FEDERAL UNIVERSITY OF CARIRI
RESEARCH, POSTGRADUATE, AND INNOVATION
PRO-RECTORY
MASTERING IN SUSTAINABLE REGIONAL DEVELOPMENT

Flávio Murilo de Carvalho Leal

A framework for automation of data recording, modelling, and optimal statistical control of production lines

Crato

March 2020

Flávio Murilo de Carvalho Leal

A framework for automation of data recording, modelling, and optimal statistical control of production lines

Dissertation of Post Graduation Program in Sustainable Regional Development of Cariri Federal University.
Research line: modelling and technology
Advisor: Paulo Renato Alves Firmino
Co-Advisor: Francisco Alixandre Ávila Rodrigues

Federal University of Cariri - UFCA

Research, Postgraduate, and Innovation Pro-Rectory

Mastering in Sustainable Regional Development

Crato

March 2020

Cataloging in Publication
Universidade Federal do Cariri
Library System

L471f Leal, Flávio Murilo de Carvalho.
A framework for automation of data recording, modelling, and optimal statistical control of production lines / Flávio Murilo de Carvalho Leal . – 2020.
91 f.: il. color.30 cm.

Dissertation (master’s degree) - Universidade Federal do Cariri, Programa de pós-graduação em Desenvolvimento Regional Sustentável, Crato, 2020.

Orientação: Prof. Dr. Paulo Renato Alves Firmino.
Coorientação: Prof. Dr. Francisco Alixandre Ávila Rodrigues.

1. Automation. 2. Quality control. 3. Statistical modelling. 4. Embedded systems. I. Title.

CDD 338.4

Bibliotecária: Glacinésia Leal Mendonça
CRB 3/ 925

Flávio Murilo de Carvalho Leal

A framework for automation of data recording, modelling, and optimal statistical control of production lines

Dissertation of Post Graduation Program in Sustainable Regional Development of Cariri Federal University.
Research line: modelling and technology
Advisor: Paulo Renato Alves Firmino
Co-Advisor: Francisco Alixandre Ávila Rodrigues

Paulo Renato C A

Advisor - Prof. Dr. Paulo Renato Alves Firmino
UFCA

Francisco Alixandre Ávila Rodrigues

Co-Advisor - Prof. Dr. Francisco Alixandre Ávila Rodrigues
UFCA

Márcio J. C. Moura

Prof. Dr. Márcio José das Chagas Moura
UFPE

Luiz Araujo Filho

Prof. Dr. Luiz Medeiros de Araujo Lima Filho
UFPB

Crato

March 2020

I dedicate this dissertation to my father José Leal Neto and my grandfather Francisco Leal Sobrinho (in memoriam).

Acknowledgements

To God for supporting me and renewing my strength every day. Without Him nothing would be possible.

To my mother Maria Eunice Leal, strong and hardworking womam, and my father José Leal Neto who welcomed me with love and care.

To my wife Maria Kelvia Vieira da Silva Leal for her love and understanding in the hardest moments.

To my grandfather Francisco Leal Sobrinho, source of inspiration by his life style and will power.

To my biological mother Marilene Augusto by the gift life and for being ever present, and my biological father Flávio Carvalho who fate has brought us together again.

To my brothers, sisters, nephews, nieces, uncles, aunts, father-in-law (José Osmar da Silva), brothers-in-law, sisters-in-law and to my whole family, including Akira, Ozzy and Logan, who provided a therapeutic company.

To my Advisor prof. Paulo Renato Alves Firmino for the inspiration, patience, availability and help in making the best decisions.

To my Co-Advisor prof. Francisco Alixandre Ávila Rodrigues for the teachings and monitoring in the research.

To the members of the examination board prof. Márcio José das Chagas Moura and prof. Luiz Medeiros de Araujo Lima Filho.

To all member of the MESOR research group.

To the UFCA and the PRODER, all professors and colleagues.

To the Technological Education Center Institute, specially the Faculty of Technology of Cariri for the opportunity to join the academic career, specially to the director professor Emerson Lacerda de Sousa for the advices.

To the friends Reginaldo Vaz Saraiva, Cícero de Lima Santos, Samuel Torres Brasil and all my professors who became colleagues.

To Luciene for lead me in my first work experience in CENTEC.

To the Domum School for the most loving and captivating experience I have ever had in teaching, specially to the Lidia (in memoriam) and Darc.

To the University Center Dr. Leão Sampaio, specially to Thiago Bessa Pontes, and to the University Center Vale do Salgado, for the rich experience in teaching.

To all others CENTEC, Unileão and Domum employees, cleaning ladies, janitors, cookers, librarians.

To my pupils who in the act of teaching allowed me to learn.

To everyone who accompanied me and believed in me.

"It is not dying that frightens us. It's living without ever having done our best."

Castlevania

Abstract

Unarguably, the automation of data collection and subsequent statistical treatment enhance the quality of industrial management systems. The rise of accessible digital technologies has enabled the introduction of the Industry 4.0 pillars in Cariri local companies. Particularly, such practice positively contributes to the triple bottom line of sustainable development: People, Environment, and Economy. The present work aims to provide a general automated framework for data recording and statistical control of conveyor belts in production lines. The software has been developed in three layers: graphical user interface, in `PHP` language; database collection, search, and safeguard, in `MySQL`; computational statistics, in `R`; and hardware control, in `C`. The computational statistics are based on the combination of artificial neural nets and autoregressive integrated and moving average models, via minimal variance method. The hardware components are composed by open source hardware as `Arduino` based boards and modular or industrial sensors. Specifically, the embedded system is designed to constantly monitor and record a number of measurable characteristics of the conveyor belts (e.g. electric consumption and temperature), via a number of sensors, allowing both the computation of statistical control metrics and the evaluation of the quality of the production system. As a case study, the project makes use of a laminated limestone production line, located at the Mineral Technology Center, Nova Olinda, Ceará state, Brazil.

Keywords: Automation, Quality control, Statistical modelling, Embedded systems.

Resumo

Indiscutivelmente, a automação da coleta de dados e o subsequente tratamento estatístico aumentam a qualidade dos sistemas de gestão industrial. O surgimento de tecnologias digitais acessíveis possibilitou a introdução dos pilares da Indústria 4.0 nas empresas locais do Cariri. Particularmente, tal prática contribui positivamente para o triplo resultado do desenvolvimento sustentável: Pessoas, Meio Ambiente e Economia. O presente trabalho tem como objetivo fornecer um *Framework* geral automatizado para registro de dados e controle estatístico de esteiras transportadoras em linhas de produção. O software foi desenvolvido em três camadas: interface gráfica do usuário, em linguagem `PHP`; coleta, pesquisa e proteção de banco de dados em `MySQL`; estatística computacional, em `R`; e controle de hardware, em `C`. As estatísticas computacionais são baseadas na combinação de redes neurais artificiais e modelos autorregressivos integrados e de média móvel, via método de mínima variância. Os componentes de *hardware* são compostos por *hardware open source* como placas baseadas em `Arduino` e sensores modulares ou industriais. Especificamente, o sistema embarcado é projetado para monitorar e registrar constantemente uma série de características mensuráveis das esteiras transportadoras (por exemplo, consumo elétrico e temperatura), por meio de uma série de sensores, permitindo tanto o cálculo de métricas de controle estatístico quanto a avaliação da qualidade do sistema de produção. Como estudo de caso, o projeto utiliza uma linha de produção de calcário laminado, localizada no Centro de Tecnologia Mineral, Nova Olinda, Ceará, Brasil.

Palavras-chave: Automação, Controle da qualidade, Modelagem estatística, Sistemas embarcados.

List of Figures

List of Tables

Glossary

ADC Analogical Digital Converters. 47

ANN Artificial Neural Networks. 33

APM Aggregate Performance Metric. 37

ARV Average Relative Variance. 37

BCNF Boyce–Codd Normal Form. 49

COOPEDRAS Cooperativa de Mineração dos Produtores da Pedra Cariri. 24

CRUD Create, Read, Update, and Delete. 48

CTMC Centro de Tecnologia Mineral do Cariri. 19

DBMS Database Management System. 50

DPNM Departamento Nacional de Produção Mineral. 24

DS Development Sustainable. 53

FINEP Financiadora de Estudos e Projetos. 24

FUNCAP Fundação Cearense de Apoio ao Desenvolvimento Científico. 24

GSA Generalised Simulated Annealing. 39

I4 Industry 4.0. 53

IC Integrated Circuit. 46

ID Index of Disagreement. 37

IoT Internet of Things. 20

ISA International Society of America. 43

ISO International Organization for Standardization. 30

LCL Lower Limit. 31

LDR Light Depent Resistor. 46

LED Light-Emitting Diode. 46

LPA Local Productive Arrangement. 17, 23

MAPE Mean Absolute Percentage Error. 37

MSE Mean Squared Error. 37

MVC Model-View-Controller. 17, 48

Reg_Intercept Intercept of the linear fit between $\hat{u}_{t,i}$ and u_t. 37

Reg_Slope Slope Coefficient of the linear fit between $\hat{u}_{t,i}$ and u_t. 37

RMS Root Mean Square. 47

SCADA Supervisory Control. 48

SENAI Serviço Nacional de Aprendizagem Industrial. 21

SESI Serviço Social da Indústria. 21

TBL Triple Bottom Line. 22

Theil Theil's U. 37

TQM Total Quality Management. 30

UCL Upper Limit. 31

URCA Universidade Regional do Cariri. 24

WiFi Wireless Fidelity. 43

WPOCID Wrong Prediction on Change of Direction. 37

WR^2 Indeterminacy Coefficient of the linear fit between $\hat{u}_{t,i}$ and u_t. 37

List of Symbols

$\bar{x}$	Sample average
χ^2	Chi square statistic
$^{\circ}C$	Celsius degree
θ	Bias
A	Ampers
a	Artificial Neuron output result
dB	Decibel
Hz	Hertz
kHz	Kilohertz
mA	Milliampers
MHz	Megahertz
Rx	Data receiver bus
s	Standard deviation
s^2	Sample variance
Tx	Data transmitter bus
V	Volts
w_{θ}	Bias weight
w_i	Artificial Neuron inputs weight
x_i	Artificial Neuron inputs values

Contents

1 Introduction

The sustainable development has emerged as a solution to the main contemporary issues of economic and social development in harmony with environmental promotion [1, 2]. However, it is claimed that such target is only reachable when operational solutions play the role in production systems in general [3, 4]. This idea is also advocated by the United States of America National Research Council [5]. They bring a synthesis of needs to be sustained (nature, life support, and community) and needs to be developed (people, economy, and society). In turn, Robert et al. [6] contextualise the productive systems as a common point to the several dimensions covered by the sustainability philosophy. Thus, the technological development can be paramount to remedy imbalances [7, 8].

The present work aims at providing a framework for enhancing the quality of manufacturing production systems in general, and conveyor belts in particular. In this way, software and hardware solutions are developed and customised to automatically monitor, collect, safeguard, and model time series that reflect the condition of the production system under study, in terms of a number of performance characteristics (e.g. on- and off-line electric consumption and temperature). A cheap and easy-to-use technology is able to compute quality control metrics as well as eventual alerts regarding the inadequate condition of the production system. The main hypothesis of the study is that the proposed automated control system will enhance the quality of the production system taken into account. This can be quantitatively verified after the taking decision stage conducted through the statistical analysis.

There are many works involving the use of Statistical Control Processes approach [9, 10, 11]. In general terms, one can find works based on cumulative sum (CUSUM) control charts [12, 13, 14, 15]. Also, some case studies can be considered [16, 17, 18, 19, 20, 21]. The usefulness of the proposed approach will be evaluated by means of a case study involving the Mineral Technology Center of Cariri (CTMC). The center is located at Nova Olinda city, Ceará, Brazil, and actuates in the mineral extraction branch. Its main product is the laminated limestone, composed by 95% of calcium oxide.

1.1 Objectives

General objective

To develop an automated system for recording, modelling, and controlling performance characteristics of conveyor belts production lines.

Specific objectives

(i) To study alternatives of sensing, capturing, storing, and modelling relevant characteristics of conveyor belts.

(ii) To install sensors and capture relevant data from the production system taken as case study, according to the findings from (i).

(iii) To develop a storage and recuperation system from the data obtained from (ii).

(iv) To statistically model data stored from (iii) for pattern learning and recognition, according to time series formalisms.

1.2 Justification

The scientific research focused on industrial development has rapidly evolved since the beginning of the XXI century. In the current information age, the interrelationship among themes such as Statistical Quality Control, Industrial Management, and Diagnosis has thus been paramount to the establishment of the named Industry 4.0. These three themes touch directly pillars for Sustainable Development of the organisations of the first, second, and third sectors. Considering that the 2030 agenda for Sustainable Development was established in sense of promoting a tripod equilibrium, each one of these three technological themes is related with one or more objectives of the agenda [22]. In this way, software development has reached a global level [23, 24, 25, 26], demanding the development of strategies for good practices of implementation [27, 28, 29]. On the other hand, hardware development has taken intensity with the introduction of concepts as Internet of Things (IoT) [30, 31] and Home Automation [32, 33]. In turn, statistical quality control has been important for the management of production systems in various aspects [34, 35].

In this scenario, the intention of providing a cheap and easy-to-use system that automatically collect, safeguard, and statistically manipulate data from industrial conveyor belts is fully justifiable. The solution is particularly attractive for countries in development, like Brazil, with emphasis on its Northeast region. To date, this region has been the scene of a broad economic transformation. In specific local areas, such as Cariri, at Ceará state, a number of companies in many fields (from food to durable goods industries) have been installed. Few units, however, employ standard practices of industry 4.0 [36]. The Industry 4.0 pillars presented by authors like Santos et al. [37], for instance, deal directly with the use of sensors as a basis for Big Data Analysis and Industrial IoT [38].

In Brazil, the named S System, a set of nine institutions related to professional categories, has been an important contributor to industrial and commercial development.

The system is based on the Article 149 of the Federal Constitution [39]. Specially, the S system involves institutions related to the National Industry Confederation (National Industrial Learning Service - SENAI [40, 41, 42] and Industry Social Service - SESI) [43, 44]. These institutions have facilitated the most varied range of industries in less developed places. In common to these production systems, one can highlight the presence of conveyor belts as well as elementary monitoring and control systems. Such phenomena seems to reflect the so-called Brazilian Growth Acceleration Program, which did include housing programs, education, health, and industrial development, among others [45]. Despite the remarkable benefits, the industrial sector still lacks cheap and easy-to-use technology solutions similar to the present one.

In this way, the proposed tool also aims to overcome the evident deficiencies arising from the lack of connection between scientific and technological knowledge, management, and diagnosis issues in researching and practical exercises of continuous improvement cycles. In fact, through a textual analysis from scientific production since XX century, it can be verified a separation among these three disciplines. For instance, it is possible to find works exclusively directed to anomalies detection [46], or management [47, 48, 49], or statistical control [50, 51, 52]. The present project claims that these works represent the existing polarisation among the aforementioned technological tripod dimensions. The work is thus an attempt to promote a better balance in this way, favouring a solid and faster local industrial sustainable development.

The rest of the work is organised as follows. In Chapter 2 it is presented the literature review with classical works that base the proposed framework and the tripod of sustainability (environmental, economic, and social dimensions). Chapter 3 presents how innovative technologies and methods presented in the literature review will be combined and applied to solve the perceived gaps with respect to the proposed approach. Chapter 4 brings a discussion of the state of the art with the works related to "Statistical Control", "Embedded Systems", and "Signal and Image Retrieval". A textual analysis was performed to summarise the interaction among these themes. The Chapter 5 shows obtained results. The Chapter 6 presents the Conclusion with the limitations and difficulties, and future works.

2 Literature Review

This chapter presents some of the main concepts regarding the basis for the development of the project, notably sustainable regional development, production control and management, and statistical control processes.

2.1 Sustainable Regional Development

The United Nations, through the report of the World Summit on Sustainable Development (Johannesburg Declaration) [53], emphasise the importance of strengthening the Triple Bottom Line (TBL), i.e. the harmonic and long-term interrelationship among economic, social, and environmental issues. Further, events like the United Nations Conference on Environment and Development - Rio 92 (The Earth Summit) have contributed to the establishment of The Global Goals For Sustainable Development, introduced in the Resolution adopted by the General Assembly on September 25th, 2015 [54]. In agreement, Sen [55] considers that the imbalance among these three dimensions are prejudicial, leading each one to become predatory with respect to each other. A classic example is when the economic development leads to the degrading exploitation of environmental resources, causing a general collapse in the long run. On the other hand, the harmony of characteristics of the TBL allow a "friendly process". Supported by the tripod idea, [56, 57, 58] accentuate the negative effects caused by the imbalance among these three aspects as the poverty increase, environmental degradation, and the violence growth.

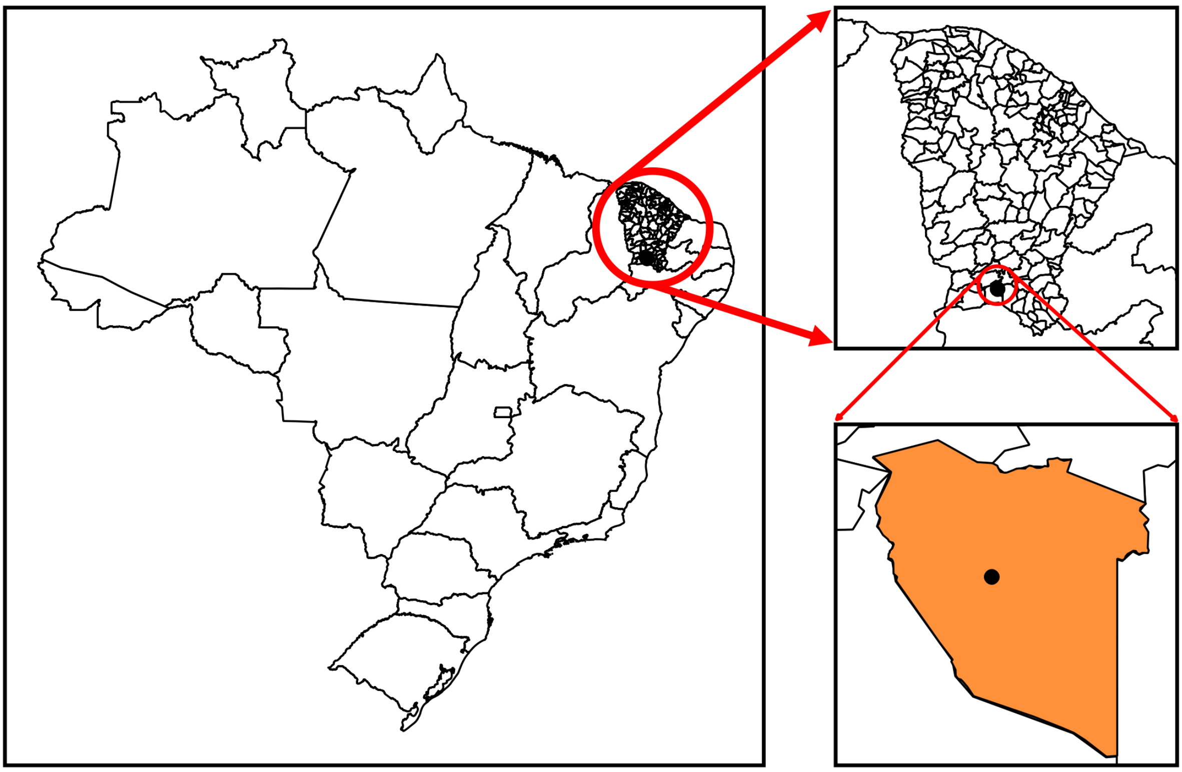
Furthermore, Cornel and Mirela [58] bring the need to treat the sustainable development at local level. In spite of conducting the local work from the perspective of another country or community, adjustments from the basic framework might be necessary. Buarque's approach to strategies for promoting sustainable local development [59] emphasises the identification and exploration of opportunities offered by the specific realities. Regarding Cariri, in Ceará State, much research has been done with respect to sustainability. From the environmental point of view, for instance, one can quote works focused on the Araripe National Forest and in their native species [60, 61, 62]. Economic aspects are mainly treated with emphasis on ecotourism and industrial activity [63, 64, 65]. In turn, the familiar agriculture and living at semiarid are the main subjects approached from the social point of view [66, 67, 68].

At all, Cariri involves the Paraíba, Pernambuco, and Ceará states. In Ceará, Cariri is a Microregion located at the Southern State Mesoregion, composed by cities like Barbalha, Crato, Jardim, Juazeiro do Norte, Missão Velha, Nova Olinda, Porteiras, and Santana do Cariri. The total territorial extension is 5460 km^2, with a population of 601817 inhabitants

[69]. Among the Cariri natural riches, Oliveira et al. [70] bring a wide variety of flora, emphasising medicinal application. The Cariri fauna has rare and exclusive species as Araripe Manakin (little soldier of Araripe) [71]. The fauna is also composed of a wide variety of lizards, snakes, amphibians, and others species [72, 73, 74, 75].

Specifically, Nova Olinda, the target city of study, was founded in April 14th, 1957. Its distance with respect to the State capital (Fortaleza) is 560 km. Figure 1 sketches the localisation of the Brazilian city. Its territorial extension is 284.404 km^2, with a population of approximately 15520 inhabitants [76]. The city is situated at the superior part of the named *Chapada do Araripe*, a geological formation [77]. In addition to paleontological riches, Nova Olinda is known for housing institutes of artistic development such as the *Fundação Casa Grande - Memorial do Homem Cariri*. It is also the birthplace of traditional culture artists such as Espedito Seleiro, the former artisan leather craftsman [65].

Figure 1 – Geographic localisation of Nova Olinda City, Ceará, Brazil. The city is situated at the metropolitan region of Cariri.



Source: The author - Elaborated on `QGis`

2.1.1 Local Productive Arrangement (LPA) of Mineral Extraction

The plateau where Nova Olinda city is located have a total area of 9.000 km^2 and is rich in rock formations of approximately 120 million years. Such resource is useful in various perspectives, including construction, handicraft, and museum collection formation. The products originated from the extraction is the limestone, used in finishing parts in civil construction, cutouts with fossils destined to exposition in museum or trading, limestone powder used in mixing for support structures, and the residues harnessed to handicraft activities [77]. With the goal to protect the geological heritage, the State Government, via Regional University of Cariri (URCA), has created the Geopark Araripe, in 2006. In fact, it was the former Geopark created in the South America. The entity is thus responsible for protecting and valuating geological, paleontological, historical, cultural, and ecological patrimony [78].

Anyway, the mineral extraction is the main industrial activity in Nova Olinda city and one of the most significant economic sources in the entire Cariri region. According to the CTMC [79], the exploration occurs for over 40 years. The activity involves around 270 extraction points and employs about 1500 people. Although not being the biggest Local Productive Arrangement (LPA) in the Metropolitan Region of Cariri [63], this sector has expressive importance in economic, social, and historic terms. Yet the sector has faced precarious development conditions, though under increasing external demand. In other words, no balance among social, economic, and environmental issues seems to exist due the unlawful exploration that occurs aggressively, among others. It must be highlighted that these LPA of mining have been implemented according to plans from Foundation to Support Scientific and Technological Development of Ceará (FUNCAP) and execution from CTMC in 2004, based on financing of R$499 500.00 by the part of FINEP [80].

A cooperative (COOPEDRAS) was then formalised to facilitate the regularisation of the associates to the National Department of Mineral Production (DPNM). The agency aggregates 33 companies. From the "Problems x Solutions" Matrix [81] that motivates the cooperative formalisation, it is possible to identify the bottlenecks in which the industrial technology proposed in this work can contributes, in terms of facilitating the appointed solution:

* Problem 3 - Lack of technical and socioeconomic information on activity in the region;
* Problem 6 - High level of loss in the mining and beneficiation stages;
* Problem 9 - Poor finished product quality.

The CTMC is a initiative of the Science, Technology, and Higher Education Secretary of Ceará State Government. The unity is built in an area of 8525m^2 and aims to facilitate the product beneficiation, ensure wastes harnessing, and lodge the cooperative. Among its activities, one can highlight professional training, production, and technological extension. Besides, bureaucratic questions of cooperative are handled. Figure 2 introduces images of the building facade.

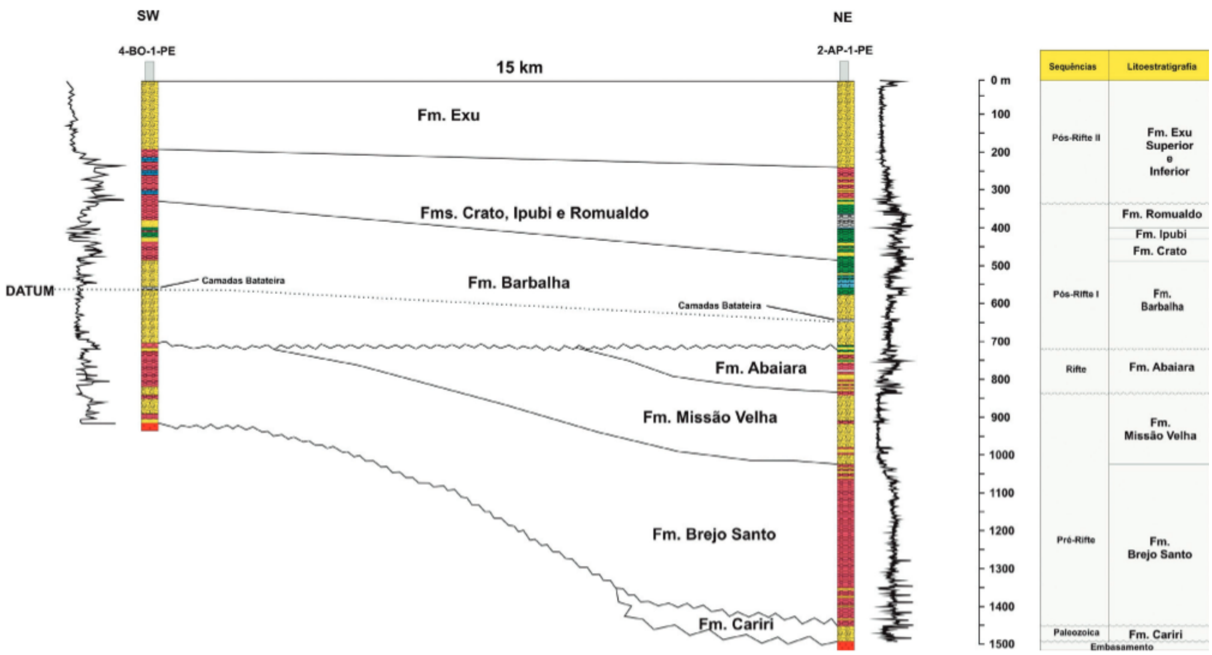
Figure 2 – Center for Mineral Technology of Cariri, located at Santana do Cariri, Ceará, Brazil.

Source: Centro de Tecnologia Mineral do Cariri - CTMC

2.1.2 Limestone production process

The laminated limestone production process consists of removing slabs by cutting. However, the layer of interest for extraction lies under other rock formations. Until the desired level is reached, newer layers of earth and rock formations are removed. Technically, the layer composed by laminated limestone are called Crato Member, a formation of about 113 million years. Over it one has Ipubi Member, Romualdo Member, and Exu formation, from oldest to newest [82]. Figure 3 brings an overview.

Figure 3 – Geological formation of the laminated limestone under study.



Source: [83]

Before the LPA formalisation, the stones were manually extracted. The activity was based only on a chisel and a hammer. The organisation through the LPA allowed the machines acquisition. By using electricity to move a blade, the stone separation became easier. However, the natural situation where the operator is submitted dramatically interferes in health and life quality. Most of the time the workers are under intense sun. The water used to cool cutting process keep in touch directly with the electric cable and the sound reaches detrimental levels [84]. It is believed that a review of the production process could allows better working conditions. Figure 4 illustrates one of the these situations.

Figure 4 – Mining front of limestone as executed currently. The high ambient temperature, landslides, and electric shock are the main risks underlying operators activity.

Source: The author

Since the beginning of the extraction process, the blocks opening are executed by using chisel and hammer on the piece width so as to separate in smaller pieces. In big pieces it is necessary a larger number of workers for simultaneously hitting, in order to avoid the part for breaking. The activity is equally unhealthy. Up to this point, it is unusual the use of personal protective equipment. The risk of impacts from the pieces with the body are also high and can cause serious problems. Figure 5 illustrates the manual separation of blocks. A shock of slabs with body parts can cause serious damage.

Figure 5 – Manual separation of blocks for limestone production.

Source: The author

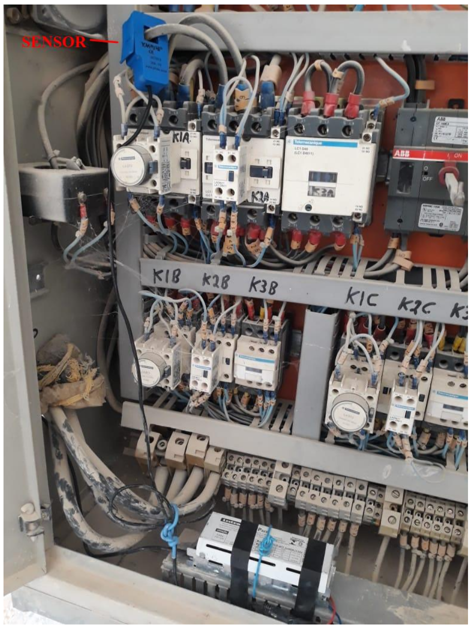
The limestone finishing is realised with more industrial methods than the other phases. Stones with more rustic characteristics are squared and polished. The most common application is in the finishing civil construction. Smaller scales are also manufactured to specific cement alloys. It is also very common to use in road construction bases. Deviating from the industrial sector, works are made of finer craftsmanship, mainly intended for tourists. Figure 6 introduces the limestone beneficiation production line.

Figure 6 – Limestone beneficiation production line. Usually machines to square and calibrate the stones are used. The machine (b) adjust stone thickness. The sensor of (c) register temperature and was installed on main calibrator engine. The sensor of (d) register temperature and was installed on main calibrator engine.

(a) Production line

(b) Calibrator

(c) Main engine



(d) Calibrator control panel

Source: The author

2.2 Statistical Control Process and Quality

Since 1776 the concept of production management has been shaped from authors like Adam Smith, considered the father of the modern economy [85]. These researchers have focused on elucidating how the production systems are organised [85]. More recently, authors like Deming have presented scientific contributions of Japanese senior executives in order to make the country a reference in innovative and high quality level products [86, 87, 88, 89]. In this way, works like "Out of the Crisis", in which 14 points to lead to the Total Quality are introduced, must be emphasised [90]. It was a precursor for the concept of Total Quality Management (TQM) [90]. These and the works that emerged during this period are essential to industrial management as well as the remaining administration areas.

Since their history beginning, humans have produced tools according to their specific needs. As stated by Shewhart and Deming [91], the need for larger-scale production arose and, as a result, the importance of standardisation among items. This historical context is still aligned with the modern concept of quality, under which “Quality is inversely proportional to variability” [92]. The variability regarding the acceptance or rejection patterns is verified through partial data collection (sampling) or full inspection. Though more expensive, the latter enhances the performance of the quality control system [93]. In special, control charts have been successfully applied in quality control exercises.

2.2.1 Quality management

Over three decades ago it was common and viable to consider quality as conformance to requirements in most cases [94]. The population growth and increased demand from several sectors over the years, however, resignified the concept of quality. It was approached to standard requirements, elaborated according to technical issues [95]. The consumers conformity was turned on conformism in the sense of adjusting individual preferences according to the majority preference or the manufacturer’s impositions [96]. The increased data flow in production systems is proportional to this evolution in standardisation.

Some of the Deming’s 14 points [90] directly match the benefits of automating data collection. For instance, the third point treats the reduction of inspection costs, that can be substituted by the automated data synthesis, modelling, and pattern recognition. The twelfth point addresses the welfare of the workers. The fifth point indicates that the quality can be increased with the production system improvement, besides reducing costs and wastes. The aggregation of these points naturally favour the TBL components. The Deming’s points are an important theoretical contribution and, despite being ignored by companies in some sense, they basis important normative or even legal provisions such as the ISO 9001 [95]. This international standard is dedicated to quality issues, basing audits to certify companies.

2.2.2 Control charts

Control charts involve the modelling of variables that reflect the performance of the production system, according to a set of parameters. Thus, one can monitor the production system and to infer eventual disturbances. In this way, auto-correlation, upper, and lower control limits are provided with respect to a given number of statistical metrics. The widely known control charts for the mean, for instance (Eq. 2.1 and 2.2), can make use of mean and standard deviation estimates (e.g. Eq. 2.3 and 2.4) and the sample size adopted for inspection (n) when computing the respective control limits [92]. These charts also are useful visualisation tools for dynamic monitoring of the control variables.

$$UCL = \bar{x} + q_{1-\alpha/2} \times \frac{s}{\sqrt{n}} \tag{2.1}$$

$$LCL = \bar{x} - q_{1-\alpha/2} \times \frac{s}{\sqrt{n}} \tag{2.2}$$

in which

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.3}$$

is the sample mean,

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{2.4}$$

is the sample standard deviation, and $q_{1-\alpha/2}$ is the quantile that reflects the $100 \times (1-\alpha)\%$ confidence level adopted by the decision maker when computing the control limits of the mean of the variable of interest. Thus, in a hypotheses testing framework, the system might be wrongly considered as out of control with a probability of α. It is yet possible that $\bar{x}$ and s be replaced by nominal values that reflect the planned (or target) mean and standard deviation of the variable, say μ and σ, respectively.

Another perspective for using control charts is to consider the sample variance (s^2 control chart).

$$UCL = \frac{\bar{s}^2}{n-1} \chi^2_{\alpha/2,n-1} \tag{2.5}$$

$$LCL = \frac{\bar{s}^2}{n-1} \chi^2_{1-(\alpha/2),n-1} \tag{2.6}$$

In this case, the average variance ($\bar{s}^2$) of previously considered samples is taken as the basis. The upper and lower limits (Eq. 2.5 and 2.6) are defined by $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-(\alpha/2),n-1}$ in terms of percentage $\alpha/2$ of the chi-square distribution with $n-1$ degrees of freedom.

It must be highlighted that both mean and variance control charts are usually designed to use only the most recent information regarding the control variable (e.g. the last sample observation), thus ignoring any information given by the entire sequence of points. Moreover, the Shewhart control charts are not the most suitable alternative to monitor small variations along the series. In order to address these drawback, the cumulative sum (CUSUM) control chart might be considered. In this way, one works on the statistic In this way, one works on the statistic

$$C_i = \sum_{j=1}^{i} (\hat{\theta}_j - \theta_0) = C_{i-1} + (\hat{\theta}_j - \theta_0), \tag{2.7}$$

in which $C_0 = 0$, $\hat{\theta}_j$ is the estimate of the parameter of interest θ according to the j^{th} sample obtained from the production system, and θ_0 is the target value considered for θ. The control limits of the CUSUM approach can be based on the following statistics

$$C_i^+ = max\{0, C_{i-1}^+ + \hat{\theta}_i - (\theta_0 + k)\}, \tag{2.8}$$

$$C_i^- = min\{0, C_{i-1}^- + \hat{\theta}_i - (\theta_0 - k)\}, \tag{2.9}$$

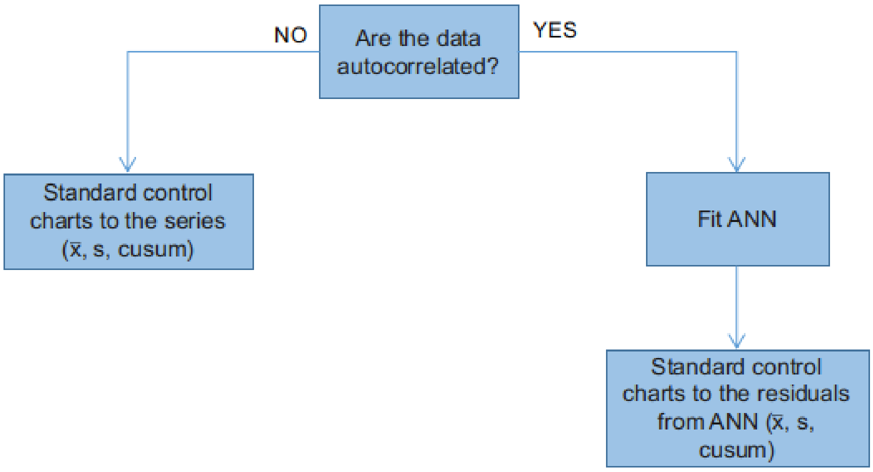
in which $C_0^+ = C_0^- = 0$ and k is usually called the reference (or the allowance) value. The constant k is a function of the distance between the target θ_0 and the out-of-control value of the parameter, say θ_1. Thus, the manager seeks to quickly detect distances greater than k [92]. In this way, if at least one, C_i^+ or $|C_i^-|$, exceeds a given threshold h then the process is considered out-of-control. It has been usual to adopt $k = 0.5$ and $h = 5$.

Anyway, control charts are made up of time series, a data set indexed in the time (Eq. 2.10), with or without autocorrelated (temporal dependence on successive residual values) pattern [97].

$$U_t = \{u_t \in R | t = 1, \ldots, n\}, n \in N \tag{2.10}$$

As stated by some authors [92, 98], in the cases in which the time series of the control variable does reflect an autocorrelated process, modelling alternatives might be taken into account, as seen in Figure 7. The idea is to consider, instead of the pattern of the control variable itself, the behaviour of the residuals provided by the fitted time series models when forecasting such variable. A number of time series modelling formalisms can be found in the literature. As follows, Artificial Neural Nets, one of the most promising frameworks, are introduced.

Figure 7 – Diagram for univariate control chart selection. CUSUM is indicated to non autocorrelated variables with small shift size and autocorrelated data. ARIMA is an alternative to the second case.



Font: Adapted from [92]

2.2.3 Autoregressive Integrated Moving Average

The architecture of Autoregressive Integrated Moving Average model (ARIMA) makes use of three order parameters (p, d, q). The order p is the number of autoregressive (AR) terms, while d is the number of non seasonal differences until reach a stationary (I), and q is the number of past forecasting errors considered (MA). Eq. 2.11 describes ARIMA model:

$$\hat{u}_t^d = \gamma_0 + \sum_{i \in I_u} \gamma_i^{ar} u_{t-t_i}^d + \sum_{j \in I_{\hat{u}}} \gamma_i^{ma} e_{t-t_j}^d, \tag{2.11}$$

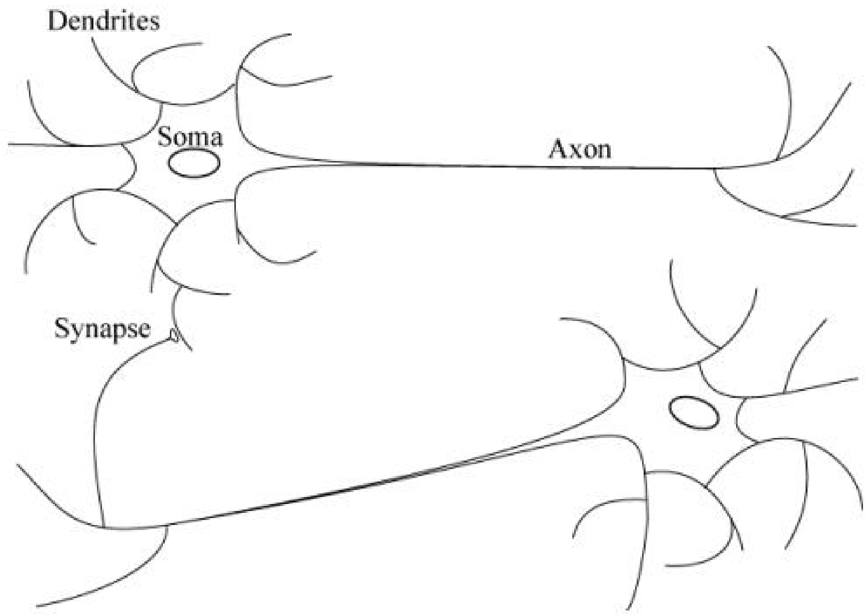
in which $\hat{u}_t^d$ is the estimated value of u_t^d, d is the differentiation necessary to turn the series stationary, γ_i^{ar} are the Autoregressive coefficients, γ_i^{ma} are the Moving Average coefficients, and $e_{t-t_j}^d$ are the residuals of the model [99]. One can also involve seasonal arguments in the ARIMA framework, leading to SARIMA$(p, d, q)(P, D, Q)[S]$, in which the triple (P, D, Q) respectively represent the seasonal order of autoregressive, integrated, and moving average coefficients whilst S implies the seasonality to be taken into account. Tools like the `forecast` package [100] of `R` software [101] contemplate alternatives to achieve near optimal SARIMA models in the face of a given time series.

2.2.4 Artificial Neural Networks

Some authors [102] describe Artificial Neural Networks (ANN) as a computational model that resembles the human brain by its parallel processing capability. ANN use simple processing units capable of mimicking the knowledge acquisition when subject to experimentation.

Neuroscience, the area of science that studies the nervous system, has inspired the formulation of the artificial neuron based on the shape of the nerve cell that is composed of a branching cell body (dendrites and axon). Dendrites are responsible for receiving nerve stimuli. Located at the end of the axon, the synapses connect to the dendrites of other neurons allowing the propagation of signals among them through an electrochemical reaction [103]. Figure 8 sketches these components.

Figure 8 – Biological Neuron.



Font: [104]

The Artificial Neuron developed by McCulloch and Pitts in 1943 is a mathematical model that simulates the nervous system plasticity and allows a linear classification when values combinations in entry exceeds a threshold. An ANN is thus formed by the connection of a range of artificial neurons [103]. Based on the flexibility enabled by ANNs, it is possible to teach patterns to the machine designed to process the captured data so that the same system fits different situations. Thus identification can range from level checking in a container with liquids such as bottles, ampules, and vials; to checking for an imperfection in painting or cutting a shoe material as well as to predict time series of interest.

The output result (u_t) depends on an activation function (Eq. 2.12) that is modelled according to the problem nature. The input variable of this function is a summation of products of all neuron inputs (say $p = (u_{t-1}, \ldots, u_{t-m})$) by the respective weights (Eq.

2.13).

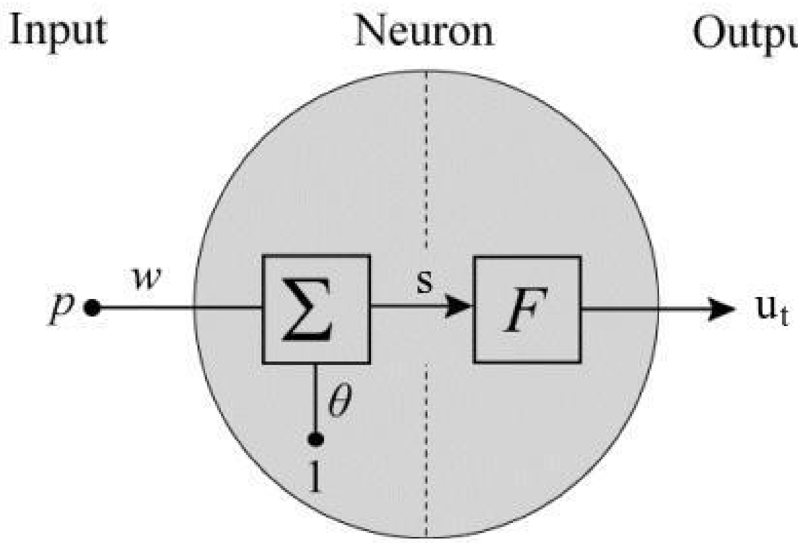
$$u_t = f(s), \tag{2.12}$$

in which

$$s = \sum_{i=0}^{m} w_i u_{t-i} + w_\theta \theta. \tag{2.13}$$

Figure 9 illustrates an artificial neuron.

Figure 9 – Artificial Neuron



Font: Adapted from [104]

There are some alternatives to activation function. Among them, one can mention the logistic function (Eq. 2.14)

$$f(s) = \frac{L}{1 + e^{-k(s-s_0)}}, \tag{2.14}$$

in which L is the max curve value, k is the slope of the curve, and s_0 is the value of s at the midpoint of the sigmoid curve; and the hyperbolic tangent (Eq. 2.15)

$$f(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}}. \tag{2.15}$$

Inadequate output values indicate that weights require adjustment. Among the learning paradigms, the supervised learning is based on response stimuli inserted by an external agent with expected results. The unsupervised learning is based on the stimuli received by the neural network itself, only [102]. Most commonly, backpropagation is the weight correction algorithm most usual. It is operated via Eq. 2.16.

$$w_i = w_i - \eta \frac{\partial (u_t - \hat{u}_t)}{\partial w_i}, \tag{2.16}$$

in which η is the learning rate.

Besides backpropagation, one can also mention the resilient backpropagation (rprop) [105] and the modified globally convergent version (grprop) [101]. The original rprop aims to improve the ANN training according to a returning weight approach, named rprop+ [106]. A new version of rprop has also been introduced, disregarding from the returning weights, named rprop- [106]. The grprop is based on the resilient backpropagation without weight backtracking and additionally modifies one learning rate, either the learning rate associated with the smallest absolute gradient (sag) or the smallest learning rate (slr) itself. The learning rates in the grprop algorithm are limited to predefined boundaries [101]. Tools like the `neuralnet` package [107] of `R` software [101] contemplate all of these ANN alternatives.

2.2.5 Minimal Variance Combiner

In order to aggregate ARIMA and ANN, the minimal variance approach (cMV) is taken into account, following [108]. Thus, from the supposed unbiased estimate of u_t based on the i^{th} model, say $\hat{u}_{t,i}$, $i = 1, 2, \ldots, k$, it is also supposed the joint distribution of the predictors might be a multivariate normal distribution:

$$g(\hat{u}_{t,1}, \hat{u}_{t,2}, \ldots, \hat{u}_{t,k} | \mathbf{u}_t, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(\hat{\mathbf{u}}_t - \mathbf{u}_t)^T \Sigma^{-1} (\hat{\mathbf{u}}_t - \mathbf{u}_t)\right), \tag{2.17}$$

in which $\mathbf{u}_t = \begin{pmatrix} \widehat{u}_{t,1} \\ \widehat{u}_{t,2} \\ \vdots \\ \widehat{u}_{t,k} \end{pmatrix}$, $\mathbf{u}_t = \begin{pmatrix} u_t \\ u_t \\ \vdots \\ u_t \end{pmatrix}$, and $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \sigma_{ij} & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}$ is the covariance matrix for the unbiased predictors. It is straightforward to show that the maximum likelihood unbiased estimate for u_t in the light of $\mathbf{y}_t$ is given by the linear combination

$$\widehat{u}_t^{cMV} = \sum_{i=1}^{k} w_i \cdot \hat{u}_{t,i} \tag{2.18}$$

in which

$$w_i = \frac{\sum_{j=1}^{k} a_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}}$$

and a_{ij} is the j^{th} element of the i^{th} row of the inverse matrix Σ^{-1}.

For instance, if $k = 2$ (the case of the present project),

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \cdot \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

in which $\rho = \sigma_{12}/\sigma_1\sigma_2$ is the correlation between the unbiased predictors $\hat{U}_{t,1}$ and $\hat{U}_{t,2}$. Thus, it becomes simple to obtain the maximum likelihood estimate of u_t in the light of $\hat{u}_{t,1}$ and $\hat{u}_{t,2}$:

$$\widehat{u}_t^{cMV} = w_1 \cdot \hat{u}_{t,1} + w_2 \cdot \hat{u}_{t,2} \tag{2.19}$$

where,

$$w_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \quad \text{and} \quad w_2 = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

In the case of unbiased and normally distributed predictors for u_t, this approach coincides with the minimal variance combiner.

2.2.6 Performance measures

With respect to the quality of the predictors, a number of measures can be taken into account. As follows, some performance metrics are presented. In this way, let $\hat{u}_t$ be the forecast for the target u_t and let N be the number of observations of the time series. There are a number of metrics for evaluating the discrepancies between u_t e $\hat{u}_t$, for $t = 1, 2, \ldots, N$. In this project, the following metrics are taken into account for evaluating the quality of the ARIMA, ANN, and cMV models: Mean Squared Error (MSE); Mean Absolute Percentage Error (MAPE); Average Relative Variance (ARV); Index of Disagreement (ID); Theil's U (Theil); Wrong Prediction on Change of Direction (WPOCID); Intercept of the linear fit between $\hat{u}_{t,i}$ and u_t (Reg_Intercept); Slope Coefficient of the linear fit between $\hat{u}_{t,i}$ and u_t (Reg_Slope); Indeterminacy Coefficient of the linear fit between $\hat{u}_{t,i}$ and u_t (WR^2) [108, 109, 110]; and an Aggregate Performance Metric (APM). (Eq. 2.20-2.28). The greater the value of a given metric, the worse the model is.

The MSE brings compromise with both accuracy and efficiency [111]:

$$MSE = \frac{\sum_{t=1}^{N}(u_t - \hat{u}_t)^2}{N}, \tag{2.20}$$

In turn, MAPE measures the model accuracy is a relative value:

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{u_t - \hat{u}_t}{u_t}.\right| \tag{2.21}$$

In turn, ARV compares the performance of the predictor with the one of the simple mean of the past values of the series.

$$ARV = \frac{\sum_{t=1}^{N}(u_t - \hat{u}_t)^2}{\sum_{t=1}^{N}(\hat{u}_t - \bar{u}_t)^2}, \tag{2.22}$$

in which $\bar{u}_t = \frac{\sum_{i=1}^{t} u_i}{t}$.

The ID disregards from the measure unit, with values in the interval $[0, 1]$.

$$ID = \frac{\sum_{t=1}^{n}(\hat{u}_t - u_t)^2}{\sum_{t=1}^{n}(|\hat{u}_t - \bar{u}_t| + |u_t - \bar{u}_t|)^2}. \tag{2.23}$$

Theil'U compares the performance of the predictor with the one of the Random Walk model (in which u_t is inferred by u_{t-1}).

$$Theil = \frac{\sum_{t=2}^{N}(u_t - \hat{u}_t)^2}{\sum_{t=2}^{N}(u_t - u_{t-1})^2}. \tag{2.24}$$

Further, WPOCID measures the model quality in forecasting the tendency of the target time series.

$$WPOCID = 1 - \frac{\sum_{t=2}^{N} D_t}{N-1}, \tag{2.25}$$

$$D_t = \begin{cases} 1, & \text{if} \quad (u_t - u_{t-1})(\hat{u}_t - \hat{u}_{t-1}) \geq 0 \\ 0, & \text{if} \quad (u_t - u_{t-1})(\hat{u}_t - \hat{u}_{t-1}) < 0. \end{cases} \tag{2.26}$$

In turn, $\mathrm{WR}^2 = 1 - R^2$, as well as Reg_Intercept and Reg_Slope are related to the linear model adjusted to the pairs u_t e $\hat{u}_t$, via minimal squared estimation. In this way, on can consider the general equation $u_t = Reg\_Intercept + Reg\_Slope \cdot \hat{u}_t$. Thus, Reg_Intercept and Reg_Slope coefficients represent the additive and multiplicative errors of the forecasts $\hat{u}_t$ of u_t, respectively. In this case, there is a constant error Reg_Intercept, independent from the forecast, and a proportional error Reg_Slope related to the prediction. In turn, R^2, the determination coefficient, reflects the performance of the model in capturing the variability of the time series [112]. R^2 is defined as

$$R^2 = \frac{\sum_{t=1}^{n}(u_t - \hat{u}_t)^2}{\sum_{t=1}^{n}(u_t - \bar{u})^2}, \tag{2.27}$$

in which $\bar{u}$ is the average of the observed series. Thus, an ideal predictor would present $\mathrm{WR}^2 = 0$, $Reg\_Intercept = 0$ and $Reg\_Slope = 1$, leading to, $u_t = 0 + 1 \cdot \hat{u}_t$.

In order to provide a general analysis of these metrics, it is proposed the aggregate performance metric:

$$APM = \frac{\sum_{i=1}^{m}(NMetric_i)}{m}, \tag{2.28}$$

in which $NMetric_i$ is the $metric_i$, from the aforementioned ones, normalised according to their values regarding the models taken into account (e.g. ARIMA, ANN, and cMV) and m is the number of metrics ($m = 8$ in the present project, reflecting MSE, MAPE, ARV, ID, Theil'U, WPOCID, |Reg_Intercept| and ||Reg_Slope|-1|). APM is based on the reasoning that the near to zero the value of APM is, the better the model is. Thus, for APM, $|Reg\_Intercept|$ and $||Reg\_Slope| - 1|$ are adopted instead of $Reg\_Intercept$ and Reg_Slope. Thus, APM is the simple average of the normalised version of the previous mentioned metrics, according to

$$NMetric_i = \frac{Metric_i - min_i}{max_i - min_i}, \tag{2.29}$$

in which min_i and max_i are, respectively, the observed minimal and maximal values of $Metric_i$ among the adjusted models under study (here, ARIMA, ANN, and cMV).

2.2.6.1 Information criterion

Further, information criteria are used in order to assist the selection of single and combined forecasting models. Here, Bayesian Information Criterion (BIC) will be used (Eq. 2.30) [113].

$$BIC = -2\log(ML) + k \cdot \log(N), \tag{2.30}$$

in which k is the number of parameters of the model, ML is the value of the maximum likelihood function underlying the residuals of the predictor in the face of the corresponding probability density function adjustment, and N is the sample size taken into account. In the present project, BIC is considered for optimising the ANN modelling. Specifically, BIC is adopted as the fitness function of an probabilistic optimisation method (e.g. Generalised Simulated Annealing - GSA), introduced as follows.

2.2.7 ANN-oriented simulated annealing algorithm

The simulated annealing process starts with the system being optimised at a high effective temperature, then lowering the temperature by slow stages until the system "freezes" and no further changes occur [114]. In mechanical statistics, such systems usually represent a collection of atoms physically disposed in the space, in which the probability of rearrangement of these atoms is proportional to the temperature imposed to the system. Thus, at a particular low temperature, the atom reconfiguration stops and the system reaches its equilibrium state. At this stage, the resulting atom configuration achieves a near-optimal solution in terms of minimisation of the thermodynamic energy of the system. Theoretically, as time runs to infinity, simulated annealing methods enable, with probability one, the identification of global optimal solutions [115]. Further, these algorithms may deal with (un)constrained optimisation problems, with qualitative and quantitative decision variables, and do not require restrictive mathematical conditions (e.g. continuously differentiable equations).

To solve general optimisation problems, the atoms represent a set of n_A decision variables underlying the system, say $\mathbf{p} = (p_1, \ldots, p_z, \ldots, n_A)$, and the system performance (energy) is mathematically described as a function of a configuration (instance) of $\mathbf{p}$, say $\mathbf{p}_{instance}$. In the herein addressed problem of ANN modelling, the system is the ANN model, the energy function to be minimised is the BIC of the model residuals in the face of a training phase of the series $\mathbf{u}$, say $\hat{e}_{ANN}(\mathbf{p})$, and the system configuration $\mathbf{p}$ involves the parameters of the ANN framework. Thus, BIC is considered for computing the fitness of a given solution ($\hat{e}_{ANN}(\mathbf{p})$) during SA-based optimisation. For ANN, $\mathbf{p}$ is given by:

- pAR - autoregressive order;
- $pARS$ - seasonal autoregressive order;
- $seasonalty$ - seasonality;
- $nNodes$ - # nodes of the ANN intermediate layer;
- $learningAlgorithm$ - learning algorithm of the ANN modelling phase (e.g. backprop, rprop+, rprop-, sag, slr);
- $actiovationFunction$ - activation function of the ANN modelling phase (e.g. logistic, tanh);

In general terms, the ANN-oriented simulated annealing works as sketched in Algorithm 1. A major difficulty may arise when determining algorithms for computing (i) the perturbation function $fp(\cdot)$ of the current instance of $\mathbf{p}$, $\mathbf{p}_{current}$, through iterations (step 7), (ii) the moving probability $fm(\cdot)$ from $\mathbf{p}_{current}$ to a new candidate configuration $\mathbf{p}_{new}$ (step 20), and (iii) the freezing function $ft(\cdot)$ which operates at the current temperature T (step 25). If no sophistication is introduced in these steps, simulated annealing may lose in computational efficiency and reduce the chance of reaching the global optimal solution for $\mathbf{p}$ in finite time. In the so-called classical SA [115] of Geman and Geman [116], $fp(\mathbf{p})$ sequentially updates each of the elements of $\mathbf{p}$ according to a conditional Gaussian distribution, whilst $fm(\Delta, T) = \exp\left(\frac{\Delta}{T}\right)$ due to the Boltzmann (Gibbs) distribution, in which Δ is the resulting change in the energy (performance) of the system, and $ft(T, q, time) = \frac{q}{log(1+time)}$, in which the pair $(q, time)$ represents a positive constant and the simulation iteration [114], in this order. In the present paper, $\Delta = \hat{e}_{ANN}(\mathbf{p}_{new}) - \hat{e}_{ANN}(\mathbf{p}_{current})$. The only necessary and sufficient condition for having probability one of reaching a global minimum via classical simulated annealing is that $ft(\cdot)$ decreases logarithmically over time [116]. This pointwise convergence property makes the algorithm a powerful, though simple, probabilistic optimisation method.

In order to enhance SA performance in terms of both time efficiency and statistical accuracy, some authors have dedicated attention to provide alternatives for $fm(\cdot)$, $fp(\cdot)$, and $ft(\cdot)$, leading to the GSA [115].

Algoritmo 1: General Simulated Annealing Algorithm for ANN modelling optimisation (adapted from SA literature [114, 115, 117])

. **Input:**
- $\mathbf{u}$ /* sample of N observation of the time series */
- T_0 /* initial temperature for SA */
- T_{min} /* frozen temperature for SA */
- $\mathbf{P}_0$ /* initial configuration for the ANN models */
- n_m /* number of moves per temperature level */
- ϵ_{sc} /* threshold for a significant change (reduction) in the system energy */
- m_{nnc} /* maximum number of non-significant changes per temperature level */
- $fm(\cdot)$ /* function to compute the moving probability */
- $fp(\cdot)$ /* function to compute a perturbation in a given configuration (solution) */
- $ft(\cdot)$ /* function to compute a decreasing in the current temperature */

Output:
- $\mathbf{P}_{min}$ /* near-optimal configuration for the ANN model */

$T = T_0$; $\mathbf{P}_{current} = \mathbf{P}_0$; $n_{nsc} = 0$; $time = 0$;
while $T > T_{min}$ *and* $n_{nsc} < m_{nnc}$/* system not yet frozen if its temperature (T) has not reached the minimal and the number of non-significant energy changes in $T(n_{nsc})$ has been acceptable */
do
    $i = 0$; $n_{nsc} = 0$
    while $i \leq n_m$ **do**
        Let $\mathbf{P}_{new} = fp(\mathbf{P}_{current})$; /* $\mathbf{P}_{new}$ is a perturbation of $\mathbf{P}_{current}$ */
        Let $\Delta = \hat{e}_{ANN}(\mathbf{P}_{new}) - \hat{e}_{ANN}(\mathbf{P}_{current})$; /* Δ is the resulting change in the energy (it depends on $\mathbf{O}$) */
        if $\Delta > -\epsilon_{sc}$ **then**
            $n_{nsc} = n_{nsc} + 1$/* significant change in energy is verified */
        else if $\Delta < 0$/* certain downhill move */
        then
            Set
            $\mathbf{P}_{min} = \mathbf{P}_{new}$, /* $\mathbf{P}_{new}$ is better than $\mathbf{P}_{current}$ (it minimises $\hat{e}_{ANN}$ */
            $\mathbf{P}_{current} = \mathbf{P}_{new}$,
            $i = i + 1$
        else if $\Delta \geq 0$ /* possible uphill move */
        then
            Let u be a random number ; /* $u \in [0.0, 1.0]$ */
            Let $p_{move} = min(1, fm(\mathbf{P}_{current}, \mathbf{P}_{new}))$; /* p_{move} is computed via $fm(\cdot)$ */
            if $p_{move} > u$ **then**
                Set
                $\mathbf{P}_{current} = \mathbf{P}_{new}$; /* avoid local optimal solutions */
                $i = i + 1$;
    Set $time = time + 1$ and decrease T via $ft(T)$; /* the temperature is decremented via $ft(\cdot)$ */
Return $\mathbf{P}_{min}$; /* near-optimal solution according to the error rate metric */

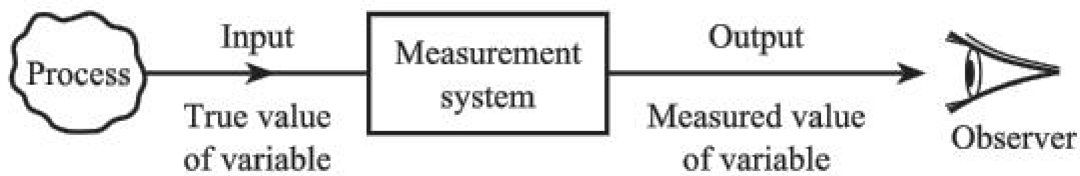
2.3 Industrial Instrumentation

One can find a number of industrial automation applications in the literature since XX century. However, one of the oldest automatic control system, the Heron's regulator, dates from 10 BC to 80 BC. The Heron's facility was a mechanism able to open a temple doors by using the heat of the flame burning in an altar. In turn, Dorf and Bishop [118] report a miscellaneous of applications employed in the period between 300 AC and 1 AC, that used floating regulators to fluid level control. Anyway, Ogata [119] mentions a device developed to control the velocity of a steam machine in XVIII century as the first automatic system with significant importance. More generally, Sharma [120] lists the automation technologies evolution emphasising the pneumatic systems used between the decades 1940 and 1960 as well as analogical systems between 1960 and 2000, personal digital systems between 1980 and 1990, and open digital systems since 2000.

Groover [121] defines automation as "a technology in which a process or procedure is performed without human assistance". He also highlights that advanced automation systems exert, among other functions, the "error detection and signal retrieving", without any physical intervention execution in some cases. This is common in situations when instrumentation allows the automatic collection of variables like temperature, pressure, velocity, volume, and flow rate, among others, offering support to decision making.

The use of instruments is thus essential for monitoring and saving data from control variables. In this way, the process constantly generates data that will serve as input to a measurement system, connecting the beholder to measured objects [122]. With the utilisation of cameras and sensors connected to a computer, for example, one can study the characteristics of interest and implement control systems. The human work is yet necessary, mainly for observing and intervening against disturbances identified by the machine. Figure 10 illustrates this relationship.

Figure 10 – Measurement Systems. The used tools in this process should present values to observer with smallest possible error.



Source: [122]

To Dunn [123] the instrumentation is the basis of automatic process control and an important framework for the implementation of computational systems. These authors argue that these tools can be comprised of simple residential as well as complex industrial systems, evolving in such a way that the techniques used today were not possible in the past, as that in the future will have a higher level of accuracy, precision, and performance, only possible with more complex controllers than the ones that exist today.

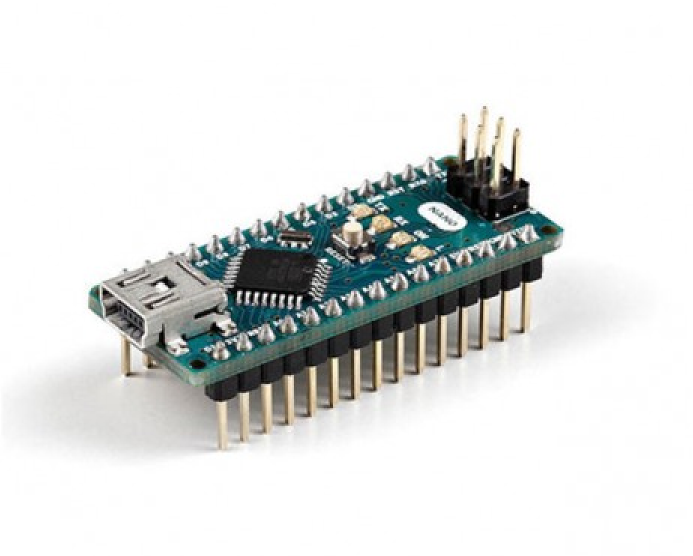
Particularly, cameras are powerful tools from an instrumentation standpoint due to their adaptability to capturing diverse information for later digital processing. Placko [124] states that there is no universal classification of instruments, though some countries accept the ISA5.1 [125] standard that makes a succinct differentiation, which can highlight the primary instruments in which the cameras are framed, which maintain a direct contact with what is being observed. Secondary instruments handle the signal from the primary element, such as the computer. Digital image processing, although a technique originally applicable in other scenarios, is an excellent artefact for industrial instrumentation and is commonly linked to the use of supervisory systems.

2.3.1 Microcontrollers

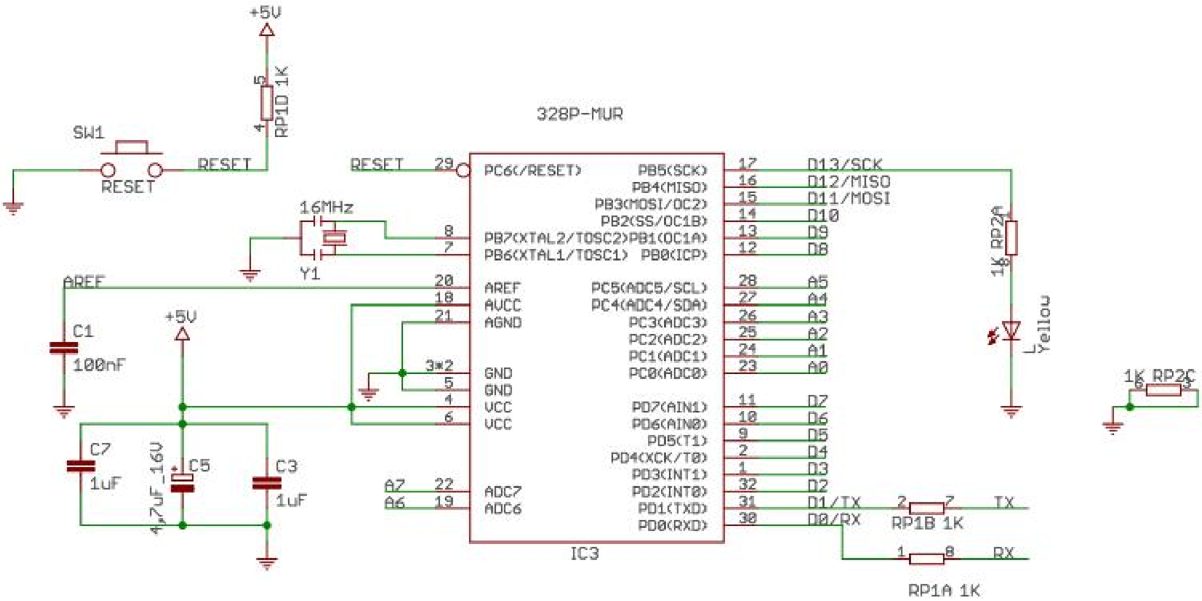
Microcontrollers have all the basic components of a computer (memory, Central Processing Unit and Input/Output bus) in an only chip. This is what differentiates microcontrollers and microprocessors [126].

The brand Arduino, specifically, has a increasing participation in most diverse use to embedded systems around the world. The facility on the hardware assembly is the main reason of technology diffusion. In addition, it is one of the few companies that joined to the open hardware concept. Arduino Nano (Figure 11(a)) is a viable option to applications in which the limited space is a condition. Though being a small board, the Arduino Nano has practically the same applicability of bigger boards, as Arduino Uno. There is no native WiFi connectivity, but has varied amount of pins (Figure 11(b)). Among the pins, there are eight analog-digital options (A0 to A7, pins 19, and 22 to 28) that can be connected to a sensor. Two pins (30 and 31) allow serial communication (RS232). Among other characteristics, the Nano version is powered in a voltage range from 7V to 12V. The operating voltage is 5V, obtained by using voltage regulators. The processor operates with a clock speed of 16MHz. Furthermore, the power consumption is 19mA, only. A pin sends data (Tx - pin 31) and other receives data (Rx - pin 38). These pins are used to send data to other boards with native WiFi connection (like ESP8266) to register the online database [127].

Figure 11 – Arduino Nano. The most popular open source electronic prototyping platform. It allows a simplified form of embedded system development.



(a) Arduino Nano



(b) Arduino Nano Schematic

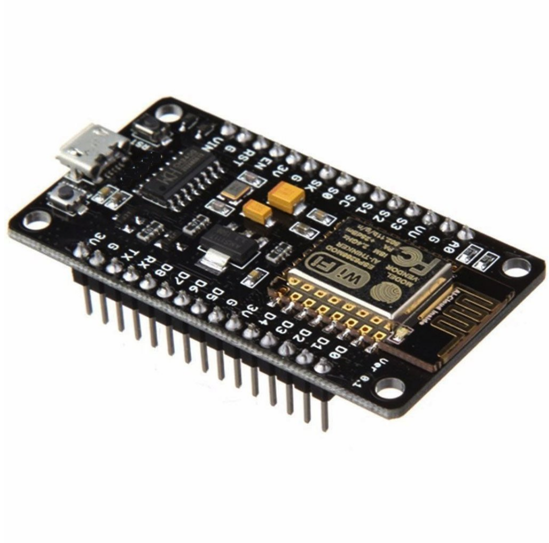
Source: [127]

Other brands, like Espressif Systems, are based on Arduino. The ESP8266 module, adapted with a NodeMCU version (Figure 12(a)), allows connectivity via WiFi. Like Nano Version, the ESP8266 is advantageous by the size. Adaptations allow receiving the data of an Arduino Nano via serial communication.

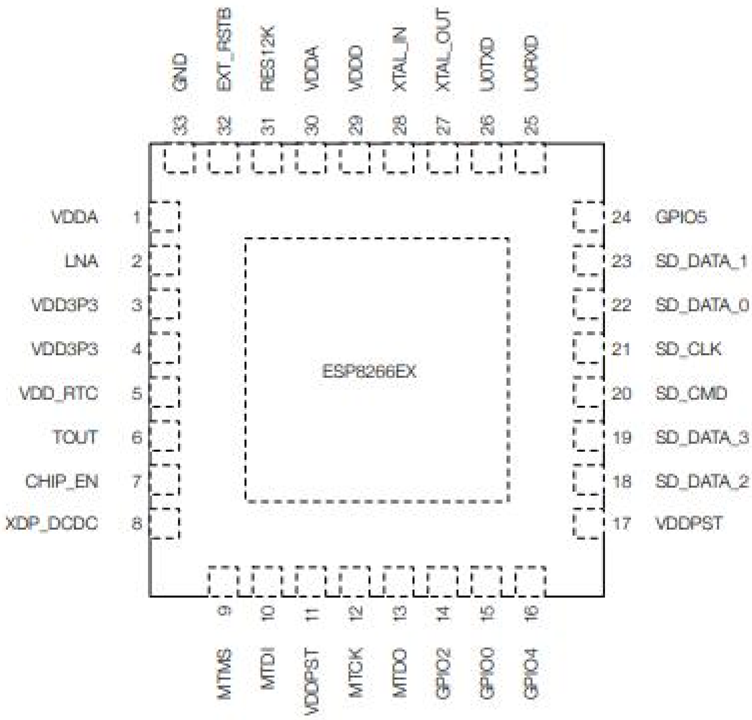
Among the pins, it is available only one analogical-digital pin (TOUT, pin 6) which can be connected to a single sensor. Two pins (25 and 26) that use serial communication

(RS232) are available (Figure 12(b)). Further, the NodeMCU is powered in a voltage range from 4.5V to 9V. The operating voltage is 3.3V to digital pins and 1V to analogical pin that is obtained by using voltage regulators. The Wireless pattern is the 802.11 b/g/n. In turn, the power consumption is 200mA, only.

Figure 12 – NodeMCU/ESP8266. Similar to Arduino, the board has a simplified handling. The WiFi connection is your main attribute.



(a) NodeMCU



(b) ESP8266 Schematic

Source: [128]

2.3.2 Sensors

Technological sensors are compared with the human senses when stimulated by some physical phenomena. Specifically, sensors allow one to monitor physical variables related to a given machine or component. They are classified by the data source (Proprioceptive or Exteroceptive) or with respect to the capacity of reading signals (Passive or Active) [129]. A synthesis of this classification is presented in Table 1.

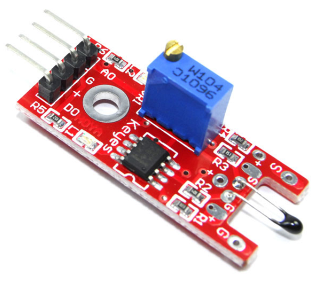
Table 1 – Sensors classification modes. A single sensor can be framed in more than one classification (e.g. a thermopar is exteroceptive passive).

Type	**Examples**	**Applications**
Proprioceptive	Magnetics, signal generator, voltmeter	Angle, motor speed and battery voltage measurement
Exteroceptive	Light Dependet Resistor (LDR), microphone, ultrassom	Luminosity, sound amplitud, distance measurement
Passive	Thermocouple	Temperature measurement
Active	Infra-red	Presence detection

Source: [129]

Depending on the sensor, a mixed classification can be given. The sensors of Figure 13, (a) and (b) are exteroceptive passive. They do not generate the own read signals and only receives signals generated by the external environment. It generates and receives its own signal, internally. The sensor of Figure 13 (a) is not specifically designed to industrial application. In fact, they are Arduino modules to prototyping. They have two Light-Emitting Diodes (LEDs): one to power indication and other to indicate the level of the measured variable. They use a LM393 Integrated Circuit (IC). The IC has two independent precision voltage comparators [130].

Figure 13 – Sensors employed to basic physical variables meditations.



(a) Temperature

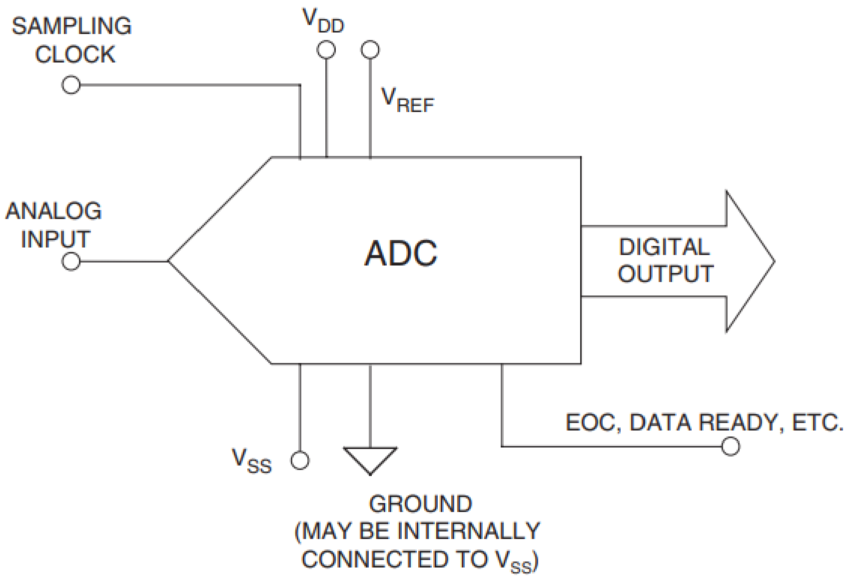
(b) Electric Current

The sensor of Figure 13 (a) (KY-028) uses a NTC Thermistor. The thermistor sends to LM393 a voltage proportional to its conductivity that varies with the temperature. This is possible due to powdered metal oxides in its composition. This materials present conductivity volatility when subjected to temperature variations. The temperature measurement ranges from -50°C to +125°C [131].

The sensor of Figure 13 (b) (SCT 013) reads the electromagnetic field around the wire in which it wraps. To its use it is not required to cut off the measured wire. The set works as a simple transformer. The electromagnetic field of measured wire works as a primary side. A transformer in its internal frame works as a secondary side. The reference brings a suffix that indicates the max Root Mean Square (RMS) input current. Commercially one can find models with readability from 0A to 5A and from 0A to 100A [132].

Usually, sensors collect analogical data, disabling a direct treatment via personal computers, for instance. In order to overcome this challenge, electronic structures are considered for performing a conversion to digital data. These structures are also named Analogical Digital Converters (ADCs). Figure 14 summarises the structure of an ADC. Basically, sensors release a quantity of energy depending on the intensity presented by the measured physical phenomenon. The voltage received by the analogical input is compared with a Reference Voltage (V_{REF}). The resolution depends on the quantity of bits of the ADC. An 8 bit resolution ADC is capable of capturing 256 signal levels, while a 16 bits ADC can handle 65536 levels. This conversion process is called discretisation [133].

Figure 14 – Basic ADC diagram. All continuous variables measured by sensors are discretised in levels according to the resolution of the converter.



Source: [133]

2.3.3 Supervisory systems

Supervisory systems are useful for obtaining data in controlled environments. They can be faced as a primary solution for automating the process of gathering information from a production line. The Supervisory Control and Data Acquisition (SCADA) refers to systems able to monitoring and controlling process by using input and output devices, Programmable Logic Controllers or Microcontrollers. Their scope with respect to the process can be total or partial.

Thus, SCADA software can be supported by several modular subsystems dedicated to a number of functions, allowing the adaptation to specific circumstances and meeting the demands of different environments. This paradigm makes the classification of automated systems intelligible as to their adaptability. Flexible systems are feasible in industries that work with seasonally varying product standards, as opposed to fixed systems that are only set up in a dedicated way. In both cases, a good level of performance in the verification of standards is a contributing factor to achieve a good level of product quality.

2.4 Software Engineering

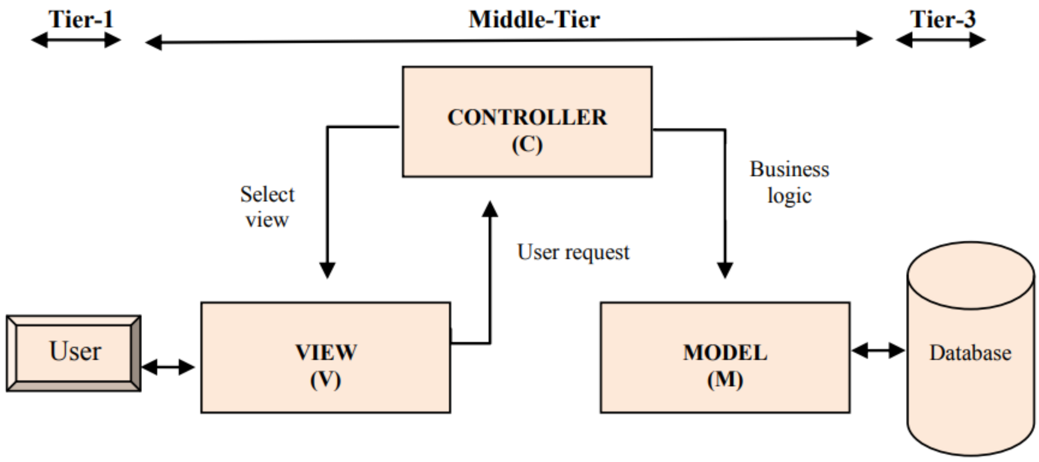
2.4.1 Model-View-Controller (MVC architecture)

The MVC is a pattern to software development created beginning in 1979 by Trygve Reenskaug at Xerox Palo Alto Research Center (Xerox PARC) [134]. It is proposed the software development scope division in three layers.

The layer View consists of structuring the user interface to facilitate his interaction with the system. This level allows accessibility to data entry or print. At this point, it does not matter what methods are used for data processing. When interpreted for user actions, controller forwards them to the Model layer. It also interprets the modelling layer response to discern what to show to the user. The Modelling layer involves the database management by performing creation, reading, updating, and deleting (CRUD [135]) operations [136].

Among the advantages of this paradigm, one has the ease of code reuse. It can also be pointed the applicability independent from the adopted language. The possibility of files organisation is potentiated by the folder division by architecture layers [137]. Figure 15 summarises the MVC architecture.

Figure 15 – MVC architecture block diagram. The user interacts with the system through the View that sends requisitions to Controller to manipulates the database with the Model.



Source: [137]

2.4.2 Database normalisation

A normalisation process is required to avoid inconsistency and data redundancy. Basically the normalisation occurs in levels where the third level is a base to Boyce–Codd Normal Form (BCNF). The BCNF is the desired level to modelling databases.

The first normal form consists of attributes separation. It is common the application of the first normal form in the case of storage of addresses, that are composed by street, number, neighbourhood, city, state, and so on. Each one of these attributes should be allocated in a specific column. The same is applied to multivalorated attributes, as telephone, for personnel studies.

The second normal form consists in partial functional dependencies elimination. It is common the application of the second normal form in which a non primary key attribute does not fully depends on the primary key. An example is when operators , one can cite characteristics stored in the same table where machines information are stored. In this case, it should be created a table to each one of the operators and machines.

The BCNF consists of turning all candidate keys in primary keys. It is common the application of the BCNF when a key that could be a primary key is not used as such. An example is when a tag of a machine is not used as a primary key. In this case, the tag of the machine should become a primary key [138].

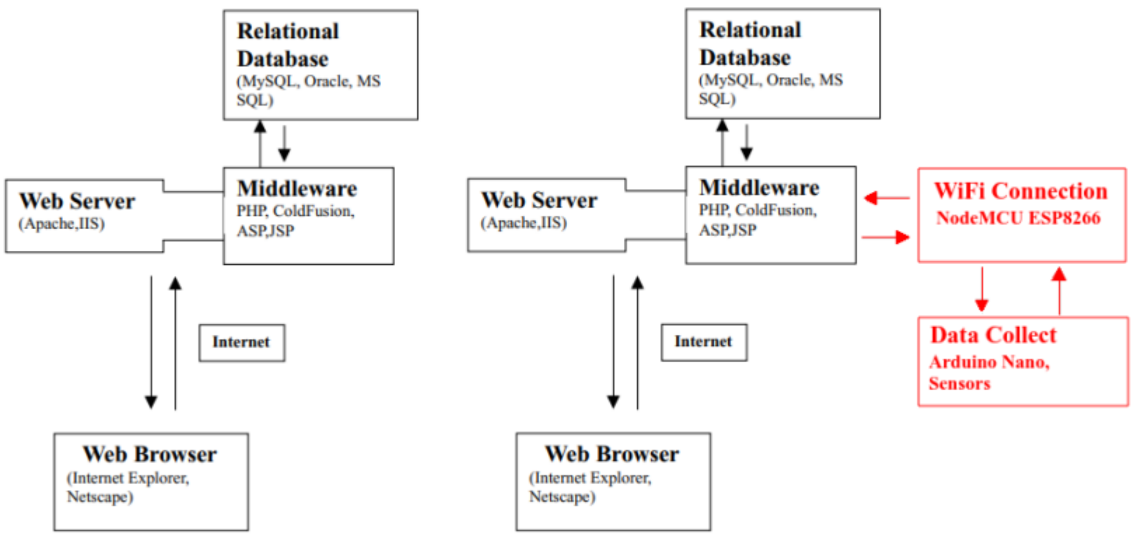
2.4.3 PHP, MySQL, and PHPMyAdmin

PHP is a acronym for Hypertext Preprocessor. It is an open source script language that can be merged with Hypertext Markup Language (HTML). The language has all usual mathematical functions of programming languages. It involves data input and output commands. Specially, PHP is a server side language. Thus, it allows the user to see only an equivalent HTML through Web Browser (Figure 16(a)). It promotes security in the cases that require the use of keywords to database connections, or others situations that contains confidential information. The language can also execute SQL queries directly, merging together with HTML and PHP codes [139, 140]. The current version of PHP is PHP7 [141].

MySQL is an Open Source Database tool. It makes possible to manage relational tables. The data are disposed in different tables depending on their class characteristics. These tables are interconnected, matching their respective indexes to aggregate characteristics in specific queries. It allows one to decrease the system operation cost for saving e manipulating data. With SQL (Structured Query Language), it is possible to execute operations as insertion, deletion, update, and selection of data [142].

PHPMyAdmin is the most popular Database Management System (DBMS). It allows to model databases through SQL commands or relational diagram drawing. Is is also possible to import and export ".sql" files from/to others DBMSs [140]. Figure 16 brings a map of the adopted architecture for web applications.

Figure 16 – Architecture of Web Applications. Eventually, other types systems can be integrated to core technologies of a web application.



(a) Architecture of Web Applications (b) Basic Architecture of Web Applications integrating the Arduino and NodeMCU (adapted from (a))

Source: [143]

3 Methodology

An exploratory, descriptive, and explicative research have been developed focused on the activity of laminated limestone production. Therefore, important variables inherent to the process and how their variability behave have been automatically registered (e.g. temperature and electrical consumption). Measurements have been made mainly by using sensors allowing the construction of time series and the respective control charts for the CUSUM statistics of the variables of interest. In this way, turned-off and turned-on phases of the machinery have been studied.

The architecture of the created time series database is summarised in the Appendix B.1. The Table "registers" was directed to the sensors detected values and is composed of continuous quantitative variables with your respective date and time. Particularly, in the explicative level, control charts for the CUSUM of the residuals of time series cMV (based on ARIMA and GSA-based ANN single predictors) models fitted to the temperature and energy consumption time series were learned from data collected via sensors installed in the extent of the production system. The statistical layer of the software has been implemented in `R` language. Thus, the temperature of the main calibrator (Figure 6(b)) engine has been recorded per minute since August 30th, 2019 until September 26th, 2019. The measure unity is Celsius Degree (°C). A total of 35729 observations were recorded for this variable. In turn, the electrical energy consumption of the same calibrator has been collected since September 9th, 2019 until January 10th, 2020, in time periods of five minutes (aggregating five values of the sensed quantities every minute before storage in a database) per sensor data capture. The measure unity is Amper (A). A total of 20911 observations were recorded for this variable. Both variables were aggregated in total time ranges of ten minutes. To this, the temperature was grouped by every ten registers stored in the database and the energy consumption was grouped by every two registers.

Therefore, a multiparadigm software have been developed for modelling the stored information based on the combination of Web Language (`PHP`) and computational statistics language (`R`) capable of statistically modelling the variables of interest. The data have been stored in a `MySQL` database. The server that shelts the physical memory is located at Federal University of Cariri. Table 2 relates the languages and equipment that have been utilised.

Table 2 – Software Languages, Tools, and Hardware Equipment adopted for implementing the solution

Sensors Control	C
Statistical Processing	R
Web Visualization	HTML, CSS, PHP, Javascript, JQuery
Structured Query Language	SQL
Data Base Management System	MySQL
Server	Apache
File Transfer Protocol Client	Filezilla
Sensors Lecture Hardware	Arduino Nano V3
WiFi Conection and Values Insertions	NodeMCU ESP8266 V2/V3
Temperature Capture	KY-028 Arduino Module
Current Capture	SCT 013

In order to provide a basis for developing the automated solution, a systematic literature review has also been taken into account. It was assumed as a main question “How deeply statistical control, image studies, and embedded systems have interact to each other?". Thus, it was performed an online search of the terms: "Statistical Process Control", "Statistical Quality Control", "Image Processing", "Image Retrieval", "Embedded systems". In order to study the contemporary literature, only papers from 2000 year have been considered. The data base used for searching was Google Scholar The exclusion criterion has been the nature of the publication. Thus, patents, higher school course syllabus, and books have been discarded. Regarding the content of each document, the title, abstract, keywords, introduction, development, and conclusion have been considered whilst the bibliography has been suppressed.

The resulting textual corpus has been statistically analysed via the free software for statistical processing *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (`IRaMuTeQ`) [144]. `IRaMuTeQ` is a Python based software and anchored on `R` [101] language, a based statistical processing open source software able to accomplish basic operations and other more complex functions. A formalism used in this study is the descendent hierarchical analysis (DHA) based on Reinert’s method [145], in which the similarity of frequency of citations of terms among text segments is measured, according to the Pearson’s Chi-Squared test [146]. Further, correspondence factorial analyses (CFAs) have been conducted, focusing on the associations among words, documents, and DHA classes. These features are easily aplicable on `IRaMuTeQ`. Many works adopt this methodology for a range of textual analyses, as review articles, interviews, and document analysis [147, 148, 149, 150, 151, 152, 153]. As a result of this study, one has the lexical and content analyses presented as follows, in the next chapter.

4 Discussion of the State of the Art

The present chapter argues that in the current information age, the interrelationship among axes such as Statistical Quality Control, Industrial Management, and Diagnosis, are paramount for the parsimonious establishment of DS and I4. It is thus claimed that these three axes directly touches the basis of I4 as well as the triple bottom line of DS of any organisation, regardless of belonging to the first, second or third sector. For instance, each one of these axes is related to at least one of the DS objectives, to be achieved until 2030 [22].

Supposing that the interrelationship among the three axes might accelerate the establishment of both, DS and I4, the present paper aims to study how deeply statistical quality control, image studies, and embedded systems have interact to each other. In this way, a textual corpus involving XXI century literature with respect to Statistical Quality Control, Image Processing, and Embedded systems has been built. The lexical analyses that based the content study has been performed via descendent hierarchical analysis (DHA) [145], in which the similarity of frequency of citations of terms among text segments is measured, according to the Pearson's Chi-Squared test [146]. Further, correspondence factorial analyses (CFAs) have been conducted, focusing on the associations among words, documents, and DHA classes.

The remaining of the section is divided in two subsections. Subsection 4.1 brings information regarding the review (e.g. adopted keywords, time period, database, and inclusion/exclusion criterion). A summary of the elected documents is also presented. In turn, Subsection 4.2 presents details and inferences with respect to the lexical and content analyses of the obtained textual corpus.

4.1 Preliminaries

For literature stratification, it was performed an online search of the terms in Google Scholar: "Statistical Process Control", "Statistical Quality Control", "Image Processing", "Image Retrieval", "Embedded systems" with the following combinations:

- "Statistical Quality Control" "Image Processing" "Embedded systems";
- "Statistical Process Control" "Image Processing" "Embedded systems";
- "Statistical Quality Control" "Image Retrieval" "Embedded systems";
- "Statistical Process Control" "Image Retrieval" "Embedded systems";

In order to study the contemporary literature, only papers since 2000 have been considered. The starting database used for searching was Google Scholar. The exclusion criterion has been the nature of the publication. Thus, patents, higher school course syllabus, and books have been discarded. Regarding the content of each document, the title, abstract, keywords, introduction, development, and conclusion sections have been appended.

The resulting textual corpus has been statistically analysed via the free software for statistical processing `IRaMuTeQ` (version 0.7 alpha 2). `IRaMuTeQ` is a `R` (version 3.5.1) based statistical processing open source software able to accomplish basic operations and other more complex functions.

At the end of the search for contemporary literature regarding statistical process/quality control, image processing/retrieval, and embedded systems, a corpus involving 25 documents have been collected. Table 3 is a summary of the corpus. The main approach of each manuscript is also highlighted, varying from machine vision to time series and education, for instance.

Table 3 – Papers selected for the systematic literature review. The manuscript are sorted by the year of publication.

Index	Main approach	Title	Year
1	Machine vision	Advanced machine vision systems and application examples [154]	2003
2	Time series	Suppressing temporal data in sensor networks using a scheme robust to aberrant readings [155]	2009
3	Statistical process	A context tree method for multistage fault detection and isolation with applications to commercial video broadcasting systems [52]	2009
4	Image analysis	Detecting External Disturbances on the Camera Lens in Wireless Multimedia Sensor Networks [156]	2010
5	Sinal processing	Real-time assessment of the reliability of welds in steel strips [51]	2010
6	Astronomy	A novel autonomous low-cost on-board data handling architecture for a pin-point planetary lander [157]	2011
7	Graphic processors	GPUSync: A framework for real-time GPU management [158]	2013
8	Smart sensors	An intelligent surveillance platform for large metropolitan areas with dense sensor deployment [159]	2013
9	Automation	Liquid level control of Coca-Cola bottles using an automated system [160]	2014
10	Education	Effective knowledge exchange with modern didactic concepts [161]	2015
11	Water monitoring	EARNPIPE: A Testbed for Smart Water Pipeline Monitoring Using Wireless Sensor Network [162]	2016

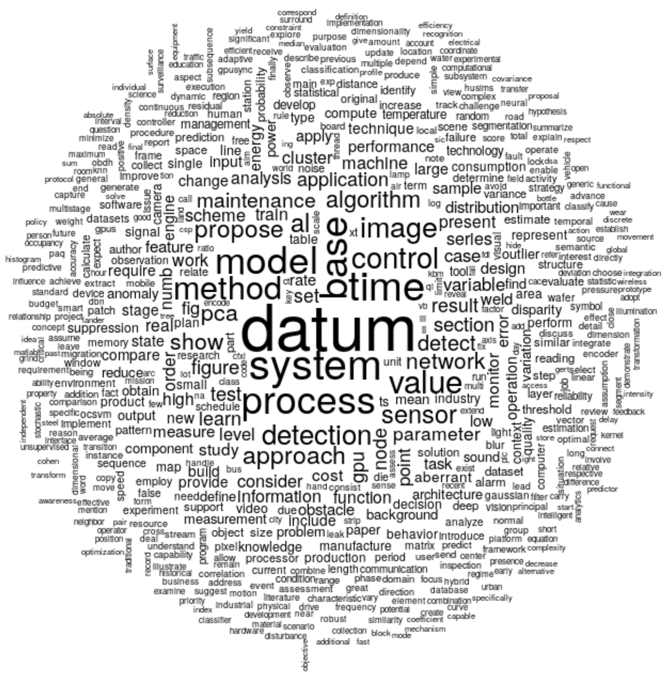
12	Street illumination	Study of smart illumination control system with M2M communication technology [163]	2017
13	Process control	Process and Operations Control in Modern Manufacturing [47]	2017
14	Time series/Algorithms	Domain agnostic online semantic segmentation for multi-dimensional time series [164]	2018
15	Image Anomalies	Image Anomalies: a Review and Synthesis of Detection Methods [46]	2018
16	Buildings	Iot for green building management [165]	2018
17	PCA	Principal component analysis: a natural approach to data exploration [50]	2018
18	Decomposition	Decomposition and Analysis of Process Variability Using Constrained Principal Component Analysis [166]	2008
19	Mobile computing	Mobile computing: Challenges and opportunities for autonomy and feedback [167]	2013
20	GERT Analysis	Evaluation of Continuous Sampling Plan (CSP-5) Parameters Using GERT Technique and MATLAB [168]	2017
21	Sequential Probability Ratio Test	COMPARATIVE ANALYSIS OF BURR TYPE III WITH PARETO TYPE II MODEL USING SPRT: ORDER STATISTICS [169]	2017
22	Industry 4.0	Research of Industry 4.0 Awareness: A Case Study of Turkey [49]	2018
23	Maintenance	PriMa: a prescriptive maintenance model for cyber-physical production systems [48]	2019
24	Obstacle detection	Unsupervised obstacle detection in driving environments using deep-learning-based stereovision [170]	2018
25	Intelligent trasportation	Obstacle Detection for Intelligent Transportation Systems Using Deep Stacked Autoencoder and-Nearest Neighbor Scheme [171]	2018

As follows, exploratory, descriptive, and explicative analyses of the terms inside these documents are performed.

4.2 Textual Analysis

A filtering of the word classes was done before processing. Defined and indefinite articles, figures of speech, conjunctions, unrecognized forms, auxiliary forms, prepositions, onomatopoeia and all types of pronouns were disregarded. Verbs, complementary verbs, adverbs, complementary adverbs, common names, complementary names and all types of adjectives were considered. From an exploratory and descriptive perspective, Figure 17 exhibits the frequency of occurrence of the main terms used through the documents. The greater the font size the greater the number of citations of the term, in the word cloud. Thus, the most frequent word is "Datum", a concatenation of equivalent terms related to "data". This word appears 1150 times, making clear its importance in the problem context (i.e. the interaction among statistical control, image studies, and embedded systems). The remaining evident terms are easily perceived in the the word cloud, e.g. system, process, model, and so on.

Figure 17 – Word cloud regarding the frequency of occurrence of the main terms used by the contemporary literature related to statistical control, image studies, and embedded systems.



Considering an explicative level of study, the DHA suggests three classes. The terms are highly associated to each other inside each class, contributing to the same latent variable. In order to name each one of these latent variables, hereafter one has “Industrial Management” for Class 1, “Statistical Processing” for Class 2, and “Diagnosis/Classification” for Class 3. These denominations have been chosen in an effort to generically represent the terms involved in each class. Thus, Class 1 directly treats the administration of production processes functions as planning and maintenance, including some physical infrastructure development aspects (e.g. sensors, actuators, and prototypes). In turn, Class 2 involves the mathematical modelling, statistical, and data treatment dedicated to quality control and processes domain. Finally, Class 3 addresses data treatment for diagnosis and classification, including sometimes software use or development. The most important words for each construct are listed in Table 4. The χ^2 column allow one to infer the magnitude of the contribution of each term to each class, indicating the associative strength between them. As higher the chi-square value, more is the association strength between the term and its respective class. Thus, maintenance, principal component analysis (PCA), and image respectively head “Industrial Management”, “Statistical Processing”, and “Diagnosis/Classification” (This last one referring to processes for detection in general terms, according to the subject addressed by each article).

Table 4 – Top 15 words list by class with the most frequent terms in all article. The Chi-Square distribution in descending order indicates the most prevalent forms. The p-value to all forms is less than 0.0001.

Class 1	χ^2	**Class 2**	χ^2	**Class 3**	χ^2
maintenance	342.39	pca	353.38	image	295.48
control	298.69	variable	186.35	detection	245.26
system	208.43	weld	153.05	anomaly	186.66
industry	195.23	variance	107.24	obstacle	177.53
machine	171.42	principal	104.29	background	171.09
production	155.51	variation	102.63	detect	170.04
knowledge	141.88	temperature	97.33	disparity	157.63
management	139.3	symbol	97.33	deep	131.35
process	116.33	component	85.34	scene	118.97
product	113.52	die	82.98	patch	118.68
manufacture	110.53	stage	80.27	false	116.93
plan	104.39	measurement	73.97	algorithm	115.56
predictive	90.68	value	72.77	map	100.45
develop	86.43	wafer	71.67	learn	99.66
integrate	84.98	ct	70.34	free	95.67

It is important to emphasize that words have meanings according to the context in which they are presented in the texts. For example, terms as "wafer" are from the context of integrated circuits fabrication where "die" means the device used to cut the wafers. Further, via CFA, the association among the words and documents inherent to each class are sketched in Figure 19 and Figure 18, respectively. One can see that two factors are sufficient to represent 100% of the variability of the frequency of citations of the terms among text segments and documents. Factor 1 has absorbed 56.59% of this variability whilst Factor 2 has encapsulated the remaining one. Thus, the CFA has performed well. Terms and documents belonging to “Industrial Management", “Statistical Processing", and “Diagnosis/Classification" are presented in red, green, and blue colour, in this order. The word clouds regarding the importance of each word and document for the classes are disposed, in this case, with a clear dissimilarity among them. Further, few words (documents) with low importance to each DHA class tend to come closer to words (documents) from other classes, indicating some possible relations. As an example, one can see the crossing among terms (e.g. model vs approach, data set vs datum, and datum vs process) as well as among documents (e.g. doc_14 vs doc_11).

One can see (Figure 19) that the main document of Class 1 (paper 13) deals directly with manufacture evolution. It shows how this change affects the quality, reliability and productivity concepts. Similarly the doc 23 proposes a model of prescriptive maintenance that allows a custom methodology. Both papers portray well the class aspects. In the Class 2, the most representative is the doc 17 that deals with Exploratory Data Analysis through Principal Component Analysis (PCA). The doc 15 makes a reviews about techniques for image anomalies detection, thus representing the Class 3. The docs 24 and 25 are equally relevant to each other for class representation. Both deal with detection methods.

Figure 18 – CFA for HDA regarding the frequency of occurrence of the main terms used by the contemporary literature related to statistical control, image studies, and embedded systems. Terms belonging to “Industrial Management", “Statistical Processing", and “Diagnosis/Classification" are presented in red, green, and blue colour, in this order.

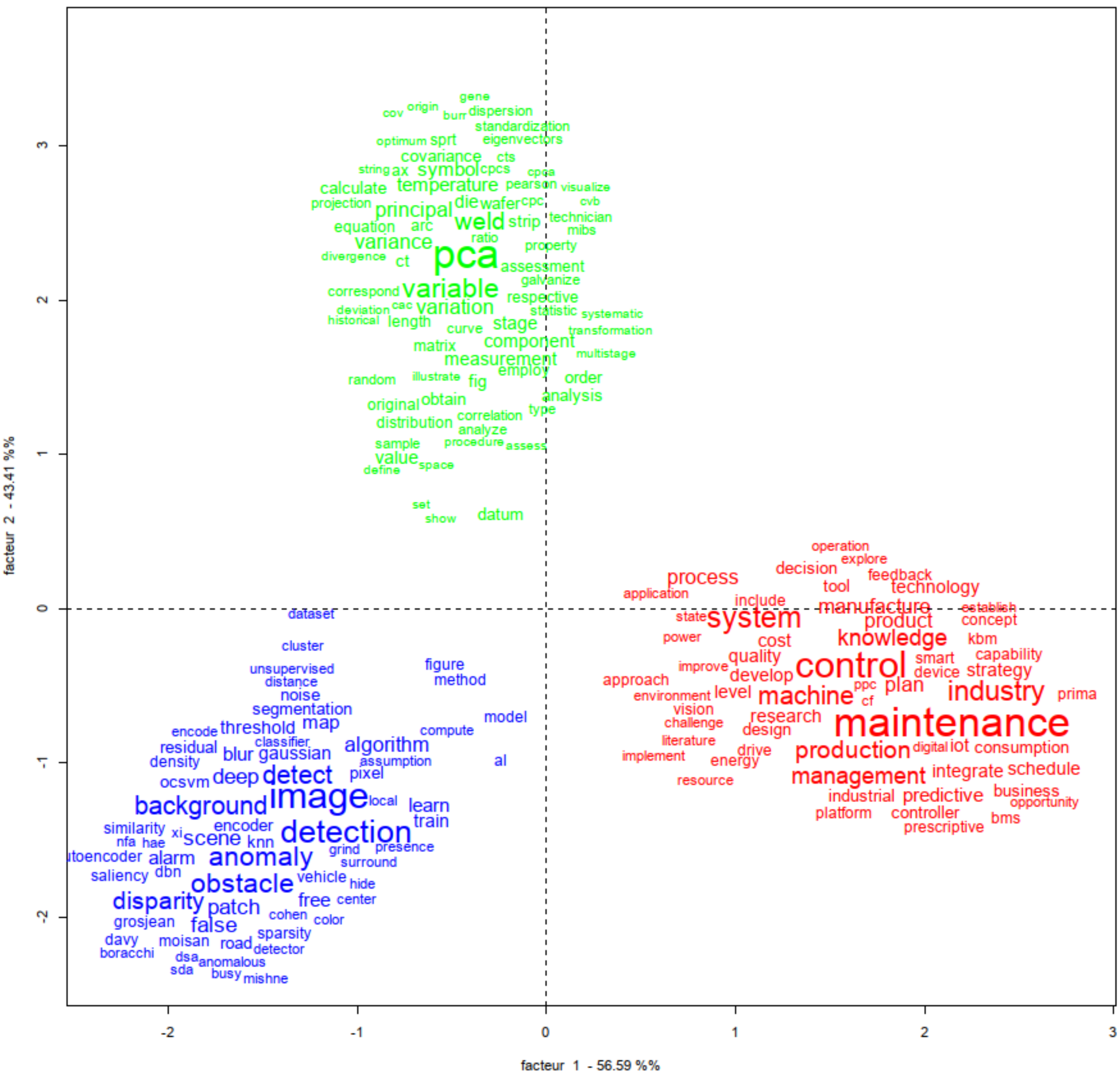
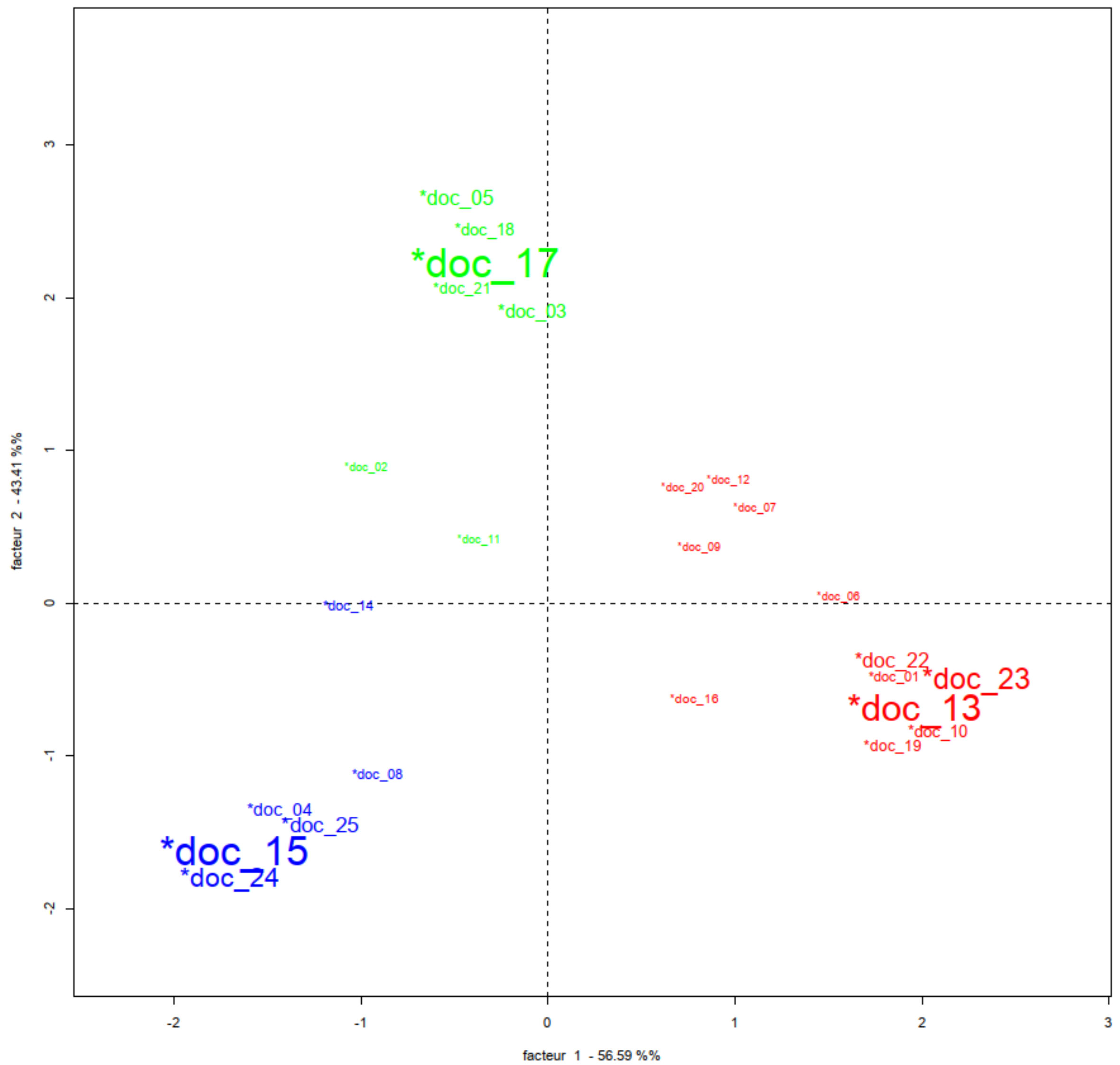


Figure 19 – CFA for HDA regarding documents from the contemporary literature related to statistical control, image studies, and embedded systems. Documents belonging to “Industrial Management", “Statistical Processing", and “Diagnosis/Classification" are presented in red, green, and blue colour, in this order.



Thus, Factor 1 and Factor 2 are latent variables that represent associations among terms and documents. Figure 18 and Figure 19 reflect these relationships. Particularly, Factor 1 exclusively receives positive (negative) contributions from the “Industrial Management" (“Diagnosis/Classification") class whilst “Statistical Processing" seems to slightly impact negatively such factor. Thus, Factor 1 (X axis) can be named “Current Industrial Management Philosophy", yet needing to evolve in terms of statistical diagnosis and classification practices. On the other hand, Factor 2 (Y axis) exclusively receives positive (negative) impact from “Statistical Processing" (“Diagnosis/Classification"). Further, this factor is slightly negatively influenced by “Industrial Management". Thus, Factor 2 can be named “Basic Research" and deserves more attention in terms of application and technology transference.

5 Experiment

It is expected that a mix of sensors to monitor the production lines, statistical modelling formalisms of relevant variables, and the use of industrial management techniques can lead to effective decision-making in contributing to wastes reduction, environmental degradation softening, and human life quality increase.

In this chapter, results regarding the study of turned-on machinery energy consumption (ENER_ON), turned-off machinery energy consumption (ENER_OFF), turned-on machinery temperature (TEMP_ON), and turned-off machinery temperature (TEMP_OFF) time series, with respect to CTMC, are taken into account. Specifically, the performance of ARIMA, ANN, and cMV formalisms is studied. The ANN modelling was optimised through a simulated annealing (SA)-based method hybridising the `R` packages `neuralnet` and `GenSA`. The possible parameters to be chosen by the simulated annealing are the standards parameters of the package.

Table 6 summarises the architecture of the ANN models per case. Autoregressive models with orders from 2 (regarding TEMP_ON series) to 23 (for TEMP_OFF series) have been achieved. In turn, the number of nodes of the intermediate hidden layer of the ANNs varies from 2 (TEMP_ON) to 14 (ENER_ON). The hyperbolic tangent has been considered as the best activation function. Further, the respective cMV residuals-based CUSUM control charts are presented. To verify the best model for each time series, the aggregated performance measures (APM) of MSE, MAPE, ARV, ID, Theil, WPOCID, Reg_Intercept, Reg_Slope, and WR^2 metrics (see Equation 2.28) is considered.

The tuning parameters for ARIMA and ANN optimisation are presented in Table 5. Thus, assuming N as the size of the observed time series, the training sample size, that is, the sample adopted for fitting the models (based on training%), was $0.8 \times N$; whilst the non-seasonal autoregressive and the seasonal autoregressive order was (from AR_ARS) upper bounded by, respectively, $0.05 \times N$ and $\frac{0.05}{2} \times N$.

Regarding ANN modelling, the threshold for the partial derivatives of the error function as stopping criterion of the ANN models was (ANN.threshold=) 0.001 and the maximum steps for the training of the ANN was (ANN.stepmax=) 1000000. Each ANN modelling has been based on the `neuralnet` package [107], and the GSA-based ANN modelling optimisation process has been performed via `GenSA` package [172] of `R` software. In turn, the maximum number of models tested in the ARIMA optimisation process (based on the `forecast` package [100] of `R`). Regarding `GenSA` [172], the maximum number of calls of the BIC-based fitness function was (GSA.max.call=)5000, the maximum running time was (GSA.max.time=) 600 seconds, the maximum number of iterations of the algorithm was (GSA.max.it=) 5000, the initial value for temperature was (GSA.temperature=) 5000, and the algorithm would stop when there were no improvement after (GSA.nb.stop.improvement=) 20 steps.

training%	AR_ARS%	ANN.threshold	ANN.stepmax	ARIMA.nmodels
0.8	0.05	0.001	1000000	5000

GSA.max.call	GSA.max.time	GSA.maxit	GSA.temperature	GSA.nb.stop.improvement
5000	600	5000	5000	20

Table 5 – Tuning parameters of the single forecasting models of ENER_ON, ENER_OFF, TEMP_ON, and TEMP_OFF series.

Specific results for each time series are presented as follows. In this way, the study of the machinery in turned-on and turned-off periods are taken into account.

5.1 Models architecture and performance

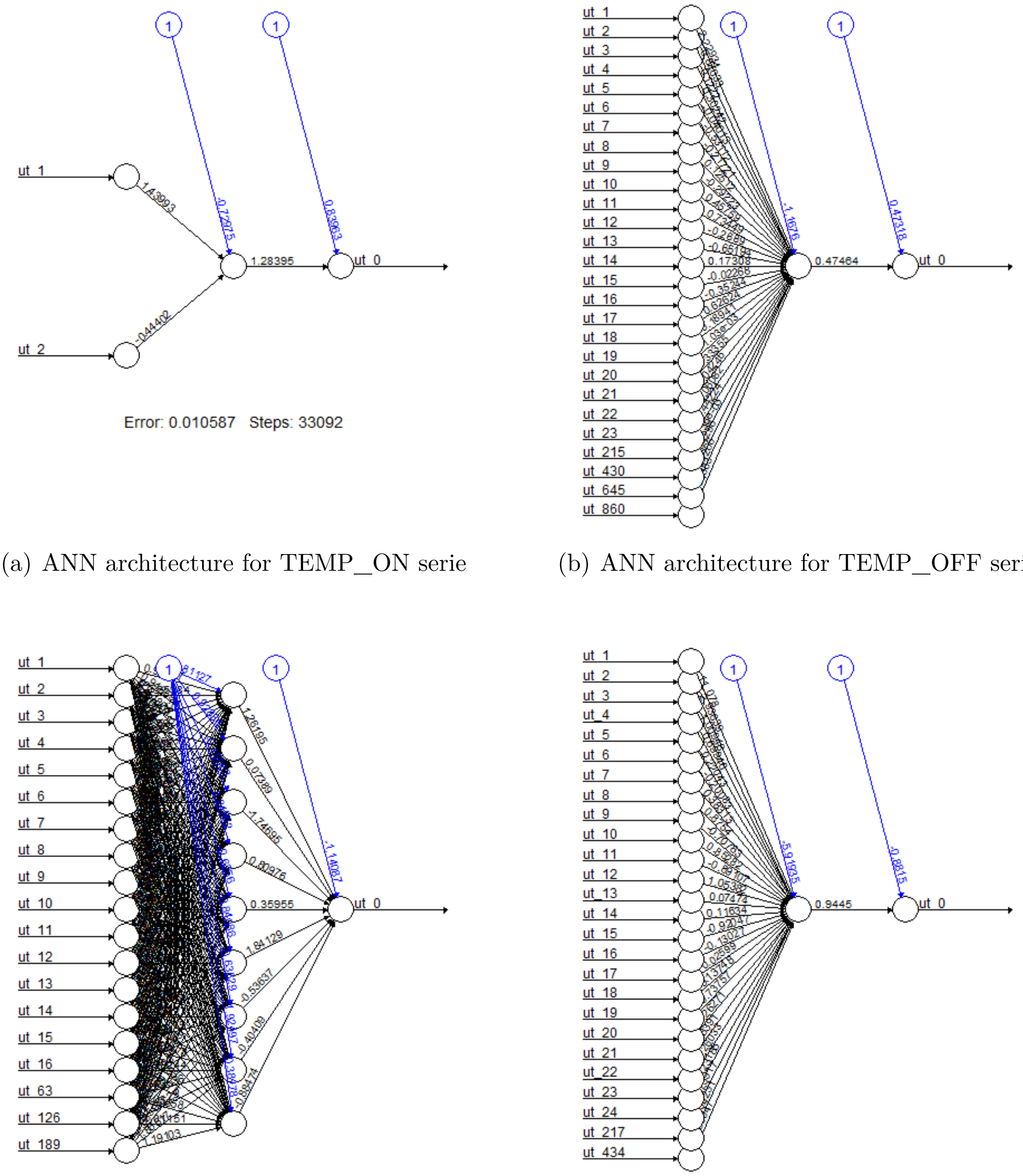
Table 6 brings ANN architecture details for ENER_ON, TEMP_ON, ENER_OFF, TEMP_OFF series. The near-optimal GSA-based ANN models designed to energy consumption series did involve autoregressive orders (pAR) greater than temperature series, ranging fro 2 (TEMP_ON) to 24 (ENER_OFF). All series have presented seasonality, but TEMP_ON. To date, the ANN model of TEMP_OFF has demanded a seasonal autoregressive order of (pARS=) 4, with length (seasonality=) 215. The hidden layer are composed of (nNodes=) 1 node, but ENER_ON (with 9 nodes). The ANN machines were fitted via a backpropagation (backprop) learning algorithm (turned-on-based series) or smallest absolute gradient (sag) method (turned-off-based series). In turn, the logistic activation function was considered to energy consumption series and the tangent hyperbolic (tanh) one to temperature series. The learning rate of 5×10^{-3} was considered when necessary. The most computer demanding series has been TEMP_OFF, consuming near 45 seconds for ANN modelling. Figure 20 exhibits the architecture of each ANN model.

characteristic	ENER_ON	TEMP_ON	ENER_OFF	TEMP_OFF
pAR	16	2	24	23
pARS	3	0	2	4
seasonality	63	-	217	215
nNodes	9	1	1	1
learningAlgorithm	backprop	backprop	sag	sag
activationFunction	logistic	tanh	logistic	tanh
learningRate	5e-03	5e-03	5e-03	5e-03
modelling time (sec)	17.7321	10.0645	37.7263	45.4434

Table 6 – Architecture of the SA-based near-optimal single ANN model for each time series taken into account (ENER_ON, TEMP_ON, ENER_OFF, TEMP_OFF).

[1]

Figure 20 – ANN architectures.



(a) ANN architecture for TEMP_ON serie (b) ANN architecture for TEMP_OFF serie

(c) ANN architecture for ENER_ON serie (d) ANN architecture for ENER_OFF serie

Source: The author

In turn, the `auto.arima` function has achieved an ARIMA(0,1,0) model for series involving temperature measures. Thus, a random walk model was considered. On the other hand, the `auto.arima` function has achieved an ARIMA(2,1,1) model for ENER_ON

[1] praf: rever nomes das subfiguras na Figura 20... acho que alguns estão errados... conferir se o texto correspondente estah ok...

and an ARIMA(1,1,2) for ENER_OFF. The ENER_OFF series has been the most time consuming, demanding near 0.28 seconds in the ARIMA modelling.

The characteristics of the computer used to execute the modelling directly influence in the modelling time. In this case it was used a PC with Windows 10 Home (64 bits) operational system, Intel i7 processor with 2.6GHz and 8GB RAM memory.

order	ENER_ON	TEMP_ON	ENER_OFF	TEMP_OFF
p	2	0	1	0
d	1	1	1	1
q	1	0	2	0
P	0	0	0	0
D	0	0	0	0
Q	0	0	0	0
modelling time (sec)	0.0907	0.016	0.2753	0.0259

Table 7 – Architecture of the `auto.arima`-based near-optimal single ARIMA model for each time series taken into account (ENER_ON, TEMP_ON, ENER_OFF, TEMP_OFF).

Finally, Table 8 presents the weights of ARIMA and ANN in the cMV combination of the energy consumption and temperature time series. In the case of combining two models, these weights reflect the variability of the predictors residuals in the training series. One can see that ANN has performed much better than ARIMA in the training series when the machinery was turned-on, though the later has some contribution to the cMV combination. When the machinery was turned-off, the ANN and ARIMA contribution to the cMV combination were nearly similar. The time consumption of the cMV modelling has been negligible (at most 1×10^{-3} seconds).

model	ENER_ON	TEMP_ON	ENER_OFF	TEMP_OFF
ANN	0.9824	1.0613	0.451	0.5106
ARIMA	0.0176	-0.0613	0.549	0.4894
modelling time (sec)	0	1e-03	0	0

Table 8 – Weights of the near-optimal Minimal Variance combination model for each time series taken into account (ENER_ON, TEMP_ON, ENER_OFF, TEMP_OFF).

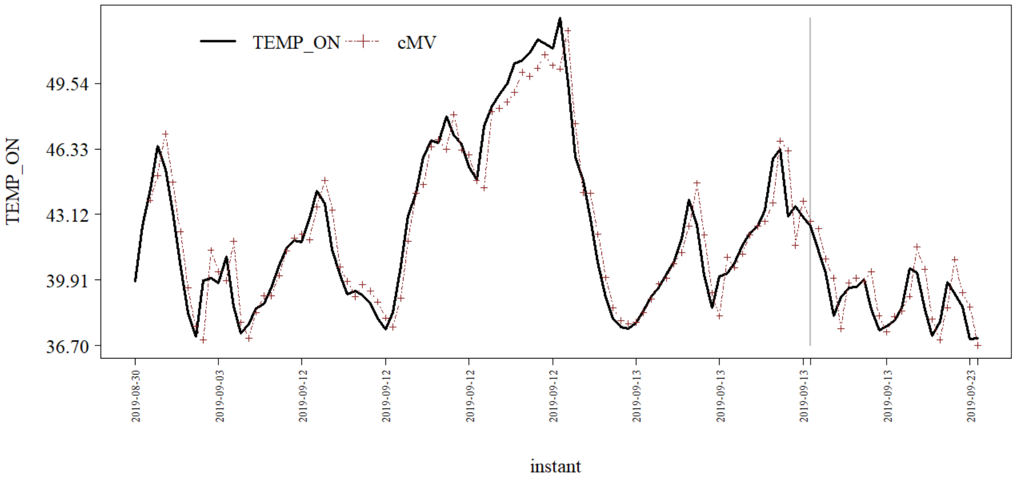
Table 9 summarises the performance of the predictors when forecasting the training set of ENER_ON, TEMP_ON, ENER_OFF, and TEMP_OFF series. ARIMA has always presented the worst results in terms of APM (Equation 2.28). In fact, the poor performance of ARIMA in modelling the training series has depreciated cMV in comparison with ANN in two occasions (TEMP_ON and ENER_OFF). Anyway, cMV is preferred once the combiner aggregates ARIMA and ANN formalisms. Additional details regarding the performance of the models are presented at Appendix A.2.

series	Worst	Best	ANN	ARIMA	cMV
ENER_ON	ARIMA	cMV	0.3401	0.6667	0.3352
TEMP_ON	ARIMA	ANN	4.1e-03	1	0.011
ENER_OFF	ARIMA	ANN	0.2222	0.7145	0.4096
TEMP_OFF	ARIMA	cMV	0.2229	0.8889	0.197

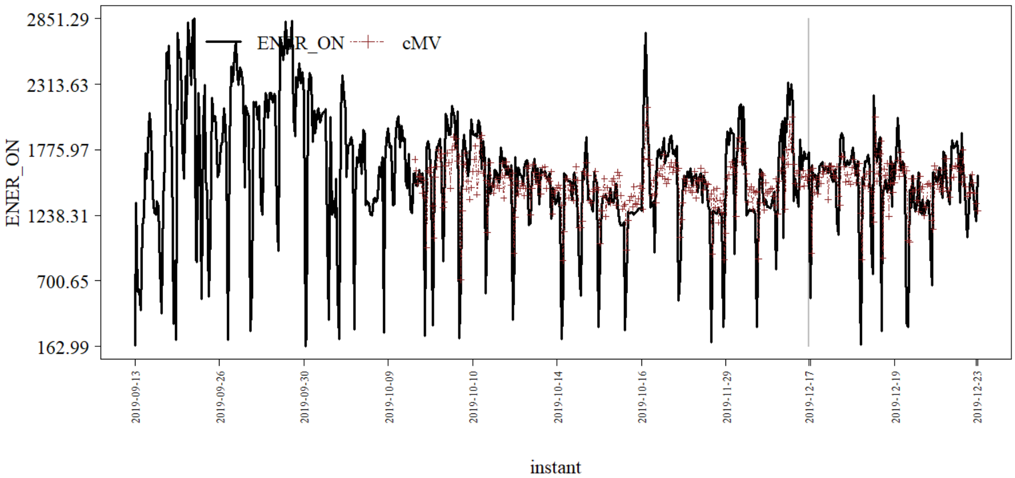
Table 9 – Aggregate mean normalised performance of the forecasting models (ANN, ARIMA, cMV) when predicting ENER_ON, TEMP_ON, ENER_OFF, TEMP_OFF time series (Training phase).

Figures 21 and 22 sketch the forecasts of the cMV model for turned-on- and turned-off-based series, in this order. On can highlight the challenge faced by the model in capturing the variability pattern of the machinery temperature in turned-on periods, leading to a subsequent high-tolerance control chart for the model residuals. On the other hand, it seems cMV has performed well when forecasting the remaining series. Appending figures exhibiting single models forecasts are given at Appendix A.1.

Figure 21 – Forecasts of cMV for turned-on machinery condition, in terms of temperature (a) and electrical energy consumption (b). The vertical grey line separates training and test phases.



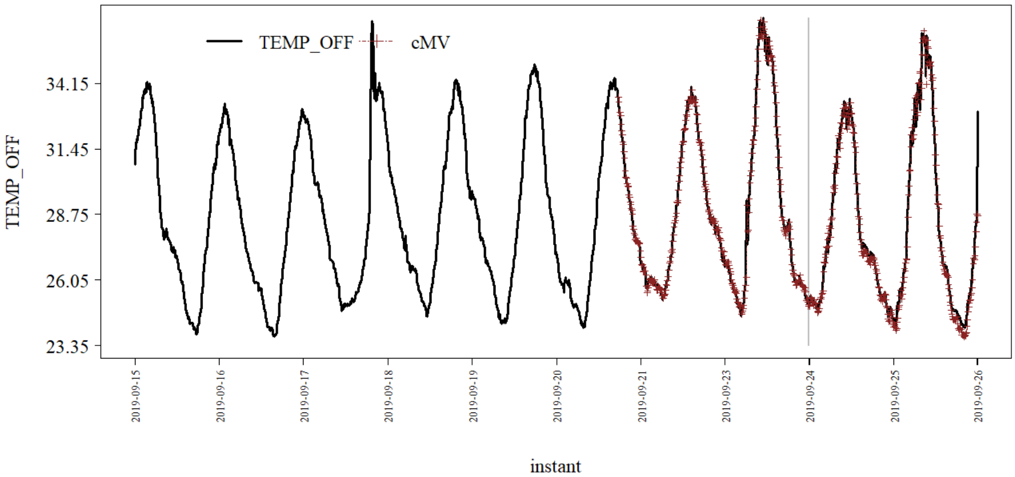
(a)



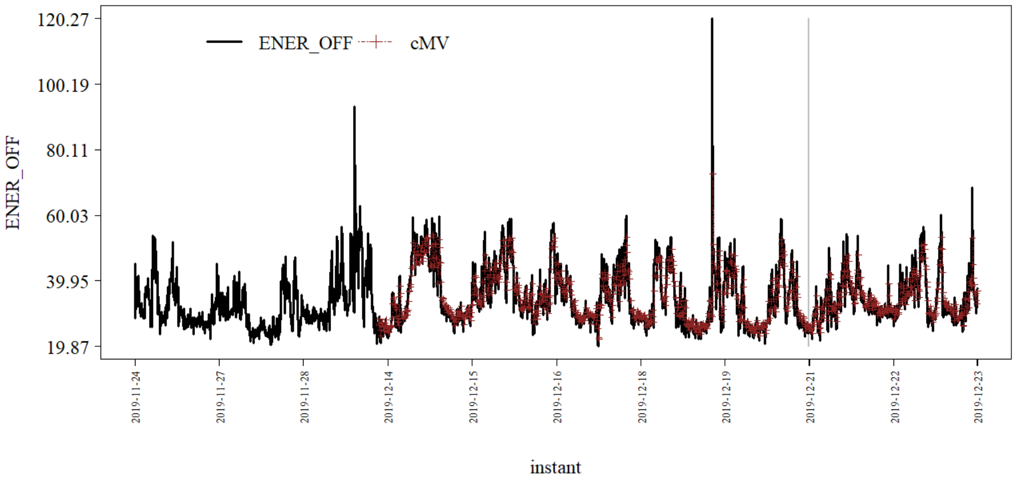
(b)

Source: The author

Figure 22 – Forecasts of cMV for turned-off machinery condition, in terms of temperature (a) and electrical energy consumption (b). The vertical grey line separates training and test phases.



(a)



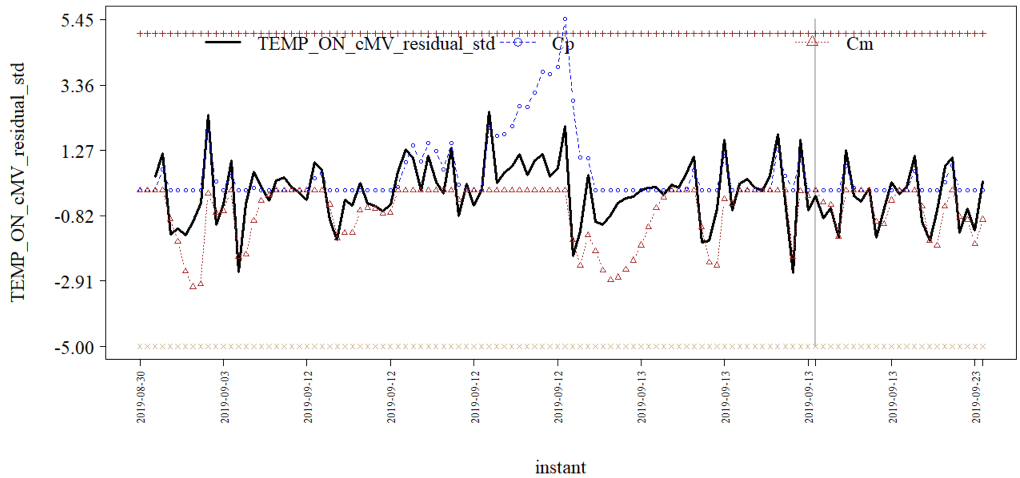
(b)

Source: The author

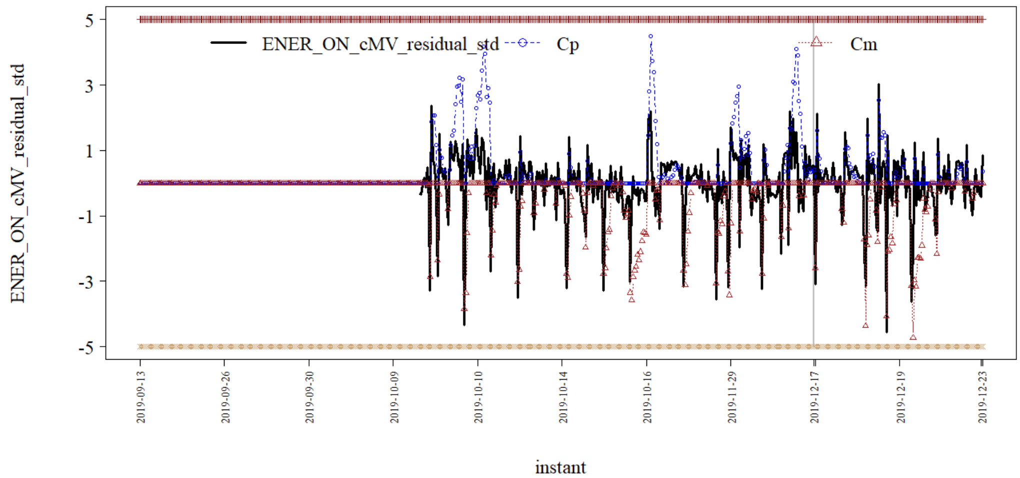
5.2 Quality control process

Figure 23 illustrates the control chart of cMV residuals CUSUM for the temperature and energy consumption time series that reflect the turned-on machinery periods. One can infer that both series remain under control through all period under study, but TEMP_ON around 2019-09-12. In this period, cMV has systematically underestimated the target temperature values of TEMP_ON (positive residuals have been registered). It might indicate a misuse of the machine by, for instance, using a inadequate blade height, leading to a lesser rolling stage though requiring greater mechanical effort.

Figure 23 – CUSUM control chart (based on $k = 0.5$ and $h = 5$) with respect to the residuals of cMV when forecasting turned-on machinery condition, in terms of temperature (a) and electrical energy consumption (b). The vertical grey line separates training and test phases.



(a)



(b)

Source: The author

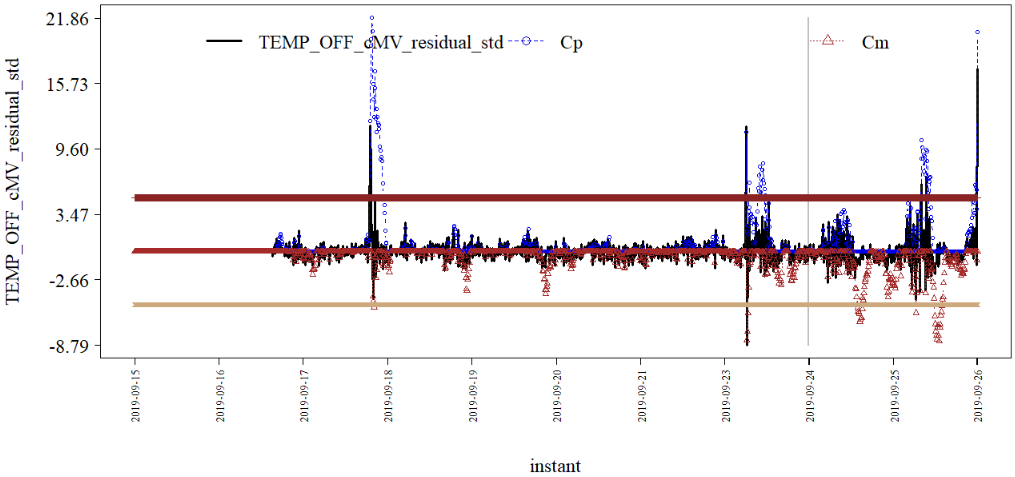
Table 10 summarises the inferred out-of-control instants of TEMP_ON series. The N^+ (N^-) informs the number of previous subsequent and relevant positive (negative) residuals until out-of-control signal. Thus, considering the first instant at which TEMP_ON would cross the lower control limit ($h = -5$), $t = 57$ (i.e. 2019 September 12th), one has that relevant positive residuals have been accumulated since ($N_t^+ =$) 11 steps ago. It is worthwhile to highlight that the temperature data were registered at time units of 10 minutes; thus, the out-of-control pattern has in fact started near two hours before its effective appearance. In this way, one could expect that the out-of-control pattern has started at $t - N_t^+ = 57 - 11 = 46$ time units. The reasons for eventual disturbances might thus be better studied.

time series	time index	date	cMV_residual_std	C+	N+	C-	N-
TEMP_ON	57	2019-09-12	2.04	5.47	11	0	0

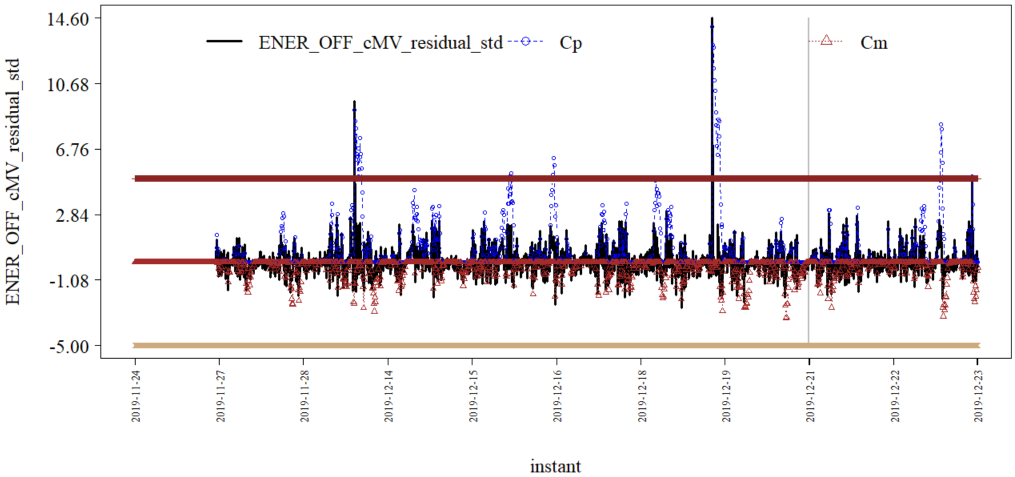
Table 10 – Summary of the inferred out-of-control period of the machinery turned-on period (TEMP_ON and ENR_ON series). The CUSUM approach with respect to the residuals of the Minimal Variance Combiner of the `auto.arima`-based and SA-based near-optimal single ANN models, fitted to the time series, is taken into account.

Figure 24 presents the behaviour of the CUSUM control chart related to the cMV residuals in the face of turned-off periods (TEMP_OFF and ENER_OFF). Table 11 exhibits specific details. Considering TEMP_OFF series, one can see an out-of-control pattern involving an unusual temperature increasing followed by a fast fall. It can be seen when comparing Figure 24 with Figure 22. Such a phenomenon has lead to positive and then negative systematic bias of the process and thus to the extrapolation of the control limits in 15 occasions (see Table 11). In turn, with respect to the ENER_OFF series, one can see that out-of-control signals correspond to energy spikes, always. In fact, in some circumstances, e.g. time indexes 392 (2019-12-13) and 1028 (2019-12-19), the disturbances have started at one step back (N^{+}=1), only. This behaviour might be an indicative of energy leak and thus reveal the need of improvements in the electrical energy system of the plant.

Figure 24 – CUSUM control chart (based on $k = 0.5$ and $h = 5$) with respect to the residuals of cMV when forecasting turned-off machinery condition, in terms of temperature (a) and electrical energy consumption (b). The vertical grey line separates training and test phases.



(a)



(b)

Source: The author

time series	time index	date	cMV_residual_std	C+	N+	C-	N-
TEMP_OFF	420	2019-09-18	11.74	12.23	4	0	0
TEMP_OFF	1089	2019-09-23	11.67	11.17	1	0	0
TEMP_OFF	1094	2019-09-23	3.16	6.43	6	0	0
TEMP_OFF	1109	2019-09-23	1.80	5.04	4	0	0
TEMP_OFF	1117	2019-09-23	3.28	7.66	12	0	0
TEMP_OFF	1123	2019-09-23	2.74	5.26	18	0	0
TEMP_OFF	1287	2019-09-24	-0.88	0	0	-5.32	9
TEMP_OFF	1380	2019-09-25	4.25	5.20	7	0	0
TEMP_OFF	1391	2019-09-25	-4.51	0	0	-5.74	3
TEMP_OFF	1400	2019-09-25	6.28	8.70	2	0	0
TEMP_OFF	1409	2019-09-25	6.94	9.48	11	0	0
TEMP_OFF	1414	2019-09-25	3.05	6.48	16	0	0
TEMP_OFF	1424	2019-09-25	-2.13	0	0	-6.04	8
TEMP_OFF	1496	2019-09-26	2.20	6.26	17	0	0
TEMP_OFF	1500	2019-09-26	17.11	20.59	21	0	0
ENER_OFF	392	2019-12-13	9.62	9.12	1	0	0
ENER_OFF	667	2019-12-15	2.14	5.12	5	0	0
ENER_OFF	671	2019-12-15	2.21	5.32	9	0	0
ENER_OFF	744	2019-12-16	1.24	5.03	6	0	0
ENER_OFF	1028	2019-12-19	14.60	14.10	1	0	0
ENER_OFF	1434	2019-12-22	2.17	6.55	6	0	0
ENER_OFF	1490	2019-12-23	4.96	5.11	6	0	0

Table 11 – Summary of the inferred out-of-control period of the machinery turned-off period (TEMP_OFF and ENR_OFF series). The CUSUM approach with respect to the residuals of the Minimal Variance Combiner of the `auto.arima`-based and SA-based near-optimal single ANN models, fitted to the time series, is taken into account.

6 Conclusion

This chapter brings a synthesis of the manuscript in terms of how the adopted methodologies can contribute to reach the automation of data recording and modelling for controlling conveyor belts production lines, in special, and production systems in general. The project has claimed that paradigms like the combination of optimised single time series forecasting models, the predictors residuals-based CUSUM control charts, the system condition-based control frameworks, and the automated collection, register, and control charts update must be prioritised. Such a framework has been experimented at a real-world limestone production line. Thus, the work aimed at directly contributing to the main sustainable development dimensions ("Economic", "Social", and "Environmental") by offering tools dedicated to balance three technological axes ("Statistical Control", "Industrial Management", and "Detection and Diagnosis Techniques"). These axes have been identified during literature review regarding the the state of the art of statistical process/quality control, image processing/retrieval, and embedded systems. This contributions are grounded by the all built framework and lead to the following results:

(i) Software development to statistical data modelling with friendly and responsive interface;

(ii) Sensors installation in a mineral processing unity to data collect;

(iii) Reports generation able to bring all aspects of interest of inherent variables in the process;

(iv) Registration of data collection, processing, and modelling software;

(v) Paper submission of state of the art textual analysis.

6.1 Limitations and difficulties

The scarcity of infrastructure resources represents the most of limitations of this research. The studied machine has not outlets to accessories implementations. To perform sensors installation, it was necessary an external adaptation to add outlets to supply them. The way chosen (extension cable) is not the best solution, due the fixation difficult and hazard exposition. The Mineral Centre adopted as study case is held by the government, and all acquisitions depend on bidding processes, making difficult the addition of outlets in the wall next to the machine at short term. Other aspect related to infrastructure was the internet signal quality. The router was positioned with some physical barriers and even the shape of the machine, making difficult the connection in some periods (i. e. electromagnetic phenomenon). Further, the presence of the sensors in the plant have also caused strangeness to operators due to their unfamiliarity. In fact, the machine has not been operated by the same operators at all time. Therefore, it was necessary warn about the need of not disconnecting the electrical extension cable, for instance.

In terms of data storage, it would be ideal an another server to data mirroring containing all necessary technologies (`R`, `MySQL` and `PHP`) beyond the limited computer used in the research. One must notice that all paid servers in the cloud, though involving `MySQL` and `PHP`, do not work with `MySQL` and `R` tools. A Virtual Private Server service would allows the installation of `R` and any other required technology, but with an impracticable cost, according to the available financial resources.

In addition to the physical aspects, the nature of the machines operation (working on demand) does not allow data recording without the need to distinguish data according to the machines status (on or off). The data separation was done through the subjective definition of a threshold. In the case of temperature when the machine was immediately turned-off, the classification of the real state of the machine might be confusing due to the slow natural cooling, leading to the illusion that the machine remains turned-on, for the sake of ilustration.

6.2 Future Works

In order to improve the quality of the research, the following actions reflect the ongoing step of the work:

- (i) To study alternative criteria for dealing with different states of the production system (e.g. turned-on and turned-off periods). Alternatives like temporal-window framework for modelling and forecasting time series [112] might be adopted;
- (ii) To apply alternative formalisms to data processing and analysis;
- (iii) To formally request the opening of a bidding process to electrical adaptation to improve the sensors disposition;
- (iv) To promote a training program to elucidate questions regarding the sensors usage and classical questions related to machine operation, such as health and safety at work;
- (v) To implement a software version to mobile devices;
- (vi) To study the behaviour of other variables related to the condition of the machines, based on noise, air pollution, water consumption, and vibration sensors. for instance.

Bibliography

[1] J. Alcamo, M. Chenje, A. Ghai, F. Keita-Ouane, S. Leonard, M. Niamir-Fuller, and C. Nobbe, "Embedding the environment in sustainable development goals," tech. rep., Unep post-2015 discussion paper 1, UNEP, Version 2 19 July, 2013.

[2] G. Assembly, "sustainable development goals," *SDGs), Transforming our world: the*, vol. 2030, 2015.

[3] M. Redclift, *Sustainable development: Exploring the contradictions*. Routledge, 2002.

[4] A. Irwin, *Citizen science: A study of people, expertise and sustainable development*. Routledge, 2002.

[5] N. R. Council *et al.*, "Policy division, board on sustainable development.(1999)," *Our common journey: A transition toward sustainability*.

[6] K. W. Robert, T. M. Parris, and A. A. Leiserowitz, "What is sustainable development? goals, indicators, values, and practice," *Environment: science and policy for sustainable development*, vol. 47, no. 3, pp. 8–21, 2005.

[7] I. Dincer, "Renewable energy and sustainable development: a crucial review," *Renewable and sustainable energy reviews*, vol. 4, no. 2, pp. 157–175, 2000.

[8] A. M. Omer, "Energy, environment and sustainable development," *Renewable and sustainable energy reviews*, vol. 12, no. 9, pp. 2265–2300, 2008.

[9] A. A. Ahmed and N. Q. Alwan, "Comparison between (x –r) charts in traditional and deviation nominal methods with usage process capability analysis," *Association of Arab Universities Journal of Engineering Sciences*, vol. 25, no. 5, pp. 415–428, 2018.

[10] P. BIANCONE, S. F. SECINARO, V. Brescia, and D. Calandra, "Data quality methods and applications in health care system: A systematic literature review," 2019.

[11] A. Di Bucchianico and E. R. van den Heuvel, "Shewhart's idea of predictability and modern statistics," in *Frontiers in Statistical Quality Control 11*, pp. 237–248, Springer, 2015.

[12] B. Zaman, N. Abbas, M. Riaz, and M. H. Lee, "Mixed cusum-ewma chart for monitoring process dispersion," *The International Journal of Advanced Manufacturing Technology*, vol. 86, no. 9-12, pp. 3025–3039, 2016.

[13] M. Faisal, R. F. Zafar, N. Abbas, M. Riaz, and T. Mahmood, "A modified cusum control chart for monitoring industrial processes," *Quality and Reliability Engineering International*, vol. 34, no. 6, pp. 1045–1058, 2018.

[14] R. A. Sanusi, M. Riaz, N. Abbas, *et al.*, "Combined shewhart cusum charts using auxiliary variable," *Computers & industrial engineering*, vol. 105, pp. 329–337, 2017.

[15] D. Prajapati, "Effectiveness of conventional cusum control chart for correlated observations," *International Journal of Modeling and Optimization*, vol. 5, no. 2, p. 135, 2015.

[16] Z. Wen, H. Liang, J. Liang, Q. Liang, and H. Xia, "Evaluation of the learning curve of laparoscopic choledochal cyst excision and roux-en-y hepaticojejunostomy in children: Cusum analysis of a single surgeon's experience," *Surgical endoscopy*, vol. 31, no. 2, pp. 778–787, 2017.

[17] D. Boullosa-Falces, M. A. Gomez-Solaetxe, Z. Sanchez-Varela, S. García, and A. Trueba, "Validation of cusum control chart for biofouling detection in heat exchangers," *Applied Thermal Engineering*, vol. 152, pp. 24–31, 2019.

[18] A. Idrissa and N. O. Nwazor, "Optimisation of energy management in a process industry: A case study," in *2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON)*, pp. 1075–1082, IEEE, 2017.

[19] S. Gezer, A. Avcı, and M. Türktan, "Cusum analysis for learning curve of videothoracoscopic lobectomy," *Open Medicine*, vol. 11, no. 1, pp. 574–577, 2016.

[20] N. Abbas, I. A. Raji, M. Riaz, and K. Al-Ghamdi, "On designing mixed ewma dual-cusum chart with applications in petro-chemical industry," *IEEE Access*, vol. 6, pp. 78931–78946, 2018.

[21] M. Keshavarz, S. Asadzadeh, and S. T. A. Niaki, "Controlling autocorrelated data in multistage manufacturing processes with an application to textile industry," *Quality and Reliability Engineering International*, vol. 35, no. 7, pp. 2314–2326, 2019.

[22] U. Desa *et al.*, "Transforming our world: The 2030 agenda for sustainable development," 2016.

[23] E. Ó. Conchúir, P. J. Ågerfalk, H. H. Olsson, and B. Fitzgerald, "Global software development: where are the benefits?," *Communications of the ACM*, vol. 52, no. 8, pp. 127–131, 2009.

[24] J. D. Herbsleb, A. Mockus, T. A. Finholt, and R. E. Grinter, "An empirical study of global software development: distance and speed," in *Proceedings of the 23rd*

international conference on software engineering, pp. 81–90, IEEE Computer Society, 2001.

[25] R. D. Battin, R. Crocker, J. Kreidler, and K. Subramanian, "Leveraging resources in global software development," *IEEE software*, vol. 18, no. 2, pp. 70–77, 2001.

[26] A. Mockus and J. Herbsleb, "Challenges of global software development," in *Proceedings seventh international software metrics symposium*, pp. 182–184, IEEE, 2001.

[27] E. Carmel and R. Agarwal, "Tactical approaches for alleviating distance in global software development," *IEEE software*, vol. 18, no. 2, pp. 22–29, 2001.

[28] C. Ebert and P. De Neve, "Surviving global software development," *IEEE software*, vol. 18, no. 2, pp. 62–69, 2001.

[29] E. Hossain, M. A. Babar, and H.-y. Paik, "Using scrum in global software development: a systematic literature review," in *2009 Fourth IEEE International Conference on Global Software Engineering*, pp. 175–184, Ieee, 2009.

[30] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[31] I. Lee and K. Lee, "The internet of things (iot): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015.

[32] A. Brush, B. Lee, R. Mahajan, S. Agarwal, S. Saroiu, and C. Dixon, "Home automation in the wild: challenges and opportunities," in *proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2115–2124, ACM, 2011.

[33] V. S. Gunge and P. S. Yalagi, "Smart home automation: a literature review," *International Journal of Computer Applications*, vol. 975, p. 8887, 2016.

[34] B. Mason and J. Antony, "Statistical process control: an essential ingredient for improving service and manufacuring quality," *Managing Service Quality: An International Journal*, vol. 10, no. 4, pp. 233–238, 2000.

[35] R. Aliverdi, L. M. Naeni, and A. Salehipour, "Monitoring project duration and cost in a construction project by applying statistical quality control charts," *International Journal of Project Management*, vol. 31, no. 3, pp. 411–423, 2013.

[36] A. Ustundag and E. Cevikcan, *Industry 4.0: managing the digital transformation*. Springer, 2017.

[37] K. Santos, E. Loures, F. Piechnicki, and O. Canciglieri, “Opportunities assessment of product development process in industry 4.0,” *Procedia Manufacturing*, vol. 11, pp. 1358–1365, 2017.

[38] J. Wan, S. Tang, Z. Shu, D. Li, S. Wang, M. Imran, and A. V. Vasilakos, “Software-defined industrial internet of things in the context of industry 4.0,” *IEEE Sensors Journal*, vol. 16, no. 20, pp. 7373–7380, 2016.

[39] C. F. do BRASIL, “Constituição da república federativa do brasil de 1988,” 2010.

[40] M. T. Müller, “O senai e a educação profissionalizante no brasil,” *Revista HISTEDBR On-Line*, vol. 10, no. 40, pp. 189–211, 2010.

[41] L. P. da Silva, “Formação profissional no brasil: o papel do serviço nacional de aprendizagem industrial-senai,” *História (São Paulo)*, vol. 29, no. 1, pp. 394–417, 2010.

[42] M. T. Müller, “A educação profissionalizante no brasil–das corporações de ofícios à criação do senai,” *Revista Estudos do Trabalho. Ano III*, no. 05, pp. 01–31, 2009.

[43] S. S. da Indústria, “Departamento nacional,” *O SESI, o trabalhador ea indústria: um resgate histórico*, 2008.

[44] S. S. da Indústria, “Serviço nacional de aprendizagem industrial,” *Brasília: CNI*, 2008.

[45] T. Bacelar, “Nordeste: desenvolvimento recente e perspectivas,” 2014.

[46] T. Ehret, A. Davy, J.-M. Morel, and M. Delbracio, “Image anomalies: A review and synthesis of detection methods,” *Journal of Mathematical Imaging and Vision*, pp. 1–34, 2019.

[47] D. Djurdjanovic, L. Mears, F. A. Niaki, A. U. Haq, and L. Li, “Process and operations control in modern manufacturing,” *Volume 3: Manufacturing Equipment and Systems. ASME*, p. V003T04A057, 2017.

[48] F. Ansari, R. Glawar, and T. Nemeth, “Prima: a prescriptive maintenance model for cyber-physical production systems,” *International Journal of Computer Integrated Manufacturing*, pp. 1–22, 2019.

[49] F. Ersoz, D. Merdin, and T. Ersoz, “Research of industry 4.0 awareness: A case study of turkey,” *Economics and Business*, vol. 32, no. 1, pp. 247–263, 2018.

[50] F. L. Gewers, G. R. Ferreira, H. F. de Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. d. F. Costa, “Principal component analysis: a natural approach to data exploration,” *arXiv preprint arXiv:1804.02502*, 2018.

[51] R. Usamentiaga, J. Molleda, and D. F. Garcia, "Real-time assessment of the reliability of welds in steel strips," *IEEE Transactions on Industry Applications*, vol. 46, no. 1, pp. 81–88, 2009.

[52] P. Brice and W. Jiang, "A context tree method for multistage fault detection and isolation with applications to commercial video broadcasting systems," *IIE Transactions*, vol. 41, no. 9, pp. 776–789, 2009.

[53] ONU, "Declaração de johanesburgo sobre desenvolvimento sustentável," 2002.

[54] U. G. Assembly, "Resolution adopted by the general assembly on 25 september 2015: 70/1," *Transforming our world: the*, vol. 2030, 2015.

[55] A. K. Sen, *Development as Freedom.* Alfred A. Knopf, Inc., 1 ed., 2000.

[56] A. R. Romeiro, "Sustainable development: an ecological economics perspective," *estudos avançados*, vol. 26, no. 74, pp. 65–92, 2012.

[57] P. Utting, "Business responsibility for sustainable development," tech. rep., Geneva 2000 Occasional Paper, 2000.

[58] L. Cornel and L. Mirela, "The quantification of the sustainable development at local level," *WSEAS Transactions on Business and Economics*, vol. 6, no. 5, pp. 310–319, 2008.

[59] S. C. Buarque, *Construindo o desenvolvimento local sustentável: metodologia de planejamento.* Editora Garamond, 2002.

[60] C. C. E. Alves, L. M. A. Bezerra, and A. C. da Costa Matias, "A importância da conservação/preservação ambiental da floresta nacional do araripe para a região d cariri–ceará/brasil.," *Revista Geográfica de América Central*, vol. 2, pp. 1–10, 2011.

[61] A. P. N. BANDEIRA, P. H. D. S. NUNES, and M. G. D. S. LIMA, "Gerenciamento de riscos ambientais em municípios da região metropolitana do cariri (ceará)," *Ambiente & Sociedade*, vol. 19, no. 4, pp. 65–84, 2016.

[62] K. V. Linhares, W. A. de Girão, *et al.*, "Soldadinho-do-araripe, símbolo da conservação das águas e florestas úmidas do cariri cearense," *Cadernos de Cultura e Ciência*, vol. 13, no. 2, pp. 37–50, 2015.

[63] W. de Lima Paiva and A. L. Cavalcante, "Padrão locacional da indústria cearense: algumas evidências," *Revista Econômica do Nordeste*, vol. 40, no. 3, pp. 605–618, 2009.

[64] A. M. N. Cordeiro and F. de Holanda Bastos, “Potencial geoturístico do estado do ceará, brasil,” *CULTUR: Revista de Cultura e Turismo*, vol. 8, no. 2, pp. 86–113, 2014.

[65] C. Gabrielli, “Planejamento turístico no cariri cearense: Integração e desenvolvimento responsável,” *Tour. Hosp. Int. J*, vol. 3, pp. 242–258, 2014.

[66] D. P. Lindoso, “Vulnerabilidade e adaptação da vida às secas: desafios à sustentabilidade rural familiar nos semiáridos nordestinos,” 2013.

[67] N. P. Damasceno, A. S. Khan, and P. V. P. S. Lima, “O impacto do pronaf sobre a sustentabilidade da agricultura familiar, geração de emprego e renda no estado do ceará,” *Revista de Economia e Sociologia Rural*, vol. 49, no. 1, pp. 129–156, 2011.

[68] R. Sacheto, “Agropolos: sustentabilidade para agricultura familiar,” *Inovação Uniemp*, vol. 2, no. 1, pp. 20–21, 2006.

[69] C. Populacional, “Instituto brasileiro de geografia e estatística (ibge),” 2018.

[70] I. G. de Oliveira, M. A. P. da Silva, *et al.*, “Plantas medicinais utilizadas na farmacopéia popular em crato, juazeiro e barbalha-ceará, brasil,” *Revista Brasileira de Biociências*, vol. 5, no. S1, pp. 189–191, 2007.

[71] W. Girao, “Soldadinho-do-araripe: símbolo da conservação da natureza no cariri,” *A Província*, vol. 1, pp. 39–46, 2009.

[72] S. C. Ribeiro, F. S. Ferreira, S. V. Brito, G. G. Santana, W. L. S. Vieira, R. R. Nóbrega, and W. O. Almeida, “The squamata fauna of the chapada do araripe, northeastern brazil.,” *Cadernos de Cultura e Ciência*, vol. 1, no. 1, 2009.

[73] I. J. Roberto, S. C. Ribeiro, M. M. de Sousa Delfino, and W. de Oliveira Almeida, “Reptilia, colubridae, helicops angulatus: distribution extension and rediscovery in the state of ceará,” *Check List*, vol. 5, no. 1, pp. 118–121, 2009.

[74] L. A. Souza and D. Grangeiro, “• primeiro registro de crustáceos terrestres (isopoda, oniscidea) para a chapada do araripe, ceará, brasil,” *Cadernos de Cultura e Ciência*, vol. 1, no. 1, 2009.

[75] F. Olmos, W. A. d. G. Silva, and C. G. Albano, “Aves em oito áreas de caatinga no sul do ceará e oeste de pernambuco, nordeste do brasil: composição, riqueza e similaridade,” *Papéis Avulsos de Zoologia*, vol. 45, no. 14, pp. 179–199, 2005.

[76] C. Populacional, “Instituto brasileiro de geografia e estatística (ibge),” 2018.

[77] F. W. H. Vidal, N. F. Castro, and C. C. Peiter, “The cariri stone local productive cluster,” 2008.

[78] N. F. Mochiutti, G. B. Guimarães, J. C. Moreira, F. F. Lima, and F. I. d. Freitas, "Os valores da geodiversidade: geossítios do geopark araripe/ce," 2012.

[79] C. de Tecnologia Mineral do Cariri, "Desenvolvimento de tecnologias sustentáeis e inovadoras para o arranjo produtivo local da pedra cariri no ceará," 2018.

[80] F. F. de Estudos e Projetos, "Relatǿrio de atividades finep," 2004.

[81] "Matriz problemas x soluções do apl da pedra cariri," 2018.

[82] J. C. M. Fanton, F. Ricardi-Branco, D. Dilcher, and M. E. C. B. de Oliveira, "New gymnosperm related with gnetales from the crato palaeoflora (lower cretaceous, santana formation, araripe basin, northeastern brazil): preliminary study," *Geosciences= Geociências*, vol. 25, no. 2, pp. 205–210, 2006.

[83] G. L. Fambrini, V. H. d. M. L. Neumann, C. L. de Barros, S. M. O. A. da Silva, P. C. Galm, and J. A. B. de Menezes Filho, "Análise estratigráfica da formação brejo santo, bacia do araripe, nordeste do brasil: implicações paleogeográficas," *Geologia USP. Série Científica*, vol. 13, no. 4, pp. 3–28, 2013.

[84] C. de Tecnologia Mineral do Cariri, "Apresentação do centro de tecnologia mineral do cariri," 2018.

[85] A. Smith, *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*. Harriman House Limited, 2010.

[86] W. E. Deming and D. W. Edwards, *Quality, productivity, and competitive position*, vol. 183. Massachusetts Institute of Technology, Center for advanced engineering study . . . , 1982.

[87] W. E. Deming, *The new economics for industry, government, education*. MIT press, 2018.

[88] W. A. Shewhart and W. E. Deming, *Statistical method from the viewpoint of quality control*. Courier Corporation, 1986.

[89] W. E. Deming, "Statistical adjustment of data.," 1943.

[90] D. W. Edwards, "Out of the crisis," *MIT Center for Advanced Engineering Study*, 1986.

[91] W. A. Shewhart and W. E. Deming, *Statistical method from the viewpoint of quality control*. Courier Corporation, 1939.

[92] D. C. Montgomery, *Introduction to statistical quality control*. John Wiley & Sons (New York), 2009.

[93] K. Ishikawa, *Guide to quality control.* No. TS156. I3713 1994., 1982.

[94] P. B. Crosby, *Quality Without Tears: The Art of Hassle-free Management.* A Plume Book, 1984.

[95] E. ISO, "9001: 2000," *Quality management systems-Requirements (ISO 9001: 2000)*, 2000.

[96] R. B. Caldini and N. J. Goldstein, "Social influence: Compliance and conformity," *Annual Review Psychology*, 2004.

[97] P. A. Morettin and C. d. C. Toloi, "Previsão de séries temporais," 1985.

[98] C.-W. Lu and M. R. Reynolds Jr, "Cusum charts for monitoring an autocorrelated process," *Journal of Quality Technology*, vol. 33, no. 3, pp. 316–334, 2001.

[99] J. D. Cryer and K.-S. Chan, *Time Series Analysis: With Aplications in R.* Springer, 2008.

[100] R. J. Hyndman, Y. Khandakar, *et al.*, *Automatic time series for forecasting: the forecast package for R.* No. 6/07, Monash University, Department of Econometrics and Business Statistics ..., 2007.

[101] R. C. Gentleman and G. R. Ihaka, *R Language*, 2019.

[102] S. Haykin, *Neural Networks - A Comprehensive Foundation, Second Edition.* Prentice Hall, 2 ed., 1998.

[103] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach.* Malaysia; Pearson Education Limited, 2016.

[104] F. Garrosa, Eva; Ladstätter, *Prediction of burnout : an artificial neural network approach.* Diplomica Verlag, 2008.

[105] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning the rprop algorithm," *Proceedings of the IEEE international conference on neural networks*, p. 586–591, 1993.

[106] C. Igel and M. Hüsken, "Empirical evaluation of the improved rprop learning algorithms," *Neurocomputing*, vol. 50, pp. 105–123, 2003.

[107] F. Günther and S. Fritsch, "neuralnet: Training of neural networks," *The R journal*, vol. 2, no. 1, pp. 30–38, 2010.

[108] P. R. A. Firmino, P. S. de Mattos Neto, and T. A. Ferreira, "Correcting and combining time series forecasters," *Neural Networks*, vol. 50, pp. 1–11, 2014.

[109] A. R. Lima Junior, D. A. Silva, P. S. Mattos Neto, and T. A. Ferreira, "An experimental study of fitness function and time series forecasting using artificial neural networks," in *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*, pp. 2015–2018, ACM, 2010.

[110] P. S. de Mattos Neto, F. Madeiro, T. A. Ferreira, and G. D. Cavalcanti, "Hybrid intelligent system for air quality forecasting using phase adjustment," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 185–191, 2014.

[111] G. Casella and R. L. Berger, *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA, 2002.

[112] P. S. de Mattos Neto, G. D. Cavalcanti, P. R. Firmino, E. G. Silva, and S. R. V. Nova Filho, "A temporal-window framework for modeling and forecasting time series," *Knowledge-Based Systems*, p. 105476, 2020.

[113] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[114] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[115] C. Tsallis and D. A. Stariolo, "Generalized simulated annealing," *Physica A*, vol. 233, pp. 395–406, 1996.

[116] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[117] U. Diwekar, *Introduction to Applied Optimization.* New York: Springer, 2nd ed., 2008.

[118] R. C. Dorf and R. H. Bishop, *Modern control systems.* Pearson, 2011.

[119] K. Ogata, *Modern Control Engineering.* Prentice hall India, 2010.

[120] K. Sharma, *Overview of industrial process automation.* Elsevier, 2016.

[121] M. P. Groover, "Automation, production systems, and computer-integrated manufacturing 2nd ed.," *Assembly Automation*, vol. 22, no. 3, pp. 298–299, 2002.

[122] J. P. Bentley, *Principles of measurement systems.* Pearson Education India, 1988.

[123] W. Dunn, *Fundamentals of industrial instrumentation and process control.* McGraw-Hill,, 2005.

[124] D. Placko, *Fundamentals of instrumentation and measurement.* John Wiley & Sons, 2013.

[125] A. N. S. ISA-5.1, *ANSI/ISA-5.1-2009: Instrumentation Symbols and Identification.* 2009.

[126] I. Susnea and M. Mitescu, *Microcontrollers in practice*, vol. 18. Springer Science & Business Media, 2005.

[127] Arduino, "Arduino nano." https://store.arduino.cc/usa/arduino-nano.

[128] E. Systems, "Esp8266ex, datasheet,"

[129] A. K. Solutions, *Robotics.* Jones & Bartlett Learning, 2010.

[130] L. LM193, "Lm393 dual differential comparator data sheet," *Texas Instruments.*

[131] Sawers, "Ky-028 temperatur sensor modul (thermistor),"

[132] M. Electronics, "Sct-013 datasheet,"

[133] W. Kester and A. D. I. Engineeri, *Data conversion handbook.* Newnes, 2005.

[134] T. Reenskaug, "Mvc xerox parc 1978-79," *Trygve/MVC*, 1979.

[135] I. Robinson, J. Webber, and E. Eifrem, *Graph databases.* " O'Reilly Media, Inc.", 2013.

[136] O. Vogel, I. Arnold, A. Chughtai, and T. Kehrer, *Software architecture: a comprehensive framework and guide for practitioners.* Springer Science & Business Media, 2011.

[137] I. H. Sarker and K. Apu, "Mvc architecture driven design and implementation of java framework for developing desktop application," *International Journal of Hybrid Information Technology*, vol. 7, no. 5, pp. 317–322, 2014.

[138] H. K. Abraham Silberschatz and S. Sudarshan, *Database System Concepts*, vol. 7. McGraw-Hill Education, 2020.

[139] H. E. Williams and D. Lane, *Web Database Applications with PHP and MySQL: Building Effective Database-Driven Web Sites.* " O'Reilly Media, Inc.", 2004.

[140] M. Delisle, *Mastering phpMyAdmin 3.4 for effective MySQL management.* Packt Publishing Ltd, 2012.

[141] P. Group, "Php: phpversion - manual." https://www.php.net/manual/pt_BR/function.phpversion.php.

[142] M. Widenius, D. Axmark, and K. Arno, *MySQL reference manual: documentation from the source.* " O'Reilly Media, Inc.", 2002.

[143] J. Greenspan and B. Bulger, *MySQL/PHP database applications.* John Wiley & Sons, Inc., 2001.

[144] P. Ratinaud, *IRaMuTeQ - Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*, 2014.

[145] R. Max, "Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte," *Les cahiers de l'analyse des données*, vol. 8, no. 2, pp. 187–198, 1983.

[146] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[147] A. M. B. da Silva and S. R. F. Enumo, "Description and analysis of psychological intervention with dancers by the software iramuteq," 2017.

[148] M. Tendero and C. Bazart, "Results from two complementary textual analysis software (iramuteq and tropes) to analyze social representation of contaminated brownfields," *JADT'18*, p. 771.

[149] L. da Silva Silva, N. L. Resende, C. K. C. S. Damasceno, and C. M. M. de Sousa, "Knowledge of nursing professionals about the risk management process," *Revista Prevenção de Infecção e Saúde*, vol. 5, 2019.

[150] F. Michelot and B. Poellhuber, "Foundations of information literacy trends: A textual data based analysis on international frameworks," in *The Sixth European Conference on Information Literacy (ECIL)*, p. 70.

[151] O. Blanchard, P. Le Quéau, and P. Menanteau, *The socio-economic analysis of collective behavior and decision making of co-owners to invest in thermal retrofits in private multi-family buildings.* PhD thesis, Grenoble Applied Economics Laboratory (GAEL); PACTE-Université Grenoble Alpes, 2019.

[152] J. Longhi, C. Marinica, Z. Després, and C. Plancq, "The flow of ideologies between a political figure and a militant community: A cmc corpora analysis," in *of the 6th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC-corpora 2018)*, p. 45.

[153] C. Comelli, "An analysis of drug issues coverage in french print news media using lexicometric approaches," 2019.

[154] P. G. Ranky, "Advanced machine vision systems and application examples," *Sensor Review*, vol. 23, no. 3, pp. 242–245, 2003.

[155] I. A. Reis, G. Câmara, R. Assunção, and A. M. V. Monteiro, "Suppressing temporal data in sensor networks using a scheme robust to aberrant readings," *International Journal of Distributed Sensor Networks*, vol. 5, no. 6, pp. 771–805, 2009.

[156] C. Alippi, G. Boracchi, R. Camplani, and M. Roveri, "Detecting external disturbances on the camera lens in wireless multimedia sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 11, pp. 2982–2990, 2010.

[157] T. Vladimirova, M. Fayyaz, M. N. Sweeting, and V. I. Vitanov, "A novel autonomous low-cost on-board data handling architecture for a pin-point planetary lander," *Acta Astronautica*, vol. 68, no. 7-8, pp. 811–829, 2011.

[158] G. A. Elliott, B. C. Ward, and J. H. Anderson, "Gpusync: A framework for real-time gpu management," in *2013 IEEE 34th Real-Time Systems Symposium*, pp. 33–44, IEEE, 2013.

[159] J. Fernández, L. Calavia, C. Baladrón, J. Aguiar, B. Carro, A. Sánchez-Esguevillas, J. Alonso-López, and Z. Smilansky, "An intelligent surveillance platform for large metropolitan areas with dense sensor deployment," *Sensors*, vol. 13, no. 6, pp. 7414–7442, 2013.

[160] M. M. G. Ramirez, J. C. V. Rincon, and J. F. L. Parada, "Liquid level control of coca-cola bottles using an automated system," in *2014 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pp. 148–154, IEEE, 2014.

[161] C. Ungermanns and W. Werth, "Effective knowledge exchange with modern didactic concepts," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 954–958, IEEE, 2015.

[162] F. Karray, A. Garcia-Ortiz, M. W. Jmal, A. M. Obeid, and M. Abid, "Earnpipe: a testbed for smart water pipeline monitoring using wireless sensor network," *Procedia Computer Science*, vol. 96, pp. 285–294, 2016.

[163] C. Hong, Z. Yue, W. RiChi, and W. JianDe, "Study of smart illumination control system with m2m communication technology," in *2017 International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, pp. 110–115, IEEE, 2017.

[164] S. Gharghabi, Y. Ding, C.-C. M. Yeh, K. Kamgar, L. Ulanova, and E. Keogh, "Matrix profile viii: domain agnostic online semantic segmentation at superhuman performance levels," in *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 117–126, IEEE, 2017.

[165] W. Tushar, N. Wijerathne, W.-T. Li, C. Yuen, H. V. Poor, T. K. Saha, and K. L. Wood, "Iot for green building management," *arXiv preprint arXiv:1805.10635*, 2018.

[166] C. Cho, D. D. Kim, J. Kim, J.-O. Plouchart, D. Lim, S. Cho, and R. Trzcinski, "Decomposition and analysis of process variability using constrained principal component analysis," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 1, pp. 55–62, 2008.

[167] O. J. Mengshoel, B. Iannucci, and A. Ishihara, "Mobile computing: Challenges and opportunities for autonomy and feedback," in *Presented as part of the 8th International Workshop on Feedback Computing*, 2013.

[168] K. P. Veerakumari and R. Resmi, "Evaluation of continuous sampling plan (csp-5) parameters using gert technique and matlab," *Journal of Applied Mathematics, Statistics and Informatics*, vol. 13, no. 2, pp. 63–75, 2017.

[169] K. Sobhana, D. R. S. Prasad, and R. K. Kumar, "Comparative analysis of burr type iii with pareto type ii model using sprt: Order statistics," *International Journal of Innovative Computer Science & Engineering*, vol. 4, no. 3, pp. 132–144, 2017.

[170] A. Dairi, F. Harrou, M. Senouci, and Y. Sun, "Unsupervised obstacle detection in driving environments using deep-learning-based stereovision," *Robotics and Autonomous Systems*, vol. 100, pp. 287–301, 2018.

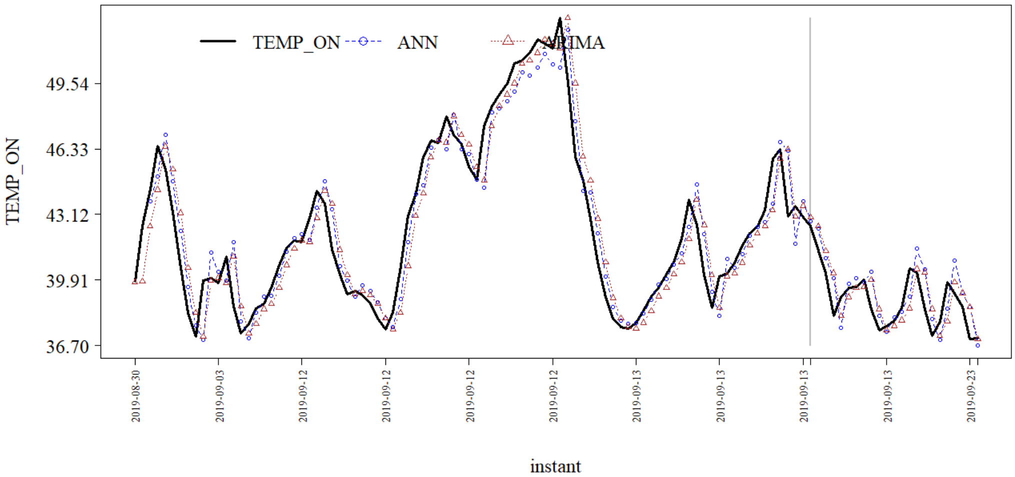
[171] A. Dairi, F. Harrou, Y. Sun, and M. Senouci, "Obstacle detection for intelligent transportation systems using deep stacked autoencoder and k-nearest neighbor scheme," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5122–5132, 2018.

[172] Y. Xiang, S. Gubian, B. Suomela, and J. Hoeng, "Generalized simulated annealing for global optimization: The gensa package.," *R Journal*, vol. 5, no. 1, 2013.

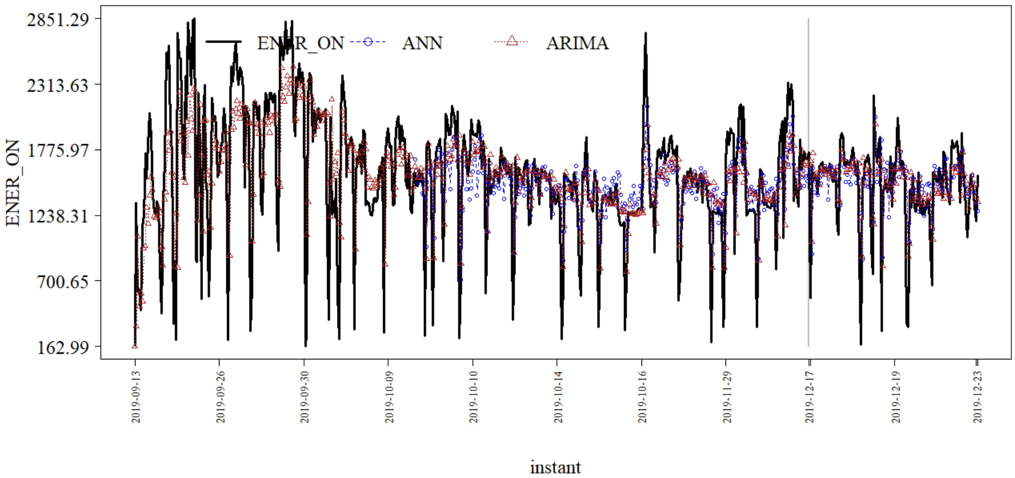
A Appendix

A.1 Single Models Forecasting

Figure 25 – Single forecasting for machine turned-on period.



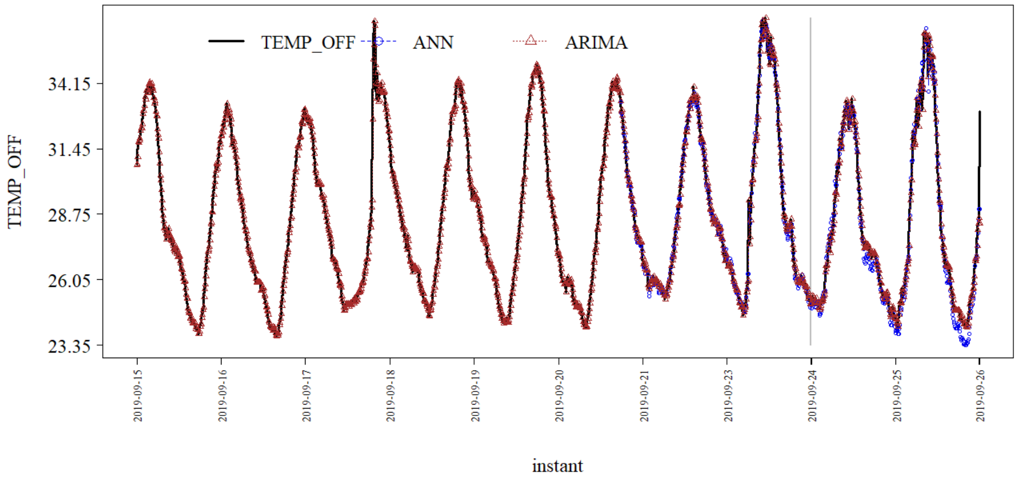
(a)



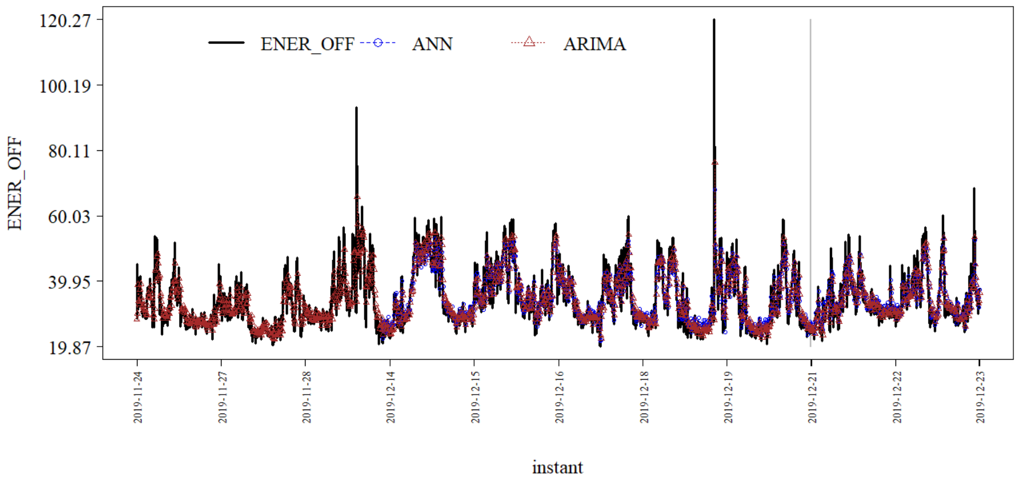
(b)

Source: The author

Figure 26 – Single forecasting for machine turned-off period.



(a)



(b)

Source: The author

A.2 Performance Metrics

Metric	Worst	Best	ANN	ARIMA	cMV
MSE	ARIMA	cMV	116825.422	199699.314	116775.835
MAPE	ARIMA	cMV	0.295	0.376	0.295
ARV	cMV	ARIMA	2.874	2.216	2.883
ID	cMV	ARIMA	0.379	0.336	0.379
Theil	ARIMA	cMV	0.702	0.750	0.701
WPOCID	ARIMA	cMV	0.523	0.571	0.519
Reg_Intercept	ARIMA	ANN	-61.502	240.897	-63.904
Reg_Slope	ARIMA	ANN	1.041	0.859	1.043
Reg_WR2	ANN	ARIMA	0.726	0.699	0.726
n.MSE	ARIMA	cMV	0.001	1.000	0.000
n.MAPE	ARIMA	cMV	0.003	1.000	0.000
n.ARV	cMV	ARIMA	0.986	0.000	1.000
n.ID	cMV	ARIMA	0.994	0.000	1.000
n.Theil	ARIMA	cMV	0.006	1.000	0.000
n.WPOCID	ARIMA	cMV	0.071	1.000	0.000
n.Reg_Intercept	ARIMA	ANN	0.000	1.000	0.013
n.Reg_Slope	ARIMA	ANN	0.000	1.000	0.016
n.Reg_WR2	ANN	ARIMA	1.000	0.000	0.987
n.Mean	ARIMA	cMV	0.340	0.667	0.335
n.Sd	ARIMA	ANN	0.490	0.500	0.495

Table 12 – Performance of the forecasting models (ANN, ARIMA, cMV) when predicting ENER_ON time series (Training phase).

Metric	Worst	Best	ANN	ARIMA	cMV
MSE	ARIMA	cMV	1.461	1.961	1.459
MAPE	ARIMA	cMV	0.022	0.026	0.021
ARV	ARIMA	ANN	0.105	0.132	0.106
ID	ARIMA	ANN	0.025	0.033	0.025
Theil	ARIMA	cMV	0.773	1.000	0.772
WPOCID	ARIMA	ANN, cMV	0.276	0.281	0.276
Reg_Intercept	ARIMA	ANN	-0.244	2.735	-0.355
Reg_Slope	ARIMA	ANN	1.006	0.937	1.008
Reg_WR2	ARIMA	cMV	0.089	0.118	0.089
n.MSE	ARIMA	cMV	0.004	1.000	0.000
n.MAPE	ARIMA	cMV	0.026	1.000	0.000
n.ARV	ARIMA	ANN	0.000	1.000	0.009
n.ID	ARIMA	ANN	0.000	1.000	0.001
n.Theil	ARIMA	cMV	0.002	1.000	0.000
n.WPOCID	ARIMA	ANN, cMV	0.000	1.000	0.000
n.Reg_Intercept	ARIMA	ANN	0.000	1.000	0.045
n.Reg_Slope	ARIMA	ANN	0.000	1.000	0.045
n.Reg_WR2	ARIMA	cMV	0.005	1.000	0.000
n.Mean	ARIMA	ANN	0.004	1.000	0.011
n.Sd	cMV	ARIMA	0.009	0.000	0.019

Table 13 – Performance of the forecasting models (ANN, ARIMA, cMV) when predicting TEMP_ON time series (Training phase).

Metric	Worst	Best	ANN	ARIMA	cMV
MSE	cMV	ANN	37.797	37.853	38.630
MAPE	ANN	cMV	0.113	0.112	0.112
ARV	ANN	ARIMA	0.672	0.647	0.653
ID	ARIMA	ANN	0.138	0.145	0.138
Theil	ARIMA	ANN	0.753	0.807	0.769
WPOCID	cMV	ANN	0.610	0.624	0.631
Reg_Intercept	ARIMA	ANN	-0.161	2.854	1.280
Reg_Slope	ARIMA	ANN	1.005	0.916	0.963
Reg_WR2	ARIMA	ANN	0.398	0.428	0.406
n.MSE	cMV	ANN	0.000	0.068	1.000
n.MAPE	ANN	cMV	1.000	0.721	0.000
n.ARV	ANN	ARIMA	1.000	0.000	0.266
n.ID	ARIMA	ANN	0.000	1.000	0.030
n.Theil	ARIMA	ANN	0.000	1.000	0.306
n.WPOCID	cMV	ANN	0.000	0.640	1.000
n.Reg_Intercept	ARIMA	ANN	0.000	1.000	0.415
n.Reg_Slope	ARIMA	ANN	0.000	1.000	0.404
n.Reg_WR2	ARIMA	ANN	0.000	1.000	0.265
n.Mean	ARIMA	ANN	0.222	0.714	0.410
n.Sd	ANN	cMV	0.441	0.409	0.364

Table 14 – Performance of the forecasting models (ANN, ARIMA, cMV) when predicting ENER_OFF time series (Training phase).

Metric	Worst	Best	ANN	ARIMA	cMV
MSE	ARIMA	ANN	0.083	0.084	0.084
MAPE	ARIMA	cMV	0.006	0.006	0.006
ARV	ARIMA	ANN	0.008	0.008	0.008
ID	ARIMA	ANN	0.002	0.002	0.002
Theil	ARIMA	ANN	0.746	1.000	0.746
WPOCID	ANN	ARIMA	0.375	0.346	0.354
Reg_Intercept	ARIMA	ANN	0.001	0.099	0.037
Reg_Slope	ARIMA	ANN	1.000	0.996	0.998
Reg_WR2	ARIMA	cMV	0.008	0.008	0.008
n.MSE	ARIMA	ANN	0.000	1.000	0.131
n.MAPE	ARIMA	cMV	0.969	1.000	0.000
n.ARV	ARIMA	ANN	0.000	1.000	0.349
n.ID	ARIMA	ANN	0.000	1.000	0.184
n.Theil	ARIMA	ANN	0.000	1.000	0.001
n.WPOCID	ANN	ARIMA	1.000	0.000	0.276
n.Reg_Intercept	ARIMA	ANN	0.000	1.000	0.366
n.Reg_Slope	ARIMA	ANN	0.000	1.000	0.467
n.Reg_WR2	ARIMA	cMV	0.038	1.000	0.000
n.Mean	ARIMA	cMV	0.223	0.889	0.197
n.Sd	ANN	cMV	0.432	0.333	0.177

Table 15 – Performance of the forecasting models (ANN, ARIMA, cMV) when predicting TEMP_OFF time series (Training phase).

B Appendix

B.1 Entity-Relationship Diagram (ERD)

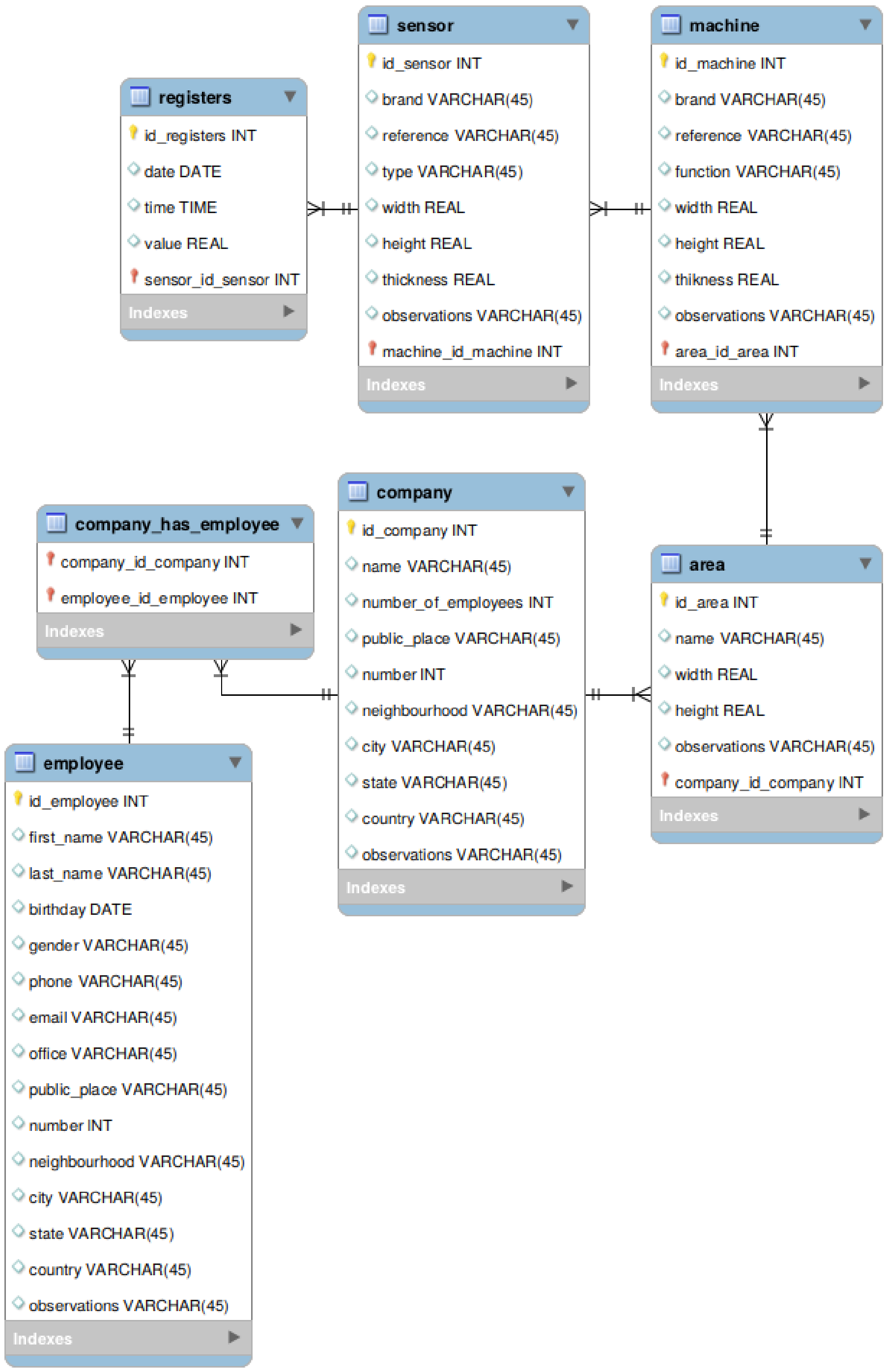


Figure 27 – Entity-Relationship Diagram (ERD) describing the relationship of the elements that can involved in any study cases with your respective cardinalities