\frac{p}{2}, \frac{p}{2}\right)$$

In the case of extrapolation, the RUL signal has to be modified. It is done similarly to interpolation.

Interpolation and extrapolation are illustrated for an application on images on the Figure 2.19 from the paper of DeVries and Taylor 2017a.

Similarly to random transformations, the pattern mixing technics can be applied to the frequency domain if pertinent for the application, which was not done here due to the datasets considered.

## 2.5.4 Decomposition family

Finally, the last family is decomposition, consisting in decomposing the signals in sub-signals and applying simple data transformations to one or more of the decomposed signals. The purpose is to keep the support of the information to ensure the pertinence of the generated trajectories.

**STL** decomposition consists in splitting the signals in three components representing the seasonality (S) of the signal, *i.e.* its periodic part, then its trend (T), *i.e.* global change of the

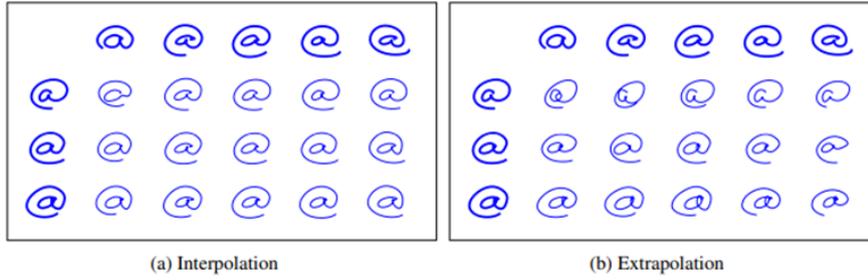


Figure 2.19: Interpolation (a) and extrapolation (b) between handwritten characters. Character  $(0,i)$  is interpolated/extrapolated with character  $(j,0)$  to form character  $(i,j)$ , where  $i$  is the row number and  $j$  is the column number. Original characters are shown in bold. Image and text from DeVries and Taylor 2017a.

average value, finally its remainder ( $R$ ), *i.e.* the rest of the signal not explained in the first two components. Data augmentation can therefore be applied classically on the remainder to modify the random noise, like in Bergmeir, Hyndman, and Benítez 2016, or on the trend to modify the degradation profile. The seasonality signal could also be modified with frequency-based transformations.

**ICA (independent component analysis)** decompose the signals in Independent Components, on which DA transformations can be applied, before reconstructing new signals. Transformations such as interpolation can be especially efficient as it could fill the space between samples in the latent space of the decomposition. It was successfully applied in Eltoft 2002.

The decomposition technics are however not studied in this work because we want to investigate perturbation from technical origins. They remain an interesting topic for further work on data augmentation for *small dataset* on prognostics.

Finally, Table 2.7 references papers for each mentioned data augmentation transformation. In our study, all the above-described transformations are defined by the two parameters  $n$  and  $p$ , whose value should be optimized to achieve the best performances on a given dataset. The next section addresses this issue when a single transformation is applied.

Table 2.7: List of references for each DA transformations

Family	Transformation	References
<b>Rand. Magnitude Transformation</b>	Jittering	Um et al. 2017; An 1996; Arslan et al. 2019
	Rotation	Um et al. 2017; Rashid and Louis 2019
	Scaling	Um et al. 2017
	Mag. Warping	Um et al. 2017
<b>Rand. Time Transformation</b>	Slicing	Le Guennec, Malinowski, and Tavenard 2016
	Time Warping	Rashid and Louis 2019; Um et al. 2017
<b>Pattern Mixing</b>	Interpolation	Chawla et al. 2002; Arslan et al. 2019
	Extrapolation	
<b>Decompositions</b>	STL	Bergmeir, Hyndman, and Benítez 2016
	ICA	Eltoft 2002

## 2.6 Single-transformation data augmentation

The data augmentation (DA) transformations being identified and defined, the current section focuses on evaluating them individually. It first defines the test methodology for the evaluation of data augmentation performances for prognostics on *small datasets*. Then it applies it on the C-MAPSS dataset with the selected prognostics models: the LSTM-based, ROCKET and MiniROCKET.

The objectives are:

- To measure by how much performances are improved when a model is trained on an augmented dataset, as well as the dependency on the dataset sampling
- To evaluate the dependency of the improvement on the *small dataset* distribution, given the previously-observed dispersion of performances among sub-samplings (see 2.4).
- To evaluate the impact of data augmentation hyper-parameters ( $n$  and  $p$ ) on performances, as well as question their optimization in an industrial context
- To determinate if the combination of DA transformations can improve even further performances.

To do so, the test methodology is defined in the following sections.

### 2.6.1 Single-transformation data augmentation methodology

Given a full dataset, it is sub-sampled to get subsets considered as *small datasets*, accordingly to the methodology proposed in section 2.2. The C-MAPSS FD001 dataset is sub-sampled down to 25%, *i.e.* 25 run-to-failure trajectories. Sub-sampling picks are done randomly. To ensure representativity of the results, multiple picks are considered according to the values computed in subsection 2.4.1 and summarized in Table 2.5. Like in subsection 2.2.2, to ensure the repeatability of experiments, sub-samplings are picked with a fixed random seed. Fixing the random seed before picking ensure that one random pick is always the same among experiments.

In a first round, DA transformations are tested individually. Later we will consider combining multiple data-augmentation transformation. Therefore, given one sub-sampling, one DA transformation is applied with parameters  $n$  and  $p$ , generating  $n$  new trajectories. The new dataset of size  $25 + n$  is then used to train the prognostics model, which is then tested on the test set to obtain the RMSE corresponding to the data augmentation. As both data augmentation and model initialisation are random processes, this process is repeated  $n_{augm} = 10$  times for every sub-samplings. Therefore, the number of training is  $n_{picks} * n_{augm}$  from which are determined the average RMSE and boxplot of the DA transformations.

The augmentation and training process is summarized visually on the schema of Figure 2.20.

Finally, data augmentation is parametrized by two parameters  $p$  and  $n$  previously introduced. Their impact on performance and their optimization is a topic of importance for performances improvement. It is then the subject of the next section.

### 2.6.2 Hyper-parameters optimization

Parameters optimization has to be done with great care. Indeed, in a *small dataset* industrial context, no absolute test set is available to evaluate efficiently prognostics performances for each tested values of  $n$  and  $p$ , contrarily to the public datasets used for this study. In industrial situation, hyper-parameters are optimized on the available data history (training data) with

## 2.6. Single-transformation data augmentation

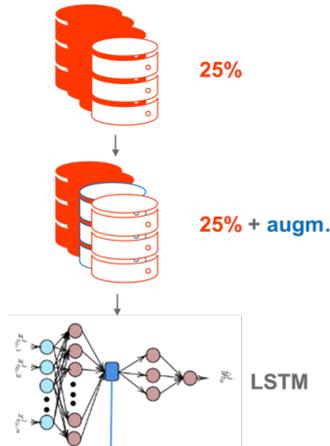


Figure 2.20: Schema of the data augmentation test process. Step 1 is the reduction of the size of the dataset into sub-samplings, repeated  $n_{picks}$  times. Step 2 is the augmentation of each sub-samplings with one augmentation (out of  $m$  transformations), repeated  $n_{augm}$  times per DA transformation. Step 3 is training the prognostics model with the  $n_{picks} * n_{augm} * m$  obtained augmented subsets and measure their performances on the test set.

the objective to obtain the best performances on future data (test data, not seen in training). Selecting a test set is usually done but there are no way to confidently know if it accurately represents the true data distribution. Therefore, particular attention has to be paid on two main points of the test methodology:

- Firstly, parameters optimization is run considering exclusively the training dataset.
- Secondly, to get an idea of in-use performances on new data, best parameters are evaluated on a previously-unseen test dataset. It represents prognostics performances on new data once the prognostics model would be implemented on-line.

To match these constraints, the following test methodology is used to optimize hyper-parameters on each sub-sampling.

Given a random sub-sampling of the turbofan dataset, the tuning parameters of DA transformations are optimized using a 5-fold cross-validation (train set). The optimization research is done using Optuna, a Python library developed by Akiba et al. 2019. Parameters optimization range are  $p \in ]0, 1]$  and  $n \in ]0, 200]$ . For each parameters' values tested, data transformation and model training are repeated 10 times ( $n_{augm}$ ) to ensure results independence against the randomness due to data transformation and model weights initialization. The average RMSE value out of the ten trainings is used as the **cross-validation RMSE** for the considered parameters' values.

Finally, the parameters' "best values" are used to train the prognostics model on the whole sub-sampling and test it on the test set to obtain the **test RMSE**. This last test is repeated  $n_{augm}$  times like all previous results.

### 2.6.3 Single-transformation data augmentation results

Results in this section focuses on individual picks performance improvements. Indeed, results are not consistent among them. This leads to the question of performance dependence to samples distribution in the context of *small datasets*.

Results then focus on the difference of performances between cross-validation and test set, which degrade sometimes importantly. This is an important topic in industrial context considering that model optimization must be done on available data (through cross validation) to be performant on future data (test set). Especially in the case of prognostics, which results will be used to support impactful decision-making.

Then, results focus on the optimal values of hyper-parameters and their variability among optimizations and picks. Firstly their impact on performances is asserted. Once it is verified they impact performances, they are optimized on the sub-samplings. However, it will be shown that it is not the case and that their values are largely spread. The impact of this dispersion will be questioned.

The next subsections will first do a thorough study of those questions on the C-MAPSS FD001 dataset with ROCKET, with extended analysis, to tackle precisely all interrogations linked to the methodology and results. The work will be extended to LSTM model as well, to generalize the analysis on multiple cases, in a shorter manner so to keep the manuscript concise and efficient.

### Hyper-parameters optimization

One should be careful not using the results of this part for parameters choice, as performances are measured on the test set which is considered not available during optimization step. However, test-set performances are useful in this study context to estimate the gap between performances obtained during optimisation and performances obtained on future new data.

First experiments are done to observe the impact of parameters on the performances and are shown in Figures 2.21a and 2.21b. Interpolation and time slicing, two of the most performant transformations for LSTM are presented with various values of  $n$  and  $p$ . The figures show the performances with  $n$  fixed to  $n = 65$ , and  $p$  varying from 0 to 1.

Those results confirm hyper-parameters impact on data augmentation performances and therefore the importance of their optimization. For interpolation, average performances improvement achieved between 16% and 13% depending on the value of  $p$ , which is not significant. However, for time slicing they achieved up to 3% but could also degrade performances as much as -9.7% when an extreme window size is selected (slice of 10% of the time series). The selection of the best value of  $p$  requires to apply the described methodology using cross-validation.

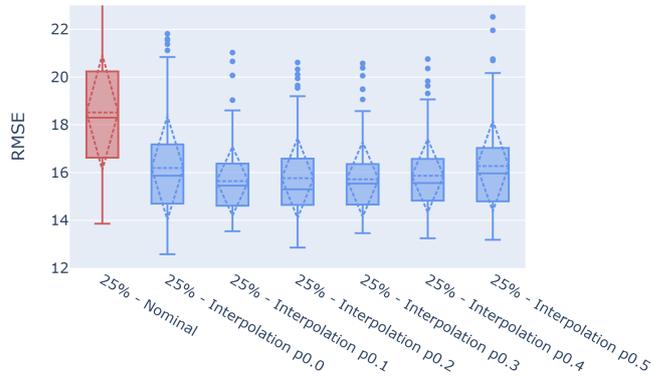
### Results on Cross Validation

*The optimization of hyper-parameters like described in the methodology section 2.6.1 is done using ROCKET for its learning speed. Indeed, the hyper-parameters optimization lead to many trainings as many hyper-parameters values are tested. Therefore, the fastest model was preferred for a thorough exploration, keeping larger models for further experiments.*

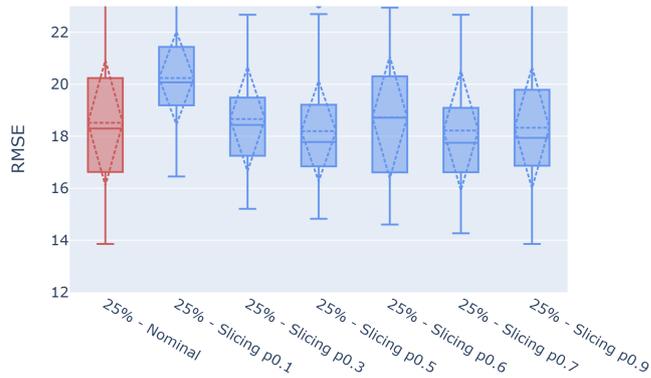
Data augmentation hyper-parameters are optimized on ten independent 25% sub-samplings of the C-MAPSS dataset. Prognostics performances obtained with cross-validation are summarized in Figure 2.22. The obtained results firstly show the high variability of prognostics performance among small datasets, with cross-validation RMSE varying from 15 to 17 on un-augmented data. Similarly, performance improvements from data transformation vary from a subsampling to another.

All transformations improve prognostics performances, but not to the same scale as improvements achieve from 2% up to 15% gain. Data augmentation transformations ranking is consistent across sub-samplings, with extrapolation, interpolation, time warping and jittering systematically among the best transformations. Meanwhile, scaling transformations (both uniform and signal-specific) are hardly changing prognostics performances compared to un-augmented dataset.

## 2.6. Single-transformation data augmentation



(a) Interpolation performances for various values of its hyper-parameters  $p$ .



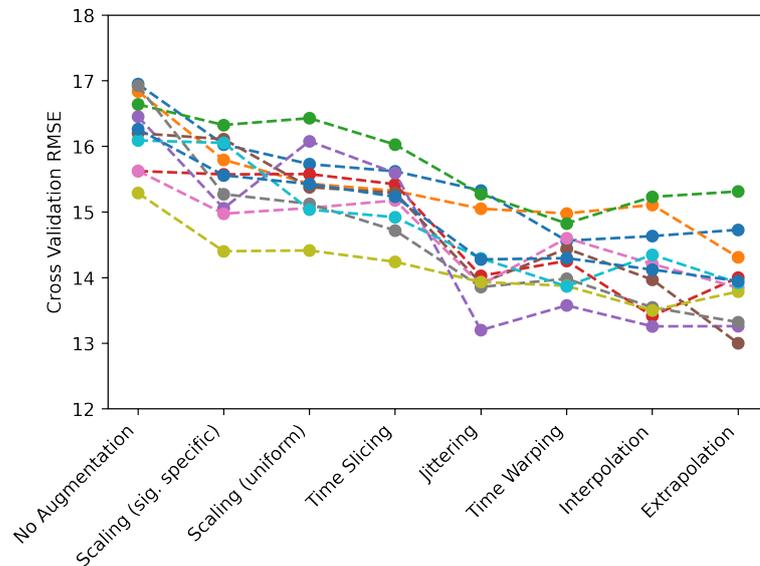
(b) Time slicing performances for various values of its hyper-parameters  $p$ .

Figure 2.21: Performances of LSTM model trained on C-MAPSS dataset augmented with (a) interpolation and (b) time slicing, with various values of their hyper-parameters  $p$ , given  $n = 65$ . Those results were not obtained with cross-validation, and therefore should be considered with care as of their absolute value.

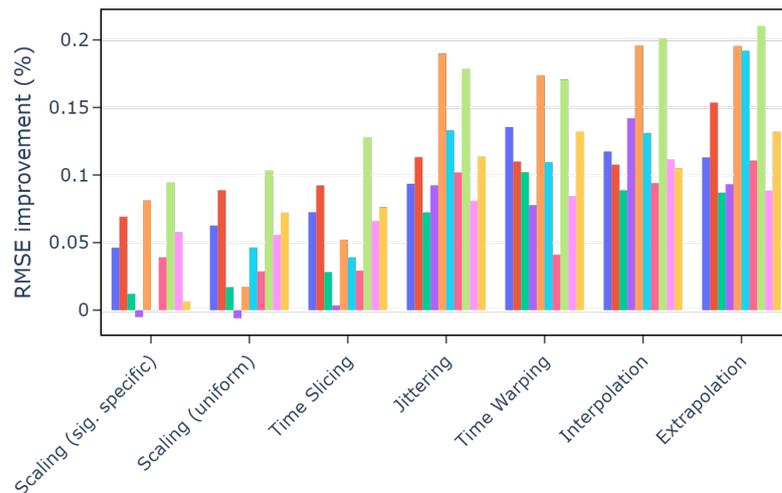
Cross-validation results are encouraging and may be good enough to allow for decision-making improvement, as will be studied in chapter 4. However, those results might be biased as the displayed RMSE were used by the optimizer as objective function. Therefore, the next step is to use those parameters on the test set to validate performances gain.

### Results on Test set

ROCKET is trained on the ten independent 25% sub-samplings augmented with their respective optimal hyper-parameters and evaluated on the test set. This methodology is representative of an industrial usage of data transformation, the test set acting like future model performance while in



(a) Average RMSE per pick. Far-left points are un-augmented RMSE.



(b) RMSE improvement from un-augmented dataset. Each bar corresponds to one pick, measuring the improvement from un-augmented dataset in percentage.

Figure 2.22: Cross validation performance gains obtained with data augmentation applied to 25% of C-MAPSS FD001 for ROCKET training. The evaluation is done on 10 independent sub-samplings of the dataset, optimization being done for each of them through cross-validation. (a) shows average performances obtained for each pick for each optimised DA transformation. (b) shows performance gain obtained for each pick for each optimised DA transformation.

application. Prognostics performances obtained are summarized in Figure 2.23, firstly measured as the obtained average RMSE and secondly as performances improvements.

Contrarily to cross-validation, results on the test set are more mitigated and dependent to picks. Indeed, some transformations improve prognostics performances while other reduce them, sometimes drastically, depending of the dataset pick. The most efficient transformations on cross-validation (interpolation, extrapolation and jittering) still achieve positive improvements on the test set. However, two transformations (time slicing and time warping) reduce (in average) prognostics performances on the test set whereas they improved them on cross validation with a large dependence to the pick dataset.

Finally, two transformations are systematically improving test performances (interpolation and jittering), with improvements from 5% to 10%. Other transformations performances are not consistent among sub-samplings. For instance, extrapolation average improvement is 5% but decreases performances down of 7% for one sub-sampling. Those results confirm test results are largely dependent on the distribution of the 25% subsampling used for training. This confirms the results already obtained in Gay, Voisin, et al. 2022.

#### 2.6.4 Analysis of hyper-parameters optimization results.

The dependence of results to the sub-samplings implies that hyper-parameters optimal values may vary among sub-samplings. To this extent, the current subsection analyses hyper-parameters optimization. It compares optimal hyper-parameters between sub-samplings. As well as the dependence of performances on hyper-parameters values inside one sub-sampling.

The optimization of hyper-parameters among various sub-samplings highlights the following questions:

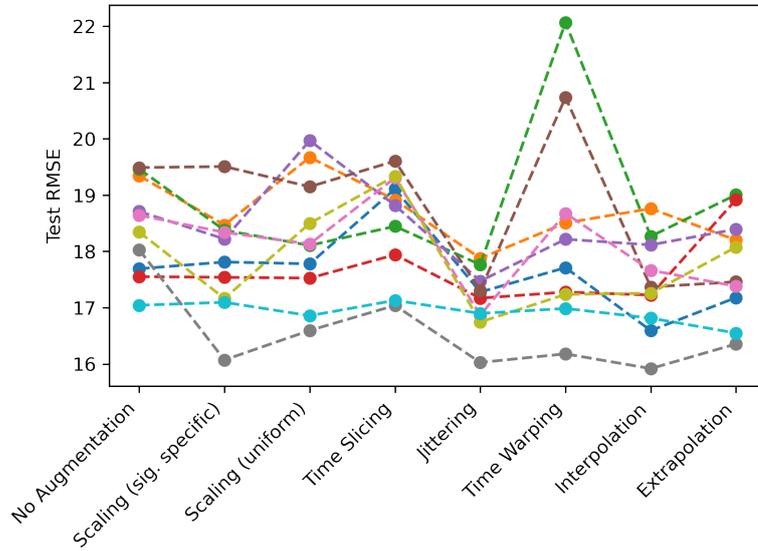
1. Given a DA transformation and a subsampling, how are prognostic performances distributed for various values of the hyper-parameters?
2. Further, is this distribution similar among sub-samplings?
3. Finally, are those observations similar among transformations? Or are some transformations more prone to variability of their optimal parameters?

To answer the first two questions, performance distributions are plotted in Figure 2.24. The figure shows the distribution of performances with **cross-validation** for the ten sub-samplings used previously among two dimensional axis corresponding to the two optimised parameters  $n$  and  $p$ . In the following, this two-dimensional representation will be called *the  $n - p$  space*.

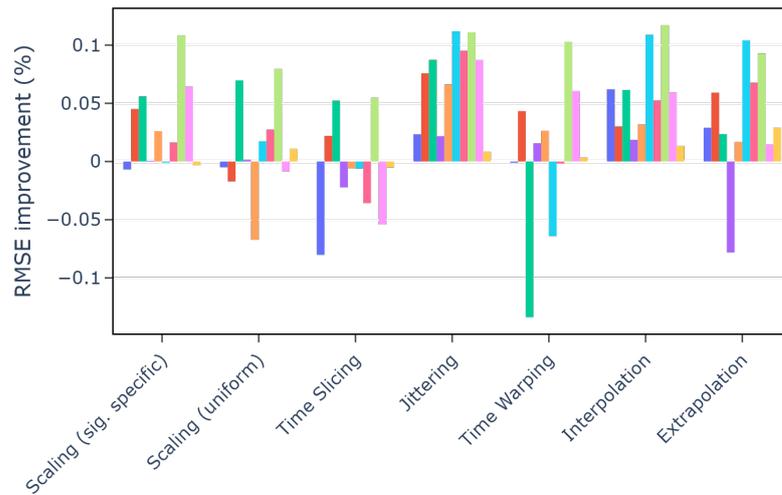
The Figure shows that optimal values are not consistent among sub-samplings, nonetheless some observations and rules are shared among them. Indeed, best performances appears in different regions of the  $n - p$  space depending of the sub-sampling. However, some general rules can be extracted, for instance  $n < 50$  mostly provides low performances, as well as  $p > 0.8$ . Nevertheless, sub-sampling 4 gets its best performances as  $n \simeq 190$  and  $p \simeq 0.1$  (see Figure 2.24) while the same values provides poor performances for sub-samplings 0, 2, 6. This disparity is caused by the diversity of trajectories among sub-samplings and partially explain the mitigated performances obtained on the test set.

## 2.7 Multi-transformation data augmentation

After obtaining results on individually-applied data augmentation, it is ought to evaluate the performances gains when DA transformations are combined. Indeed, to face the mitigated results



(a) Absolute RMSE. Far-left points are un-augmented RMSE.



(b) Performances improvement from un-augmented dataset.

Figure 2.23: Test performance gains obtained with data augmentation applied to 25% of C-MAPSS FD001 for ROCKET training. The evaluation is done on 10 independent sub-samlings of the dataset, using optimized hyper-parameters. (a) shows average performances obtained for each pick for each optimised DA transformation. (b) shows performance gain obtained for each pick for each optimised DA transformation.

## 2.7. Multi-transformation data augmentation

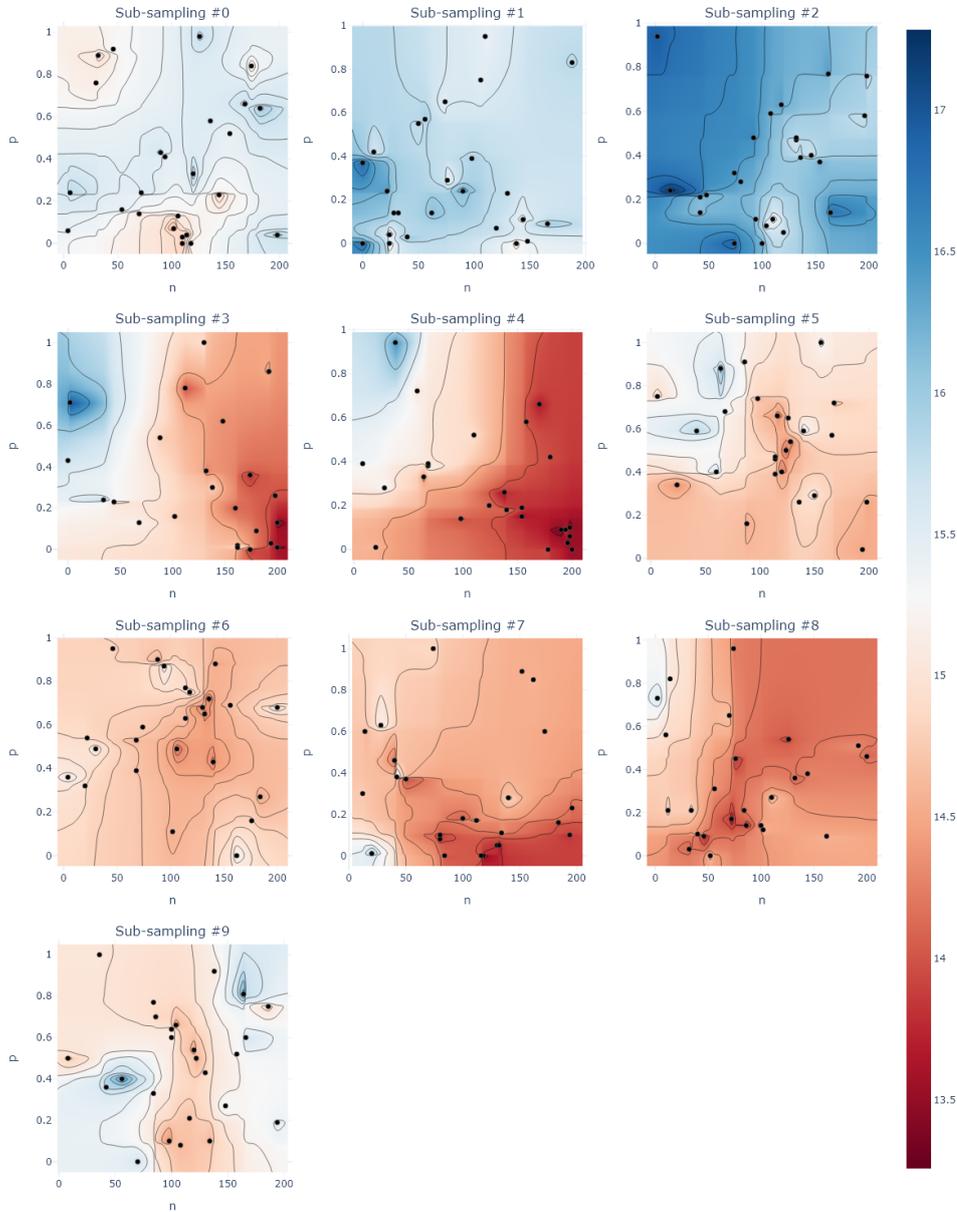


Figure 2.24: Distribution of **interpolation** performances among sub-samplings, depending on parameters values, extrapolated from optimizer results. Each graph represents a sub-sampling from the C-MAPSS dataset. Best performances are brighter, colors are relative to each graph, the darkest and brightest blue are for the worst and the best performance for the given sub-sampling. Black points corresponds to evaluated values, contours are extrapolated around them. This Figure illustrates the variability of optimal values from a sub-sampling to another.

on the test set using a single transformation, we shall consider to improve the diversity of the generated trajectories using several DA transformation in sequence. This method has been shown to be effective in the literature. For instance in Um et al. 2017, the authors obtained the best results on single transformation with rotation and an accuracy of 82.62, and went up to 86.88 when using both rotation, permutation and time warping. However, adding transformations does not always perform better than single DA transformations as they obtained only 85.60 when using the same rotation + permutation + time warping and magnitude warping, whereas magnitude warping improved results when used single. This result is still better than the baseline.

### 2.7.1 Multi-transformation methodology

Using data augmentation transformations concurrently poses the question of the order of application. For instance, interpolation takes the hypothesis that the two original trajectories are run-to-failure. So if time slicing is applied before it, this requirement is not met anymore. Resulting samples are then not representative of run-to-failure trajectories, but are considered as such.

Therefore, the first step is to define the order of application of the DA transformations. This order could be considered as a hyper-parameter to optimize. However, this would lead to costly tests in terms of computation power, and would require too much time for experiments. Therefore, we define the application order of data augmentations transformation. This order is used in the multi-augmentation experiments.

Data augmentation implementation is done in sequence such that a transformation placed after another one takes as “input data” the dataset augmented by the previous transformation. Given a transformation  $\mathcal{A}$  coming before transformation  $\mathcal{B}$ ,  $\mathcal{A}$  augments only the original trajectories, while transformation  $\mathcal{B}$  augments both original trajectories and trajectories generated by  $\mathcal{A}$ . For instance, if magnitude scaling is set after time warping, resulting trajectories will be (1) original ones, (2) scaled ones, (3) warped ones and (4) scaled + warped ones.

We fixed the sequence of DA application in order to avoid disturbance among transformations. The following constraints are identified:

1. Interpolation and extrapolation are set in parallel, meaning they take same input trajectories and the generated trajectories are concatenated afterward. To limit the risk of interpolation / extrapolation creating irrelevant trajectories as a mix of an original and a (potentially irrelevant) generated trajectory, they are set first.
2. Time warping is set before magnitude transformations.
3. Magnitude transformations (scaling uniformly and per signal) are set in parallel to avoid over-modifications of magnitudes.
4. Jittering is set at the end, such that noise is added at the end of the process, limiting the risk of perturbation of other transformations based on statistical characteristics of the signals.
5. Time slicing is set last, because any following transformations would be applied to truncated signals, whose characteristics are not representative of the whole signal.

The final data augmentation process is represented in Figure 2.25.

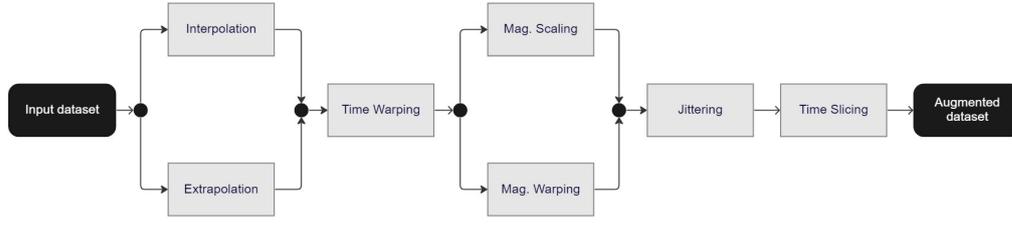


Figure 2.25: Schema of the multiple-data augmentation process, indicating in which order are applied the transformations. Each transformation uses as original trajectories to augment the trajectories supplied by the previous step in the chart. In the case of parallel transformations, generated trajectories are merged after application.

### 2.7.2 Multi-transformation evaluation perspective

Multi-transformation data augmentation evaluation requires a great amount of computational power to obtain robust results. Indeed, hyper-parameters optimization is as much, if not more, important as with single-transformation in order to get efficient results. The required number of runs to optimize DA hyper-parameters is much larger than for single-transformation as they all have to be optimized in the same time. Indeed, considering two parameters  $p$  and  $n$ , the search space is two-dimensional for single-transformations but it consists of 14 dimensions when considering the seven transformations in the meantime. Unfortunately, not enough computational time was available to obtain robust results on multi-transformation DA for this study. It was however attempted to use the optimal parameters identified ofr single-transformations as base values, which unfortunately did not achieve pertinent results: prognostics performances were degraded.

Therefore, the thorough exploration of multi-transformation DA remains an issue to tackle for future works on the subject of data augmentation for prognostics on *small datasets*.

## 2.8 Conclusion

This chapter aims to evaluate the impact of *small datasets* on prognostics performances and study the capability of agnostic data augmentation as a solution to improve them without using system-knowledge. It is evaluated on the C-MAPSS dataset with two prognostics models, an LSTM-based acting as state-of-the-art on the dataset and ROCKET, a promising model considered for fastening results computation. To this end, thorough test methodologies are proposed to simulate *small datasets* and confidently evaluate data augmentation.

Firstly, *small datasets* are obtained as random independent sub-samplings of a *full dataset*. Sub-sampling being a source of variance on the results, it is repeated multiple times to ensure the representativity of the results. Similarly, as model initialization is a source of variance, especially on *small datasets*, model initialization and training is also repeated multiple times on each sub-sampling.

Secondly data augmentation transformations are optimized by an industrial-based methodology using cross-validation. Indeed, considered data augmentation transformations are parametrized by two hyper-parameters. Their impact on prognostics performances being asserted, they are

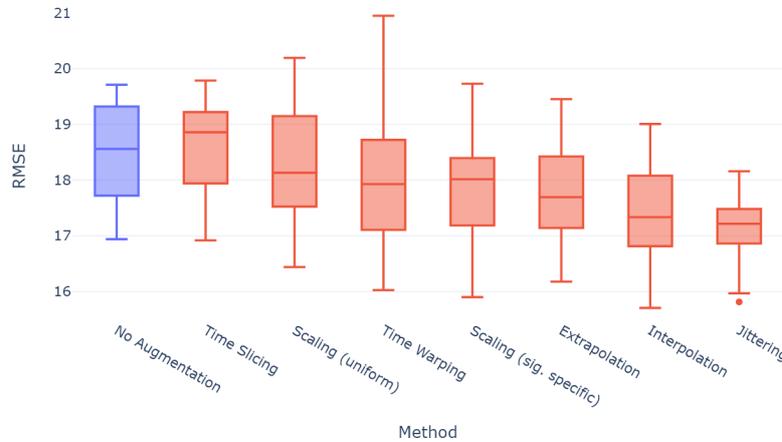


Figure 2.26: Best results obtained on a *small dataset* for various data augmentation transformations. The *small dataset* is obtained as a 25% subset of the C-MAPSS dataset. Prognostics performances are measured as a RMSE (Root Mean Square Error, to be minimized). Prognostics is done with the ROCKET model. The box in red corresponds to the performances on the un-augmented dataset.

optimized for each sub-sampling. This optimization is done through a cross-validation approach as test set should not be used for such a task. A test set is finally used for performances evaluation. This optimization is considered for individually-applied transformations.

Thirdly it is defined an order for multiple data augmentation transformations applied simultaneously. Indeed, the order of application of multiple transformations impacts their final performances. It is defined a fixed order based on the interaction between methods to reduce the risk of performance degradation.

A first contribution successfully links prognostics performances to dataset size through an equation. The proposed equation, inspired by Vapnik, Levin, and Cun 1994, is written in Equation (2.5). It can be used further to predict expected performance gain from additional data when only *small datasets* are available.

A second contribution confirms the improvement of prognostics performance through the application of individual agnostic data augmentation up to 10%. All selected transformations do not achieve similar performances improvements, with some of them decreasing performances. Great improvements are systematically obtained on cross-validation, but they are not systematically observed on the test set (which act as future data for industrial systems). Finally, an opening is proposed on the simultaneous application of multiple DA transformations. All these results are summarized in Figure 2.26.

Finally, this chapter highlights the required care when applying data augmentation to industrial systems, which can be addressed with the exploitation of expert knowledge on the systems. These results point out the importance as well as difficulty of DA transformation selection and optimization for when applied to *small dataset*. Proposals to address this issue are mentioned in Fink et al. 2020. One of them is the exploitation of further knowledge thanks to experts when available on industrial systems. For instance, the two most efficient transformations identified are jittering and interpolation, which are consistent with the physics of the system. Indeed, jittering adds noise to already noisy data, while interpolation generates new trajectories as mix of existing

ones, increasing their potential to correspond to real data.

Therefore, Expert Knowledge capacity for both DA transformation selection and optimization is the subject of the next Chapter 3.



## Chapter 3

# Knowledge-based data augmentation for prognostics

### 3.1 Introduction

To face industrial *Small datasets* for prognostics, the previous Chapter 2 proposed to use data augmentation. Data augmentation has been applied with no use of prior knowledge on the datasets neither on the system, to get general results and methodologies, called “agnostic data augmentation”. Agnostic data augmentation has proven efficient, although its implementation requires care as to ensure prognostics improvements on future data. Indeed, prognostics performances improvements present high variability, which can even result in prognostics degradation.

This chapter ought to face this reliability issue of data augmentation for prognostics through the use of expert knowledge. Expert knowledge has the potential to fit augmentation methods more closely to datasets. Two sources of knowledge are considered in the following. First, knowledge available from experts of the industrial systems, such as physics models, functional or dysfunctional analysis. Second, statistical knowledge gathered from advanced data analysis used to identify system’s data behaviour and empirical rules.

These two sources of knowledge are used to tune data augmentation transformations used in the previous chapter to meet system (/dataset) specificities. Those methods are then evaluated using the test methodologies defined in the previous chapter and compared to agnostic transformations.

To this extent, this chapter proposes three **upgrades** to data augmentation transformations.

1. Exclusion of constant and discrete signals from magnitude-based transformations.
2. Clustering of signals by similarity to apply similar scaling factor in magnitude-scaling transformations.
3. Definition of a measure of the distance between augmented and original dataset to identify and discard aberrant generated trajectories.

The implementation of these upgrades in the chapter does not improve significantly prognostics performances but successfully avoid performances degradation under unfavorable conditions. Indeed, under some conditions agnostic data augmentation degrades prognostics performances. This is the case for some random sub-samplings and some values of the hyper-parameters  $n$  and  $p$  of the transformations as defined in the previous chapter. It is seen in the current chapter that the simultaneous application of the three proposed upgrades reduces drastically the occurrence

of prognostics performance reduction due to data augmentation. To this extent, the chapter is composed as follows.

The first section 3.2 lists, describes and formalizes the available knowledge for prognostics on the C-MAPSS dataset. Firstly, knowledge is considered as expert knowledge in subsection 3.2.2. Secondly, it is considered as statistical knowledge in subsection 3.2.3. This section formalizes three upgrades of data augmentation transformations to exploit this knowledge.

The second section 3.3 evaluates the performances of the three proposed upgrades of data augmentation transformations on the C-MAPSS dataset and compares the results with agnostic data augmentation. The evaluation is done accordingly to the test methodologies defined in the previous chapter.

## 3.2 Description and formalization of available knowledge on dataset

This section summarizes the knowledge available for optimizing prognostics on the C-MAPSS FD001 dataset.

Based on the work proposed in our paper published during the 2021 ICSRS conference (Gay, Benoit Iung, et al. 2021), the exploitable knowledge has been grouped in two categories: “expert” knowledge and statistical knowledge. Expert knowledge corresponds to knowledge possessed by industrial experts (line and maintenance engineer, R&D researchers, operators, . . . ) concerning the behavior of the system. One of the main way to exploit it, as mentioned in Gay, Benoit Iung, et al. 2021, is the exploitation of functional and dysfunctional analysis to create coherent data augmentation. Statistical knowledge corresponds to the exploitation of advanced statistical analysis of the dataset to tune data augmentation transformations to fit the data profile. It requires less knowledge of the system or physics, but rather data analysis inquiry in the dataset. It can be more easily applied on datasets when experts are not available. It is therefore particularly adapted to the C-MAPSS dataset as few expert knowledge on turbofan is available on our side.

Next subsection 3.2.1 shortly describes some expert knowledge integration in data augmentation from the literature. Then, the two following subsections describe the two types of knowledge (“expert” in 3.2.2 and “statistical” in 3.2.3). Finally, the transposition of this knowledge into data augmentation methods upgrades is done in subsection 3.2.4.

### 3.2.1 Introduction to expert-knowledge for data augmentation

Data augmentation has previously been applied with expert knowledge, such approaches are known as physics-based data augmentation. It is for instance done manually in the image processing field to add defaults on pictures based on expert knowledge. Such works were proposed in the medical field, Dahiya et al. 2021 generates a full dataset of high quality computed tomography images for segmentation based on medical knowledge on the field from low quality recorded ones. Their approach proves efficient to improve dataset quality and therefore prediction accuracy. Similarly, Alam et al. 2021 use medical knowledge to guide artifact detection on real data. Such approaches relate to model-based simulation, like is done in Pan et al. 2022, where a degradation is simulated on a system to extend the available data samples.

Expert knowledge about the system is also used to adapt augmentation methods to fit the expected behaviour of a system. Such methodology is applied by Omigbodun et al. 2019 to transform signals according to what is expected in considered default cases. Unfortunately, their implementation does not perform better than standard data augmentation. Similarly, Singla

et al. 2022 applies a physics-informed selection of standard augmentation transformations to an ultrasound dataset to improve model accuracy.

### 3.2.2 “Expert” knowledge on turbofan systems for the C-MAPSS dataset

This study is conducted on the C-MAPSS dataset to allow comparison with previous results, therefore expert knowledge has to be acquired on the dataset. In the case of real-life applications, expert knowledge comes from people working with or conceiving the systems. In the case of the public dataset, such experts are not available directly, therefore another source of knowledge has to be sought. Hence, it is tried to gather expert knowledge from the description of the dataset generation supplied in the original paper of Abhinav Saxena et al. 2008. The extracted knowledge comes from the C-MAPSS simulation tool structure and from the scenarios used to generate the dataset. Both sources are discussed in the following paragraphs.

The dataset is generated using a turbofan simulation software called *C-MAPSS* developed using Matlab and Simulink, which structure can be exploited as a source of knowledge. The simulation is further described in its authors’ user guide Frederick, DeCastro, and Litt 2007, where the system’s components and interactions are shortly described and illustrated, summarized in the schema of the Figure 3.1. The simulation description could have been a source of information about the link between components and therefore signals once linked with sensors. This knowledge would have been similar to what could be obtained from a functional analysis, like mentioned in Gay, Benoit Iung, et al. 2021. Those causal links between signals could have been exploited as rules to be asserted on data augmentation generated trajectories. Unfortunately, this approach could not be applied as the dataset authors did not specify in their paper nor documentation the signification of the supplied signals. Therefore, it is impossible to link signals and components to exploit the system functional analysis. However, in ArcelorMittal use-cases, functional analysis and physics models are available and could be used for such applications.

The second source of expert knowledge on the C-MAPSS dataset is the original paper Abhinav Saxena et al. 2008 describing the run-to-failure trajectories creation. The paper describes the degradation model applied in the simulation. The authors use “an exponential term while modeling changes of health parameters in C-MAPSS” resulting in a health equation written as  $h(t) = 1 - \exp(at^b)$  with  $h(t)$  the degradation index such as 1 corresponds to a healthy system and 0 to the threshold of the failure happening. This knowledge of the exponential nature of

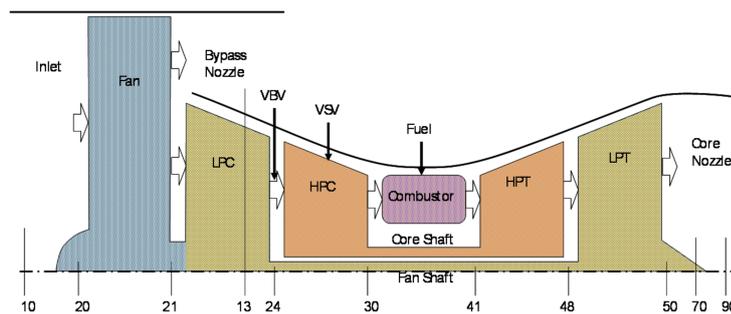


Figure 3.1: Schematic diagram showing the components and station numbers of the C-MAPSS engine. LPC: low pressure compressor, HPC: high pressure compressor, HPT: high pressure turbine, LPT: low pressure turbine. Source Frederick, DeCastro, and Litt 2007

the degradation could be used to apply data augmentation through decomposition as mentioned in Iwana and Uchida 2021 but previously excluded, by exploiting a decomposition based on exponential properties. Experiments are conducted in this direction.

Firstly, exponential curves in the form of  $y(t) = a * \exp(b * t) + c$  are fitted to signals, like illustrated in Figure 3.2, with the expectation to use  $a$ ,  $b$  and  $c$  as a lower dimension space of trajectories, in which new trajectories can be sampled (using gaussian mixtures for instance). Unfortunately, least square optimization on non-linear curve fitting do not achieve a consistent optimization of the three parameters, as the fitting does not converge on all trajectories. Therefore obtained projection could not be exploited confidently enough for data generation. This approach is therefore rejected, however further work could still be attempted to improve the method and confidence in its results.

The exponential nature of the signal can otherwise be exploited with exponential smoothing methods, like ARIMA (Autoregressive integrated moving average, Box and Jenkins 1970). It is successfully attempted to fit ARIMA to the C-MAPSS signals time series, like shown for one example in Figure 3.3. Then, weights of the fitted ARIMA can be used for dimension reduction from which new weights can be sampled (gaussian mixture) and used to generate new trajectories.

Unfortunately, the reconstruction of trajectories from ARIMA weights is not achieved in this study and this approach is still to be tested for confidence and pertinence.

Expert knowledge is still a credible way to improve prognostics performances on further use-cases where it is more largely available. In ArcelorMittal descaling use-case, functional analysis of the system are available. They describe the link between components of the descaling (valves, pumps, pipes, nozzles, etc.) under the form of flows and functions, such that the propagation of the effect of the failure of one component can be modeled. They can be used for various applications, such as proposed in Ben Hassen et al. 2015 to cluster sensors. Similarly, functional analysis combined with domain ontologies could be used to create graph-based representations of the datasets to be exploited for prediction models. Additionally, physics models of fluid mechanics describe the behavior of water within the system, which can be used to detect anomalies and deviations from nominal state in the data, based on expected values. Therefore, the mitigated results obtained by this section are to be considered in the light of more expert knowledge available

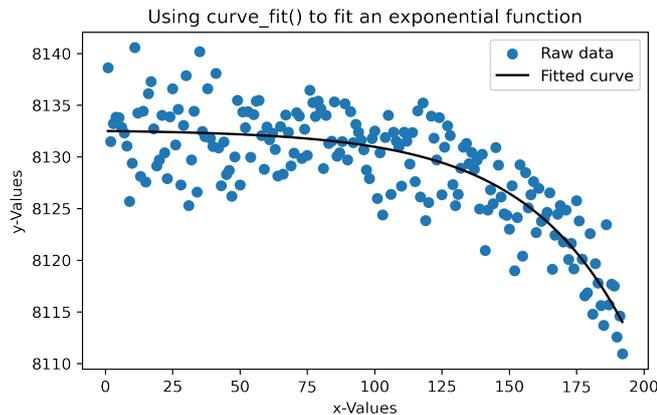


Figure 3.2: Fitted exponential function on one signal of the C-MAPSS FD001 dataset. The fitted equation coefficient for this trajectory is  $y(t) = -0.12 * \exp(.026 * t) + 8133$ . If this fitting would have been used for dimension reduction, the coordinates in the projected space would have been  $(a, b, c) = (-0.12, 0.26, 8133)$ .

### 3.2. Description and formalization of available knowledge on dataset

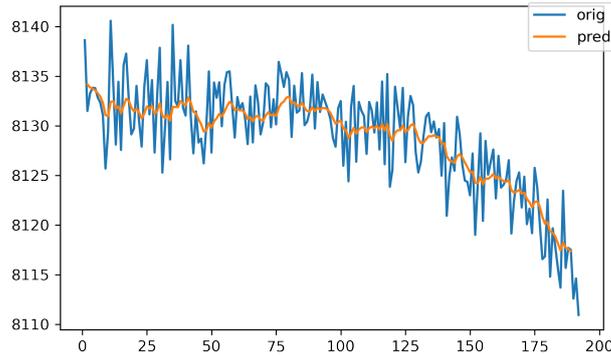


Figure 3.3: Fitted ARIMA estimator to one signal of the C-MAPSS FD001 dataset.

on other use-cases.

#### 3.2.3 Statistical knowledge on turbofan systems for the C-MAPSS dataset

The second source of knowledge is statistical knowledge on the dataset, which can be gathered more easily especially when no expert are available. Therefore, it is an important source of knowledge on the C-MAPSS dataset.

The first proposed approach is to observe all signals to detect global characteristics which can lead to data augmentation tuning. To do so, available training trajectories are observed and analyzed. Eleven trajectories are plotted signal per signal on the Figure 3.4. The following observation are therefore made, leading to knowledge about the dataset and therefore the concerned system:

- Some signals are taking a single constant value over all trajectories. They are *set\_3*, *sig\_1*, *sig\_5*, *sig\_10*, *sig\_16*, *sig\_18*, *sig\_19*. Therefore, modifying their magnitude with data augmentation transformation creates “false” information. Thus, those signals shall be excluded from magnitude-based data augmentation transformations (jittering, magnitude scaling). Time-based transformations are not concerned as they do not alter the magnitude of signals.
- The signal *sig\_6* only takes two possible values. It is addressed in the following as a “discrete signal”, because it takes discrete values (as opposed to continuous ones) Therefore, as for constant signals, it shall be avoided to change its magnitude.

The second proposed approach is to exploit the common evolution of signals through correlation analysis. First proposal consists in using correlations to group similar signals, *i.e.* signals evolving conjointly. Therefore, those signals should be augmented conjointly, in order to keep their conjoint evolutions. This ensures that if signals degrade in the same way when facing a degradation / failure, augmented trajectories maintain this relationship. Second proposal consists in computing signals correlations on the trajectories generated through data augmentation and to reject them if they are too different from the original correlations. Indeed, it would mean that the new trajectories do not respect the original distribution of data and might break physics links between signals.

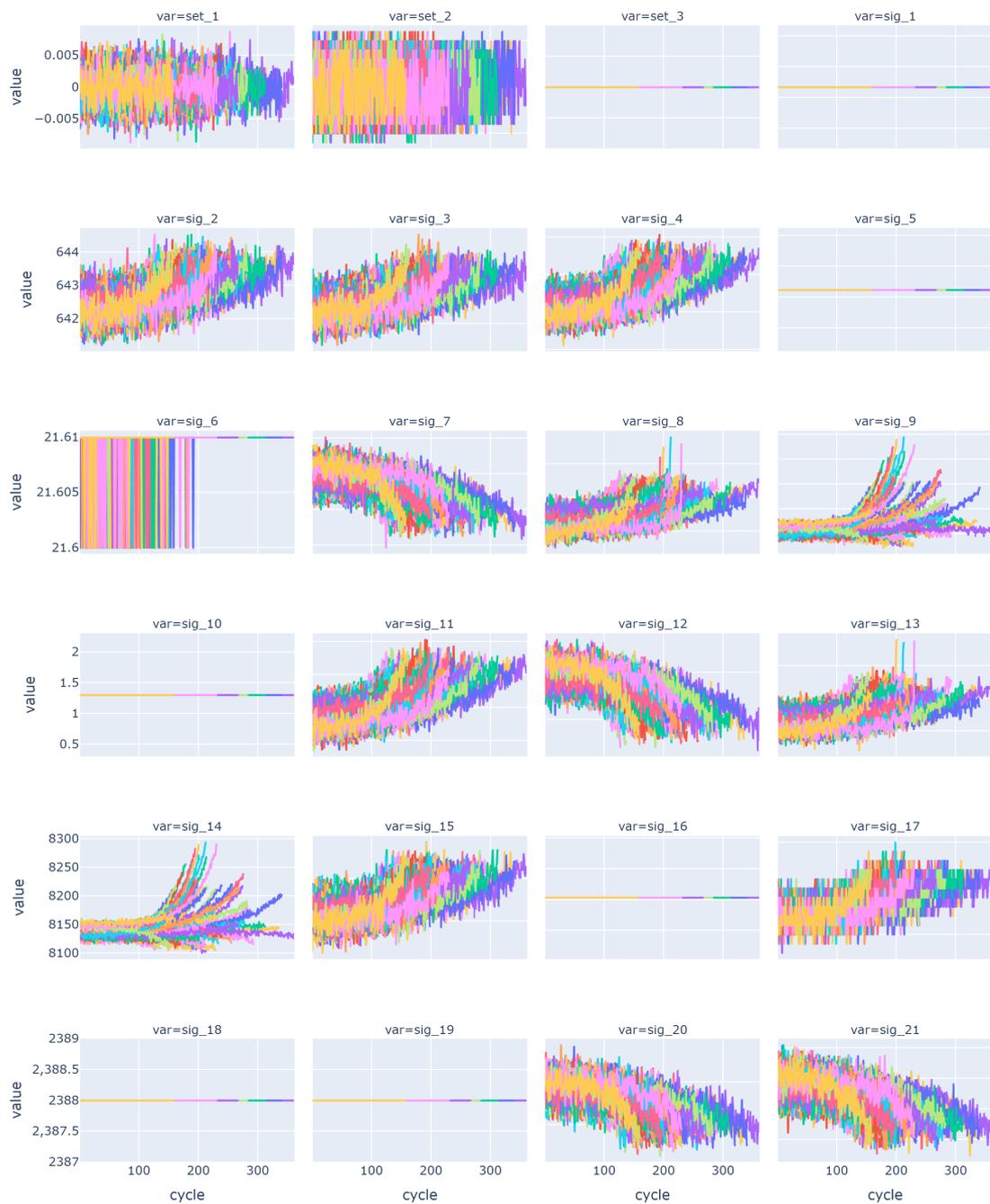


Figure 3.4: All the signals from the C-MAPSS dataset (FD001) plotted for eleven run-to-failure trajectories. Scales are not the same across signals, but time axis is similar.

### 3.2. Description and formalization of available knowledge on dataset

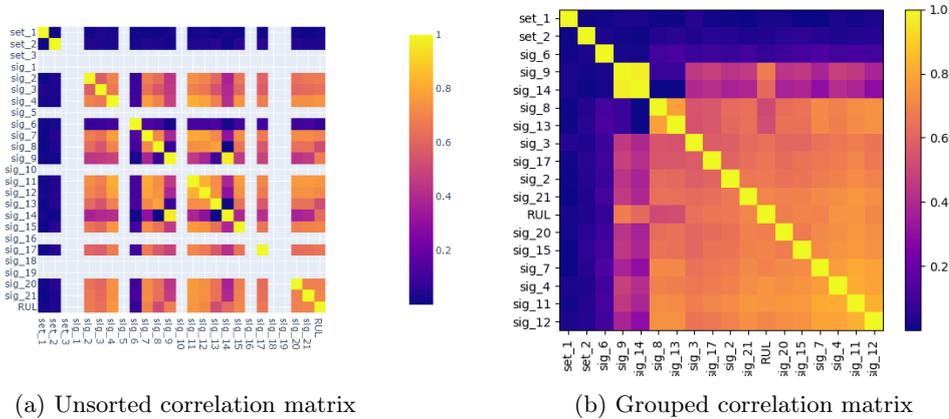


Figure 3.5: Correlation matrix for C-MAPSS FD001 dataset. Values correspond to the absolute values of the correlations between signals. On the Figure (b) they are grouped by similarity. The empty lines and rows in Figure (a) correspond to signals always constant. The correlation is computed on the integrality of the 100 train trajectories of the dataset.

The correlations are thus computed for all signals of the C-MAPSS dataset and are plotted in Figure 3.5a, as well as in Figure 3.5b on which signals are grouped by similarity according to their correlations. The grouping is done using a dendrogram, presented later. The correlation matrix shows that many signals are correlated, which could be expected from signals originating from a physical system.

Those correlations are used to group signals, using a dendrogram-based method shown in Figure 3.6.

Like the Figure shows, signals can easily be split into three groups, which could already be guessed on the sorted correlation matrix. The groups are the following:

- *sig\_9, sig\_14*: As shown on Figure 3.7 those two signals evolve systematically conjointly. Among other signals, they are slightly more correlated to the RUL, which the Figure is hinting. Therefore, they could be particularly important to track.
- *sig\_8, sig\_13*: in a lesser way than the previous group, those two signals are more correlated among each other than the rest and could be considered separately. They are shown on Figure 3.8. Nevertheless, they could also be joined in the general group of the rest of signals.
- *set\_1, set\_2, sig\_6*: those three signals are correlated to neither of the other signals available. This can be explained as they are mostly noisy, as can be seen on Figure 3.9.
- *The rest of signals*: The rest of the signals are moderately correlated among each other, and are clearly identified as a group on the dendrogram. Those signals include the RUL, therefore they might be interesting for its prediction. They shall nevertheless be considered together when applying data augmentation. Due to their number, they were not plotted. Those signals are *sig\_3, sig\_17, sig\_2, sig\_20, sig\_21, sig\_15, sig\_7, sig\_4, sig\_11, sig\_12*

The rest of signals are not represented in the dendrogram as their correlations were null with any other signals. Those signals (and sets) are constant over all trajectories, therefore they should be excluded from magnitude-based data augmentation transformations.

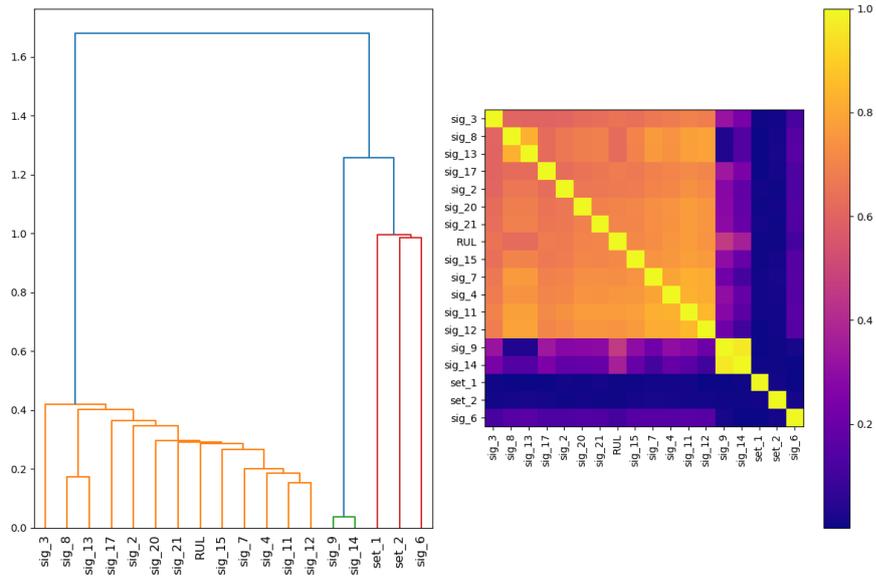


Figure 3.6: Hierarchical clustering applied to the C-MAPSS FD001 signals. Signals can be grouped by cutting the branches of the dendrogram at a selected height.

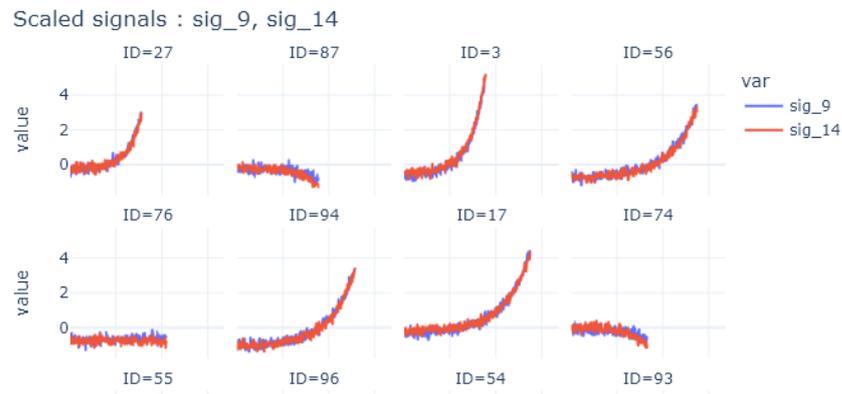


Figure 3.7: Plot of twelve trajectories of the two signals of the first group, sig\_9 and sig\_14 from C-MAPSS dataset. These two signals are highly correlated together and less with the rest. The signals on the plots have been scaled (using Standard Scaler) to set them to the same scale.

To conclude, out of statistical analysis based on correlations, three approaches are identified. Firstly, constant signals are excluded from magnitude-based data augmentations, as well as signals with discrete values. Secondly, signals are grouped in three groups based on their mutual correlations, which shall be used to augment similarly (/proportionally) signals when applying magnitude-based data augmentation transformations. Thirdly, it shall be ensured that augmented trajectory correlation matrices are not too different from the original dataset correlation matrix. This can be implemented by dropping the generated trajectories not respecting this condition.

### 3.2. Description and formalization of available knowledge on dataset

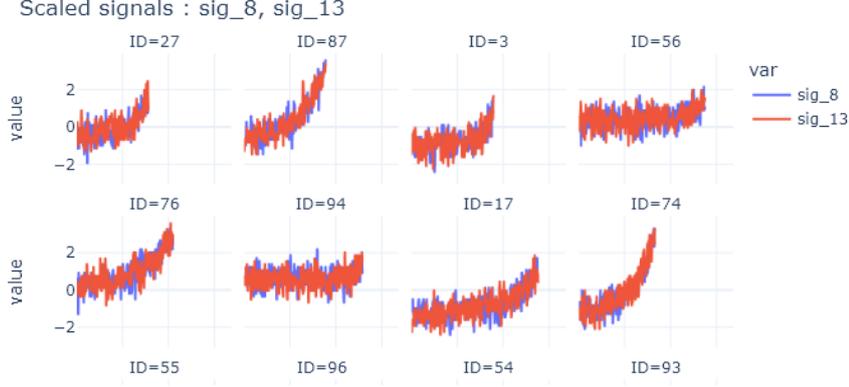


Figure 3.8: Plot of twelve trajectories of the two signals of the first group, sig\_8 and sig\_13 from C-MAPSS dataset. These two signals are highly correlated together and less with the rest. The signals on the plots have been scaled (using Standard Scaler) to set them to the same scale.

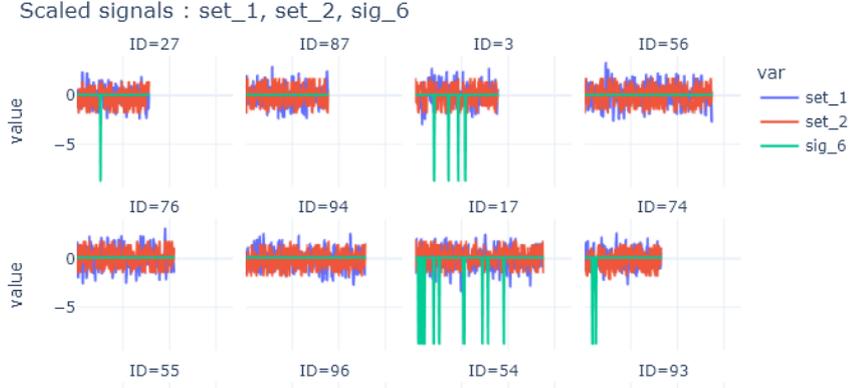


Figure 3.9: Plot of twelve trajectories of the three signals of the last group from C-MAPSS dataset. These three signals have no correlations with either other signals, due to their very noisy shape. The signals on the plots have been scaled (using Standard Scaler) to set them to the same scale.

#### 3.2.4 Knowledge formalization for data augmentation on the C-MAPSS dataset

The proposed approaches are applied to magnitude-based data augmentation transformations. Both equations and algorithms are modified to fit the new constraints.

Firstly, the augmentation algorithm now excludes constant and discrete signals. Secondly, the per-signal magnitude scaling equation is edited such as scaling factors  $\varepsilon_k$  are shared among signals from same groups. The new equation is:

$$(\underline{x}^{(j)})_i^{per-signal\ scaling} = (\underline{x}^{(j)} * (1 + \varepsilon_k))_i \quad \text{with} \begin{cases} \text{signal } j \text{ in group } k \\ \varepsilon_k \sim \mathcal{U}(-p, p), \forall k \end{cases}$$

Thirdly, the algorithm compares the correlation matrix of the augmented dataset ( $R_a$ ) with

the one of the original dataset ( $R_o$ ). The augmentation is re-done if their difference is superior to a threshold. The difference between correlation matrices is computed through the Correlation Matrix Distance (CMD) such as introduced in Herdin and Bonek 2005; Herdin, Czink, et al. 2005. The CMD equation is written in Equation (3.1). This metric “becomes zero if the correlation matrices are equal up to a scaling factor and one if they differ to a maximum extent” (Herdin, Czink, et al. 2005). Finally, the threshold value is identified based on the values of CMD between sub-samplings of the dataset. These CMD correspond to acceptable values as they are directly obtained from real subsets of the original dataset. This methodology adapts the threshold to the considered dataset specificity.

$$d_{corr}(R_o, R_a) = 1 - \frac{tr\{R_o R_a\}}{\|R_o\|_f \|R_a\|_f} \in [0, 1], \quad \text{with } \|\cdot\|_f \text{ the Frobenius norm} \quad (3.1)$$

### 3.3 Implementation and evaluation of knowledge-based data augmentation for prognostics

The data augmentation transformations upgrades mentioned in the previous sections are implemented in the Python code as new augmentation transformations. They are then evaluated using the previous test methodologies, defined in section 2.2 and subsection 2.7.1. All following results are obtained on the C-MAPSS dataset with ROCKET prognostics model for the sake of rapidity of execution.

The methodology for hyper-parameters optimization through cross-validation is applied on all augmentation methods. In a first time, only the signal exclusion and the signal grouping are applied, results are presented in subsection 3.3.1. Secondly, the validation of augmentation through correlation matrix distance is applied in addition to the two previous methods. Results are presented in subsection 3.3.2.

#### 3.3.1 Knowledge-based data augmentation through signal exclusion and grouping

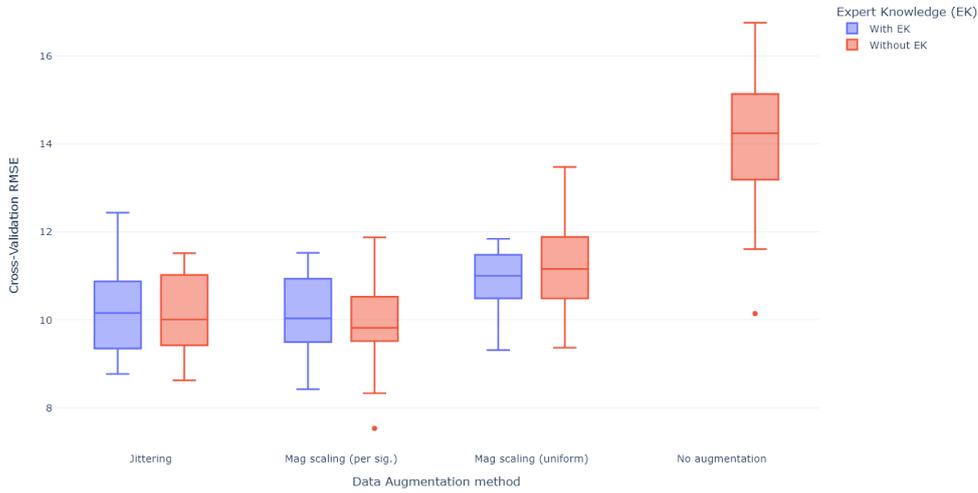
Table 3.1: List of T-test p-values obtained between knowledge-based and standard version of augmentation methods, for various methods and evaluation sets. Those results were obtained on 25% sub-samplings of the C-MAPSS dataset with the MiniROCKET model.

Augmentation	Evaluation set	p-value for $H_0: \widehat{rmse}_{noEK} = \widehat{rmse}_{EK}$
Jittering	Test set	0.83
	Cross-Validation	0.74
Mag scaling (per sig.)	Test set	0.92
	Cross-Validation	0.30
Mag scaling (uniform)	Test set	0.59
	Cross-Validation	0.36

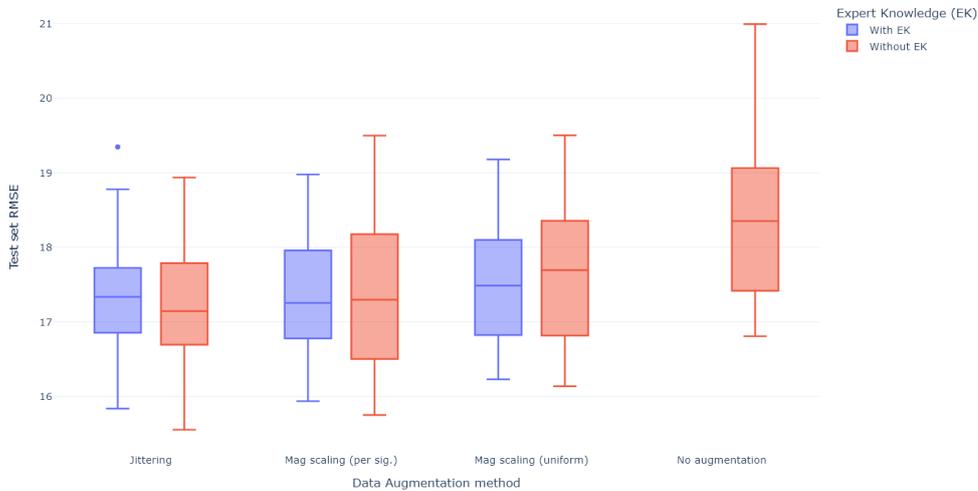
Performances comparison of methods with and without signal exclusion and grouping are shown on the Figure 3.10a for their Cross-validation performances and on Figure 3.10b for their performances on the Test set. Cross-validation and test results have similar results and conclusion. Expert knowledge integration (based on statistical knowledge) does not impact performances of

### 3.3. Implementation and evaluation of knowledge-based data augmentation for prognostics

the data augmentation methods on both DA with or without expert knowledge (EK). Indeed, performances of the knowledge-modified methods are similar to standard augmentation methods. This can be seen on the figures and is confirmed by t-test evaluations, which p-values are all largely superior to 0.05 as summarized in Table 3.1.



(a) Expert knowledge DA performances through Cross-Validation.



(b) Expert knowledge DA performances on the test set.

Figure 3.10: Distribution of performances of data augmentation methods (a) through cross-validation and (b) on the test set of the C-MAPSS dataset with hyper-parameters optimized. Those results are obtained with the MiniROCKET model. Methods presented here are the random magnitude transformations (in red) and their knowledge-based counterparts (in blue).

### 3.3.2 Knowledge-based data augmentation through correlation matrices distance

The third upgrade proposed by statistical knowledge exploitation is to reject augmented trajectories when they are too different from the original dataset. It is proposed to exploit correlation matrices to measure the similarity to the original dataset. To do so, the correlation matrix distance has been introduced in Equation (3.1). It is used to measure the distance between augmented dataset and original dataset, if the distance exceeds a defined threshold, the augmented dataset is rejected and a new augmented dataset is generated until its CMD is below the threshold. To avoid infinite loop, a maximum number of iteration is allowed. The proposed algorithm is defined in Algorithm 1.

---

**Algorithm 1** Data Augmentation with Correlation Matrix filtering
 

---

**Require:**  $thres \leftarrow$  distance threshold

**Require:**  $n_{loops} \leftarrow$  maximum number of iteration before abort

$X \leftarrow$  Original dataset ▷ *i.e.* the current 25% sub-sampling of C-MAPSS  
▷ Counter so the loop is not infinite

**for**  $i \leftarrow 0; i \leq n_{loops}; i \leftarrow i + 1$  **do**

$X_{da} \leftarrow$  Augmentation( $X$ )

$X_{new} \leftarrow X + X_{da}$

**if**  $cmd(corr(X), corr(X_{new})) \leq thres$  **then**

return  $X_{new}$  ▷ If distance condition is met, augmented dataset is returned

**end if**

**end for**

return  $X$  ▷ After n unsuccessful loops: aborts and returns un-augmented  $X$

---

To be applied, the algorithm requires the distance threshold to be defined, as well as maximum number of iterations. To define the threshold, Figure 3.11 shows the distribution of CMDs between random sub-samplings of the C-MAPSS dataset considered in the previous section for the sake of testing. This distribution illustrates the values that can be found when comparing real data from the dataset, so they can be considered as acceptable values of distance for additional data. It can therefore be used to define the threshold for the algorithm. The maximum value of the CMD being 0.02, it is decided to set the threshold as  $thres = 0.02$ . This threshold is high enough to allow: (1) not to reject all data augmentations and (2) to ensure nevertheless a diversity of augmented trajectories required to increase the small dataset, which lack this diversity. This value should ensure that generated data are more relevant thanks to the original dataset and thus less likely to generate aberrant data.

To illustrate the impact of this method, Figure 3.12 shows example of rejected and accepted trajectories with knowledge-based<sup>6</sup> jittering.

The CMD-based jittering is applied to the C-MAPSS dataset along with signal exclusion, obtained results are plotted in Figure 3.13. The main impact of CMD-based algorithm 1 is to reject the augmented dataset when expected to degrade performances. It does not improve performances when hyper-parameters are correctly optimized, but avoid performance degradation when it is not possible to determine precisely optimized values. However, it can improve performances in the cases where they were degraded due to un-adapted settings of hyper-parameters. Indeed, CMD validation rejects all trajectories for  $p > 1.05$ , as shown on Figure 3.14. To verify this hypothesis, jittering with and without CMD-based validation are evaluated for various values

---

<sup>6</sup>*i.e.* with signal selection

### 3.3. Implementation and evaluation of knowledge-based data augmentation for prognostics

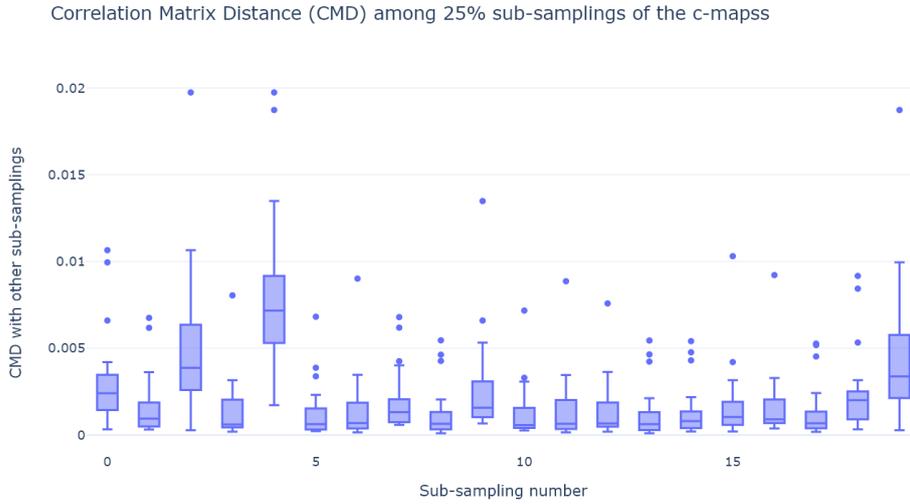


Figure 3.11: Distributions of Correlation Matrix Distances (CMD) among 25% sub-samplings of the C-MAPSS dataset. Each box represents the distribution of CMDs computed among the considered sub-samplings and the 20 others. It is seen that distances are low, as could be expected. From these data it is seen that 90% of distances are below 0.0065 and 95% below 0.0085, with maximum value at  $d_{max} = 0.020$ .

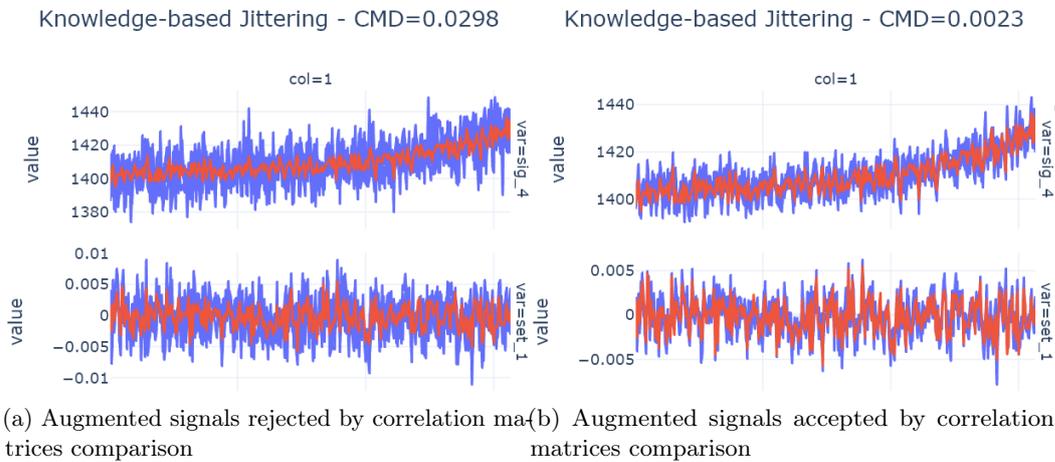


Figure 3.12: Graphs of sample augmented trajectories through knowledge-based jittering from C-MAPSS dataset. Red signals are original (un-augmented) signals. Blue signals are augmented signals. Augmented signals on the figure (a) are rejected by the correlation matrices comparison. Their Correlation Matrix Distance (CMD) is  $d \approx 0.030 > 0.02$ . Augmented signals on the figure (b) were accepted with the same method. Their Correlation Matrix Distance (CMD) is  $d \approx 0.003 < 0.02$ .

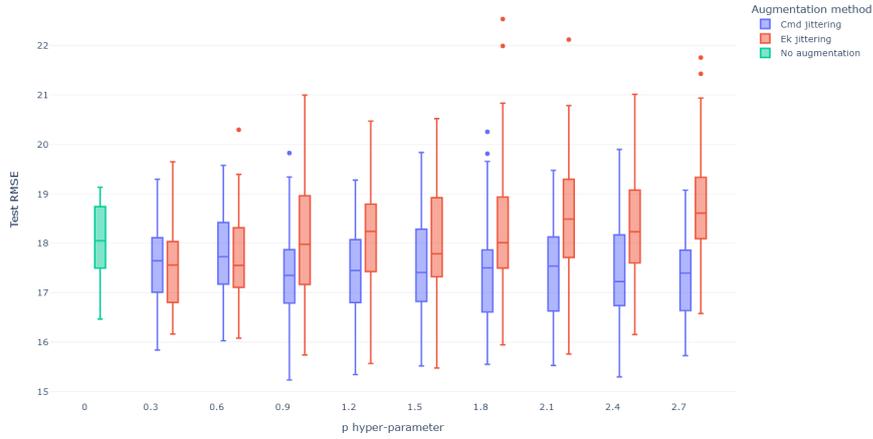


Figure 3.13: Performances of CMD-based jittering on C-MAPSS dataset with MiniROCKET for various values of its  $p$  hyper-parameter, compared to knowledge-based jittering. Knowledge-based jittering implements the knowledge-based signal selection by not augmenting constant and discrete signals. CMD-based jittering implements both signal selection and CMD-based Algorithm 1 of augmentation rejection if generated signals are too different of original ones.

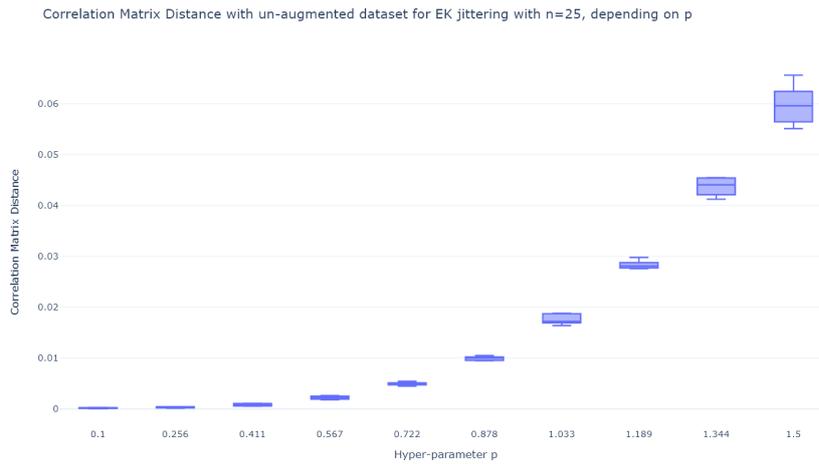


Figure 3.14: Correlation Matrix Distance between augmented and original datasets with EK jittering for various values of  $p$ . Values obtained on the C-MAPSS dataset with MiniROCKET and  $n = 25$ .

of  $p$  and plotted in Figure 3.13. Firstly, for  $p < 1.05$ , both methods (with and without CMD validation) improve prognostics performances compared to un-augmented dataset. Then, it is seen that without CMD validation, prognostics performances on the test set are degraded by data-augmentation for  $p > 1.05$ . Whereas with CMD validation, prognostics performances are capped to the performances with un-augmented dataset as soon as  $p > 1.05$ . Therefore, the implementation of CMD validation avoids prognostics performances loss for higher values.

In conclusion, the CMD-based approach does not offer performance improvements but reduces the risks of degrading performances with data augmentation transformations. This property is useful, especially in the light of the possible performances degradation identified in Chapter 2.

## 3.4 Conclusion

This chapter focuses on the integration of Expert Knowledge in data augmentation transformation to improve their performances on small datasets for industrial systems. This approach is supported by the literature. It has a high potential because performances on small dataset are dependent on the new information brought by additional data in the learning process. Therefore, optimizing data augmentation transformations with expert knowledge can increase added information. In this chapter, knowledge is exploited from two sources: expert knowledge and statistical knowledge.

Knowledge is gathered on the previously used C-MAPSS dataset from expert sources and statistical analysis to propose upgrades of data augmentation transformations. Few expert knowledge is found in the literature concerning this simulation-based dataset. It is attempted to exploit the exponential form of the degradation to formulate an appropriate dimension reduction method. It can then be used as a dimension reduction method and combined with decomposition-based augmentation transformations. Unfortunately, the decomposition could not be successfully fitted and this approach was not applied further.

Statistical knowledge is therefore exploited to propose three upgrades of magnitude-based data augmentation transformations (jittering and magnitude scaling). First, constant and discrete signals are excluded from data augmentation transformations based on the results of data analysis. Second, signals are grouped based on signals correlations to apply magnitude scaling factors by groups rather than by signals to keep consistency. Third, augmented trajectories are validated by exploiting the distance between correlation matrix before and after augmentation to reject aberrant augmentations based on a threshold.

The three proposed upgrades achieve mitigated results, with no notable improvements nor degradation of performances, though they improved the resilience of transformations. The approaches are tested through the test methodology proposed in the previous Chapter 2. Signal exclusion does not modify performances in the tested conditions (MiniROCKET applied to 25% of C-MAPSS). Signal grouping did not impact performances neither. However, distance-based validation of data augmentation improves the resilience of data augmentation to performance reduction (notably due to hyper-parameters values).

Hopefully, by improving prognostics models resilience to training, knowledge exploitation make them more eager to support maintenance decision-making. Indeed, as discussed in the global introduction, decision-making requires accurate prognostics predictions as well as reduced uncertainty on the prognostics. Reducing the risks of inefficient model fitting and overfitting of the training sets ensure to decision-making that data documentation cannot degrade the performances of prognostics. Therefore, decision-making can be more reliable and trusted.

This unsuccessful attempt at exploiting expert knowledge lead to the formulation of perspectives for further system knowledge exploitation for data augmentation. Functional analysis of systems could be used to define link between signals and so guide data augmentation into generating new signals more coherent with the real behaviour of the system. Similarly to signal grouping done after in statistical knowledge, but endorsed by knowledge and system models. Knowledge could also be extracted from dysfunctional analysis, like HAZOP, to guide the generation of data augmentation methods. The recent paper Han et al. 2022 propose for instance to use causality graphs for expert knowledge integration, which could be used to represent the information contained in (dys)functional analysis. This would be particularly efficient to face

new degradation modes, not or less represented in training sets. This issue was not considered in this work, which focused on single degradation mode datasets.

Further improvements could also be implemented in the proposed knowledge-based methods. Correlation matrices have been used for signal grouping and divergence assessing of newly-generated trajectories. The variations of correlation matrices could be used to identify trajectories ought to vary the most to focus data augmentation on them. Such approach ought to apply data augmentation at a “microscopic” scale rather than “macroscopic” as done by agnostic transformations.

The next chapter will focus on the link between prognostics and decision-making, focusing on ArcelorMittal use case specificities.

## Chapter 4

# Impact of prognostics performances on maintenance decision-making

### 4.1 Introduction

Maintenance decision-making is at the core of predictive maintenance cost reduction for industrial assets. Decision-making is the last step of the PHM process (Atamuradov et al. 2017). It directly inputs prognostics results (Skima et al. 2019), which takes the form of Remaining Useful Lifetime (RUL) estimation in the current study, as presented in Chapter 2. Decision-making optimization is an ongoing research problem, as mentioned in Wesendrup and Hellingrath 2020.

Predictive maintenance decision-making efficiency is therefore dependant of prognostics. Previous works have studied the prognostics decision-making for various applications, like Skima et al. 2019; J. Zuo et al. 2021; A. Hoang, Do, and Benoît Iung 2015. Some considered the joint optimization of prognostics and decision-making to get prognostics results pertinent for decision-making, like proposed in K. Nguyen and Do 2019.

ArcelorMittal plants' decision-making is constrained by steel-making specificities, which are not addressed thoroughly in previous works. As presented in Chapter 1, steel-making is a linear continuous process. The linearity of the process make any failure critical on the whole steel-making line. Being continuous, few opportunities are available for maintenance operations. Therefore, maintenance decision-making must ensure the availability of assets until the next maintenance slots. Maintenance decision-making efficiency is measured through the average maintenance cost, which maintenance optimization ought to reduce.

This chapter proposes an evolution of maintenance decision-making cost adapted to the considered context. Maintenance cost is expressed as a function of the performances of prognostics, based on a probabilistic formulation of the prognostics..

The proposed cost formulation is used to measure cost reduction with the implementation of prognostics and with the improvement of prognostics performances. The implementation of prognostics provides a cost improvement of about 20% compared with conventional scheduled maintenance. Further, prognostics performance improvement can improve further decision cost of few additional percents.

To this extent, the chapter is organized as follows. First, section 4.2 formalizes ArcelorMittal industrial constraints on decision-making. Then, section 4.3 proposes an evolution of the literature cost formulation to take into account the formalized constraints. Finally, section 4.4 proposes a numerical application of the proposed cost formulation to (i) validate by comparison to a specific

Python simulation and (ii) measure the cost evolution with prognostics implementation and prognostics improvement.

This section contains numerous specific variables. For the ease of the reader, their definition are recalled on the next pages in a chapter-specific glossary.

## 4.2 ArcelorMittal decision-making constraints

Based on ArcelorMittal steel-making context, the manuscript introduction defined the following decision-making constraints:

**Constraint 5:** planned maintenance operations can only be done during regular planned maintenance slots. Works in the literature propose advanced decision-making taking advantage of opportunistic maintenance slots to do operations, like proposed in Derigent et al. 2009. For instance, they can exploit available downtimes when corrective maintenance is performed on another failed system (Benoit Iung, Do, et al. 2016). However, this solution is not compatible with the current functioning of ArcelorMittal maintenance organisation. Indeed, production continuity must be ensured and the systems never stop outside of these time slots, and scheduling maintenance is not possible outside these slots. On the Hot Strip Mill line, two days of stop every three weeks are dedicated for these maintenance slots.

**Constraint 4:** un-planned maintenance stops (between maintenance slots) have to be avoided. It implies **Constraint 4.1** that at a given maintenance stop, preventive operations must be taken to ensure the functioning of the system until the next stop. Indeed, as explained before, no maintenance opportunities are available between planned maintenance slots. Other works such as Lian, Zheng, and X. Gao 2022 propose to reschedule steel-making operations to face potential failures and other process constraints as oxygen supply changes. It represents another possibility to face failure occurrence before next maintenance slot. Production planing modification is however out of the scope in ArcelorMittal use-case, as it depends on too many external constraints to be modified. Therefore, prognostics is used to assert the remaining useful lifetime of the considered asset is larger than the time before next maintenance slot, i.e;  $T_{step}$ .

**Constraint 2:** if a failure happens on one system of the Hot Strip Mill process, the whole line has to be stopped for the time of the corrective maintenance. This constraint is linked to the linear nature of steel-making process. Running a Hot Strip Mill is both expensive and energy consuming, therefore it implies **Constraint 2.1** that failures have a high corrective maintenance operation cost due to the required production stop (compared to preventive maintenance operation).

In this chapter, maintenance does not consider multi-components system, but focuses on a single component. In the case of ArcelorMittal descaling, it would consist in focusing on one valve of the system. Advanced decision-making strategies can account for multiple components simultaneously, in a system-wide approach. They allow for maintenance grouping, either for multi-component systems like proposed in Do 2015 or to optimize time-limited opportunities like in Do 2013. Such approaches are ignored in this study as maintenance operation duration is not a constraint. Furthermore, as this study consists in a first proposal in considering prognostics performance in maintenance cost, it also simplifies cost formulation. Multi-component remains a pertinent perspective for future works and extensions of the proposed contribution.

## 4.3 Prognostics-based maintenance decision-making cost

### 4.3.1 Introduction to maintenance cost modeling

Maintenance cost optimization is a matter of balancing the maintenance cost with the system downtime cost. Maintenance cost in the case of preventive maintenance considers both operation cost and the cost of the operation duration. For instance, Hou-bo, Jian-min, and Chang-an 2011 formulates the cost of scheduled operations as  $c = c_1 + c_2$  with “ $c_1$  the cost of repairing the component, including man-hour, spare parts and other related costs” and “ $c_2$  the cost of the interruption by the scheduled maintenance”. Additionally, Feldman, Jazouli, and P. A. Sandborn 2009 considers both maintenance operation cost and out-of-service cost of the considered component for different types of operations. Their maintenance cost is formulated as:

$$c = fC_{comp} + (1 - f)C_{comprepair} + fT_{replace}V + (1 - f)T_{repair}V$$

With:

- $C_{comp}$  the cost of procuring a new component, while
- $C_{comprepair}$  the cost of repairing it.
- $f$  is the fraction of operations requiring a replacement rather than reparation.
- $T_{replace}$  and  $T_{repair}$  are the duration of both operations and  $V$  the value of time out of service.

However, considering system out-of-service cost as an additional cost do not account that a corrective (un-scheduled) maintenance operation also postpone future scheduled maintenances. This can be accounted by modeling the maintenance cost as an average cost over time. Therefore, maintenance operation cost are divided by the asset runtime preceding the operation, like proposed in H. Wang and Pham 2006. Therefore, maintenance operation cost are divided by the asset runtime preceding the operation.

### 4.3.2 Notations

- $n \in \mathbb{N}^*$  is the number of maintenance slots corresponding to preventive maintenance periods slots.
- $T_p$  is the preventive maintenance period, *i.e.* the period at which preventive maintenance operation is done,  $T_p$  is a multiple of  $T_{step}$ , *i.e.*  $T_p = n \cdot T_{step}$ .
- $C(T_p)$  is the average maintenance cost over time.
- $r(t)$  is the probability density function (PDF) of the Weibull distribution, *i.e.* the instantaneous probability of failure of the component.
- $R(T) \in [0, 1]$  is the reliability of the system over time period  $T$ . It corresponds to the probability of the system to still be running at the end of the period  $T$ .
- $C_p$  is the preventive maintenance operation cost.
- $C_c$  is the corrective maintenance operation cost.

- $T_c$  is the average failure time of the system under preventive maintenance strategy. It is expressed as the expected value of the failure probability distribution:

$$T_c = \int_0^{T_p} t \cdot r(t) dt$$

- $Sp(T_p)$  is the prognostics' specificity. It corresponds to the true negative rate, *i.e.* the probability to correctly predict that no failure will occur before  $T_p$ .
- $Rc(T_p)$  is the prognostics recall, also called sensitivity. It corresponds to the true positive rate, *i.e.* the probability to correctly predict an upcoming failure will occur before  $T_p$ .
- $T_{prog}$  is the average maintenance time for maintenance initiated by prognostics, *i.e.* it corresponds to the cases when the preventive maintenance operation is anticipated based on prognostics prediction.
- $T_i = i \cdot T_{step}$  is the time at which maintenance time slot number  $i$  occurs.

- $\Phi(x)$  is the cumulative distribution function (CDF) of the standard normal distribution, defined as:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-w^2/2} dw$$

- $\overline{rul}$  is the "ground truth" remaining useful lifetime. It is considered as the time of failure in the equations, *i.e.*  $\overline{rul} = t$  means the component will fail at time  $t$  (and not that it will fail in  $t$  units of time). Even though this formulation is not classical, it is useful as it simplifies further formulation of equations.
- $\widehat{rul}$  is the prognostics predicted remaining useful lifetime (RUL). It is considered as the predicted time of failure in the equations, *i.e.*  $\widehat{rul} = t$  means the component is predicted to fail at time  $t$ .

### 4.3.3 Preventive maintenance cost modeling

Scheduled preventive maintenance cost is modeled based on the insights mentioned in the previous subsection. Similarly to Feldman, Jazouli, and P. A. Sandborn 2009, the operation cost is formulated as the sum of scheduled and un-scheduled operations, pondered by the probability of their occurrence. This probability of occurrence is modeled based on reliability, *i.e.* the probability of the asset to function until a given time. To account for the reduced operation time due to corrective maintenance, the cost is formulated as the average cost, by dividing cost by the average maintenance period, *i.e.* the duration of asset's runtime since last maintenance operation, like proposed in H. Wang and Pham 2006. In the case of reliability-based modeling, it is formulated as:

$$C(T_p) = \frac{R(T_p)C_p + (1 - R(T_p))C_c}{T_p} \quad (4.1)$$

where

- $T_p$  is the preventive maintenance period, *i.e.* the period at which preventive maintenance operation is done. **Constraint 5** imposes  $T_p$  to be a multiple of the maintenance stop period:  $T_p = n \cdot T_{step}$ .

### 4.3. Prognostics-based maintenance decision-making cost

- $C(T_p)$  is the average maintenance cost over time.
- $R(T) \in [0, 1]$  is the reliability of the system over time period  $T$ . It corresponds to the probability of the system to still be running at the end of the period  $T$ .
- $C_p$  is the cost of preventive maintenance operation.
- $C_c$  is the cost of corrective maintenance operation. In this specific equation, it includes the cost of unavailability of the asset until the preventive maintenance operation. It is considered that  $C_c > C_p$ , and **constraint 2.1** imposes that  $C_c \gg C_p$ .

This equation is compatible with the considered ArcelorMittal use-case preventive maintenance. Nevertheless, given **constraint 2**, the equation is edited to take into account that a failure requires an immediate maintenance operation and resets the system health. Indeed, it is required that the system has to be repaired immediately at failure time. Therefore, the average maintenance cost has to be divided by the average lifetime of the system rather than  $T_p$  to get the average maintenance cost over time. Thus, the denominator is modified to correspond to the average time at which the life cycle ends, either by failure or preventive maintenance. The modified equation is:

$$C(T_p) = \frac{R(T_p)C_p + (1 - R(T_p))C_c}{R(T_p)T_p + (1 - R(T_p))T_c} \quad (4.2)$$

where

- $T_c$  is the average failure time of the system under preventive maintenance strategy. It is expressed as the expected value of the failure probability distribution:

$$T_c = \int_0^{T_p} t \cdot r(t) dt$$

Given  $C_c$  and  $C_p$ , the optimal value of  $T$  corresponds to the equilibrium such that the un-avoided failures cost less than the additional preventive maintenance which is required to avoid them. It is done by finding the optimal preventive maintenance period  $T_p$  minimizing the global cost  $C$ :

$$T_p = \arg \min_{T=n \cdot T_{step}} \left\{ C(T) = \frac{R(T)C_p + (1 - R(T))C_c}{R(T)T + (1 - R(T))T_c} \right\} \quad (4.3)$$

This equation can be represented as a probabilist tree, synthesized in Figure 4.1. The logic behind the tree is the following, given preventive maintenance is done every  $T$ . The reliability of the system between two maintenance operations is  $R(T_p)$ . In such case, the maintenance cost over the period  $T_p$  is  $C_p$ , the preventive operations is done at the end of the period. But with probability  $(1 - R(T_p))$ , the system fails before  $T_p$  and corrective maintenance has to be applied. In such case, the maintenance cost over the period  $T_c$  is  $C_c$ , the cost of the corrective operations done at failure time. This cost must include every implied cost, like stop time.

Equation (4.3), does not consider the effect of predicting the RUL. For that purpose, we will extend it thanks to the integration of the prognostic consideration in the decision process and its effect on the decision tree proposed in Figure 4.1.

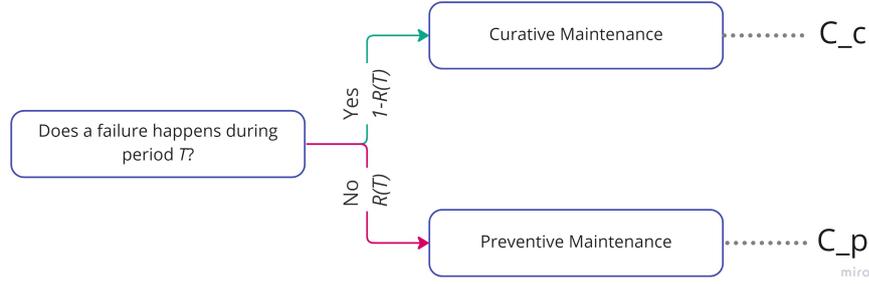


Figure 4.1: Schema of the decision tree implied by Equation (4.2).

#### 4.3.4 Maintenance cost model based on prognostics

Given the defined constraints in subsection 4.2, predictive maintenance takes the form of prognostics-based decision at maintenance slots time.

The proposition is to model decision-making process as a decision tree at maintenance slots time. At a given maintenance slots, prognostics model predicts if a failure could happen before the next maintenance slot. Based on ArcelorMittal descaling, prognostics and decision-making are considered for a single-component, because components of the descaling system have few impact on each other. If the prediction is positive, a preventive maintenance operation is conducted at the current slot. Otherwise, no operation is done and the decision is postponed to the next maintenance slot.

This model of prognostics-based decision-making is then included in the preventive decision tree proposed earlier in Figure 4.1. It ought to be used concurrently with the preventive operation planned at  $T_p$ . Therefore, additionally to the preventive maintenance operation planned at  $T_p$ , at each maintenance slot available before  $T_p$ , the prognostics model predicts if a failure could happen before the next maintenance slot. If a failure is predicted, preventive maintenance is done, otherwise nothing is done. Prognostics is included in the preventive decision tree as represented in Figure 4.2 and described as follows:

- If a failure will happen (branch  $1 - R(T_p)$  of Figure 4.2)
  - If the failure is predicted (**True positive**) maintenance cost is  $C_p$  because the preventive operation is done in time.
  - Otherwise (the failure is not predicted before  $T_p$ , **False negative**), the failure happens and maintenance cost is  $C_c$ .
- If no failure will happen (branch  $R(T_p)$  of Figure 4.2)
  - If no failure is predicted (**True negative**) maintenance cost is  $C_p$  as the preventive operation is done at time  $T_p$ <sup>7</sup>.
  - However, if a failure is falsely predicted (**False positive**), preventive maintenance is done “too soon” and maintenance cost is also  $C_p$ , but the system lifetime is shorter. This case mostly impacts the equation denominator, as done before.

<sup>7</sup>It is considered that preventive maintenance is not postponed.

### 4.3. Prognostics-based maintenance decision-making cost

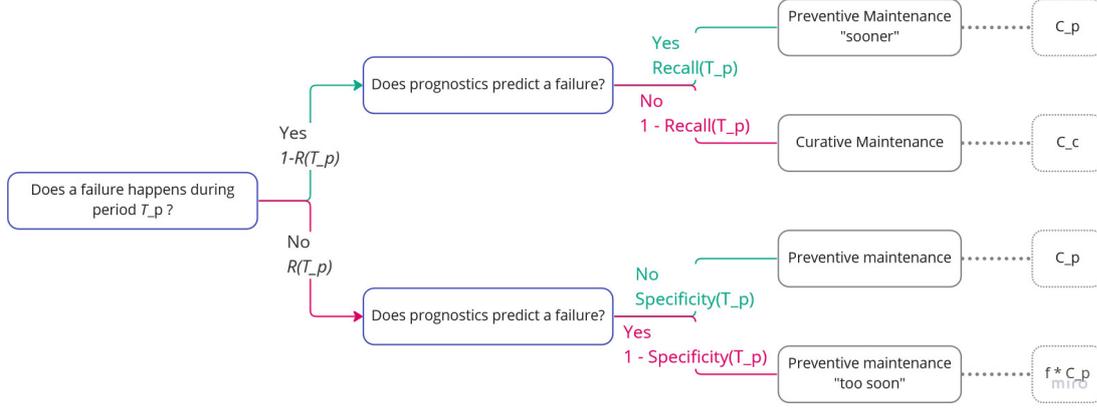


Figure 4.2: Maintenance decision tree integrating prognostics in decision-making for a single component.  $Recall(T_p)$  is the recall indicator of the pronostics model. It corresponds to the true positive rate.  $Specificity(T_p)$  is the specificity indicator of the pronostics model. It corresponds to the true negative rate.  $f$  is the factor corresponding to the lost product remaining useful life when preventive maintenance is done “too soon” due to False Positive from the prognostics model. It is included in the equation by the impact of the Specificity on the average  $T$ .

This decision tree is then mathematically formulated to evaluate the cost of predictive maintenance by adapting the previous Equation (4.2). This equation was formulated as:

$$\text{maintenance cost} = \frac{\text{average cost}}{\text{average maintenance period}}$$

Both numerator *average cost* and denominator *average maintenance period* evolve based on the proposed tree-based formulation of Figure 4.2. First, the numerator is expressed as:

$$\text{average cost} = R(T_p)C_p + (1 - R(T_p)) [Rc(T_p)C_p + (1 - Rc(T_p))C_c] \quad (4.4)$$

where

- $Sp(T_p)$  is the prognostics specificity. It corresponds to the true negative rate, *i.e.* the probability to correctly predict that no failure is arriving.
- $Rc(T_p)$  is the prognostics recall, also called sensitivity. It corresponds to the true positive rate, *i.e.* the probability to correctly predict an upcoming failure is arriving.

Second, the denominator is expressed as:

$$\begin{aligned} \text{average maintenance period} = & R(T_p) [Sp(T_p)T_p + (1 - Sp(T_p))T_{prog}] \\ & + (1 - R(T_p)) [Rc(T_p)T_{prog} + (1 - Rc(T_p))T_c] \end{aligned} \quad (4.5)$$

where

- $T_{prog}$  is the average maintenance time for maintenance initiated by prognostics, *i.e.* it corresponds to the cases when the preventive maintenance operation is anticipated based on prognostics prediction.

The final equation is written in Equation (4.6).

$$C = \frac{R(T_p)C_p + (1 - R(T_p)) [Rc(T_p)C_p + (1 - Rc(T_p))C_c]}{R(T_p) [Sp(T_p)T_p + (1 - Sp(T_p))T_{prog}] + (1 - R(T_p)) [Rc(T_p)T_{prog} + (1 - Rc(T_p))T_c]} \quad (4.6)$$

Under the hypothesis of un-biased prognostics and slow growth of  $r$ ,  $T_{prog}$  can be approximated by  $T_c - \frac{T_{step}}{2}$ . Where  $r$  is the probability density function (PDF) of the Weibull distribution, see Glossary in subsec 4.3.2. The un-biased prognostics  $\widehat{rul}$  averaged over multiple lifetimes of the component is equal to average failure time, considering a single-component prognostics and averaging over . Prognostics-decided maintenance operations happen during the maintenance slot preceding the predicted failure. Considering a slow growth of  $r$  between maintenance slots, *i.e.*  $r(t + T_{step}) - r(t) \ll r(t), \forall t \in [0, T_p]$ , a uniform distribution of failure time between maintenance slots can be considered. Therefore, the offset between the average failure time  $T_c$  and the preceding maintenance slot is  $\frac{T_{step}}{2}$ .

### Recall and Specificity formulation

The evaluation of the maintenance cost evolution with the integration of prognostics requires to evaluate the Specificity and Recall of the prediction algorithm. In the first part of this work, we focused on prognostics as a regression model predicting the RUL. Therefore, it is required to define the link between recall and specificity and RUL prediction error, the RMSE.

**In order to be able to derive the cost equations, the hypothesis is taken that the RUL prediction model has no bias and that its error follows a normal distribution.** The unbiased hypothesis is strong, especially for prognostics models, which may be trained specifically to underestimate the RUL. For instance, Abhinav Saxena et al. 2008 defined an asymmetric training score for prognostics evaluation. Nevertheless, the impact on this chapter's results is small. Indeed, prognostics performances in the rest of the section mostly impact maintenance cost by its average error. Therefore, the impact of a biased prognostics can be taken into account by increasing the average error considered. Meanwhile, this hypothesis largely simplifies the formulation of the problem. Therefore, this hypothesis is taken for the rest of the work.

For the equations, this hypothesis means that, given a "true"  $\overline{rul}$ , the variability of the prediction  $\widehat{rul}$  follows a normal distribution with mean equals to  $\overline{rul}$  and standard deviation equals the prognostics RMSE, *i.e.*  $\widehat{rul} \sim \mathcal{N}(\overline{rul}, rmse)$ . Indeed, considering the prediction with no bias, therefore standard deviation formulation is the same as the RMSE:

$$\sigma = \frac{1}{N} \sqrt{\sum_i^N (\mu - \widehat{rul})^2} = \frac{1}{N} \sqrt{\sum_i^N (\overline{rul} - \widehat{rul})^2} = rmse$$

**Recall** Recall can be expressed as the probability of the (prognostics-)predicted  $\widehat{rul}$  to be lower than  $T_p$  knowing that real  $\overline{rul}$  is lower than  $T_p$ . This corresponds to the probabilist formulation of recall definition "probability to correctly predict an upcoming failure is occurring". This can be written as follows:

$$\begin{aligned} \text{Recall} &= \mathbb{P}(\text{failure predicted} \mid \text{failure happened}) \\ &= \mathbb{P}(\text{failure predicted} \mid \overline{rul} < T_p) \\ &= \frac{\mathbb{P}(\text{failure predicted} \cap \overline{rul} < T_p)}{\mathbb{P}(\overline{rul} < T_p)} \end{aligned}$$

### 4.3. Prognostics-based maintenance decision-making cost

The numerator of the equation is then decomposed as an integral over all the preventive maintenance period, such that:

$$\mathbb{P}(\text{failure predicted} \cap \overline{rul} < T_p) = \int_0^{T_p} \mathbb{P}(\text{failure predicted} \cap \overline{rul} = t) dt$$

As stated before, prognostics-based decision-making is conducted at each maintenance slot. Hence, prognostics is evaluated at each one of them. This behavior is presented in Figure 4.3.

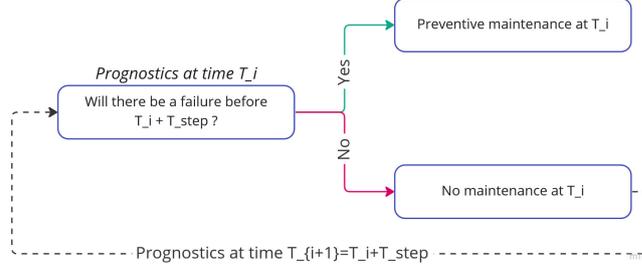


Figure 4.3: Schema of the basic decision tree linked to the application of prognostics to decision making.  $T_i$  is the time of the decision, during a maintenance slot.  $T_{step}$  is the maintenance slots period.

Following this model, the integral can then be decomposed as a sum of sub-integrals between consecutive maintenance slots. The probability of the failure to be detected can be expressed as the probability for the predicted  $\widehat{rul}$  to be inferior to the next maintenance slot. Given  $n = T_p/T_{step}$ ,  $n \in \mathbb{N}$ , the number of maintenance slots corresponding to preventive maintenance periods slots, *i.e.*  $T_p = n \cdot T_{step}$ . The previous integral is decomposed as:

$$\begin{aligned} & \int_0^{T_p} \mathbb{P}(\text{failure predicted} \cap \overline{rul} = t) dt \\ &= \sum_{i=0}^{n-1} \int_{T_i}^{T_{i+1}} \mathbb{P}(\text{failure predicted} \cap \overline{rul} = t) dt \quad \text{with} \begin{cases} T_i = i \cdot T_{step} \\ T_p = n \cdot T_{step} \end{cases} \\ &= \sum_{i=0}^{n-1} \int_{T_i}^{T_{i+1}} \mathbb{P}(\widehat{rul} < T_{i+1} \cap \overline{rul} = t) dt \\ &= \sum_{i=0}^{n-1} \int_{T_i}^{T_{i+1}} \mathbb{P}(\widehat{rul} < T_{i+1} \mid \overline{rul} = t) \mathbb{P}(\overline{rul} = t) dt \end{aligned}$$

The prognostics is defined as  $\widehat{rul} \sim \mathcal{N}(\overline{rul}, rmse)$ , therefore  $\mathbb{P}(\widehat{rul} < T_{i+1} \mid \overline{rul} = t)$  is equal to  $\Phi\left(\frac{T_{i+1}-t}{rmse}\right)$ , with  $\Phi$  the cumulative distribution function (CDF) of the normal law. The shape of  $\Phi$ , the CDF of the normal distribution, is shown in Figure 4.4 for a mean ( $t$ ) of 0 and a scale ( $rmse$ ) of 1. Finally, recall is expressed as:

$$\text{Recall} = \frac{\sum_{i=1}^n \int_{T_i}^{T_{i+1}} \Phi\left(\frac{T_{i+1}-t}{rmse}\right) \mathbb{P}(\overline{rul} = t) dt}{\mathbb{P}(\overline{rul} < T_p)} \quad (4.7)$$

where

- $\Phi(x)$  is the cumulative distribution function (CDF) of the standard normal distribution, defined as:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-w^2/2} dw$$

- $\mathbb{P}(\overline{rul} = t)$  and  $\mathbb{P}(\overline{rul} < T)$  depends on the probability model chosen for the system reliability
- $n = \frac{T_p}{T_{step}} \in \mathbb{N}^*$  is the number of maintenance slots corresponding to preventive maintenance periods slots, *i.e.*  $T_p = n \cdot T_{step}$
- $T_i = i \cdot T_{step}$  is the time at which maintenance time slot number  $i$  occurs.

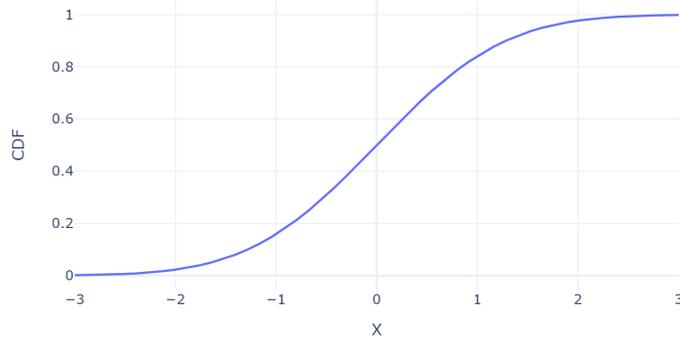


Figure 4.4: Cumulative distribution function (CDF) of the standard normal distribution, named  $\Phi$ , for a law  $\mathcal{N}(0, 1)$

**Specificity** Similarly, specificity can be expressed as the probability of the predicted  $\widehat{rul}$  to be greater than  $T$  knowing that real  $\overline{rul}$  is greater than  $T$ . This corresponds to the probabilist formulation of recall definition “probability to correctly predict that no failure is arriving”.

$$\begin{aligned} \text{Specificity} &= \mathbb{P}(\text{no failure predicted} \mid \text{no failure happened}) \\ &= \mathbb{P}(\text{no failure predicted} \mid \overline{rul} \geq T_p) \\ &= \frac{\mathbb{P}(\text{no failure predicted} \cap \overline{rul} \geq T_p)}{\mathbb{P}(\overline{rul} \geq T_p)} \end{aligned}$$

The numerator of the equation is then decomposed as an integral over all the time period where  $\overline{rul}$  is found, such that:

### 4.3. Prognostics-based maintenance decision-making cost

$$\begin{aligned}
& \mathbb{P}(\text{no failure predicted} \cap \overline{rul} \geq T_p) \\
&= \int_{T_p}^{+\infty} \mathbb{P}(\text{no failure predicted} \cap \overline{rul} = t) dt \\
&= \int_{T_p}^{+\infty} \mathbb{P}(\text{no failure predicted} \mid \overline{rul} = t) \mathbb{P}(\overline{rul} = t) dt
\end{aligned}$$

Therefore, the event of not prognosticating the failure is decomposed as the intersection of the failure not being prognosticated at any of the maintenance slots available during the preventive period  $T_p$ . This is similarly to what is done to recall and illustrated in Figure 4.3. The probability is therefore expressed as:

$$\begin{aligned}
& \mathbb{P}(\text{no failure predicted} \mid \overline{rul} = t) \\
&= \mathbb{P}\left(\text{nfp}_{T_0} \cap \text{nfp}_{T_1} \cap \dots \cap \text{nfp}_{T_{n-1}} \mid \overline{rul} = t\right) \\
&\text{with } \begin{cases} \text{nfp}_{T_i} = \text{no failure predicted at maintenance slot } T_i \\ T_p = n \cdot T_{step} \end{cases}
\end{aligned}$$

The hypothesis is made so that the prognostics at each maintenance slot is independent of the prognostics at other maintenance slots. This hypothesis is strong, since prognostics at each maintenance slots similarly decrease when  $\widehat{rul}$  increases. Nevertheless, as we will show later in this chapter, its impact on the result of the computation is limited. Given this hypothesis, the probability of the intersection (of the failure not being prognosticated at any of the maintenance) slots can be expressed as the product of individual probabilities:

$$\begin{aligned}
& \mathbb{P}(\text{no failure predicted} \mid \overline{rul} = t) \\
&= \prod_{i=0}^{n-1} \mathbb{P}(\text{nfp}_{T_i} \mid \overline{rul} = t) \quad \text{with } \begin{cases} T_i = i \cdot T_{step} \\ T_p = n \cdot T_{step} \end{cases} \\
&= \prod_{i=0}^{n-1} \mathbb{P}(\widehat{rul} \geq T_{i+1} \mid \overline{rul} = t)
\end{aligned}$$

As explained before  $\widehat{rul} \sim \mathcal{N}(\overline{rul}, rmse)$ , therefore  $\mathbb{P}(\widehat{rul} \geq T_{i+1} \mid \overline{rul} = t)$  is equal to  $\left[1 - \Phi\left(\frac{T_{i+1} - t}{rmse}\right)\right]$ , with  $\Phi$  the cumulative distribution function of the normal law. Finally, specificity is expressed as:

$$\boxed{\text{Specificity} = \frac{\prod_{i=0}^{n-1} \int_{T_p}^{+\infty} \left[1 - \Phi\left(\frac{T_{i+1} - t}{rmse}\right)\right] \mathbb{P}(\overline{rul} = t) dt}{\mathbb{P}(\overline{rul} \geq T_p)}} \quad (4.8)$$

## 4.4 Numerical application for ArcelorMittal descaling valve use-case

This section ought to apply the proposed maintenance cost formulation to the introduced ArcelorMittal descaling use-case. To this extent, numerical values are selected to correspond to the order of magnitude of the hot strip mill maintenance. Degradation models are selected based on literature to fit the behavior of descaling valves. Valves are critical descaling components exposed to continuous wear and regular failure. They are currently exposed to scheduled replacements but failure still happens occasionally. In the following, costs are expressed in relative values for confidentiality reasons.

### 4.4.1 Weibull distribution for component failure rate

The valves failure behaviour is modeled with a Weibull distribution. The probability density function of a Weibull random variable is (Rinne 2008; Weibull 1951):

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

It depends on two parameters  $\lambda$  and  $k$ .  $k$  defines the shape of the distribution, which sets the symmetry of the distribution, as represented on Figure 4.5a. While  $\lambda$  defines the scale of the distribution, influencing the mean and median values for a given  $k$ , as represented on Figure 4.5b.

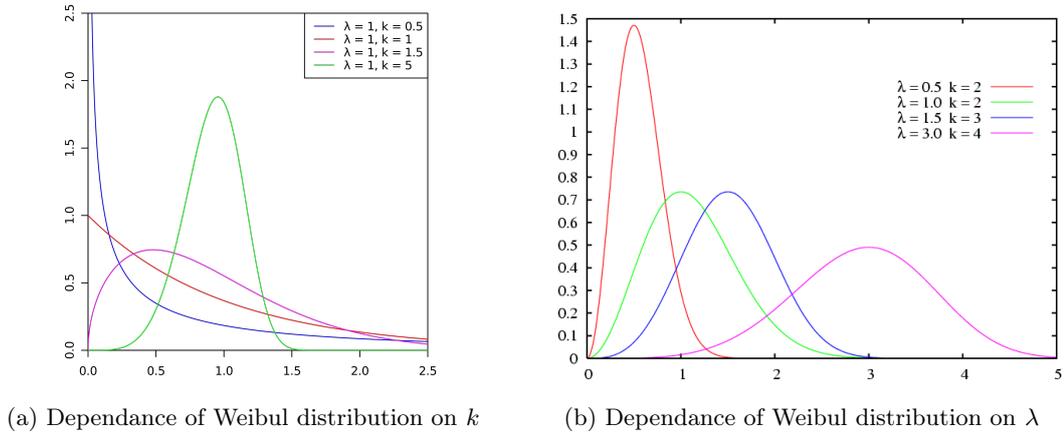


Figure 4.5: Weibull distribution probability density function (PDF) depending on the two parameters  $k$  (a) and  $\lambda$  (b). Source: wikimedia.org

The Weibull distribution is largely used in the literature to represent the failure probability of an asset. Its shape, controlled by  $k$ , can be adapted to match various degradation scenarios (Kızılersü, Kreer, and Thomas 2018; Papoulis and Pillai 2002). Small values of  $k < 1$  represent assets likely to fail at the early stage of its lifetime. Medium values of  $k \approx 1$  represent assets whose rate of failure is fairly constant. Larger values of  $k > 1$  represent assets whose failure becomes more likely as time goes on.

The Weibull distribution is useful as it can fit most components behavior in ArcelorMittal plants, including for descaling. Indeed, degradation of component is mostly due to long-term

wear and failure at the start of lifetimes are not likely to happen. This behavior is well-fitted with a Weibull distribution with  $k > 1$ .

#### 4.4.2 Assumptions and used data

To measure the cost evolution depending on the prognostics rmse, the following hypothesis on the system modeling are considered:

- The degradation of one component (a valve) is considered
- The component failure rate is modeled with a Weibull distribution.
- The Mean Time To Failure (MTTF) of the component is set to  $mttf = 200$  days and its form factor is  $k = 4$ . This ends up with  $\lambda = \frac{mttf}{\Gamma(1+\frac{1}{k})} = 221$ . The corresponding probability density function (PDF) and cumulative distribution function (CDF) are plotted on Figure 4.6
- Maintenance slots are available every three weeks (21 days). It is considered that components do not degrade during a maintenance slot. Indeed, valves are degraded when high pressure water passes through them while they open or close. The maintenance slots are short enough (2 days) to neglect degradation due to ageing. Therefore, slots duration will be neglected in the models.
- Maintenance operations on valves consist in replacement of the degraded pieces, therefore they are perfect and bring the component in as good as new state, *i.e.* degradation is reset to 0 when maintenance is done.
- Maintenance costs are defined as follows:
  - $C_p = 1,000$  is the standard maintenance operation value
  - $C_c = 20,000$  is the corrective cost. It is set  $C_c$  much higher than the  $C_p$  accordingly to the constraints of steel plants. Indeed, corrective maintenance makes an unexpected stop of the line in operation, which is extremely expensive as furnaces are not stopped and consume in continuous large amount of gas.

**Cumulative distribution function (CDF) approximation** The cost formula relies on the cumulative distribution function (CDF) of the normal distribution  $\Phi(x)$ , which value cannot be expressed in terms of elementary functions. In this study, we will simply use the one proposed by the Python library Scipy (Virtanen et al. 2020), offering a high degree of precision.

However, many numerical approximations exist for the normal distribution CDF. For instance Abramowitz and Stegun 2013 proposed the following approximation for  $x > 0$ , with the absolute error  $|\varepsilon(x)| < 7.5 \cdot 10^{-8}$ :

$$\Phi(x) = 1 - \varphi(x) (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5), \quad t = \frac{1}{1 + b_0 x} \quad (4.9)$$

where

- $\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$  is the standard normal probability density function (PDF)
- $b_0 = 0.2316419, b_1 = 0.319381530, b_2 = -0.356563782, b_3 = 1.781477937, b_4 = -1.821255978, b_5 = 1.330274429$

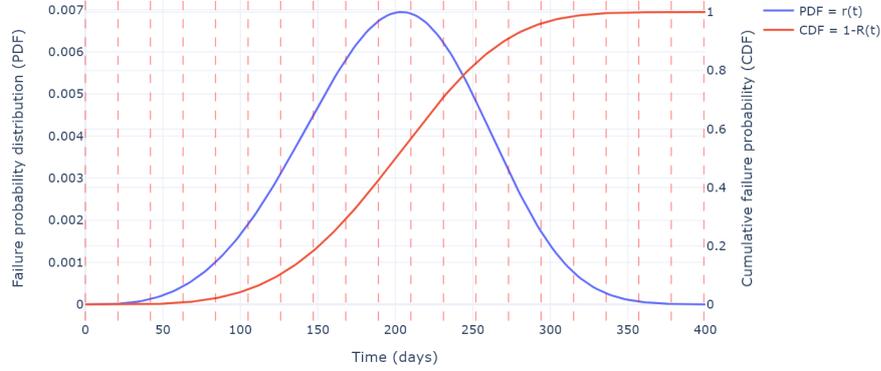


Figure 4.6: Probability density function (PDF) and cumulative distribution function (CDF) of the Weibull distribution used to model the failure of the considered component (a valve). Vertical red dash lines correspond to the available maintenance stops.

Meanwhile, Marsaglia 2004 proposed this simple to remember algorithm:

$$\Phi(x) = \frac{1}{2} + \varphi(x) \left( x + \frac{x^3}{3} + \frac{x^5}{3 \cdot 5} + \frac{x^7}{3 \cdot 5 \cdot 7} + \frac{x^9}{3 \cdot 5 \cdot 7 \cdot 9} + \dots \right)$$

Other approximations exists, many are listed in Hart 1978.

### 4.4.3 Simulation model

#### Decision Making simulation

To support the mathematical modeling of the maintenance cost, a simulation model of the maintenance has been developed. In the literature, decision-making evaluation often requires simulations of the assets' maintenance because evaluating maintenance strategies on real systems is expensive and exposes critical systems to risks, which may require a theoretical validation before being used. Moreover, simulation can evaluate long-term impacts immediately.

Simulations simulate only the decision process based on artificial degradation data. In such case, incremental approaches are mostly used for simulating the systems, such as Meissner, Rahn, and Wicke 2021 which proposes the development of a discrete event simulation framework for determining the expected benefit of maintenance strategy. Similarly, a Java simulation is developed by Skima et al. 2019 to show the importance of the post-prognostics decision-making in the system. The paper by Skima et al. 2019 developed a simulation of their parallel conveyor belts. The paper J. Zuo et al. 2021 developed a simulation to validate the proposed maintenance strategy. It consists in a single component degradation simulation through Gamma process, for which the component degradation increases at each time step, this incremental degradation following a random distribution. Failure of systems can also be represented as probability distributions. For instance, Feldman, Jazouli, and P. A. Sandborn 2009 proposes a simulation-based methodology to calculate the return on investment (ROI) of PHM implementation. They use the simulation proposed by P. Sandborn and Wilkinson 2007. Their simulation model treats all inputs as

probability distributions, exploited by a stochastic analysis implemented as a Monte Carlo simulation. Various maintenance interval and PHM approaches are distinguished by how sampled TTF values are used to model PHM RUL forecasting distributions.

Similarly to these previous works, the proposed simulation model iteratively generates the maintenance over one life cycle of one component (a valve) of ArcelorMittal descaling, according to the stated constraints. Running it a large number of times therefore supplies an experimental average maintenance cost. It is used in the following to validate the results of the maintenance cost obtained through the equations.

The proposed model is based on the valve maintenance formalized in subsection 4.4.2, the corresponding algorithm is presented in Algorithm 2. It consists in randomly picking a failure time from the component failure probability distribution. Then for all maintenance slot, prognostics is simulated to get the estimated  $\widehat{rul}$  (more details are given later). If the predicted  $\widehat{rul}$  is lower than the time before the next step, maintenance is done. The average maintenance cost is the preventive cost divided by the time of the current slot. If the predicted  $\widehat{rul}$  is greater, then no maintenance is done. In this case if the failure happens before the next slot (*i.e.* prognostics failed to predict it), the failure occurs and average maintenance cost is the corrective cost divided by the time of failure. Finally, if the preventive period time is reached without failure nor prognostics, classical periodic maintenance is done. The algorithm is implemented in Python.

---

**Algorithm 2** Simulation of one life cycle of a component and returning average maintenance cost

---

**Require:**

- $C_p$  – Preventive maintenance cost
- $C_c$  – Corrective maintenance cost (including line stop)
- $T_p$  – Preventive maintenance period
- $T_{step}$  – Maintenance slots period

```

1:  $T_{fail} \leftarrow$  Time of occurrence of failure, Sampled from failure probability distribution.
2: for  $T_i \geq T_{step}$  to  $T_p$  with step  $T_{step}$  do    ▷ Going through maintenance slots ( $T_p$  excluded).
3:    $\widehat{rul} \leftarrow$  Prognostics prediction at  $T_i$ 
4:   if  $\widehat{rul} < T_{step}$  then
5:     return  $C_p/T_i$     ▷ If prognostics predicts failure before next slot, maintenance is done
6:   end if
7:   if  $T_{fail} < T_i + T_{step}$  then
8:     return  $C_c/T_{fail}$     ▷ If failure happens before next slot and prognostics failed
9:   end if
10: end for
11: return  $C_p/T_p$ 

```

---

#### 4.4.4 Maintenance cost - Scheduled (preventive) maintenance

Given the considered hypothesis, the preventive maintenance cost (Equation (4.2)) can be computed and the optimal preventive period  $T_{opt}$  can be obtained using Equation (4.3). Figure 4.7 plots the evolution of the maintenance cost for various values of the preventive period. Among all possible values, the available ones are represented as red points on the Figure.

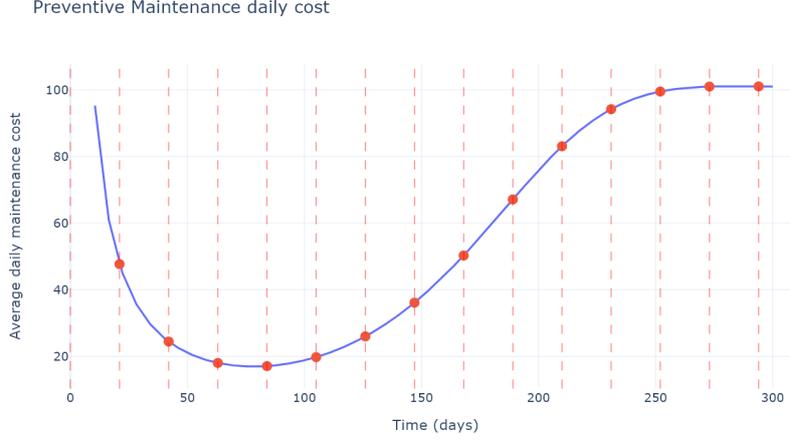


Figure 4.7: Evolution of the preventive maintenance cost defined in Equation (4.2). Units are daily cost accordingly to arbitrary units defined earlier ( $C_p = 1,000$  and  $C_c = 20,000$ ). Red lines correspond to the available maintenance stops. Red points correspond to the value that the preventive period can take.

The resolution of Equation (4.3) gives  $T_{opt} = 4 \cdot T_{step} = 84$  days and an average daily cost of  $C_{prev} = 17.1$ .

#### 4.4.5 Maintenance cost - Predictive maintenance

Considering the hypothesis that the component failures follows a Weibull distribution, the probabilities of failure can be expressed as follows:

$$\mathbb{P}(\overline{rul} < T) = 1 - R(T) = \begin{cases} 1 - e^{-(T/\lambda)^k}, & T \geq 0, \\ 0, & T < 0. \end{cases}$$

$$\mathbb{P}(\overline{rul} = t) = f(t; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k}, & t \geq 0, \\ 0, & t < 0, \end{cases}$$

Therefore, Recall and Specificity can be written as:

$$\text{Recall} = \frac{\sum_{i=1}^n \int_{T_i}^{T_{i+1}} \Phi\left(\frac{T_{i+1}-t}{rmse}\right) \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} dt}{1 - e^{-(T_p/\lambda)^k}}$$

$$\text{Specificity} = \frac{\prod_{i=0}^{n-1} \int_{T_p}^{+\infty} \left[1 - \Phi\left(\frac{T_{i+1}-t}{rmse}\right)\right] \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} dt}{e^{-(T/\lambda)^k}}$$

With these equations, it is possible to evaluate the average maintenance cost with predictive maintenance using Equation (4.6). For the sake of numerical application, values of  $rmse = 13$

#### 4.4. Numerical application for ArcelorMittal descaling valve use-case

(days) and  $rmse = 18.3$  (days)<sup>8</sup> are considered, such as obtained in Chapter 2 on the C-MAPSS dataset for respectively 100% and 25% of the dataset.

Table 4.1: Numerical application of decision-making maintenance cost when prognostics is applied, for two values of  $rmse$ . Those two values correspond to LSTM performances on the C-MAPSS dataset for various dataset sizes.

Dataset percentage	100%	25%
<b>RMSE</b>	<b>13</b>	<b>18.3</b>
<b>Recall</b> = $\mathbb{P}(\widehat{rul} < T   \overline{rul} < T)$	0.73	0.68
<b>Specificity</b> = $\mathbb{P}(\widehat{rul} \geq T   \overline{rul} \geq T)$	0.93	0.93
<b>Average daily cost</b>	13.6	13.9
<b>Cost reduction compared to preventive</b>	20.5%	18.7%

Specificity obtains particularly high result. It is however logical. Indeed, due to  $C_c$  being largely greater than  $C_p$ , preventive maintenance is conservative. Therefore, the probability of the failure occurring after  $T_{opt} + rmse$  is high. And in such case, the probability of the prediction being after  $T_p$  is high. Thus, specificity ends up with high value.

These numerical applications confirm the added value of prognostics for maintenance in the considered industrial scenario. Indeed, average maintenance cost in this representative scenario improve from  $C_{prev} = 17.1$  to  $C_{pred} = 13.9$  when learned on a small dataset, with poor performances of the prognostics and down to  $C_{pred} = 13.6$  with better prognostics performances.

The daily cost values estimated through the equations are confirmed by the developed simulation. Indeed, among 10,000 simulations, the average daily cost for  $rmse = 13$  is 13.74, while the equation predicted 13.6. For  $rmse = 18.3$ , the simulation returned 13.92 for 13.9 for the equations.

The question of the reduction of maintenance cost linked to the gain of prognostics performance is yet to be thoroughly addressed. The following subsection focuses on this topic.

#### 4.4.6 Maintenance cost - Evaluation of the gains with RMSE optimization

Previous Chapters 2 and 3 improved prognostics through data augmentation, the decision-making models can measure their impact on maintenance cost. These improvements mostly consisted in improving prognostics rmse. They achieve RMSE improvement of 10%. This improvement can further improve maintenance cost. This subsection explores two approaches for maintenance cost improvement:

1. Recall and Specificity improvement
2. Update of optimal preventive maintenance period  $T_{opt}$

Firstly, recall and sensitivity increase proportionally when the  $rmse$  decreases, and therefore the cost decreases, according to Equation (4.6). The cost depending on the rmse is shown in

<sup>8</sup>The value of the RMSE can be roughly interpreted as the average value of the error. This is an approximation allowing to give an order of magnitude to interpret the value.

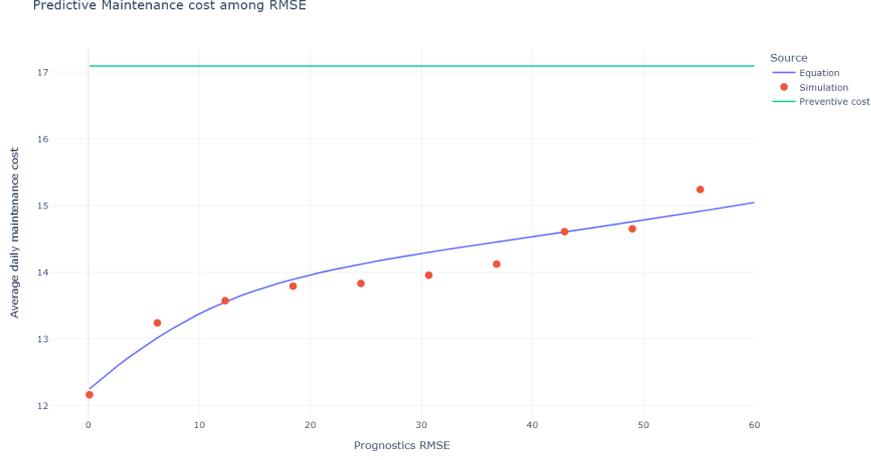


Figure 4.8: Average daily cost of maintenance as a function of the RMSE. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21\text{days}$ ,  $T_{opt} = 4 \cdot 21\text{days}$  and a Weibull distribution with  $k = 4$ ,  $\lambda = 221$ ).

Figure 4.8, it is compared to the preventive cost. It confirms that the cost is proportionally increasing with the RMSE. Indeed, with a lower  $rmse$ , failures to occur are more regularly correctly predicted. Therefore, decision-making reduces the number of failures occurring, reducing by this way the average maintenance cost. It improves faster for smaller values of RMSE ( $rmse < 15$ ). Given that data augmentation on small datasets allowed for an RMSE improvement of 10%, Figure 4.9 plots the relative reduction of maintenance cost for various initial RMSE values. The Figure also plots the cost reduction in the hypothetical case of 20% RMSE improvement.

These first results illustrate that  $rmse$  improvement can enhance maintenance efficiency thanks to better failure prediction. Chapter 2 reduced the  $rmse$  by 10% thanks to data augmentation. This 10% improvement of  $rmse$  can lead to a maintenance cost reduction of 0.6%. This is a small yet positive improvement. It is particularly small compared to the improvement brought by the introduction of prognostics in maintenance. Indeed, this led to a cost reduction of 20%.

Secondly, the determination of the optimal value of the preventive maintenance period  $T_{opt}$ , which was based on preventive cost (4.2) can be reevaluated based on predictive cost (4.6). Indeed, efficient prognostics can reduce the failures occurrence and therefore postpone preventive maintenances. This can therefore further reduce the average daily maintenance cost efficiently. The new value of the optimal preventive maintenance period is therefore defined as follows:

$$T_{opt}(rmse) = \arg \min_{T=n \cdot T_{step}} \left\{ C = \frac{1}{T} \left[ R(T)[Sp(T)C_p + (1 - Sp(T))fC_p] + (1 - R(T))[Rc(T)C_p + (1 - Rc(T))C_c] \right] \right\} \quad (4.10)$$

For instance, Figure 4.10 shows the average maintenance cost for various values of  $T_p$  given  $rmse = 13$ . It is seen that the optimal value of  $T_p$  is  $5 \cdot T_{step}$  in this case while it was  $4 \cdot T_{step}$  with preventive maintenance.

Improving values of  $rmse$  leads to an increase of the optimal value for preventive maintenance period. Indeed, the optimal value of  $T$  is an equilibrium between regular maintenance cost and the cost of the un-avoided failures. With the improvement of the  $rmse$ , more failures are predicted

#### 4.4. Numerical application for ArcelorMittal descaling valve use-case

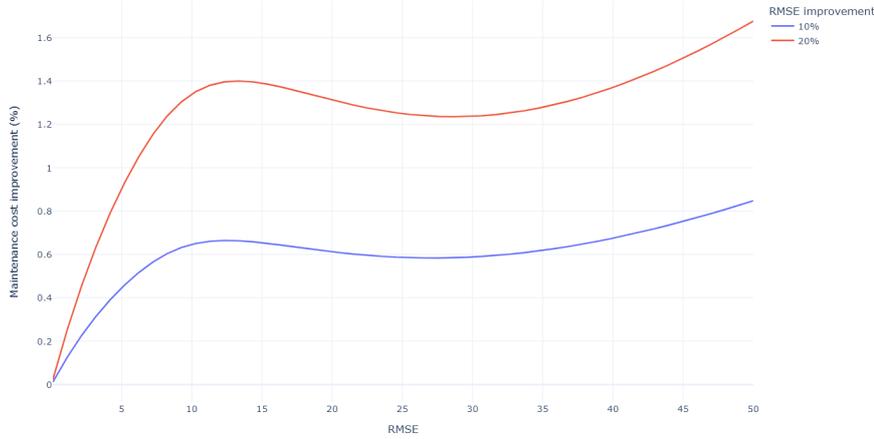


Figure 4.9: Relative improvement of average daily maintenance cost for a 10% or 20% improvement of the RMSE, as a function of the original RMSE. The graph can be read as follows: given an initial  $rmse = 13$ , if it was improved by 10% (getting to  $rmse = 11.7$ ), the average daily maintenance cost would improve by 0.6%. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21$  days,  $T_{opt} = 4 \cdot 21$  days and a Weibull distribution with  $k = 4$ ,  $lambda = 221$ ).

and avoided. Therefore, the cost of un-avoided failures decreases and the equilibrium can move to higher values of  $T$ . Figure 4.11 plots the evolution of cost depending on both  $rmse$  and  $T$ . It can be observed the dependence of the optimal value of  $T_p$  on the  $rmse$ . Indeed, as  $rmse$  decreases,

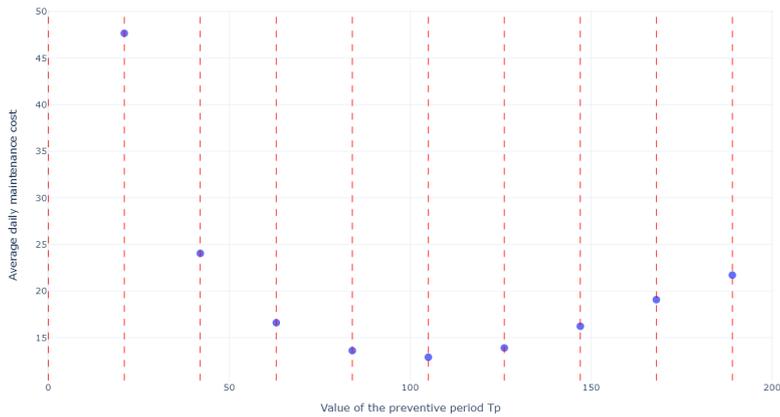


Figure 4.10: Average daily cost of maintenance as a function of the preventive period  $T$ . This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21$ days,  $T_{opt} = 4 \cdot 21$ days and a Weibull distribution with  $k = 4$ ,  $lambda = 119$ ).

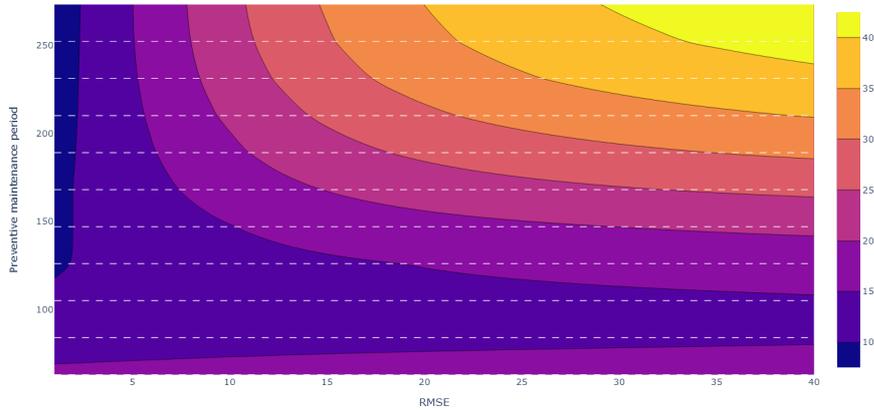


Figure 4.11: Evolution of the maintenance cost as a function of  $T$  and the  $RMSE$ . The white dash lines corresponds to available values for  $T$ , *i.e.* maintenance slots. The color corresponds to the maintenance cost value, purple corresponding to smaller cost values.

the minimal value of  $T$  increases, which can be seen on the Figure with the “valley” going “up” on the left side of the contour plot.

Value of  $T_{opt}$  is therefore determined using Equation (4.10). Figure 4.12 plots the evolution of the average maintenance cost when the preventive period is updated according to predictive maintenance cost.

It is seen that for  $rmse$  inferior to 5, the optimal preventive maintenance period is postponed to the 6th maintenance slot. Average maintenance cost is therefore additionally reduced compared to fixed  $T_{opt}$ . For instance, given an  $rmse = 4$ , the cost is reduced by 14% with the updated value of  $T_{opt}$  ( $c \approx 12.8$ ) compared to the fixed value of  $T_{opt}$  ( $c \approx 1$ ).

## 4.5 Current limits and perspectives

The proposed equations are validated by the simulation for  $rmse$  values below 60, however they diverge from it for greater values. Indeed, previous Figure 4.8 shows an accordance between simulated and calculated maintenance costs for  $rmse$  under 60. However, Figure 4.13 shows that above this value simulated and calculated values diverge. This is problematic as the simulated cost goes over the preventive cost for an  $rmse \geq 90$ . However, this corresponds to high values of the  $rmse$ , which do not concern prognostics algorithms. Indeed, it concerns average errors of 90 days when maintenance slots are distant of 21 days. Therefore, this approximation is conserved and previous results are still considered valid.

Nevertheless, future work on the subject could focus on the precision of the formulas to correct this error. It might be due to the Recall definition. Indeed, two types of “true positive” prognostics are to be considered: (1) failure is predicted at the maintenance slot preceding it (2) failure is predicted at a maintenance slot preceding this one, occurring in early maintenance. These “False True positives”<sup>9</sup> occur more often when  $rmse \gg T_{step}$ , which could explain the divergence above  $rmse > 90$ .

<sup>9</sup>The failure is correctly predicted but not at the right maintenance slot

#### 4.5. Current limits and perspectives

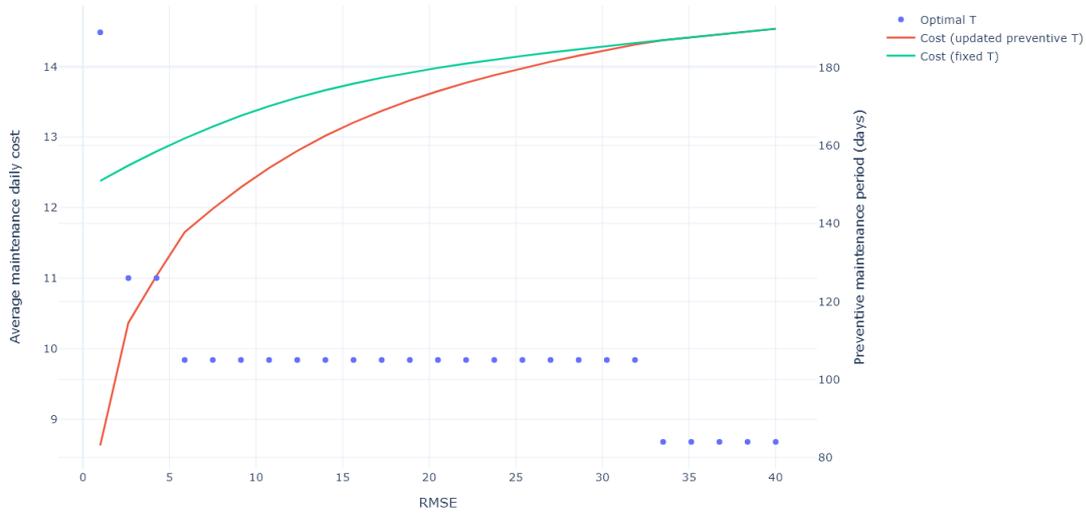


Figure 4.12: Average daily cost of maintenance as a function of the RMSE, with and without updated  $T_{opt}$ . Updated  $T_{opt}$  means that the optimal value of preventive maintenance period  $T$  is computed based on the cost considering prognostics, see Equation (4.10). The blue points corresponds to the optimal value of  $T$  associated to the  $rmse$  value.  $T$  can only take values proportionnal of the maintenance slots period  $T_{step}$ . This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21\text{days}$ ,  $T_{opt} = 4 \cdot 21\text{days}$  and a Weibull distribution with  $k = 4$ ,  $lambda = 119$ ).

## 4.6 Conclusion

This chapter aims to evaluate the impact of prognostics performances on the maintenance cost in the context inspired from ArcelorMittal plants and its specific constraints.

This chapter proposes a formalization of the prognostics developed previously through statistical modeling. Prognostics  $\widehat{rul}$  is modeled as a stochastic variable obeying a normal distribution of mean  $\overline{rul}$  the “real” RUL and of variance the  $rmse$  of the prognostics model. Then the component failure is modeled as a stochastic variable obeying a Weibull law.

These formalizations of prognostics and context are used to define the average maintenance cost based on statistical considerations. This formulation ends up with the expression of the cost in Equation (4.6). This definition is based on the expression of the recall and specificity of the prognostics. Under the considered context, and based on a prognostics formalized as a regression, both of them are developed thoroughly. They end up defined in Equations (4.7) and (4.8).

In parallel, a simulation model has been developed to verify/compare with the theoretical results and ensure the formulations’ relevance. It is developed in Python using an incremental method and is run tens of thousands of times to approximate maintenance costs. It successfully validates the values predicted by the models for considered values of  $rmse$  for which the prognostic model would have been rejected. The equation’s results start diverging from the simulation values only for large values of  $rmse > 60$ , which correspond to un-real values of the  $rmse$ . Therefore, the formulations is considered as valid.

Finally, the maintenance cost formulation based on prognostics characteristics (including  $rmse$ ) is used to calculate the gain in maintenance cost by (i) applying prognostics-based predictive maintenance rather than preventive maintenance and (ii) by improving prognostics performances as proposed in Chapters 2 and 3 (*i.e.* improving its  $rmse$ ). The experiments are parametrized with numerical values based on a valve of the descaling system of ArcelorMittal, described thoroughly in the Chapter 1. It is shown that the implementation of predictive maintenance reduces maintenance cost up to 18-20%. Finally, it is shown that improving prognostics performances of 10% reduces

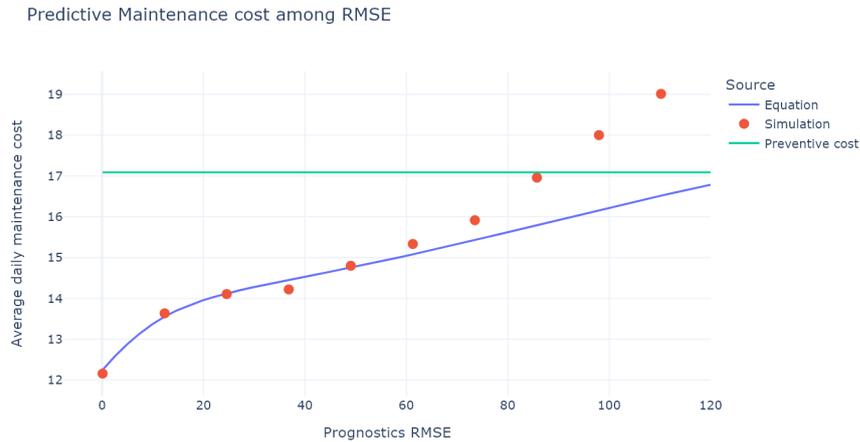


Figure 4.13: Average daily cost of maintenance as a function of the RMSE. This cost is computed using Equation (4.6) with the previously defined variables ( $C_p = 1,000$ ,  $C_c = 20,000$ ,  $T_{step} = 21\text{days}$ ,  $T_{opt} = 4 \cdot 21\text{days}$  and a Weibull distribution with  $k = 4$ ,  $lambda = 119$ ).

maintenance cost up to 1%. Improving prognostics performances of 20% reduces maintenance cost up to 2%. In Chapter 2, prognostics performances were improved by 10% with data augmentation.

Finally, it is also considered the impact of updating of the preventive maintenance period with prognostic efficiency. Indeed, preventive period was initially fixed to optimize preventive cost. Following the implementation of prognostics, preventive period can be updated to be done less often and take advantage of the prognostics ability to avoid failures. This dynamic period optimization reduces even further maintenance cost.



## Chapter 5

# Conclusion and perspectives

This joint PhD project between ArcelorMittal and the CRAN laboratory aimed at optimizing industrial maintenance decision-making through the exploitation of the available sources of information, *i.e.* industrial data and expert knowledge, under the industrial constraints inherited from the steel-making context. According to the industrial context and scientific issues identified, it raises two main problematics, the *small datasets* impact on prognostics performances and the *decision-making* optimization within a constrained steel-making environment. They lead to four scientific contributions, each one addressing one of the scientific issue raised along with the industrial objectives.

The **first scientific issue** this PhD aims to address is to formulate the link between dataset size and prognostics performances to predict the performances improvement brought by additional data. It relates to industrial *small datasets* induced by the ongoing implementation of Industry 4.0 technologies, leading to a low availability of measurements history.

The first contribution of the PhD successfully links prognostics performances to dataset size through an equation. The proposed equation is inspired by Vapnik, Levin, and Cun 1994. It can be used further to predict expected performance gain from additional data when only *small datasets* are available. It is formulated as follows, with parameters  $RMSE^*$ ,  $c_1$  and  $c_2$  to fit:

$$RMSE(n) = RMSE^* + c_1 * \frac{\log(c_2 * n) + 1}{\frac{1}{2} * c_2 * n}$$

Its pertinence is showed on the C-MAPSS public dataset (Abhinav Saxena et al. 2008) via a thorough test methodology defined in the Chapter 2. This methodology consists in simulating *small datasets* through sub-samplings of the full dataset. Furthermore, the methodology aims at considering the sources of variability. Sub-sampling being a source of variance on the results, it is repeated multiple times to ensure the representativity of the results. Similarly, as model initialization is a source of variance, especially on *small datasets*, model initialization and training are also repeated multiple times on each sub-sampling.

The **second scientific issue** this PhD aims to address is the exploitation and optimization of data augmentation on Time Series data to improve prognostics performances when applied to *small datasets* gathered from industrial applications. It relates to the industrial requirement to implement dynamic predictive maintenance as soon as possible, which requires efficient data-driven prognostics algorithms despite the *small datasets* constraint.

The second contribution of the PhD evaluates the improvement of prognostics performance through the application of individual agnostic data augmentation, achieving improvements up

to 10%. All selected transformations do not achieve similar performances improvements, with some of them decreasing performances. Great improvements are systematically obtained on cross-validation, but they are not systematically observed on the test set (which acts as future data for industrial systems).

Data augmentation for prognostics is evaluated on the C-MAPSS dataset via the extension of the methodology developed for the first scientific issue. It is upgraded to conduct data augmentation transformations optimization through a cross-validation approach as test set should not be used for such a task. A test set is finally used for performances evaluation.

The **third scientific issue** this PhD aims to address is the incorporation of expert knowledge into data augmentation methodologies to improve even further the performances and reliability of prognostics for industrial applications. It relates to the capability of ArcelorMittal experts, either from plants or R&D, to supply additional knowledge able to improve prognostics models if used wisely.

Knowledge is gathered on the C-MAPSS dataset from expert sources and statistical analysis to propose upgrades of data augmentation transformations. Few expert knowledge is found in the literature concerning this simulation-based dataset. Statistical knowledge (from statistical analysis) is therefore exploited.

The third contribution of the PhD proposes three upgrades of magnitude-based data augmentation transformations through statistical knowledge integration. First, constant and discrete signals are excluded from data augmentation transformations based on the results of data analysis. Second, signals are grouped based on signals correlations to apply magnitude scaling factors by groups rather than by signals to keep consistency. Third, augmented trajectories are validated by exploiting the distance between correlation matrix before and after augmentation to reject aberrant augmentations based on a threshold.

The three proposed upgrades improved the resilience of transformations though they did not achieve significant improvements nor degradation of performances. In particular, distance-based validation of data augmentation improves the resilience of data augmentation to performance reduction (notably due to hyper-parameters values).

So, the **fourth scientific issue** this PhD aims to address is the evaluation of prognostics performance impact on decision-making under the industrial context provided by ArcelorMittal use-cases. The goal is to evaluate how the prognostics improvement, thanks to data augmentation, will be reflected in the the decision-making. It integrates prognostics uncertainty into maintenance cost formulation, leading to a feedback between prognostics improvement and the final maintenance cost, applied in the ArcelorMittal use-case.

The fourth contribution of the PhD proposes a formulation of the average maintenance cost considering prognostics performances within the ArcelorMittal maintenance decision-making context. This formulation is based on a few modelling hypotheses. The first hypothesis is that the component failure is modelled as a stochastic variable obeying a Weibull law. The second hypothesis is that the Prognostics output, *i.e.*  $\widehat{rul}$ , is modeled as a stochastic variable obeying a normal distribution centered on the “true” RUL and of variance the *rmse* of the prognostics model. These assumptions allow us to include in the cost (Equation (4.6)) the performance of the prognostics thanks to the formulation of its recall and specificity.

In parallel, a simulation model is developed to confirm the validity of the theoretical results and ensure the formula relevance. The simulated maintenance costs successfully confirm the values predicted by the theoretical model. The experiments are parametrized with numerical values based on the real descaling use-case of ArcelorMittal hot strip mill plant, described thoroughly in the Chapter 1.

Finally, the maintenance cost formulation based on prognostics performance is used to calculate the improvement in the maintenance cost obtained (i) by applying prognostics-based predictive

maintenance instead of scheduled preventive maintenance and (ii) by improving prognostics performances with data augmentation. It is shown, for descaling use-case of ArcelorMittal, that the implementation of predictive maintenance reduces maintenance cost up to 18–20%. In addition, it is shown that improving prognostics performances of 10% reduces maintenance cost up to 1%; Improving prognostics performances of 20% reduces maintenance cost up to 2%.

## 5.1 Perspectives

Above the four contributions proposed by this PhD’s work, some scientific questions were left un-answered and others appeared during the development. They constitute promising perspectives for further work on the subject.

The results obtained on *small datasets* and with data augmentation by the first and second contributions can be extended in various directions.

Firstly, both studies could be extended to additional datasets. It is evaluated on the C-MAPSS FD001 dataset, acting as a reference for the PHM community, but additional results would extend the pertinence of the conclusion.

Secondly, it shall be examined if applying data augmentation to a *full dataset* can improve further prognostics performances. This would be coherent with the expected performances predicted by the equation proposed in the first contribution.

Then, the data augmentation family of decomposition-based methods presented by Iwana and Uchida 2021 could be evaluated on prognostics *small datasets*. It would require more thorough data analysis to identify pertinent parametrization. This is sketched in the development of the third contributions on the C-MAPSS dataset with exponential decomposition without success. However, these methods have potential for performances improvements, and further development could take inspiration from data processing proposed in the literature to identify appropriate decompositions.

The mitigated results at exploiting expert knowledge of the third contribution lead to the formulation of perspectives for further system knowledge exploitation for data augmentation.

Functional analysis of systems could be used to define link between signals and so guide data augmentation into generating new signals more coherent with the real behaviour of the system. Similarly to signal grouping done after in statistical knowledge, but endorsed by knowledge and system models.

Knowledge could also be extracted from dysfunctional analysis, like FMECA and HAZOP, to guide the generation of data augmentation methods. The recent paper Han et al. 2022 propose for instance to use causality graphs for expert knowledge integration, which could be used to represent the information contained in (dys)functional analysis. This would be particularly efficient to face new degradation modes, not or less represented in training sets. This issue was not considered in this work, which focused on single degradation mode datasets.

Further improvements could also be implemented in the proposed knowledge-based methods. Correlation matrices have been used for signal grouping and divergence assessing of newly-generated trajectories. The variations of correlation matrices could be used to identify trajectories ought to vary the most to focus data augmentation on them. Such approach ought to apply data augmentation at a “microscopic” scale rather than “macroscopic” as done by agnostic transformations.

At a “microscopic” scale, expert knowledge could also be used to customize data augmentation transformations to each individual signal. This would end up in the formulation of application-specific data augmentation transformations directly based on signals individual characteristics.

## *Chapter 5. Conclusion and perspectives*

Finally, the fourth contribution proposed a first insight into prognostics integration in maintenance cost, and many directions remains to be explored.

Firstly, additional use-cases could be considered with different settings to measure the cost gain from preventive to prognostics-based decision, and evaluate if different context would be more dependent on prognostics performances.

Secondly, the formulation could be extended to biased prognostics, as the current formulation considered an un-biased prognostics estimator. Such consideration is even more pertinent as prognostics models are often biased towards “too soon” predictions, as illustrated by the score proposed in Abhinav Saxena et al. 2008.

Thirdly, the proposed equation could be extended to advanced maintenance strategies, such as multi-components considerations and imperfect maintenance.

# Annexes

## 1 Student's t-test values

The following table lists values for  $t$ -distributions with  $v$  degrees of freedom for a range of one-sided or two-sided critical regions.

The first column is  $v$ , the percentages along the top are confidence levels, and the numbers in the body of the table are the  $t_{\alpha, n-1}$  factors described in the section on confidence intervals.

The last row with infinite  $v$  gives critical points for a normal distribution since a  $t$ -distribution with infinitely many degrees of freedom is a normal distribution.

$v$	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690

Annexes

v	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
inf	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
v	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

# Bibliography

- Abramowitz, Milton and Irene A. Stegun, eds. (2013). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. 9. Dover print.; [Nachdr. der Ausg. von 1972]. Dover Books on Mathematics. New York, NY: Dover Publ. ISBN: 978-0-486-61272-0.
- Ahmad, Jamil, Khan Muhammad, and Sung Wook Baik (Aug. 2017). “Data Augmentation-Assisted Deep Learning of Hand-Drawn Partially Colored Sketches for Visual Search”. In: *PLOS ONE* 12.8, e0183838. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0183838. (Visited on 12/01/2022).
- Akiba, Takuya et al. (July 2019). “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '19. New York, NY, USA: Association for Computing Machinery, pp. 2623–2631. ISBN: 978-1-4503-6201-6. DOI: 10.1145/3292500.3330701. (Visited on 10/24/2022).
- Alam, Sadeqh R. et al. (Mar. 2021). “Generalizable Cone Beam CT Esophagus Segmentation Using Physics-Based Data Augmentation”. In: *Phys. Med. Biol.* 66.6, p. 065008. ISSN: 0031-9155. DOI: 10.1088/1361-6560/abe2eb. (Visited on 01/25/2023).
- Almuet, Mohammad and Juhana Salim (June 2014). “From a Literature Review to a Conceptual Framework for Automation Knowledge Acquisition in Supply Chain Management”. In: *Journal of Theoretical and Applied Information Technology* 64, pp. 730–745. URL: <https://www.researchgate.net/publication/264420376>.
- An, Guozhong (Apr. 1996). “The Effects of Adding Noise During Backpropagation Training on a Generalization Performance”. In: *Neural Computation* 8.3, pp. 643–674. ISSN: 0899-7667. DOI: 10.1162/neco.1996.8.3.643. (Visited on 12/07/2022).
- Arslan, Mehmet et al. (Sept. 2019). “SMOTE and Gaussian Noise Based Sensor Data Augmentation”. In: *2019 4th International Conference on Computer Science and Engineering (UBMK)*, pp. 1–5. DOI: 10.1109/UBMK.2019.8907003.
- Assaf, Roy et al. (Feb. 2018). “Wear Rate–State Interactions within a Multi-Component System: A Study of a Gearbox-Accelerated Life Testing Platform”. In: p. 10.
- Atamuradov, Vepa et al. (Dec. 2017). “Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation”. In: *International Journal of Prognostics and Health Management* 8, p. 31.
- Backman, Jere, Vesa Kyllönen, and Heli Helaakoski (Jan. 2019). “Methods and Tools of Improving Steel Manufacturing Processes: Current State and Future Methods”. In: *IFAC-PapersOnLine*. 9th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2019 52.13, pp. 1174–1179. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2019.11.355. (Visited on 03/06/2023).
- Barros, A., C. Berenguer, and A. Grall (Dec. 2003). “Optimization of Replacement Times Using Imperfect Monitoring Information”. In: *IEEE Transactions on Reliability* 52.4, pp. 523–533. ISSN: 1558-1721. DOI: 10.1109/TR.2003.821944.

## BIBLIOGRAPHY

- Bazart, Loïc et al. (Aug. 2014). “Model-Based Sticker Detection in Continuous Casting”. In: DOI: 10.13140/2.1.3458.6885.
- Behera, Sourajit (2021). “Multiscale Deep Bidirectional Gated Recurrent Neural Networks Based Prognostic Method for Complex Non-Linear Degradation Systems”. In: *Information Sciences*, p. 25.
- Ben Hassen, Wafa et al. (2015). “Distributed Sensor Fusion for Wire Fault Location Using Sensor Clustering Strategy”. In: *International Journal of Distributed Sensor Networks* 11.4, p. 538643.
- Béranger, Gérard et al., eds. (1996). *The Book of Steel*. Trans. by James H. Davidson. Andover, UK: Intercept. ISBN: 978-2-7430-0022-6.
- Bergmeir, Christoph, Rob J. Hyndman, and José M. Benítez (Apr. 2016). “Bagging Exponential Smoothing Methods Using STL Decomposition and Box–Cox Transformation”. In: *International Journal of Forecasting* 32.2, pp. 303–312. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2015.07.002. (Visited on 06/03/2021).
- Bole, Brian et al. (2011). “Adaptive Load-Allocation for Prognosis-Based Risk Management”. In: *PHM\_CONF* 3.1. ISSN: 2325-0178. DOI: 10.36001/phmconf.2011.v3i1.2019. (Visited on 06/21/2021).
- Bonissone, P. and Naresh Iyer (July 2008). “Knowledge and Time: A Framework for Soft Computing Applications in PHM”. In: *Uncertainty and Intelligent information Systems* July 2008, pp. 333–344. ISSN: 978-981-279-234-1. DOI: 10.1142/9789812792358\_0024.
- Bougacha, Omar, Christophe Varnier, and Nouredine Zerhouni (2020). “A Review of Post-Prognostics Decision-Making in Prognostics and Health Management”. In: *International Journal of Prognostics and Health Management* 11.15, p. 31. URL: <https://hal.archives-ouvertes.fr/hal-03107810> (visited on 03/31/2022).
- Bougacha, Omar, Christophe Varnier, Nouredine Zerhouni, and Sonia Hajri-Gabouj (Dec. 2018). “A Post-Prognostic Decision Approach for Production and Maintenance Planning”. In: *48th Conference on Computers & Industrial Engineering*.
- Box, George EP and Gwilym M Jenkins (1970). *Time Series Analysis Forecasting and Control*. Tech. rep. WISCONSIN UNIV MADISON DEPT OF STATISTICS.
- Bozkurt, Erdoğan, İlhami M. Orak, and Yasin Tunçkaya (2021). “Performance Analysis of Hot Metal Temperature Prediction in a Blast Furnace and Expert Suggestion System Proposal Using Neural, Statistical and Fuzzy Models”. In: *Metall. Res. Technol.* 118.3, p. 321. ISSN: 2271-3646, 2271-3654. DOI: 10.1051/meta1/2021043. (Visited on 05/31/2021).
- Byrne, Nick et al. (2021). “A Persistent Homology-Based Topological Loss Function for Multi-class CNN Segmentation of Cardiac MRI”. In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by Esther Puyol Anton et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 3–13. ISBN: 978-3-030-68107-4. DOI: 10.1007/978-3-030-68107-4\_1.
- Cao, Hong, Vincent Y. F. Tan, and John Z. F. Pang (Dec. 2014). “A Parsimonious Mixture of Gaussian Trees Model for Oversampling in Imbalanced and Multimodal Time-Series Classification”. In: *IEEE Trans. Neural Netw. Learning Syst.* 25.12, pp. 2226–2239. ISSN: 2162-237X, 2162-2388. DOI: 10.1109/TNNLS.2014.2308321. (Visited on 12/05/2022).
- Carrasco, Jacinto et al. (Oct. 2021). “Anomaly Detection in Predictive Maintenance: A New Evaluation Framework for Temporal Unsupervised Anomaly Detection Algorithms”. In: *Neurocomputing* 462, pp. 440–452. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2021.07.095. (Visited on 10/08/2021).
- Catal, Cagatay and Banu Diri (Mar. 2009). “Investigating the Effect of Dataset Size, Metrics Sets, and Feature Selection Techniques on Software Fault Prediction Problem”. In: *Information Sciences* 179.8, pp. 1040–1058. ISSN: 0020-0255. DOI: 10.1016/j.ins.2008.12.001. (Visited on 12/20/2021).

- Chao, Manuel Arias et al. (Oct. 2020). “Fusing Physics-based and Deep Learning Models for Prognostics”. In: *arXiv*. arXiv: 2003.00732. URL: <http://arxiv.org/abs/2003.00732> (visited on 02/03/2021).
- Chaoub, Alaaeddine, Christophe Cerisara, et al. (June 2022). “Deep Learning Representation Pre-training for Industry 4.0”. In: *PHME\_CONF* 7.1, pp. 571–573. ISSN: 2325-016X, 2325-016X. DOI: 10.36001/phme.2022.v7i1.2784. (Visited on 04/03/2023).
- Chaoub, Alaaeddine, Alexandre Voisin, et al. (June 2021). “Learning Representations with End-to-End Models for Improved Remaining Useful Life Prognostic”. In: *PHM Society European Conference* 6.1, pp. 1–8. ISSN: 2325-016X. DOI: 10.36001/phme.2021.v6i1.2785. (Visited on 09/22/2021).
- Chawathe, Sudarshan S. (Jan. 2019). “Condition Monitoring of Hydraulic Systems by Classifying Sensor Data Streams”. In: *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0898–0904. DOI: 10.1109/CCWC.2019.8666564.
- Chawla, N. V. et al. (June 2002). “SMOTE: Synthetic Minority Over-sampling Technique”. In: *Journal of Artificial Intelligence Research* 16, pp. 321–357. ISSN: 1076-9757. DOI: 10.1613/jair.953. (Visited on 10/21/2022).
- Chen, X (2021). “Application of Data-Driven Models to Predictive Maintenance: Bearing Wear Prediction at TATA Steel”. In: *Expert Systems With Applications*, p. 11.
- Dahiya, Navdeep et al. (2021). “Multitask 3D CBCT-to-CT Translation and Organs-at-Risk Segmentation Using Physics-Based Data Augmentation”. In: *Medical Physics* 48.9, pp. 5130–5141. ISSN: 2473-4209. DOI: 10.1002/mp.15083. (Visited on 01/25/2023).
- de Beer, Jeroen, Ernst Worrell, and Kornelis Blok (1998). “Future Technologies for Energy-Efficient Iron and Steel Making”. In: *Annual Review of Energy and the Environment* 23.1, pp. 123–205. DOI: 10.1146/annurev.energy.23.1.123. (Visited on 03/21/2023).
- Dempster, Angus, François Petitjean, and Geoffrey I. Webb (Sept. 2020). “ROCKET: Exceptionally Fast and Accurate Time Series Classification Using Random Convolutional Kernels”. In: *Data Min Knowl Disc* 34.5, pp. 1454–1495. ISSN: 1573-756X. DOI: 10.1007/s10618-020-00701-z. (Visited on 03/31/2022).
- Dempster, Angus, Daniel F. Schmidt, and Geoffrey I. Webb (Aug. 2021). “MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification”. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. KDD ’21. New York, NY, USA: Association for Computing Machinery, pp. 248–257. ISBN: 978-1-4503-8332-5. DOI: 10.1145/3447548.3467231. (Visited on 02/01/2023).
- Derigent, W. et al. (Jan. 2009). “Opportunistic Maintenance Based on Fuzzy Modelling of Component Proximity”. In: *CIRP Annals* 58.1, pp. 29–32. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2009.03.079. (Visited on 05/20/2021).
- DeVries, Terrance and Graham W. Taylor (Feb. 2017a). “Dataset Augmentation in Feature Space”. In: *arXiv:1702.05538 [cs, stat]*. arXiv: 1702.05538 [cs, stat]. URL: <http://arxiv.org/abs/1702.05538> (visited on 06/03/2021).
- (Nov. 2017b). *Improved Regularization of Convolutional Neural Networks with Cutout*. arXiv: 1708.04552 [cs]. URL: <http://arxiv.org/abs/1708.04552> (visited on 03/07/2023).
- Diez-Oliván, Alberto (June 2017). “Data-Driven Prognostics Using a Combination of Constrained K-means Clustering, Fuzzy Modeling and LOF-based Score”. In: *Neurocomputing* 241, pp. 97–107. DOI: 10.1016/j.neucom.2017.02.024.
- Ding, Peng, Minping Jia, and Xiaoli Zhao (June 2021). “Meta Deep Learning Based Rotating Machinery Health Prognostics toward Few-Shot Prognostics”. In: *Applied Soft Computing* 104, p. 107211. ISSN: 1568-4946. DOI: 10.1016/j.asoc.2021.107211. (Visited on 04/03/2023).
- Dinh, Duc-Hanh, Phuc Do, and Benoit Iung (June 2020a). “Degradation Modeling and Reliability Assessment for a Multi-Component System with Structural Dependence”. In: *Computers &*

## BIBLIOGRAPHY

- Industrial Engineering* 144, p. 106443. ISSN: 0360-8352. DOI: 10.1016/j.cie.2020.106443. (Visited on 04/14/2021).
- Dinh, Duc-Hanh, Phuc Do, and Benoit Iung (Jan. 2020b). “Maintenance Optimisation for Multi-Component System with Structural Dependence: Application to Machine Tool Sub-System”. In: *CIRP Annals* 69.1, pp. 417–420. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2020.04.004. (Visited on 04/14/2021).
- Do, Phuc (2013). “Dynamic Grouping Maintenance with Time Limited Opportunities”. In: p. 9. — (Oct. 2015). “Maintenance Grouping for Multi-Component Systems with Availability Constraints and Limited Maintenance Teams”. In: p. 12. DOI: 10.1016/j.res.2015.04.022.
- Do, Phuc et al. (Jan. 2015). “A Proactive Condition-Based Maintenance Strategy with Both Perfect and Imperfect Maintenance Actions”. In: *Reliability Engineering & System Safety* 133, pp. 22–32. ISSN: 0951-8320. DOI: 10.1016/j.res.2014.08.011. (Visited on 03/23/2021).
- Ellefsen, Andre Listou (2019). “A Comprehensive Survey of Prognostics and Health Management Based on Deep Learning for Autonomous Ships”. In: *IEEE TRANSACTIONS ON RELIABILITY*, p. 23.
- Eltoft, T. (May 2002). “Data Augmentation Using a Combination of Independent Component Analysis and Non-Linear Time-Series Prediction”. In: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN’02 (Cat. No.02CH37290)*. Vol. 1, 448–453 vol.1. DOI: 10.1109/IJCNN.2002.1005514.
- Emmanouilidis, Christos et al. (2010). “Condition Monitoring Based on Incremental Learning and Domain Ontology for Condition-Based Maintenance”. In: *Proceedings of APMS 2010 International Conference on Advances in Production Management Systems*. URL: <https://cris.vtt.fi/en/publications/condition-monitoring-based-on-incremental-learning-and-domain-ont> (visited on 02/08/2021).
- Erhan, Dumitru et al. (2009). “The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training”. In: p. 8.
- Feldman, Kiri, Taoufik Jazouli, and Peter A. Sandborn (June 2009). “A Methodology for Determining the Return on Investment Associated With Prognostics and Health Management”. In: *IEEE Transactions on Reliability* 58.2, pp. 305–316. ISSN: 1558-1721. DOI: 10.1109/TR.2009.2020133.
- Fields, Tonya, George Hsieh, and Jules Chenou (Dec. 2019). “Mitigating Drift in Time Series Data with Noise Augmentation”. In: *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 227–230. DOI: 10.1109/CSCI49370.2019.00046.
- Fink, Olga et al. (June 2020). “Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications”. In: *Engineering Applications of Artificial Intelligence* 92, pp. 103678–103692. DOI: 10.1016/j.engappai.2020.103678. (Visited on 02/08/2021).
- Frederick, D., Jonathan A. DeCastro, and J. Litt (Oct. 2007). “User’s Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)”. In: URL: <https://www.semanticscholar.org/paper/User%27s-Guide-for-the-Commercial-Modular-System-Frederick-DeCastro/e4a0750e98db745eacececf8baa7efe3aeb20c3> (visited on 01/11/2023).
- Frigui, F. et al. (2018). “Global Methodology for Damage Detection and Localization in Civil Engineering Structures”. In: *Engineering Structures* 171, pp. 686–695.
- Gao, Jingkun et al. (Sept. 2021). *RobustTAD: Robust Time Series Anomaly Detection via Decomposition and Convolutional Neural Networks*. DOI: 10.48550/arXiv.2002.09545. arXiv: 2002.09545 [cs, eess, stat]. (Visited on 12/05/2022).

## BIBLIOGRAPHY

- Gao, Robert X. et al. (Jan. 2020). “Big Data Analytics for Smart Factories of the Future”. In: *CIRP Annals* 69.2, pp. 668–692. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2020.05.002. (Visited on 11/30/2021).
- Garga, Amulya et al. (Feb. 2001). “Hybrid Reasoning for Prognostic Learning in CBM Systems”. In: *2001 IEEE Aerospace Conference Proceedings (Cat. No.01TH8542)*. Vol. 6, pp. 2957–2969. ISBN: 978-0-7803-6599-5. DOI: 10.1109/AERO.2001.931316.
- Gay, Antonin, Benoit Iung, et al. (Nov. 2021). “A Short Review on the Integration of Expert Knowledge in Prognostics for PHM in Industrial Applications”. In: *2021 5th International Conference on System Reliability and Safety (ICSRS)*, pp. 286–292. DOI: 10.1109/ICSRS53853.2021.9660645.
- Gay, Antonin, Alexandre Voisin, et al. (Jan. 2022). “Data Augmentation-based Prognostics for Predictive Maintenance of Industrial System”. In: *CIRP Annals* 71.1, pp. 409–412. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2022.04.005. (Visited on 09/07/2022).
- Gouriveau, Rafael, Kamal Medjaher, and Noureddine Zerhouni (Oct. 2016). *From Prognostics and Health Systems Management to Predictive Maintenance 1: Monitoring and Prognostics*. John Wiley & Sons. ISBN: 978-1-119-37102-1.
- Guo, J., Z. Li, and M. Li (Sept. 2020). “A Review on Prognostics Methods for Engineering Systems”. In: *IEEE Transactions on Reliability* 69.3, pp. 1110–1129. ISSN: 1558-1721. DOI: 10.1109/TR.2019.2957965.
- Han, Yinghua et al. (Oct. 2022). “A Novel Knowledge Enhanced Graph Neural Networks for Fault Diagnosis with Application to Blast Furnace Process Safety”. In: *Process Safety and Environmental Protection* 166, pp. 143–157. ISSN: 09575820. DOI: 10.1016/j.psep.2022.08.014. (Visited on 03/28/2023).
- Hao, Songhua, Jun Yang, and Christophe Berenguer (Oct. 2020). “Condition-Based Maintenance with Imperfect Inspections for Continuous Degradation Processes”. In: *Applied Mathematical Modelling* 86, pp. 311–334. ISSN: 0307-904X. DOI: 10.1016/j.apm.2020.05.013. (Visited on 04/05/2023).
- Hart, John F., ed. (1978). *Computer Approximations*. Reprint of the ed. published by Wiley, New York, 1968. Huntington, N.Y: Krieger. ISBN: 978-0-88275-642-4.
- Helwig, Nikolai, Steffen Klein, and Andreas Schütze (Jan. 2015). “Identification and Quantification of Hydraulic System Faults Based on Multivariate Statistics Using Spectral Vibration Features”. In: *Procedia Engineering*. Eurosensors 2015 120, pp. 1225–1228. ISSN: 1877-7058. DOI: 10.1016/j.proeng.2015.08.835. (Visited on 02/11/2020).
- Heng, Aiwina et al. (Apr. 2009). “Rotating Machinery Prognostics: State of the Art, Challenges and Opportunities”. In: *Mechanical Systems and Signal Processing* 23.3, pp. 724–739. ISSN: 08883270. DOI: 10.1016/j.ymsp.2008.06.009. (Visited on 04/10/2019).
- Herdin, M. and E. Bonek (2005). “A MIMO Correlation Matrix Based Metric for Characterizing Non-Stationarity”. In: URL: <https://www.semanticscholar.org/paper/A-MIMO-Correlation-Matrix-based-Metric-for-Herdin-Bonek/c95d1d6cde187f1cf81822aaa9bcf3251ebf6151> (visited on 01/25/2023).
- Herdin, M., N. Czink, et al. (May 2005). “Correlation Matrix Distance, a Meaningful Measure for Evaluation of Non-Stationary MIMO Channels”. In: *2005 IEEE 61st Vehicular Technology Conference*. Vol. 1, 136–140 Vol. 1. DOI: 10.1109/VETECS.2005.1543265.
- Herr, Nathalie, Jean-Marc Nicod, and Christophe Varnier (Aug. 2014). “Prognostics-Based Scheduling in a Distributed Platform: Model, Complexity and Resolution”. In: *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 1054–1059. DOI: 10.1109/CoASE.2014.6899455.

## BIBLIOGRAPHY

- Hoang, Anh (Feb. 2016). “Energy Efficiency Performance-Based Prognostics for Aided Maintenance Decision-Making: Application to a Manufacturing Platform”. In: *Journal of Cleaner Production*, p. 20.
- Hoang, Anh, Phuc Do, and Benoît Iung (Oct. 2015). “Prognostics on Energy Efficiency Performance for Maintenance Decision-Making: Application to Industrial Platform TELMA”. In: *2015 Prognostics and System Health Management Conference (PHM)*, pp. 1–7. DOI: 10.1109/PHM.2015.7380096.
- Hongzhou, Wang and Pham Hoang (2006). *Reliability and Optimal Maintenance*. Springer Series in Reliability Engineering. Springer London. ISBN: 978-1-84628-324-6. DOI: 10.1007/b138077. (Visited on 03/24/2023).
- Hou-bo, He, Zhao Jian-min, and Xu chang-an (Jan. 2011). “Cost-Benefit Model for PHM”. In: *Procedia Environmental Sciences*. 2011 3rd International Conference on Environmental Science and Information Application Technology ESIAAT 2011 10, pp. 759–764. ISSN: 1878-0296. DOI: 10.1016/j.proenv.2011.09.123. (Visited on 03/15/2023).
- Hua, Zhiguang (2021). “Remaining Useful Life Prediction of PEMFC Systems under Dynamic Operating Conditions”. In: *Energy Conversion and Management* 231, p. 113825. DOI: 10.1016/j.enconman.2021.113825.
- Huynh, K. T., A. Grall, and C. Bérenguer (Mar. 2017). “Assessment of Diagnostic and Prognostic Condition Indices for Efficient and Robust Maintenance Decision-Making of Systems Subject to Stress Corrosion Cracking”. In: *Reliability Engineering & System Safety* 159, pp. 237–254. ISSN: 0951-8320. DOI: 10.1016/j.ress.2016.11.022. (Visited on 04/04/2022).
- Iung, Benoît, Phuc Do, et al. (2016). “Opportunistic Maintenance Based on Multi-Dependent Components of Manufacturing System”. In: *CIRP Annals* 65.1, pp. 401–404. ISSN: 00078506. DOI: 10.1016/j.cirp.2016.04.063. (Visited on 04/13/2021).
- Iung, Benoît, Thomas Laloix, and Alexandre Voisin (2018). “Predictive Maintenance-based Part Quality Control from Joint Product-Process-Machine Requirements: Application to a Machine Tool Subsystem”. In:
- Iwana, Brian Kenji and Seiichi Uchida (July 2021). “An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks”. In: *PLOS ONE* 16.7, e0254841. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0254841. (Visited on 09/10/2021).
- Iyer, N., K. Goebel, and P. Bonissone (Mar. 2006). “Framework for Post-Prognostic Decision Support”. In: *2006 IEEE Aerospace Conference*. DOI: 10.1109/AERO.2006.1656108.
- Jaitly, Navdeep and Geoffrey E Hinton (2013). “Vocal Tract Length Perturbation (VTLP) Improves Speech Recognition”. In: *Proceedings of the 30 Th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013*. P. 5.
- Jardine, Andrew K.S., Daming Lin, and Dragan Banjevic (Oct. 2006). “A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance”. In: *Mechanical Systems and Signal Processing* 20.7, pp. 1483–1510. ISSN: 08883270. DOI: 10.1016/j.ymsp.2005.09.012. (Visited on 04/10/2019).
- Javed, Kamran, Rafael Gouriveau, and Nouredine Zerhouni (Sept. 2017). “State of the Art and Taxonomy of Prognostics Approaches, Trends of Prognostics Applications and Open Issues towards Maturity at Different Technology Readiness Levels”. In: *Mechanical Systems and Signal Processing* 94, pp. 214–236. DOI: 10.1016/j.ymsp.2017.01.05.
- Jensen, Uwe (1996). “Stochastic Models of Reliability and Maintenance: An Overview”. In: *Reliability and Maintenance of Complex Systems*. Ed. by Süleyman Özekici. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3–36. ISBN: 978-3-642-08250-4. DOI: 10.1007/978-3-662-03274-9\_1. (Visited on 03/24/2023).

- Jiao, Ruihua, Kaixiang Peng, and Jie Dong (July 2021). “Remaining Useful Life Prediction for a Roller in a Hot Strip Mill Based on Deep Recurrent Neural Networks”. In: *IEEE/CAA Journal of Automatica Sinica* 8.7, pp. 1345–1354. ISSN: 2329-9274. DOI: 10.1109/JAS.2021.1004051.
- Johansson, Carl-Anders (2014). “Context Driven Remaining Useful Life Estimation”. In: *Procedia CIRP* 22, pp. 181–185. DOI: 10.1016/j.procir.2014.07.129.
- Kalgren, Patrick W. et al. (Sept. 2006). “Defining PHM, A Lexical Evolution of Maintenance and Logistics”. In: *2006 IEEE Autotestcon*, pp. 353–358. DOI: 10.1109/AUTEST.2006.283685.
- Karpatne, Anuj et al. (Feb. 2018). “Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling”. In: *arXiv:1710.11431 [physics, stat]*. arXiv: 1710.11431 [physics, stat]. URL: <http://arxiv.org/abs/1710.11431> (visited on 03/24/2021).
- Karras, Tero et al. (2020). “Analyzing and Improving the Image Quality of StyleGAN”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8110–8119. URL: [https://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Karras\\_Analyzing\\_and\\_Improving\\_the\\_Image\\_Quality\\_of\\_StyleGAN\\_CVPR\\_2020\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2020/html/Karras_Analyzing_and_Improving_the_Image_Quality_of_StyleGAN_CVPR_2020_paper.html) (visited on 12/01/2022).
- Kermanpur, Ahmad et al. (2008). “Artificial Neural Network Modeling of High Pressure Descaling Operation in Hot Strip Rolling of Steels”. In: *ISIJ International* 48.7, pp. 963–970. DOI: 10.2355/isijinternational.48.963.
- Kerschke, Pascal et al. (Mar. 2019). “Automated Algorithm Selection: Survey and Perspectives”. In: *Evolutionary Computation* 27.1, pp. 3–45. ISSN: 1063-6560. DOI: 10.1162/evco\_a\_00242. (Visited on 05/07/2021).
- Kim, Tae San and So Young Sohn (Dec. 2021). “Multitask Learning for Health Condition Identification and Remaining Useful Life Prediction: Deep Convolutional Neural Network Approach”. In: *J Intell Manuf* 32.8, pp. 2169–2179. ISSN: 1572-8145. DOI: 10.1007/s10845-020-01630-w. (Visited on 04/03/2023).
- Kızılersü, Ayşe, Markus Kreer, and Anthony W. Thomas (Apr. 2018). “The Weibull Distribution”. In: *Significance* 15.2, pp. 10–11. ISSN: 1740-9705, 1740-9713. DOI: 10.1111/j.1740-9713.2018.01123.x. (Visited on 03/15/2023).
- Laloix, Thomas (Nov. 2018). “Machine Health Check Methodology to Help Maintenance in Operational Condition: Application to Machine Tool from Its Kinematic Monitoring”. PhD thesis.
- Laloix, Thomas et al. (2019). “Parameter Identification of Health Indicator Aggregation for Decision-Making in Predictive Maintenance: Application to Machine Tool”. In: DOI: 10.1016/j.cirp.2019.03.020.
- Le Guennec, Arthur, Simon Malinowski, and Romain Tavenard (Sept. 2016). “Data Augmentation for Time Series Classification Using Convolutional Neural Networks”. In: *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*. URL: <https://halshs.archives-ouvertes.fr/halshs-01357973> (visited on 06/03/2021).
- Lee, Jay et al. (Jan. 2014). “Prognostics and Health Management Design for Rotary Machinery Systems—Reviews, Methodology and Applications”. In: *Mechanical Systems and Signal Processing* 42.1, pp. 314–334. ISSN: 0888-3270. DOI: 10.1016/j.ymsp.2013.06.004. (Visited on 03/20/2023).
- Lei, Xin et al. (June 2015). “PHM Based Predictive Maintenance Optimization for Offshore Wind Farms”. In: *2015 IEEE Conference on Prognostics and Health Management (PHM)*, pp. 1–8. DOI: 10.1109/ICPHM.2015.7245027.
- Leukel, Joerg, Julian González, and Martin Riekert (Oct. 2021). “Adoption of Machine Learning Technology for Failure Prediction in Industrial Maintenance: A Systematic Review”. In: *Journal of Manufacturing Systems* 61, pp. 87–96. ISSN: 02786125. DOI: 10.1016/j.jmsy.2021.08.012. (Visited on 10/08/2021).

## BIBLIOGRAPHY

- Li, Guangtao et al. (May 2021). “Ensemble Learning Based Methods for Crown Prediction of Hot-Rolled Strip”. In: *ISIJ Int.* 61.5, pp. 1603–1613. ISSN: 0915-1559, 1347-5460. DOI: 10.2355/isijinternational.ISIJINT-2020-639. (Visited on 05/17/2021).
- Li, Yanrui and Chunjie Yang (Apr. 2021). “Domain Knowledge Based Explainable Feature Construction Method and Its Application in Ironmaking Process”. In: *Engineering Applications of Artificial Intelligence* 100, p. 104197. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2021.104197. (Visited on 02/23/2021).
- Lian, Xiaoyuan, Zhong Zheng, and Xiaoqiang Gao (Dec. 2022). “Steel Plant Rescheduling Considering the Balance of Oxygen Supply and Demand under Converter Failure”. In: *Computers & Chemical Engineering* 168, p. 108064. ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2022.108064. (Visited on 12/14/2022).
- Łomowski, Roman and Szczepan Hummel (July 2020). “A Method to Estimate the Remaining Useful Life of a Filter Using a Hybrid Approach Based on Kernel Regression and Simple Statistics”. In: *PHME\_CONF* 5.1. ISSN: 2325-016X. URL: <https://papers.phmsociety.org/index.php/phme/article/view/1316> (visited on 02/15/2021).
- Lyu, Jianhua et al. (Oct. 2020). “Remaining Useful Life Estimation with Multiple Local Similarities”. In: *Engineering Applications of Artificial Intelligence* 95, p. 103849. DOI: 10.1016/j.engappai.2020.103849.
- Manan, Zainuddin Abdul et al. (Nov. 2017). “Advances in Process Integration Research for CO2 Emission Reduction – A Review”. In: *Journal of Cleaner Production* 167, pp. 1–13. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2017.08.138. (Visited on 03/06/2023).
- Marsaglia, George (2004). “Evaluating the Normal Distribution”. In: *J. Stat. Soft.* 11.4. ISSN: 1548-7660. DOI: 10.18637/jss.v011.i04. (Visited on 02/09/2023).
- Matsokis, Aristeidis et al. (Jan. 2010). “An Ontology-based Model for Providing Semantic Maintenance”. In: *IFAC Proceedings Volumes*. 1st IFAC Workshop on Advanced Maintenance Engineering, Services and Technology 43.3, pp. 12–17. ISSN: 1474-6670. DOI: 10.3182/20100701-2-PT-4012.00004. (Visited on 02/17/2021).
- McDonnell, Darren (July 2018). “Predicting the Unpredictable – Consideration of Human and Organisational Factors in Maintenance Prognostics”. In: *Journal of Loss Prevention in the Process Industries* 54, pp. 131–145. DOI: 10.1016/j.jlp.2018.03.008.
- Medina-Oliva, Gabriela et al. (Sept. 2014). “Predictive Diagnosis Based on a Fleet-Wide Ontology Approach”. In: *Knowledge-Based Systems*. Enhancing Experience Reuse and Learning 68, pp. 40–57. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2013.12.020. (Visited on 05/06/2019).
- Meissner, Robert, Antonia Rahn, and Kai Wicke (Oct. 2021). “Developing Prescriptive Maintenance Strategies in the Aviation Industry Based on a Discrete-Event Simulation Framework for Post-Prognostics Decision Making”. In: *Reliability Engineering & System Safety* 214, p. 107812. ISSN: 0951-8320. DOI: 10.1016/j.ress.2021.107812. (Visited on 06/13/2022).
- Merkert, Lennart et al. (Jan. 2015). “Scheduling and Energy – Industrial Challenges and Opportunities”. In: *Computers & Chemical Engineering*. A Tribute to Ignacio E. Grossmann 72, pp. 183–198. ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2014.05.024. (Visited on 03/06/2023).
- Moghaddam, Mohsen, Qiliang Chen, and Abhijit V. Deshmukh (Feb. 2020). “A Neuro-Inspired Computational Model for Adaptive Fault Diagnosis”. In: *Expert Systems with Applications* 140, p. 112879. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2019.112879. (Visited on 02/01/2023).
- Narkhede, Meenal V., Prashant P. Bartakke, and Mukul S. Sutaone (Jan. 2022). “A Review on Weight Initialization Strategies for Neural Networks”. In: *Artif Intell Rev* 55.1, pp. 291–322. ISSN: 1573-7462. DOI: 10.1007/s10462-021-10033-z. (Visited on 03/06/2023).

- Nguyen, Huu-Tri et al. (Sept. 2021). “Machine Learning for Integrating Combustion Chemistry in Numerical Simulations”. In: *Energy and AI* 5, p. 100082. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100082. (Visited on 05/17/2021).
- Nguyen, Khanh and Phuc Do (Apr. 2019). “Joint Optimization of Monitoring Quality and Replacement Decisions in Condition-Based Maintenance”. In.
- Nguyen, Kim-Anh, Phuc Do, and Antoine Grall (Dec. 2017). “Joint Predictive Maintenance and Inventory Strategy for Multi-Component Systems Using Birnbaum’s Structural Importance”. In: *Reliability Engineering & System Safety* 168, pp. 249–261. ISSN: 09518320. DOI: 10.1016/j.ress.2017.05.034. (Visited on 04/13/2021).
- Nikolić, Danko (Oct. 2017). “Why Deep Neural Nets Cannot Ever Match Biological Intelligence and What to Do about It?” In: *Int. J. Autom. Comput.* 14.5, pp. 532–541. ISSN: 1476-8186, 1751-8520. DOI: 10.1007/s11633-017-1093-8. (Visited on 11/06/2019).
- Núñez, David Lira and Milton Borsato (June 2017). “An Ontology-Based Model for Prognostics and Health Management of Machines”. In: *Journal of Industrial Information Integration* 6, pp. 33–46. ISSN: 2452-414X. DOI: 10.1016/j.jii.2017.02.006. (Visited on 02/17/2021).
- Omigbodun, Akinyinka O. et al. (2019). “The Effects of Physics-Based Data Augmentation on the Generalizability of Deep Neural Networks: Demonstration on Nodule False-Positive Reduction”. In: *Medical Physics* 46.10, pp. 4563–4574. ISSN: 2473-4209. DOI: 10.1002/mp.13755. (Visited on 01/25/2023).
- Ooi, Ching Sheng, Meng Hee Lim, and Mohd Salman Leong (2019). “Self-Tune Linear Adaptive-Genetic Algorithm for Feature Selection”. In: *IEEE Access* 7, pp. 138211–138232. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2942962.
- Pan, Yan et al. (Jan. 2022). “Knowledge-Based Data Augmentation of Small Samples for Oil Condition Prediction”. In: *Reliability Engineering & System Safety* 217, p. 108114. ISSN: 0951-8320. DOI: 10.1016/j.ress.2021.108114. (Visited on 01/25/2023).
- Papoulis, Athanasios and S. Unnikrishna Pillai (2002). *Probability, Random Variables, and Stochastic Processes*. 4th ed. Boston: McGraw-Hill. ISBN: 978-0-07-366011-0 978-0-07-112256-6.
- Park, Daniel S. et al. (Sept. 2019). “SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition”. In: *Interspeech 2019*. ISCA, pp. 2613–2617. DOI: 10.21437/Interspeech.2019-2680. (Visited on 12/05/2022).
- Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio (May 2013). “On the Difficulty of Training Recurrent Neural Networks”. In: *Proceedings of the 30th International Conference on Machine Learning*. PMLR, pp. 1310–1318. URL: <https://proceedings.mlr.press/v28/pascanu13.html> (visited on 03/24/2023).
- Pedregosa, Fabian et al. (Nov. 2011). “Scikit-Learn: Machine Learning in Python”. In: *MACHINE LEARNING IN PYTHON*.
- Peng, Ying, Ming Dong, and Ming Jian Zuo (Sept. 2010). “Current Status of Machine Prognostics in Condition-Based Maintenance: A Review”. In: *Int J Adv Manuf Technol* 50.1, pp. 297–313. ISSN: 1433-3015. DOI: 10.1007/s00170-009-2482-0. (Visited on 01/17/2020).
- Peng, Z.K. and F.L. Chu (Mar. 2004). “Application of the Wavelet Transform in Machine Condition Monitoring and Fault Diagnostics: A Review with Bibliography”. In: *Mechanical Systems and Signal Processing* 18.2, pp. 199–221. ISSN: 08883270. DOI: 10.1016/S0888-3270(03)00075-X. (Visited on 03/22/2023).
- Pérès, François and Daniel Noyes (2003). “Evaluation of a Maintenance Strategy by the Analysis of the Rate of Repair”. In: *Quality and Reliability Engineering International* 19.2, pp. 129–148.
- Porzio, Giacomo Filippo et al. (Dec. 2013). “Reducing the Energy Consumption and CO2 Emissions of Energy Intensive Industries through Decision Support Systems – An Example of Application to the Steel Industry”. In: *Applied Energy* 112, pp. 818–833. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2013.05.005. (Visited on 03/06/2023).

## BIBLIOGRAPHY

- Rashid, Khandakar M. and Joseph Louis (May 2019). “Time-Warping: A Time Series Data Augmentation of IMU Data for Construction Equipment Activity Identification”. In: *ISARC Proceedings*, pp. 651–657. URL: [http://www.iaarc.org/publications/2019\\_proceedings\\_of\\_the\\_36th\\_isarc/time\\_warping-a\\_time\\_series\\_data\\_augmentation\\_of\\_imu\\_data\\_for\\_construction\\_equipment\\_activity\\_identification.html](http://www.iaarc.org/publications/2019_proceedings_of_the_36th_isarc/time_warping-a_time_series_data_augmentation_of_imu_data_for_construction_equipment_activity_identification.html) (visited on 10/21/2022).
- Raudys, S.J. and A.K. Jain (Mar. 1991). “Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.3, pp. 252–264. ISSN: 1939-3539. DOI: 10.1109/34.75512.
- Rinne, Horst (Nov. 2008). *The Weibull Distribution: A Handbook*. New York: Chapman and Hall/CRC. ISBN: 978-0-429-14257-4. DOI: 10.1201/9781420087444.
- Rizzolo, Ludovic et al. (Aug. 2011). “Aggregation of Health Assessment Indicators of Industrial Systems”. In: p. 8.
- ROCKET (2023). *ROCKET: Exceptionally Fast and Accurate Time Series Classification Using Random Convolutional Kernels*, URL: <https://slideslive.com/38932458/rocket-exceptionally-fast-and-accurate-time-series-classification-using-random-convolutional-kernels?ref=og-meta-tags> (visited on 03/07/2023).
- Roemer, Michael and Gregory Kacprzynski (2001). “Assessment of Data and Knowledge Fusion Strategies for Prognostics and Health Management”. In: *2001 IEEE Aerospace Conference Proceedings (Cat. No.01TH8542)*. Vol. 6, pp. 2979–2988. DOI: 10.1109/AERO.2001.931318.
- Ruiz-Tagle Palazuelos, Andrés and Enrique López Drogue (Feb. 2021). “System-Level Prognostics and Health Management: A Graph Convolutional Network-Based Framework”. In: *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 235.1, pp. 120–135. ISSN: 1748-006X. DOI: 10.1177/1748006X20935760. (Visited on 02/15/2021).
- Sahli, Aymane, Richard Evans, and Arthi Manohar (Jan. 2021). “Predictive Maintenance in Industry 4.0: Current Themes”. In: *Procedia CIRP*. 54th CIRP CMS 2021 - Towards Digitalized Manufacturing 4.0 104, pp. 1948–1953. ISSN: 2212-8271. DOI: 10.1016/j.procir.2021.11.329. (Visited on 11/30/2021).
- Sandborn, Peter and Chris Wilkinson (Dec. 2007). “A Maintenance Planning and Business Case Development Model for the Application of Prognostics and Health Management (PHM) to Electronic Systems”. In: *Microelectronics Reliability* 47, pp. 1889–1901. DOI: 10.1016/j.microrel.2007.02.016.
- Sankararaman, Shankar and Kai Goebel (July 2015). “Uncertainty in Prognostics and Systems Health Management”. In: *International Journal of Prognostics and Health Management* 6. DOI: 10.36001/ijphm.2015.v6i4.2319.
- Satish, B. and Sarma Nuthalapati (July 2005). “A Fuzzy BP Approach for Diagnosis and Prognosis of Bearing Faults in Induction Motors”. In: *IEEE Power Engineering Society General Meeting, 2005*. Vol. 3, 2291–2294 Vol. 3. DOI: 10.1109/PES.2005.1489277.
- Saxena, A. and K. Goebel (2008). *Turbofan Engine Degradation Simulation Data Set*. <http://ti.arc.nasa.gov/project/data-repository>. URL: <http://ti.arc.nasa.gov/project/prognostic-data-repository> (visited on 09/14/2021).
- Saxena, Abhinav et al. (Oct. 2008). “Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation”. In: *2008 International Conference on Prognostics and Health Management*, pp. 1–9. DOI: 10.1109/PHM.2008.4711414.
- Schneider, Tizian, Nikolai Helwig, and Andreas Schütze (May 2018). “Automatic Feature Extraction and Selection for Condition Monitoring and Related Datasets”. In: *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6. DOI: 10.1109/I2MTC.2018.8409763.

- She, D. et al. (June 2020). “Wasserstein Distance Based Deep Multi-Feature Adversarial Transfer Diagnosis Approach under Variable Working Conditions”. In: *J. Inst.* 15.06, P06002–P06002. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/06/P06002. (Visited on 02/18/2021).
- Shorten, Connor and Taghi M. Khoshgoftaar (July 2019). “A Survey on Image Data Augmentation for Deep Learning”. In: *Journal of Big Data* 6.1, p. 60. ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. (Visited on 10/24/2022).
- Shukla, Bibhudhendu, Ip-Shing Fan, and Ian Jennions (July 2020). “Opportunities for Explainable Artificial Intelligence in Aerospace Predictive Maintenance”. In: *PHME\_CONF* 5.1, pp. 11–11. ISSN: 2325-016X. URL: <https://papers.phmsociety.org/index.php/phme/article/view/1231> (visited on 02/15/2021).
- Sikorska, J. Z., M. Hodkiewicz, and L. Ma (July 2011). “Prognostic Modelling Options for Remaining Useful Life Estimation by Industry”. In: *Mechanical Systems and Signal Processing* 25.5, pp. 1803–1836. ISSN: 0888-3270. DOI: 10.1016/j.ymssp.2010.11.018. (Visited on 03/20/2023).
- Singla, Rohit et al. (Dec. 2022). “Speckle and Shadows: Ultrasound-specific Physics-based Data Augmentation for Kidney Segmentation”. In: *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*. PMLR, pp. 1139–1148. URL: <https://proceedings.mlr.press/v172/singla22a.html> (visited on 01/25/2023).
- Skima, Haithem et al. (Mar. 2019). “Post-Prognostics Decision Making in Distributed MEMS-based Systems”. In: *J Intell Manuf* 30.3, pp. 1125–1136. ISSN: 1572-8145. DOI: 10.1007/s10845-017-1312-8. (Visited on 05/06/2021).
- Souto, Nelson et al. (2022). “Performance Analysis of Work-Roll Wear Models on Hot Rolling”. In: *Key Engineering Materials* 926, pp. 621–631. ISSN: 1662-9795. DOI: 10.4028/p-q6v323. (Visited on 09/09/2022).
- Um, Terry T. et al. (Nov. 2017). “Data Augmentation of Wearable Sensor Data for Parkinson’s Disease Monitoring Using Convolutional Neural Networks”. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ICMI ’17. New York, NY, USA: Association for Computing Machinery, pp. 216–220. ISBN: 978-1-4503-5543-8. DOI: 10.1145/3136755.3136817. (Visited on 09/13/2021).
- Vapnik, Vladimir, Esther Levin, and Yann Le Cun (Sept. 1994). “Measuring the VC-Dimension of a Learning Machine”. In: *Neural Computation* 6.5, pp. 851–876. ISSN: 0899-7667. DOI: 10.1162/neco.1994.6.5.851. (Visited on 01/05/2022).
- Vasilyeva, Natalia, Elmira Fedorova, and Alexandr Kolesnikov (May 2021). “Big Data as a Tool for Building a Predictive Model of Mill Roll Wear”. In: *Symmetry* 13.5, p. 859. DOI: 10.3390/sym13050859. (Visited on 05/17/2021).
- Vieira, Rodrigo J. A., Matteo Lietti, and Miguel A. Sanz-bobi (Jan. 2012). “New Variable Health Threshold Based on the Life Observed for Improving the Scheduled Maintenance of a Wind Turbine”. In: *IFAC Proceedings Volumes*. Vol. 45. 2nd IFAC Workshop on Advanced Maintenance Engineering, Services and Technology, pp. 187–192. DOI: 10.3182/20121122-2-ES-4026.00032. (Visited on 05/14/2021).
- Virtanen, Pauli et al. (Mar. 2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nat Methods* 17.3, pp. 261–272. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-019-0686-2. (Visited on 02/13/2023).
- Vogel-Heuser, Birgit and Dieter Hess (2016). “Guest Editorial Industry 4.0—Prerequisites and Visions”. In: *IEEE Transactions on automation Science and Engineering* 13.2, pp. 411–413.
- Vogel-Heuser, Birgit, Daniel Schütz, et al. (Oct. 2014). “Model-Driven Engineering of Manufacturing Automation Software Projects – A SysML-based Approach”. In: *Mechatronics*. 1. Model-Based Mechatronic System Design 2. Model Based Engineering 24.7, pp. 883–897. ISSN: 0957-4158. DOI: 10.1016/j.mechatronics.2014.05.003. (Visited on 04/05/2023).

## BIBLIOGRAPHY

- Vogl, Gregory W, Brian A Weiss, and M Alkan Donmez (2014). “Standards for Prognostics and Health Management (PHM) Techniques within Manufacturing Operations”. In: p. 13.
- Voisin, Alexandre, Jean-Baptiste Léger, et al. (May 2014). “Health Monitoring and Prognostic Assessment in a Fleet Context”. In: *MFPT 2014 Conference: Technology Solutions for Affordable Sustainment*.
- Voisin, Alexandre, Gabriela Medina-Oliva, et al. (July 2013). “Fleet-Wide Diagnostic and Prognostic Assessment”. In: p. 11.
- Wang, Hongzhou and Hoang Pham (Oct. 1996). “A Quasi Renewal Process and Its Applications in Imperfect Maintenance”. In: *International Journal of Systems Science* 27.10, pp. 1055–1062. ISSN: 0020-7721, 1464-5319. DOI: 10.1080/00207729608929311. (Visited on 03/24/2023).
- (Feb. 2006). “Availability and Maintenance of Series Systems Subject to Imperfect Repair and Correlated Failure and Repair”. In: *European Journal of Operational Research* 174, pp. 1706–1722. DOI: 10.1016/j.ejor.2005.03.030.
- Wang, Rongxi et al. (Mar. 2020). “A Dilated Convolution Network-Based LSTM Model for Multi-Step Prediction of Chaotic Time Series”. In: *Comp. Appl. Math.* 39.1, p. 30. ISSN: 2238-3603, 1807-0302. DOI: 10.1007/s40314-019-1006-2. (Visited on 02/11/2020).
- Wang, Yaqing et al. (June 2020). “Generalizing from a Few Examples: A Survey on Few-shot Learning”. In: *ACM Comput. Surv.* 53.3, 63:1–63:34. ISSN: 0360-0300. DOI: 10.1145/3386252. (Visited on 04/03/2023).
- Wang, Zhao-Qiang et al. (Feb. 2018). “Remaining Useful Life Prediction of Degrading Systems Subjected to Imperfect Maintenance: Application to Draught Fans”. In: *Mechanical Systems and Signal Processing* 100, pp. 802–813. ISSN: 0888-3270. DOI: 10.1016/j.ymsp.2017.08.016. (Visited on 03/23/2021).
- Weibull, Waloddi (Sept. 1951). “A Statistical Distribution Function of Wide Applicability”. In: *Journal of Applied Mechanics* 18.3, pp. 293–297. ISSN: 0021-8936, 1528-9036. DOI: 10.1115/1.4010337. (Visited on 03/15/2023).
- Wesendrup, Kevin and Bernd Hellgrath (July 2020). “A Process-based Review of Post-Prognostics Decision-Making”. In: *PHME\_CONF* 5.1, pp. 12–12. ISSN: 2325-016X. URL: <https://papers.phmsociety.org/index.php/phme/article/view/1203> (visited on 02/15/2021).
- Wu, Zekun and Kaiwei Wu (Apr. 2022). *Prognostic Classification Based on Random Convolutional Kernel*. DOI: 10.48550/arXiv.2204.04527. arXiv: 2204.04527 [eess]. (Visited on 10/20/2022).
- Xi, Han et al. (Aug. 2021). “Artificial Intelligent Based Energy Scheduling of Steel Mill Gas Utilization System towards Carbon Neutrality”. In: *Applied Energy* 295, p. 117069. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.117069. (Visited on 05/20/2021).
- Xiahou, T., Z. Zeng, and Y. Liu (Apr. 2021). “Remaining Useful Life Prediction by Fusing Expert Knowledge and Condition Monitoring Information”. In: *IEEE Transactions on Industrial Informatics* 17.4, pp. 2653–2663. ISSN: 1941-0050. DOI: 10.1109/TII.2020.2998102.
- Xu, Chumbao (Charles) and Da-qiang Cang (Mar. 2010). “A Brief Overview of Low CO2 Emission Technologies for Iron and Steel Making”. In: *Journal of Iron and Steel Research, International* 17.3, pp. 1–7. ISSN: 1006-706X. DOI: 10.1016/S1006-706X(10)60064-7. (Visited on 03/21/2023).
- Yoon, Jinsung, Daniel Jarrett, and Mihaela van der Schaar (2019). “Time-Series Generative Adversarial Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc. URL: <https://proceedings.neurips.cc/paper/2019/hash/c9efe5f26cd17ba6216bbe2a7d26d490-Abstract.html> (visited on 02/02/2023).
- Yu, Wennian, II Yong Kim, and Chris Mechefske (Aug. 2019). “Remaining Useful Life Estimation Using a Bidirectional Recurrent Neural Network Based Autoencoder Scheme”. In: *Mechanical*

## BIBLIOGRAPHY

- Systems and Signal Processing* 129, pp. 764–780. ISSN: 0888-3270. DOI: 10.1016/j.ymsp.2019.05.005. (Visited on 04/13/2021).
- Zhang, Liangwei et al. (2019). “A Review on Deep Learning Applications in Prognostics and Health Management”. In: 7, p. 24.
- Zio, Enrico (Feb. 2022). “Prognostics and Health Management (PHM): Where Are We and Where Do We (Need to) Go in Theory and Practice”. In: *Reliability Engineering & System Safety* 218, p. 108119. ISSN: 0951-8320. DOI: 10.1016/j.ress.2021.108119. (Visited on 01/31/2023).
- Zuo, Jian et al. (May 2021). “Post-Prognostics Decision Making Strategy to Manage the Economic Lifetime of a Two-Stack PEMFC System”. In: *2021 Annual Reliability and Maintainability Symposium (RAMS)*. Orlando, FL, USA: IEEE, pp. 1–7. ISBN: 978-1-72818-017-5. DOI: 10.1109/RAMS48097.2021.9605719. (Visited on 06/13/2022).