TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2020

68

Nurseda Yildirim Yurusen

Wind Farm Management Decision Support Systems For Short Term Horizon

Departamento

Centro de Investigación de Recursos y Consumos Energéticos (CIRCE)

Director/es

Melero Estela, Julio Javier

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

WIND FARM MANAGEMENT DECISION SUPPORT SYSTEMS FOR SHORT TERM HORIZON

Autor

Nurseda Yildirim Yurusen

Director/es

Melero Estela, Julio Javier

UNIVERSIDAD DE ZARAGOZA

Centro de Investigación de Recursos y Consumos Energéticos (CIRCE)

2019

WIND FARM MANAGEMENT DECISION SUPPORT SYSTEMS FOR SHORT TERM HORIZON

Dissertation in partial fulfillment of the requirements for the degree of

Programa de Doctorado Energías Renovables y Eficiencia Energética

Instituto Universitario de Investigación Mixto CIRCE
UNIVERSIDAD DE ZARAGOZA

NURSEDA YILDIRIM YÜRÜŞEN

October 1, 2019

WIND FARM MANAGEMENT DECISION SUPPORT SYSTEMS FOR SHORT TERM HORIZON

Dissertation in partial fulfillment of the requirements for the degree of

Programa de Doctorado Energías Renovables y Eficiencia Energética
Universidad de Zaragoza
Campus Rio Ebro

Approved by:

Supervisor, Julio Javier Melero

Examiner Committee,

International Reviewers,

October 1, 2019

List of Publications

1. N Y Yürüşen, Julio J Melero, Importance Ranking for Revenue of a Wind Farm Case Study: Spain. (Poster Presentation in Wind Europe conference 2016)

2. N Y Yürüşen, Julio J Melero, Probability density function selection based on the characteristics of wind speed data. (Poster Presentation in TORQUE conference 2016) (Full paper in Journal of Physics: Conference Series)

   DOI:10.1088/1742-6596/753/3/032067

3. Elana Gonzalez, Emmanouil Nanos, Helene Seyr, Laura Valldecabres, Nurseda Y Yürüşen, Ursula Smolka , Michael Muskulus , Julio J. Melero, Key performance indicators for wind farm operation and maintenance. (Poster Presentation in 14th Deep Sea Offshore Wind R&D Conference in 2017) (Full paper in Energy Procedia)

   DOI:10.1016/j.egypro.2017.10.385

4. N Y Yürüşen, Maik Reder, Failure Event Definitions & their Effects on Survival and Risk Analysis of Wind Turbines. (Oral Presentation in ESREL 2017) (Full paper in Safety and Reliability – Theory and Applications, CRC Press Taylor & Francis)

   DOI:10.1201/9781315210469-93

5. N Y Yürüşen, Jannis Tautz-Weinert, Simon J Watson, Julio J. Melero, The Financial Benefits of Various Catastrophic Failure Prevention Strategies in a Wind Farm: Two market studies (UK-Spain). (Oral Presentation in WindEurope 2017) (Full paper in Journal of Physics: Conference Series)

   DOI: 1088/1742-6596/926/1/012014

6. N Y Yürüşen, Maik Reder, Julio J Melero, Data-driven learning framework for associating weather conditions and wind turbine failures, Reliability Engineering and System Safety 169 pp. 554–569, ISSN: 09518320, 2018.

   DOI: 10.1016/j.ress.2017.10.004

7. N Y Yürüşen, Jannis Tautz-Weinert, Simon J Watson, Julio J. Melero, Sensitivity study of a wind farm maintenance decision - a performance and revenue analysis, Renewable Energy, 132, 93-105,2019.

   DOI:10.1016/j.renene.2018.07.110

8. N Y Yürüşen, Paul N Rowley, Simon J Watson, Julio J. Melero, Advanced wind turbine maintenance scheduling, Submitted to Reliability Engineering and System Safety Journal.

9. N Y Yürüşen, Julio J. Melero, Timely decision making for an operation and maintenance visit. (Oral Presentations in Wind Energy Science Conference 2019, Cork, Ireland).

10. N Y Yürüşen, Julio J Melero, Wind farm maintenance decisions, key factors contribution: electricity market price (In progress).

Abstract

Wind energy is one of the fastest growing energy sources and its technology maturity level is already higher than the majority of other renewables. Therefore, many countries started to change their financial support policies in an unfavourable way for the wind energy. This unsubsidised new era forces the wind industry to re-visit its expenditure components and to make improvements in operating strategies in order to minimise operational and maintenance (O&M) costs. The classical maintenance strategies focus on a year advanced programming of calendar based maintenance visits and corrective interventions. In this classical approach the maintenance programming flexibility is quite limited, since this kind of programming ignores dynamic environment of the wind farm and real time data-driven indicators. Then, downtimes, and corresponding revenue losses, due to wind turbine inaccessibility occur because wind turbines are exposed to challenging dynamic environmental conditions and located in remote areas. Low accessibility is one of the predominant problems, and remote control not always solves the problems. The cost optimal O&M strategies for the wind energy must consider condition based maintenance and a timely programming of wind turbine visit. Thus, an elaborate and flexible approach, which is capable of considering condition and accessibility of wind turbines using meteorological measurements and operational records is highly needed for the wind farm O&M management. The core objective of this thesis is the investigation of decision making processes in wind farm management, and the generation of Decision Support Systems (DSSs) for O&M of wind farms. In order to develop practical and feasible DSSs, the research is conducted prioritising data-driven approaches. There still exist various inefficiently used data sources in an operational wind farm, therefore there is a room for an improvement to use efficiently available data. Generally, in a wind farm, two types of condition monitoring data can be collected as online inspection and offline inspection data. Online inspec-

tion data can be obtained from both condition monitoring system (CMS) and Supervisory Control and Data Acquisition (SCADA). CMS data require an additional investment in the turbines while, on the contrary, SCADA data are already available in the turbines. As a third source, offline inspection data consist of the records of all O&M visits to the wind farm, which are available but poorly recorded. In this study, the answer for the question of how to change a classical O&M strategy to an enhanced one using only the existing data sources without the need for an additional investment is searched.

Firstly, analysis of key factors influencing in wind farm maintenance decisions is performed. In this regard, exploratory data analysis was considered to understand the monthly seasonality and the dependencies of day ahead hourly electricity market price, which is one of the decisive parameters for the wind farm revenue. Then, the connection between wind turbine failures, atmospheric variables and downtime is studied in order to provide additional information to a maintenance team and a maintenance planner for the intervention day. For the first part, well structured and analysed electricity market price, electricity generation and demand data are needed. Therefore, the existing databases are reviewed for the case countries and a relevant analysis period is chosen. The electricity market data can be easily interpreted as time series data. To exhibit the characteristics of different electricity markets, various time series comparison tools are combined as an analysis guideline. By using this guideline, the drivers of the electricity market price are summarised for each case country. For the second part, available atmospheric and failure data for the relevant wind turbine components are gathered and combined. Then, convenient approaches among unsupervised learning models are selected. By combining the available tools and considering the needed information level for different purposes, the failure rules of prior to failure occurrence per month, in hours and in ten minutes increments are mined.

Then, what-if analysis for revenue tracking of maintenance decisions is performed in order to generate a DSS for the evaluation of the major maintenance decisions taken in wind farms. To this purpose, the impact of country dynamics and subsidy frameworks considering the electricity market conditions are modelled. The impact of the intervention timing is analysed and the sensitivity of financial losses to environmental causes of under performance are estimated.

Finally, generation of decision support tool for planning of a maintenance day is studied to provide a useful maintenance DSS for in situ applications.

The safe working rules considering the wind speed constraints for the accessibility to the wind turbine are reviewed taking into account the turbine manufacturer's O&M guidelines. The characteristics of the maintenance visits are summarised. Wind turbine accessibility trials using numerical weather prediction forecasting techniques for wind speed variable and synthetic forecasts for wind speed and wind gust variables are presented. An intervention decision pool considering safe working rules is generated, containing a list of plans capable of providing the optimal sequence of various tasks and ranked for revenue prioritised timing.

This work has been part of the "Advanced Wind Energy Systems Operation and Maintenance Expertise" project, a European consortium with companies, universities and research centres from the wind energy sector. Parts of this work were developed in collaboration with other fellows in the project.

Resumen

La energía eólica es una de las fuentes energéticas con mayor tasa de crecimiento habiendo alcanzado una madurez tecnológica superior al resto de fuentes renovables. Por este motivo, muchos países han cambiado sus políticas de subvenciones legislando de forma no favorable a la energía eólica. Esta nueva era sin subvenciones está forzando a la industria eólica a reevaluar sus costes y realizar mejoras en las estrategias de operación para minimizar los costes de operación y mantenimiento (O+M). Las estrategias clásicas de mantenimiento se centran en programación de las actividades de forma anual o visitas periódicas, así como acciones correctivas. Con esta aproximación clásica, la flexibilidad en la programación de las tareas de mantenimiento está bastante limitada puesto que se está ignorando la dinámica existente en el emplazamiento de los parques y la información basada en registros en tiempo real. Los tiempos de parada y, por tanto, las correspondientes pérdidas de beneficios, aumentan debido a la inaccesibilidad de las turbinas que están expuestas a entornos hostiles y de difícil acceso. La baja accesibilidad es uno de los problemas predominantes y el control remoto no siempre resuelve los problemas. Las estrategias de operación y mantenimiento óptimas en energía eólica tienen que considerar el mantenimiento basado en la condición y la programación periódica de las visitas a las turbinas. Por tanto, queda claro que una aproximación elaborada y flexible, capaz de considerar la condición de las turbinas y su accesibilidad, utilizando medidas meteorológicas y registros de operación, puede ser muy útil para la gestión del O+M de los parques eólicos. El objetivo principal de esta tesis es la investigación de procesos de toma de decisión en la gestión de parques eólicos y la generación de sistemas de ayuda para la toma de decisiones para la O+M de parques eólicos. Para desarrollar sistemas de toma de decisión prácticos y factibles, la investigación se ha realizado dando prioridad a aproximaciones basadas en datos. Todavía existen ineficiencias en el uso que se da a los datos generados en los par-

ques eólicos en operación y, por tanto, queda espacio para mejora en su uso eficiente. Generalmente, en un parque eólico, la información relativa la monitorización de estado genera dos tipos de datos, datos registrados on-line de forma remota y datos registrados en la turbina u off-line. Los datos on-line pueden proceder a su vez de dos fuentes; sistemas de monitorización de estado, que suelen requerir una inversión extra para su instalación, o el propio sistema SCADA (Supervisory Control and Data Acquisition), que suele estar presente en las turbinas y los parques eólicos. Una tercera fuente de datos, disponibles "off-line" son los registros de las visitas de O+M al parque eólica, que suelen estar disponibles, pero generalmente con baja o dudosa calidad.

En este estudio, se busca la respuesta a cómo cambiar la estrategia clásica de O+M a una mejorada utilizando solamente las fuentes de información existentes sin necesidad de inversión adicional.

En primer lugar, se realiza el análisis de los factores clave que influyen en las decisiones de mantenimiento de parques eólicos. Para ello, se aplica el análisis exploratorio de datos para entender la estacionalidad mensual y las dependencias del precio horario de la electricidad a un día, que es uno de los parámetros decisivos para la cuenta de beneficios de un parque eólico. Posteriormente se estudia la conexión existente entre los fallos en las turbinas, las variables atmosféricas y los tiempos de parada con el objetivo de disponer de información adicional, relativa al día de la intervención, que puede ser utilizada tanto por el planificador de las tareas de mantenimiento como por el propio equipo de intervención. Para la primera parte es necesario analizar datos bien estructurados del precio de mercado de la electricidad, generación y demanda eléctricas. Se han revisado las bases de datos existentes para los países estudiados y se ha elegido un periodo de análisis relevante. El precio de mercado de la electricidad se puede intepretar fácilmente como una serie temporal de datos. Para mostrar las características de los diferentes mercados de electricidad estudiados se utilizaron varias herramientas de comparación de series temporales junto con una guía para el análisis. Utilizando dicha guía, se han identificado los componentes más importantes responsables del precio del mercado eléctrico para cada país estudiado.

En la segunda parte, se han recopilado y combinado datos meteorológicos y atmosféricos disponibles junto con datos de fallos de los componentes más importantes de las turbinas. Se han aplicado modelos de aprendizaje supervisado y no supervisado que han permitido obtener reglas de fallo para sucesos previos a un fallo por mes, en horas y en incrementos de diez minutos. A continuación, se realiza un análisis de hipótesis que considera los beneficios de las decisiones de mantenimiento con el objetivo de generar un sistema de apoyo a la toma de decisiones para la evaluación de las grandes decisiones de mantenimiento tomadas en parques eólicos. Para ello, se incluyen en el modelo el impacto de la dinámica del país y los distintos modelos de subvenciones existentes en los mercados de la electricidad considerados. Se analiza también el impacto de la fecha de ejecución de la intervención y se estima la sensibilidad de las pérdidas financieras a causas ambientales de bajo rendimiento.

Finalmente, se estudia la forma de generar una herramienta de apoyo a la toma de decisiones para la planificación del mantenimiento diario y se desarrolla una herramienta de apoyo a la toma de decisiones para aplicaciones in-situ. Se han revisado las normas de seguridad en el trabajo que regulan las restricciones, por velocidad de viento, para la accesibilidad a las turbinas, teniendo en cuenta también las directrices de O+M de los fabricantes. Se han revisado y resumido las características de las visitas de mantenimiento. Utilizando tanto técnicas numéricas de predicción meteorológica para la velocidad de viento como predicciones sintéticas simuladas para velocidad de viento y ráfagas de viento, se han generado pruebas de acceso a la turbina. Combinando las normas de seguridad en el trabajo con las pruebas de accesibilidad, se ha generado un conjunto de decisiones posibles que contiene una lista de planes de intervención, con la secuencia óptima a seguir para ejecutar las tareas asignadas, clasificados y priorizados temporalmente en función del beneficio económico.

Este trabajo forma parte del proyecto AWESOME, "Advanced Wind Energy Systems Operation and Maintenance Expertise", un consorcio europeo formado por investigadores de universidades, compañías privadas y centros de investigación del sector de la energía eólica. Parte del trabajo se desarrolló en colaboración con otros compañeros del proyecto.

First of all, my favourite word in Spanish is *vale*.

En el primer lugar, mi palabra favorita en Castellano es *vale*.

Three poets, three languages:

"Kitap rüzgar olmalı, perdeyi kaldırmalıdır."
Nazım Hikmet Ran

"Have we, or have we not, an analogical right to the inference that this perceptible Universe - that this cluster of clusters - is but one of a series of clusters, the rest of which are invisible through distance - through the diffusion of their light being so excessive, ere it reaches us, as not to produce upon our retinas a light-impression."
Edgar Allan Poe

"El viento es un caballo: óyelo cómo corre por el mar, por el cielo."
Pablo Neruda

ACKNOWLEDGEMENTS

I would like to thank my supervisor Julio J. Melero for his guidance and support every now and then throughout the course of this research. I am aware that I was a student from hell with many grammatical errors and disorganised figures for a perfectionist supervisor. Then, I would like to thank specially to AWESOME fellows Jannis Tautz-Weinert and Maik Reder.
Thanks also go to my friends and colleagues and staff for making my time at CIRCE a great experience. Specially, I would like to thank to Ana Patricia Talayero Navales, Roberto Lázaro Gastón, Elena Gonzalez and Enrique Telmo Martínez.
Particular gratitude is expressed to Javier Gracia Bernal, Lucas García Pérez, Marta Heras Heras and Miguel Angel Hernández Lucas for their friendly guidance during my industrial secondment and even after its finalisation when they continued to respond my never ending questions patiently.
I learnt a lot and spent a great time in Loughborough. I would like to express my very great appreciation for a productive academic secondment to Simon J. Watson, Paul N. Rowley, Sibel Yılmaz and Farhad Anvari-Azar.
Then, I must send my gratitudes to researchers who kindly responded questions on the models and techniques that I got lost: Bertrand Kerres, Hasan Engin Duran, Bahri Uzunoğlu, Mehmet Remzi Doğar, JQ Veenstra, Paulo Cortez, Daniel C. Bowman, J.M. Yusta , Gang (John) Xie and all other academicians who kindly answered to my e-mails, questions.
Finally, I would like to thank a lot to my family and my husband for their unlimited support and understanding. I would like to dedicate this study to my husband, my mother and my sister for their encouragement.

Thanks to the AWESOME project, I managed to survive during my PhD. This thesis study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642108, known as the AWESOME consortium.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background, justification and aim

Modern society has a high and continuous electricity demand. In order to be able to address this demand, there exist many electricity generation technologies. While electricity generation with conventional energy sources (coal, natural gas, etcetera.) still dominates the electricity market, these technologies create high carbon footprints and have unfavourable impacts on the environment. Therefore, electricity generation with renewable energy sources is being prioritised and supported by EU countries. As one of the main renewable energy contributors, wind energy is capable to respond to the needs for a secure, clean, sustainable and cheap power supply. If the carbon footprints of electricity generation technologies are considered, wind energy is one of the lowest [1]. Moreover, the levelized cost of electricity of the wind is one of the fast decreasing electricity generation technologies [2, 3].

The wind industry has approximately 1600 jobs per each produced TWh by 2018 [4], and, according to employment statistics and estimations, the operation and maintenance (O&M) sector will increase its share drastically (from 20% in 2013 to 40% in 2029) [5]. On the contrary, it is anticipated that other sub-sectors will decrease their employment volume due mainly to the technological developments. Then, a considerable number of wind energy jobs are already in the O&M sector which, being one job intensive sector, is expected to create even more employments in the future.

The classical approaches for O&M, such as planning a maintenance day in a bank holiday, or managing electricity generation and its auction by taking into account demand forecasts, are sufficient for the generators which can be switched on/off according to demand. However, efficient O&M of wind farms is a more challenging task, since wind turbines are exposed to environmental conditions. Low accessibility is one of the predominant problems, and remote control not always solves the problems. Thus, an elaborate and flexible approach is highly needed for the wind farm O&M management.

Nowadays, the renewable energy sector is evolving with the cycle of technology improvements and developments for cost reduction. In a recent report published by the International Renewable Energy Agency, one of the main drivers of cost reductions for renewable energy is stated as "optimised operation and maintenance (O&M) practices and the use of real-time data to allow improved predictive maintenance, reducing O&M costs and generation loss from planned and unplanned outages" [6]. The same document also highlights the fact that "onshore wind is one of the most competitive sources of new generation capacity" and the global O&M market for wind power is estimated to be 2.25 times more than its market size in 2016 by 2026 [6].

For a such dense O&M labouring sector, risks due to working environment and importance of in advance scheduling must be considered and studied wisely. O&M of wind turbines requires working at high altitudes in harsh environmental conditions. In this sector, making decisions in short term related to in situ interventions is a common practice. In order to minimize the risk of working in unfavourable conditions and the revenue losses due to unoptimized scheduling, health and safety rules prioritized data-based decision support systems are needed.

Management of wind farms requires various types of critical decisions, which are interdependent on each other. The wind farm maintenance optimisation differs from other wind farm focused data driven analyses, which include performance, availability and failure root cause studies. As indicated in the complexity level chart of the IEA Wind Task 33 [7], the wind farm maintenance optimisation is ranked as the most challenging category because of the amount of required input data and the variety of the mandatory analyses [8].

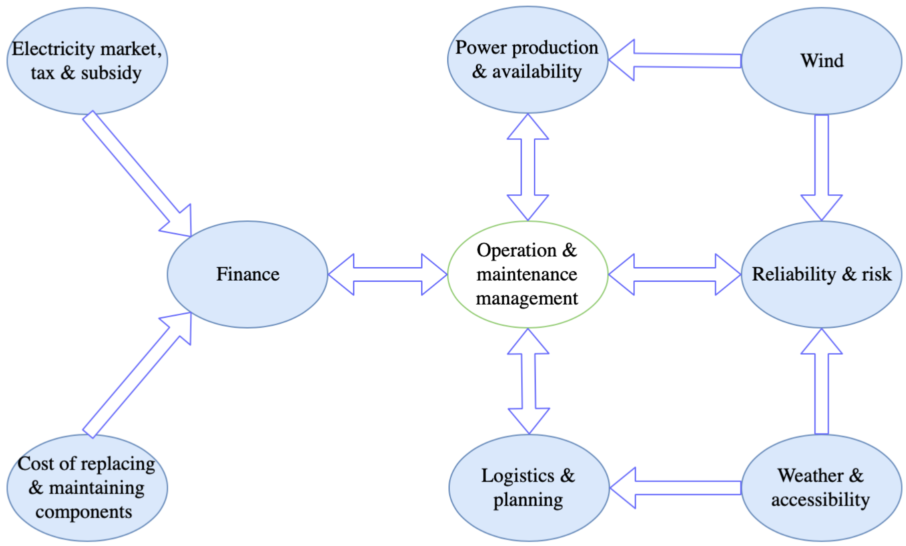


Figure 1.1: Management decisions' dependencies.

Because the decision making mechanism is a very complex one, the problem statement is done following a bottom-up approach. In contrast to existing simulation based studies, the connection to real data is always prioritised, and available data and in situ observations are grouped for the management process as given in Figure 1.1.

O&M costs can reach up to 30% of overall life cycle cost break down of a wind turbine [9]. A recent survey reported that by 2030, wind energy experts are estimating a reduction by 24% - 30% in levelised cost of energy (LCOE) and, to achieve this value, one of the needs is to accomplish 9% decrease in onshore and offshore O&M expenses considering the 2014 baseline scenario [10]. It can be said that there is room for an improvement in O&M management and technology, which will be worth a 9% average of O&M costs [10].

Thus, increasing interest can be found in the literature for providing decision support systems (DSSs), which target O&M cost minimisation [11, 12].

However, real data prioritisation in DSSs requires a focus on specific aspects. Therefore, sequential and single decision support systems are proposed in this PhD study. The primary research questions investigated are listed below.

- How should the available O&M data be cleaned and organised?
- What are the contributing factors to a maintenance decision?
- How should a maintenance day be planned?
- What are the effects of a maintenance decision on the revenue of a wind farm?

1.2 Objectives

In this section, the objectives are presented. The core objective of this study is the investigation of decision making processes in wind farm management, and the generation of DSSs for O&M of wind farms. For the consecution of the core objective, the next specific objectives are proposed.

Regarding the analysis of key factors influencing in wind farm maintenance decisions, reviewed in Chapter 3, the specific objectives are:

- To provide information on preferable maintenance months, from revenue perspective, to O&M engineers, who can consider the dependencies of the electricity market price as information sources for maintenance management decisions in one year in advance O&M plans.
- To investigate the contributing factors to a maintenance decision and the dependencies of hourly electricity market price considering country specific features.
- To provide wind turbine health status related information to O&M engineers without demanding frequent wind farm maintenance visits resulting in costly downtimes.
- To find out the failure rules of prior to failure occurrence per month, in hours and in ten minutes increments.

The what-if analysis for revenue tracking of maintenance decisions studied in Chapter 4, have the next specific objectives:

- To investigate the financial outcome of a maintenance decision.

- To generate what-if scenarios for the financial outcomes of the alternative maintenance decisions.

- To analyse the impact of the timing, environmental causes and the country specific policies for the financial outcomes of the alternative maintenance decisions.

Finally, the development of a decision support tool for planning of a maintenance day, covered in Chapter 5, is done based on the next specific objectives:

- To design a maintenance planning DSS for in situ applications, which can reduce scheduling downtimes and corresponding costs.

- To test the sensitivity of the maintenance planning DSS using the range of input wind speed forecasts.

The research and work conducting to this PhD thesis was done at Research Centre for Energy Resources and Consumption (CIRCE)- University of Zaragoza, CETASA in Soria and Centre for Renewable Energy Systems Technology (CREST) Wolfson School of Mechanical, Electrical and Manufacturing Engineering in Loughborough University.

1.3 Literature review

In this section, besides the introduction of the research field, the existing information on the wind turbines O&M is reviewed and evaluated in order to clarify the contributions of this thesis. Here, existing wind farm DSSs in the literature, maintenance plans & prerequisites of a service (information obtained from wind farm owner internal procedures: industrial secondment, personal communications etcetera) and wind farm maintenance decisions and their sensitivity (using both the literature and the information obtained from wind farm owner internal procedures) are reviewed in the subsequent sections.

1.3.1 Introduction

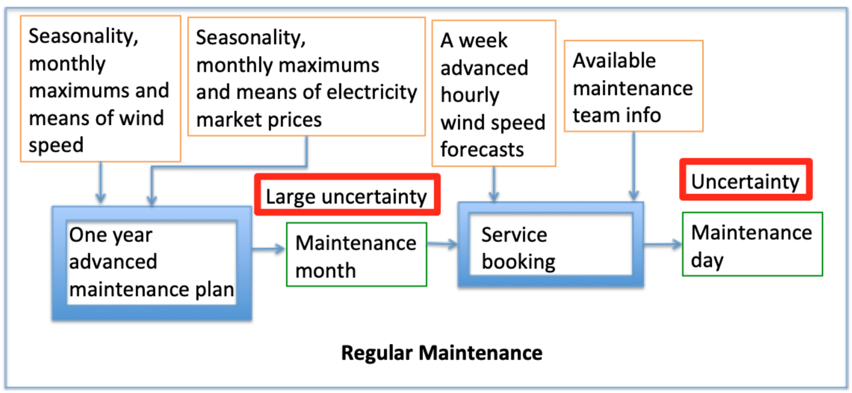
Initial attempts to research wind farm DSSs can be traced back to the end of 1990's [13, 14]. However, these tools were not designed; for operating wind farms as their main purpose was simply to support the planning phase. O&M costs were held constant in the calculations, rather than considered as dynamic. In parallel with the increase in global cumulative installed wind energy capacity (7.6 GW in 1997 to 539.2 GW in 2017 [15, 16]), the need for DSSs with various purposes increased over time.

A very comprehensive review study, providing a summary for offshore wind energy decision support systems emphasising O&M, exists in the literature [17], although, in most of the referred DSSs' detailed descriptions and software information are confidential. Unfortunately, the literature is more scarce when the maintenance strategy aspect for short term decision making is searched. As an example, 54 DSSs are reviewed in [17] and only one of them covers this aspect [18]. Remaining 53 DSSs are designed for either long term or life cycle analysis.

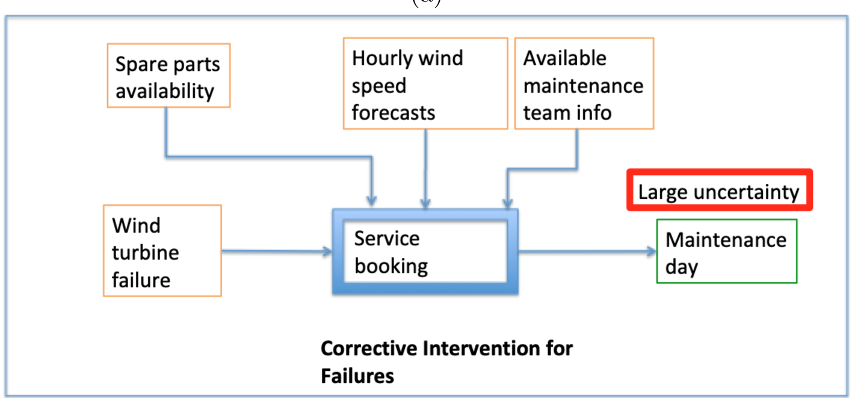
In later offshore wind energy DSS studies focusing on short term decisions, the common goal is to develop applications for 'opportunistic' maintenance [19–21]. When the literature for land based (or onshore) wind farms is reviewed, it is also observed that DSSs applications for opportunistic maintenance attract the attention of the researchers [11, 22, 23]. Opportunistic maintenance is defined in the literature as " whenever a failure occurs in the wind farm, the maintenance team is sent to the site to perform corrective maintenance, and take this opportunity to simultaneously perform preventative maintenance on the other components in the failed turbines and the running turbines and their components which show relatively high risks " [19]. Performing such an intervention requires sequential and simultaneous decision making. Thus, apart from environmental conditions, holistic, instant information of wind turbines and maintenance status is required in order to perform the opportunistic maintenance. The appropriate indicators must be selected and kept up-to-date.

1.3.2 Maintenance plans & prerequisites of a service

Wind turbines can be subjected to corrective and preventative actions, and long term maintenance policies must address both of them. The corrective maintenance policy is based on services after fault recognition, whereas preventative maintenance can be performed according to calendar based predetermined intervals. These are biannual, annual, biennial and quinquennial periods [24–28]. The number of tasks and the duration of scheduled maintenance differ from case to case and depend on the technology, model, and capacity of the wind turbine. Furthermore, even the working practices of the maintenance provider may change these variables, making difficult to precisely estimate the number and duration of tasks. Scientific literature and manufacturer maintenance guides provide numbers that vary from few lubrication tasks with a duration of 3 hours to 16 different tasks lasting up to 18 hours [27, 28] for the biannual maintenance visit. This makes the development of detailed maintenance plans difficult, especially where the number of tasks and their duration have to be accounted for.



(a)



(b)

Figure 1.2: Maintenance scheduling procedure (a) preventative policy and (b) corrective intervention.

Preventative maintenance is usually planned a year in advance on an annual basis for onshore wind farms [29]. A typical flow diagram of this type of advanced planning is shown in Figure 3.1a where account needs to be taken of the weather and electricity market prices as well as the availability of a maintenance team. Requirements of preventative maintenance can also be seen even when planning corrective actions as shown in Figure 3.1b. Both, regular and corrective maintenance, involve uncertainties, particularly concerning the weather and the ability to forecast it. The typical limiting factor for executing maintenance actions is the wind speed. Regulations and manufacturers' good practices set the maximum values of the wind speed which allow work at different locations on the turbine. This information has been used in previous research works to develop maintenance frameworks. For example, one study fixed the wind speed limit as 10 m/s for accessing the whole turbine [24], while another based the safe working limit on cut-in wind speed, i.e., the turbine was only considered maintainable when the wind speed was lower than cut-in [30]. Furthermore, current regulations and maintenance guides include dynamic safety limits taking into account not only the mean ten-minute wind speed value but also the gust value, when a crane usage is required for such a case like major component replacement. The definition of gust is a short-duration (seconds) maximum of the fluctuating wind speed [31].

The maximum permissible wind gust speed for crane usage depends on various factors such as mean wind speed, intervention height and weight of the load [32]. Therefore, the corresponding wind gust restriction for any intervention requires timely and case based controls. Moreover, high gust values cause more restrictive wind turbine component specific accessibility rules reducing the highest allowed mean wind speed. In this case, the mean wind speed limit for safe working has to be decreased by 2 m/s when the wind speed gust is above 5 m/s for operation requiring a crane usage.

Taking into account only wind speed limits, the safe working rules are also different depending on turbine maker and model. For example, in the case of MADE AE 46 turbines, preventative maintenance requires calmer winds than 20 m/s at the nacelle, however changing the whole nacelle requires the wind speed to be not more than 5 m/s. If we check the requirements for NEG Micon NM 52 turbines, working in the hub requires wind speeds below 15 m/s while working in the nacelle roof is allowable until 12 m/s and generator alignment should not be performed for wind speeds above 10 m/s. Finally, for the Vestas V 90 3.0 MW model, generator alignment intervention can't be done for wind speeds above 8 m/s, changing pitch angle requires wind speed values smaller than 6 m/s and working in the drive train is allowed up to 7 m/s.

Various tasks must be completed in a wind turbine within a work shift, and also comply with time and labour force restrictions. As stated in [33], it is almost impossible to generate a flawless maintenance plan in regards to production loss, since it is difficult to find a period where the turbine is not producing. The solution is to use weather forecasts and schedule the maintenance with an acceptable uncertainty [34]. The effects of these imprecisions and modelled uncertainties must be analysed, and their contribution to 'weather downtime' must be reported in order to make required improvements, which can then decrease the downtime and the intervention risk for the technicians.

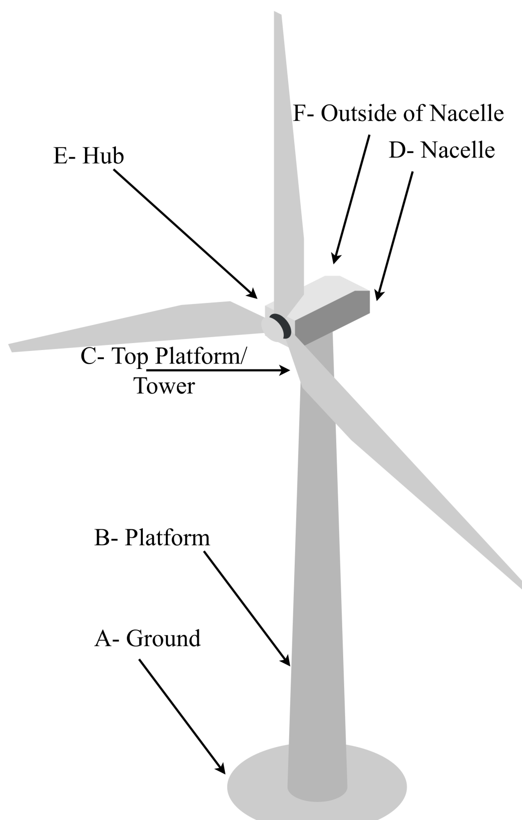


Figure 1.3: Example of turbine working zones.

Figure 5.1 presents an example of the turbine working zones, where different HSE rules exist. Rigorous HSE rules must be fulfilled for the service visit depending on the height of the working zone, intervention type and environmental factors.

1.3.3 Wind farm maintenance decisions

The interpretation of an optimal combination of particular maintenance decisions is known as maintenance strategy or policy. In general, maintenance decisions can first be grouped based on whether the component needs to be replaced or repaired, then further grouped based on an intervention date. For making these meta decisions, high level information is needed, such as requirements for the stock of spare parts and consumable materials , along with wind speed and electricity market price forecasts during a specific time. In this manner, decision groups are given as "replace or repair" as binary choice cases, assuming that "do nothing" is not an option, intervention timing, transportation type selection and usage of the sources (optimal number of maintenance personal, spare parts stocks, optimal composition of maintenance strategies, in situ intervention timing and prioritisation).

Replace or repair

The difference between restoring an item from a failed state to a functional state through reparation, or substitution of the failed part, versus replacement of the whole item, leads us to the term repairability. In practice, an item's repairability depends on technical limitations and its owner's financial holdings [35]. This kind of decision making can be structured as an evaluation of mutually exclusive projects or replacement analysis. Before the purchase of any fixed asset, the decision of the expenditure must be evaluated by taking into account connected possible future profits (estimated values) [36], which represent the cash flow. If the acceptance of one project causes the rejection of the remaining alternatives, these projects are known as mutually exclusive. The net present value (NPV) ranking can be used to make a decision between mutually exclusive projects [37].

Although replacement projects are a subcategory of the mutually exclusive ones, they do possess unique characteristics that allow scientists and analysts to use specialised concepts and analysis techniques in their evaluation. When faced with a failed component, one of the main decisions an operator or owner needs to consider is whether to replace it with the same component or to substitute it with a different or even newer technology. Replacement analysis is a theory which evaluates whether a faulty system is to be replaced by one of the same brand and technology, or by a different set up. The decision maker evaluates an existing asset, which is called a defender, against a new component technology, named a challenger [38].

In this thesis, replace or repair decisions are analysed in Chapter 4 considering replacement and repair timings.

Optimal time for intervention

Although replacement analysis can provide an estimation for the replacement time of a component, its weakness is determining the optimal intervention time for the routine maintenance, inspection, and so forth. One way to address this problem is to set up appropriate scenarios and simulate the effects of either postponing the intervention, or keeping the maintenance intervals between service visits short [24, 39].

The literature provides various approaches for maintenance decision making. These include an imperfect maintenance model using the proportional age set back process, and the reliability life analysis using the Weibull distribution to estimate the longest period. The Weibull distribution is used for performing analysis in order to keep the component reliable and the maintenance costs minimized [40–42].

In this thesis, analysis for cost optimal preventative intervention time considering electricity market price peak values is provided in Chapter 3.

Transportation type selection

Remoteness and low accessibility of wind farms are well known maintenance plan problems. Several researchers have worked on the development of cost optimisation of the vessel selection for offshore wind farms [43, 44]. These tools are concerned with maintenance procedures, rather than questioning about the necessity of maintenance. Since, the available maintenance data are obtained from an onshore wind farm and transportation type selection is not a prioritised decision as it is in the offshore wind farms, in this thesis this decision is not directly studied. In order to provide wind turbine health status related information to an O&M engineer, the necessity of maintenance is partially studied in Chapter 3. The maintenance procedures for only the crane usage (helicopters, sea vessels are out of topic) and the accessibility of the wind turbines are studied in Chapter 5.

Optimal number of maintenance personal

According to accepted practice, focusing on maintenance procedure requires an optimisation of labour force. As such, the optimal number of teams to be sent to a wind farm has become a valid research question, and corresponding DSSs have been developed [24, 45]. In this thesis, a maintenance team consists of two technicians considering the information obtained during the industrial secondment of the researcher. Here, it is assumed that the studied case maintenance visits require intervention of one maintenance team at a single maintenance visit.

Spare parts stocks

Another important aspect of the maintenance procedure is the warehouse management strategy. In practice, execution of maintenance is influenced by spare parts availability. Specially risk-based wind farm warehouse management DSSs exist in the literature [46, 47]. In this work, the needed spare parts are assumed to be available in all the studied interventions.

Optimal composition of maintenance strategies

An evaluation of long term maintenance strategies is given in this subsection. In comparison with wind, there exists more literature in mining and transportation fields [48, 49]. These concern selective maintenance and maintenance efficiency to recommend preventative maintenance, imperfect maintenance, minimal repair or corrective maintenance. It was shown that for small scale wind turbines, it is possible to avoid complex maintenance plans during the first 10 years of their life time [50]. However, the development of long term maintenance strategies depends on the wind turbine technology and the site specific features. Furthermore, the inability to assess WT health through sensor technologies, such as condition monitoring systems (CMS), requires preventative and corrective maintenance policies for the older wind turbines. Nevertheless, the complexity of newer WT technologies generated the need for predictive techniques. For new wind farms, condition monitoring systems and gathered online and offline inspection data provided additional information for the maintenance decisions and a new type of visit, known as inspection visit [51]. For wind farm operators, this resulted in hybrid-opportunistic maintenance plans [11, 19] composed of preventative, predictive and corrective actions. These three maintenance approaches hold different recommended shares among hybrid-opportunistic maintenance approaches presented in the literature. Some studies suggest an increased share of corrective maintenance, while others find that augmenting the percentage of preventative actions results in better overall maintenance [24, 52]. The uneven distribution of different maintenance techniques is also found in current industrial practice, where the intervention decisions are highly dependent on the preferences of the operators or maintenance providers. Thus, for each wind farm, the percentage of the three different maintenance approaches can vary greatly, causing some degree of incomparability among different farms.

In situ intervention timing and prioritisation

Wind turbines are complex systems consisting of approximately 8000 parts [53], and a single farm consists of many wind turbines. Prior to a service visit, a team must be well informed regarding the turbines in need of maintenance, and the parts of each subject turbine. In other fields, there exists an overall throughput effectiveness metric for assigning maintenance priority to components of the system [54]. In the field of wind energy, the number of turbines that can be maintained in one visit is addressed various times for both onshore and offshore cases [30, 33, 55, 56] by focusing on daily travel limitations and technicians availability. As soon as the components' conditions are identified as critical, the maintenance management needs to immediately decide whether or not to carry out the maintenance actions. Literature shows that 'what-if' analyses and case based simulations are often required to decide whether to postpone the intervention or execute it immediately; as well as whether to take minor or major actions such as repair and replacement [39, 57]. However, after the arrival of the maintenance team to the wind farm, additional information is required to prioritise the executable tasks. Such information has a high value, as it decreases downtime and production losses due to weather.

In Chapter 5 an in situ intervention timing and prioritisation tool is designed and explained in detail.

1.3.4 Sensitivity of decisions

Gaining profit from a wind turbine is a complex and multi-variable dependent process. Well-known parameters such as electricity market price, wind turbine health status, production efficiency and regional renewable energy policies have been deemed the main influential factors for generating revenue. Although they all contain a significant level of unpredictability due to both their own complex working principles and limitations of observation and measurement, it is still possible to track their influences through response of the revenue. Evaluation of the economic behaviour of wind parks has shown that the wind farm feasibility is strongly influenced by the capacity factor and electricity market price fluctuations, whereas the nominal power and inflation rate were found as only slightly influential on the payback period of an investment [13].

Not only for the initial investment decisions but also for the decisions through life time of a wind farm, electricity market price and policy are two of the driving factors.

- Electricity Market Price: The day-ahead market is often a spot market with contracts for energy generation in an hourly time resolution. Spot markets can differ by country, e.g. half hourly trading in the UK, the use of flexible block lengths in the Netherlands and the possibility of complex bids for generators with load gradients and scheduled stops in Spain [58, 59]. Electricity market price is also dependent on complex drivers such as consumer behaviour, which varies with time and date, electricity generation from renewables, changing with the weather and the season, and the fossil fuel prices.

- Policy: Both, regional and country specific policies can contribute to the attractiveness of wind energy. The policy represents tax rates, subsidy frameworks and energy targets in this context. As an example, the 2020 national target for Spain has been set so that renewable resources will cover 20 % of total energy consumption [60]. By the end of 2017, 18.4% of electricity demand of Spain was met by wind energy[61]. According to statistics from Spanish Wind Energy Association (AEE), wind energy is ranked as the second source for the generation of electricity, despite the 2015 halt on the objective to increase installed capacity [61, 62]. Some of the reasons behind the halt are that the subsidy scheme was changed many times, and the premium for wind energy ended because of the tariff deficit [63],[64]. These frequent regulation updates can also be seen in other countries. Although there has been some recent attempt to harmonise and liberalise state aid in the EU [65], there are still various subsidy frameworks for wind energy in force such as fixed feed-in tariffs, premiums, green certificates and tax exemption rules. Figure 1.4 illustrates the history of subsidies for Spain, Netherlands and UK [66–68]. By following a similar comparative approach, the tax rate varies in the countries being 25-28 % for Spain, 25% in Netherlands and 20-21% for UK depending on the year [69]. As a rule of thump, higher tax rates have negative impacts on the investor's behaviours. This statement is also valid for the wind energy business.

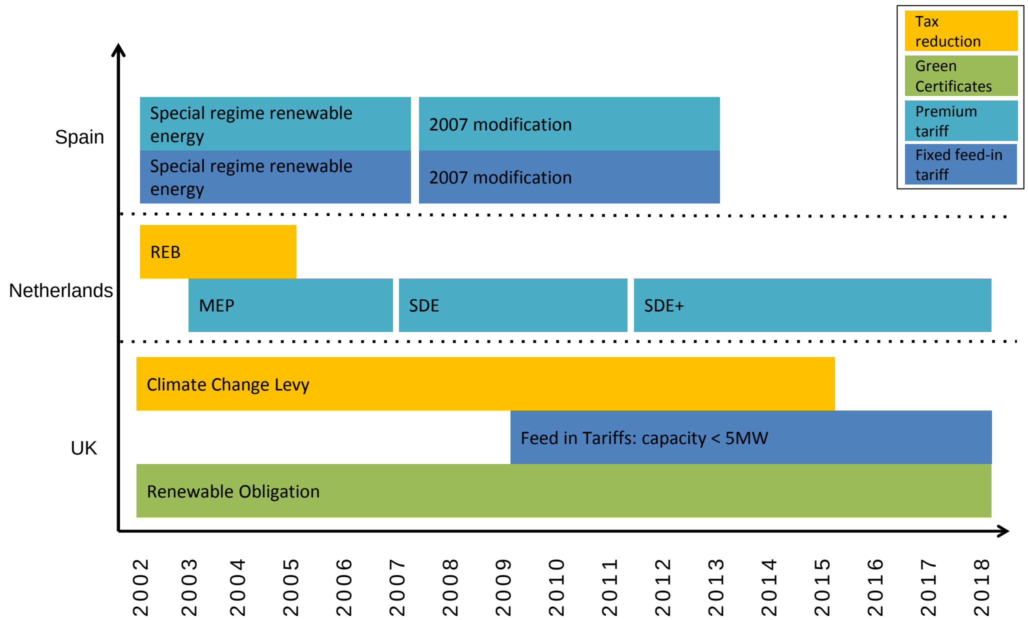


Figure 1.4: Simplified history of wind energy subsidies in Spain, Netherlands and UK.

During the life time of a wind farm, continuous decision making is mandatory for operation and maintenance activities. The increasing importance of O&M costs has directed attention toward sensitivity studies on maintenance policies. Wind farm maintenance simulation and optimisation tools have been developed, and results show that turbine availability is sensitive to the shift length of the service team and failure rates of components [70–72]. Repair time, inspection timing and inspection accuracy were found to strongly affect the performance of the maintenance strategy (valid for all types: corrective, preventative or predictive) [51, 73–75]. Kerres *et al.* [52] stated that corrective maintenance is the most cost-effective strategy for the components of the drive train. Leigh and Dunnet [24] showed that periodical replacements of subsystems significantly decrease the number of required corrective maintenance visits. A maintenance decision is also highly dependent on the environment, since environmental variables are significantly correlated with failure occurrences [76] and accessibility is also dependent on the weather [77]. According to the estimations obtained with Monte Carlo simulations for offshore wind turbines, the uncertainty surrounding weather forecasts can increase maintenance costs by 1.74 times in the case of corrective maintenance strategy, 0.56 times for the calendar based and 0.35 times for the condition monitoring based preventative maintenance strategies [75].

In addition to previously discussed factors, the model structure for maintenance principles is also important. High maximum-age threshold decreases maintenance costs, whereas the effect of the low minimum age threshold increases preventative maintenance cost [21]. Case studies based on real data can give an insight into the complexity and sensitivity of decisions that simulation tools cannot provide.

1.4 Structure of the thesis

The thesis is structured in the order that follows. Problem statement and objectives of the research are already presented in Chapter 1. It is worthwhile to highlight that the literature review is also presented in Chapter 1. Chapter 2 provides the mathematical foundations of the employed methodologies. In Chapter 3, components of a wind farm revenue are analysed and the modelling alternatives to determine key factors contribution on the decision making are studied. Chapter 4 focuses on an application study considering the influence of the maintenance decisions on wind farm revenue. In the Chapter 5, the design of a decision support tool for the maintenance intervention planning is explained and tests of the proposed DSS are given. Finally, the summary of the contributions, comments on closeness to goals and future work are given in Chapter 6.

Chapter 2

Mathematical Foundations

In this chapter, the theoretical basis of all the techniques used in the thesis will be presented. The techniques are grouped under the sections named: unsupervised learning, search algorithms for scheduling decisions, time series analysis and financial tools.

2.1 Unsupervised learning

Data driven models and studies are very promising for engineers and researchers from different areas, due to various reasons. The main reason behind data driven studies' popularity is its potential for both the identification of existing trends and for making accurate predictions about the future. While these identifications and predictions are possible with either supervised or unsupervised learning tools, this section presents the unsupervised learning tools and principal component regression (a hybrid tool that uses both unsupervised and supervised techniques) that are used in this thesis. In the following subsections apriori rule mining, agglomerative clustering, K-means clustering and principal component regression are presented.

2.1.1 Apriori association rule mining

Apriori Association Rule Mining (AARM) is a well structured method used to find relations and frequent patterns among variables in databases. Mining of these relations provides viable information for decision support systems. The rule generation can be done using frequent itemset searching algorithms such as the Apriori algorithm [78], the Eclat algorithm[79], etcetera.

The AARM is frequently used by researchers who need to deal with big maintenance and failure data [76, 80–82]. In this thesis, the AARM algorithm is used in Chapter 3 associating environmental parameters and component failures. This algorithm is easy to implement and has already many applications in the existing literature such as fault cause analysis for power distribution systems [82], improvements of wind speed forecasts [83], wind power ramp detection [84] and labelling & classification of failure occurrences [76].

The AARM algorithm, which is based on counting process, mines frequent patterns from the data set that contains transactions. The details of this algorithm can be found in [85].

The main principles of the AARM can be summarised in three major steps:

(a) Grouping of items as itemsets and generation of transactions within databases

(b) Counting of the combination of frequent itemsets

(c) Filtering and ranking by using rule metrics

The characteristics of each rule are summarised with three metrics: *support*, *confidence* and *lift*. They are given in Table 2.1 for A (an item on the left side of the rule), B (an item on the right side of the rule) and N (the total number of transactions). $Freq$ stands for the absolute frequency of appearance of the example rule.

Table 2.1: Rule metrics and their formulations for the Apriori Ruling [85–87]

Metric	Formula
Support	$supp(A \to B) = \frac{Freq(A \to B)}{N} = Prob(A \wedge B)$
Confidence	$conf(A \to B) = \frac{supp(A \to B)}{supp(A)} = Prob(A \mid B)$
Lift	$lift(A \to B) = \frac{supp(A \to B)}{supp(A) supp(B)}$

The aim of the AARM is to discover all the rules that exceed pre-set values of support and confidence in a given set of transactions [88]. In addition to rule generation process, the significance of an association rule is also determined according to the values of these metrics. As an example, for items A and B, a user defined support value of 0.5 is a decisive parameter to check whether the support of $A \rightarrow B$ equals or exceeds 0.5. If it does not equal or exceed, $A \rightarrow B$ is not an association rule. If the comparison between user defined support level and $supp(A \rightarrow B)$ confirms that $A \rightarrow B$ is a rule, then the difference between these two support values can be used to rank the rules. In a similar way, more elaborated metrics, confidence and lift contribute to rule generation and rule ranking processes.

2.1.2 Agglomerative nesting

One of the challenges to make a research using operation and maintenance data, is the need of considering several data sources simultaneously. With each data source considered, the number of signals, the number of variables and the complexity of the final database change, then the need for an elaborated data treatment increases. As an example, maintenance logbooks, Supervisory Control and Data Acquisition (SCADA) signals, weather data and electricity market price data are some of the sources used in this thesis. In order to simplify such data mixtures, labelling and grouping are recommended in the literature [89, 90]. As a data processing tool, Agglomerative nesting is used in Chapter 5 in order to group some subsets into broader categories.

Agglomerative nesting (Agnes) is a data-mining tool and a sub-category of hierarchical clustering. The Agnes algorithm is known as a bottom-up process, since it first samples a separate cluster assignation for each observation, and then samples the merging among these clusters. Thus, hierarchy is defined from the bottom and then moves up.

Agglomerative nesting is chosen to avoid cluster structure's dependence from the assigned initial centroids and the number of clusters selected in advance [91]. Although this method is more computationally expensive than others, this is not an issue for relatively small databases. To set-up this approach, the distance between clusters and their merging rule must be defined in advance. Various distance definitions, such as Euclidean, Manhattan and Mahalanobis, can be found in the literature. This study uses the Euclidean distance. As a first step, agglomerative nesting assigns each observation to a separate cluster. Then, the Euclidean distance is estimated, which is defined as the square root of the straight line (shortest) distance between two clusters. Further, clusters are merged according to the Ward Algorithm [92], whereby the sum of squared Euclidean distances is minimised. The Ward algorithm is based on the minimum variance principle and it uses an objective function, which forms the distances as error sum of squares, given below.

$$ESS = (\Sigma_{i=1}^{n} S_i^2) - \frac{1}{n}(\Sigma_{i=1}^{n} S_i)^2 \tag{2.1}$$

where ESS represents the Error sum of squares and S_i represents the score of i^{th} individual.

The simplified flowchart of the original version published in 1963 [92], reproduced by using [93] is shown in Figure 2.1.

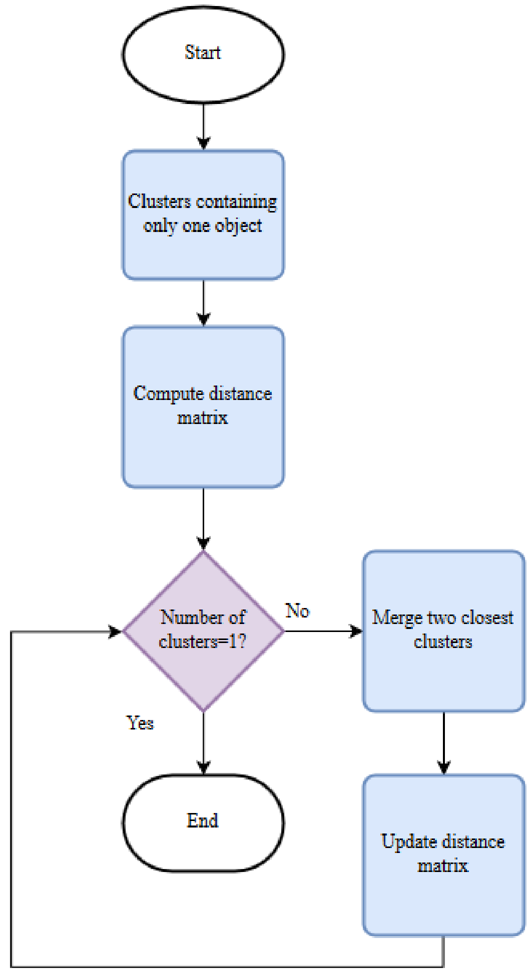


Figure 2.1: The Ward algorithm

A dendrogram illustration is the common way to show the arrangement of clusters that are generated by agglomerative nesting [94–96]. One of the drawbacks of the process is the difficulty for the identification of the number of clusters just from the dendrogram only. As it is recommended in the literature, the selection of the relevant number of clusters is made by considering the agreement between various indexes taking into account the majority rule, the number of clusters proposed by the majority of indexes [97]. For an example data set, if 27 indexes are considered and 19 of them proposed 5 clusters as the best number, 4 of them proposed 1 cluster, 3 of them proposed 2 clusters and 1 of them proposed 8 clusters, according to the majority rule the best number of clusters is 5.

In the dendrogram visualisation, height represents the value of the Euclidean distance between clusters. To estimate this distance, input data must be scaled. As an example, consider an input consisting of 100 rows and 2 columns, where the first column indicates the variable E and the second one stands for the variable U. After scaling the input data each observation is firstly assigned to a temporary cluster. Following this procedure, in the first step there exist 100 clusters and, for instance, the Euclidean distance between Cluster 1 and Cluster 2 can be obtained as;

$$Euc_{dist} = \sqrt{(E_{C_1} - E_{C_2})^2 + (U_{C_1} - U_{C_2})^2} \tag{2.2}$$

This distance calculation must be repeated for all 100 clusters. Afterwards, the Ward algorithm groups these clusters according to the minimisation principle of the Euclidean distances. This is repeated as many times as needed finishing when only one cluster remains.

2.1.3 K-means clustering

As it is introduced in the Agnes section, the needed final database for operation and maintenance research requires serious data treatment processes. In order to simplify the available data and to have a well established input database, labelling and grouping of variables are recommended in the literature [89, 90]. As a data processing tool, K-means clustering is used in Chapter 3 to discover the weather conditions at certain point of time or just before a failure.

The K-means algorithm is known as the point-assignment type partitioning method, and draws from multivariate means estimation in the Euclidean space. The K-means algorithm is known as one of the oldest and most widely used algorithms [98]. The main difference with the Agnes algorithm is the working principle of the K-means algorithm for the initial assignment phase of the clusters. In Agnes, the process starts by assigning a cluster for each observation, whereas the K-means algorithm requires as one input the desired number of clusters specified by user, k. The working principle of the K-means is illustrated in Figure 2.2.

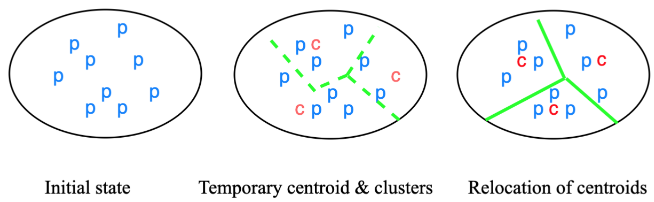


Figure 2.2: K-means algorithm, p represents points and c stands for centroids.

At first, clusters are randomly formed, therefore their initial centroids (c) are at random locations. Each cluster must contain some of the observations. For example, an observation p is assigned to a cluster j based on p's closeness to the centroid of cluster j in the Euclidean space. In the second step, the coordinates of cluster j's participants (member observations) are used to estimate the new centroid of cluster j. Then, in order to find the distances between new cluster centroids and observations the calculation is repeated, which means that the members of a cluster might change according to the distance and the centroid re-calculations. Every time a cluster's member changes, its centroid also changes. This continues until there are no more member changes between the clusters [94].

Suppose $D = \{p_1, ..., p_n\}$ is the observation set to be clustered:

K-means can be formulated by an objective function [98]:

$$min \sum_{j=1}^{K} \sum_{p \in C_j} \pi_p dist(p, c_j) \tag{2.3}$$

where π_p is the weight of p, c_j is the centroid of cluster C_j. k is the number of clusters. The function $dist$ calculates the distance between the object p and the centroid c_j.

2.1.4 Principal component regression

In wind farms and electricity markets, a common practice, is to record signals in 10-minutes intervals or one hour resolution, resulting in big generic databases that can't be used directly in particular models. For such databases, it is difficult to know, which signals exert influence on a selected response variable. Therefore, for a research focused on providing a data-driven analysis and to determine key variables for timely decision making as aimed in Chapter 3, it is important to define properly useful variables and to eliminate unnecessary ones for the sake of modelling accuracy and to decrease computational time. In order to simplify and interpret such data, dimensionality reduction is recommended in the literature [99, 100]. Principal component analysis (PCA) is a useful procedure (one of the widely used machine learning methods) to summarise data, which have many features and various interpretation summaries [101]. The PCA summarises the data based on combinations of the existing features to capture the nature of data with fewer aspects, and in a more comprehensive way. Moreover, the PCA is a remedy for multicollinearity problem of regression models. Multicollinearity is known as existence of high correlations among the explanatory variables of a regression. This phenomenon generates overall erroneous good fit statistics with inaccurate regression coefficients for the explanatory variables [102]. Therefore, under multicollinearity the regression statistics can indicate that the generated regression equation is successful while the estimation accuracy for the response variable is very weak. It is worthwhile to highlight that linear regression models are known as supervised learning methods and the PCA is an unsupervised learning method. When these two methods are combined in order to address multicollinearity, the resulting technique is named as Principal Component Regression (PCR). Therefore, the PCR is a regression technique based on the PCA. In order to explain the PCR, the multiple linear regression and the PCA are presented in the subsequent sections.

Multiple linear regression

The supervised component of the PCR is the multiple linear regression (MLR). The MLR aims to model the relationship between explanatory variables and a response variable. The mathematical representation of a multiple linear regression model can be written as[102]:

$$Y = X\beta + \epsilon \tag{2.4}$$

where β symbolises fixed and unknown regression coefficients, ϵ is the random error. X stands for the vector of the exploratory variables and Y is the response variable. If the equations are written for n observations of the h explanatory variables, the matrix notation can be used:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & x_{31} & \dots & x_{h1} \\ 1 & x_{22} & x_{32} & \dots & x_{h2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2n} & x_{3n} & \dots & x_{hn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_h \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The β coefficients can be obtained using matrix operations:

$$\beta = (X'X)^{-1}X'Y \tag{2.5}$$

where $()'$ is the transpose of the matrix given in parenthesis and $()^{-1}$ indicates the inverse of the matrix given in parenthesis.

Principal components

The unsupervised parts of the PCR are the principal components. Using the same notation presented in the multiple linear regression explanation, it can be said that $X'X$ is the correlation matrix. As it is already said, the PCA is a procedure that summarises the data based on combinations of existing features to capture the nature of data with fewer aspects. The outcome of the PCA is a set of new variables $\gamma_i, i = 1, ..., h$.. The transformation of the exploratory variables to the new set of variables can be written as [102]:

$$\gamma_i = v_{i1}x_1 + v_{i2}x_2 + ... + v_{ih}x_h, \text{i=1,2,...,h} \tag{2.6}$$

$$\Gamma = XV \tag{2.7}$$

where V is a matrix of coefficients (v_{ij}) and Γ is a linear transformation of the exploratory variables.

According to the equation 2.6, an infinite number of linear transformations exist, but the principal components transformation has unique features such that:

- The new set of variables are uncorrelated, then $\Gamma'\Gamma$ is a matrix with the elements in the diagonal having non-zero values (λ_i) and all the remaining elements, out of the diagonal, equal to zero.

- The variance across the principal components distributes in descending order.

$$\begin{bmatrix} \gamma_1 \rightarrow \text{largest variance} \\ \gamma_2 \rightarrow \text{second largest variance} \\ \vdots \\ \gamma_i \rightarrow \text{smallest variance} \end{bmatrix}$$

In order to find the principal components, the characteristic values and vectors ($\lambda_1, \lambda_2, ..., \lambda_h$) of the covariance matrix (for similar variable scales) or correlation matrix (when the variables are on different scales) must be found using several matrix operations via the PCA [102] or techniques such as singular value decomposition [103, 104].

The combination of the PCA and the MLR

If we revisit the MLR equation;

$$Y = X\beta + \epsilon \tag{2.8}$$

In the PCR, after the dimension numbers of data are reduced and the principal components Γ are calculated, the principal components are used as variables in a multiple linear regression [105]. If we redefine the MLR equation by using principal components of the exploratory variables, it can be written as:

$$Y = \Gamma V + \epsilon \tag{2.9}$$

where V is the vector of regression coefficients estimated by the least squares principle. In this way, the exploratory variables have been replaced by their principal components in the regression model [106].

In order to estimate the coefficients by least squares:

$$\hat{V} = (\Gamma'\Gamma)^{-1}\Gamma' Y \tag{2.10}$$

Remembering that the principal components are uncorrelated, therefore $(\Gamma'\Gamma)$ is a diagonal matrix. Thus, the multicollinearity does not affect the variances of the regression coefficients [102].

In Chapter 3, the PCR is combined with Data-Based sensitivity analysis [107, 108] in order to investigate relative importance ranking of the electricity generation and demand variables on the electricity market price.

2.2 Search algorithm for scheduling decisions

While many operation and maintenance (O&M) decision support systems (DSS) have been already proposed, there is still a serious research need for the area of wind farm O&M scheduling, which is a challenging task due to the fact that turbines are frequently located in difficult to access locations such as offshore or in mountainous terrain. Maintenance teams must follow specific procedures when performing their service. Chapter 5 is devoted to find the optimal intervention time and the most effective execution order for different maintenance tasks. For such a purpose, the operations research field should be visited, where scheduling problems and their solution algorithms are one of the major topics of concern. In this field, the scientific perspective of decision making involves the application of mathematical representation for actual situations and then, the usage of optimisation models for choosing the best option amongst given possibilities. As it is said in [109], "an optimisation model seeks to find values of the decision variables that optimise (maximise or minimise) an objective function among the set of all values for the decision variables that satisfy the given constraints". In an optimisation model, objective function, decision variables, and constraints must be defined properly.

The mathematical representation can be written as:

For a variable X
Optimise $F(X)$
Subject to: $\{X \in S_1, G(X) \in S_2\}$

where X is a vector of decision variables and $F(X)$ is the objective function subjected to restrictions X belonging to S_1 and $G(X)$ belonging to S_2.

After defining the scheduling problem in detail, and generating all possible combinations, a search algorithm must be used to find the optimal solution. Therefore an extensive decision pool, containing a prioritised list of (all) possible scheduling combinations, can be provided to the decision maker. For such a decision pool, all scheduling combinations must be generated considering problem-specific heuristics. This straight forward way is known as brute force search by its definition in literature [110]. Brute-force search is simple to implement and it always finds a solution, if there exists at least one. This search is practical when there are problem-specific heuristics for the problem which are capable of reducing the set of candidates.

Specifically, if the brute-force method is used for finding an item or sets of items in a table by checking all entries sequentially, the search process is known as linear search [111].

2.3 Time series analysis

Physical variables can be measured and recorded in time, then the collection of these records is called time series. As a fundamental concept, it has applications as regression, forecasting and correlation analysis. Finance, meteorological, and operation and maintenance data of wind energy are widely expressed as time series. Therefore, wind farms' data driven analyses always contain methodologies developed for time series [112–115]. In this thesis, the majority of used data are in time series form. In order to compare different time series, the Kendall's rank correlation coefficient, and the Granger causality are used. Seasonal trend decomposition methods and several relative importance metrics are also applied. When there are no forecasts, synthetic time series data are generated using Auto Regressive Moving Average algorithm. In the following subsections the working principles of these methods are presented.

2.3.1 Kendall's rank correlation coefficient

When there exist many signals expressed as time series, it is important to summarise their interrelations in an efficient way. Statistical correlation analysis is capable of evaluating the strength of relationship between two time series. There exist several correlation analysis techniques in the literature to examine the relationships between variables. Among the existing techniques

Pearson's correlation analysis is preferred in most of the cases [116, 117].

However, Pearson is a parametric correlation method and it has very strict prerequisites and assumptions such as a Gaussian distribution of the data [118]. For other situations, non-parametric Kendall's rank correlation coefficient is suggested in the literature [119]. Especially the electricity market data don't necessarily have a Gaussian distribution. In Chapter 3, to overcome the Gaussian prerequisite of the Pearson method, the non-parametric Kendall correlation technique has been considered. As a robust and an efficient way to identify monotone relationships between two time series, the Kendall's rank correlation coefficient is used. It can be explained as follows:

A set of N observations of the variables J and D can be written as $(j_1, d_1), (j_2, d_2), ..., (j_N, d_N)$. Given an example pair of observations (j_1, d_1) and (j_2, d_2), if both $j_1 > j_2$ and $d_1 > d_2$ or $j_1 < j_2$ and $d_1 < d_2$, this pair is labelled as a concordant pair. On the contrary, if $j_1 > j_2$ and $d_1 < d_2$ or $j_1 < j_2$ and $d_1 > d_2$ this pair is a discordant pair. Lastly, If $j_1 = j_2$ and $d_1 = d_2$, this pair does not have a label as concordant or discordant. After labelling all possible pairs, the Kendall's rank correlation coefficient can be expressed as [120]:

$$\frac{\text{(number of concordant pairs)-(number of discordant pairs)}}{\frac{N(N-1)}{2}} \tag{2.11}$$

A variable is named as a ranked variable whenever every observation of the variable can be put in an ordinal order $(1^{st}, 2^{nd}, ..., etcetera)$. This ranking procedure does not say anything about the values of the observations or the comparison between them. Instead, it informs about which observation comes after the other. For variable J consisting from N data points, the first point, associated to the observation j_1, has the highest rank and the last one j_N, has the lowest rank. Similarly, for variable D, observation d_1 has the highest rank and observation d_N has the lowest rank. A positive result of the Kendall's rank coefficient means that higher ranks of the J variable are associated with higher ranks of D variable. A negative result indicates that lower ranks of the J variable are related to higher ranks of D variable (or the opposite). Zero or close to zero results indicate that there is no linear relation between these two variables [120, 121].

2.3.2 Granger causality

The correlation between two time series does not ensure the causality between them. In Chapter 3, in order to define exploratory variables for a selected response variable, causality analysis is performed with statistical tests. There exist several techniques, but the most frequently used test is Granger-causality [122–126].

The Granger causality test provides a similarity comparison focusing on cause-effect relationship between time series. Before introducing step by step the process, it is worth mentioning the main terminology of statistical tests, since the Granger causality is one of them.

A null hypothesis indicates that there is no relationship, no association among variables under investigation [127]. When the null hypothesis is statistically tested and the result of the test, p_{result}, is obtained, the interpretation of the result requires a comparison between the p_{result} and the preselected significance level. As a rule of thumb, $p_{result} \leq p_{significance}$ represents a statistically significant result, where $p_{significance}$ is the preselected significance level. The interpretation of the result is "there exist a relationship between the variables under investigation".

The mathematical test procedure consists of three main steps [128]:

1. Assume a bivariate linear autoregressive model M_{null}, on X and Y time series, here Y is dependent on the past records of X and Y. Secondly, consider a univariate linear autoregressive model $M_{restricted}$, on Y, where Y is dependent only on its own past records.

   $M_{null} : Y_t = \alpha_0 + \alpha_1 Y_{t-1} + ... + \alpha_l Y_{t-l} + \beta_1 X_{t-1} + ... + \beta_l X_{t-l} + \epsilon_t$
   $M_{restricted} : Y_t = \alpha_0 + \alpha_1 Y_{t-1} + ... + \alpha_l Y_{t-l} + \epsilon_t$

   where l is the number of lagged observations, ϵ_t is the error. α and β are the coefficients of each lagged observation.

2. Define the null hypothesis $H_0 :$ X does not cause Y

   Considering the notation given in previous step, the null hypothesis indicates that all β coefficients equal to zero.

3. Evaluate the null hypothesis

   (a) Calculate the residual sum of squares Res_{null} and $Res_{restricted}$ for M_{null} and $M_{restricted}$ respectively.

(b) Perform the F-test

(c) Interpret the test value obtained from the F-test, by keeping in mind: H_0 indicates that model M_{null} does not fit to data better than $M_{restricted}$. By evaluating p_{result} obtained from the F-test, it could be possible to reject H_0. Then, the test can be concluded as X causes Y.

2.3.3 Seasonal trend decomposition procedure

Typically, a time series of data shows three main components, trend, seasonality (systematic cycles), and noise (instant fluctuations). Among them, the trend shows a general direction, long-term movements and tendency in the series [112]. In order to evaluate the similarity between two time series, one possibility is providing a trend comparison. The similarity between time series analyses require the identification of the common features for the variables under investigation. As a handy filtering process Seasonal Trend Decomposition of Time series by LOESS method (STL) is used in Chapter 3. As it name says, this method uses LOESS (locally estimated scatter plot smoothing) and bases on two recursive looping sub procedures called inner and outer loops [129].

Simply said, STL decomposes an additive time series into three components named trend, seasonality and residuals.

$$Y_t = T_t + S_t + R_t \tag{2.12}$$

where Y_t is the time series, T_t the trend component, S_t the seasonal component and R_t stands for the remainder or residual term.

In order to visualise the decomposition outputs, an example analysis is shown in Figure 2.3.

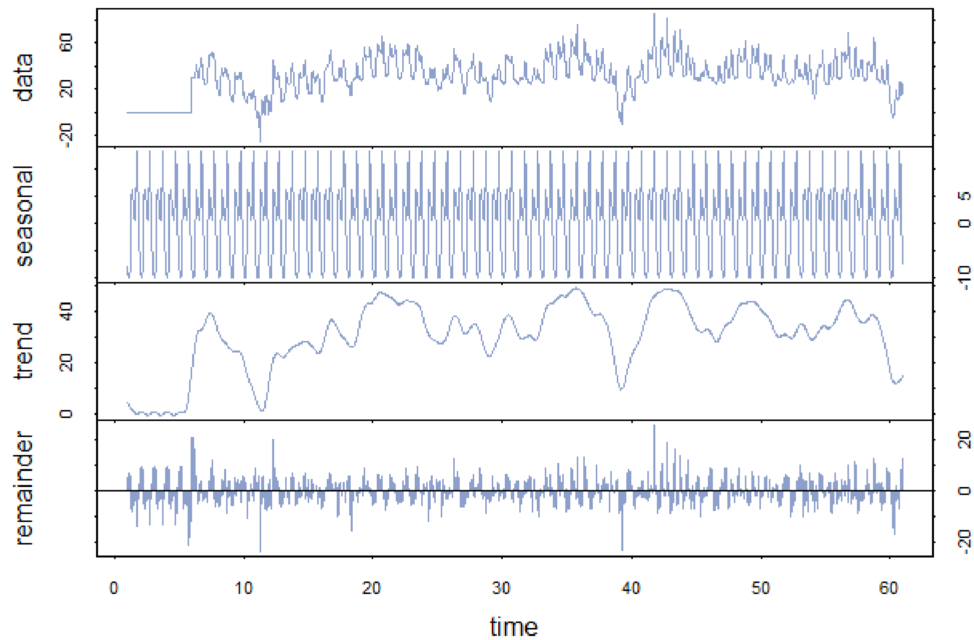


Figure 2.3: The components of an example time series (daily seasonality)

In Figure 2.3, during 60 days hourly recorded data are analysed and daily seasonality, trend and residuals (remainder) features of the time series data are presented.

Considering the purpose of the researcher, these features can be used for understanding the different components of the data under investigation [130] or in an anomaly detection [131] or ensuring whether the data under investigation satisfies statistical modelling and forecasting prerequisites or determining the needs of different data treatment procedures [132].

2.3.4 Autoregressive Integrated Moving Average

According to the existing state of art, wind speed forecasting error can be obtained with Auto regressive models [133–135]. These models, proposed by Box and Jenkins [136], have been extensively used for short term forecasting. They are commonly known by the generic name of ARIMA (Autoregressive

Integrated Moving Average) [137, 138]. The future value of the variable is described by a linear function of the previous data and a random error. ARIMA models are generally denoted by ARIMA(p,d,q) where the parameters p, d, and q are non-negative integers, representing the order of the autoregressive model, the degree of differencing and the order of the moving average respectively. When d equals to zero and p, q equal to 1 the ARIMA(p,d,q) becomes the ARIMA(1,0,1).

Mathematically, ARIMA(1,0,1) can be written as:

$$y_t = \phi y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \tag{2.13}$$

where ϕ is the autoregression operator of order $p = 1$, θ is the moving operator of order $q = 1$ and ϵ_t is the error term obtained with a Gaussian distribution with average equal to zero and standard deviation σ at time t.

ARIMA (1,0,1) is the equivalent of Auto Regressive Moving Average (ARMA) (1,1) model. The ARMA model is capable to mimic the correlation between errors in time, as it exists in the wind data [134]. Hence, according to the state of the art [135], wind speed scenario generation can be obtained with the ARMA (1,1) by just simulating the error instead of the point wind speed forecasts. The simulation data stands for the generated noise that will later be added to the historical data in order to produce synthetic forecasts with the desired error. In Chapter 5, forecasts are needed in order to test a decision support tool. Therefore, in Chapter 5, the ARMA (1,1) is used to generate the simulation data as a replacement of the forecasting errors. Then, the simulated errors are added to the measurement, in order to obtain synthetic time series with a known forecast error.

2.3.5 Relative importance metrics

In Chapter 3 a study for the identification of the most important regressors for the electricity market price is performed. Relative importance estimations and ranking among the variables change significantly depending on the underlying regression method and variable importance metrics. In order to obtain a robust result, relative importance rankings obtained using different metrics should be considered all together.

The procedure requires modelling of a response variable with independent variables. Each regression model is subject to a question of how well the model fits the data. Here the answer can be given with the findings of goodness-of-fit statistics, such that R-Square. In practice, the importance metrics use R-Square values as input. However there exist differences between test designs and importance estimation procedures. In another words, each metric considers R-Square in a different manner in order to estimate the importance of independent variables on the response variable.

In this study, metrics First, Last, Betasq and Pratt are used due to their simplicity. In the following a description of the procedure is given.

For the estimated response value $\hat{y}$ and the mean response value y_{mean}:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y_{mean})^2} \tag{2.14}$$

In order to explain how each metric considers R-Square, now we assume that there exist p variables and then p regression models. The procedure will be explained with the notation of a linear regression model where y is the response of object i with regressor values $x_{i1} : x_{ip}$, model coefficients are β, the estimated values for the model coefficients are $\hat{\beta}$ and modelling error ϵ. The metrics considered in Chapter 3 use R-Square and $\hat{\beta}$ in different ways in order to evaluate relative importance ranking of the exploratory variables.

Regarding the details of the metrics:

1. *Metric First* provides information on what amount of response variable can be explained by each individual exploratory variable. Here the importance ranking among exploratory variables is performed by comparing R^2 obtained from each regression model designed with one exploratory variable, [139].

   Example estimation procedure for the relative importance of variable X_1:

   1. $y_i = \beta_0 + x_{i1}\beta_1 + \epsilon_i => R_1^2$

   Example estimation procedure for the relative importance of variable X_2:

   2. $y_i = \beta_0 + x_{i2}\beta_2 + \epsilon_i => R_2^2$

   $\vdots$

   For X_p variable:

p. $y_i = \beta_0 + x_{ip}\beta_p + \epsilon_i => R_p^2$

Metric First for variable X_1: R_1^2

2. *Metric Last* provides information on each exploratory variable's capability to explain the response variable in addition to all other independent variables. In this metric with each regressor, increase in R^2 is evaluated. This metric is sometimes called usefulness, [139].

   Example estimation procedure for the relative importance of variable X_1 is given in the following steps. Regression of p variables with an exemption of variable X_1:

   $y_i = \beta_0 + x_{i2}\beta_2 + ... + x_{ip}\beta_p + \epsilon_i => R_{rest}^2$

   Regression of p variables:

   $y_i = \beta_0 + x_{i2}\beta_2 + ... + x_{ip}\beta_p + x_{i1}\beta_1 + \epsilon_i => R_{rest+1}^2$

   Metric Last for variable X_1: R_{rest+1}^2-R_{rest}^2

3. *Betasq* is the squared standardized coefficient of $\hat{\beta}$, [139].

   This metric measures the relative importance by comparing $\hat{\beta}_{p,standardized}^2$ obtained for each independent variable. $\hat{\beta}_{p,standardized}^2$ is calculated as follows [140]:

   $\hat{\beta}_{p,standardized}^2 = (\hat{\beta}_p \frac{S_{Xp}}{S_Y})^2$

   where S_{X_p} and S_Y are the empirical standard deviation of the variables X_p and Y.

4. *Metric Pratt* is the product of $\hat{\beta}_{p,standardized}^2$ and the empirical Pearson correlation coefficient ($H_{Xp,Y}$) for variable X_p [140]. The metric Pratt is known as natural decomposition of R^2 [139].

   $Pratt_p = \hat{\beta}_{p,standardized}^2 H_{Xp,Y}$

Using the relative importance estimations obtained from all these listed metrics, it would be possible to obtain an overall influence picture of the regressor variables on the response variable. The metric First tends to give strong preference for correlated regressors. The metric Last is weak to show direct influence and a good option for showing combined effects of the regressors. Finally, it is expected to see a strong agreement between Betasq and Pratt metrics because both of them are based on the squared standardised coefficient.

2.4 Financial tools

Chapter 4 is devoted to the financial evaluation of O&M decisions. In order to select the most appropriate financial tools, a short methodology summary is provided here. Any decision that results in a business expense can be classified by considering anticipated future profits. Most of the time, engineers take a responsible role as a decision maker in such cases. Financial assessment of engineering decisions can be grouped as [36]:

1. Equipment or process selection among various options.
2. Replacement of the worn-out equipment.
3. New product design or expansion.
4. Cost reduction with an investment such as purchase of new machinery for the production line, which can reduce manual operations.
5. Improvement in after-sale service and the quality of the final product.

The majority of O&M decisions can be counted among the first or second groups and, for them, evaluation of mutually exclusive projects is required, which is introduced in the next subsection.

2.4.1 Cash flow analysis and evaluation of mutually exclusive projects

In Chapter 4 more than 30 scenarios are analysed and these scenarios are defined as mutually exclusive projects. In order to present a comparison and a selection procedure between them, a financial evaluation indicator must be used. This financial indicator is estimated for each scenario using the cash flow analysis. In these scenarios, the estimated profits are derived in a discounted cash flow analysis that considers the time value of money. Interest represents the compensation accepted when risking loss of opportunities for the present sum of the money due to an uncertain future [141]. In this case, the decision threshold for an investor is known as the minimum attractive rate of return (MARR), which is used in discounting and refers to hurdle rate. The selection of MARR can be done based on long term interest rates and inflation rate (consumer price indexes)[39, 142].

For the evaluation of the alternative scenarios, candidate indicators can be listed as Net Present Value (NPV), payback period, internal rate of return (IRR) and inflation adjusted rate of return (IARR).

Table 2.2 shows the advantages and disadvantages of the indicators with NPV emerging as the most suitable measure for the study presented in Chapter 4.

Table 2.2: Financial indicator selection [36, 37]

Indicator	Advantages	Disadvantages
IRR	independent of the accuracy of interest rate	misleading for the selection among the projects, difficult to compute
IARR	shows the effect of the inflation	dependent on IRR
Payback period	simple	low resolution, no time value of money
NPV	tracks the direct impact of the project	dependent on the accuracy of interest rate estimation, long computing time

The generic formulas of the net present value (NPV) is given in the equation that follows [36],

$$NPV = \sum_{t=0}^{N} \frac{C_t}{(1+i)^t} \tag{2.15}$$

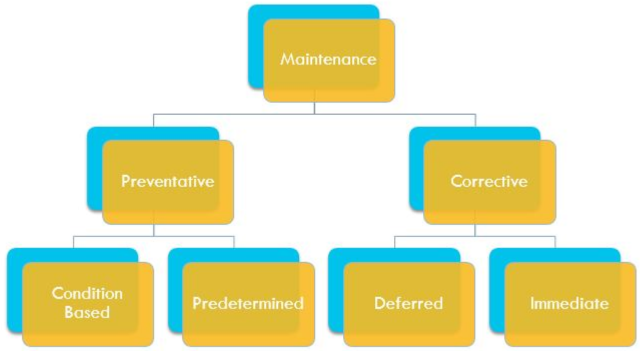
where t is the time step, C_t is the cash flow in step t, i is the interest rate and N is the total number of time steps. Common time steps are one year, one quarter, or one month. While, annual cash flow is most popular, it requires an annually averaged interest rate and does not reflect the timing in the year.

The NPV decision rule states that the selection from alternatives can be made according to the ranking of NPVs. NPVs can be used to make a decision among mutually exclusive projects, which means selection of one causes the exclusion of others [37].

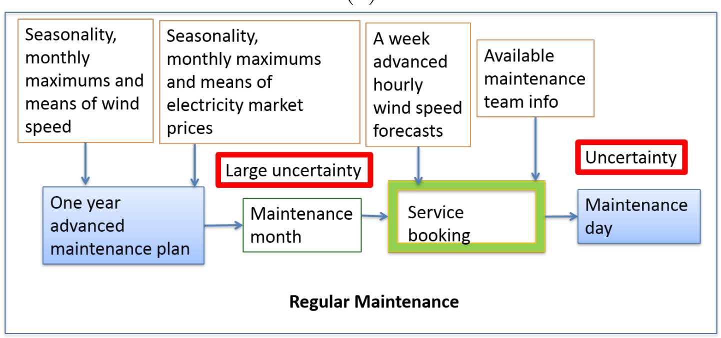
Chapter 3

Wind Farm Maintenance Decisions, Key Factors Contribution for Preferable Scheduling

The execution of any maintenance policy requires in advance scheduling performed by O&M engineer or planner (whether or not a calendar based policy considered). Figure 3.1a shows a classical diagram, which is valid for all technologies. If maintenance interventions are executed before a fault detected, these interventions are result of a preventative policy. Preventative maintenance consists of two sub groups, condition based and predetermined. Condition based maintenance interventions may require a condition monitoring system to be installed in the wind turbines. Here the intervention decision depends on the findings obtained from online measurements and/or offline inspection information. If these intervention are carried out following a forecast derived from the analysis and evaluation of the online measurements and/or offline inspection information obtained form the wind farm, maintenance policy is named as predictive maintenance. On the contrary predetermined interventions are performed solely based on calendar and fixed intervals. However, for both of them the intervention day must be scheduled in advance. If maintenance interventions are executed after a fault detected, these interventions are result of a corrective policy. Corrective interventions could be whether deferred or immediate. Both of these intervention types also require scheduling of a maintenance day in advance.



(a)



(b)

Figure 3.1: Maintenance scheduling procedure (a) maintenance policies (b) application of preventative policy

In Figure 3.1b, it is shown the preventative maintenance policy planning diagram. O&M engineer or planner must consider peaks of electricity market prices, especially for predetermined preventative maintenance (calendar based) and deferred corrective maintenance, when generating one year advanced maintenance plan. In the next chapters, real data based interventions are studied and all these interventions are performed in May, July and October in Spain as a result of one year advanced maintenance plan which was executed through a year. In this chapter, at first we will study electricity market prices and evaluate which months have maximum revenue flows. In the second section, we will propose an analysis framework for the preferable inspection data type and collecting periods by associating environmental data with major component failures resulting in different downtime durations. According to the standard of the maintenance terminology [143], if a maintenance intervention decision is made considering only the pre-selected intervention intervals, without considering the condition of the asset, this intervention is related to the predetermined maintenance policy. For the wind farms, if the condition of the assets is neglected for the intervention decision, then the electricity market prices and the electricity generation possibilities can be considered in order to generate a cost effective preventative maintenance plan. Energy estimations can be performed using the wind resource assessment tools, which are widely covered in the literature. Energy yield estimations became one of the well-established standardised procedures. However, the literature is scarce for the electricity market price behaviour considerations in predetermined wind farm O&M planning. Therefore the first section of this chapter contributes to one year advanced maintenance plan preparation. Whereas, if a prescribed criteria is considered on the evaluation of the status of an asset, resulting intervention decision is related to condition based maintenance policy. In the second section such criteria are investigated, this section provides useful information for a maintenance day, as shown in the second blue shadowed block of Figure 3.1b.

In this chapter, the major decisive parameters for the optimal maintenance strategies are presented. For a grid connected wind farm, revenue is obtained selling the generated electricity according to one (or the combination of some) of the following options: market based pricing, subsidy frameworks (feed in tariff or premiums, green certificates) and power purchase agreements [144]. Nowadays, onshore wind has started to be classified as a mature technology without any need of a subsidy among EU countries and the most popular choice is the market-based pricing. If we face to O&M decisions, the main aim is always focused to ensure selling the maximum amount of energy possible at the time of the highest electricity market price. To this purpose, dynamics and dependencies of the electricity market price are analysed in the first section. Then, the associations between weather, major component failures and wind turbine downtimes are investigated in the second section.

3.1 Electricity market price

The main research questions that will be studied here are listed below.

- Which months are better to perform a scheduled component replacement or a maintenance visit considering electricity market price?

- For electricity market price estimations, which variables can be tracked as information providers?
- Which one is important for the peak and the bottom electricity market prices in a country?
  - Electricity generation and consumption figures of a country
  - Role of a country in transnational electricity trade
- What are the influences of the generator technologies on electricity market price regarding their ability to respond timely to demand?
- How do the drivers of the electricity market price vary among different EU countries?

Firstly, general information regarding the electricity market price calculation procedure in EU, and the reasoning behind the selection of the studied case countries will be given. In this study Spain, Germany, UK, France and the Netherlands are investigated. A short summary of the influential variables that are reported in the literature for the electricity market price is provided, before introducing data and variables used in this study. Data section covers the information regarding available data sources and the details of the used variables such as resolution and units. Secondly the key figures, such as electricity demand and generation of electricity per country will be discussed. Then, demand-price trend analysis will be presented as this information is useful for the evaluation of the long-term direction and the similarity investigation between time series of electricity demand and electricity generation versus that of price of electricity. In order to perform a detailed visual analysis of the electricity generation with different technologies and the electricity demand, level plots are generated. However, all these time series comparison analyses do not necessarily assure the existence of cause-effect relationships between investigated variables and the electricity market price. Exploratory causality analysis can be performed with experimental set-up or statistical tests. There exist several techniques, but the most frequently used test for such investigations is Granger-causality [122–126]. Therefore, Granger-causality analysis will be used to test statistically those cause-effect relationships. Then, overall and monthly correlation coefficients for generation by technologies per country, demand and price per country, and importance ranking for demand and electricity generation by different technologies relative to electricity market price for each case country will be reported in the remainder of the study.

Electricity market

According to a briefing by the European Parliamentary Research Service (EPRS), the electricity market is one of two the constituent mechanisms for the electricity system. The other one can be defined as the physical infrastructure (generation and transportation). The physical infrastructure or the grid represents the flow of the electricity, whereas the electricity market stands for the flow of the money. Various stakeholders get involved in this flow, such as suppliers, who are in between electricity generators and consumers, and wind farm owners, who belong to the group of electricity generators. The consumers are the end-users, or those who pay the bills [145].

Each electricity market shows principal working differences depending on its participants and its working zone. From a historical perspective, the national electricity markets have been evolving to transnational regional ones by market coupling, a statement which can easily be confirmed by tracking quarterly reports on European electricity markets [146–148]. The findings on these reports pave the way through an interconnected, single, internal energy market, which consists of gas and electricity markets for all European Union (EU)[149]. With this aim, national monopolies were converted into liberalised markets. As a result of market coupling efforts and need for functional transnational trade, for the majority of electricity prices in the EU, day-ahead, hourly bidding working principles are in use (half-hourly, block, and flexible bidding also exist). The price and volume of energy for a specific hour are determined by the meeting point of the supply and demand curves according to the marginal pricing model for all European markets, EU Pan-European Hybrid Electricity Market Integration Algorithm (EUPHEMIA) [150]. The pricing approach needs one day in advance load and generation forecasts and then, prices can be addressed in a timely fashion with the help of the balancing markets' actual demand.

The present study uses the historical electricity market price data from selected EU countries. The interpretation of the findings of electricity market price analyses is given from a wind farm owner's perspective.

Country Selection

Since the wind farm owner's perspective is prioritised in this study, it can be assumed that the EU countries which correspond to significant wind power generation must be focused. It will be easier to respond to the interests of a wind farm owner when his/her major asset's location is taken into account.

Figure 3.2 shows the EU country ranking for cumulative onshore and offshore wind power. It can be seen that Germany, Spain, the UK and France are placed in the top four.

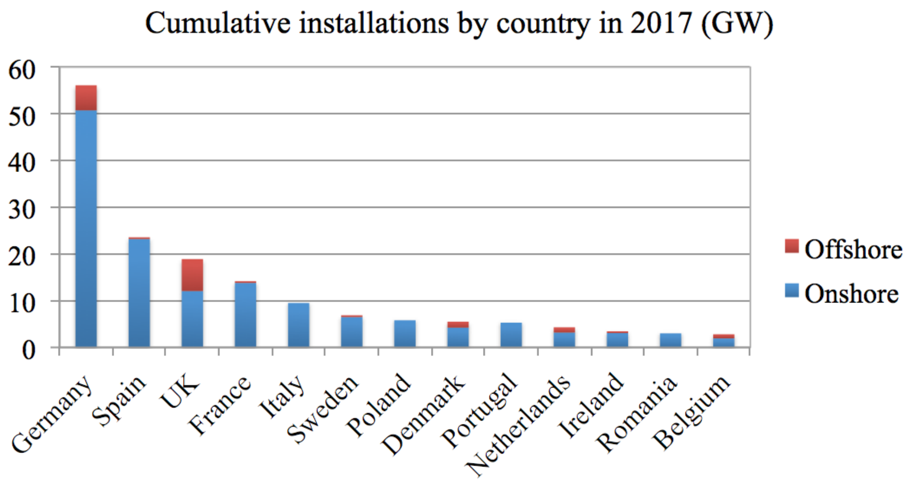


Figure 3.2: Country ranking for cumulative onshore and offshore installations in 2017 [151]

The biggest cumulative power (Germany), one from the south (Spain), one balanced onshore and offshore (UK), one with long term nuclear energy investments (France), and one emerging (Netherlands) are selected. The Netherlands is selected because its wind energy contribution might change drastically according to 2030 energy scenarios. In onshore wind, its place is estimated as the eighth, however in offshore it is expected to reach the third place [152].

Literature based known relationships

The electricity market price of a country is affected by various dynamic parameters. These influential parameters can be listed as energy mixture, energy production by technologies, and total electricity demand of the country. It is possible to make a deeper analysis by studying into oil prices, electricity import/export trends, and even the consumer's daily routine to understand peak load times. But in this study, we are not aiming to provide such a detailed report.

For an hourly analysis, demand (load) is one of the decisive parameters. Different types of power plants contribute to the energy generation for addressing this electricity demand. The classification of the generators can be done by their technology as well as their preferred usage time based on the amount of the electricity demand. By technology, generators can be grouped as fossil (coal, petroleum, natural gas, etc.), renewable (wind, solar, tidal, etc.), and nuclear [153]. With this kind of classification, the drivers for the electricity market can be highlighted as fuel costs and weather conditions.

In the second classification type, there exist three groups for power generators as base load (cheap power production, expensive shut-down and/or restart such as hydro, nuclear and coal power plants), mid-merit (such as diesel power plants, wind and solar generators) and peaking power plants (cheap to keep inactive, expensive to run, and not very efficient such as natural gas plants)[154].

In Figure 3.3, the different demand zones are shown. The electricity generation technologies can be classified being in the role of one of the supply sources for these zones.

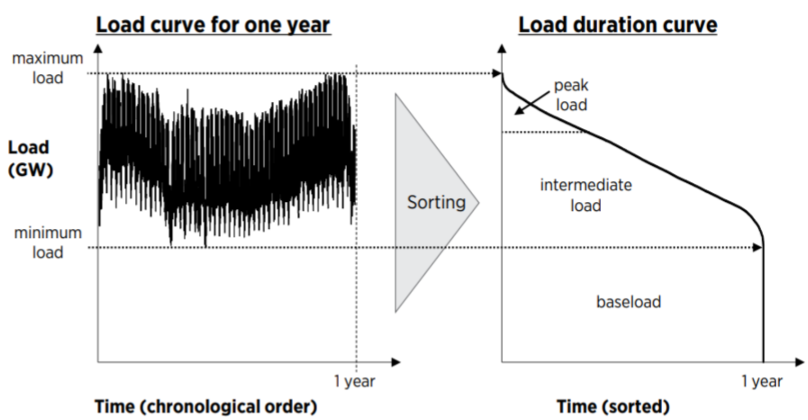


Figure 3.3: The load zones: base load, mid-merit (intermediate) load and peak load [155, 156]

The supply role of an electricity generation technology may vary depending on the country analysed. Base load, mid-merit load and peaking are the electricity demand characteristics [155–157]. Base load units are frequently reported with load factors (the percentage of hours that a power generation unit is capable to operate at its maximum capacity for a given time period) between 75% and 95%. For the mid-merit load units, these load factors range between 40% and 60%. Whereas, peaking units operate at very low annual load factors ranging from 5 percent to 15 percent. For the intermittent renewables the load factors range between 20% and 40% [158].

While wind and solar groups can also be used as mid-merit and partial peak demand supply, although their production depends on weather, they are still capable of providing electricity during afternoon peak demand times. All these power plants address the hourly and daily demands of the consumers, who have an hourly and daily pattern for working and school (mid-electricity demand), resting and dining (maximum household electricity demand), sleeptime (minimal electricity demand), etc. Therefore, the electricity demand shows severe periodicity (hourly, daily, monthly and seasonally) and country-based differences (traditional dining hours, national days, sacred days, etc. [159]). More specific results can be derived from the literature. For example, under higher levels of wind energy generation and penetration on electricity grid, reduction in average electricity prices did not observed according to simulation results, although it was anticipated to be the opposite [160]. Data-driven analysis are needed for further understanding of electricity market price mechanisms.

3.1.1 Data

The electricity market data of five EU countries obtained from the ENTSOE platform [161], are used in this section. Table 3.1 summarizes the main input groups and the resolution of these variables. As implied by their name, day-ahead market price and day-ahead total load represent singular variables as price and load. However, actual generation per production consists of a multi-variable input set.

Table 3.1: Summary of analysed data

Data	Germany			France			Spain			UK			Netherlands		
Resolution (minutes)	15	30	60	15	30	60	15	30	60	15	30	60	15	30	60
Day ahead market price			x			x			x			x			x
Actual electricity generation	x					x			x		x		x		
Day ahead total load	x					x			x		x		x		

Some of the technologies are not available in all five countries and electricity generation is limited from some specific technologies. To obtain comparable results, the common and available technologies are selected here as relevant.

As different electricity generation preferences exist among countries, there also exist fundamental trade zone differences. These trade areas are known as bidding zones which are the largest geographical areas within which market participants are able to exchange energy without capacity allocation [161]. It is worth to mention that some of the case countries do not have one single bidding zone (BZN). In this manner, the bidding zone containing Germany, Austria and Luxembourg (BZN | DE-AT-LU) is considered for Germany. BZN | GB for UK, which does not cover single electricity market of Ireland, is used. Spain, France and Netherlands have a single bidding zone in their electricity market structure.

To double check the values obtained from ENTSOE platform, additional Eurostat electricity generation statistics are used [162]. However in this source, data are not given as hourly observations. Moreover, it is only possible to reach final breakdown of energy mixture in EU countries, which is given as percentages. Energy technology details are limited, available with five generic groups such that conventional thermal, nuclear, hydro, wind, solar and geothermal & others. Regarding the completeness level and consistency of the ENTSOE data, a detailed review is available in literature, which is based on a snapshot of data from 2015 and 2016 [163]. The paper concludes that ENTSOE data are capable of serving the scientific research purposes.

In this study, the analysis covers the period from January 2015 to October 2018, because in the case of Germany, the data were available until the end of September 2018 due to the bidding zone setting changes between Germany and Austria [164], what means that from October 2018 Germany data have different demand and load contributors. Therefore up-to-date Germany data will not be comparable with other case countries' datasets in time series analysis. The analysis must be performed before or after the bidding zone setting change. However, the data after the bidding zone change were not available for at least one full year, when the present study was performed. Since, it was not possible to obtain monthly and seasonal characteristics of the data after the bidding zone change, the data before the bidding zone change were used.

3.1.2 Overall generation by technologies and demand for the studied countries

In order to visualise the major differences in electricity generation and demand, the overall sums per country and per technology are analysed in this section. The electricity demand of the case countries is given in Figure 3.4.

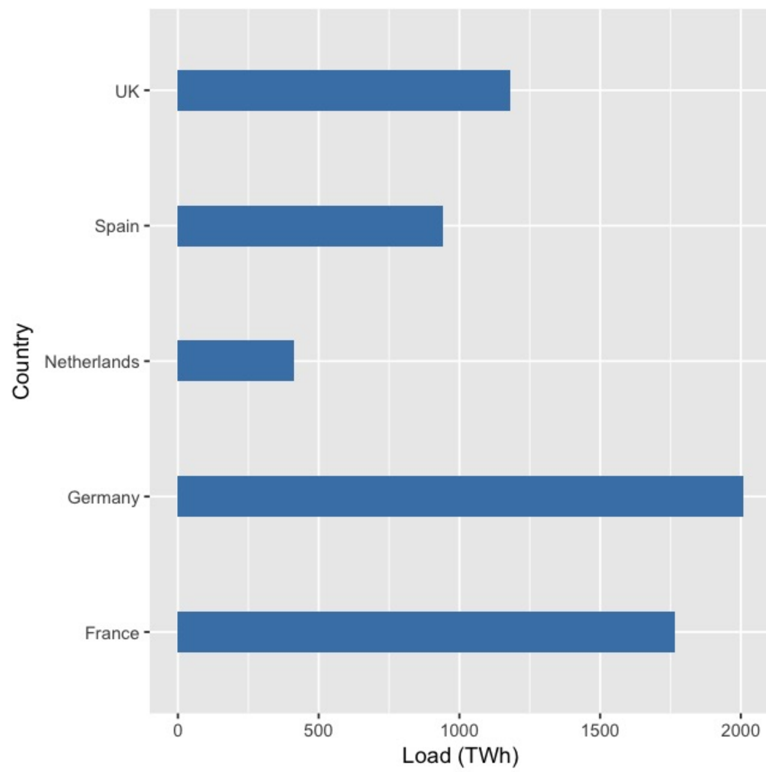


Figure 3.4: Electricity demand per country from January 2015 to October 2018

As expected, Germany is the leader and France comes in second place, the UK is third and Spain fourth. The Netherlands comes in fifth place with relatively smaller electricity consumption. Total electricity generation values for the different technologies can be found in Figure 3.5.

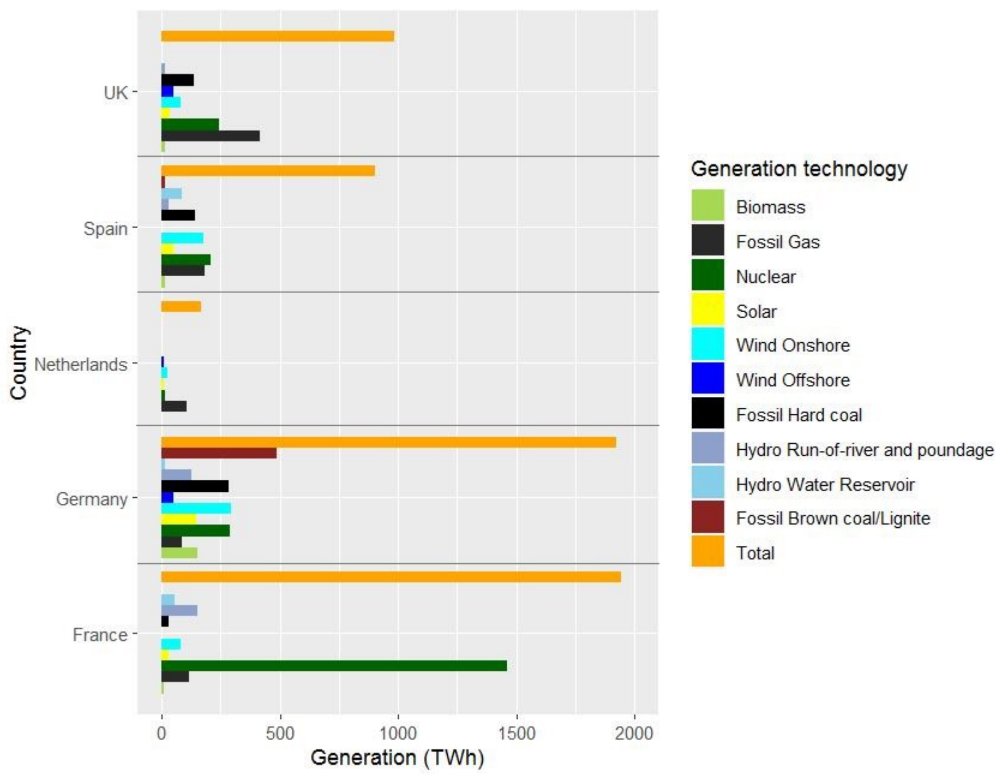


Figure 3.5: Electricity generation per country by technologies from January 2015 to October 2018

Table 3.2 provides a summary of the electricity import export balance (data taken from [165, 166]) for the case countries.

Table 3.2: Electricity Import-Export flow

Case country	2015 (GWh)	2016 (GWh)	2017 (GWh)	2018 (GWh)	Characteristic
UK	21071	18912	16391	17666	Net importer
FR	-62987	-40980	-40236	-56769	Net exporter
DE	-51788	-53744	-55367	-51138	Net exporter
ES	-133	7670	9171	10360.5	Net importer
NL	9108	5186	3897	10810.5	Net importer

The export/import balance column stands for the overall trade of the country. Therefore, this sum also covers trade with other countries which were not analysed in this study. Considering the electricity trade, indirect parameters, such as neighbouring countries' energy mixtures, demand, and long-term gas and electricity trade agreements are expected to exert an influence upon the electricity market prices of a country. Therefore, in this study it is claimed that if a county is not balanced, nor a net exporter, nor a major contributor to its bidding zone, its electricity price will fluctuate rather differently from its own generation and demand. The UK, Spain and the Netherlands are all net importers.

The difference between electricity demand per country (in Figure 3.4) and total electricity generation per country (in Figure 3.5 orange bars) stands for the electricity export/import traffic between the countries and also the excluded generation sources technologies. It is worthwhile to highlight that these countries' electricity trade has a complex structure and is not limited to the studied EU countries.

Figure 3.5 reveals the energy mixture of the case studies. In the UK the first three generation sources are fossil gas, nuclear and hydro run-of-river & pondage. The Netherlands depends a lot on fossil gas for the electricity generation. Germany shows a balanced energy mixture with various contributors. In France, the leadership of nuclear energy is proven, as anticipated. In the case of Spain, the highest contribution to overall electricity generation in the analysis period was from nuclear. Fossil gas was second and onshore wind power was ranked in third place. One can expect that onshore wind generation has a significant effect on electricity price as one of the major contributors to overall energy generation in Spain. However, this statement can not be made without further analysis due to the working principle of the day-ahead electricity market and intermittent characteristics of wind energy generation.

3.1.3 Summary statistics for electricity markets

In order to grasp general characteristics of the case countries' electricity market price, summary statistics are given in Figure 3.6 and in Table 3.3.

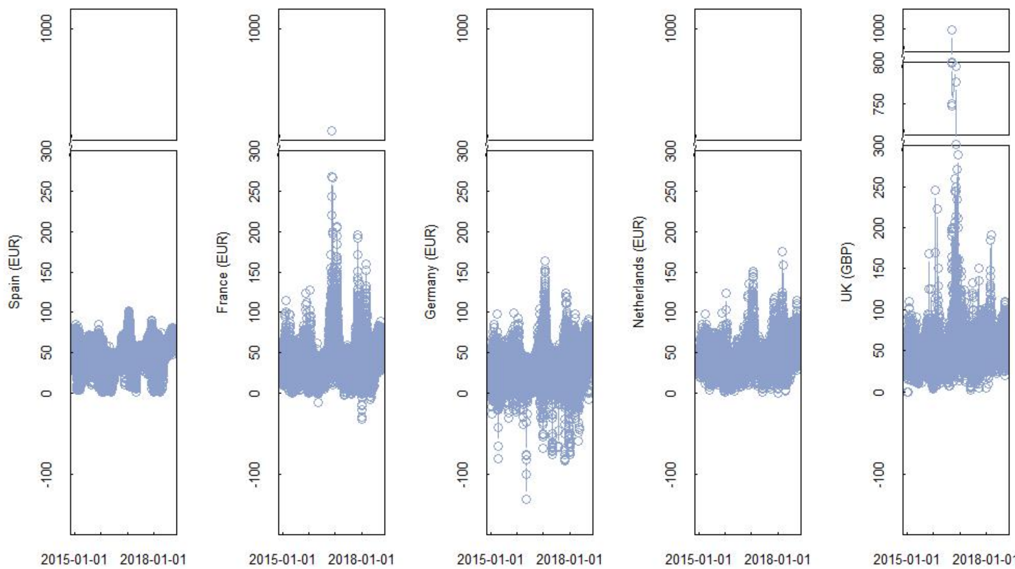


Figure 3.6: Electricity market prices per country from January 2015 to October 2018

Figure 3.6 shows that Spain and the Netherlands data do not contain high extreme observations as other countries. In the cases of the UK and France there exist prices reaching 999 GBP and 874 EUR respectively and these two extreme observations coincide in time. The high electricity trade between France and the UK could trigger such extremes to occur in both countries at the same time. In fact, it is reported that electricity supply issues were recorded due the disconnection of the French nuclear power plants for extended periods in 2016 and 2017. The supply issues reduced the ability of the UK to import the French electricity and were noted as one of the major reasons behind the price spikes occurred in both countries at the corresponding periods [167]. In the case of Germany, extreme prices are observed in negative values.

Table 3.3: Electricity market price statistics during Jan 2015-Dec 2018

Features	Spain	France	Germany	Netherlands	UK
Currency	EUR	EUR	EUR	EUR	GBP
Max	101.99	874.01	163.52	175	999
Mean	48.99	41.23	33.68	39.72	44.69
Min	2.06	-31.87	-130.09	0.55	0
Mode	40	50	30	40	50

Table 3.3 supports visual findings obtained from Figure 3.6. The lowest mean is observed in Germany and the highest mean is recorded for Spain.

3.1.4 Monthly features of electricity markets and demands

Figures 3.7, 3.8, 3.9, 3.10 and 3.11 provide a comparison between the actual demand (load) and day-ahead electricity market price. These analyses are performed by using the Seasonal Trend Decomposition of Time series by LOESS method explained in section 2.3.3. The components of a time series are known as trend, seasonality (systematic cycles), and noise (instant fluctuations). Among them, the trend shows a general direction, long-term movements and tendency in the series [112]. Here a visual similarity investigation is performed between the electricity price and the demand time series. In the monthly trend figures, the local maxima and minima of the curves are shown with points. Here, a peak is defined as an observation point greater than its neighbours, which are other elements within a certain window where the peak observation is centred. Then, using this rule a peak search method was applied [168].

In Figures 3.7, 3.8, 3.9, 3.10 and 3.11 on the top and the bottom axes, months with shorter tick lines correspond to the local minima, whereas months with longer tick lines stand for the local maxima. These figures show hourly demand data on the left axis and hourly electricity price on the right axis, the bottom axis is given for the significant peak months for the electricity market price and the top axis is given for the significant peak months for the electricity demand. When the population and the gross domestic product of each country are considered, it is anticipated that significant differences in the hourly demand values can be seen. Countries with an average hourly demand less than 40000 MWh are given in Figures 3.7 and 3.8 while Figures 3.9, 3.10 and 3.11 show those with an average hourly demand higher than 40000 MWh.

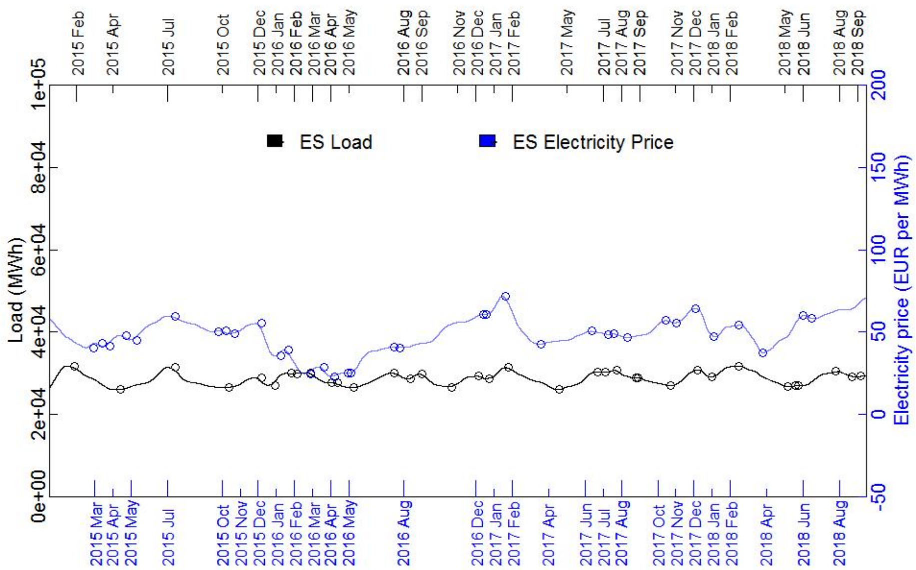


Figure 3.7: Trend analysis for demand and electricity market price Spain trend curves

Figure 3.7 shows the Spanish case. High demand months are February (2015, 2016, 2017), March (2018), July (2015), August (2016, 2018) and December (2015, 2016, 2017, 2018). In March (2016), July (2017), October (2015) and November (2016) there exists a peculiar separation between the electricity market price and the demand. The price curve shows the bottom values in January (2016, 2017, 2018), and the peak values in February (2016, 2018) and August (2015, 2016, 2017, 2018).

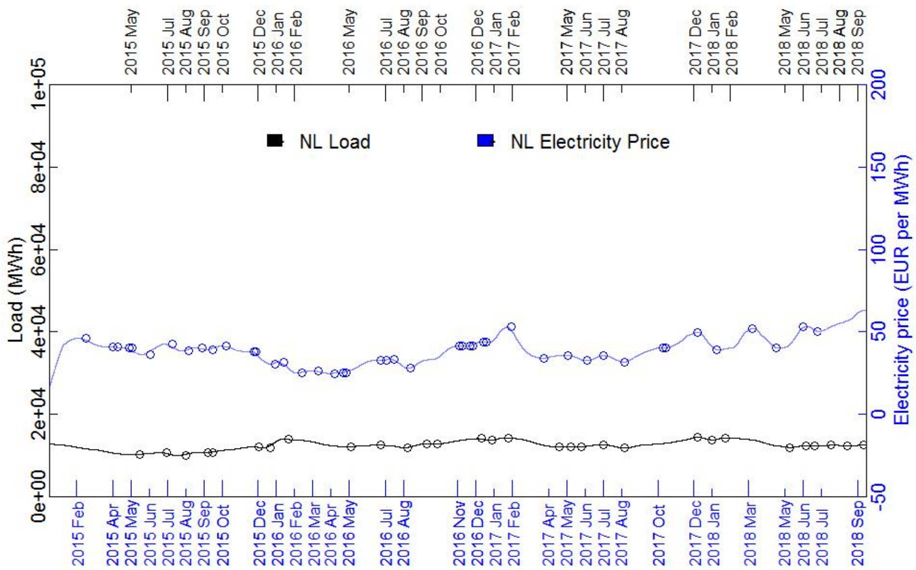


Figure 3.8: Trend analysis for demand and electricity market price the Netherlands trend curves

In Figure 3.8, the Netherlands case is given. In comparison to the electricity demand curve of Spain, the Netherlands has a flatter demand curve, while still can be observed peaks in December (2015, 2016, 2017). However, the price curve has clear peaks in February (2015, 2017) and March (20116, 2018).

Figures 3.9, 3.10 and 3.11 present the case countries with higher electricity demands.

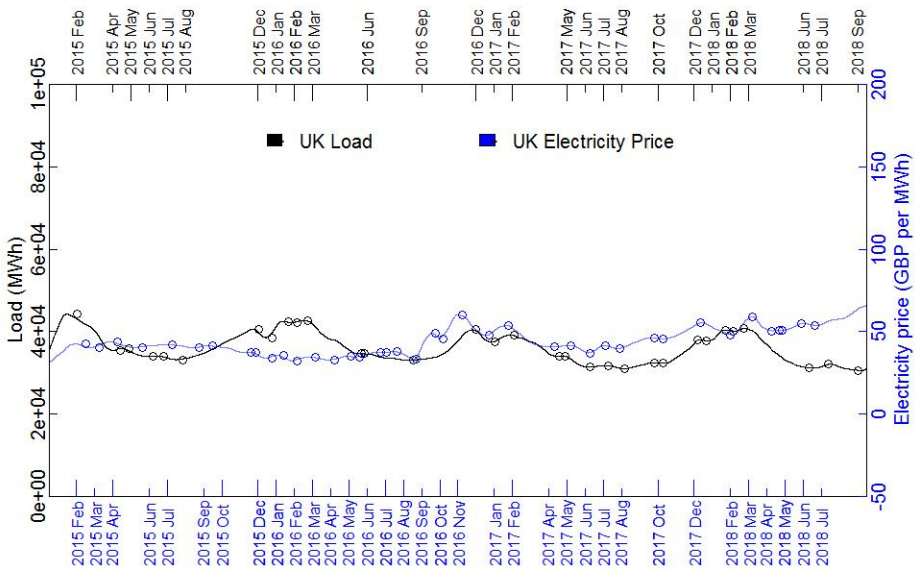


Figure 3.9: Trend analysis for demand and electricity market price the UK trend curves

In Figure 3.9, the UK case is shown. In February (2016) and June (2018), a separation between two time series is observed considering peak values. The price curve has clear peaks between November and December 2016. Whereas, the demand curve has peaks in February (2015, 2016, 2017, 2018) and March (2016, 2018).

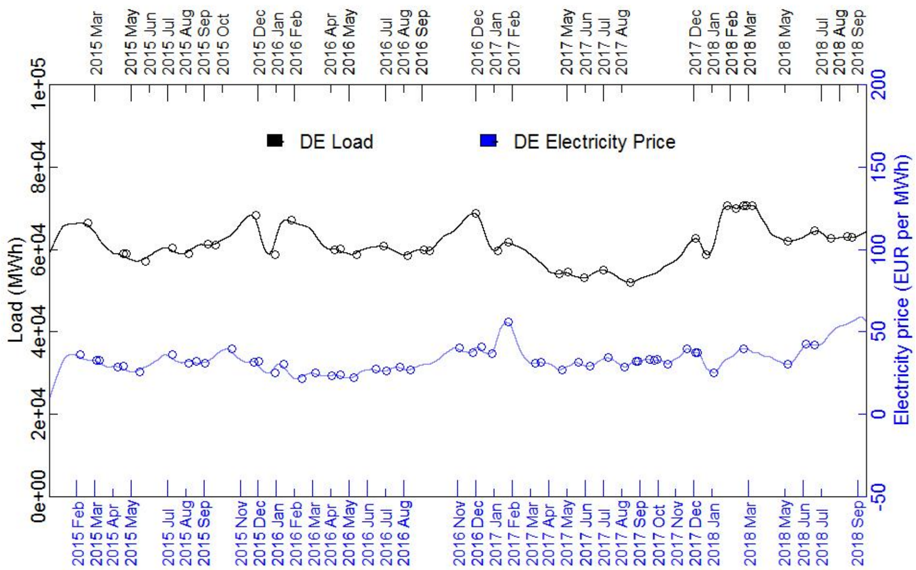


Figure 3.10: Trend analysis for demand and electricity market price Germany trend curves

The German case is shown in Figure 3.10. It must be noted that a strange phenomenon occurs in the case of Germany, because there exist negative electricity market prices. Actually, it is a known anomaly for the German case. When, Germany's electricity supply is not even with its demand, Germany pays its users for their electricity consumption to reduce the stress on the grid [169]. February (2016, 2017, 2018), March (2015), July (2015, 2016, 2017, 2018) and December (2015, 2016, 2017) all have high demand, but the most significant price peak occurs in February (2015, 2017). September 2018 has also a peak, but this month could be a peculiar case due to the infrastructure modifications for the bidding zone change.

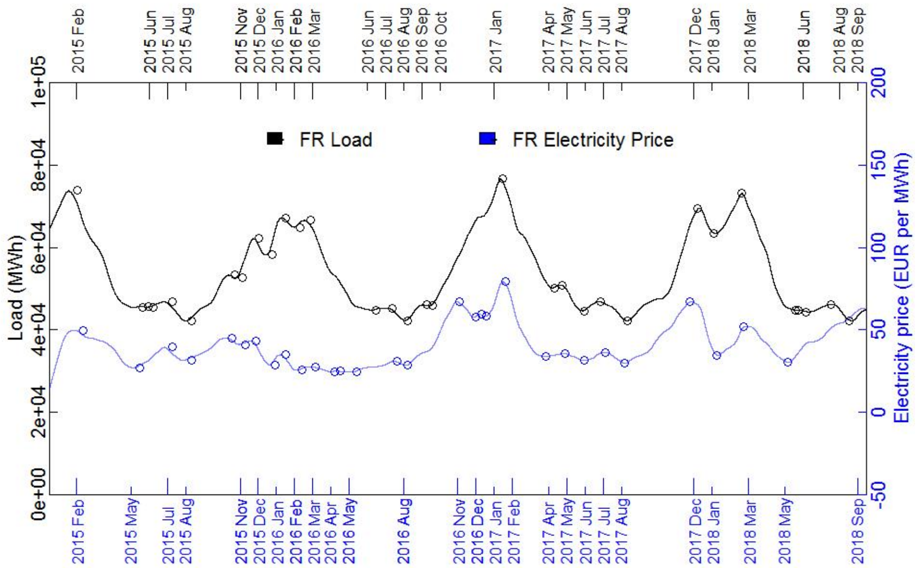


Figure 3.11: Trend analysis for demand and electricity market price France trend curves

Lastly, the analysis for France is given in Figure 3.11. This analysis reveals that France's demand almost doubled in January (2018), February (2015, 2016, 2017), March (2018) in comparison to June (2015, 2018), July (2017), August (2015) and September (2017) records. January (2017), February (2015, 2016), March (2018), July (2015, 2017), November (2015) and December (2015) stand out with high load peaks while February (2015, 2016, 2017), March (2016, 2018), July (2015, 2017), November (2015, 2016) and December (2015, 2016, 2017) have high price peaks.

From a wind farm operation and maintenance engineer's perspective, the bottom price months observed in wind turbine accessible seasons provide very useful information in order to schedule preventative maintenance activities considering revenue. The most of the time, during the winter months the accessibility of wind farms is low due to icing and high wind speeds. Then, we can sum up the most significant bottom price months for a maintenance engineer:

- Spain: April (2015, 2017, 2018) and November (2015 and 2018)

- Germany: April (2015, 2016) and May (2017, 2018)
- France: May (2015, 2018) and April (2016, 2017)
- The Netherlands: June (2015, 2017, 2018)
- The UK: June (2015, 2016, 2017)

Additionally, we saw that, the electricity demand is not alone enough to explain the electricity market price. So, we will investigate other factors such that electricity generation by sources and country policies.

3.1.5 Generation and demand time series

Hourly demand versus generation per sources time series are presented here for visual inspections to determine the impact of the technology (such as demand following, base load provider or peak demand following) of the generator units and energy mix policy of the case countries along with the investigation period. In order to display large amounts of data in hourly resolution, level plots were generated using time index of data as independent and conditioning variable [170, 171]. In these level plots, the load variation is repetitively displayed to make the visual comparison easier. For all the figures present in this section, the color scale is defined from low to high as blue to red respectively. Therefore, for a base load technology, it is expected to see mono colour gradient, a mid-merit load technology gradient changes similar to Load figure, whereas in peak following technology it is anticipated to see matching windows with Load figure considering only red dense zones.

In Figure 3.12, the interpretation of one example figure is explained.

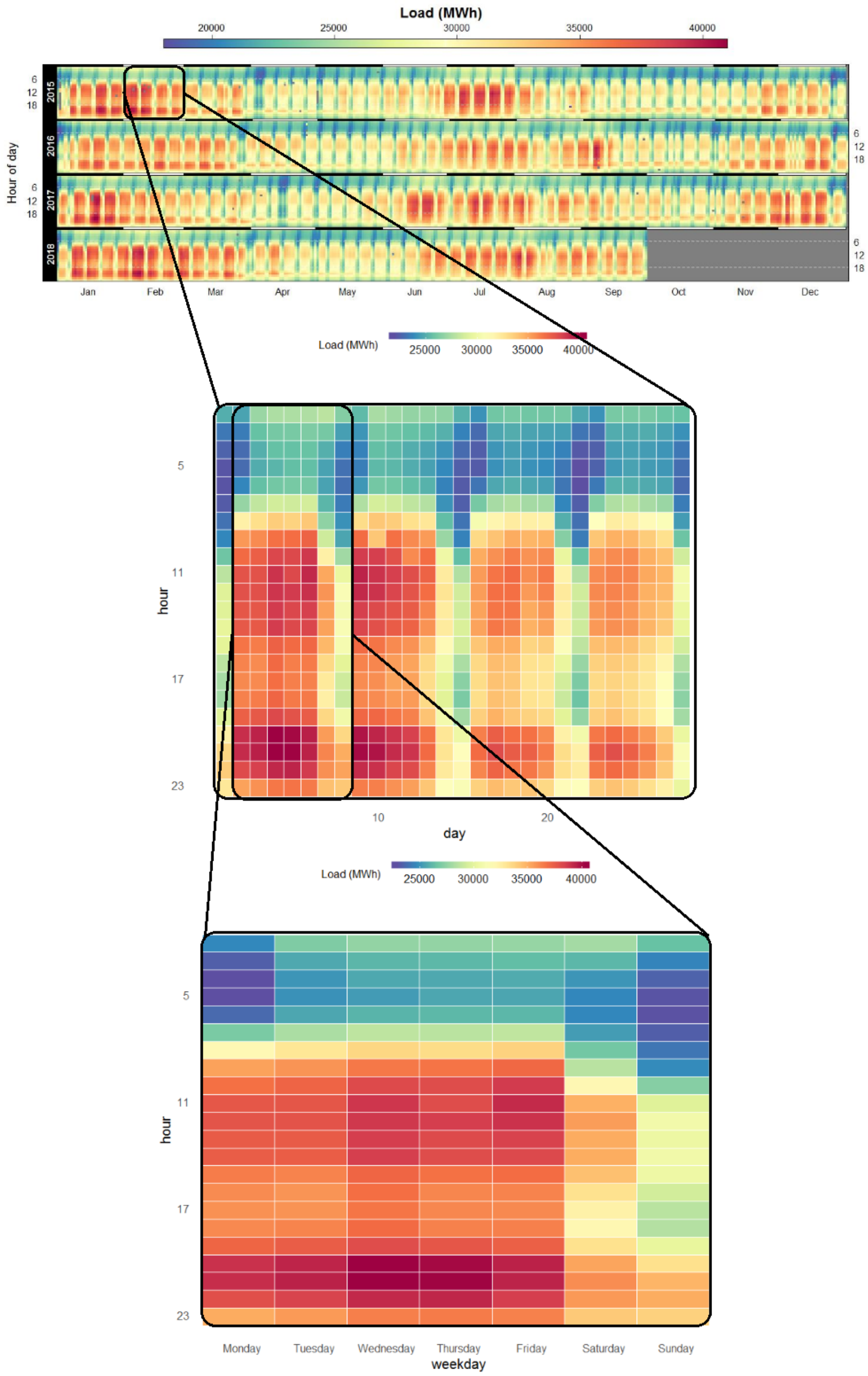
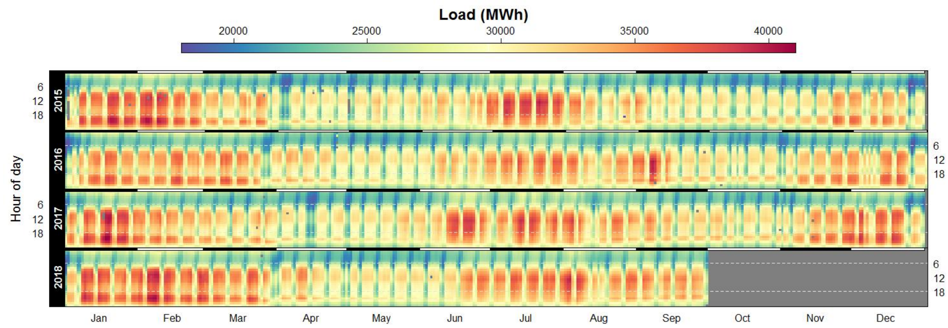


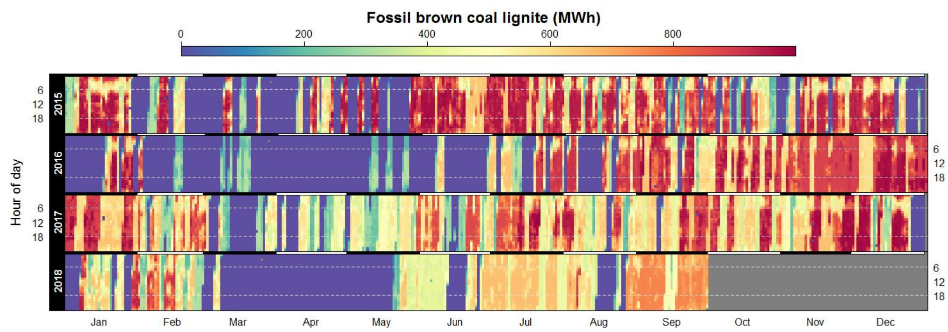
Figure 3.12: Hourly demand series for Spain

In Spain between 14:00 and 18:00, there exists a traditional break period for the majority of business. Here, the plot on top illustrates the demand time series for the investigation period. The plot in the middle stands for February 2015 and the plot in the bottom shows the period between 02^{nd} of February and 08^{th} of February, 2015.

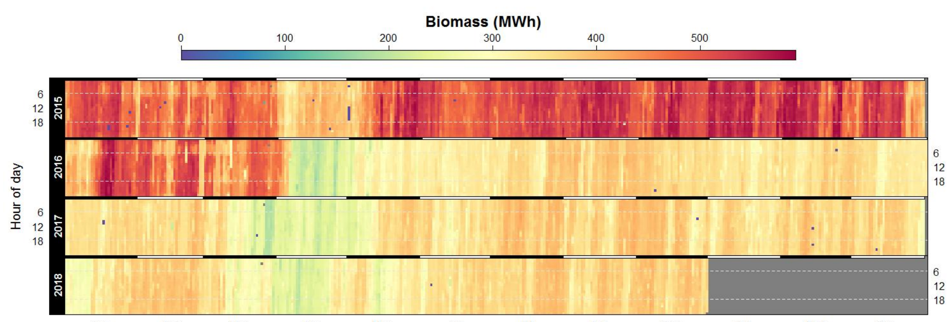
Firstly, Spain analysis are given in Figures 3.13, 3.14, 3.15, 3.16 and 3.17.



(a)



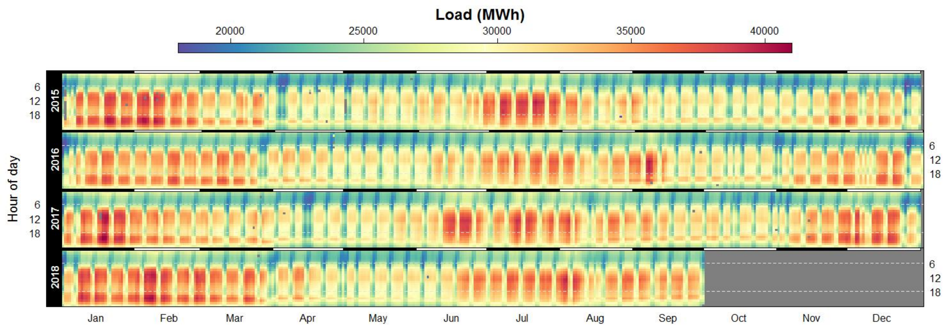
(b)



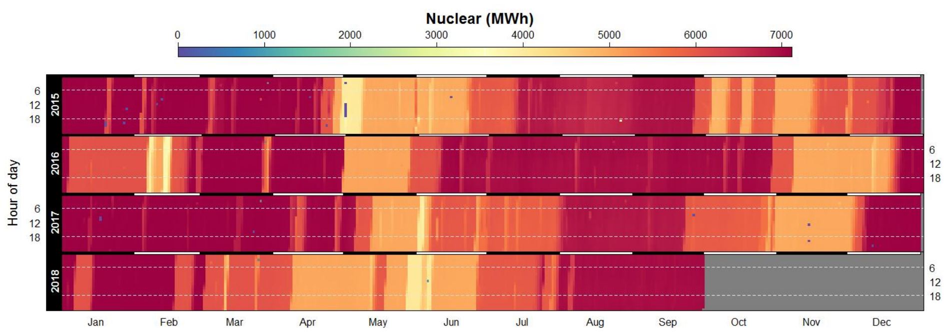
(c)

Figure 3.13: Hourly demand and generation time series for Spain: Load (a) and Electricity generation from Fossil brown coal lignite (b) and from Biomass (c)

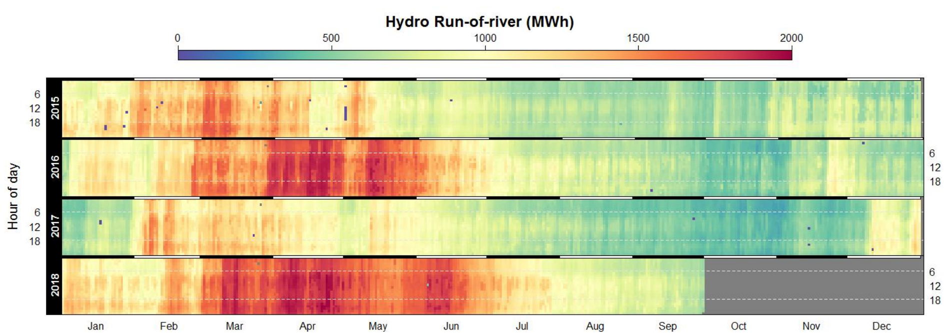
Figure 3.13 shows the hourly generation time series in comparison with hourly demand for Spain. Figure 3.13a displays that in January, February, March, July and December electricity demand is significantly higher and diurnal pattern shows that from 08:00 to 15:00 high demand, between 15:00 to 18:00 moderate demand and from 18:00 to 23:00 again high demand occur. During night time, low demand is observed as anticipated. In Figure 3.13b, electricity generation from Fossil brown coal lignite is illustrated, here it is clear that these generators do not operate continuously and instead generate electricity in high demand months, mainly when the renewables are not available, but their capability to respond hourly demand is not strong. When these generators are taken in operational mode, they generate at least during a couple of weeks or the full month. This technology is neither a base load, nor peak following electricity provider for Spain. It can be said that Fossil brown coal lignite is one of the mid-merit load following technologies. In Figure 3.13c, electricity generation from Biomass is shown and in contrast to Figure 3.13b, a continuous operation is observed. Biomass can be considered as one of the base load providers in Spain.



(a)



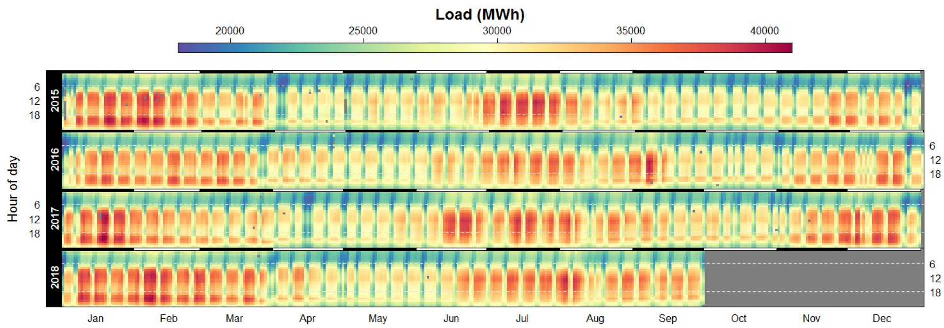
(b)



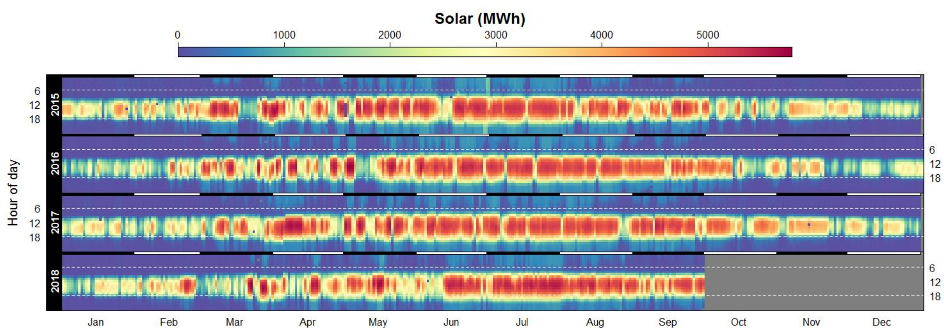
(c)

Figure 3.14: Hourly demand and generation time series for Spain: Load (a) and Electricity generation from Nuclear (b) and from Hydro Run-of-river (c)

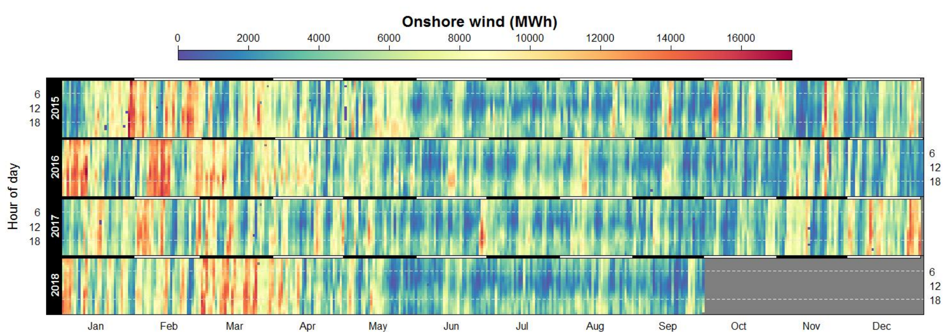
By using the same Load figure explained in the previous illustration, here the electricity generation from Nuclear and Hydro Run-of-river data are analysed. Figure 3.14b displays that Nuclear is clearly one of the base load providers in Spain. A similar statement can be also done for Hydro-Run-of-river, as can be seen in Figure 3.14c.



(a)



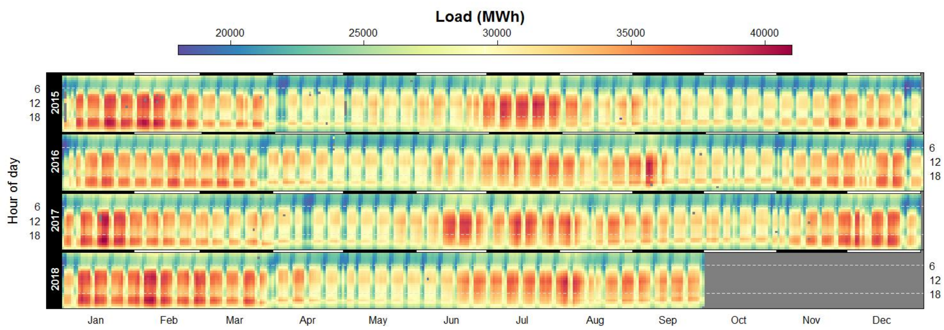
(b)



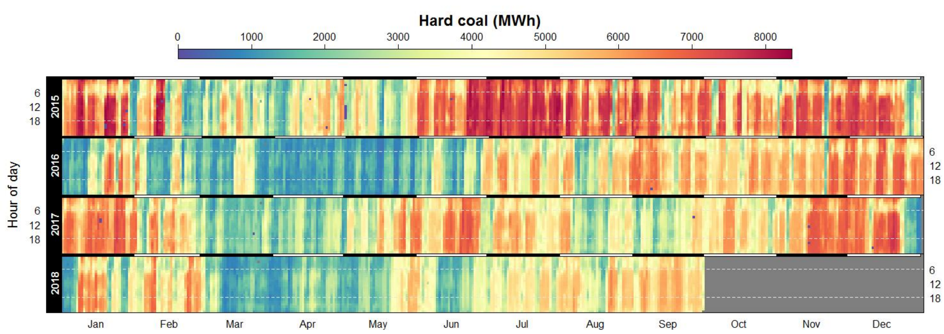
(c)

Figure 3.15: Hourly demand and generation time series for Spain; Load (a) and Electricity generation from Solar (b) and from Onshore wind (c)

Figure 3.15b shows that Solar is capable of generating electricity partially at high demand hours between April and October. In Spain, Solar can be considered one of the peak load following technologies. In contrast, Onshore wind is more of a peak following generator between December and April.



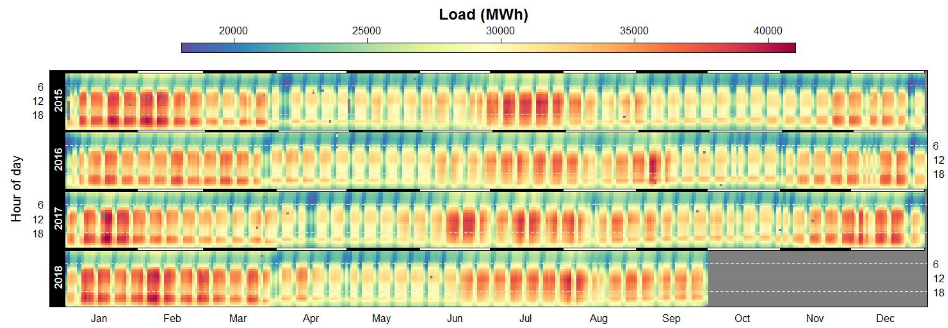
(a)



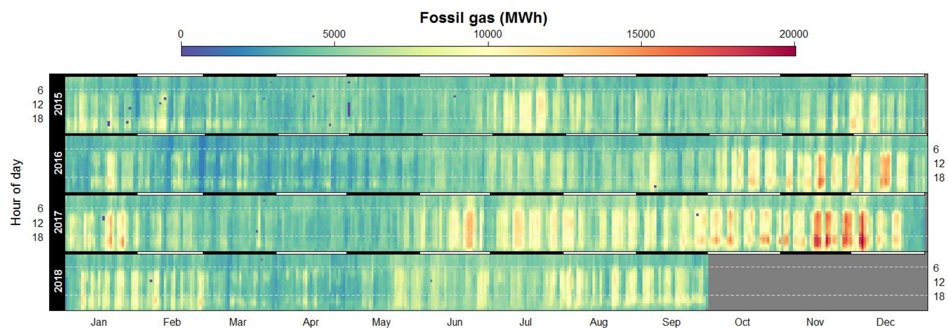
(b)

Figure 3.16: Hourly demand and generation time series for Spain: Load (a) and Electricity generation from Hard coal (b)

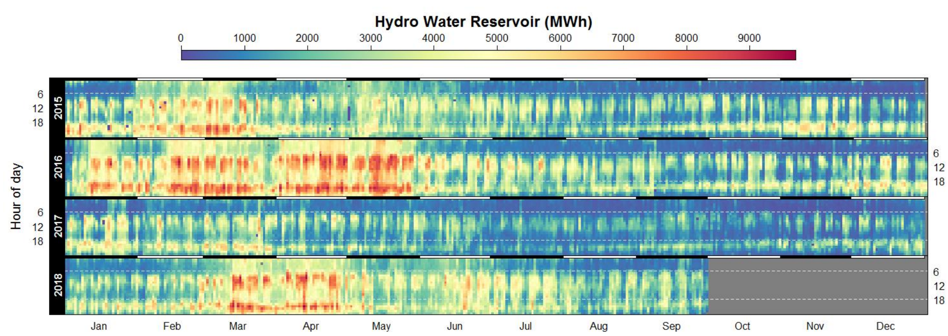
It is shown in Figure 3.16 that Fossil hard coal is capable of following the load.



(a)



(b)

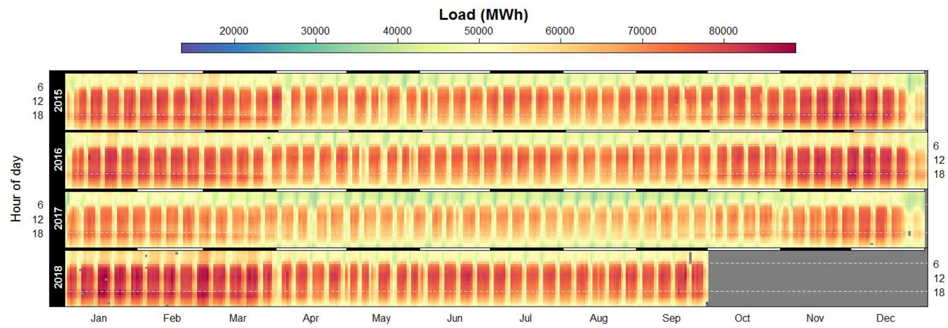


(c)

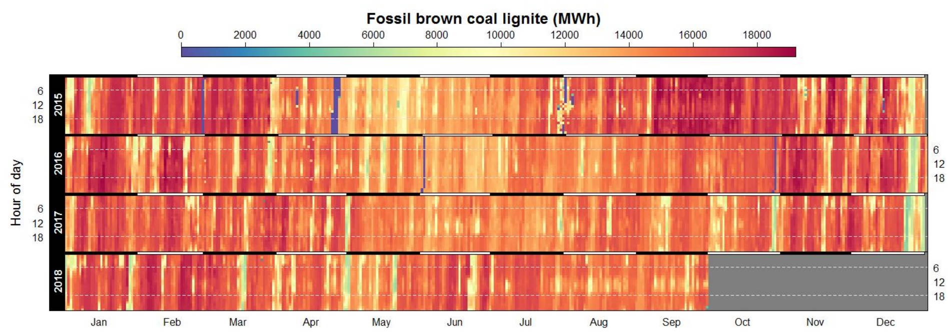
Figure 3.17: Hourly demand and generation time series for Spain: Load (a) and Electricity generation from Fossil gas (b) and from Hydro Water Reservoir (c)

Lastly for Spain, two of the peak load following technologies are shown in Figure 3.17. Both, Fossil gas and Hydro Water Reservoir are capable of responding on hourly demand. Between these two there exists a monthly balance, from January to June Hydro Water Reservoir and from October to December Fossil Gas takes the role of electricity generator technology for the peak load.

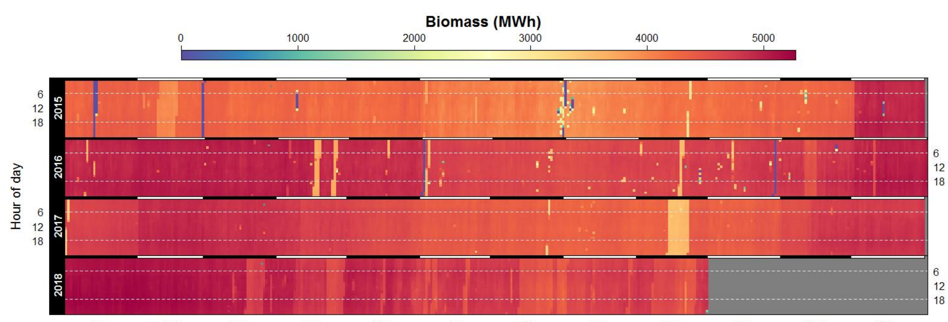
Now we continue presenting Germany analysis, which are given in Figures 3.18, 3.19, 3.20, 3.21 and 3.22.



(a)



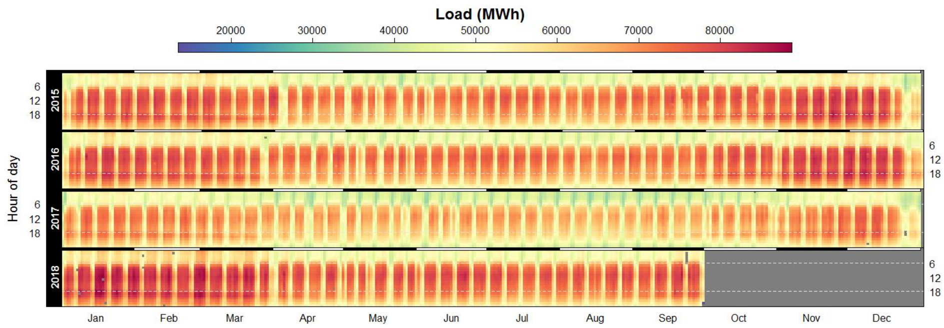
(b)



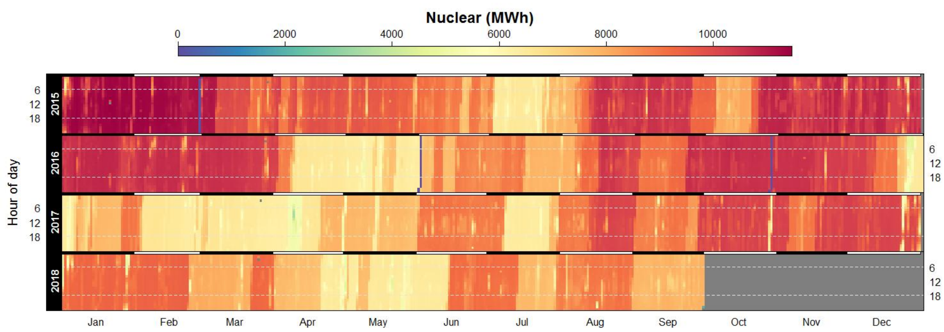
(c)

Figure 3.18: Hourly demand and generation time series for Germany: Load (a) and Electricity generation from Fossil brown coal lignite (b) and from Biomass (c)

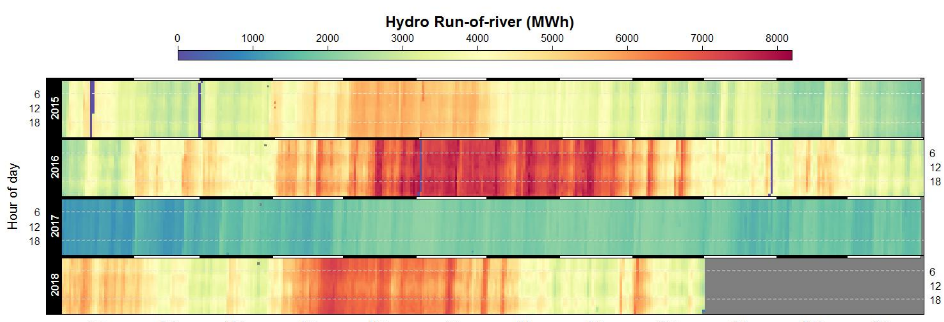
Figure 3.18 shows the German case. In comparison to Spain Load figure especially from 8:00 to 22:00 in diurnal pattern, there exist high demand periods without a clear moderate demand window. Also, the demand difference between months do not vary as much as in Spain, see Figure 3.18a. Electricity generation from Fossil brown coal lignite shows a mid-merit load following pattern, Figure 3.18b, whereas Biomass is clearly one of the base load providers, Figure 3.18c.



(a)



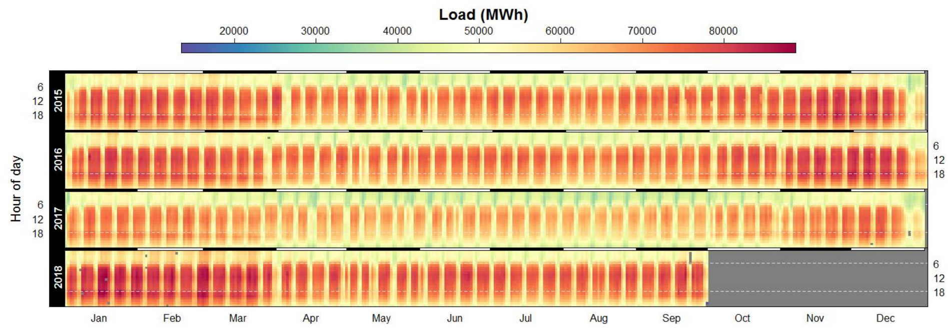
(b)



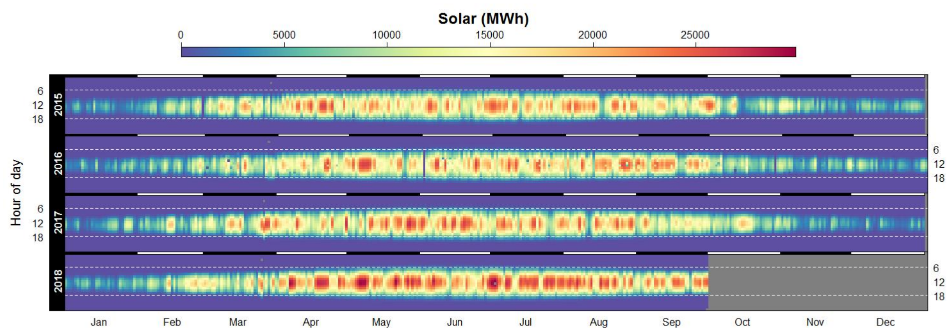
(c)

Figure 3.19: Hourly demand and generation time series for Germany: Load (a) and Electricity generation from Nuclear (b) and from Hydro-Run-of-river (c)

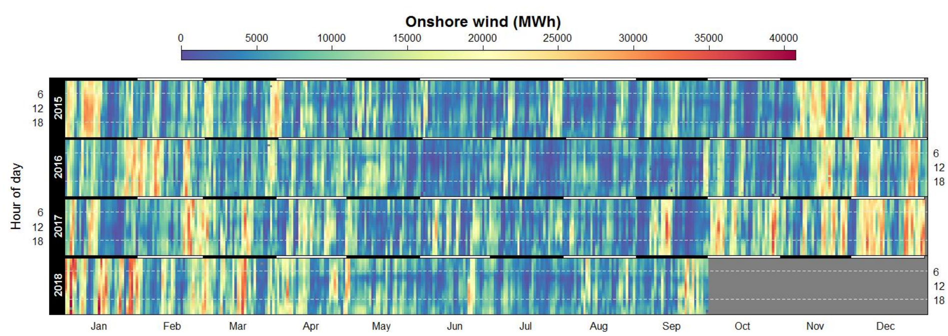
According to Figure 3.19, both Nuclear and Hydro Run-of-river are base load providers but they follow monthly demand and balance each other. Between April and September Hydro Run-of-river is higher and for the rest Nuclear generates more. It can be seen that Hydro Run-of-river shows very low production during 2017. This is also true for Hydro water reservoir, as can be seen in Figure 3.22, and is probably due to drought.



(a)



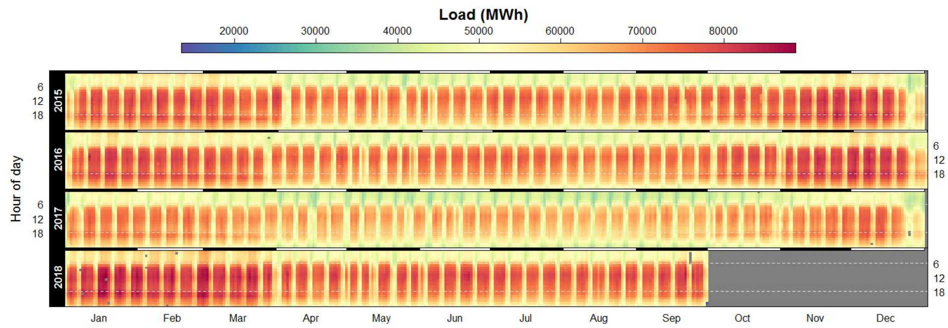
(b)



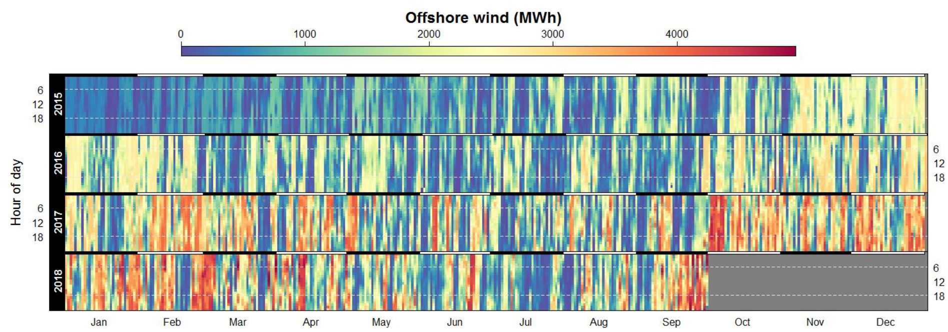
(c)

Figure 3.20: Hourly demand and generation time series for Germany: Load (a) and Electricity generation from Solar (b) and from Onshore wind (c)

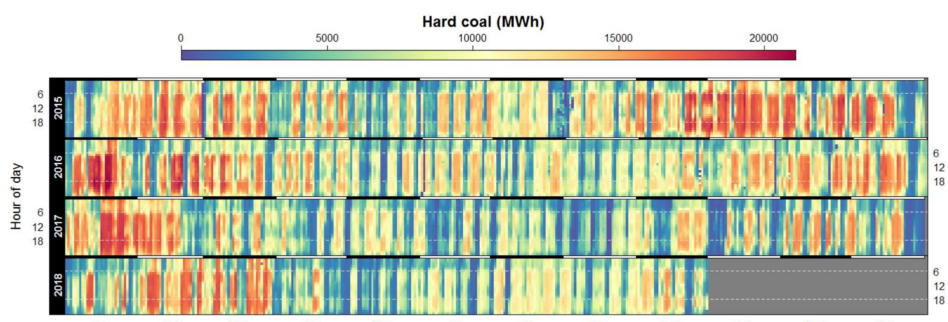
It is worthwhile to remember that although the intermittent energy sources such as wind and solar primarily are dependent on atmospheric variables, they are considered as peak following or mid-merit load technologies. Because it is still possible to operate them or stop their generation considering the demand in a less costly way than many other technologies such as Nuclear. In Figure 3.20, electricity generation from Solar (in diurnal pattern) and Onshore wind (in monthly pattern) show mid-merit load or peak load following generator characteristics.



(a)



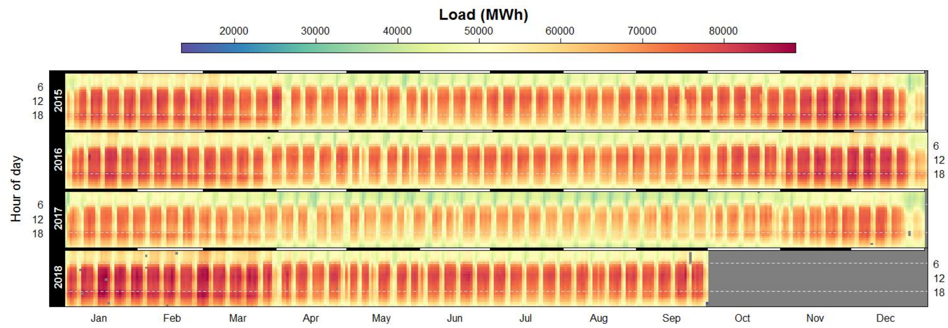
(b)



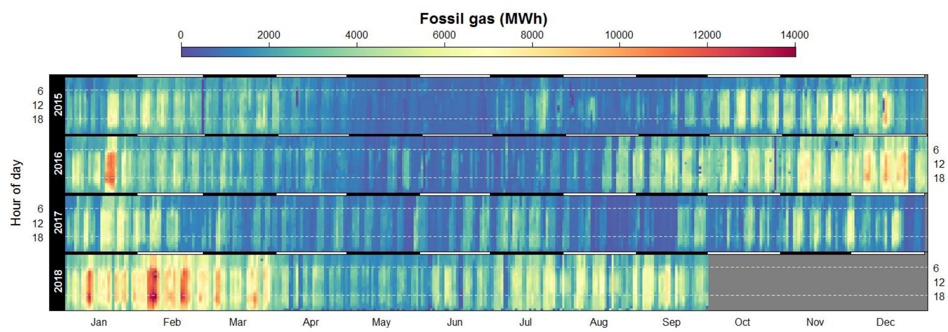
(c)

Figure 3.21: Hourly demand and generation time series for Germany: Load (a) and Electricity generation from Offshore wind (b) and from Hard coal (c)

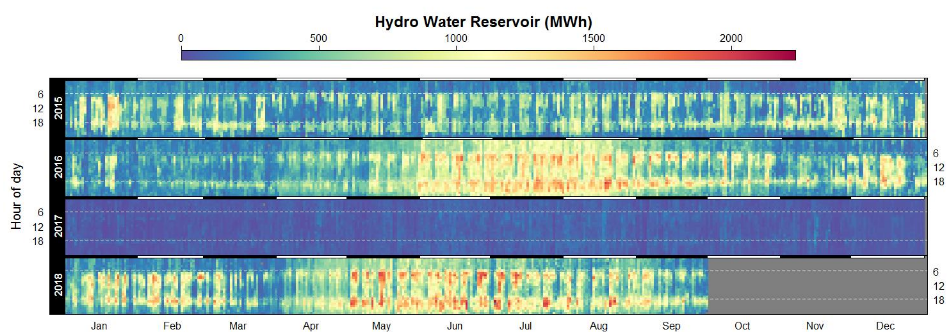
Figure 3.21 shows that Offshore wind generation was not significant in 2015 and in the following years, this technology became one of the supply sources of the mid-merit load. On the contrary, electricity generation from Hard coal shows that Hard coal partially corresponds to peak load in Germany.



(a)



(b)

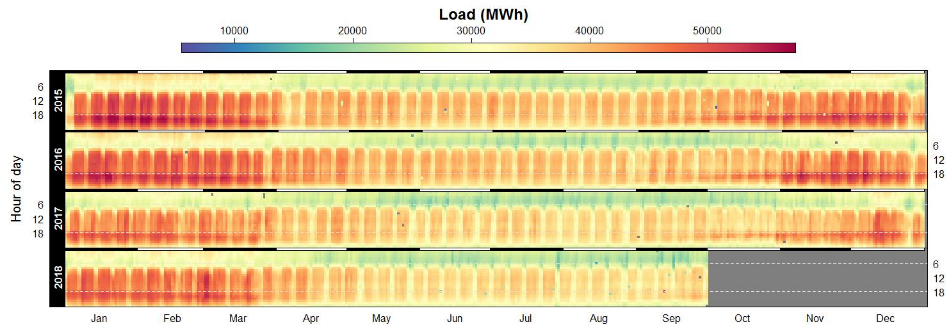


(c)

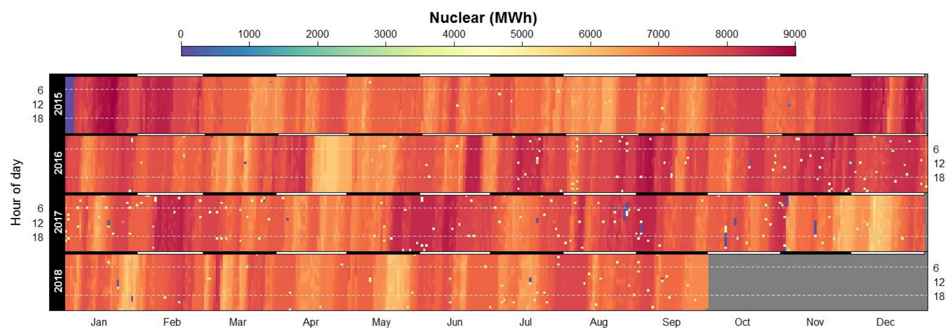
Figure 3.22: Hourly demand and generation time series for Germany: Load (a) and Electricity generation from Fossil gas (b) and from Hydro Water Reservoir (c)

In Figure 3.22, two of the peak load following technologies are. Both of Fossil gas and Hydro Water Reservoir are capable of responding on hourly demand. Between these two there exists a monthly balance, from January to April Fossil gas and from May to October Hydro Water Reservoir takes the role of electricity generator technology for the peak load. It is worth mentioning the absence of Hydro Water Reservoir generation during 2017 due probably to drought, as commented previously.

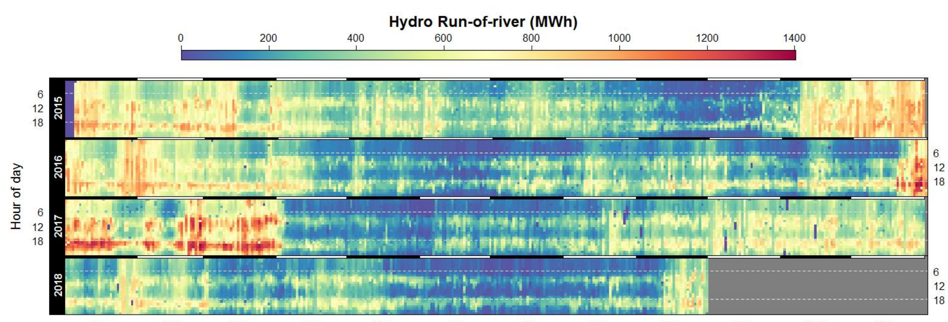
Figures 3.23, 3.24, 3.25 and 3.26 show the UK analysis. Germany, Spain, the Netherlands and France are one hour ahead of the UK due to time zone difference. This effect revealed in Figure 3.23a, high demand hours reported from 09:00 to 24:00 considering Central European Time. From April to November the electricity load is lesser than the electricity load from December to April.



(a)



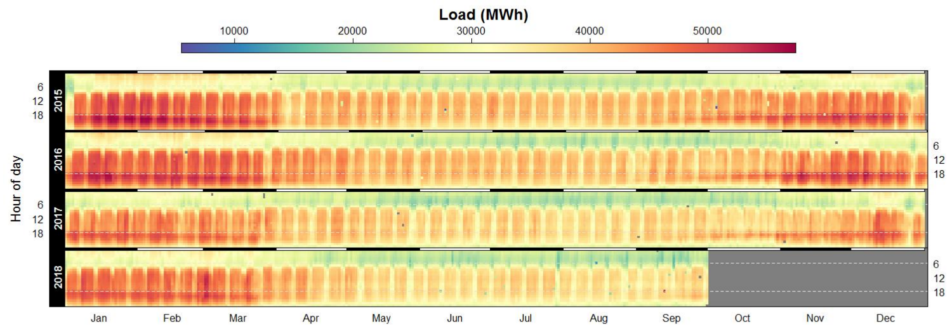
(b)



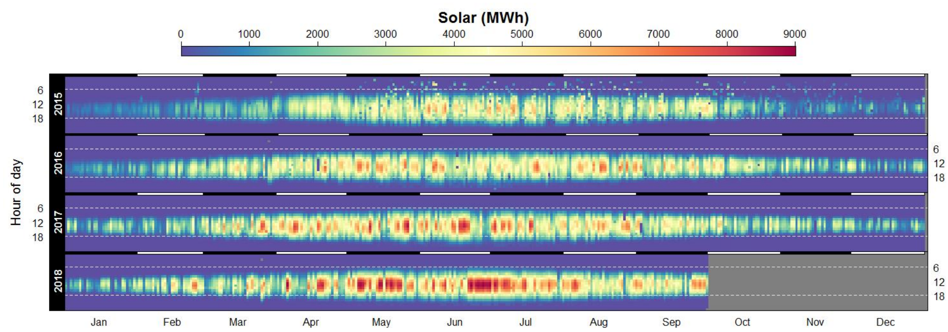
(c)

Figure 3.23: Hourly demand and generation time series for the UK: Load (a) and Electricity generation from Nuclear (b) and from Hydro Run-of-river (c)

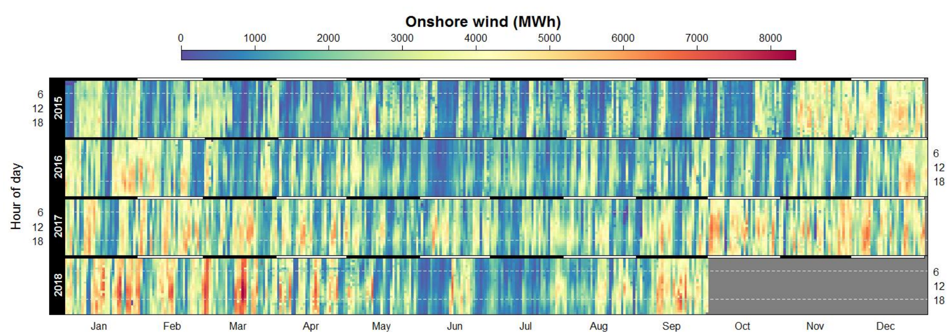
Figure 3.23 shows that Nuclear is base load generator in the UK, whereas Hydro Run-of-river generates electricity for responding peak load.



(a)



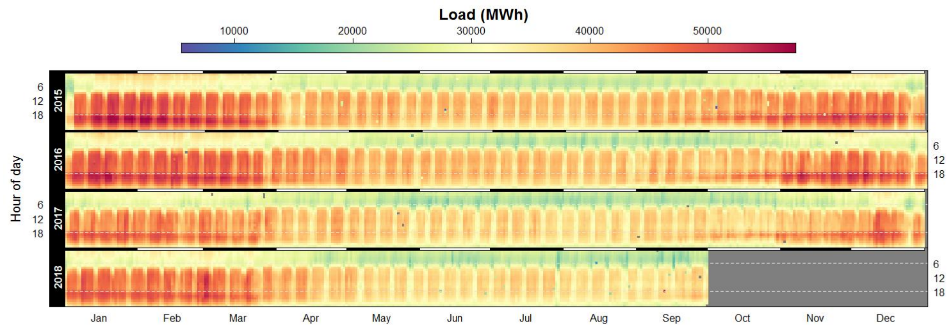
(b)



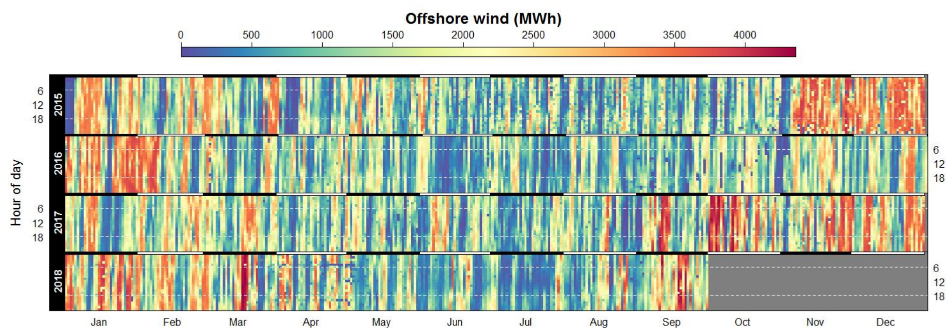
(c)

Figure 3.24: Hourly demand and generation time series for the UK: Load (a) and Electricity generation from Solar (b) and from Onshore (c) wind

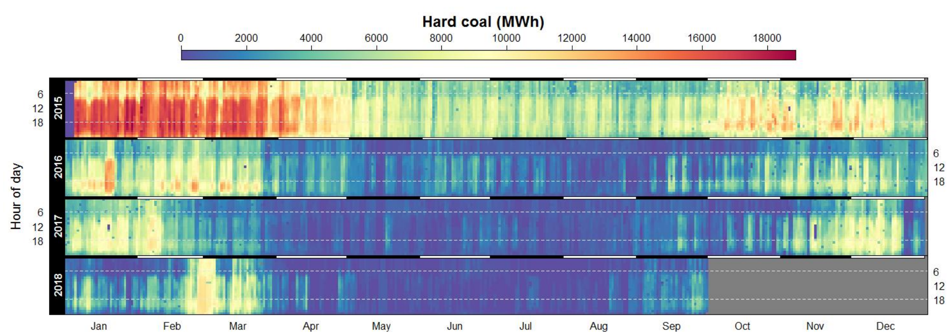
Figure 3.24 shows that the supply sources of the mid-merit load are renewable energy technologies as Solar and Onshore wind.



(a)

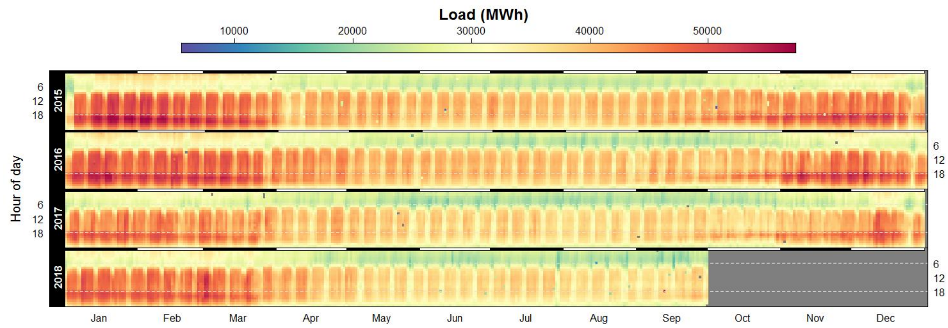


(b)

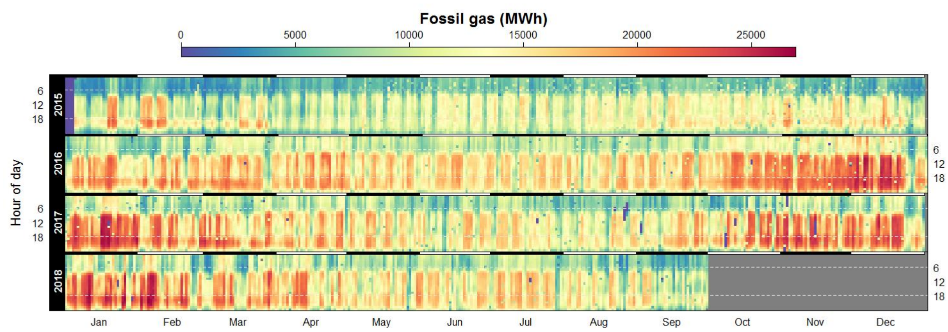


(c)

Figure 3.25: Hourly demand and generation time series for the UK: Load (a) Load and Electricity generation from Offshore wind (b) and from Hard coal (c)



(a)

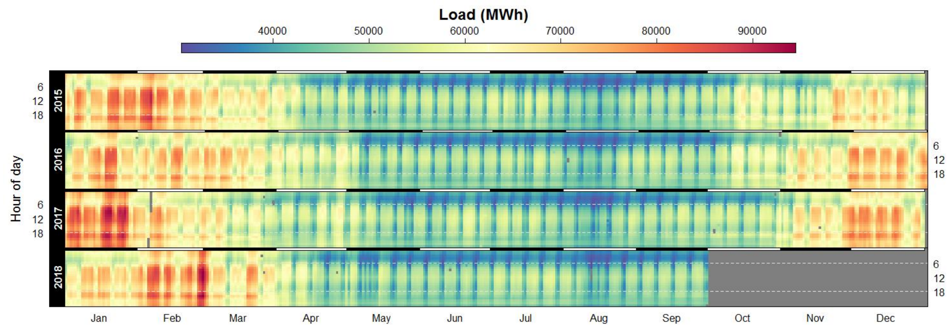


(b)

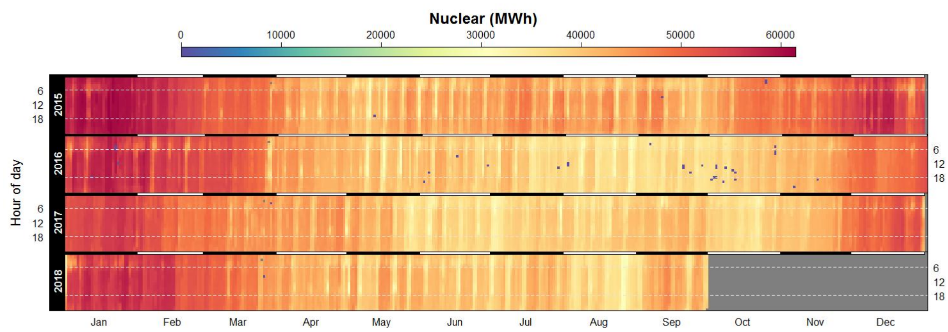
Figure 3.26: Hourly demand and generation time series for the UK: Load (a) and Electricity generation from Fossil gas (b)

Figures 3.25 and 3.26 shows the peak load following technologies as Offshore wind and Fossil gas. Although the UK started to abandon electricity generation from Hard coal from 2015, Hard coal shows still the peak load following characteristics.

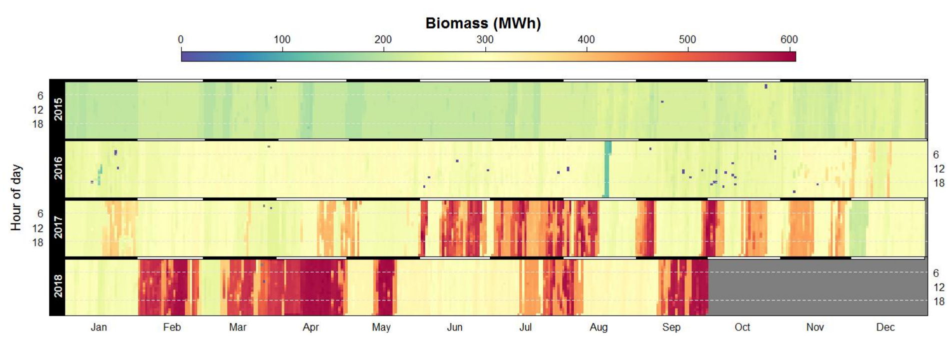
France analysis are given in Figures 3.27, 3.28, 3.29, 3.30 and 3.31.



(a)



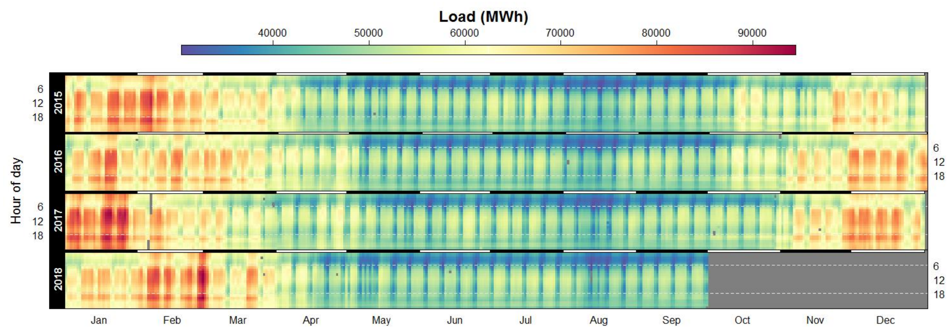
(b)



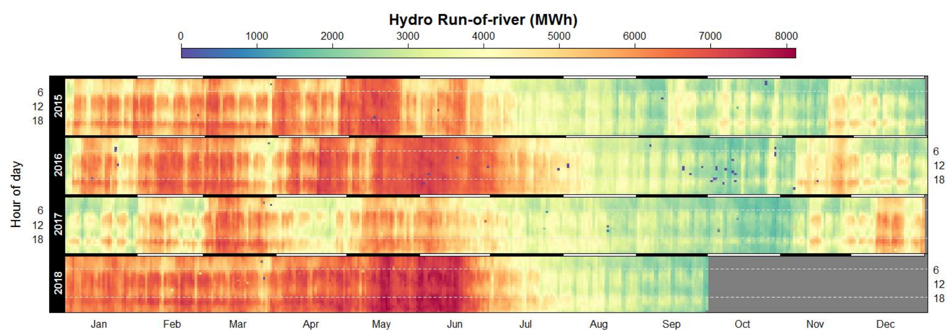
(c)

Figure 3.27: Hourly demand and generation time series for France: Load (a) and Electricity generation from Nuclear (b) and from Biomass (c)

In contrast to majority of other case countries, in France Nuclear energy don't show the characteristics of base load, whereas it behaves like a load following technology. Whereas, the generation from Biomass is very limited, see Figure 3.27. It is seen that load varies a lot between seasons, what is a consistent finding with results of trend analysis.



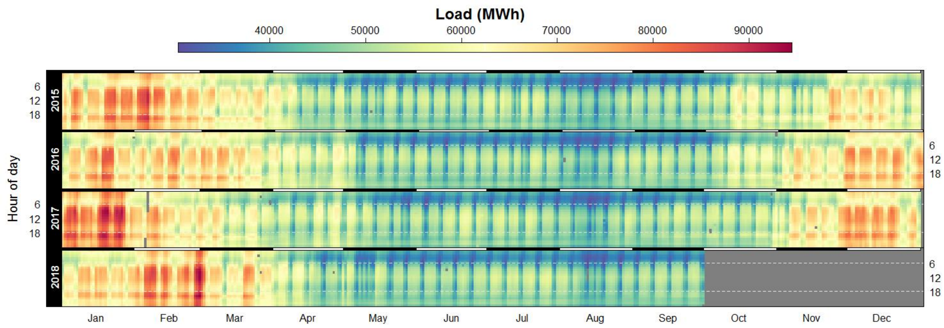
(a)



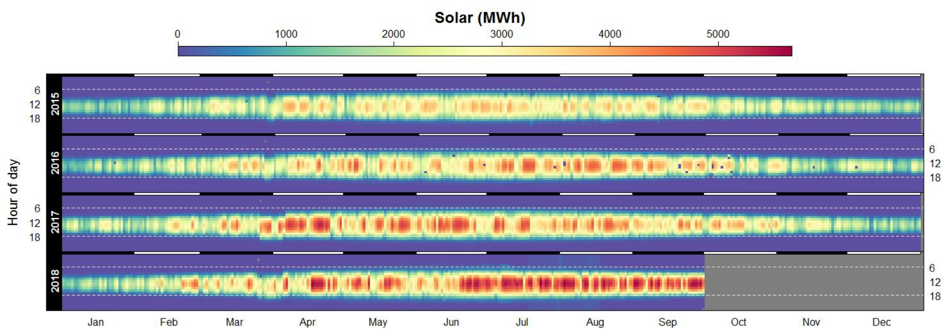
(b)

Figure 3.28: Hourly demand and generation time series for France: Load (a) and from Hydro Run-of-river (b)

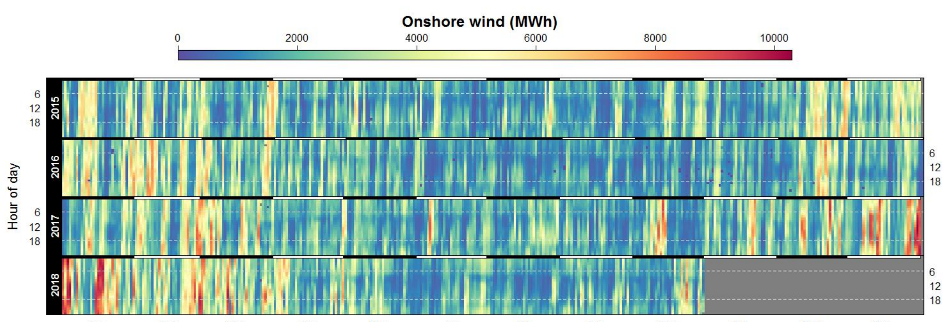
According to Figure 3.28, Hydro Run-of-river can be seen as a load following technology in France.



(a)



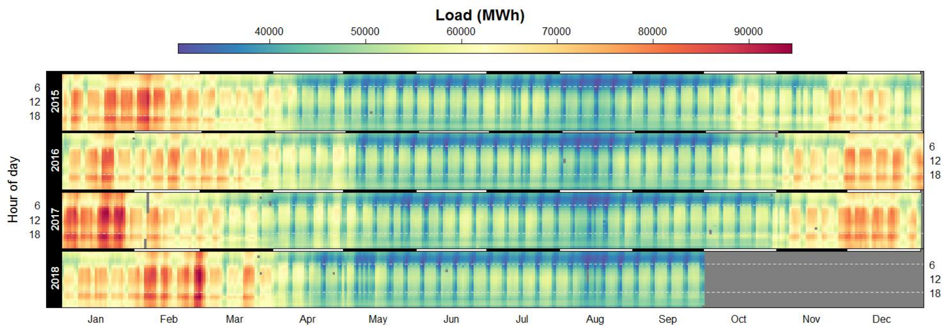
(b)



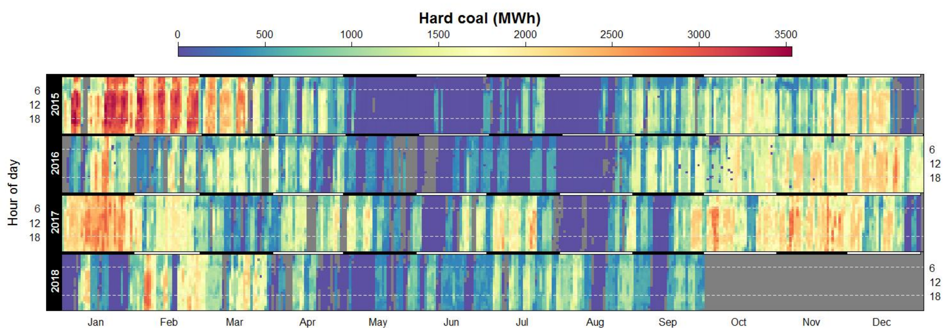
(c)

Figure 3.29: Hourly demand and generation time series for France: Load (a) and Electricity generation from Solar (b) and from Onshore wind (c)

Figure 3.24 shows the supply sources of the mid-merit load are renewable energy technologies as Solar and Onshore wind.



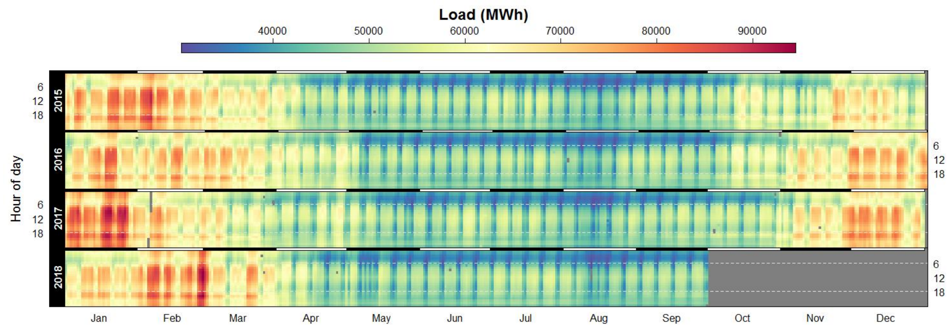
(a)



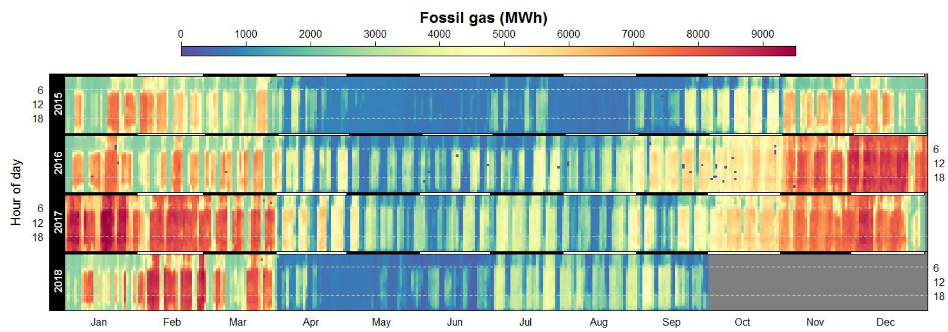
(b)

Figure 3.30: Hourly demand and generation time series for France: Load (a) and from Hard coal (b)

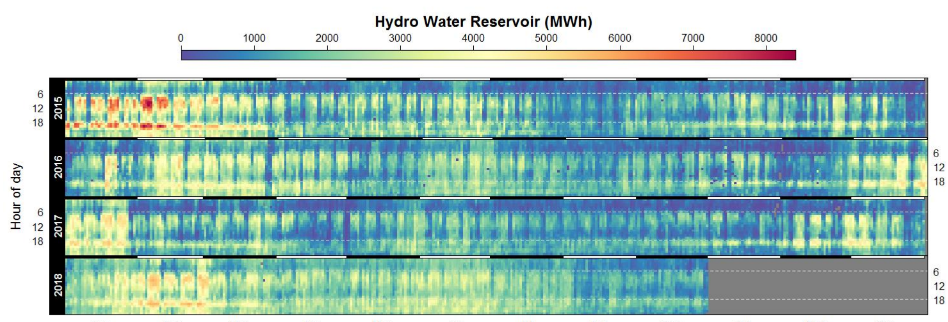
According to Figure 3.30, Hard coal is a peak load following technology in France.



(a)



(b)

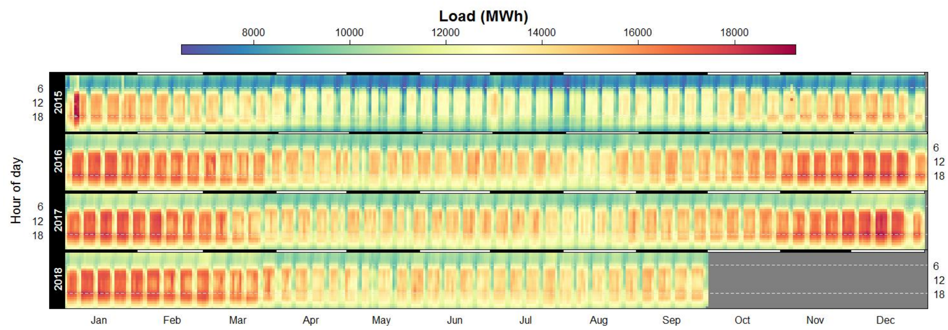


(c)

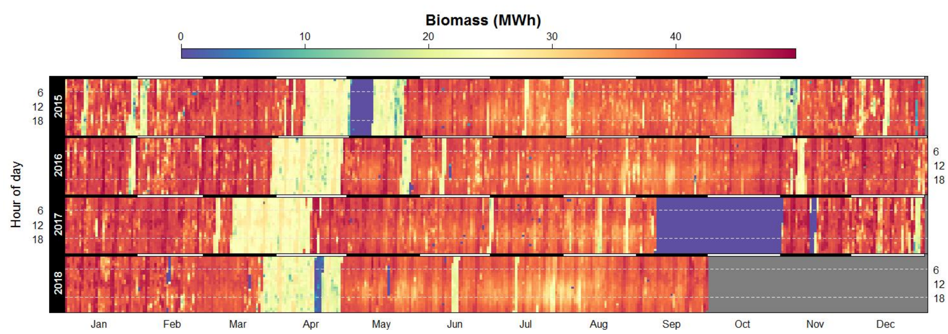
Figure 3.31: Hourly demand and generation time series for France: Load (a) and Electricity generation from Fossil gas (b) and Electricity generation from Hydro Water Reservoir (c)

Lastly, in Figure 3.31, Fossil gas and Hydro Water Reservoir are shown as technologies capable of responding peak load.

As a final case country the Netherlands analysis is shown in Figures 3.32, 3.33, 3.34 and 3.35.



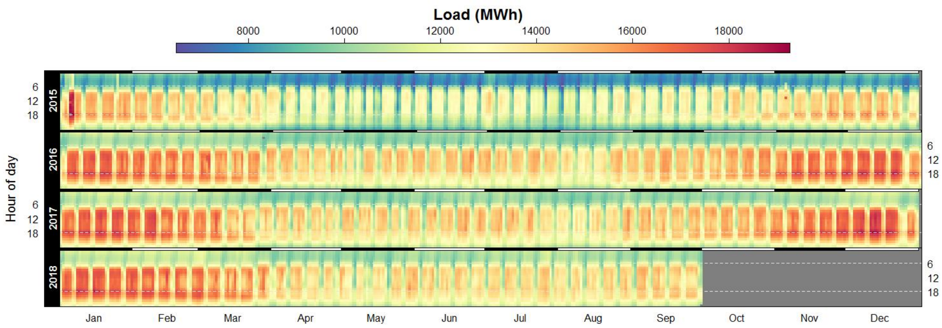
(a)



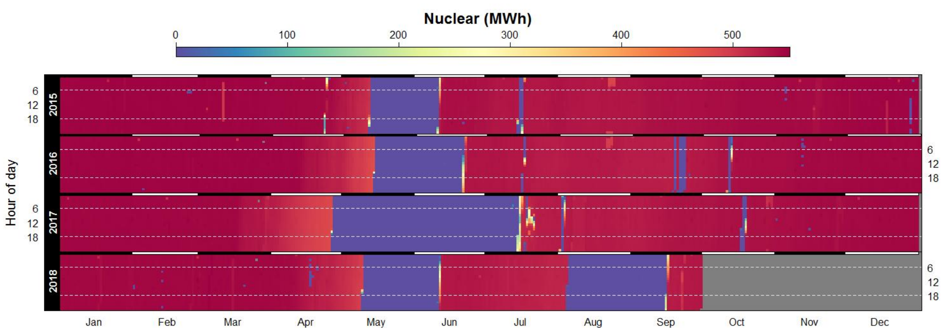
(b)

Figure 3.32: Hourly demand and generation time series for the Netherlands: Load (a) and from Biomass (b)

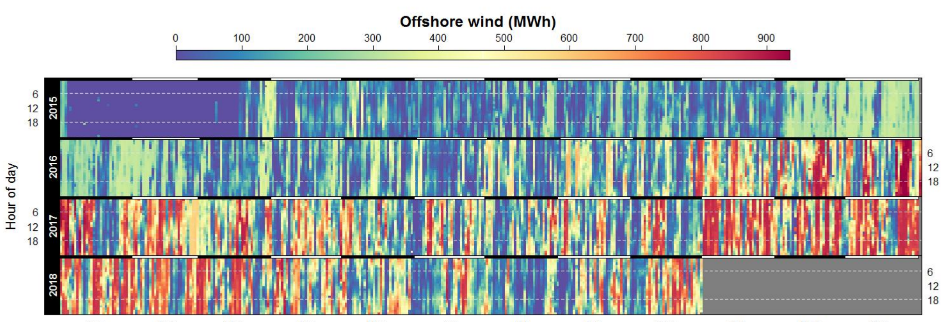
According to Figure 3.32, the electricity demand is higher during winter months and lower between April and September, also the contribution of Biomass is reported as low.



(a)



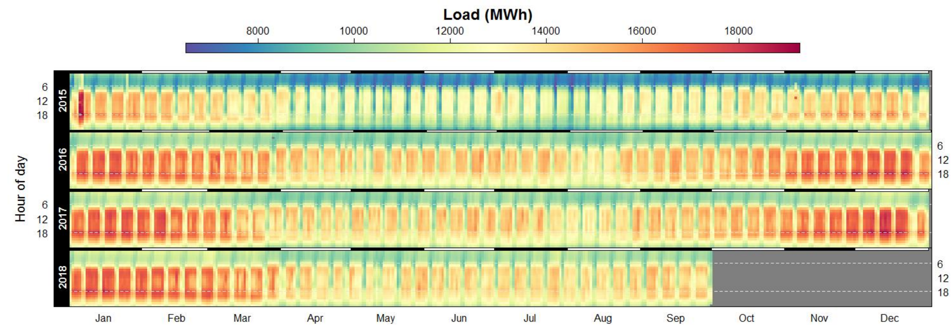
(b)



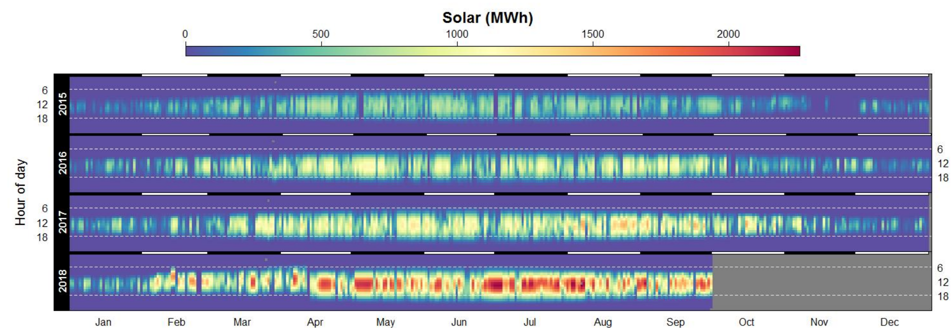
(c)

Figure 3.33: Hourly demand and generation time series for the Netherlands: Load (a) and Electricity generation from Nuclear (b) and Electricity generation from Offshore wind (c)

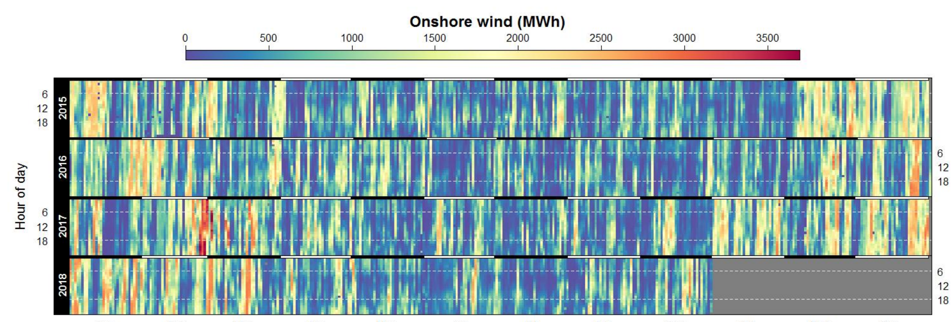
Nuclear energy shows clear base load provider features, see Figure 3.33. As, currently, there is only one nuclear plant in the country, Figure 3.33 shows clearly the periodic stoppages of the plant, so as some possible emergency shut downs. Offshore wind shows the supply source of the mid-merit load technology features.



(a)



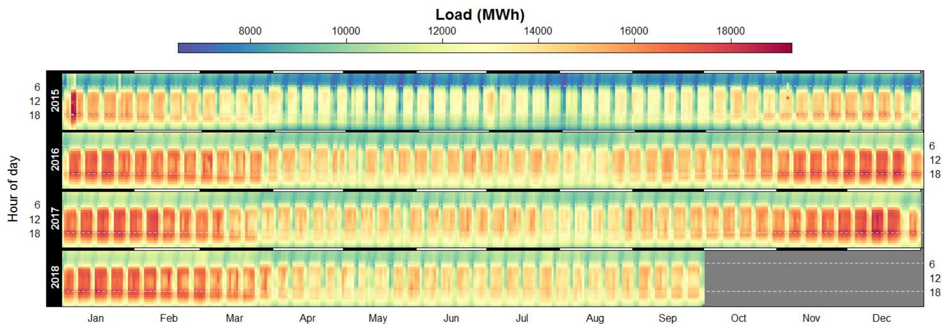
(b)



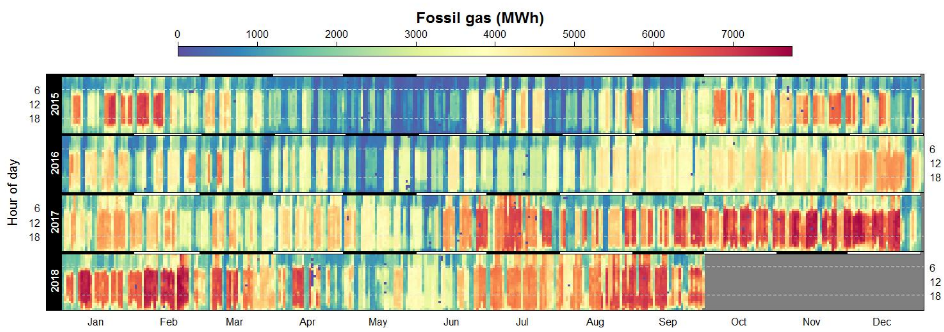
(c)

Figure 3.34: Hourly demand and generation time series for the Netherlands: Load (a) and Electricity generation from Solar (b) and from Onshore wind (c)

Here, once again Onshore wind and Solar behaves like the supply sources of the mid-merit load, see Figure 3.34.



(a)



(b)

Figure 3.35: Hourly demand and generation time series for the Netherlands: Load (a) and Fossil gas (b)

In Figure 3.35, a peak following technology Fossil gas is shown for the Netherlands.

In Table 3.4, the overall summary of the findings obtained in this section is given. It must be noted that especially Solar, Onshore, Offshore wind technologies can be considered whether in demand following or peak following groups depending on month and hours for the generation.

Table 3.4: Summary table for generation and demand time series for each country

Country	Peak following	Base load	Mid-merit load
Spain	Solar	Biomass	Fossil brown coal
	Onshore Wind	Nuclear	Hard coal
	Fossil Gas	Hydro Run.	Solar
	Hydro Reservoir		Onshore Wind
Germany	Onshore Wind	Biomass	Fossil brown coal
	Solar	Nuclear	Offshore Wind
	Hard coal	Hydro Run.	Solar
	Fossil gas		Onshore Wind
	Hydro Res.		
UK	Hydro Run.	Nuclear	Solar
	Offshore Wind		Onshore Wind
	Fossil gas		
	Hard coal		
	Solar		
	Onshore Wind		
France	Hard coal	Biomass	Nuclear
	Fossil gas		Hydro Run.
	Hydro Reservoir		Solar
	Solar		Onshore Wind
	Onshore Wind		
Netherlands	Fossil gas	Biomass	Offshore wind
	Offshore wind	Nuclear	Solar
	Solar		Onshore wind
	Onshore wind		

Table 3.4 provides a classification of the electricity generation technologies. Later the relation between the electricity market price and the electricity generation levels per each individual technology is investigated. These roles are assigned to the generation technologies resulting from the electricity generation strategy of the countries analysed. The association between the electricity market price and the generation technology, as being one of the supply sources in the roles of the base load, the mid-merit load or the peak load is discussed later.

3.1.6 Granger causality analysis

Until this subsection, only general characteristics of the data are presented. However, these general characteristics are not enough to reach a conclusion for the causality between the electricity market price, the electricity demand and the electricity generation by different technologies. Therefore in this section, exploratory causality analysis is performed with statistical tests. There exist several techniques, but the most frequently used test is Granger-causality for such investigations [122–126].

Thus, the Granger-causality statistical test is selected to identify the parameters driving the electricity market price. This test determines whether one time series is convenient in forecasting another for a predetermined significance level. Here, the analyses are performed by using the methodology explained in section 2.3.2. Hourly averaged data obtained from day-ahead markets are used, therefore the Granger test is defined with lag 24.

In this analysis, it is assumed that if a resulting test value is smaller than a given threshold, typically 0.01, there exists a causation between tested variables.

The Granger causality analysis is performed for daily lagged (24 hours) sets in order to evaluate day-ahead market dynamics. The results of the analysis of all case countries are shown in Table 3.5.

Table 3.5: Granger causality analyses for hourly data and daily lagged sets in exporter cases

Variable→Price	Germany	France	Spain	UK	Netherlands
Biomass	0.08	0.889	<0.01	0.986	<0.01
Fossil brown coal	<0.01	-	<0.01	-	-
Fossil gas	<0.01	-	<0.01	<0.01	<0.01
Fossil hard coal	<0.01	<0.01	<0.01	<0.01	-
Hydro run.	<0.01	<0.01	<0.01	<0.01	-
Hydro reservoir	<0.01	<0.01	<0.01	-	-
Nuclear	<0.01	<0.01	<0.01	0.208	0.274
Solar	<0.01	<0.01	<0.01	<0.01	<0.01
Onshore wind	<0.01	<0.01	<0.01	<0.01	<0.01
Offshore wind	<0.01	-	-	<0.01	<0.01
Load	<0.01	<0.01	<0.01	<0.01	<0.01

Considering available generation technologies in Germany and France, in both cases, based on the Granger causality, the Biomass generation has not a statistically significant effect on the electricity market price. It is worthwhile to remember (see section 3.1.5)that Biomass is one of the base load providers for Germany and France.

In the test for the Spanish case, all available generation technologies exert an influence on electricity market price. In the UK, the influences of Biomass and Nuclear on the electricity market price are found to be statistically insignificant. Again it is recommended to re-visit section 3.1.5, Nuclear energy is a based load provider for the UK and in the case of Biomass, there is very low data availability. Lastly in the Netherlands case, based on the Granger causality test the generation from the Nuclear technology does not exert an influence on the electricity market price. Here also Nuclear is a base load provider for the Netherlands.

The common characteristic of the technologies that do not exert influence on the electricity market price is their behaviour as base load providers.

3.1.7 Overall correlations

Until here, general figures and causality analyses are presented. Causality analysis generates statistical test values and compares these values to predetermined significance levels. Therefore, they can provide a binary finding for 'one parameter causes another' statements. From now on, a numeric summary for the relation between variables will be presented for different investigation periods. This summary is useful to see whether one variable decreases when another one decreases and with what degree these two variables move together.

By performing correlation analysis on the available data, the pairwise price coefficients were obtained for the full analysis period. The pairwise price coefficient stands for the correlation between a tested variable (as an example Onshore wind generation, demand, Solar generation, etc.) and the electricity market price.

Table 3.6: Electricity market price correlation coefficients Jan 2015-Dec 2018, NA indicates that the technology does not exist, - indicates that the variable is not considered according to Granger casuality analysis

Variable	Spain	France	Germany	Netherlands	UK
Biomass	0.09	-	-	-0.05	-
Fossil brown coal lignite	0.44	NA	0.34	NA	NA
Fossil gas	0.54	0.42	0.38	0.4	0.33
Fossil hard coal	0.5	0.42	0.47	NA	0.06
Hydro run-of-river & pondage	-0.17	-0.08	-0.06	NA	0.17
Hydro water reservoir	0.07	0.3	0.18	NA	NA
Nuclear	-0.07	0.18	0.01	-	-
Solar	0.08	0.05	0.09	0.17	0.13
Offshore wind power	NA	NA	-0.12	-0.05	-0.06
Onshore wind power	-0.29	-0.05	-0.25	-0.11	-0.01
Load	0.33	0.38	0.36	0.36	0.29

Among the renewable energy sources in Table 3.6, only Onshore wind power is found to negatively correlate with electricity prices almost without any exception what means that when Onshore wind power generation is high, the electricity market prices are low. In most cases, the major driving parameters for the electricity prices are Fossil gas and Load. However, in Spain and Germany Fossil hard coal is noted as the parameter with the highest correlation coefficient, indicating that electricity market prices are highly correlated with conventional energy generation for the case countries. Fossil Gas has a very high positive correlation coefficient for the electricity market price and it should be highlighted that Fossil Gas is a peak following technology for the case countries. It can be interpreted that high electricity generation from Fossil Gas results in high electricity market price. If we consider the country specific energy mixture policies, in Spain and Germany onshore wind contribute to energy mixture as peak load or mid-merit load generators, therefore Onshore wind can be noted as a good balancer for electricity market prices at peak demand hours, see section 3.1.5.

3.1.8 Monthly correlations

In order to see the effect of the seasonality of the load and the seasonality of the renewable energy resources, the monthly correlations are given in Figures 3.36, 3.37, 3.38 and 3.39.

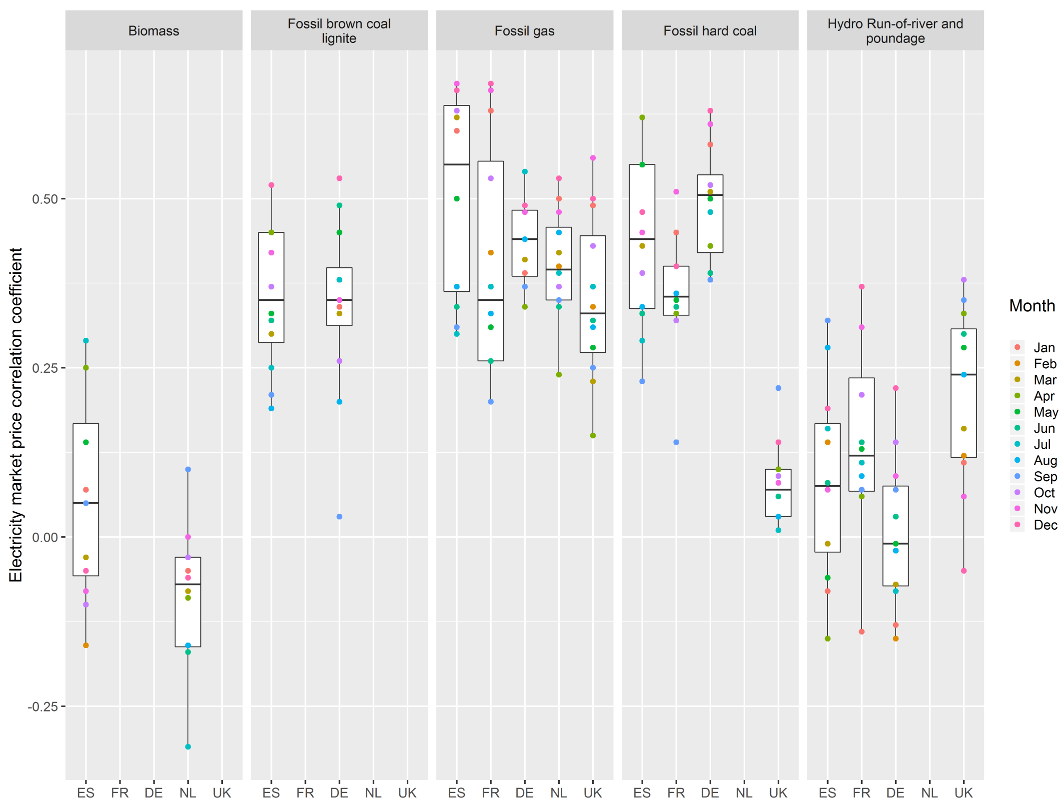


Figure 3.36: Correlation analysis per countries part 1

According to Figure 3.36, during warmer months electricity generation from the fossil fuel sources has lower correlation coefficients for electricity market price in all the case countries. Whereas, during winter months the correlation coefficient of the generation from the fossil fuel sources approximately doubled. For Hydro run-of-river and pondage, in the UK spring and summer months, in France winter months, in Germany autumn months, in Spain summer months have high correlations. In Biomass, Spain has the lowest correlation coefficient in February, whereas the Netherlands has the lowest correlation coefficient in July.

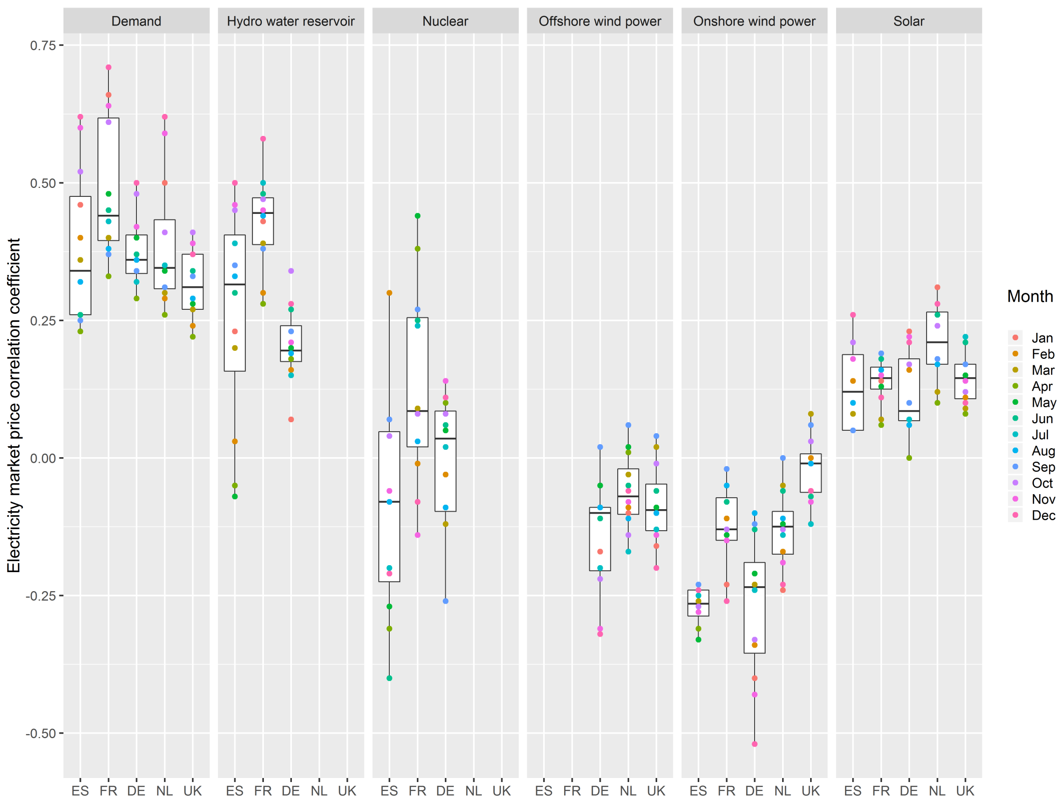


Figure 3.37: Correlation analysis per countries part 2

Continuing from Figure 3.37, in Spain, France, Germany and the Netherlands the demand has the higher correlation coefficients in December, January and February. In the case of the UK, October has the highest correlation coefficient between the demand and the electricity market price. For Nuclear, France has the highest coefficient in May, Spain in February and Germany in November. The wind energy for both onshore and offshore applications, has high correlation coefficients during summer and spring months in all case countries. Finally, for Solar, in the majority of the cases (except Spain) April has the lowest correlation coefficients. In Spain and Germany winter months have high correlation coefficients and in France and the UK summer months have high correlation coefficients. Hydro water reservoir exist in three of the case countries. In Spain, it has a larger range of the correlation coefficients (-0.1 to 0.5), whereas in France there is no such a large range observed. However in these two countries the extremes of the correlation coefficients are observed in similar months as maxima in January and minima in February and March. In the case of Germany, October has the maximum coefficient and February has the minimum coefficient.

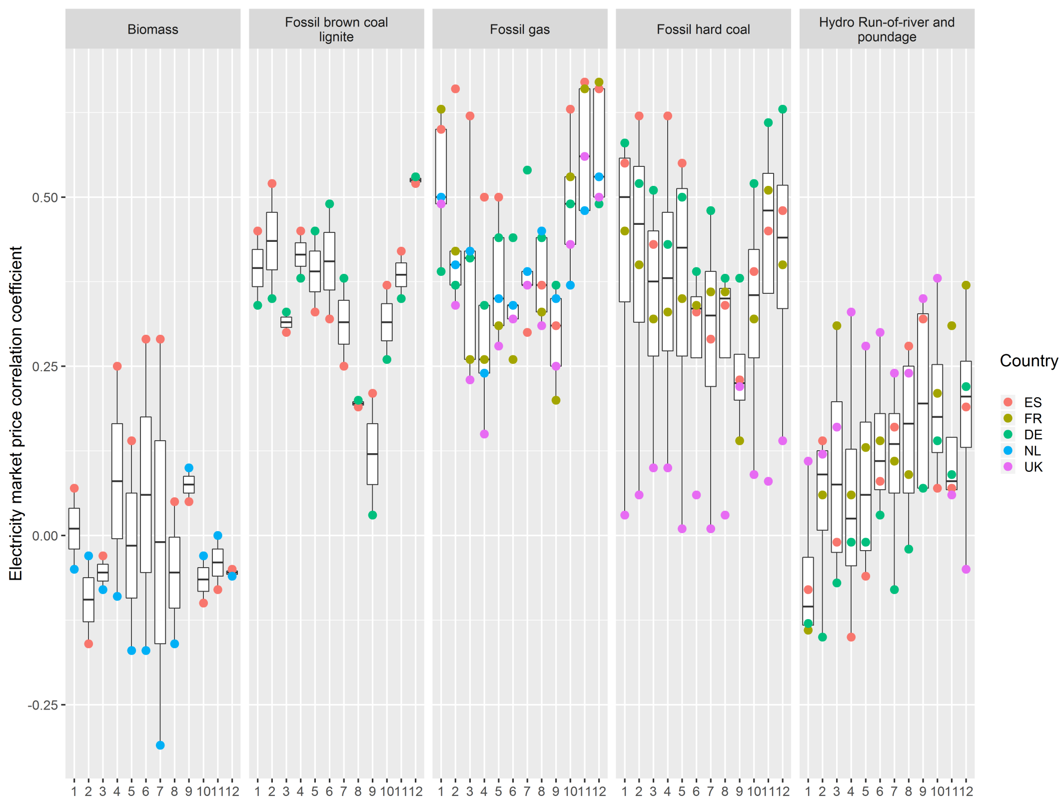


Figure 3.38: Correlation analysis per months part 1

When we look into the same data from a different perspective by grouping data per country, see Figures 3.38 and 3.39. In Biomass, Spain and the Netherlands show the two opposites in July. For the fossil brown coal lignite, September shows the lowest coefficients for Spain and Germany. In June and July the spread of the correlation coefficients of Fossil gas is too low. For

Fossil hard coal the UK shows the minimum correlation coefficients, Germany and Spain competes for the maxima. In Hydro run-of-river and pondage, the majority of the maximum correlation coefficients among months indicate the UK and the minimum correlation coefficients show Germany.

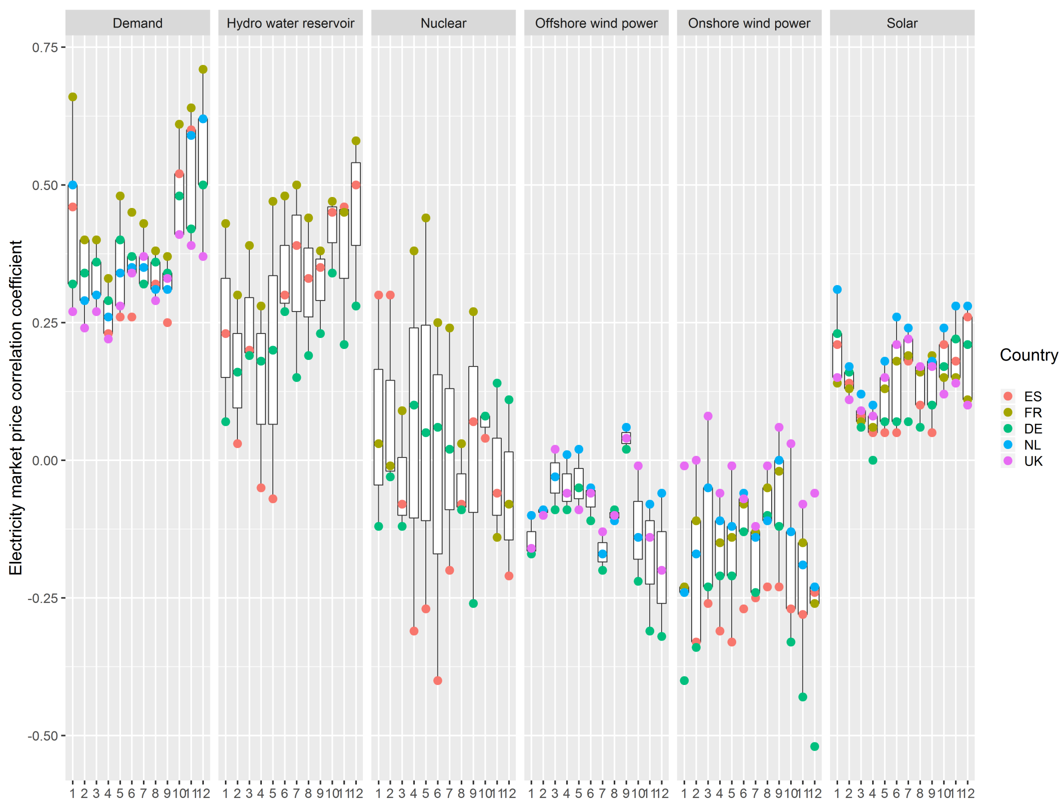


Figure 3.39: Correlation analysis per months part 2

In Figure 3.39, the correlation coefficients of the load are the highest in France and the lowest in the UK. In the case of France, Hydro water reservoir, Nuclear and Demand show the highest correlation coefficients among other countries. Offshore wind power has very narrow spread bars almost in all months excluding November and December. Whereas in Onshore wind power, the wider boxes are observed being extremes in the UK and Germany. Solar energy shows also narrow spread bars, the most of the time the highest electricity market price coefficient is reported for the Netherlands.

Re-visiting the most significant bottom price months for a maintenance engineer, see section 3.1.4:

- Spain: April (2015, 2017, 2018) and November (2015 and 2018)
- Germany: April (2015, 2016) and May (2017, 2018)
- France: May (2015, 2018) and April (2016, 2017)
- The Netherlands: June (2015, 2017, 2018)
- The UK: June (2015, 2016, 2017)

Now if we combine these with the findings obtained in this section, a maintenance engineer can track the listed electricity generation sources in order to estimate a realistic electricity market price and corresponding revenue losses by selecting variables who have higher electricity market price coefficients from | 0.25 |:

- Spain: April (Fossil brown coal lignite, Fossil gas, Fossil hard coal, Nuclear, Onshore wind power) and November (Fossil brown coal lignite, Fossil gas, Fossil hard coal, Demand, Hydro water reservoir and Onshore wind)
- Germany: April (Fossil brown coal lignite, Fossil gas, Fossil hard coal, Demand) and May (Fossil brown coal lignite, Fossil gas, Fossil hard coal, Demand)
- France: April (Fossil brown coal lignite, Fossil gas, Fossil hard coal, Demand, Nuclear) and May (Fossil brown coal lignite, Fossil gas, Fossil hard coal, Demand, Nuclear)
- The Netherlands: June (Biomass, Fossil gas, Demand, Solar)
- The UK: June (Fossil gas, Demand)

3.1.9 Importance ranking

Up until here, trend, causality and correlation analyses were presented for the electricity market data. These analyses mostly cover the pairwise tests for the time series data, which means pairwise comparisons are investigated for each variable respect to the electricity market price. In order to provide a complete picture, the interrelations between the contributor variables need to be considered. Therefore, in this section, the relative importance of each variable will be studied. In order to do this, hourly data are scaled. Then the principal component regression (PCR) analysis and the boot strapping intervals for regression models are used. Importance rankings of the electricity generation technologies and the demand on the electricity market price per country are given in Table 3.7.

Table 3.7: Variable importance ranking using PCR 2015 Jan- Oct 2018

Technology	Spain	France	Germany	Netherlands	UK
Biomass	8	-	-	5	-
Fossil brown coal lignite	10	-	4	-	-
Fossil gas	7	6	4	2	2
Fossil hard coal	4	6	5	-	3
Hydro run-of-river & pondage	9	4	3	-	3
Hydro water reservoir	3	3	6	-	-
Nuclear	6	1	4	-	-
Solar	5	7	3	4	3
Offshore wind power	-	-	7	4	4
Onshore wind power	2	5	2	3	5
Load	1	2	1	1	1

From the ranking presented for Spain in Table 3.7, it can be said that Load has the highest importance level on the electricity market price. Onshore wind power is the second most important. Germany's electricity generation mixture contains very high amount of renewables, and the findings of the Sections 3.1.2 and 3.1.5 are consistent with this characteristic of Germany. Here, peak load following and mid-merit load technologies sit on top in the importance rankings. As it is shown in Table 3.7, the second most important is the onshore wind generation, while Solar is third in Germany. Both the UK and the Netherlands showed that the important parameters for electricity market price are Load and Fossil gas. Lastly, the findings for the French case show that the three major parameters according to PCR model are Nuclear, Load and Hydro water reservoir. The electricity generation from nuclear energy is not highly ranked in many countries. France, however, does generate a very high amount of its electricity from Nuclear, generation from Nuclear equals to almost 74 % of its total demand. In a correlation coefficient analysis Nuclear was a significant energy source, but not the one with highest coefficient. In its relative importance rank, the real effect is revealed. It places first in the PCR model.

In order to provide a more detailed numerical summary and test the importance of the investigated variables, four importance metrics are considered (First, Last, Betasq and Pratt explained in Chapter 2). The underlying model for these metrics is multivariate linear regression.

Table 3.8: Rankings for the UK

UK	First	Last	Betasq	Pratt
Fossil gas	0.49	0.30	0.35	0.49
Fossil hard coal	0.03	0.12	0.09	-0.07
Solar	0.03	0.15	0.08	0.00
Offshore wind power	0.00	0.00	0.00	0.00
Onshore wind power	0.04	0.07	0.05	0.05
Demand	0.35	0.19	0.35	0.40

Table 3.9: Rankings for the Netherlands

Netherlands	First	Last	Betasq	Pratt
Biomass	0.00	0.00	0.00	0.00
Fossil gas	0.49	0.48	0.48	0.49
Solar	0.03	0.00	0.00	0.00
Offshore wind power	0.00	0.00	0.00	0.00
Onshore wind power	0.04	0.07	0.07	0.05
Demand	0.45	0.43	0.45	0.45

In Tables 3.8 and 3.9, Fossil gas and Demand have the higher relative importances for the electricity market price in the case of the UK and the Netherlands by consensus of all the models.

Table 3.10: Rankings for France

France	First	Last	Betasq	Pratt
Fossil gas	0.28	0.01	0.01	0.11
Fossil hard coal	0.25	0.02	0.01	0.09
Hydro run-of-river and pondage	0.01	0.27	0.12	0.06
Hydro water reservoir	0.16	0.2	0.12	0.25
Nuclear	0.04	0.24	0.2	-0.17
Solar	0.00	0.01	0.00	0.00
Onshore wind power	0.00	0.05	0.02	0.01
Demand	0.25	0.2	0.51	0.65

Whereas, in Table 3.10 the models are not in agreement especially for Fossil gas, Hydro run-of-river & pondage and Nuclear energy. The highest demand importance is noted in the metric pratt. Nevertheless, it is still possible to reach a summary as Demand, electricity generation from both hydro technologies, Fossil gas and Fossil hard coal and Nuclear are important variables for the electricity market price for France.

Table 3.11: Rankings for Germany

Germany	First	Last	Betasq	Pratt
Fossil brown coal lignite	0.19	0.04	0.02	0.12
Fossil gas	0.19	0.04	0.2	0.12
Fossil hard coal	0.25	0.03	0.04	-0.2
Hydro run-of-river and pondage	0.00	0.14	0.08	0.04
Hydro water reservoir	0.07	0.02	0.02	0.06
Nuclear	0.00	0.06	0.02	-0.01
Solar	0.00	0.14	0.09	0.01
Offshore wind power	0.03	0.00	0.00	0.00
Onshore wind power	0.12	0.31	0.32	0.38
Demand	0.16	0.24	0.39	0.48

Table 3.12: Rankings for Spain

Spain	First	Last	Betasq	Pratt
Biomass	0.01	0.01	0.00	-0.01
Fossil brown coal lignite	0.20	0.00	0.00	0.00
Fossil gas	0.22	0.00	0.00	0.05
Fossil hard coal	0.27	0.01	0.01	0.29
Hydro run-of-river and pondage	0.05	0.00	0.00	0.00
Hydro water reservoir	0.00	0.1	0.09	0.00
Nuclear	0.00	0.03	0.00	0.00
Solar	0.00	0.14	0.07	-0.03
Onshore wind power	0.11	0.30	0.21	0.26
Demand	0.13	0.33	0.51	0.43

In Tables 3.11 and 3.12, Fossil hard coal, Fossil gas, Onshore wind power and Demand have the higher relative importances for the electricity market price for Germany and Spain by consensus of all the metrics.

3.1.10 Summary

Here, the summary of the findings for the main research questions investigated in this section is given. Q stands for question and R is response.

- Q: For electricity market price estimations, which variables can be tracked as information providers?
- R: Demand can be considered, especially for the estimation of the peak prices, but the demand times series do not provide a good reference for the estimation of bottom prices, what a maintenance engineer would like to know in order to schedule maintenance activities. Especially, the electricity generation time series in France from Nuclear and Fossil gas, in Spain and Germany from Onshore wind and Fossil hard coal, in the Netherlands and the UK from Fossil gas can be tracked as information providers.
- Q: Which one is more important for the peak and the bottom electricity market prices in a country
  - Electricity generation and consumption figures of a country

  - Role of a country in transnational electricity trade

- R: Spain, the UK and the Netherlands are the net importer countries, whereas France and Germany are the net exporter countries. The total electricity demand for the investigated period in Germany is approximately 2000 TWh and France is approximately 1800 TWh. Whereas in the net importer case countries, the demand is almost half of the net exporter countries. General electricity market statistics show that the UK (net importer) and France (net exporter) show very similar features considering extreme occurrences as maximum and minimum values. Also, the market characteristics resembles each other in the Netherlands (net importer) and Germany(net exporter) cases. However, it is not possible to see such a similarity between electricity markets of Spain and France, which are neighbouring importer-exporter countries. The analysis must be extended to answer this question considering all trading countries.

- Q: What are the influences of the generator technologies on electricity market price regarding their ability to respond timely to demand?

- R: Energy mixture policy of a country does not stand only for the varying amounts of the electricity generation from different sources. It also covers also the electricity generation timing from different energy sources. Base load provider technologies such as Nuclear (the Netherlands, the UK, Germany, Spain) exert very weak or zero influence on the electricity market prices, but when Nuclear is considered as mid-merit load technology in the energy mixture, its influence on the electricity market price increase severely. Peak load following generators such as Fossil gas exert very high influence on electricity market prices, even in some cases its influence is stronger than the electricity demand (Spain, France, Germany). When Onshore wind is considered as peak following technology in the energy mixture (Spain and Germany), its influence increases in comparison to being in the role of the supply sources of the mid-merit load consideration (the UK, the Netherlands, France).

- Q: Which months are better to perform a scheduled component replacement or a maintenance visit considering electricity market price?

- R: It can be noted that for the first and the last months of the year the extremes of price follows the extremes of the load and the load is significantly higher in winter months. During April, May, June and November the case countries show the bottom prices, at these months a scheduled component replacement or a maintenance visit can be considered.

- How do the drivers of the electricity market price vary among different EU countries?

- R: The drivers of the electricity market price vary among the case countries depending on the countries energy mixture policies regarding load response timing, the contributor technologies of the energy generation and the level & the role of participation to the transnational electricity trade.

Additionally, monthly demand and correlation analyses provide more information in comparison to the overall summary analyses. Therefore, it is suggested to consider seasonality of electricity market prices, when estimating or evaluating electricity market price of a country. It should be noted that each case country has different influential features affecting the electricity market price and instead of using a generic model, it is recommended to generate case specific algorithms.

3.2 Failures, weather and health status relationships

It is worth to highlight that the revenue of wind farms depends on concurrent time series of two major factors. The first major factor is the electricity market price, which along with its dependencies, was discussed and analysed in the previous section. The second major factor is the produced power, which depends on intermittent wind source and reliability of wind turbines. While electricity market prices follow seasonal, monthly patterns which are highly correlated to the national electricity demand, produced power depends more on the local weather characteristics and failure occurrences of the turbines. Yet the usability of these features in decision making requires a thorough understanding of the dynamic structure of the relevant parameters. As anticipated, a wind farm owner is interested in operating a wind farm when simultaneous maxima occurred in the time series of the electricity market price and the produced power. In order to operate a wind farm at concurrent maxima of the electricity market price and the produced power, operation and maintenance engineer evaluates frequently health status of wind turbines to ensure that wind turbines are ready to generate electricity. Therefore, wind turbine health status and failure pattern related information is particularly valuable for operation and maintenance engineer and this information can be obtained in several ways as:

- Maintenance visits, visual inspections
- Physical tests such as vibration and oil analysis for gearbox
- SCADA alarms, SCADA signals

During wind farm maintenance visits, wind turbines do not produce electricity because they stop. Moreover, accessibility of a wind turbine due to harsh environmental conditions is limited to schedule frequent maintenance visits when the bottom electricity market prices expected. Thus, avoiding unnecessary visits to wind farm and performing maintenance and inspection activities in short time periods must be taken into account by a maintenance engineer during wind turbine health status information gathering. Physical tests are costly procedures, besides most of the time they require maintenance visits to intervene wind turbines for a sample collection or a test. SCADA alarms consider only the control limits (preselected thresholds) defined by the engineers or the data scientists. Only in case of the exceedance of these pre-determined limits, the wind turbine SCADA generates an alarm. Therefore, a SCADA alarm is not a strong information provider for preventative and predictive maintenance actions, since it does not provide information in advance to failure, or an anomaly in any signal considering the past records. Besides, some of SCADA alarms are false alarms that provide no value on the health status of the wind turbines. Under the costs and the limitations of information gathering via maintenance visits, physical tests and SCADA alarms, learning from the past failures occurred in the wind farm, evaluation of the SCADA signals and discovering the failure patterns related to meteorological conditions become a serious research need.

In this section, resulting downtimes from failure occurrences are investigated considering atmospheric conditions and occurrence timing of the failures. Here, when the power losses or production losses terms are used, they are referred to the following groups.

Production losses due to maintenance:

These losses stem from wind turbines that are shut down for required maintenance. Therefore, they can be associated with the repair time [143] and their cumulative effects can be measured via technical maintenance indicators [172]. Several studies in the literature report the respective downtime occurrence per wind turbine component intervention and the distribution of the repair times [173–175]. In addition to unavoidable service durations, production losses caused by low accessibility to the wind turbine (resulting from weather and coarse planning) can also be included in this group. These losses will be discussed in detail in Chapter 5. Finally, there are power production losses caused by efficiency loss, which are the result of deviations from the reference power curve. These losses will be discussed in Chapter 4 with a real case example.

Production losses due to turbine/ component failures:

The second major group of causes for wind turbine downtime and production losses are the component, system, and turbine failures. Here, not to lose focus on minor, which are easy (in a fast and cheap way) to fix failures, the investigation must be centred on major components failures considering their failure frequency and resulting downtime for the wind turbine. On the one hand downtime durations of frequently failing components can be very short (from minutes to couple of hours) and on the other hand downtime durations of very rarely failing components can be very long (from days to couple of months). For instance, although gearbox failure rates were far lower than electricity system failure rates, their downtime per failure records were reported as far more higher durations [176]. A recent study reported generator, gearbox, and blade components as the highest contributors to overall downtime [177].

In this section, by focusing on major wind turbine component failures, failure patterns related to atmospheric variables are studied, in order to provide additional information to a maintenance team.

This section is structured in the order that follows. In the first place, available data and variables are summarised. Then, using both supervised and unsupervised learning tools, data processing approach is given. The findings of the data processing step are used as input for the rule mining process. Finally the interpretation of association rules are presented.

3.2.1 Data

To perform this study, various types of data were collected, and combined. In this section, resolution, unit and type of the data used will be summarised. It is worth highlighting that not only data at failure event but also data prior to failure occurrence are used. The used data were as follows:

- *Wind Speed (WS):* SCADA 10-minutes mean wind speed values starting from 80 minutes prior to failure (WS_{80}, WS_{70},..., WS_{10}), as well as one wind speed measurement at the time of the failure occurrence (WS_{atF}).

- *Relative Humidity (RH):* The analysis includes hourly values of 10 hours prior to failure, and are indicated by RH_{atF}, RH_1, RH_2, etc.

- *Ambient Temperature (T):* This is comprised of the monthly mean T_m, maximum T_{max} and minimum temperatures T_{min} for the 30 days prior to failure, as well as the temperature at the exact time of failure T_{atF}.

- *Power Production (P):* The SCADA data also provided the power production at failure P_{atF} (last 10 minute mean value before failure occurrence) and the power production before failure P_{bF} (the previous 10 minute measurement). Since the manufacturer's power curves are not site and season specific, these values will be presented relative to the measured monthly mean power curve. This allows the turbine's relative performance to be displayed, by taking a measured monthly mean representation of power versus wind speed. The observed SCADA value P_o is divided by the power obtained from the mean power curve at the same wind speed P_m. This measures the efficiency of the turbines' power production: $P_e = P_o \ / \ P_m$. While most values fall between 0 and 1,values slightly below 0 or above 1 are possible, since the reference power curve represents only a mean value. Hence, values of $P_o/P_m \leq 0$ are assigned to 0 and $P_o/P_m \geq 1$ to 1.

- *Downtime / Severity:* The downtime per failure, which indicates the severity of each failure in terms of WT (un-) availability, is detailed in the historical failure data and cross-checked with the SCADA data.
- *Maintenance Strategy:* This indicates the availability of the maintenance personnel, which directly affects the repair time. The expert judgement of O&M strategies was used to define this term, as well as the typical working hours of nearby personnel. In this strategy, the day shift is understood from 08:00 to 18:00, with the remaining hours assigned to the night shift.

As the humidity and the temperature were not available from the wind farm's data, re-analysis datasets from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP/ NCAR) were used as the humidity and temperature inputs for the given wind farm locations and failure occurrence times. The NCEP/NCAR data set uses observations and numerical weather predictions to continuously update weather databases, and is available for free through the United State's National Oceanic and Atmospheric Administration Earth System Research Laboratory [178, 179].

Historical Failure Data: The failure database and its turbine characteristics are shown in Table 3.13. The considered turbines are modern three bladed and pitch regulated, with rated power between 660 kW and 2000kW. All these turbines belong to the same manufacturer, and are equipped with a doubly fed induction generator (DFIG). This study examined 146 of the failure events, with about 30 failures per each of the five components. Failures are defined as any stop to the wind turbine caused by component errors which requires for repair or replacement. The failure data includes the exact time and length of the failures (downtime) for each turbine, but inspections and cleaning events are excluded.

Table 3.13: Input Wind Turbine Database

Details	Case study
Number of WTs	448
Total WT Years	972
WT Technology	47 WTs with 0.66 MW, 289 WTs with 0.85 MW, 112 WTs with 2 MW
Components	frequency converter, yaw and pitch systems, generator, gearbox

3.2.2 Analysis structure

A framework, capable of handling big environmental and historical failure data, has been developed in order to quantify the impact of meteorological conditions on WT component failures. As visualised in Figure 3.40, it is divided into four stages: (1) data pre-processing, (2) data processing, (3) unsupervised rule mining, and (4) ranking and interpretation of the rules.

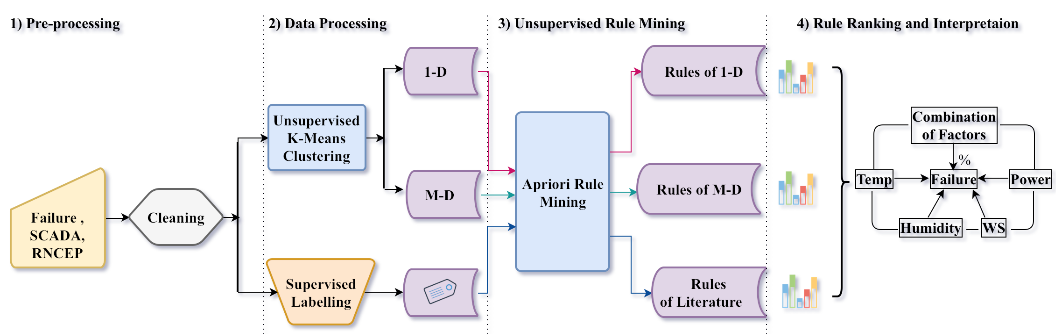


Figure 3.40: Proposed analysis methodology

The initial step represents the raw data pre-processing, consisting on raw data acquisition, time step matching, controls for the time shifts and the merging of the different sets.

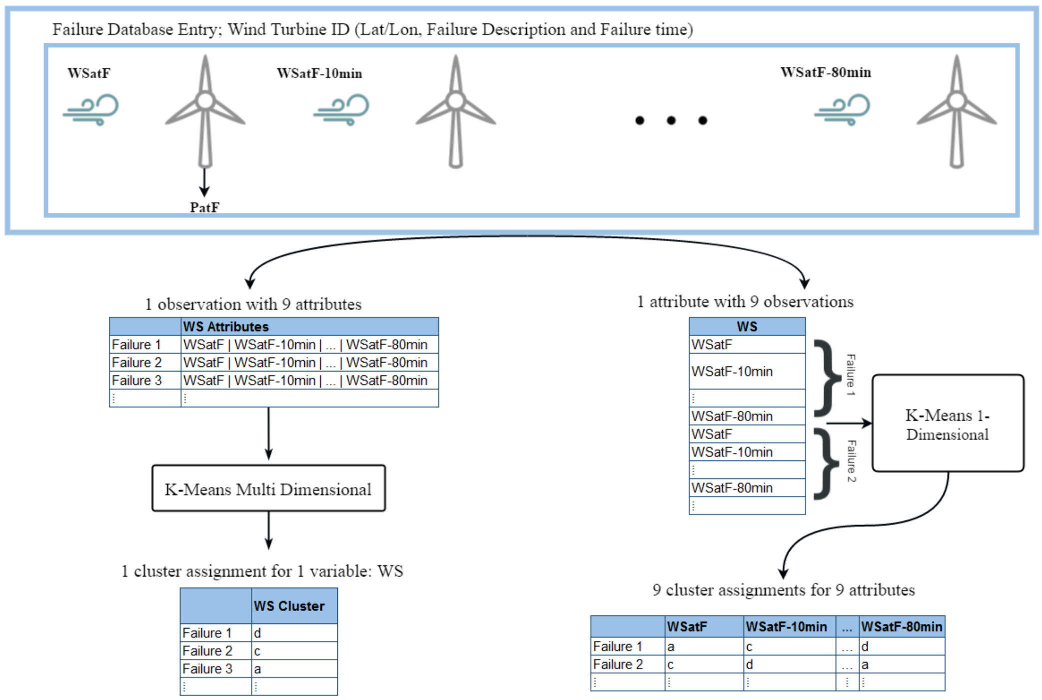


Figure 3.41: Multi-D and 1-D Data Pre-Processing example for wind speed

In the second step, two distinct data mining techniques are used: (a) unsupervised k-means clustering and (b) supervised data labelling. Each technique is carried out independently with the same data set and is expected to influence the results differently. Unsupervised k-means clustering also has two sub-methods, 1-Dimensional and Multi-Dimensional processes. Examples for the treatment of wind speed variables and attributes are given in Figure 3.41 starting from the pre-processing step. Meanwhile, supervised labelling is used to define thresholds and assign labels to the input parameters. Conducted manually, this process uses professional judgements and findings from the literature, and requires profound knowledge of common parameter classification. Supervised labelling was used in this work for the wind speed labelling, using common literature findings and the typical cut-in and cut-out wind turbine wind speeds. Wind speed range was divided into calm, low, high and stormy wind conditions.

Another example of supervised labelling is the relative humidity (RH), which can be labelled in terms of its resulting corrosiveness as indicated in Jiang et al. [182], Leygraf et al. [183] and Xiang et al. [184]. The downside is that the corrosiveness is only applicable for one specific kind of steel, and therefore difficult to apply to WT systems, since they typically consist of many different types of materials. Quantification of the RH in terms of corrosiveness for the entire system would be very complex. Nevertheless it is considered a very good indicator and it is used in the case study. Five labels, from dry air to highly corrosive and precipitation-laden air, were used to manually define the thresholds, one for each hourly value.

The output of either one of supervised or unsupervised clustering serves as the input to the third step. Here, an association rule mining algorithm - called the AARM (see Chapter 2 for the details) - is applied. The algorithm logically interconnects the environmental parameters and the component failures. On the one hand, the resulting rules give an insight into the environmental conditions that have the highest impact on WT components failures. On the other hand, they enable an evaluation of which input parameters are the more appropriate, (a) or (b), which of the techniques 1-D K-means clustering or Multi-D K-means clustering is more appropriate as input to transactions for deriving association rules.

3.2.3 Results

Here, the findings obtained from Multi-D labelling, Supervised labelling and 1-D labelling, see Table 3.14, are processed using AARM in order to discover and visualise association rules between severity of downtime, major components failure and environmental conditions. Multi-D labelling was used to discover the general conditions prior to each failure for the wind speed (WS), the relative humidity (RH), the power production (P), and the severity and ambient temperature (T) throughout the whole observation period. The labelling and 1D clustering were used to define the conditions at a certain point in time, that occurred with high frequency. For example, $WS20$ indicates the 10-minute mean value for the second time step before failure (between 30-20 min before failure). $WSatF$ represents wind speed at time of failure occurrence.

Table 3.14: Ranges for manual Labelling and results for Clustering with One-Dimensional (1-D) and Multi-Dimensional (Multi-D) Input

Minimum and Maximum Values for the Labelling

Label	WS $[\mathrm{m\,s^{-1}}]$	Label	RH $[\%]$	Label	P $[-]$	Label	Downtime $[\mathrm{h}]$	Label	T $[^\circ\mathrm{C}]$
calm	< 3	dry air	20 - 40	consumption	$P_o < 0$	Minor	< 48	freezing	-10 - 0
low	3 - 10	moist air	40 - 60	not efficient (ne)	$0 \leq P_o < P_m$	Major	≥ 48	very cold	0 - 5
high	10 - 26	corrosive	60 - 80	efficient (e)	$P_o \geq P_m$			cold	5 - 10
storm	> 26	highly corr.	80 - 98					cool	10 - 15
		precipitation	100					mild	15 - 20
								room temp.	20 - 25
								warm	25 - 30
								hot	30 - 35
								very hot	35 - 40

Minimum and Maximum Values for the Clusters obtained for the 1-D Input

Cluster	WS $[\mathrm{m\,s^{-1}}]$	Cluster	RH $[\%]$	Cluster	P $[-]$	Cluster	Downtime $[\mathrm{h}]$	Cluster	T $[^\circ\mathrm{C}]$
1	0 - 6.568	1	32.46 - 62.3	1	0 - 0.27	1	27.5 - 144	1	3.34 - 10.65
2	6.62 - 12.98	2	62.84 - 79.19	2	0.3 - 0.7509	2	151.17 - 360	2	10.68 - 14.85
3	13 - 27.08	3	79.2 - 98.46	3	0.759 - 1	3	402 - 980.66	3	14.96 - 20.15
								4	20.44 - 30.13

Centroids of the Clusters obtained for Multi-D Input[1]

Cluster	WS $[\mathrm{m\,s^{-1}}]$	Cluster	RH $[\%]$	Cluster	P $[-]$	Cluster	Downtime $[\mathrm{h}]$	Cluster	T $[^\circ\mathrm{C}]$
1	3.97	1	64.63	1	0.12	1	72.68	1	10.24
2	9.36	2	80.78	2	0.53	2	226.84	2	17.03
3	15.98					3	563.98	3	22.69

[1] As the multi-dimensional input also results in multi-dimensional clusters, the minima and maxima cannot be displayed. Thus, the cluster centroids are being displayed instead, giving an idea about the location of the clusters.

The association rules are given in the next three Figures 3.42, 3.43 and 3.44. These illustrations are known as matrix plots. Here each circle represents one rule and support value of rules are given with radius of the circles, whereas colouring of circles stands for lift value of rules. Darker red indicates the rules with high lift values and bigger circles show the rules with high support values. Then, in order to give a more clear interpretation stronger rules are listed under each figure. In this analysis, pre defined support value is set to 0.035 and confidence value is set to 0.5. The strength of a rule is enumerated considering the value of a confidence being in the range of 0.5 and 1. If a rule has confidence value equals to 1, that rule is one of the strongest rules.

If we list in detail the rules illustrated in Figure 3.42 by sorting from the rules with high confidence values to the rules with low confidence values:

Rule	LHS	RHS	confidence
[1]	{Component=Frequency Converter ,Temp=2}	=>{Downtime=1}	1.0000000
[2]	{Component=Yaw System ,Temp=3}	=>{Downtime=1}	1.0000000
[3]	{Component=Frequency Converter ,Failureoccurrencezone=night, RH=1}	=>{Downtime=1}	1.0000000
[4]	{Component=Frequency Converter , Failureoccurrencezone=night , Temp=1}	=>{Downtime=1}	1.0000000
[5]	{Component=Frequency Converter, Failureoccurrencezone=night, P=2}	=>{Downtime=1}	1.0000000
[6]	{WS=1, Component=Yaw System, Temp=3}	=>{Downtime=1}	1.0000000
[7]	{Component=Yaw System , Failureoccurrencezone=day, Temp=3}	=>{Downtime=1}	1.0000000
[8]	{Component=Yaw System , P=2, Temp=3}	=>{Downtime=1}	1.0000000
[9]	{WS=1, Component=Yaw System, RH=1}	=>{Downtime=1}	1.0000000
[10]	{WS=2, Component=Yaw System, RH=2}	=>{Downtime=1}	1.0000000
[11]	{WS=1, Component=Yaw System, Failureoccurrencezone=night}	=>{Downtime=1}	1.0000000
[12]	{WS=1, Component=Yaw System, P=2}	=>{Downtime=1}	1.0000000
[13]	{Component=Yaw System, Failureoccurrencezone=night, RH=2}	=>{Downtime=1}	1.0000000
[14]	{Component=Yaw System Failureoccurrencezone=day P=2}	=>{Downtime=1}	1.0000000
[15]	{Component=Yaw System P=2 RH=2}	=>{Downtime=1}	1.0000000
[16]	{Component=Yaw System P=2}	=>{Downtime=1}	0.9523810
[17]	{Component=Frequency Converter Failureoccurrencezone=night}	=>{Downtime=1}	0.9411765
[18]	{Component=Yaw System Failureoccurrencezone=night}	=>{Downtime=1}	0.9285714
[19]	{Component=Yaw System Failureoccurrencezone=night P=2}	=>{Downtime=1}	0.9230769
[20]	{Component=Yaw System P=2 Temp=1}	=>{Downtime=1}	0.9230769
[21]	{Component=Frequency Converter RH=1}	=>{Downtime=1}	0.9166667
[22]	{Component=Frequency Converter P=2 RH=1}	=>{Downtime=1}	0.9166667
[23]	{Component=Yaw System RH=2 Temp=1}	=>{Downtime=1}	0.9166667
[24]	{Component=Yaw System RH=2}	=>{Downtime=1}	0.9047619
[25]	{Component=Yaw System}	=>{Downtime=1}	0.9000000
[26]	{WS=1 Component=Frequency Converter Failureoccurrencezone=night}	=>{Downtime=1}	0.9000000
[27]	{WS=3 Component=Generator}	=>{Downtime=2}	0.8571429
[28]	{WS=3 Component=Generator P=2}	=>{Downtime=2}	0.8571429
[29]	{Component=Generator Temp=2}	=>{Downtime=2}	0.7777778
[30]	{Component=Generator Failureoccurrencezone=day RH=2}	=>{Downtime=2}	0.7142857
[31]	{Component=Generator Failureoccurrencezone=day}	=>{Downtime=2}	0.6250000
[32]	{Component=Generator Failureoccurrencezone=day P=2}	=>{Downtime=2}	0.5714286
[33]	{Component=Gearbox RH=2 Temp=1}	=>{Downtime=3}	0.5333333
[34]	{Component=Generator}	=>{Downtime=2}	0.5000000
[35]	{Component=Generator RH=2}	=>{Downtime=2}	0.5000000

Before the evaluation of the over all analyses, their interpretation is given with an example rule using average support and lift values.

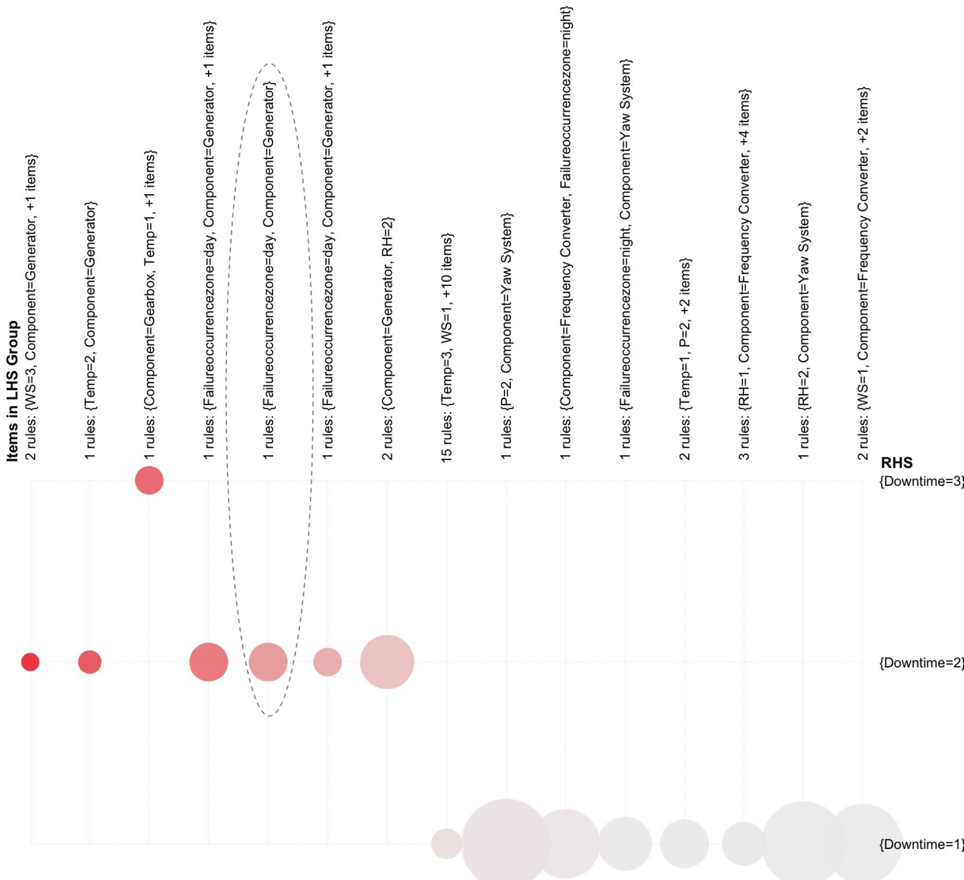


Figure 3.42: Grouped matrix used for Multi-D clustering rules for Downtime. Colour represents lift and circle size represents the support values, RHS stands for right hand side of the rule and LHS is the left hand side of the rule. Dashed ellipse indicates the example rule.

In the above given list, the example rule is in thirty-first position. Whereas, one of the strongest rules, in the list see the first rule, state that frequency converter failures occurred when temperature records were approximately 17 Celsius did not resulted with downtimes longer than 10 days.

$$Failureoccurrencezone = day, Component = Generator \rightarrow Downtime = 2$$

The example rule says that Generator failures recorded in day time causes in average 220 hours downtime.

Figure 3.42 shows the association between the most severe downtime occurrences and gearbox failures at low temperatures. Here, it is easy to see that the significant downtimes reported for generator failures might be controlled ones, since they occur during the day shift. However, they share the association between high corrosive and wind speed observations. Yaw system and frequency converter failures correspond to slightly severe downtime losses. In figures, lift and support values are shown and for the details of stronger rules, now we will present confidence metrics.

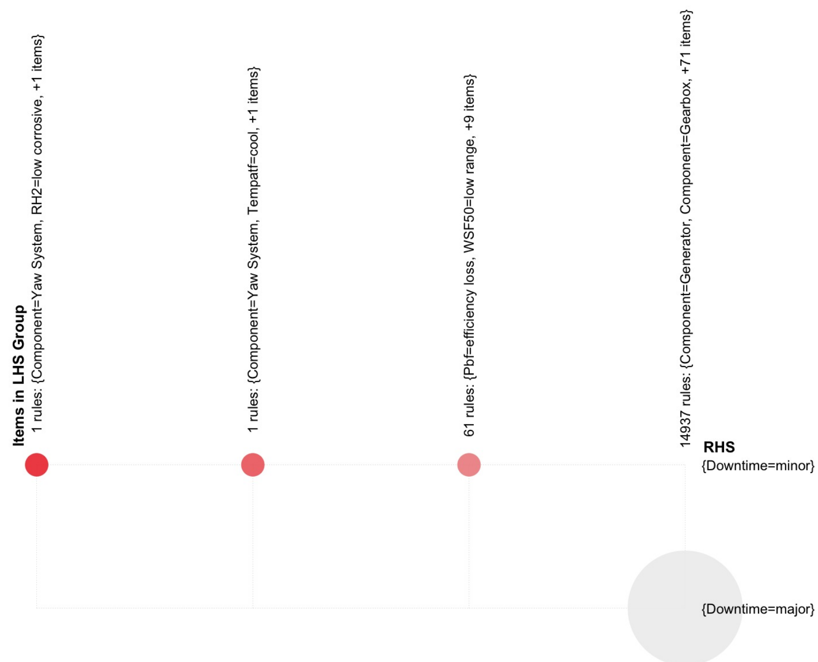


Figure 3.43: Grouped matrix used for Supervised labelling rules for Downtime. Colour represents lift and circle size represents the support values, RHS stands for right hand side of the rule and LHS is the left hand side of the rule.

Figure 3.43 gives a generic summary by stating that yaw system failures causes minor downtimes, whereas generator and gearbox failures are associated with major downtime occurrences.

Rule	LHS	RHS	Confidence
[1]	{Component=Pitch System,Tempmaxmonth=cool}	=>{Downtime=major}	1
[2]	{Component=Gearbox,Patf=efficient production}	=>{Downtime=major}	1
[3]	{Component=Yaw System,RH10=moist air}	=>{Downtime=major}	1
[4]	{Component=Generator,RH2=moist air}	=>{Downtime=major}	1
[5]	{Component=Generator,RH=moist air}	=>{Downtime=major}	1
[6]	{Component=Generator,Tempminmonth=mild}	=>{Downtime=major}	1
[7]	{Component=Frequency Converter,Tempmaxmonth=hot}	=>{Downtime=major}	1
[8]	{Component=Generator,Tempmaxmonth=hot}	=>{Downtime=major}	1
[9]	{WSF30=calm,Component=Generator}	=>{Downtime=major}	1
[10]	{Component=Gearbox,Tempmaxmonth=warm}	=>{Downtime=major}	1
[11]	{Component=Generator,Tempmaxmonth=warm}	=>{Downtime=major}	1
[12]	{Component=Gearbox,Tempatf=mild}	=>{Downtime=major}	1
[13]	{WSF20=calm,Component=Frequency Converter}	=>{Downtime=major}	1
[14]	{WSF20=calm,Component=Generator}	=>{Downtime=major}	1
[15]	{Component=Generator,Tempmeanmonth=room temperature}	=>{Downtime=major}	1
[16]	{WSF70=calm,Component=Generator}	=>{Downtime=major}	1
[17]	{WSF60=calm,Component=Generator}	=>{Downtime=major}	1
[18]	{Component=Gearbox,Pbf=consumption}	=>{Downtime=major}	1
[19]	{Component=Frequency Converter,Tempmeanmonth=cold}	=>{Downtime=major}	1
[20]	{Component=Frequency Converter,Tempatf=room temperature}	=>{Downtime=major}	1
[21]	{Component=Frequency Converter,Tempatf=cold}	=>{Downtime=major}	1
[22]	{Component=Frequency Converter,Tempminmonth=very cold}	=>{Downtime=major}	1
[23]	{Component=Frequency Converter,RH6=low corrosive}	=>{Downtime=major}	1
[24]	{WSatF=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[25]	{WSF80=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[26]	{WSF40=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[27]	{WSF70=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[28]	{WSF50=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[29]	{WSF60=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[30]	{WSF20=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[31]	{WSF10=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[32]	{WSF30=high range,Component=Frequency Converter}	=>{Downtime=major}	1
[33]	{Component=Frequency Converter,RH4=low corrosive}	=>{Downtime=major}	1
[34]	{Component=Pitch System,Tempmaxmonth=room temperature}	=>{Downtime=major}	1
[35]	{WSF50=calm,Component=Gearbox}	=>{Downtime=major}	1

Figure 3.44: Grouped matrix used for 1-D clustering rules for Downtime. Colour represents lift and circle size represents the support values. RHS stands for right hand side of the rule and LHS is the left hand side of the rule.

Although, Figure 3.44 offers a generic summary, it also provides more detailed associations compared with Figure 3.43. Here, it is easy to obtain shared characteristics in advance of failure occurrence, such as relative humidity levels 8 hours prior to failure, etc. While it should be noted that much more information could be drawn from the results,space limitations prevent a full discussion herein. Additionally, there are other environmental conditions, such as ice and snow, that can be responsible for certain failures. These have not been considered directly in this analysis, but enter indirectly through the temperatures and seasons.

Rule	LHS	RHS	Confidence
[1]	{Component=Generator,Failureoccurrencezone=day,Tmmean=3}	=>{Downtime=2}	1
[2]	{Component=Generator,RH8=3,Tmmean=3}	=>{Downtime=2}	1
[3]	{Component=Generator,Failureoccurrencezone=day,RH8=3,Tmmean=3}	=>{Downtime=2}	1
[4]	{Component=Frequency Converter,RH6=1}	=>{Downtime=1}	1
[5]	{Component=Frequency Converter,RH8=1}	=>{Downtime=1}	1
[6]	{Component=Frequency Converter,RH4=1}	=>{Downtime=1}	1
[7]	{Component=Frequency Converter,Tmmean=3}	=>{Downtime=1}	1
[8]	{Component=Frequency Converter,Tmmean=1}	=>{Downtime=1}	1
[9]	{Component=Yaw System,Tmmean=4}	=>{Downtime=1}	1
[10]	{Component=Yaw System,Tins=4}	=>{Downtime=1}	1
[11]	{Component=Yaw System,RH8=2}	=>{Downtime=1}	1
[12]	{Component=Yaw System,Patf=3}	=>{Downtime=1}	1
[13]	{Component=Yaw System,Tins=2}	=>{Downtime=1}	1
[14]	{Component=Yaw System,RH10=2}	=>{Downtime=1}	1
[15]	{Component=Yaw System,RH6=2}	=>{Downtime=1}	1
[16]	{WSatF=2,Component=Yaw System}	=>{Downtime=1}	1
[17]	{WSF10=2,Component=Yaw System}	=>{Downtime=1}	1
[18]	{Component=Frequency Converter,RH6=1,RH8=1}	=>{Downtime=1}	1
[19]	{Component=Frequency Converter,RH4=1,RH6=1}	=>{Downtime=1}	1
[20]	{Component=Frequency Converter,Patf=2,RH6=1}	=>{Downtime=1}	1
[21]	{Component=Frequency Converter,RH=2,RH6=1}	=>{Downtime=1}	1
[22]	{Component=Frequency Converter,Failureoccurrencezone=night,RH6=1}	=>{Downtime=1}	1
[23]	{Component=Frequency Converter,Pbf=3,RH6=1}	=>{Downtime=1}	1
[24]	{Component=Frequency Converter,RH4=1,RH8=1}	=>{Downtime=1}	1
[25]	{Component=Frequency Converter,Patf=2,RH8=1}	=>{Downtime=1}	1
[26]	{Component=Frequency Converter,RH=2,RH8=1}	=>{Downtime=1}	1
[27]	{Component=Frequency Converter,Failureoccurrencezone=night,RH8=1}	=>{Downtime=1}	1
[28]	{Component=Frequency Converter,Pbf=3,RH8=1}	=>{Downtime=1}	1
[29]	{Component=Frequency Converter,Patf=2,RH4=1}	=>{Downtime=1}	1
[30]	{Component=Frequency Converter,RH=2,RH4=1}	=>{Downtime=1}	1
[31]	{Component=Frequency Converter,Failureoccurrencezone=night,RH4=1}	=>{Downtime=1}	1
[32]	{Component=Frequency Converter,Pbf=3,RH4=1}	=>{Downtime=1}	1
[33]	{Component=Frequency Converter,Failureoccurrencezone=night,RH10=2}	=>{Downtime=1}	1
[34]	{Component=Frequency Converter,Tins=1,Tmmean=1}	=>{Downtime=1}	1
[35]	{Component=Frequency Converter,Failureoccurrencezone=night,Tins=1}	=>{Downtime=1}	1

3.3 Chapter conclusions

This chapter is dedicated to the review and exemplification of available databases, explanation of initial data analysis and interpretation of the findings of exploratory data analysis for the wind farm operation and maintenance decisions.

In the first section, the drivers of the electricity market price are investigated. The electricity demand can be considered especially for the estimation of the peak prices, but the demand times series do not provide a good reference especially for the estimation of bottom prices, what a maintenance engineer would like to know in order to schedule maintenance activities. Especially, the electricity generation time series in France from Nuclear and Fossil gas, in Spain and Germany from Onshore wind and Fossil hard coal, and in the Netherlands and the UK from Fossil gas can be tracked as information providers. Energy mixture policy of a country does not stand only for the varying amounts of the electricity generation from different sources. It covers also the electricity generation timing from different energy sources. Base load provider technologies such as Nuclear (the Netherlands, the UK, Germany, Spain) exert very weak or zero influence on the electricity market prices, but when Nuclear is considered as load following technology in the energy mixture, its influence on the electricity market price increases severely. Peak load following generators such as Fossil gas exert very high influence on electricity market prices, even in some cases its influence is stronger than the electricity demand (Spain, France, Germany). When Onshore wind is considered as peak following technology in the energy mixture (Spain and Germany), its influence increases in comparison to being in a role of the supply sources of the mid-merit load consideration (the UK, the Netherlands, France).

It can be noted that for the first and the last months of the year the extremes of price follows the extremes of the load and the load is significantly higher in winter months. During April, May, June and November the case countries show the bottom prices. In these months a scheduled component replacement or a maintenance visit can be considered.

As future work, detected monthly key driving factors for electricity market price can be used for monthly multivariate electricity market price models, which are capable of providing good electricity price forecasts two or three weeks ahead.

Whereas in the second section an analysis was presented for correlating wind turbine failure observations and various environmental conditions in order to provide information to the O&M team at the day of the intervention. The proposed framework is capable of providing a summary for the whole observation period and the information prior to failure occurrence at interested time steps. Therefore, it can be considered as an enhancement tool for offline inspection data collection. The proposed framework is designed as an ensemble of supervised labelling, unsupervised clustering techniques and apriori rule mining algorithm.

For prior to failure analysis 1-D clustering and supervised labelling are suggested considering the interpretation of association rules, which behave as a decision support system for a maintenance team to evaluate the health status of the components other than the subject component of the initially planned maintenance. The supervised labelling technique requires expert opinion and labours classification for each variable. 1-D clustering technique can provide information fast and without requirement of an expert guidance. Although, Multi-D clustering is weak to provide information at specific time steps, this technique is useful to make annual maintenance plans since it provides a summary for the overall downtime occurrences.

Chapter 4

Revenue Tracking for Maintenance Decisions

The main objectives of maintenance management are ensuring the availability and the durability of the assets, as clearly defined by international standards such that EN 13306:2010 Maintenance - Maintenance Terminology [143]. This entails performing the maintenance of the wind turbine at optimum costs, and verifying the quality of the product (power performance in our case) while also considering costs where necessary.

As a nature dependent energy source, wind has the characteristics of being intermittent, which makes the coarse O&M planning costly and O&M decision making process difficult. The revenue of the wind farms can be modelled considering wind resource dependent power estimations, wind turbine health status dependent operational hours and electricity demand influenced electricity market prices, which display differences in different countries. The planning of the maintenance tasks of a wind farm shall account for all these factors which are important for the revenue. Although wind is an intermittent source, these factors (wind resource, turbine health status and electricity market price) can display some common characteristics. As an example, the monthly seasonality of the electricity market price and the seasonality of the wind resource can be noted [185, 186]. In Chapter 3, until some degree, exploratory analysis for electricity market prices and failure occurrences have already been shown and discussed

Here, the contribution of these three factors (wind resource, turbine health status and electricity market price) will be examined in the maintenance decision making process.

Commercial operation and maintenance of wind farms always involves the selection of the most cost-effective solution from various options. In this chapter, the process of generating alternative scenarios is explored, as well as the evaluation criteria required to determine the optimal decision from among the produced alternatives. Indications reflecting operational, financial, and environmental factors are all examined in this evaluation. The decision making process is analysed in terms of the power performance and the net present value from the cash flow resulting from the energy sales. The case presented involves a maintenance action on a Spanish wind farm, where a blade replacement was required to prevent a catastrophic failure. According to the maintenance log books and the power generation records, the conducted replacement was discovered by an under performance resolved with the blade replacement and a later blade re-pitching.

The impact of the timing of the maintenance is evaluated in various what-if scenarios. The sensitivity to environmental causes of under performance is compared by varying the duration of blade icing, and comparing the performance of different wind directions. Country dynamics and subsidy impacts are hypothetically evaluated for the prevailing electricity market conditions as if the turbine were operating in either Spain, the Netherlands or the UK. In the following sections, firstly the description of the input data is given. Then, the studied scenarios are explained. The scenario generation procedure and the revenue modelling framework are given in detail. The findings of the scenario analysis are reported. Lastly, the major conclusions that are derived from this chapter are summarised.

4.1 Input Data

The study presented required various data-sets and a high amount of input data collection. As it was already introduced in Chapter 3, there exist various databases to be considered. A categorical summary of the input data is given in Table 4.1. The resolution and unit differences are considered in the data cleaning and preparation phases. Here, environmental data obtained from National Centers for Environmental Prediction (NCEP), nearby meteorological stations, SCADA and maintenance history from wind turbines, country specific financial indicators obtained from statistics published by Organisation for Economic Co-operation and Development (OECD) and electricity market prices obtained from European Network of Transmission

System Operators for Electricity (ENTSOE) are used. The relevant maintenance history of the studied turbine is shown in Table 4.2. This is important when conducting power performance analyses and generating reference power curves. Lastly, in Tables 4.3 and 4.4, the major parameters concerning cash flow settings such as duration of the investigation, time steps, interest index, currency and set-ups of Net Present Value (NPV) analyses such that subsidies inclusion, tax inclusion etc. are given.

Table 4.1: Summary of the case study data

Category	Variable	Resolution
Turbine SCADA	Wind speed mean	10 min
	Wind speed variance	10 min
	Active power mean	10 min
	Ambient temperature mean	10 min
	Average generator speed	10 min
	Nacelle orientation mean *	10 min
Wind farm met mast	Pressure	10 min
Met station [187]	Pressure **	1 day***
NCEP [179]	Relative humidity (at 850 mbar)	6 hours
ENTSOE [161]	Day-ahead market price Spain	1 hour
	Day-ahead market price UK	1 hour
	Day-ahead market price Netherlands	1 hour
OECD [188]	Consumer price index	1 month
	Long-term interest rate	1 month

*: Approximation used in case where the wind direction is unavailable.
**: Substitute for missing data, altitude corrected.
***: Average of daily minimum and maximum recording.

Table 4.2: Maintenance history of the investigated turbine

Number	Date	Type	Event
1	09/2012	Repair	Brake pad replacement
2	05/2013	Inspection	Blade inspection
3	07/2013	Repair	Anemometer replacement
	07/2013	Inspection	Main bearing inspection
4	09/2013	Repair	Blade repair on site
5	04/2014	Repair	Tower repair
6	08/2014	Repair	Communication repair
7	10/2014	Repair	Converter repair
8	05/2015	Major repair	Preventative blade replacement
	05/2015	Repair	Repair of brake pumps
9	09/2015	Optimisation	Re-pitching of blades

According to the maintenance history presented in Table 4.2, there were several maintenance actions previous to the major blade replacement of the wind turbine. These are important to define the signature of major power anomalies, which are used to generate case specific power curves and related energy estimations. The blade replacement is performed in May 2015, 4 months later another maintenance intervention was reported due to the need for pitch adjustment for the investigated turbine.

Table 4.3: Parameters for cash flow set-up and NPV analyses.

General parameter	Definition
Duration	May 2014 to May 2017
Time steps for cash flow (t)	Monthly
Total number of months (N)	37
Index (i)	Monthly long term interest rate - consumer price index
Currency in cash flow settings	Baseline: EUR, UK case: GBP

Table 4.3 presents the common parameters for NPV analyses and cash-flow set-up. The investigated period goes from May 2014 to May 2017 covering 37 months in total. The cash flow set-up is defined in monthly steps. In order to consider inflation, the interest rate used in the analysis is defined as

the difference between the monthly long term interest rate and the consumer price index. In this study three case countries are investigated, Spain and the Netherlands, using currency in EUR and UK, using currency in GBP.

Table 4.4: Settings for cash flow

Setting	Relevant month(s)	Cash flow type	Details
Baseline	1	cash-out	Blade costs: 70000 EUR
	[13:37]	cash-in	Monthly sum (hourly energy * hourly electricity price)
Subsidies included	[13:37]	cash-in	Monthly sum (hourly energy * subsidy)
Tax included	20,32,37	cash-out	Annual sum (hourly energy * hourly electricity price * profit rate * tax rate), annual financial closures for tax calculations are in December and in the end of the analysis.

In Table 4.4 the baseline scenario cash flow setting, used as reference for all the comparisons, consists on the cash-out in May 2014 for the spare blade procurement and the monthly cash-in considering the energy sales from May 2015 to the end of the investigation period. Subsidies included scenarios have modifications in monthly cash-in amounts, whereas tax included scenarios have cash-out in December 2015, December 2016 and May 2017.

4.2 Studied Scenarios

The studied scenarios can be grouped into three categories: Environmental conditions, maintenance timing and country dynamics. In each group, the energy estimations and the cash flow set-ups are modified by changing some variables. Then, the NPV calculations are performed separately for each case. In total 43 scenarios that are subject to univariate changes are studied, being 22 in maintenance timing, 6 in environmental conditions and 15 in country dynamics including the baseline scenario. Additionally, country dynamics and maintenance timing parameters combined to study multivariate changes in 180 scenarios. The details of the scenarios are given in Table 4.5.

Table 4.5: List of evaluated scenarios

Category	Scenario group	Baseline value	Tested values
Maintenance timing	Optimisation delay	141 days	0, 15, 30, 60, 90, 120, 180 days
	Additional downtime	0 days	15, 30, 60, 90 days
	Shifting of the preventative intervention	0 months	1, 2, 3, ..., 11 months
Environmental condition	Icing	19.7 days	0, 26.7, 33.7, 40.7 days
	Wind direction	124/203 days WSW/NNW	327 days WSW, 327 days NNW
Country dynamics	Country	Spain	Spain, Netherlands, UK
	Taxed revenue	0 %	10 %, 20 %, 30 %
	Subsidy	no	scheme 1, scheme 2

In Table 4.5, the maintenance timing category focuses on the shifting of either re-pitching or major repair. The baseline scenario represents 141 days optimisation delay without any downtime corresponding to the interval in the real case between blade replacement and blade re-pitching. In order to estimate the financial consequences of a possible downtime occurrence, four scenarios corresponding to hypothetical downtime duration in time horizons of half month (15 days), one month (30 days), two months (60 days) and three months (90 days) are examined. Shifting of the preventative intervention is studied for 11 scenarios corresponding to 11 months. For the investigation of different timings of the intervention in the year, the duration of the underperformance was unchanged, i.e. the underperformance was only 'shifted' by a number of months.

In the second category, the effects of environmental conditions such as icing and the leading wind direction are studied. Here, in the baseline scenario the leading wind direction is found to be north-north-west (NNW) corresponding to 203 days of the investigation period and the second major wind direction is reported as west-south-west (WSW) corresponding to 124 days of the investigation period. The analysis period for the energy estimations covers approximately 654 days. In order to evaluate the influence of the changes in the leading wind direction in energy estimations, two scenarios are defined as wind flows 327 days from WSW and 327 days from NNW. Although, there are no icing and precipitation data available for this study, for the baseline scenario, humidity and temperature measurements [190] are considered to estimate icing occurrence. Then, it is estimated that 19.7 days of the analysis period correspond to icing phenomena. Four scenarios are defined to test the effect of icing phenomena as 0 days (icing-free), 26.7 days (1 week added to the baseline), 33.7 (2 weeks added to the baseline) and 40.7 (three weeks added to the baseline). The last category consists of 15 scenarios, being 6 of them defined for Spain (including the baseline), 4 for the Netherlands and 5 of them defined for the UK. In these set-ups, country dynamics are considered using the differences in subsidy, tax and electricity price market for the three EU countries.

4.3 Framework

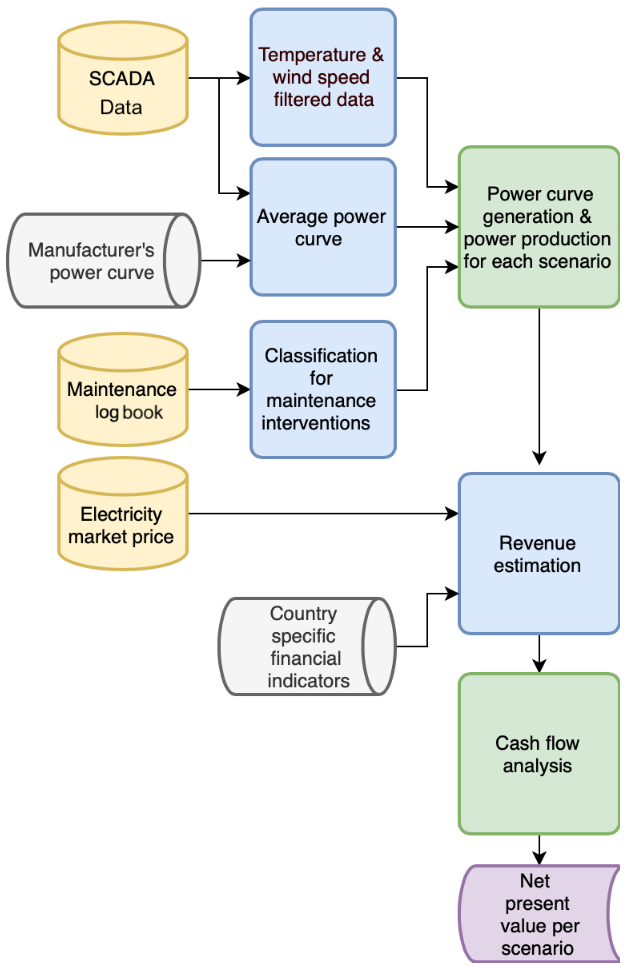


Figure 4.1: Methodology flowchart

A complete framework has been developed for performing the study. The process flowchart is given in Figure 4.1, where the required data and filtering operations as well as financial calculations are illustrated. The analysis is conducted in MATLAB and R [191]. There are two connected major processes, which are coloured as green blocks. The first one stands for the generation of the trained power curve for each scenario. The second covers the cash flow analysis for the NPV estimations.

The power curve has been obtained following the IEC standard for wind turbines [192] using the mean values of the power production. The averages should be calculated for each 0.5 m/s bin, including wind speeds in the relevant operation range of the turbine. Two years of wind speed and power production records are used to define the energy estimation of the baseline sensitivity scenario. For the sensitivity study of the environmental and intervention timing effects, the scenario design requires alterations in energy estimations, therefore performance of the investigated turbine was modified by conducting the following steps:

1. Determine which condition (icing, leading direction, temporary repair shift etc.) matches with the investigated time period.

2. Generate reference and modified power curves and define them for each relevant condition of the turbine.

3. Use wind speed measurements and reference power curves to interpolate within the produced power estimation.

4. Use wind speed measurement as the first applicable condition to interpolate power production that is dependent on the power curve.

5. Ensure the operational ability of the turbine, and make sure the actual production is still lower than the power curve of the scenario (or still higher in case of decrease).

6. Derive difference of interpolated power production as power correction.

7. Filter the correction to permit 5 % or less of a deviation from the new power curve (with only an upper limit for power increase and only a lower limit for power decrease).

8. Repeat steps 4-7 for all conditions and apply the power correction which results in the lowest power production for each time step.

The issue of seasonality is solved through derivation of one power curve for each three month season. The impact of the wind direction was examined using 12 wind direction sectors, i.e. 30^0 each, with the derivation of one individual power curve for each one. Periods with risk of ice on the blades were filtered prior to any power performance analysis [192]. For the purposes of this study, in which precipitation and icing data are unavailable, it is assumed that icing occurs when relative humidity $> 80\,\%$ and temperature $< 2\,°C$ [190].

The maintenance actions, blade replacement, and blades re-pitching were the focus of decreasing cash flow in the financial evaluation. The complex and case specific nature of the real cash flow in a wind farm lead to the use of a simplified chronology, where an initial investment for the repair costs was paid off in the following years. Turbine procurement was disregarded, while all generated income was utilised to balance the maintenance expenses. Repair costs were backdated to 2014 while energy sales were considered beginning in May 2015, covering the periods for spare blade acquisition and blade replacement, respectively. This procedure provided an estimate of the net present value for each scenario and for each case country.

4.4 Findings

In this section, the NPVs calculated for each scenario are grouped under three sub-titles as follows, effect of environmental conditions, maintenance timing and country dynamics. The interpretation of the NPV's is done in comparison to baseline scenario's NPV value.

The baseline NPV, i.e. the Spanish case without considering subsidy or tax, emerged as 102,549 EUR. However, it should be emphasised that relative changes of the NPV are of interest in this study, as the absolute NPV value is a consequence of the assumptions taken in the cash flow setup. The baseline is compared with various scenarios of optimisation delay, additional downtime and shifting of the intervention (as listed in Table 4.5).

4.4.1 Effect of environmental conditions

Real environmental data are only available for the case of Spain. Then, the effect of environmental conditions is tested only for the Spanish case.

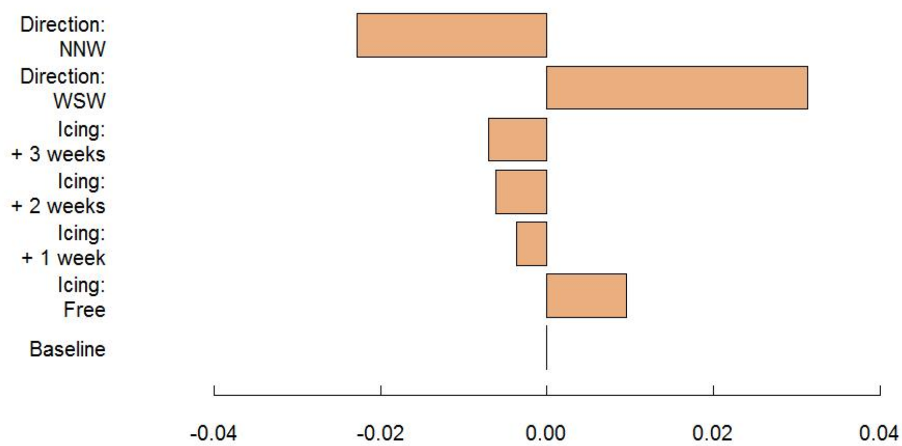


Figure 4.2: Effect of environment on NPV. NNW: north north west, WSW: west south west, + stands for added to Baseline.

The results, presented in Figure 4.2, show that the leading wind direction is found to be more important than the length of anomalous period due to icing.

4.4.2 Effect of maintenance timing

Figure 4.3 shows the resulting NPV for the maintenance timing scenarios. It can be seen that both, the duration of the optimisation delay (optimisation after 15, 30, 0, 90, 120 and 180 days from the blade replacement) and monthly shifting of the preventative intervention have a significant impact on NPV, but none of them is as dramatic as downtime. A direct optimisation (delay 0d: 0 days) results e.g. in a small NPV increase, whereas 90 days of downtime reduce the NPV more than 6 times of the NPV increase amount observed in the direct optimisation scenario. NPV does not change linearly with the length of underperformance or downtime due to the varying wind resource. The results of the NPV analysis shifting the preventative intervention to different months in the year showed that shifts of 1, 2, 9, 10 or 11 months result in a higher NPV than the baseline. Whereas, shifts of 3, 4, 5, 6, 7 and 8 months result in a lower NPV than the baseline due probably to the influence of seasonality of the electricity market prices.

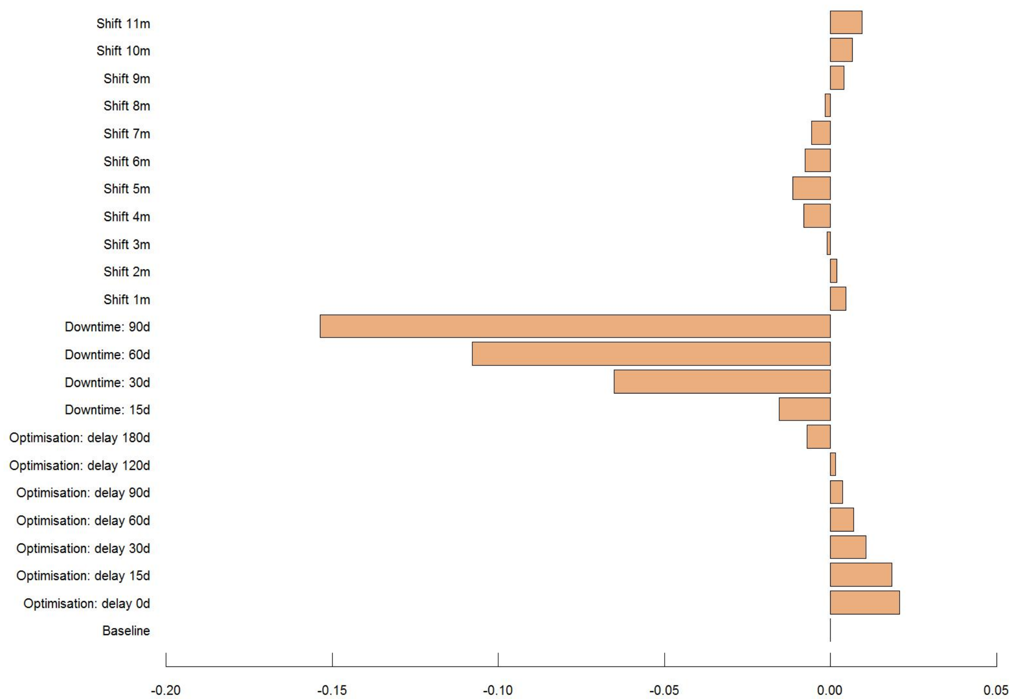


Figure 4.3: Effect of maintenance timing on NPV. Temp. is temporary, m: month, d: day

4.4.3 Effect of country dynamics

The effect of taxes and subsidies for all cases is summarised in Table 4.6.

Table 4.6: Effect of corporation tax, wholesale electricity prices and subsidies for each country

Country	Tax rate	Subsidy	NPV (1000 EUR)
Spain	none	no	102.5
	10 %	no	97.9
	20 %	no	93.3
	30 %	no	88.7
	none	premium	214.9
	none	fixed	221.7
Netherlands	none	no	104.1
	10 %	no	99.7
	20 %	no	95.3
	30 %	no	90.8
UK	none	no	150.7
	10 %	no	145.3
	20 %	no	139.8
	30 %	no	134.4
	none	premium	151.0

In Spain and the UK cases, subsidy schemes are investigated, while the effect of subsidies in Spain approximately doubles NPV, whereas in the UK it causes a smaller NPV increase. The influence of tax policy is quite similar in the Netherlands and Spain due to the same currency.

In order to evaluate the influence of calendar based preventative intervention in different countries, the shifting of the maintenance intervention is compared in Figure 4.4. In this figure, it can be seen that the ranking of the preferable intervention month varies depending on the country. We need to note that the scaled NPV values (sNVP) are obtained using as reference the NVP estimation corresponding to May (original major intervention month).

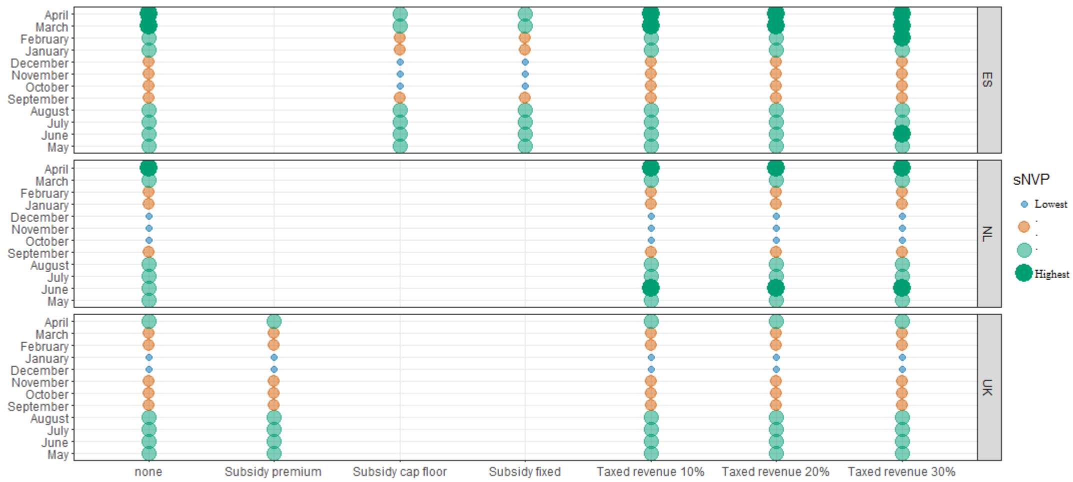


Figure 4.4: Ranking of shifting option, sNVP is the scaled NPV for each policy in respect with the NPV obtained in May, both color and size of circles show sNVP.

In the Spanish case without subsidy and without tax scenario April and March are the preferable months. The scenarios with subsidy inclusion suggest that January, February and September should be avoided for such an intervention. The influence of taxed revenue at levels of 10% and 20% have similar results. Both of them suggest March and April as the preferable months. However, taxed revenue at level of 30% shows a higher NPV separation between months, such that April, March, February and June are preferable and September, October, November, December are not recommendable with respect to May NPV.

The Netherlands case shows that April is the preferable month for all studied financial scenarios, however inclusion of tax cuts show also June as a preferable candidate for the investigated maintenance intervention.

In the UK case, there is a strong agreement on the ranking of months in all studied financial set-ups. January and December should be avoided. April, June, July and August can be considered similar with respect to May.

4.5 Chapter conclusions

The financial consequences of underperformance are shown in the NPV of the cash flow. There is an optimisation potential in terms of the timing of any maintenance that results in temporary underperformance. The analysis of shifting the maintenance period through the year shows that the actual timing was optimised to the seasonal wind resource trends. The financial results indicate that the optimal timing will change due to the different seasonality of electricity markets. However, this is affected by the complexity of the cash flow, the electricity market in each country, taxes and subsidies. In most configurations, a shifting to earlier spring appears to be more profitable.

The NPV sensitivity study with icing and wind directional performance variation demonstrated that wind direction is more important in comparison to icing scenarios studied.

The comparison for the three countries highlights that based on the electricity markets alone, the UK (approximately 150700 EUR) and the Netherlands (approximately 104100 EUR) were more attractive for the wind farm studied due to higher NPV values than Spain (approximately 102500 EUR). If the subsidies are included, the Spanish baseline is far more advantageous than the Netherlands and the UK. However, it should be noted that the attractive Spanish subsidy scheme ended and new wind farms may rely only on electricity market sales.

Lastly, in the investigation of multivariate sensitivity scenarios, the UK shows a fixed ranking for maintenance months being independent from financial policies. The Netherlands case shows small differences for taxed revenue scenarios with respect to those without subsidy and without tax case. Whereas, the Spanish case shows significant maintenance month ranking differences for the most preferable and the least preferable months when different financial set-ups (tax and subsidy) are considered.

Chapter 5

Planning of a Maintenance Day

One of the major responsibilities of maintenance management is to contemplate the safety and any other mandatory requirements associated with item (lathe machine in metal cutting, screwdriver in construction etc. and of course wind turbine in our case) [143]. In this chapter, we will discuss the implementation of this responsibility in the wind farm maintenance planning. This chapter also bases on production losses due to maintenance caused by low accessibility to the wind turbine (resulting from weather and coarse planning).

Operation and maintenance (O&M) scheduling of wind farms is a challenging task due to the fact that turbines are frequently located in relatively inaccessible locations such as offshore or in mountainous terrain. Maintenance teams must follow specific procedures when performing their service. The relation between the effectiveness of a maintenance plan and accuracy of wind speed/gust forecast data will be shown in a quantitative comparison in this chapter. Herein, an enhanced maintenance scheduling framework is developed based on real case studies, which involve an analysis of the maintenance history of wind turbines in a transition phase between stochastic 'mid-life' failures and 'end of life' failures, by taking into account mean wind speed and wind gust predictions.

The methodology proposed in this chapter is designed as a Decision Support System (DSS) to find the optimal intervention time and the most effective execution order for different maintenance tasks. The methodology was built on information from regular maintenance visit tasks and a corrective maintenance visit involving a generator replacement. The goal of the work is to enable the timely prediction of executable and not executable tasks to be carried out during pre-planned maintenance days.

In this manner, routine maintenance tasks are grouped using the findings of an agglomerative nesting analysis. Then, the DSS is tested with actual observations and error introduced synthetic forecast data sets. It is shown with the case studies that the proposed DSS is capable of preventing the planning downtime (due to weather) from a couple of hours to a day.

5.1 Data source, wind farm maintenance procedure and data

Maintenance logs, service work orders and SCADA data were obtained from a Spanish wind farm. According to the information gathered, the average durations of the biannual, annual, biennial and quinquennial visits are approximately 21, 26, 15 and 18 hours respectively. The total number of different tasks to be performed in maintenance visits is 169. Most of them, 117, are included in the biannual visit actions while the rest are distributed over the other visits. However, not all maintenance actions are carried out during each planned visit as some of them have priority based on the findings of previous service visits and the needs of the wind turbine.

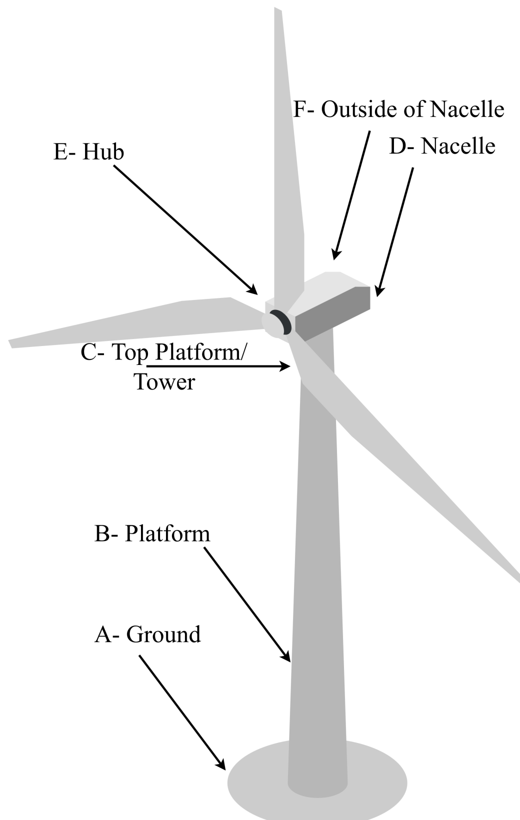


Figure 5.1: Example of turbine working zones

Table 5.1: Executed tasks for the scheduled visit

Turb. Work Zone	**Sub System**	**Task Numbers**
A-Ground	Tower	1 to 2
A-Ground	Electrical Parts	3
A-Ground	Rotor-Blades	4
B-Platform	Electrical Parts	5 to 7
C-Tower	Yaw System	8 to 14
D-Nacelle	Main Shaft and Bearing	15 to 17
D-Nacelle	Gearbox	18 to 27
D-Nacelle	Generator	28
D-Nacelle	Base Structure and Cover	29 to 31
D-Nacelle	Electrical Parts	32
E-Hub	Rotor	33 to 34
F-Outside of Nacelle	Sensors	35 to 36

Figure 5.1 shows the considered turbine working zones, while the task numbers associated to these zones are listed in Table 5.1. In this work, for the sake of simplicity, only the tasks numbers listed in Table 5.1 are used to define a case study considering a regular service visit, for the details please see Appendix 1. A second case study is based on a major intervention, which requires a crane usage. More specifically, a generator replacement is studied and more information will be provided regarding the corresponding task. To explain the working environment of the service personal for performing either a regular service or a major intervention, the seasonal and general characteristics of the subject wind farm are shown in Figures 5.2 and 5.3.

In Figure 5.2, the wind speed seasonal histograms corresponding to the studied wind farm in Spain are presented. The annual histogram is included in each graphic to highlight the seasonal contribution. It can be seen that the majority of wind speed observations lie between 0 and 10 m/s in summer months. Then, summer looks the best season for maintenance actions, but there are still a significant number of wind speed observations with values higher than 15 m/s.

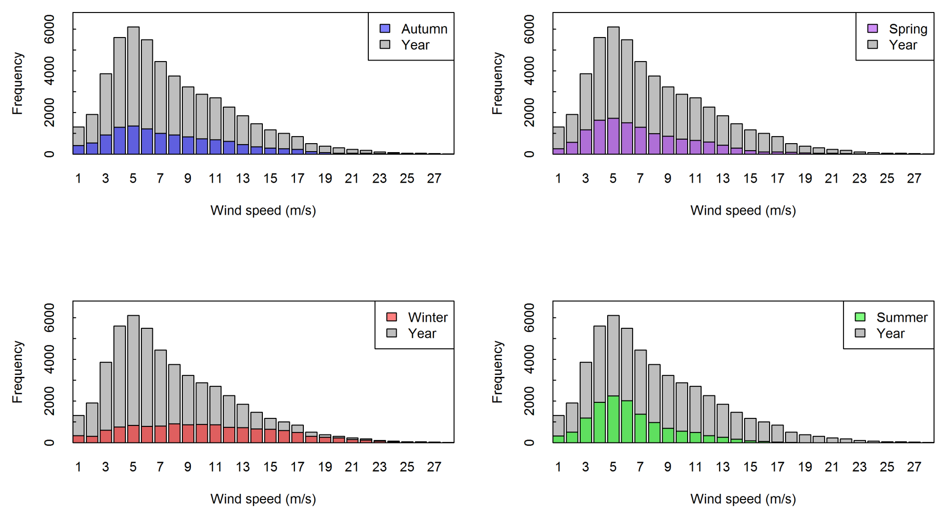


Figure 5.2: Annual versus seasonal wind speed histograms

Figure 5.2 shows the seasonal characteristics of the nacelle wind speed obtained from the analysed wind farm. It is known that the seasonal wind speed behaviour is dependent on the location of the wind farm. The annual maintenance plan must be prepared considering the seasonal wind behaviour and the electricity market prices of the country where the analysed wind farm is located. Then, the seasonal wind behaviour is an important factor for long term scheduling, which is not the aim of this study. The resulting program from the annual maintenance plan is an input to decision making support tool. Therefore, this input must be modified, when the analysed wind farm is changed.

Figure 5.3 illustrates the diurnal behaviour of the wind speed for each season during 2017 comparing the maxima of the 10 minute averages in one hour for all seasons. It can be seen that the day shift (08:00 to 18:00) in summer, with maxima of 10 minute mean wind speeds lower than 20 m/s, indicates relatively reasonable wind farm accessibility to perform a maintenance visit.

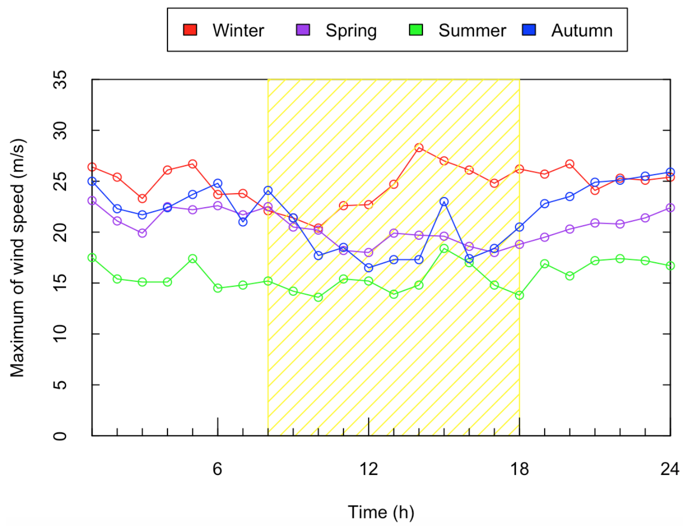


Figure 5.3: Seasonal wind speed trends as hourly maximum of 10 minute averages. This figure is obtained calculating the maxima of 10 minutes averaged wind speed measurements per hour of each day over a season in 2017. The window, which is shaded in yellow represents the day shift from 08:00 to 18:00.

The majority of scheduled maintenance interventions are planned in summer and autumn months in the case study maintenance log. For this reason, the cases are modelled for summer and autumn conditions, and input wind speed and wind gust measurements obtained from the data corresponding to the studied wind farm in Spain are shown in Figures 5.4 and 5.5.

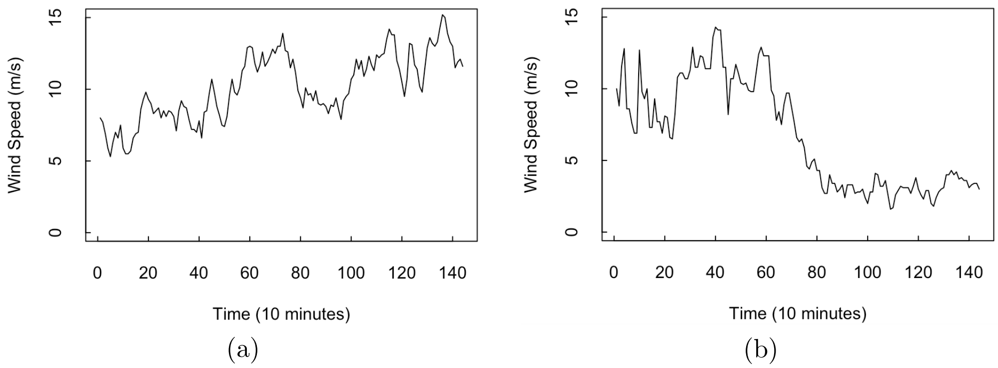


Figure 5.4: Input wind speed measurements for the routine maintenance visits (a) summer day example and (b) autumn day example

In Figure 5.4 wind speed sample data for the routine maintenance scheduling are shown. From this figure, it can be observed that these two days have almost ideal features for the wind farm accessibility as they do not have wind speed occurrences with values higher than 20 m/s.

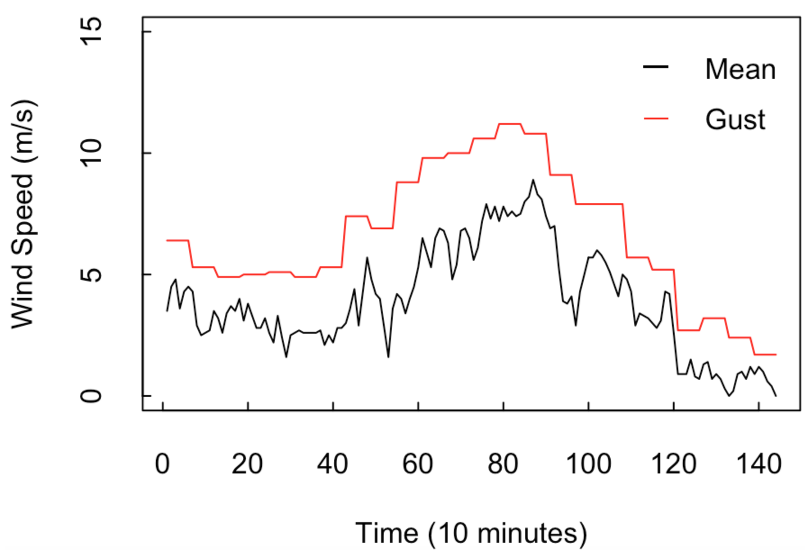


Figure 5.5: 10 minutes mean wind speed and hourly gust during 24 hours for a summer day. Gust value is repeated for six time steps, since it is only available as one maximum value per hour.

In Figure 5.5, test data for a corrective maintenance visit day are given.

Wind speed data are available in 10 minutes resolution as averaged values. Whereas, wind gust data are measured by the turbine nacelle anemometer as 1 second values during an hour. From Figure 5.5, it is clear that there is a significant difference between maximum of 10 minutes averaged data and gust measurements.

5.2 Methodology, case set-ups and proposed frameworks

In Chapter 2, agglomerative nesting, the general concept of search algorithms and ARIMA are explained. Herein, these tools are used in the developed model. Thus, firstly the mathematical formulation of the scheduling problem and the search algorithm for the generation of feasible solutions are presented. Then, the proposed framework is introduced. Finally, the test methodology developed to consider input forecasting accuracy is explained.

5.2.1 Scheduling problem

The problem studied belongs to the operations research field, where scheduling problems and their solution algorithms are one of the major topics of concern. In this field, the scientific perspective of decision making involves the application of mathematical representation for actual situations and then, the usage of optimisation models for choosing the best option amongst given possibilities. As it is said in [109], "an optimisation model seeks to find values of the decision variables that optimise (maximise or minimise) an objective function among the set of all values for the decision variables that satisfy the given constraints". In an optimisation model, objective function, decision variables, and constraints must be defined properly.

The mathematical interpretation can be written as:

For a variable X
Optimise $F(X)$
Subject to: $\{X \in S_1, G(X) \in S_2\}$

where X is a vector of decision variables and $F(X)$ is the objective function subjected to restrictions X belonging to S_1 and $G(X)$ belonging to S_2.

By following the same notation, now we will define the wind farm maintenance scheduling problem. In a single visit, there can be various tasks to be performed by a maintenance team, and each task requires a completion time and the fulfilment of the HSE rules regarding accessibility to the working zone. For n number of tasks to be completed in a single visit, each task's (and later on each cluster's) required time window for completion is C_k, where k is in $\{1, 2, ..., n\} \in \mathbb{Z}$. For a working interval W, $W = [t_1 : t_w]$ in 10-minute resolution, $t_1 = 1$ indicates the starting time step, t_w is the final time step. The assigned time slot for execution of the cluster of tasks is symbolised with A_k, Figure 5.6 shows graphically A_k and C_k.

When a maintenance team visits a wind turbine and stops it, the corresponding duration without power production is called maintenance downtime. If a maintenance team's work is interrupted due to unfavourable weather conditions, the team must wait until the safe working rules and weather conditions are met, and meanwhile the turbine remains in an idle state. This additional waiting time is called weather downtime and will be denoted as Z in this study.

For the first assigned task k during interval W, the corresponding weather downtime Z_k, equals the difference Δt_k, between t_1 and the starting time step of the first assigned task. The k^{th} weather downtime occurs between the completion time of task k and the starting time of the next assigned task and so on. Then, using the given notation, the total duration of a single visit, L, will include the function of each task k and its corresponding weather downtime:

$$Z_1 + A_1 + Z_2 + A_2 + ... + Z_n + A_n = L \tag{5.1}$$

With these assumptions, $W \geq \sum_{k=1}^{n} C_k$ and $W \geq L$ guaranteeing the execution of the tasks with duration L, during interval W in multiple ways. It is assumed that once C_k is assigned to a window, its execution requires continuous work without any break or interruption.

In a compact form, L can be written as:

$$L_r = \sum_{k=1}^{n} [Z_k + A_k]_r \tag{5.2}$$

where r stands for the task completion sequence, $r = \{1, 2, ..., n!\}$. As an example, for a task pool containing four candidates, r=1 represents sequential assignment of (C_1, C_2, C_3, C_4) as (A_1, A_2, A_3, A_4) into W, whereas r=2 stands for the assignment of (C_1, C_3, C_2, C_4) as (A_1, A_2, A_3, A_4) into W.

Placement of C_k into A_k is done using the decision vector p_k of the same length (in order to fit the tasks in W) containing the value of b_k. b_k is constructed for each task and each time step as:

$$b_k = \begin{cases} 1, & v_t < V_k \;\;\&\;\; g_t < G_k \\ 1, & v_t < V_k' \;\;\&\;\; g_t \geq G_k \\ 0, & v_t \geq V_k \;\;\&\;\; g_t < G_k \\ 0, & v_t \geq V_k' \;\;\&\;\; g_t \geq G_k \end{cases} \tag{5.3}$$

where v is the wind speed, V_k is the HSE wind speed limit of task k, g is the wind gust, G_k the wind gust limit of task k, and V_k' is the HSE wind speed limit of a task k when the wind gust is higher than its limit ($V_k' = V_k$-2 m/s).

In this problem our variable is the the total duration of the scheduled tasks, L, and the objective function for this maintenance scheduling problem is:

$$\min\{L_r\} \;\; \forall \text{ r; r} \in S_n \tag{5.4}$$

where S_n represents all permutations of the elements of the task completion sequence, r.

This objective function is subject to the following constraints:

$$\begin{array}{l} \sum_{k=1}^{n} A_k = \sum_{k=1}^{n} C_k \\ \forall C_k, \exists (p_k = 1) \text{ for } W = [t_1 : t_w] \\ W \geq \sum_{k=1}^{n} C_k \\ W \geq L \end{array}$$

By finding the optimal configuration for elements of r, it is possible to perform a maintenance visit with minimum total duration.

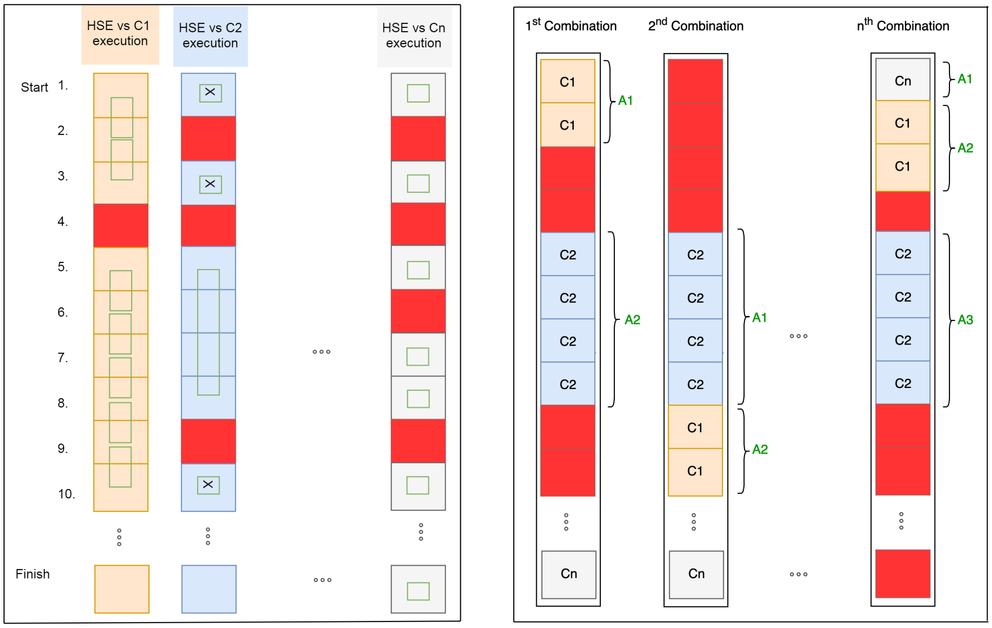
5.2.2 Search for the optimal time window

After defining the scheduling problem in detail and generating all possible combinations, a search algorithm must be used to find the optimal one.

Therefore an extensive decision pool, containing a prioritised list of all possible scheduling combinations, can be provided to the decision maker. For such a decision pool, all scheduling combinations must be generated considering problem-specific heuristics. This straight forward way is known as brute force search by its definition in literature [110]. Brute-force search is simple to implement and it always finds a solution, if it exists.

As an example, a maintenance visit can include the execution of 36 tasks grouped in 4 clusters, whereas another one could be defined just with 4 tasks. Tasks and clusters have two common features: the execution duration and the corresponding wind and/or gust related safety restrictions. Specifically, a task is the fundamental element and a cluster consists of many tasks. A task has its own safety restrictions and execution duration, whereas the cluster execution duration is the summation of its member tasks' execution durations. The safety restriction for a cluster corresponds to the most restrictive wind speed limit found for its member tasks. The optimal schedule is then chosen from the whole set of execution combinations. Furthermore, the selection criteria for the optimal solution depend on the minimum execution time, the starting time and the work shift. It is worth highlighting that the minimum execution time implies the execution of all tasks avoiding downtime due to weather restrictions as much as possible. The process of generating the combinations for the clusters is as follows.

The algorithm uses the wind speed forecast, the wind speed working restrictions and the clustered tasks. Initially, the algorithm starts matching the wind speed limit of each cluster with the wind speed forecast for the whole period (typically one day) obtaining the allowed forecast windows for each cluster, as can be seen in Figure 5.6a.



(a) (b)

Figure 5.6: Search algorithm working principles: (a) Forecast-Safety rule matching (b) Clusters allocation

Hereinafter, the execution of n number of clusters for a single visit will be examined. In Figure 5.6, the squares stand for 10 minutes accessibility periods. Each red square represents non-executable period for a corresponding cluster. As an example, let's assume that the execution of the Cluster 1 (C_1) requires 20 minutes accessibility (i.e., two ten-minutes time steps) to the corresponding wind turbine location, whereas the needed time for the Cluster 2 (C_2) is 40 minutes and for the final cluster (C_n) is 10 minutes. In Figure 5.6a, it can be observed that the execution of the Cluster 1 can be performed from the first step ($Start$) until the 3^{rd} as it can be placed in two different manners in that interval or from the 5^{th} to the 10^{th} (in this interval five different options are available). Regarding the Cluster 2, although there exist time steps confirming the HSE requirement (time steps 1 and 3), its length is not enough to perform all the tasks of the Cluster. Therefore, these tasks can only be executed from the 5^{th} to the 8^{th} time steps.

In the second part of the process, the clusters are allocated together into the allowed forecast windows based on their duration, see Figure 5.6b. *A* symbolises the assigned task in Figure 5.6b, A_1 starts from the first time step in 1^{st} to n^{th} combinations but in the second combination, A_1 starts at fifth time step. This illustrates why the scheduling differences occur among combinations. Via cluster permutation, the assignment is done as many times as possible whilst changing the allocation order and obtaining all possible execution windows.

The whole process is then repeated increasing the starting time in order to shift the forecasting assignments by one time-step. The process finishes when there is no room to allocate the minimum execution duration. In this way, a solution plan pool, which consists of many maintenance plans (combinations), is generated. The best combination minimises the downtime occurrence (red blocks), and it must reflect the most appropriate start and finish time according to the decision maker's preferences.

5.2.3 Proposed framework

The proposed methodology is graphically explained in Figure 5.7. The initial step is to provide information on the type of the intervention, safe working rules and wind forecasts. Then, it is required to decide if wind gust measurements and estimations are needed as decision variables. The corresponding answer depends on the specific requirements of the planned intervention, such that intervention may require a crane usage.

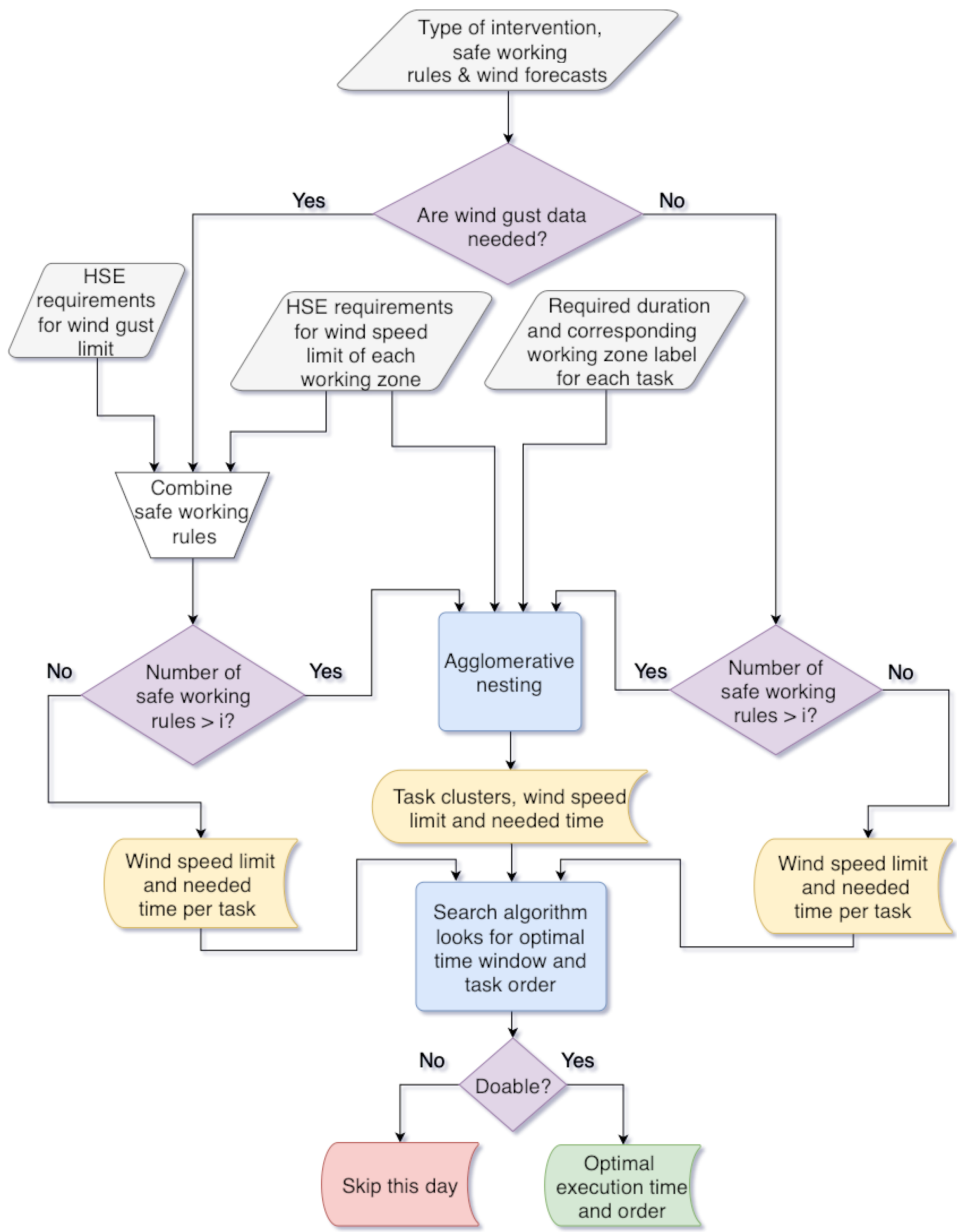


Figure 5.7: Flowchart of proposed solution, HSE: Health-Safety and Environment regulations, i: user defined limit for entering agglomerative nesting process.

Here, we assumed that a maintenance task can be done within a minimum of four stages such as: access to working area, access to failed component, removement of failed component and placement of the new component. Then, for a case that each stage requires a unique safe working rule, the minimum total number of safe working rules is 4. Therefore, predetermined comparison value, i, is set to 4 .

For an intervention consisting of more than four tasks or requiring the fulfilment of more than four safety rules related to wind speed, forecasts must be used along with the outcomes of the agglomerative nesting as input in the search process. The gust forecasts are necessary if the intervention is performed using a crane, which requires reduction of the wind speed limits due to the high gust values. Lastly, the search process scans the available time windows during the intended maintenance day to find the optimal time window for the work to take place. If the maintenance intervention can be executed during the pre-planned day, optimal execution time and order of the tasks are determined. If not, a change in the pre-planned day is suggested.

This methodology can also be used for offshore applications, but it is very important to update HSE requirements considering wave height and offshore operations specific rules. Moreover, intervention type, required duration, outputs of annual maintenance, etc. must be updated considering the technology type and the working environment.

5.2.4 Forecasting accuracy measurement and synthetic data generation

As explained previously, the wind speed forecasts are the fundamental input variables of the process. Here, we will test the developed algorithm's sensitivity to input forecasting accuracy. To do this, appropriate forecasting accuracy measurements must be selected and test data must be generated.

In order to assess the accuracy of the forecasts, one of the widely used metrics is selected: root mean square error (RMSE) [137, 193].

Recent work in wind speed forecasting suggested that novel methodologies capable of generating forecasts with high accuracy (less than 0.5 m/s in RMSE with a forecasting of 3 time steps ahead, i.e. 30 minutes), could be useful for operational scheduling [194, 195]. When the forecast lead time increases, accuracy levels decrease. Moreover, at least one day ahead hourly forecasts might be required for operational planning [196]. In a recent study, the accuracy of 1 hour ahead forecasted hourly wind speed predictions was reported within 0.93 m/s and 1.01 m/s in RMSE. However, the accuracy ranged from 1.82 m/s to 2.35 m/s for 6 hours ahead forecasted hourly predictions [197]. Then, in this work, we will consider the forecasting accuracy for wind speed ranging from 1 m/s to 1.8 m/s. For the wind gust value, it is difficult to select a forecasting accuracy interval due to the lack of wind gust forecasting studies. Nevertheless, one study was found presenting a Wind Gust Estimation (WGE) method which states that the typical error range is about 5 m/s for the gust estimations and presents estimation accuracies for different cases with RMSE values 2 m/s and 5 m/s [198]. Considering the limited literature in this field, we assumed that statistics for the difference between the mean wind speed and the gust wind speed can be used in order to define accuracy margins for the synthetic wind gust data generation. The closest statistics to the existing literature were the maximum and the mean. Therefore, a maximum difference 5.1 m/s and a mean difference 2.7 m/s between mean wind speed and gust wind speed were chosen.

According to the existing state of art, wind speed forecasting error can be obtained with Auto regressive models. These models, proposed by Box and Jenkins [136], have been extensively used for short term forecasting. They are commonly known by the generic name of ARIMA (Autoregressive Integrated Moving Average) [137, 138].

ARIMA is one of the most simple and easy to implement methods for synthetic data generation [133–135]. Here the purpose of the synthetic data is not to generate forecasts, but to test the influence of accuracy loss of weather prediction in the developed DSSs. In practice, wind farm O&M engineers use commercial weather predictions, which are usually obtained using Numerical Weather Prediction (NWP) models. Therefore, in order to imitate the forecasts with a certain level of accuracy, the synthetic data generated with the ARIMA model are sufficient for the purposes of the work.

The future value of the variable is described by a linear function of the previous data and a random error. ARIMA models are generally denoted by ARIMA(p,d,q) where the parameters p, d, and q are non-negative integers, representing the order of the autoregressive model, the degree of differencing and the order of the moving average respectively. When d equals to zero and p and q equal to 1 the ARIMA(p,d,q) becomes the ARIMA(1,0,1).

ARIMA (1,0,1) is the equivalent of ARMA (1,1) model. The ARMA model is capable to mimic the correlation between errors in time, as it exists in the wind data [134]. Hence, according to the state of the art [135], wind speed scenario generation can be performed with the ARMA (1,1) by just simulating the error instead of the point wind speed forecasts. The simulation data stands for the generated noise that will later be added to the historical data in order to produce the synthetic forecast with the desired error.

Therefore, in this study, the ARMA (1,1) is used to generate the simulation data as a replacement of the forecasting errors. Then, the simulated error are added to the measurement, in order to obtain synthetic time series with a known forecast error.

The white noise ϵ, and n number of random observations generated using the normal distribution with mean value equal to 0 and standard deviation equal to sd_h. These observations are used as input for ARMA (1,1) model defined with the autoregression operator ϕ_h and the moving operator θ_h. In the next step, the second data set was generated having the features of sd_{h+h}, ϕ_{h+h} and θ_{h+h}, where h is a constant incremental amount, which takes the initial value 0. The resulting simulation data are added to the original measurements, symbolised as mes. The RMSE calculations are performed using the mes and the ARMA(1,1) simulation data. This process is repeated within a recursive loop, until the desired RMSE levels obtained. The resulting ARMA(1,1) simulation data sets are used as the synthetic data sets. By following this mathematical notation, the synthetic data generation process can be expressed as:

$$\begin{aligned}
arma11 &= RNG(n, sd_h) \\
\epsilon &= RNG(n, sd_h) \\
arma11_t &= arma11_{t-1} * \phi_h + arma11_{t-1} * \theta_h + \epsilon_t \\
Set_0 &= mes_t + arma11_t \\
RMSE_0 &= \sqrt{(}mean((Set_0 - mes_t)^2)) \\
RMSE_{goal} &=? RMSE_h \ \textit{if yes Synthetic} = Set_0 \\
RMSE_{goal} &=? RMSE_h \ \textit{if no} \\
h = h + h \ & \textit{rerun the process using the final h value}
\end{aligned}$$

5.3 Case 1: Routine Maintenance

Clustering

The problem of planning a high number of tasks is simplified by applying the agglomerative nesting methodology to the pool of 36 tasks. Clustering was performed using the Euclidean distance as similarity criterion. It was calculated using the wind speed limit and the corresponding turbine working zone of each operation. Figure 5.8 shows how the tasks are grouped forming a total number of 4 clusters (represented with different colors) as a function of the restrictions, wind turbine working zone and wind speed.

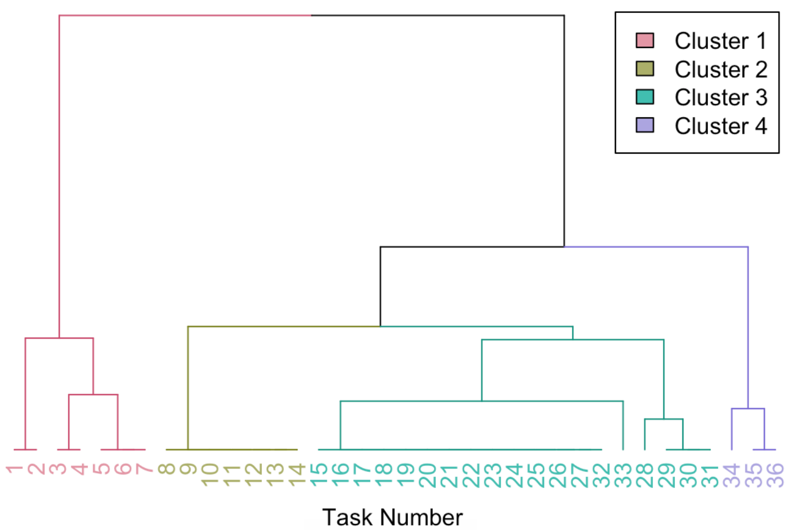


Figure 5.8: Graphical representation of the clustering process. Different colours represent the different clusters of tasks.

A summary of the clustering results is given in Table 5.2 where the cluster duration and its wind speed limit are shown. As maintenance tasks are usually accomplished by two technicians, which will require half the time, and the required resolution for the planning schedule is based on 10-minute steps, the rounded duration per person on a 10-minute scale is also provided.

Table 5.2: Clustering results

Cluster	**Duration (min)**	**Per person (10 mins)**	**v_{lim} (m/s)**
1	66	4	20
2	106	6	15
3	491	25	12
4	50	3	10

24 hours evaluation for executable/not executable windows

Now by applying the procedure, explained in Section 5.2.2, with measured wind speed data of test days (the first day was 24^{th} July 2017 and the second, 7^{th} October 2017) executable and not executable periods for the maintenance clusters are determined. Figure 5.9 and Figure 5.10 show the allowed intervention starting times for each of the clusters found in the previous section.

Execution of the maintenance service is only possible, if the starting time of the intervention is within the green dots. Here green dots represent valid periods for both wind speed safe working limit and the availability of a window to accomplish the task within its minimum required completion duration. In these figures, v_{lim} represents wind speed limit and Dur stands for the required duration for the execution of the corresponding cluster.

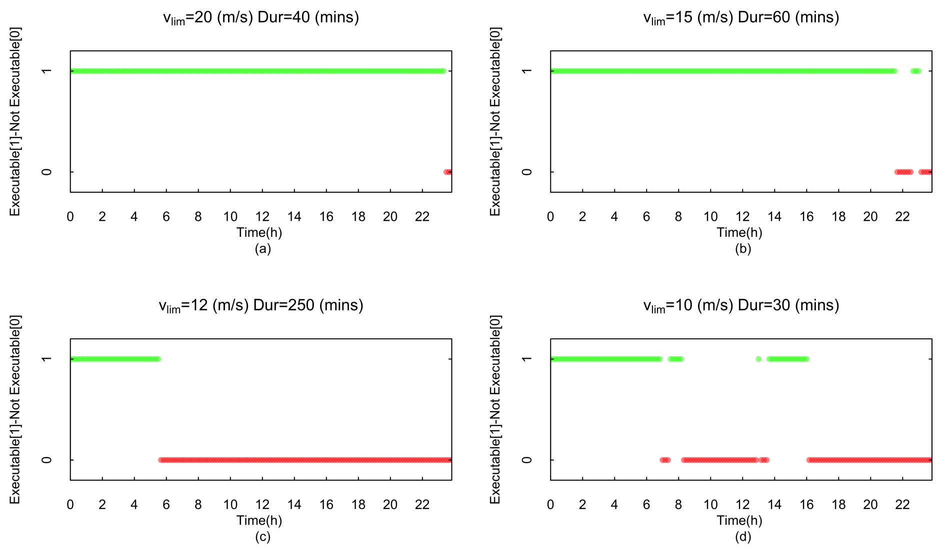


Figure 5.9: Routine maintenance evaluation with actual input data for 24^{th} July 2017 (a) cluster 1, (b) cluster 2 (c) cluster 3 (d) cluster 4

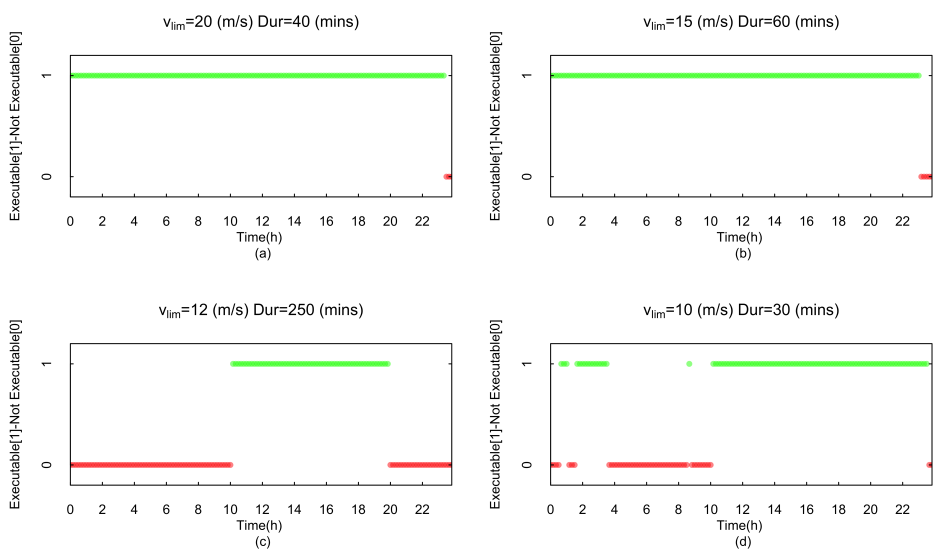


Figure 5.10: Routine maintenance evaluation with actual input data for 7^{th} October 2017 (a) cluster 1, (b) cluster 2 (c) cluster 3 (d) cluster 4

The results for each of the clusters were:

Cluster 1: Tasks are executable during both analysed days, since the corresponding wind speed restriction is very flexible and its duration is relatively low, see Figures 5.9a and 5.10a.

Cluster 2: Tasks are mostly executable for both days, see Figures 5.9b and 5.10b. There is a small non-executable window in the first test day, see Figure 5.9b.

Cluster 3: Tasks are mostly non-executable for both days, see Figures 5.9c and 5.10c. Cluster 3 tasks are the most challenging group, because they require a longer time with major wind speed restrictions. During the first test day the executable period is limited for starting times at the morning hours, however during the second test day the executable window starts from mid-day and continues during the day shift.

Cluster 4: The execution of Cluster 4 tasks depends mostly on the most restrictive wind speed limit. Nevertheless, it can be seen that there exist some executable windows, since the execution of this cluster requires the lowest duration. During the first test day, allowed windows were observed early in the morning and after lunch-time, see Figure 5.9d. In contrast, the cluster was mostly executable during the autumn day after 11:00, see Figure 5.10d.

Table 5.3 and Table 5.4 present the numerical characteristics of the allowed executable windows, including the total number of solutions, the starting time for each cluster, the cluster execution order and the total duration for both studied days. The first and last optimal solutions are shown in both tables. It is worth highlighting that the selection of the best solutions depends on the preferences of the decision maker. In this manner, the first solution implies that the earliest possible execution plan, which starts at the earliest hour, and last solution stands for the latest execution plan, which finishes at the latest hour. Here the evaluation of the earliest and the latest execution plans depends on the decision maker's preference. If the decision maker considers a night shift, the first test day is a valid day for the maintenance visit. Otherwise, only the second day (during the day shift) is a valid day for the maintenance visit.

Table 5.3: Cluster execution starting time and execution order for actual data in the first test day, total number of solutions=201

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Execution order**	**Total Dur(min)**
Earliest plan	00:00	00:40	01:40	05:50	1234	380
Latest plan	04:50	09:40	05:30	04:20	4132	380

Table 5.4: Cluster execution starting time and execution order for actual data in the second test day, total number of solutions=592

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Execution Order**	**Total Dur(min)**
Earliest plan	08:30	09:10	10:10	14:20	1234	380
Latest plan	19:00	18:00	19:40	17:30	4213	380

Summarising the results, it can be stated that a routine maintenance visit for the first test day should start very early making use of the night shift, see Table 5.3, due to the Cluster 3 restrictions, as shown in Figure 5.9. In contrast, a routine visit is possible during the day shift for the second test day as shown in Table 5.4 and Figure 5.10. If these plans are compared to default visits starting at 08:00, in the first day it may not be possible to execute the plan at all due to inaccessibility for the Cluster 3 tasks, see Figure 5.9 c. Regarding the second test day the Cluster 1 and the Cluster 2 are executable, however the execution of the Cluster 3 and the Cluster 4 tasks may be subject to a couple of hours downtime, see Figure 5.10c and Figure 5.10d.

Using hourly electricity market price data obtained from ENTSOE platform [199] and energy losses for each plan estimated from measured wind speed values and manufacturer's power curve, the revenue prioritised decision pools can be obtained as given in Figures 5.11 and 5.12. This DSS is prepared as a computational tool and the visualisation of the reporting module is given in 5.13, where the alternative plans and the revenue evaluation procedure are exemplified.

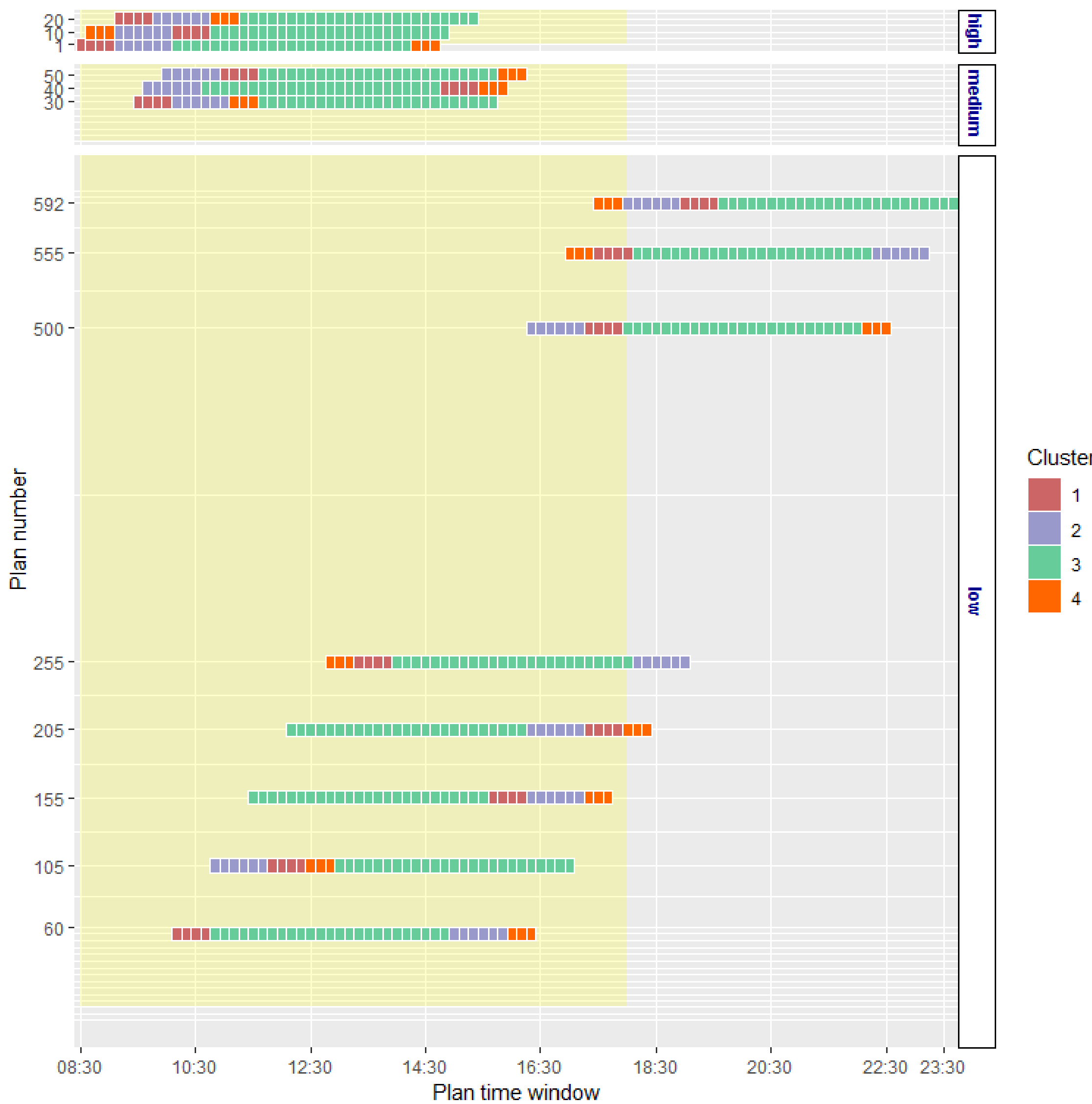


Figure 5.11: Decision pool for routine maintenance visit scheduling for the second test day, the second y axis stands for the grouping according to the revenue losses, yellow shaded window shows the day shift. For the sake of clarity, only selected cases are plotted when a clear trend can be observed (14 of 592). Example interpretation of the plan 1, which matches with the standard procedure: At 08:30 the intervention starts with the cluster execution order 1,2,3,4 and results with high revenue losses in comparison to the alternative plans.

In the next figure, Figure 5.11 is re-generated considering only the plans which are executable during the day shift.

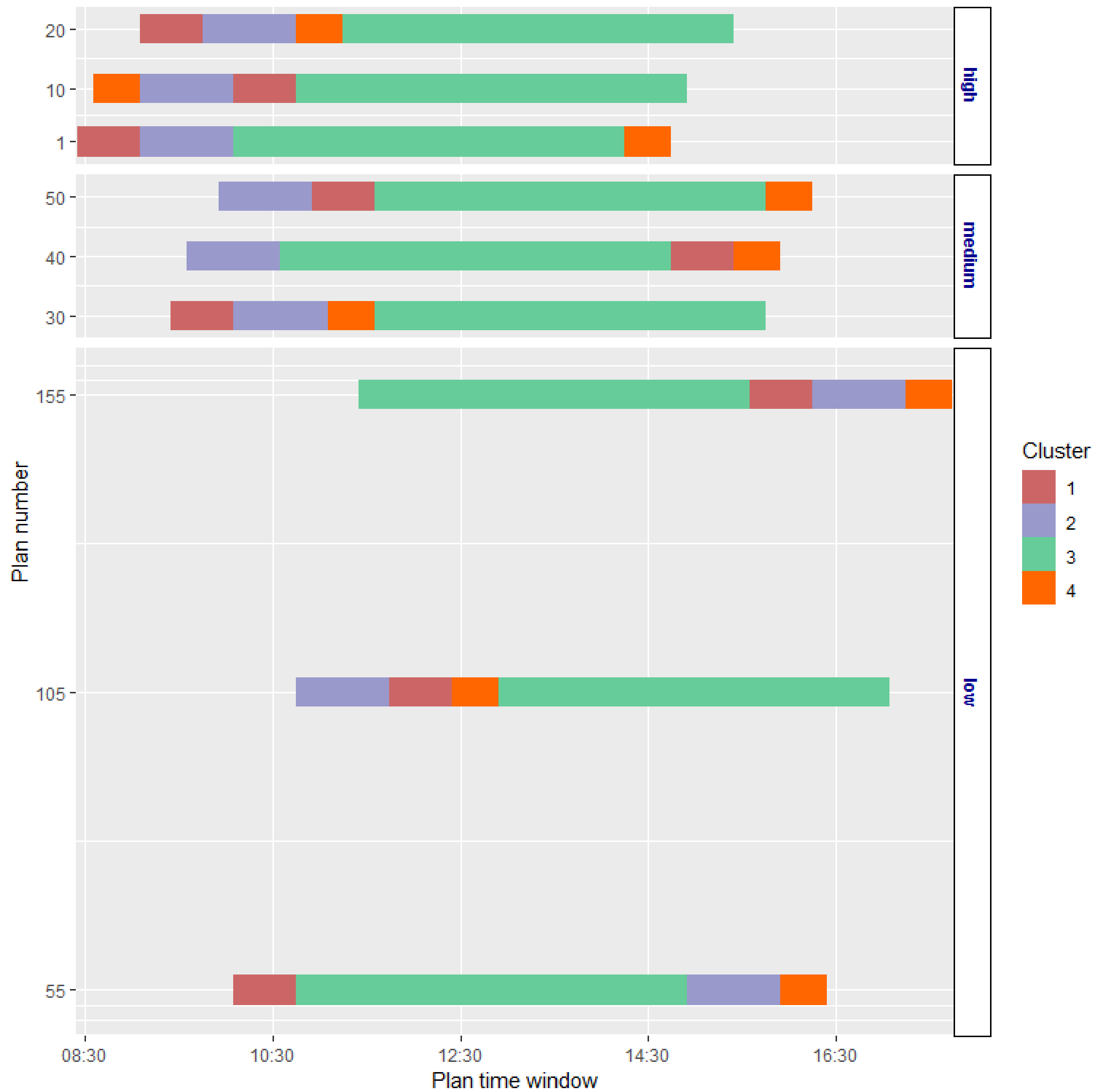


Figure 5.12: Decision pool of day shift working hours for routine maintenance visit scheduling for the second test day, the second y axis stands for the grouping according to the revenue losses. For the sake of clarity, only selected cases are plotted when a clear trend can be observed (9 of 592).

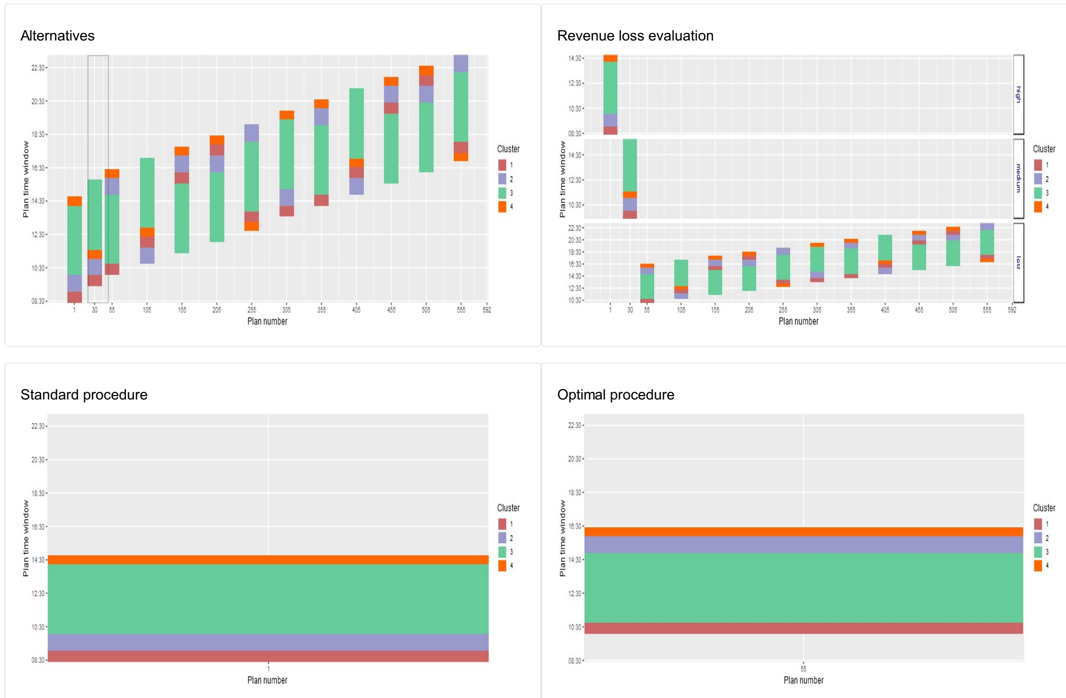


Figure 5.13: Maintenance plan evaluation tool

The reporting module of the developed DSS is given in Figure 5.13. This module consists of four zones.

- Alternatives: it is possible to see all the alternatives, which confirm the safe working rules. Among them, it is possible to request only a sample, in this example only 13 of them are shown.

- Revenue loss evaluation: In this window, for the plans that are given in the ‘Alternatives’, the revenue losses are estimated and the plans are grouped as high, medium and low.

- Standard procedure: This window refers to the default procedure, regarding the day shift which starts at 08:30 and the execution of the clusters performed in an order of 1,2,3 and 4.

- Optimal procedure: This is the optimal plan obtained with the proposed methodology, which results in the minimum revenue loss.

Wind speed forecasting

As displayed in the previous section, scheduling a maintenance day consists of challenging decisions even with knowing real wind speed values as it is displayed in the previous section. Moreover, the scheduling must be performed in advance, which means without knowing the exact wind speed values, so wind speed forecasts with an acceptable accuracy must be used. To generate these forecasts, various models were tested and their accuracy comparison is presented here for the same test days previously introduced as 24^{th} July 2017 and the second, 7^{th} October 2017.

For longer than 6 hours forecast horizons, Numerical Weather Prediction (NWP) models need to be used. In this work, the forecasts from two NWP models have been used. These are the European Centre for Medium Range Weather Forecasting (ECMWF) model [200, 201] and the National Centers for Environmental Prediction Global Forecast System model (GFS)[202].

ECMWF and GFS forecasts provide 6 hourly and 3 hourly resolution respectively. As the final resolution needed was 10-minutes, all models were over-sampled repeating the last available value as many times as needed to fit to the final resolution. Regarding their lead times GFS and ECMWF forecasts were available 10 days ahead. The accuracy ranking among the numerical weather prediction methodologies is given in Tables 5.5 and 5.6.

Table 5.5: Wind speed forecasting accuracy of the numerical models for the first test day

Model	**RMSE (m/s)**
ECMWF	2.65
GFS	3.56

Table 5.6: Wind speed forecasting accuracy of the numerical models for the second test day

Model	**RMSE (m/s)**
ECMWF	2.09
GFS	2.98

Influence of the forecasting accuracy in the routine intervention: NWP forecasts

In addition to actual data, the decision support system (DSS) was also tested with NWP wind speed forecasts to track the effect of the forecasting accuracy in the generated maintenance plans. In order to verify this effect, the downtime due to weather restrictions is analysed here. As previously highlighted, the tests were performed for two days.

In Figures 5.14 and 5.15, the results for the maintenance intervention using NWP forecasts obtained from ECMWF are showed for two test days. In these analyses the forecast horizons are 24 hours. The RMSE values are calculated using each forecast value in 10-minute resolution during 24 hours of the analysed day and observed measurement values of the same day.

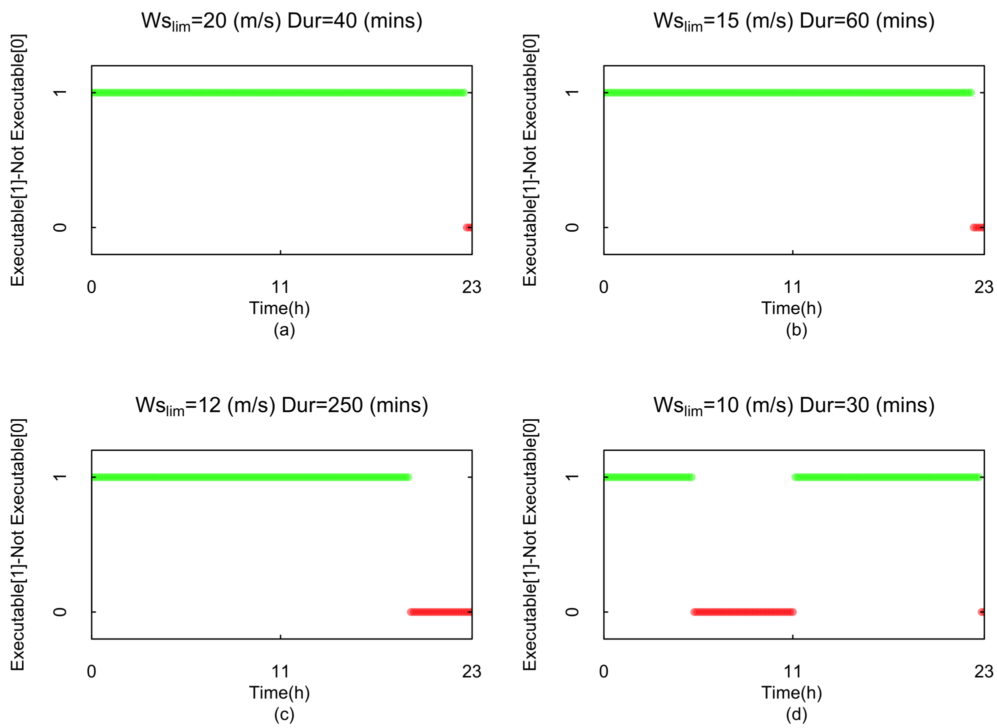


Figure 5.14: Routine maintenance evaluation with ECMWF forecasts for 24^{th} July 2017 (a) Cluster 1 tasks, (b) Cluster 2 tasks (c) Cluster 3 tasks, and (d) Cluster 4 tasks

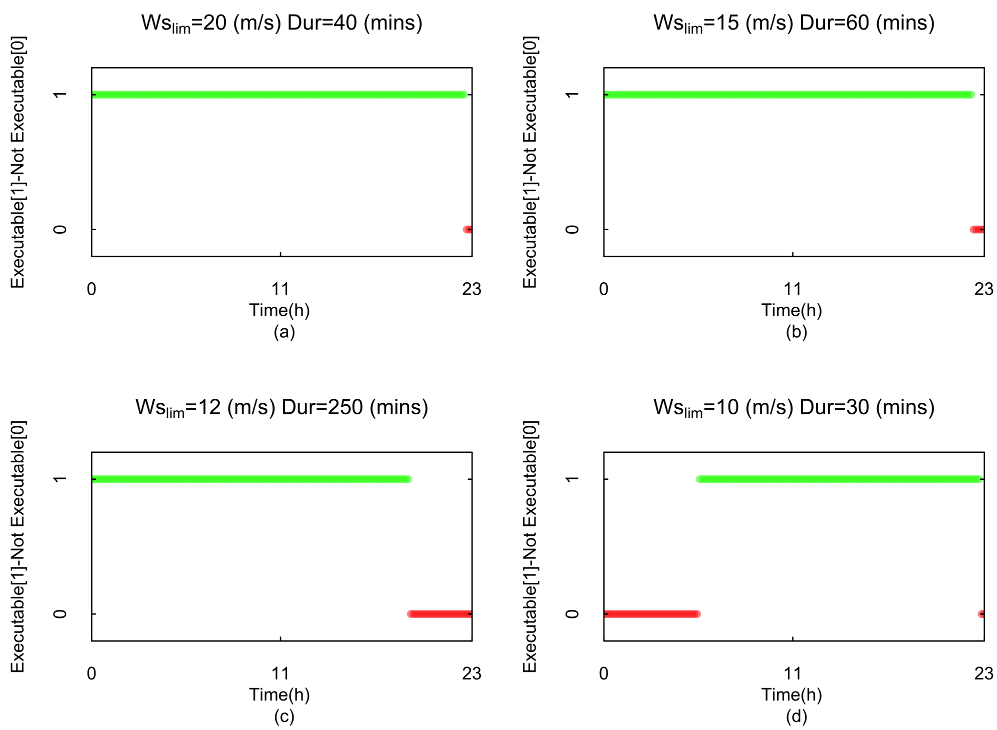


Figure 5.15: Routine maintenance evaluation with ECMWF forecasts for 7^{th} October 2017 (a) Cluster 1 tasks, (b) Cluster 2 tasks (c) Cluster 3 tasks, and (d) Cluster 4 tasks

Here the summary of the wind farm accessibility deviation in Figures 5.14 and 5.15 from Figures 5.9 and 5.10 due to NWP forecasts is given by clusters.

Cluster 1: Overall matching with the actual data is good for both test cases, (see Figures 5.14a and 5.15a for comparison to actual data evaluation).

Cluster 2: Overall matching with the actual data is fair for both test cases. Since the non-executable window in both tests cannot be caught toward the end of the day, (see Figures 5.14b and 5.15b for comparison to actual data evaluation).

Cluster 3: The matching with the actual data is bad for both tests. The ECMWF forecasts could not catch the majority of non-executable windows in this cluster, (see Figures 5.14c and 5.15c in comparison to actual wind speed evaluation). Moreover, these forecasts give the decision maker wrong information showing the non-executable windows as executable and vice versa.

Cluster 4: The matching with the actual data is better than in the Cluster 3 sub-case. However, it is not as good as the Cluster 1 and the Cluster 2 sub-cases. These forecasts provide misleading information especially after mid-day for the first test of this sub-case, (see Figure 5.14d). Regarding the second test day, there is an information loss in early hours of the day, (see Figure 5.15d).

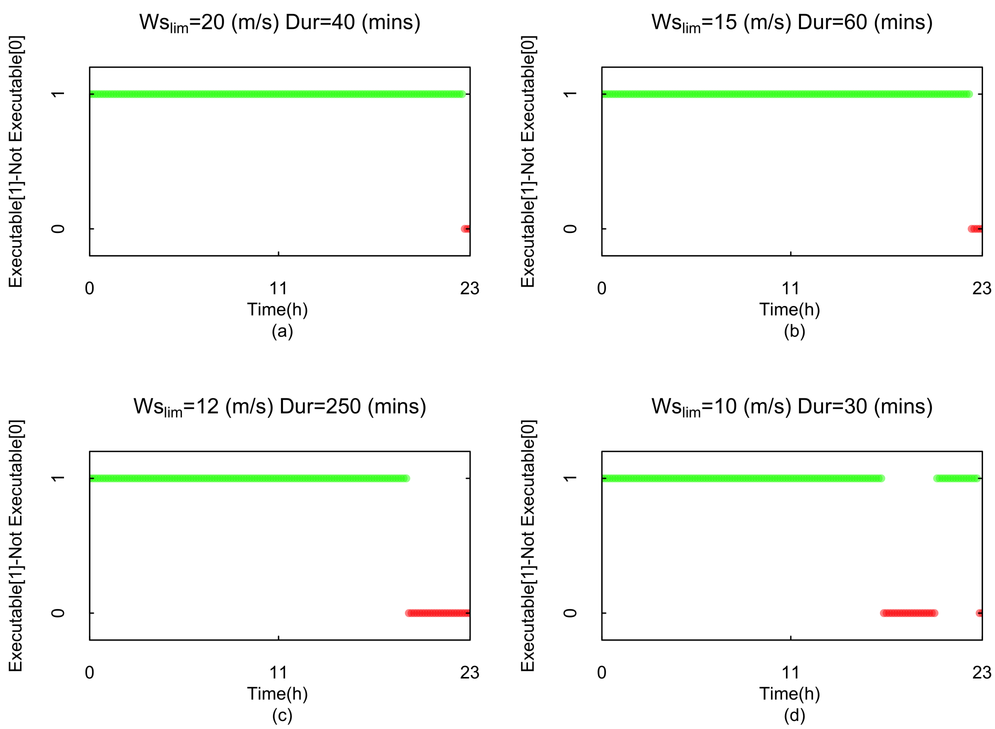


Figure 5.16: Routine maintenance evaluation with GFS forecasts for 24^{th} July 2017 (a) Cluster 1 tasks, (b) Cluster 2 tasks (c) Cluster 3 tasks, and (d) Cluster 4 tasks

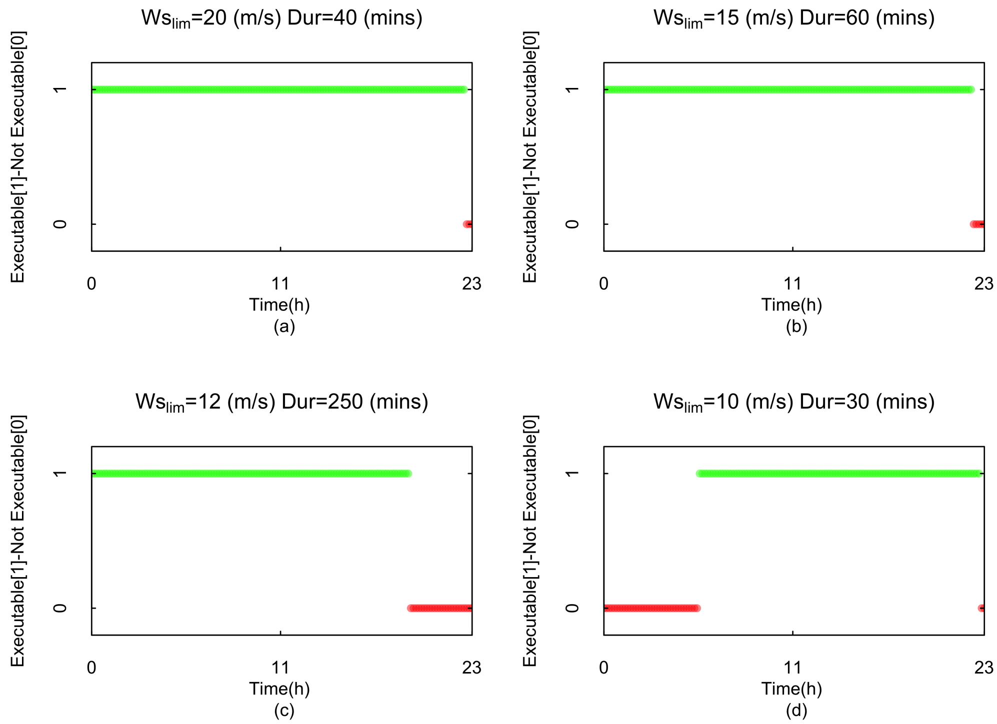


Figure 5.17: Routine maintenance evaluation with GFS forecasts for 7^{th} October 2017 (a) cluster 1 tasks, (b) cluster 2 tasks (c) cluster 3 tasks (d) cluster 4 tasks

In order to compare performance of two different NWP models, now, we perform the same analysis using the GFS forecasts for the same two test days.

Cluster 1: Over all matching with the actual data is good for both test cases, (see Figures 5.16a and 5.17a for comparison to actual data evaluation).

Cluster 2: Overall matching with the actual data is fair for both test cases. Because, towards the end of day non-executable window in the both tests can not be caught, (see Figures 5.16b and 5.17b for comparison to actual data evaluation).

Cluster 3: The matching with the actual data is bad for both tests. The GFS forecasts also could not catch the majority of non-executable windows in this cluster, (see Figures 5.16c and 5.17c for comparison to actual wind speed evaluation). The statement of the ECMWF case is valid also here.

Cluster 4: The matching with the actual data is better than the Cluster 3 sub-case. However, it is not good as the Cluster 1 and the Cluster 2 sub-cases. These forecasts provide more misleading information especially mid-day hours in for the first test of this sub-case, (see Figure 5.16d. Regarding the second test day, there is an information loss in early hours of the day, see Figure 5.17d).

Influence of the forecasting accuracy in the routine intervention: synthetic sets

In addition to actual data and the NWP forecasts, the decision support system (DSS) was also tested with synthetic forecasts with higher accuracy than the NWP forecasts to track the effect of the forecasting accuracy in the generated maintenance plans. In order to verify this effect, the downtime due to weather restrictions is analysed here. As previously highlighted, the tests were performed for the same two days as before. Three time series with different accuracy, represented by RMSE values 1 m/s, 1.4 m/s and 1.8 m/s were generated for each test day. They were labelled as synthetic 1, synthetic 2 and synthetic 3, respectively.

The time differences for the first and the last optimal solutions, according to the start time in both studied days are reported in Table 5.7 and Table 5.8. To interpret Tables 5.7 and 5.8, an example is given. In the second test day, the first optimal solution was reported with Cluster 1 tasks starting at 08:30, the Cluster 2 tasks at 09:10, the Cluster 3 tasks at 10:10 and Cluster 4 tasks at 14:20 by using actual measurements. Now, if the Synthetic 1 data are used instead of the actual data, the plan becomes starting Cluster 1 tasks at 08:10, Cluster 2 tasks at 08:50, Cluster 3 tasks at 09:50 and Cluster 4 tasks at 14:00.

This plan causes 20 minutes weather downtime between the execution of Cluster 2 and Cluster 3 tasks. According to Figure 5.10, if the team starts the maintenance at 08:10 they can execute the Cluster 1 tasks and the Cluster 2 tasks, however they must wait for the execution of the Cluster 3 tasks. When they complete the maintenance accordingly to the generated plan with the Synthetic data 1, the total duration is 400 minutes. Whereas, it ought to be 380 minutes in the optimal case according to the actual data. This difference is a planning downtime, which occurs due to inaccuracy in weather forecasts, of approximately 0.33h as reflected in Table 5.8. Therefore, columns referring to a time difference in Tables 5.7 and 5.8 show the deviation from the optimal plans.

Table 5.7: Sensitivity of the scheduling with synthetic data for the first test day

Input	**Time difference for the first optimal solution**	**Time difference for the last optimal solution**
Synthetic 1	0	0
Synthetic 2	0	0.33 h
Synthetic 3	1 h	day loss

Table 5.8: Sensitivity of the scheduling with synthetic data for the second test day

Input	**Time difference for the first optimal solution**	**Time difference for the last optimal solution**
Synthetic 1	0.33 h	0
Synthetic 2	0.33 h	0
Synthetic 3	0.66 h	0

5.4 Case 2: Generator replacement

24 hours evaluation for executable/not executable windows

In this case study, the generator replacement is investigated for the proposed scheduling process. To replace the generator, a crane must be used. Figure 5.18 illustrates the required lifting and unloading processes. Firstly, the nacelle cover must be removed and then the failed generator must be taken out. These removals are followed by installation of the new generator and re-installation of the original nacelle cover. In other words, this intervention requires two types of lifting /unloading tasks. Safety requirements with regards to wind speed vary due to the gust values. The mean wind speed limit for safe working has to be decreased by 2 m/s when the wind speed gust is above 5 m/s for operation requiring a crane usage (wind farm owner internal procedures: personal communication, 14^{th} December 2018), from 10 m/s, for a gust lower than 5 m/s, to 8 m/s for a gust higher than 5 m/s in the case of nacelle cover and from 8 m/s to 6 m/s, for the same gust values, in the case of the generator. It is worth mentioning here that the gust limit, to the authors knowledge, has never been considered in previous scientific studies.

Another difference, regarding routine maintenance plan, is the requirement to follow a fixed task order, as obviously, it would not be possible to perform removal of old generator before removing the nacelle cover. Therefore, the maintenance execution order is fixed for this problem.

It is rather easy to highlight the importance of the fine-tuning with a simple comparison. If this maintenance day is planned as a visit anticipated to start at 8 AM, planning downtime becomes 9 hours due to a long inaccessibility period for task 3, replacement of the new generator, after 12:50, see Figure 5.19c.

Figure 5.18: Photographs of a generator replacement in the wind farm investigated (copyright, CETASA).

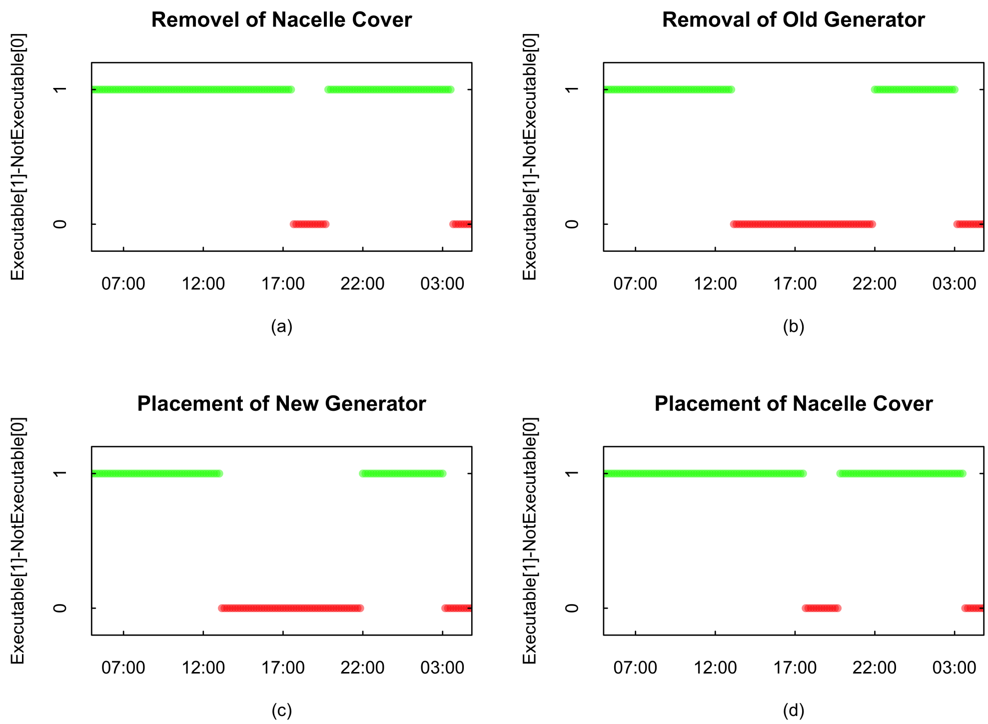


Figure 5.19: Evaluation of generator replacement when considering dynamic safe working limits due to gust. Executable (green)-Not executable (red) time windows during the day: 120 min search for generator 8-6 m/s mean wind speed limit due to gust value (a),(d); 90 min search for nacelle (b),(c) 10-8 m/s mean wind speed limit due to gust value.

The details of the optimal maintenance plans for a generator replacement, which require on average 7 hours accessibility, are given in Table 5.9.

Table 5.9: Optimal plan and task execution starting time estimated by using actual mean wind speed and gust data obtained for 24 hours from 5 AM 20th July 2017

	Removal nacelle cover	**Removal old generator**	**Placement new generator**	**Placement nacelle cover**
Earliest plan	05:00	06:30	08:30	10:30

Influence of the forecasting accuracy in the generator replacement

To test the effects of gust and mean wind speed accuracy together, similarly to the routine maintenance case, synthetic data sets were generated for both variables for 24 hours (from 05:00, 20^{th} July 2017).

The combinations of noise introduced on the mean wind speed and noise introduced on the gust inputs are tested together in the decision support tool.

Table 5.10: Syntetic data sets

Groups	Details
Synthetic 4	Gust RMSE 2.7 m/s and Mean wind speed RMSE 1.0 m/s
Synthetic 5	Gust RMSE 2.7 m/s and Mean wind speed RMSE 1.8 m/s
Synthetic 6	Gust RMSE 5.1 m/s and Mean wind speed RMSE 1.0 m/s
Synthetic 7	Gust RMSE 5.1 m/s and Mean wind speed RMSE 1.8 m/s

Table 5.11: Sensitivity of the scheduling

Input	Time difference for the first optimal solution
Synthetic 4	0
Synthetic 5	day loss
Synthetic 6	0
Synthetic 7	day loss

According to the reported results in Table 5.11 and considering the current test settings given in Table 5.10, the mean wind speed forecasting accuracy shows a higher influence on the proposed planning tool in comparison to the effect of the gust accuracy. This difference occurs because the mean wind speed is the primary parameter of a safe working rule, which concerns with the accessibility. The wind speed can directly influence the accessibility, whereas the wind gust has a secondary effect on it.

5.5 Chapter conclusions

This study presents maintenance intervention scheduling challenges and possible solutions for two different maintenance cases, routine and corrective. Safety restrictions for wind farm maintenance visits are studied in detail. It is found that in addition to wind speed, wind gust is also a limiting parameter for wind turbine accessibility and maintainability. A flexible decision support tool is proposed, which is capable of considering dynamic safe working rules and requires mean wind speed and gust forecasts. Typically, industrial practice neglects the planning of the execution order of the tasks and uses input forecasts with average accuracy performance. This can cause planning downtime from 9 hours (in corrective visit) to a day loss (in the first test case for routine maintenance). It is shown that obtaining an efficient maintenance plan with the available NWP forecasts could be quite challenging.

The complexity of the safety requirements due to the high number of task types was simplified by using agglomerative nesting in the routine case. Wind speed and wind gust forecasts were needed to test the decision support tool. As the aim of this study was not the generation of wind speed and wind gust forecasts, a range of accuracy levels were tested with synthetic data sets for the developed planning tool. It was observed that the accuracy of the input forecasts strongly determines the effectiveness of the planning tool. As the forecasting accuracy gets worse, the weather downtime increases as it is not possible to access to the turbine.

A possible extension of this study would be to schedule a maintenance plan for multiple wind turbines in a single visit by taking into account short term forecasts. Moreover, gust variable might be used as a more serious contributor in the decision support tool by including crane usage permissible wind speed limits. Then, the combination of dynamic safe access prerequisites for wind turbine and crane could be simulated together for a maintenance day by using both wind speed and wind gust forecasts.

Chapter 6

Conclusions

In this thesis, decision making processes in wind farm management were investigated. Practical and feasible DSSs have been developed prioritising data-driven approaches.

Firstly, exploratory analysis of key factors influencing in wind farm maintenance decisions is performed to understand the dependencies of hourly electricity market price, which is one of the decisive parameters for the wind farm revenue, and the connection between wind turbine failures and atmospheric variables in order to provide additional information to a maintenance team and a maintenance planner. In this regard, two frameworks are established. The first one serves as the information source for one year in advance maintenance planning and the second one provides information to the O&M team on the day of intervention. The drivers of the electricity market price are investigated for five EU countries, Spain, Germany, the UK, France and the Netherlands. It is found that the electricity demand can be considered especially for the estimation of the peak prices, but the demand times series do not provide a good reference especially for the estimation of bottom prices, what a maintenance engineer would like to know in order to schedule maintenance activities. Especially, the electricity generation time series in France from Nuclear and Fossil gas, in Spain and Germany from Onshore wind and Fossil hard coal, and in the Netherlands and the UK from Fossil gas can be tracked as information providers. Energy mixture policy of a country does not stand only for the varying amounts of the electricity generation from different sources. It covers also the electricity generation timing from different energy sources. Base load provider technologies such as Nuclear (the Netherlands, the UK, Germany, Spain) exert very weak or zero influence on the electricity market prices, but when Nuclear is considered as load following technology in the energy mixture, its influence on the electricity market price increases severely. Peak load following generators such as Fossil gas exert very high influence on electricity market prices, even in some cases its influence is stronger than the electricity demand (Spain, France, Germany). When Onshore wind is considered as peak following technology in the energy mixture (Spain and Germany), its influence increases in comparison to mid-merit load consideration (the UK, the Netherlands, France). During April, May, June and November the case countries show the bottom prices. In these months a scheduled component replacement or a maintenance visit can be considered. As future work of this line, detected monthly key driving factors for electricity market price can be used for monthly multivariate electricity market price models, which are capable of providing good electricity price forecasts two or three weeks ahead.

Then in this study, associations between wind turbine failure observations and various environmental conditions are discovered in order to provide information to the O&M team at the day of the intervention. The resulting learning framework is capable of providing a summary for the whole observation period and the information prior to failure occurrence at interested time steps. Therefore, it can be considered as an enhancement tool for offline inspection data collection. The proposed framework is designed as an ensemble of supervised labelling, unsupervised clustering techniques and apriori rule mining algorithm. For prior to failure analysis 1-D clustering and supervised labelling are suggested considering the interpretation of association rules, which behave as a decision support system for a maintenance team to evaluate the health status of the components other than the subject component of the initially planned maintenance. The supervised labelling technique requires expert opinion and labours classification for each variable. 1-D clustering technique can provide information fast and without requirement of an expert guidance. Although, Multi-D clustering is weak to provide information at specific time steps, this technique is useful to make annual maintenance plans since it provides a summary for the overall downtime occurrences. As future work of this line, usage and practicability of the obtained association rules can be measured and evaluated by O&M engineers. These rules can be enriched with the feedbacks of O&M engineers.

Next, what-if analysis for revenue tracking of maintenance decisions is performed in order to generate a DSS for the evaluation of the major maintenance decisions taken in wind farms. To this purpose, the impact of country dynamics and subsidy frameworks considering the electricity market conditions are modelled for three case countries Spain, the UK and the Netherlands. The impact of the intervention timing is analysed and the sensitivity of financial losses to environmental causes of under performance are estimated. The financial consequences of underperformance are shown in the NPV of the cash flow. 43 scenarios in total are studied for univariate sensitivity analysis. Additionally, 180 scenarios are evaluated for multivariate sensitivity analysis. There is an optimisation potential in terms of the timing of any maintenance that results in temporary underperformance. The analysis of shifting the maintenance period through the year shows that the actual timing was optimised to the seasonal wind resource trends. The financial results indicate that the optimal timing will change due to the different seasonality of electricity markets. However, this is affected by the complexity of the cash flow, the electricity market in each country, taxes and subsidies. In most configurations, a shifting to earlier spring appears to be more profitable. The NPV sensitivity study with icing and wind directional performance variation demonstrated that wind direction is more important in comparison to icing scenarios studied. The comparison for the three countries highlights that based on the electricity markets alone, the UK (approximately 150700 EUR) and the Netherlands (approximately 104100 EUR) were more attractive for the wind farm studied due to higher NPV values than Spain (approximately 102500 EUR). If the subsidies are included, the Spanish baseline is far more advantageous than the Netherlands and the UK. However, it should be noted that the attractive Spanish subsidy scheme ended and new wind farms may rely only on electricity market sales. In the investigation of multivariate sensitivity scenarios, the UK shows a fixed ranking for maintenance months being independent from financial policies. The Netherlands display small differences for taxed revenue scenarios with respect to those without subsidy and without tax case. Whereas, the Spanish case shows significant maintenance month ranking differences for the most preferable and the least preferable months when different financial set-ups (tax and subsidy) are considered. As an extension of O&M scenario analysis in wind farms, replacement analysis considering remaining useful lifetime can be performed. The sensitivity study can be enriched by considering more variables with multi variate modifications in the scenarios.

Finally, the generation of a decision support tool for planning of a maintenance day is studied to provide a useful maintenance DSS for in situ applications. The safe working rules considering the wind speed constraints for the accessibility to the wind turbine are reviewed taking into account the turbine manufacturer's O&M guidelines. The characteristics of the maintenance visits are summarised. Wind turbine accessibility trials using numerical weather prediction forecasting techniques for wind speed variable are presented. An intervention decision pool considering safe working rules is generated, containing a list of plans capable of providing the optimal sequence of various tasks and ranked for revenue prioritasied timing. Maintenance intervention scheduling challenges and possible solutions for two different maintenance cases (routine and corrective) are investigated. Safety restrictions for wind farm maintenance visits are studied in detail. It is found that in addition to wind speed, wind gust is also a limiting parameter for wind turbine accessibility and maintainability. A flexible decision support tool is proposed, which is capable of considering dynamic safe working rules and requires mean wind speed and gust forecasts. Typically, industrial practice neglects the planning of the execution order of the tasks and uses input forecasts with average accuracy performance. This can cause planning downtime from 9 hours (in corrective visit) to a day loss (in the first test case for routine maintenance). It is shown that obtaining an efficient maintenance plan with the available NWP forecasts could be quite challenging. The complexity of the safety requirements due to the high number of task types was simplified by using agglomerative nesting in the routine case. Wind speed and wind gust forecasts were needed to test the decision support tool. As the aim of this study was not the generation of wind speed and wind gust forecasts, a range of accuracy levels were tested with synthetic data sets for the developed planning tool. It was observed that the accuracy of the input forecasts strongly determines the effectiveness of the planning tool. As the forecasting accuracy gets worse, the weather downtime increases as it is not possible to access to the turbine. A possible extension of this research line would be to schedule a maintenance plan for multiple wind turbines in a single visit by taking into account short term forecasts. Moreover, gust variable might be used as a more serious contributor in the decision support tool by including crane usage permissible wind speed limits. Then, the combination of dynamic safe access pre-requisites for wind turbine and crane could be simulated together for a maintenance day by using both wind speed and wind gust forecasts.

As an overall conclusion of this thesis, the effects of environmental conditions on failure occurrences and wind farm accessibility need to be considered for in situ decision support systems. Local features of electricity market price are important drivers in maintenance optimisation. Timely decision making can decrease the revenue losses due to coarse scheduling practices and weather downtimes.

Chapter 7

Conclusiones Generales de la Tesis

En esta tesis, se ha investigado sobre procesos de toma de decisión en la gestión de parques eólicos. Se han desarrollado sistemas de apoyo a la toma de decisión prácticos y factibles priorizando las aproximaciones basadas en datos. Inicialmente, se ha utilizado un análisis exploratorio de datos para estudiar los principales factores que influyen en la toma de decisiones para el mantenimiento de parques eólicos, facilitando la comprensión de las dependencias del precio horario del mercado de la electricidad, uno de los factores decisivos en los beneficios de parques eólicos, y la conexión entre los fallos de los aerogeneradores y las variables atmosféricas con el objetivo de proporcional información adicional a los equipos de trabajo y al gestor del mantenimiento. De esta forma, se han desarrollado dos entornos de trabajo. El primero de ellos sirve como fuente de información para la planificación de mantenimiento a un año vista y el segundo proporciona información útil al equipo de operación y mantenimiento en el día de la intervención. Se han investigado los responsables de los precios del mercado de la electricidad para cinco países de la Unión Europea, España, Alemania, Reino Unido, Francia y Holanda. Se ha encontrado que la demanda de la electricidad debe tenerse en cuenta de forma particular para la estimación de los precios pico, pero la serie temporal de la demanda no proporciona una buena referencia para la estimación de los precios valle, dato importante a conocer por un ingeniero cuando programa actividades de mantenimiento. En particular, las series temporales de generación nuclear y de gas en Francia, eólica terrestre y carbón en España y Alemania, y de gas en Holanda y Reino Unido, pueden considerarse como proveedoras de información. La política de mix energético de un país no solo corresponde a las cantidades variables obtenidas de distintas fuentes, sino que también tiene en cuenta la disponibilidad temporal de las diversas fuentes energéticas que lo componen. Las tecnologías que sustentan la demanda básica como la nuclear (Holanda, Reino Unido, Alemania, España) ejercen una influencia muy débil e incluso nula en los precios del mercado de la electricidad, pero cuando la generación nuclear se considera como una tecnología de seguimiento de demanda en el mix energético, su influencia en el precio de la electricidad se incrementa de forma importante. Los generadores que cubren los picos de demanda, tal como las plantas de gas natura, ejercen una influencia muy alta en el precio de la electricidad, siendo su influencia incluso más alta que la de la demanda en algunos casos (España, Francia y Alemania). Si consideramos la energía eólica terrestre como tecnología para cubrir picos de demanda en el mix energético (España y Alemania), su influencia aumenta respecto a considerarla como generación que cubre la carga media (Reino Unido, Holanda y Francia). Durante abril, mayo, junio y noviembre, todos los países estudiados mostraron los precios de la electricidad más bajos. En dichos meses, se puede considerar sin problemas reemplazar un componente o realizar una visita de mantenimiento. Como trabajo futuro de esta línea, se pueden utilizar los resultados relativos a los principales factores que modulan el precio de la electricidad para desarrollar modelos multivariable capaces de proporcionar buenas predicciones del precio de la electricidad con un horizonte de dos o tres semanas.

Posteriormente, en este estudio se ha estudiado la relación existente entre los fallos observados en aerogeneradores y las condiciones ambientales con el fin de proporcionar información al equipo de operación y mantenimiento el día de la intervención. El entorno desarrollado es capaz de proporcionar un resumen para el intervalo total de observación, así como información previa al suceso del fallo en intervalos temporales de interés. Puede por tanto considerarse como una herramienta de mejora para la recogida de datos de inspección off-line. El entorno de trabajo propuesto se ha diseñado como un conjunto de algoritmos de clasificación y agrupamiento, supervisado y no supervisado (unidimensional y multidimensional), así como de un algoritmo de descubrimiento de reglas que asocian los fallos de las turbinas con las variables ambientales. Para el análisis previo al fallo, el sistema utiliza agrupamiento y clasificación supervisada unidimensional junto con el algoritmo de descubrimiento de reglas de asociación. Esto permite al equipo de mantenimiento evaluar, y corregir si es necesario, el estado de salud del resto de componentes de la turbina y no solo el de aquel componente para el que se había planificado la intervención. La técnica de clasificación supervisada requiere intervención técnica, así como un laborioso trabajo para realizar la clasificación de cada variable. El agrupamiento unidimensional no supervisado, por su parte, puede proporcionar información de forma rápida y sin requerir intervención externa. Aunque el agrupamiento multidimensional no es lo suficientemente robusto para proporcionar información en intervalos de tiempo específicos, si que es útil para realizar planes de mantenimiento anuales puesto que proporciona un resumen de las paradas de las turbinas. Como trabajo futuro de esta línea, sería interesante que ingenieros de operación y mantenimiento evaluaran de forma práctica la usabilidad de las reglas que asocian los fallos con las variables meteorológicas obtenidas en este trabajo. Estas reglas podrían enriquecerse con la opinión de dichos expertos.

A continuación, se ha realizado un análisis tipo "¿qué pasa si?" para el seguimiento del beneficio de las decisiones de mantenimiento con el objetivo de desarrollar un entorno de trabajo para la evaluación de decisiones de grandes mantenimientos a considerar en un parque eólico. Para ello, se consideran el impacto de la dinámica del país y las condiciones de subsidios existentes en el precio de la electricidad en tres países, España, Reino Unido y Holanda. Se ha analizado el impacto de la temporización de la intervención y se ha estimado la sensibilidad de las pérdidas financieras rente a causas de baja producción por motivos ambientales. Las consecuencias financieras de la baja producción se muestran en el valor actual neto, VAN, del flujo de caja. Se utilizaron un total de 43 escenarios para el análisis de sensibilidad monovariable. Adicionalmente se evaluaron también 180 escenarios para un análisis de sensibilidad multivariable. Se ha encontrado una posibilidad de optimización en términos de la temporización de cualquier operación de mantenimiento que suponga una pérdida de rendimiento. El análisis de desplazar el periodo de la intervención de mantenimiento durante todo el año demostró que la temporización utilizada estaba optimizada respecto a las tendencias del recurso eólico estacional. Los resultados financieros indican que la temporización óptima cambiará debido a la diferente estacionalidad de los mercados eléctricos. En cualquier caso, estos resultados están influenciados por la complejidad del flujo de caja, el mercado de la electricidad de cada país, los impuestos y los subsidios. En muchas configuraciones, un desplazamiento de la intervención a principios de la primavera parece ser más beneficioso. El estudio de la sensibilidad del VAN respecto a heladas y dirección del viento demostró que la dirección de viento es más importante que las heladas en los escenarios utilizados. Comparando los tres países estudiados únicamente desde el punto de vista del mercado eléctrico, se observa que el Reino Unido (aproximadamente 150700 EUR) y Holanda (aproximadamente 104100 EUR) eran más atractivos para el parque eólico estudiado debido a los mayores valores del VAN que en España (aproximadamente 102500 EUR). Si se incluyen los subsidios, la línea base de España es bastante más ventajosa que las de Holanda y la del Reino Unido. Sin embargo, debe notarse que el atractivo esquema de subsidios en España finalizó y que los parques nuevos deben basarse únicamente en el precio del mercado eléctrico. En la investigación de los escenarios de sensibilidad multivariable, el Reino Unido muestra el mismo orden de prioridad para los meses en los que se realiza el mantenimiento, independientemente de las políticas financieras. Holanda, por su parte, muestra pequeñas diferencias para los escenarios de beneficios después de impuestos con respecto aquéllos con subsidios o sin impuestos. Por último, el caso de España muestra diferencias significativas en el orden de prioridad para los meses en los que se ejecuta la intervención, fundamentalmente el más favorable y el más desfavorable, cuando se consideran las condiciones financieras relativas a impuestos y subsidios. Como trabajo futuro de esta línea, podría considerarse una extensión del análisis de escenarios de operación y mantenimiento considerando la vida útil estimada para la turbina. El análisis de sensibilidad se podría enriquecer asimismo considerando más variables en los distintos escenarios.

Para finalizar, se ha estudiado y desarrollado un sistema de apoyo a la toma de decisiones para la planificación de un día de mantenimiento. Se consideran las normas de seguridad en el trabajo teniendo en cuenta los requisitos de operación y mantenimiento del fabricante de la turbina. Se han presentado y descrito las características de las visitas de mantenimiento. Se han obtenido ensayos de accesibilidad a las turbinas utilizando técnicas de predicción numérica para la velocidad de viento. Finalmente, se genera un conjunto de decisiones de intervención, considerando las normas de seguridad en el trabajo, que contiene una lista de planes de trabajo, proporcionando la secuencia óptima de las tareas a realizar en la turbina, priorizadas y ordenadas temporalmente en función del beneficio obtenido. Se han investigado los retos que hay que acometer en la planificación de las tareas de mantenimiento para dos casos diferentes (rutina y correctivo) considerando en detalle las restricciones de seguridad a tener en cuenta en las intervenciones. Se ha encontrado que, además de la velocidad de viento, el valor de la ráfaga de viento es un parámetro limitante para la accesibilidad a la turbina y su mantenibilidad. Se propone una herramienta flexible de apoyo a la toma de decisiones, capaz de considerar de forma dinámica las normas de seguridad en el trabajo, que requiere las predicciones de velocidad media de viento y de ráfaga de viento. La práctica en la industria no considera la planificación del orden de las tareas a ejecutar y además utiliza predicciones meteorológicas de precisión promedio. Esto puede ocasionar tiempos de parada desde 9 horas (en visitas correctivas) hasta la pérdida de un día completo. Se ha demostrado que obetner un plan de mantenimiento eficiente con las predicciones meteorológicas disponibles puede ser complicado. La complejidad de los requisitos de seguridad, debida al alto número de tareas, se ha simplificado utilizando técnicas de agrupamiento anidado en el caso de la visita rutinaria. Para verificar la herramienta desarrollada, eran necesarias las predicciones de velocidad y ráfagas de viento. Sin embargo, como el objetivo del estudio no era la generación de predicciones meteorológicas, se probaron un rango de niveles de precisión en la predicción utilizando datos sintéticos para la herramienta de planificación. Se observó que la precisión de los datos de entrada determina enormemente la efectividad de la herramienta. Conforme la precisión de las predicciones empeora, el tiempo de parada debido a la meteorología aumenta puesto que no es posible acceder a la turbina. Una posible extensión de esta línea de investigación sería generar un plan de mantenimiento para múltiples turbinas en una sola visita teniendo en cuenta predicciones de corto plazo. Además, el valor de ráfaga de viento podría usarse como un condicionante serio en la herramienta de apoyo a las decisiones incluyendo los límites de velocidad de viento permitidos para el uso de grúas. De esta forma, la combinación de los pre-requisitos dinámicos de seguridad para el acceso a la turbina junto con los de utilización de grúa podrían utilizarse para planificar un día de mantenimiento utilizando tanto la velocidad de viento como la ráfaga. Como conclusión general de esta tesis, los efectos de las condiciones ambientales en las ocurrencias de fallo de las turbinas y en la accesibilidad a los parques eólicos deben considerarse en los sistemas de ayuda a la toma de decisiones de operación y mantenimiento. Los fenómenos locales que afectan al precio de la electricidad son factores importantes que deben considerare en la optimización del mantenimiento. Finalmente, la toma de decisiones de forma oportuna puede disminuir las pérdidas de beneficios debidas a planificaciones de mantenimiento toscas y a paradas por condiciones ambientales.

Appendix 1

Table A1: Executed tasks for the scheduled visit

	A-Ground	
Num.	**Sub System**	**Task**
1	Tower	Check the status of welding joints at the entrance, flanges and holes
2	Tower	Check the status of painting and galvanizing
3	Electrical parts	Cleaning of the ventilation filters
4	Rotor-Blades	Visual inspection of the condition of the blades
	B-Platform	
Num.	**Sub System**	**Task**
5	Electrical parts	Wiring defects check
6	Electrical parts	Cleaning of the ventilation filters
7	Electrical parts	Tower lightining
	C- Tower	
Num.	**Sub System**	**Task**
8	Yaw System	Check the adjustment of the internal gear of the crown gear
9	Yaw System	Lubrication for the internal gear
10	Yaw System	Check the level of oil and the existence of leaks in geared motors
11	Yaw System	Inspection for cracked or worn teeth
12	Yaw System	Lubrication for the crown gear
13	Yaw System	Lubrication for the aligning ring bearings
14	Yaw System	Lubrication for the lip seals
	D- Nacelle	
Num.	**Sub System**	**Task**
15	Main Shaft and Bearings	Check the status of the main bearing joints
16	Main Shaft and Bearings	Lubrication for the main bearing
17	Main Shaft and Bearings	Lubrication for the lip seals
18	Gearbox	Check the level of oil and the existence of leaks
19	Gearbox	Check the existence of leaks for the sleeves and the fittings.

20	Gearbox	Check the pitting or other defects in internal bearings
21	Gearbox	Check the increase in noise levels
22	Gearbox	Change the oil filter
23	Gearbox	Take the oil sample
24	Gearbox	Check the oil pump leakages
25	Gearbox	Tighten the bolts on the gearbox
26	Gearbox	Check the condition of the silent blocks
27	Gearbox	Check the status of painting and galvanizing
28	Generator	Generator alignment
29	Base structure and cover	Check the status of painting and galvanizing
30	Base structure and cover	Check the status of the hinges and the locks
31	Base structure and cover	Check the tightness of the cover
32	Electrical parts	Replace the filters in the fan
	E- Hub	
Num.	**Sub System**	**Task**
33	Rotor-Blades	Check the damage level for the metal parts of the hub connections. (rotary union)
34	Rotor	Re-tighten of the hub blades bolts
	F- Outside of Nacelle	
Num.	**Sub System**	**Task**
35	Sensors	Solve the mechanical problems in the wind vanes, such as loose bolts
36	Sensors	Solve the mechanical problems in the anemometers, such as loose bolts

Bibliography

[1] R. C. Thomson, G. P. Harrison, Life cycle costs and carbon emissions of offshore wind power, Tech. rep., accessed 01.05.2019 (2015).
URL https://www.climatexchange.org.uk/media/1461/main_report__life_cycle_costs_and_carbon_emissions_of_offshore_wind_power.pdf

[2] Lazard's levelized cost of energy analysis, Tech. rep., LAZARD, accessed 01.05.2019 (2017).
URL https://www.lazard.com/media/450337/lazard-levelized-cost-of-energy-version-110.pdf

[3] Energy prices and costs in europe 2018, Tech. rep., European Commission, accessed 01.05.2019 (2019).
URL https://ec.europa.eu/energy/en/data-analysis/energy-prices-and-costs

[4] Renewable energy and jobs, Tech. rep., IRENA, accessed 01.05.2019 (2018).
URL https://irena.org/-/media/Files/IRENA/Agency/Publication/2018/May/IRENA_RE_Jobs_Annual_Review_2018.pdf

[5] Workers wanted: The eu wind energy sector skills gap, Tech. rep., European Wind Energy Technology Platfrom, accessed 01.05.2019 (2013).
URL https://etipwind.eu/files/reports/TPWind-Workers-Wanted.pdf

[6] Renewable power generation costs in 2017, Tech. rep., International Renewable Energy Agency, accessed 02.02.2018 (2018).

URL https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2018/Jan/IRENA_2017_Power_Costs_2018.pdf

[7] IEA, Task 33: Reliability Data: Standardizing Data Collection for Wind Turbine Reliability and Operation & Maintenance Analyses, accessed 10.07.2019, (2019).
URL https://community.ieawind.org/tasks/new-item/task33.

[8] IEA task 33, Data collection and reliability assessment for o&m optimization of wind turbines, wind Europe Summit, Hamburg (2016).

[9] A. Ioannou, A. Angus, F. Brennan, A lifecycle techno-economic model of offshore wind energy for different entry and exit instances, Applied Energy 221 (2018) 406–424.

[10] R. Wiser, K. Jenni, J. Seel, E. Baker, M. Hand, E. Lantz, A. Smith, Forecasting wind energy costs and cost drivers: The views of the world's leading experts, accessed 09.05.2018 (2016).
URL https://emp.lbl.gov/sites/all/files/lbnl-1005717.pdf

[11] A. Erguido, A. Crespo Márquez, E. Castellano, J. F. Gómez Fernández, A dynamic opportunistic maintenance model to maximize energy-based availability while reducing the life cycle cost of wind farms, Renewable Energy 114 (2017) 843–856.

[12] S. Zhong, A. A. Pantelous, M. Beer, J. Zhou, Constrained non-linear multi-objective optimisation of preventive maintenance scheduling for offshore wind farms, Mechanical Systems and Signal Processing 104 (2018) 347–369.

[13] J. Kaldellis, T. J. Gavras, The economic viability of commercial wind plants in greece a complete sensitivity analysis, Energy Policy 28 (8) (2000) 509–517.

[14] C. Kiranoudis, N. Voros, Z. Maroulis, Short-cut design of wind farms, Energy Policy 29 (7) (2001) 567–578.

[15] Gwec.net, Global wind 2006 report, accessed 09.05.2018 (2006).
URL http://gwec.net/wp-content/uploads/2012/06/gwec-2006_final_01.pdf

[16] Gwec.net, Global wind statistics 2017, accessed 09.05.2018 (2018). URL http://gwec.net/wp-content/uploads/vip/GWEC_PRstats2017_EN-003_FINAL.pdf

[17] M. Hofmann, A Review of Decision Support Models for Offshore Wind Farms with an Emphasis on Operation and Maintenance Strategies, Wind Engineering 35 (1) (2011) 1–16.

[18] F. Besnard, M. Patrikssont, A.-B. Strombergt, A. Wojciechowskit, L. Bertling, An optimization framework for opportunistic maintenance of offshore wind power system, in: PowerTech, 2009 IEEE Bucharest, IEEE, 2009, pp. 1–7.

[19] F. Ding, Z. Tian, Opportunistic maintenance for wind farms considering multi-level imperfect maintenance thresholds, Renewable Energy 45 (2012) 175–182.

[20] M. Yildirim, N. Z. Gebraeel, X. A. Sun, Integrated predictive analytics and optimization for opportunistic maintenance and operations in wind farms, IEEE Transactions on Power Systems 32 (6) (2017) 4319–4328.

[21] B. R. Sarker, T. I. Faiz, Minimizing maintenance cost for offshore wind turbines following multi-level opportunistic preventive strategy, Renewable Energy 85 (2016) 104–113.

[22] K. Atashgar, H. Abdollahzadeh, Reliability optimization of wind farms considering redundancy and opportunistic maintenance strategy, Energy Conversion and Management 112 (2016) 445–458.

[23] C. Zhang, W. Gao, S. Guo, Y. Li, T. Yang, Opportunistic maintenance for wind turbines considering imperfect, reliability-based maintenance, Renewable energy 103 (2017) 606–612.

[24] J. M. Leigh, S. J. Dunnett, Use of petri nets to model the maintenance of wind turbines, Quality and Reliability Engineering International 32 (1) (2016) 167–180.

[25] D. Mcmillan, G. W. Ault, Quantification of condition monitoring benefit for offshore wind turbines, Wind Engineering 31 (4) (2007) 267–285.

[26] I. E. M. García, A. S. Sánchez, S. Barbati, Reliability and preventive maintenance, in: MARE-WINT, Springer, 2016, pp. 235–272.

[27] BCS, Incorporated, Establishing an in-house wind maintenance program a case study of los angeles department of water and power, Tech. rep., American Public Power Association, accessed 05.06.2017 (2008).
URL `http://www.appanet.org/files/PDFs/LADWPInHouseWindMaintenanceP rog10-15-08FINAL.pdf`

[28] T. Price, J. Bunn, D. Probert, R. Hales, Wind-energy harnessing: global, national and local considerations, Applied energy 54 (2) (1996) 103–179.

[29] A. Kovács, G. Erdös, L. Monostori, Z. J. Viharos, Scheduling the maintenance of wind farms for minimizing production loss, IFAC Proceedings Volumes 44 (1) (2011) 14802–14807.

[30] A. Froger, M. Gendreau, J. E. Mendoza, E. Pinson, L.-M. Rousseau, A branch-and-check approach for a wind turbine maintenance scheduling problem, Computers & Operations Research 88 (2017) 117–136.

[31] I. Suomi, T. Vihma, Wind gust measurement techniques—from traditional anemometry to new possibilities, Sensors 18 (4) (2018) 1300.

[32] Influence of wind on crane operation, Tech. rep., Liebherr, accessed 09.07.2018 (2017).
URL `https://www.liebherr.com/shared/media/mobile-and-crawler-cranes/brochures/wind-influences/liebherr-influence-of-wind-p403-e04-2017.pdf`

[33] A. Froger, M. Gendreau, J. E. Mendoza, E. Pinson, L.-M. Rousseau, Solving a wind turbine maintenance scheduling problem, Journal of Scheduling (2016) 1–24.

[34] J. Browell, I. Dinwoodie, D. McMillan, Forecasting for day-ahead offshore maintenance scheduling under uncertainty, in: Risk, Reliability and Safety: Innovating Theory and Practice, CRC Press., 2016, pp. 1337–1144.

[35] G. Pulcini, Mechanical reliability and maintenance models (2003).

[36] C. S. Park, Contemporary engineering economics, Pearson Prentice Hall, New Jersey, 2007.

[37] J. Berk, P. DeMarzo, J. Harford, G. Ford, V. Mollica, N. Finch, Fundamentals of corporate finance, Pearson Higher Education, Boston, 2013.

[38] H. E. Kierulff, The replacement decision: Getting it right, Business Horizons 50 (3) (2007) 231 – 237.

[39] N. Y. Yürüşen, J. Tautz-Weinert, S. J. Watson, J. J. Melero, The financial benefits of various catastrophic failure prevention strategies in a wind farm: Two market studies (uk-spain), Journal of Physics: Conference Series 926 (1) (2017) 012014.

[40] S. Carlos, A. Sánchez, S. Martorell, I. Marton, Onshore wind farms maintenance optimization using a stochastic model, Mathematical and Computer Modelling 57 (7) (2013) 1884–1890.

[41] A . Sanchez, S . Martorell, M . Villamizar, I . Marton and S . Carlos, Multiobjective optimization of preventive maintenance applied to onshore wind turbines, in: Safety, Reliability and Risk Analysis Beyond the Horizon, CRC Press, 2013, pp. 775–781.

[42] I. Sagarna, J. Uribetxebarria, E. Castellano, A. Erguido, After-sales maintenance service strategies optimization. an offshore wind farm case study, IFAC-PapersOnLine 49 (28) (2016) 156–161.

[43] A. Karyotakis, On the optimisation of operation and maintenance strategies for offshore wind farms, Ph.D. thesis, UCL (University College London) (2011).

[44] I. A. Dinwoodie, D. McMillan, Operational strategies for offshore wind turbines to mitigate failure rate uncertainty on operational costs and revenue, IET Renewable Power Generation 8 (4) (2014) 359–366.

[45] H. Abdollahzadeh, K. Atashgar, M. Abbasi, Multi-objective opportunistic maintenance optimization of a wind farm considering limited number of maintenance groups, Renewable Energy 88 (2016) 247–261.

[46] U. R. Bharadwaj, V. V. Silberschmidt, J. B. Wintle, Risk based optimisation of spares inventory management., Advances in Production Engineering & Management 6 (3) (2011).

[47] Yildirim, Nurseda & Reder, Maik & Melero, Julio., Failure event definitions & their effects on survival and risk analysis of wind turbines., in: Safety and Reliability. Theory and Applications, ESREL, 2017.

[48] M. Pandey, Y. Liu, M. J. Zuo, Selective maintenance for complex systems considering imperfect maintenance efficiency, Reliability Modeling With Applications: Essays In Honor Of Professor Toshio Nakagawa On His 70th Birthday (2013) 17.

[49] de Toledo, Maria Luíza Guerra and Freitas, Marta A and Colosimo, Enrico A and Gilardoni, Gustavo L, Ara and ari imperfect repair models: Estimation, goodness-of-fit and reliability prediction, Reliability Engineering & System Safety 140 (2015) 107–115.

[50] Danish Wind Industry Assosiaction, The wind energy pioneers: The gedser wind turbine, accessed 19.02.2018, (2003).
URL http://xn–drmstrre-64ad.dk/wp-content/wind/miller/windpower%20web/en/pictures/juul.htm.

[51] J. S. Nielsen, J. D. Sørensen, Methods for risk-based planning of o&m of wind turbines, Energies 7 (10) (2014) 6645–6664.

[52] B. Kerres, K. Fischer, R. Madlener, Economic evaluation of maintenance strategies for wind turbines: a stochastic analysis, IET Renewable Power Generation 9 (7) (2015) 766–774.

[53] Gwec.net, Wind in numbers - gwec, accessed 14.08.2015, (2015).
URL http://www.gwec.net/global-figures/wind-in-numbers/.

[54] O. Duran, Throughput-centered physical assets priorization technique, in: Safety and Reliability,Theory and Applications, CRC Press, 2017.

[55] Z. Y. Zhang, Scheduling and routing optimization of maintenance fleet for offshore wind farms using duo-aco, in: Advanced Materials Research, Vol. 1039, Trans Tech Publ, 2014, pp. 294–301.

[56] M. Stålhane, L. M. Hvattum, V. Skaar, Optimization of routing and scheduling of vessels to perform maintenance at offshore wind farms, Energy Procedia 80 (2015) 92–99.

[57] A. Kovacs, G. Erdős, Z. J. Viharos, L. Monostori, A system for the detailed scheduling of wind farm maintenance, CIRP Annals-Manufacturing Technology 60 (1) (2011) 497–501.

[58] J.-M. Roldan-Fernandez, M. Burgos-Payan, J.-M. Riquelme-Santos, A.-L. Trigo-Garcia, The merit-order effect of energy efficiency, Energy Procedia 106 (2016) 175–184.

[59] APX Group, EPEX Netherlands, accessed 22.11.2017 (2017).
URL https://www.apxgroup.com/trading-clearing/apx-power-nl/

[60] EC.EUROPA EU, Progress towards 2020 targets, accessed 14.08.2018, (2016).
URL= http://ec.europa.eu/europe2020/europe-2020-in
-your-country/ espana/progress-towards-2020-targets/
index_en.htm .

[61] La Asociacion Empresarial Eolica, The figures of the spanish wind power industry, accessed 05.07.2018 (2017).
URL https://www.aeeolica.org/uploads/pictures/infografia
-windpower-EN.jpg

[62] L. A. E. Eolica, Installed power, accessed 22.11.2017 (2016).
URL http://www.aeeolica.org/en/about-wind-energy/wind-energy
-in-spain/installed-power/

[63] International Energy Agency, Spain statistics, accessed 15.11.2017 (2016).
URL http://www.iea.org/policiesandmeasures/
renewableenergy/?country=Spain

[64] La Asociacion Empresarial Eolica, Eólica 15. toda la información del año 2014 que necesitas conocer sobre el sector, accessed 14.08.2016 (2016).
URL http://www.aeeolica.org/es/aee-divulga/publicaciones/

[65] European Commission, Guidelines on environmental and energy State aid, accessed 22.11.2017 (2015).
URL http://eur-lex.europa.eu/legal-content/EN/LSU/?uri=
CELEX:52014XC0628(01)

[66] Organisation for Economic Co-operation and Development / International Energy Agency, IEA/IRENA global renewable energy joint policies and measures database, accessed 22.11.2017 (2017).
URL https://www.iea.org/policiesandmeasures/renewableenergy/

[67] J. Schallenberg-Rodriguez, R. Haas, Fixed feed-in tariff versus premium: A review of the current Spanish system, Renewable and Sustainable Energy Reviews 16 (1) (2012) 293–305.

[68] P. Agnolucci, Renewable electricity policies in The Netherlands, Renewable Energy 32 (5) (2007) 868–883.

[69] European Commission, Taxation and customs union -Taxes in Europe, accessed 14.11.2017 (2017).
URL https://ec.europa.eu/taxation_customs/index_en

[70] F. Douard, C. Domecq, W. Lair, A probabilistic approach to introduce risk measurement indicators to an offshore wind project evaluation–improvement to an existing tool ecume, Energy Procedia 24 (2012) 255–262.

[71] R. Martin, I. Lazakis, S. Barbouchi, Analysis of input factors to operations and maintenance of two offshore wind farm case studies; a screening process, in: 3rd Renewable Power Generation Conference (RPG 2014), 2014, pp. 1–6.

[72] R. Martin, I. Lazakis, S. Barbouchi, L. Johanning, Sensitivity analysis of offshore wind farm operation and maintenance cost and availability, Renewable Energy 85 (2016) 1226–1236.

[73] M. Florian, J. D. Sørensen, Wind turbine blade life-time assessment model for preventive planning of operation and maintenance, Journal of Marine Science and Engineering 3 (3) (2015) 1027–1040.

[74] M. Asgarpour, J. D. Sorensen, O&M modeling of offshore wind farms — state of the art and future developments, in: Reliability and Maintainability Symposium (RAMS), 2016 Annual, IEEE, 2016, pp. 1–6.

[75] S. Ambühl, J. Dalsgaard Sørensen, Sensitivity of risk-based maintenance planning of offshore wind turbine farms, Energies 10 (4) (2017) 505.

[76] M. Reder, N. Y. Yürüşen, J. J. Melero, Data-driven learning framework for associating weather conditions and wind turbine failures, Reliability Engineering & System Safety 169 (2018) 554–569.

[77] N. Gatzert, T. Kosub, Risks and risk management of renewable energy projects: The case of onshore and offshore wind parks, Renewable and Sustainable Energy Reviews 60 (2016) 982–998.

[78] C. Zhang, S. Zhang, Association rule mining: models and algorithms, Springer-Verlag, 2002.

[79] C. C. Aggarwal, J. Han, Frequent pattern mining, Springer, 2014.

[80] X. Jie, H. Wang, M. Fei, D. Du, Q. Sun, T. Yang, Anomaly behavior detection and reliability assessment of control systems based on association rules, International Journal of Critical Infrastructure Protection 22 (2018) 90–99.

[81] B. Grabot, Rule mining in maintenance: Analysing large knowledge bases, Computers & Industrial Engineering (2018).

[82] M. Doostan, B. H. Chowdhury, Power distribution system fault cause analysis by using association rule mining, Electric Power Systems Research 152 (2017) 140–147.

[83] Z. Guo, D. Chi, J. Wu, W. Zhang, A new wind speed forecasting strategy based on the chaotic time series modelling technique and the apriori algorithm, Energy Conversion and Management 84 (2014) 140–151.

[84] N. Yıldırım, B. Uzunoğlu, Data Mining via Association Rules for Power Ramps Detected by Clustering or Optimization, in: L. M. Gavrilova, K. C. J. Tan, A. Sourin (Eds.), Transactions on Computational Science XXVIII: Special Issue on Cyberworlds and Cybersecurity, Springer Berlin Heidelberg, 2016, pp. 163–176.

[85] R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules, in: Proceedings of the 20th International Conference on Very Large Databases, 1994, pp. 487 – 499.

[86] S. Brin, R. Motwani, J. D. Ullman, S. Tsur, Dynamic itemset counting and implication rules for market basket data, Acm Sigmod Record 26 (2) (1997) 255–264.

[87] M. Hahsler, A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules, 2015, accessed on 15.12.2016.
URL `http://michael.hahsler.net/research/association_rules/measures.html`

[88] C. R. de Sá, P. Azevedo, C. Soares, A. M. Jorge, A. Knobbe, Preference rules for label ranking: Mining patterns in multi-target relations, Information Fusion 40 (2018) 112–125.

[89] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, Cnc machine tool health assessment using dynamic bayesian networks, IFAC Proceedings Volumes 44 (1) (2011) 12910–12915.

[90] G. Bode, T. Schreiber, M. Baranski, D. Müller, A time series clustering approach for building automation and control systems, Applied Energy 238 (2019) 1337–1345.

[91] S. Tufféry, Data mining and statistics for decision making, Vol. 2, Wiley Chichester, 2011.

[92] J. H. Ward Jr, Hierarchical grouping to optimize an objective function, Journal of the American statistical association 58 (301) (1963) 236–244.

[93] I. Panapakidis, A. Dagoumas, Modeling demand-price curve: A clustering approach to derive dynamic elasticity for demand response programs, in: Meeting the Energy Demands of Emerging Economies, 40th IAEE International Conference, June 18-21, 2017, International Association for Energy Economics, 2017.

[94] J. Leskovec, A. Rajaraman, J. D. Ullman, Mining of massive datasets, Cambridge university press, 2014.

[95] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, Vol. 344, John Wiley & Sons, 2009.

[96] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, cluster: Cluster Analysis Basics and Extensions, r package version 2.0.6 — For new features, see the 'Changelog' file (in the package source) (2017).

[97] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R package for determining the relevant number of clusters in a data set, Journal of Statistical Software 61 (6) (2014) 1–36.

[98] J. Wu, Advances in K-means clustering: a data mining thinking, Springer Science & Business Media, 2012.

[99] Y. Feng, X. Huang, R. Hong, J. Chen, 1828. residual useful life prediction of large-size low-speed slewing bearings–a data driven method., Journal of Vibroengineering 17 (8) (2015).

[100] M. Žvokelj, S. Zupan, I. Prebil, Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis, Mechanical Systems and Signal Processing 24 (4) (2010) 1049–1067.

[101] P. E. Van Der MLJP, J. Van Den Hh, Dimensionality reduction: A comparative review, Tilburg, Netherlands: Tilburg Centre for Creative Computing, Tilburg University, Technical report 5 (2009) 2009.

[102] R. J. Freund, W. J. Wilson, P. Sa, Regression analysis, Elsevier, 2006.

[103] A. Naylor, Generalized frequency response concepts for time-varying, discrete-time linear systems, IEEE Transactions on Circuit Theory 10 (3) (1963) 428–440.

[104] J. Nash, L. Lefkovitch, Principal components and regression by singular value decomposition on a small computer, Journal of the Royal Statistical Society: Series C (Applied Statistics) 25 (3) (1976) 210–216.

[105] A. R. Leach, V. J. Gillet, An introduction to chemoinformatics, Springer Science & Business Media, 2007.

[106] I. T. Jolliffe, Principal component analysis, Springer, 2002.

[107] R. H. Kewley, M. J. Embrechts, C. Breneman, Data strip mining for the virtual design of pharmaceuticals with neural networks, IEEE Transactions on Neural Networks 11 (3) (2000) 668–679.

[108] P. Cortez, M. J. Embrechts, Using sensitivity analysis and visualization techniques to open black box data mining models, Information Sciences 225 (2013) 1–17.

[109] W. L. Winston, J. B. Goldberg, Operations research: applications and algorithms, Vol. 3, Thomson Brooks/Cole Belmont, 2004.

[110] A. Levitin, Introduction to the design & analysis of algorithms, Boston: Pearson,, 2012.

[111] D. E. Knuth, The art of computer programming: sorting and searching, Vol. 3, Pearson Education, 1997.

[112] A. C. Harvey, Forecasting, structural time series models and the Kalman filter, Cambridge university press, 1990.

[113] R. G. Kavasseri, K. Seetharaman, Day-ahead wind speed forecasting using f-ARIMA models, Renewable Energy 34 (5) (2009) 1388–1393.

[114] M. Reder, J. J. Melero, Time series data mining for analysing the effects of wind speed on wind turbine reliability, Safety and Reliability – Theory and Applications 1st Edition (2017) 109–119.

[115] C. Jung, High spatial resolution simulation of annual wind energy yield using near-surface wind speed time series, Energies 9 (5) (2016) 344.

[116] K. Pearson, Contributions to the mathematical theory of evolution, Philosophical Transactions of the Royal Society of London. A 185 (1894) 71–110.

[117] I. Couso, O. Strauss, H. Saulnier, Kendall's rank correlation on quantized data: An interval-valued approach, Fuzzy Sets and Systems 343 (2018) 50–64.

[118] M. Bermudez-Edo, P. Barnaghi, K. Moessner, Analysing real world data streams with spatio-temporal correlations: Entropy vs. pearson correlation, Automation in Construction 88 (2018) 87–100.

[119] W. B. Kendall, A new algorithm for computing correlations, IEEE Transactions on Computers 100 (1) (1974) 88–90.

[120] R. Nelsen, Kendall tau metric, Encyclopaedia of mathematics 3 (2001) 226–227.

[121] H. Abdi, The kendall rank correlation coefficient, Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA (2007) 508–510.

[122] V. Troster, M. Shahbaz, G. S. Uddin, Renewable energy, oil prices, and economic activity: A granger-causality in quantiles analysis, Energy Economics 70 (2018) 440–452.

[123] D. Ciuiu, M. Bădileanu, L. I. Georgescu, Granger causality test and chow breakpoint test on the romanian day ahead electricity market, Procedia Economics and Finance 22 (2015) 601–609.

[124] B. Hamdi, M. Aloui, F. Alqahtani, A. Tiwari, Relationship between the oil price volatility and sectoral stock markets in oil-exporting economies: Evidence from wavelet nonlinear denoised based quantile and granger-causality analysis, Energy Economics 80 (2019) 536–552.

[125] F. Hayat, M. D. S. Pirzada, A. A. Khan, The validation of granger causality through formulation and use of finance-growth-energy indexes, Renewable and Sustainable Energy Reviews 81 (2018) 1859–1867.

[126] K. Pinzón, Dynamics between energy consumption and economic growth in ecuador: A granger causality analysis, Economic Analysis and Policy 57 (2018) 88–101.

[127] D. C. LeBlanc, Statistics: concepts and applications for science, Vol. 2, Jones & Bartlett Learning, 2004.

[128] A. Arratia, Computational finance, An Introductory Course with R, Atlantis Studies in Computational Finance and Financial Engineering 1 (2014).

[129] R. B. Cleveland, W. S. Cleveland, J. E. McRae, I. Terpenning, Stl: A seasonal-trend decomposition, Journal of Official Statistics 6 (1) (1990) 3–73.

[130] E. Chham, A. Milena-Pérez, F. Piñero-García, M. Hernández-Ceballos, J. Orza, E. Brattich, T. E. Bardouni, M. Ferro-García, Sources of the seasonal-trend behaviour and periodicity modulation of 7be air concentration in the atmospheric surface layer observed in southeastern spain, Atmospheric Environment (2019).

[131] J. Hochenbaum, O. S. Vallis, A. Kejariwal, Automatic anomaly detection in the cloud via statistical learning, arXiv preprint arXiv:1704.07706 (2017).

[132] F. Martínez, M. P. Frías, M. D. Pérez-Godoy, A. J. Rivera, Dealing with seasonality by narrowing the training set in time series forecasting with knn, Expert Systems with Applications 103 (2018) 38–48.

[133] L. Soder, Simulation of wind speed forecast errors for operation planning of multiarea power systems, in: 2004 International Conference on Probabilistic Methods Applied to Power Systems, IEEE, 2004, pp. 723–728.

[134] A. Naimo, A Novel Approach to Generate Synthetic Wind Data, Procedia - Social and Behavioral Sciences 108 (2014) 187–196.

[135] J. Chen, C. Rabiti, Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems, Energy 120 (2017) 507–517.

[136] G. E. Box, G. M. Jenkins, Time series analysis: forecasting and control, revised ed, Holden-Day, 1976.

[137] R. J. Hyndman, forecast: Forecasting functions for time series and linear models, r package version 8.1, accessed 9.11.2017 (2017).
URL `http://github.com/robjhyndman/forecast`

[138] R. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for r, Journal of Statistical Software, Articles 27 (3) (2008) 1–22.

[139] U. Grömping, et al., Relative importance for linear regression in r: the package relaimpo, Journal of statistical software 17 (1) (2006) 1–27.

[140] J. Weissteiner, Variable importance measures in classification and regression methods, Master's thesis, Vienna University of Economics and Business (2018).

[141] H. J. Lang, D. N. Merino, The selection process for capital projects, Vol. 9, John Wiley & Sons, 1993.

[142] T. Au, T. P. Au, Engineering Economics for Capital Investment Analysis., Prentice-Hall, New Jersey, 1983.

[143] Technical Committee CEN/TC 319, En 13306:2010 maintenance - maintenance terminology (2010).

[144] P. Jain, Wind energy engineering, McGraw-Hill, New York, 2011.

[145] European Parliamentary Research Service, Understanding electricity markets in the eu, accessed 17.11.2017 (2017).
URL http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593519/EPRS_BRI(2016)593519_EN.pdf

[146] Market Observatory for Energy, Quarterly report on european electricity markets, second quarter of 2008, Tech. rep., European Commission, Brussels (2008).

[147] Market Observatory for Energy, Quarterly report on european electricity markets, third quarter of 2014, Tech. rep., European Commission, Brussels (2014).

[148] Market Observatory for Energy, Quarterly report on european electricity markets, first quarter of 2017, Tech. rep., European Commission, Brussels (2017).

[149] R. Haffner, O. Batura, K. Ryszka, van den Bergen Kimberley, Competition policy and an internal energy market, Tech. rep., Directorate General For Internal Policies, Policy Department A: Economic and Scientific Policy European Parliament, Brussels (2017).

[150] Euphemia public description pcr market coupling algorithm, Tech. rep., Price Coupling of Regions (PCR), accessed 07.12.2018 (2018).
URL https://www.nordpoolgroup.com/globalassets/download-center/pcr/euphemia-public-description.pdf

[151] WindEurope, Cumulative installations onshore and offshore by country, accessed 01.06.2018 (2017).
URL https://windeurope.org/about-wind/statistics/european/windinpower2017/#findings

[152] WindEurope, Wind energy in Europe: Scenarios for 2030, accessed 01.06.2018 (2017).
URL https://windeurope.org/wpcontent/uploads/files/about-wind/reports/WindenergyinEuropeScenariosfor2030.pdf

[153] G. Erbach, Understanding electricity markets in the EU, accessed 01.06.2018 (2016).
URL http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593519/EPRS_BRI(2016)593519_EN.pdf

[154] D. W. Edwards, Energy trading & investing: Trading, risk management, and structuring deals in the energy markets, McGraw-Hill, 2010.

[155] F. Ueckerdt, R. Kempener, From baseload to peak: renewables provide a reliable solution, International Renewable Energy Agency (2015).

[156] F. Ueckerdt, R. Brecha, G. Luderer, Analyzing major challenges of wind and solar variability in power systems, Renewable energy 81 (2015) 1–10.

[157] A. Lokhov, Load-following with nuclear power plants, NEA news 29 (2) (2011) 18–20.

[158] J. Hynes, How to compare power generation choices, accessed 19.05.2019, (2019).
URL https://www.renewableenergyworld.com/2009/10/29/how-to-comparepowergenerationchoices/#gref.

[159] A. M. De Livera, R. J. Hyndman, R. D. Snyder, Forecasting time series with complex seasonal patterns using exponential smoothing, Journal of the American Statistical Association 106 (496) (2011) 1513–1527.

[160] M. Á. Lynch, J. Curtis, The effects of wind generation capacity on electricity prices and generation costs: a monte carlo analysis, Applied Economics 48 (2) (2016) 133–151.

[161] European Network of Transmission System Operators, ENTSO-E Transparency Platform: Day-ahead prices, accessed 14.11.2017 (2017).
URL https://transparency.entsoe.eu/

[162] Eurostat, Electricity statistics files, accessed 01.07.2018 (2017).
URL https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=File:Electricity_Statistics_files,_2017_(GWh)1update23042018.xlsx&oldid=383871

[163] L. Hirth, J. Mühlenpfordt, M. Bulkeley, The entso-e transparency platform–a review of europe's most ambitious electricity data platform, Applied Energy 225 (2018) 1054–1067.

[164] Entso-e bidding zone configuration technical report 2018, Tech. rep., ENTSOE, accessed 11.01.2019 (2018).
URL https://docstore.entsoe.eu/Documents/Events/2018/BZ_report/20181015_BZ_TR_FINAL.pdf

[165] Statistical factsheet 2015, Tech. rep., ENTSOE, accessed 01.09.2018 (2015).
URL https://docstore.entsoe.eu/Documents/Publications/Statistics/Factsheet/entsoe_sfs_2015.pdf

[166] ENTSOE, Physical energy & power flows, accessed 11.01.2019 (2019).
URL https://www.entsoe.eu/data/power-stats
/physical-flows/

[167] Institution of Civil Engineers, What caused recent spikes in uk wholesale power prices?, accessed 19.05.2019, (2017).
URL https://www.ice.org.uk/newsandinsight/thecivilengineer/february2017/whatcausedtherecentspikeinpower-pricesl.

[168] P. J. Aphalo, Learn R ...as you learnt your mother tongue, Leanpub, 2016.

[169] S. Reed, Power Prices Go Negative in Germany, a Positive for Energy Users, accessed 01.07.2018 (2017).
URL https://www.nytimes.com/2017/12/25/business/energy-environment/germanyelectricitynegativeprices.html

[170] T. Appelhans, metvurst: Meteorological visualisation untilities using R for science and teaching, r package, accessed 9.11.2018 (2013).
URL https://www.rdocumentation.org/packages/metvurst

[171] Ó. P. Lamigueiro, Displaying time series, spatial, and space-time data with R, Chapman and Hall/CRC, 2014.

[172] British Standards Institution, EN 15341:2007 Maintenance Maintenance Key Performance Indicators (2017).

[173] J. Carroll, A. McDonald, D. McMillan, Failure rate, repair time and unscheduled o&m cost analysis of offshore wind turbines, Wind Energy 19 (2015) 1107–1119.

[174] S. Pfaffel, S. Faulstich, K. Rohrig, Performance and reliability of wind turbines: A review, Energies 10 (11) (2017) 1904.

[175] H. Seyr, M. Muskulus, Value of information of repair times for offshore wind farm maintenance planning, in: Journal of Physics: Conference Series, Vol. 753, IOP Publishing, 2016, p. 092009.

[176] J. Ribrant, L. M. Bertling, Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005, IEEE Transactions on Energy Conversion 22 (1) (2007) 167–173.

[177] M. Reder, E. Gonzalez, J. J. Melero, Wind turbine failures-tackling current problems in failure data analysis, in: Journal of Physics: Conference Series, Vol. 753, IOP Publishing, 2016, p. 072027.

[178] M. Kanamitsu, W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnillo, M. Fiorion, G. L. Potter, Ncep/doe amip-ii reanalysis (r-2), Bull. Amer. Meteor. Soc. 83 (2002) –12.

[179] M. U. Kemp, E. Emiel van Loon, J. Shamoun-Baranes, W. Bouten, RNCEP: global weather and climate data at your fingertips, Methods in Ecology and Evolution 3 (1) (2012) 65–70.

[180] G. Wilson, D. McMillan, Assessing Wind Farm Reliability Using Weather Dependent Failure Rates, Journal of Physics: Conference Series 524 (1) (2014) 1–10.

[181] P. J. Tavner, D. M. Greenwood, M. W. G. Whittle, R. Gindele, S. Faulstich, B. Hahn, Study of weather and location effects on wind turbine failure rates, Wind Energy 16 (2) (2013) 175–187.

[182] G. Jiang, J. Keller, P. L. Bond, Determining the long-term effects of H2S concentration, relative humidity and air temperature on concrete sewer corrosion, Water Research 65 (2014) 157–169.

[183] G. O. Lloyd, Atmospheric corrosion, in: M. Radojevic, R. Harrison (Eds.), ACIDITY - Sources, Consequences and Abatement, Elsevier Science Publishers Ltd, 2000, pp. 1–8.

[184] Y. Xiang, Z. Wang, Z. Li, W. Ni, Long term corrosion of X70 steel and iron in humid supercritical CO 2 with SO 2 and O 2 impurities, Corrosion Engineering, Science and Technology 48 (5) (2013) 395–398.

[185] P. Drobinski, C. Coulais, B. Jourdier, Surface wind-speed statistics modelling: Alternatives to the weibull distribution and performance evaluation, Boundary-Layer Meteorology (2015) 1–27.

[186] N. Y. Yürüşen, J. J. Melero, Probability density function selection based on the characteristics of wind speed data, in: Journal of Physics: Conference Series, Vol. 753, IOP Publishing, 2016, p. 032067.

[187] AEMET, AEMET OpenData, accessed 22.11.2017 (2017).
URL `http://www.aemet.es/en/datos_abiertos/AEMET_OpenData`

[188] Organisation for Economic Co-operation and Development, Long-term interest rates, inflation, accessed 14.11.2017 (2016).
URL `https://data.oecd.org/`

[189] Non-Fossil Purchasing Agency, e-ROC track record, accessed 21.11.2017 (2017).
URL `http://www.epowerauctions.co.uk/erocrecord.htm`

[190] MEASNET, Power Performance Measurement Procedure – Version 5 (2009).

[191] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2013).
URL `http://www.R-project.org/`

[192] International Electrotechnical Commission, IEC 61400-12-2:2013 Wind turbines Part 12-2: Power performance of electricity-producing wind turbines based on nacelle anemometry (2016).

[193] R. J. Hyndman, Y. Khandakar, Automatic time series forecasting: the forecast package for R, Journal of Statistical Software 26 (3) (2008) 1–22.

[194] J. Song, J. Wang, H. Lu, A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting, Applied Energy 215 (2018) 643–658.

[195] C. Li, Z. Xiao, X. Xia, W. Zou, C. Zhang, A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting, Applied Energy 215 (2018) 131–144.

[196] T. H. El-Fouly, E. F. El-Saadany, M. M. Salama, One day ahead prediction of wind speed and direction, IEEE Transactions on Energy Conversion 23 (1) (2008) 191–201.

[197] J. Browell, D. Drew, K. Philippopoulos, Improved very short-term spatio-temporal wind forecasting using atmospheric regimes, Wind Energy (2018).

[198] O. Brasseur, Development and application of a physical approach to estimating wind gusts, Monthly Weather Review 129 (1) (2001) 5–25.

[199] European Network of Transmission System Operators, ENTSO-E Transparency Platform: Day-ahead prices, accessed 14.11.2018, (2018). URL https://transparency.entsoe.eu/.

[200] T. Gintautas, J. D. Sørensen, Improved methodology of weather window prediction for offshore operations based on probabilities of operation failure, Journal of Marine Science and Engineering 5 (2) (2017) 20.

[201] European Centre for Medium-Range Weather Forecasts, Ecmwf operational regular gridded data at 1.125 degrees resolution., accessed 23.01.2018, (2012).
URL http://catalogue.ceda.ac.uk/uuid/a67f1b4d9db7b1528b800ed48198bdac.

[202] D. C. Bowman, J. M. Lees, Near real time weather and ocean model data access with rNOMADS, Computers & Geosciences 78 (2015) 88–95.