

Application of a process model to define potential operational functional failure conditions

**By
Paul Schepers**

Emmen, 19-9-2019

**Dynamics Based Maintenance (DBM) Research Group
Faculty of Engineering Technology
University of Twente**

**A thesis submitted to the University of Twente in accordance with the requirements of the degree
of Master of Science in Mechanical Engineering**

**In cooperation with
HIsarna, TATA Steel Europe – the Netherlands**

**Thesis supervisor:
Dr. Melissa Schwarz, P.Eng.**

**Examination committee:
Prof. Dr. Ir. Tiedo Tinga
Prof. Dr. Ing. Bojana Rosic
Ir. Johan van Boggelen**

Abbreviations

Δ DT	Difference between the medium dry-bulb temperature and the dew point temperature
AFAT	All-Factors-At-a-Time
CPP	Coal preparation plant
CPS	Cyber-physical systems
CPU	Central processing unit
CV-RMSE	Cross Validation – Root Mean Square Error
dp	Differential pressure
EE	Elementary Effects (Morris method)
FF	Functional failure
GCI	Granular coal injection
GCP	Granular coal preparation
GCSS	Granular coal supply silo
H	Symbolic name followed by a number for a HEXS-model input factor
HEXS	HIsmelt Excel Simulation
HGI	Hard grove index
HI _{sarna}	HIsmelt (name of the melting vessel) and Isarna (Celtic word for iron)
HMI	Human Machine Interface
IAPWS	Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam
M	Symbolic name followed by a number for a metamodel input factor
MAIN	Main drying gas circulation
MCC	Moisture-carrying capacity
Mill	Impact dryer mill (including the drying chamber)
ML	Machine Learning
MOT	Mill outlet dry-bulb temperature
OEM	Original equipment manufacturer
OFAT	One-Factor-At-a-Time
P&ID	Piping and Instrumentation Diagram
PdM	Predictive maintenance
PFFC	Potential functional failure conditions
PLC	Programmable Logic Controller
PSD	Particle size distribution
RCC	Raw coal charging
RCSS	Raw coal supply silo
RECIRC	Drying gas recirculation
RMPP	Raw material preparation plant
RMSE	Root Mean Square Error
ROM	Run-of-mine materials
SA	Sensitivity analysis
SML	Supervised machine learning
SUPREME	Smart Sensing and Predictive Maintenance in Steel Manufacturing

Summary

Today, manufacturing systems are complex with respect to interconnective technologies, operations, and maintenance processes. In Industry 4.0, the number of variables involved in complex systems is vast; as such, systems can exhibit non-linear interactive trends that may be beyond human understanding. Thus, addressing the complexity of such systems requires the utilization of advanced strategies by which data can be systematically processed into information. The information can explain variability and thereby assist in making more informed decisions such as where to allocate resources, assist in developing strategies for process control and maintenance activities. In maintenance, and specifically in predictive maintenance (PdM), variability insights can be utilized to prevent and delay system failures just-in-time.

Sensitivity analysis (SA) can assist in understanding such complex systems. SA can identify the influence of input parameters in relation to a response; thus, it can be utilized to understand the behaviour of complex systems (such as operations involved in production processes), and in succession, to facilitate the development and maintenance of these systems. However, when the process itself is poorly known, researchers caution that special care should be taken when drawing conclusions based on an SA. This study aims to analyse and illustrate how an existing process model of a complex system can be utilized for PdM.

To achieve this aim, first a generic overview is created of the process, where the system functions and the operational conditions are investigated. The system is the raw material preparation plant (RMPP) that processes coal within the HIsarna pilot plant (Tata Steel Europe). The HIsarna pilot plant is part of the sustainability initiatives undertaken by Tata Steel IJmuiden and utilizes technology to reduce energy use and emissions during ironmaking. The HIsarna plant is a combination of technology acquired from Rio Tinto's HIsarna plan and information developed in IJmuiden. Maintenance has been identified as a hurdle to the rollout of this technology. Therefore, the HIsarna RMPP is utilized as a case study in this work.

As part of the acquisitions, a process model was obtained of the original Rio Tinto's HIsarna RMPP. The process model --the HIsarna Excel simulation (HEXS) model-- describes thermodynamic process behaviour of a complex HIsarna RMPP. This model can be utilized to determine process conditions to operate the RMPP, such as material composition, flow rates, temperatures, and other settings. Upon acquisition, the HEXS model has not been validated against HIsarna's RMPP operations. Therefore, the model must first be validated against a data-driven model, before it can be utilized in a SA. Should the HEXS model be found to reasonably reflect the real-world process of the HIsarna RMPP, then it could provide insights into the parameters that influence the actual system and sub-systems of the HIsarna RMPP.

The data-driven model is constructed by means of supervised machine learning (SML) and trained with historical HIsarna RMPP operating data. The model describes a particular behavior exhibited in the HIsarna RMPP, being the mill outlet temperature (MOT), and is validated against a test set before validating the HEXS model. When both models are validated, they are utilized in the SA to identify the potential operational failure factors of the RMPP. Lastly, the HEXS model is utilized to identify potential operational functional failure conditions.

To identify a suitable regression algorithm for the data-driven model, nineteen regression algorithms are trained by k-fold cross-validation SML and evaluated on their cross-validation root means square error (CV RMSE). The predictive performance of the best performing model is assessed by validating it with the actual MOT response. This best performing data-driven model is defined by the lowest CV-RMSE of the algorithms. This best performing algorithm is then utilized to validate the HEXS model with respect to the MOT of the HIsarna RMPP for a comparable set of inputs.

During the validation experiments, the majority of validity agreements between the HEXS model and the data driven model were observed, thus indicating the validity of the HEXS model. However, a minor amount of validity disagreements exists between both models. Therefore, it cannot definitively be concluded that the HEXS model is representing the MOT of the HIsarna RMPP under all operating conditions.

An extended Morris SA method is performed on the case study of a commercial RMPP. The conventional Morris method does not provide a normalization of ratings for the relative contribution of influence with respect to other compared factors. The proposed extended Morris method accounts for this by allowing the comparison of normalized values of different models, such as the SA results from the HEXS and data-driven model. Normalization is achieved by utilizing the cumulative percentages of a Pareto plot. Furthermore, it is demonstrated that presenting SA results in a Pareto plot allows to distinguish between the “vital few” and the “useful many” factors dominantly influencing the response parameter. Although, it is not evident if the HEXS model is valid, in the SA the burner outlet temperature is identified to be the most influential factor to the MOT, supporting the assumption of repetitive validity.

Finally, this thesis demonstrates that the HEXS process model can be utilized for defining potential dominant operational functional failure factors and potential risk full operating conditions in a complex system. This work thus demonstrates that a process model can assist PdM by defining potential thermodynamically related failure conditions and by identifying their related dominant influencing factors.

Keywords: Steady state thermodynamic process model, complexity of systems, coal preparation plant, raw material preparation plant, predictive maintenance, sensitivity analysis, extended Morris method, supervised machine learning.

Table of Contents

1	Introduction	1
1.1	Industrial revolutions and the evolution of maintenance.....	2
1.2	Details of the SUPREME project	4
1.2.1	The case study.....	4
1.3	Complexity of systems	5
1.4	Context of this research.....	6
1.4.1	Project description.....	6
1.4.2	Aim and research questions	7
1.5	Research approach.....	8
1.5.1	Global strategy	8
1.5.2	Project activities	9
1.5.3	Scope	10
1.5.4	Limitations.....	10
1.6	Definitions and terms.....	11
1.7	Thesis outline.....	12
2	Background and related work	13
2.1	The HIsarna raw material preparation plant.....	13
2.2	Functional description of the RMPP	14
2.2.1	Production process.....	14
2.2.2	{1} Raw coal storage yard.....	14
2.2.3	{2–3} Raw coal charging plant	14
2.2.4	{4–9} Coal preparation plant: Main drying gas circulation.....	16
2.2.5	{10–15} Coal preparation plant: Drying gas recirculation	17
2.2.6	Inert conditions and oxygen content	18
2.2.7	Water spray.....	18
2.2.8	Makeup air flow rate	19
2.2.9	Operation of the RMPP.....	19
2.3	The HIs melt Excel simulation model (the HEXS model).....	20
2.3.1	The HEXS model.....	20
2.3.2	Evaluation, limitations, and comparison of HEXS with the HIsarna RMPP	22
2.4	Introduction to drying.....	24
2.4.1	Introduction to the convective drying of porous solids.....	24
2.5	Discussion of the commonalities	25
3	Experimental design and methodology	27
3.1	Validation strategy.....	27
3.1.1	Regression strategy	28
3.1.2	Evaluation criteria for the trained model.....	30
3.2	Sensitivity analysis	31
3.2.1	Sensitivity analysis strategy.....	31

3.2.2	Selecting an SA method.....	31
3.2.3	Local and global sensitivity	34
3.2.4	Elementary effects (Morris method)	35
3.2.5	Extensions of the Morris method.....	37
3.2.6	Applying the theory to extend the Morris method	39
4	Constructing a metamodel for the RMPP.....	41
4.1	Step mm1: Metamodel objectives	41
4.2	Step mm2: Identification of the required data	41
4.2.1	Retrieved forms of the data elements	44
4.2.2	Selecting the data retrieval form and resolution.....	45
4.3	Step mm3: Data pre-processing.....	48
4.4	Step mm4 and mm5: Sampling and training the regression algorithms.....	48
4.5	Step mm6: Select the regression algorithm.....	49
5	Execution, results and discussion.....	51
5.1	Goodness of fit, comparison of the observed versus the predicted.....	51
5.2	Validation of the HEXS model	54
5.2.1	Initialization of the validation experiment for (1) and (2).....	54
5.2.2	(1) Results: Are the signs of the linear regression equation similar?	56
5.2.3	(2) Results: Are the response behaviours as expected?	57
5.2.4	(3) Results: Are the SA results similar for the HEXS and metamodel?.....	59
5.3	Metamodel sensitivity analysis	61
5.4	HEXS model sensitivity analysis	64
5.5	Potential operational functional failure condition (PoC).....	67
5.5.1	Special note for a potential safety hazard	69
5.6	General discussions	70
5.6.1	Constructing the metamodel	70
5.6.2	Validation.....	70
5.6.3	Sensitivity analysis.....	71
5.6.4	Potential operational functional failure condition (PoC).....	72
5.7	Remarks and observations.....	72
6	Conclusions and recommendations	75
6.1	Answers to the research sub-questions	75
6.2	Answers to the main research question.....	77
6.3	Recommendations	77
	Bibliography	79
	Appendices	83

1 Introduction

Technological advancement has led to complex systems, that may be beyond human understanding; complex with respect to interconnective technologies, operations, and maintenance processes. In this work it is examined how a process model can be utilized to obtain valuable insights in a complex system for predictive maintenance (PdM). Sensitivity analysis (SA) performs a significant role in the understanding and investigation of the complex system. SA assist to identify the influence of input operation condition in relation to the systems output, that may be critical for the system operations.

From a maintenance and process perspective, it is beneficial to identify the critical conditions and parameters that influence product quality, production uptime, and plant safety; as the availability of such knowledge can be used to improve the availability, reliability, and cost effectiveness of complex technical systems.

This study was conducted in the context of the Smart Sensoring and Predictive Maintenance in Steel Manufacturing (SUPREME) project [16], and was executed in cooperation with Tata Steel IJmuiden. Tata Steel IJmuiden possess a complex HIsarna raw material preparation plant that is used as a case study for the SUPREME project.

This introduction starts with providing a brief introduction to maintenance and to changes that have occurred to maintenance practices due to industrial revolutions. The following subsection then introduces the SUPREME project. Thereafter, an introduction to the challenges posed by the complexity of modern systems is presented. Furthermore, Section 1.3 emphasizes the urgent need for a tool (in the form of a maintenance strategy) that can be adapted to the fourth industrial revolution; this tool should be able to address the ever-increasing complexity of systems. Thereafter, Section 1.4 establishes the context of this thesis and presents its problem statement and objectives; in addition Section 1.5 describes the approach adopted, states limitations for the research, offers key definitions of terms and outlines the remainder of this thesis.

1.1 Industrial revolutions and the evolution of maintenance

An industrial revolution marks a major turning point in history. It refers to a period of time in which daily life and human well-being are significantly affected as a result of technological developments. In particular, economic, technological, social and cultural changes can impact a society's standard of living to such a degree that their impact can be compared to that of the shift from a hunter-gatherer to a farming society [1]. The following highlights some aspects that affect the standard of living in industrialized countries: an increase in the human development index, an increase in life expectancy, an increase in gross domestic product, an increase computational power, an increase of product quality for decreased prices, and a decline in working hours per year (as productivity increases, working hours decrease) [2][3][4][5].

The first industrial revolution emerged at the end of the 18th century as a result of the development of the ability to harness steam for power, which made human labour more productive [6]. In the 1870s, the second industrial revolution dawned as a result of the development of electrically powered systems and the establishment of the first mass production line [7]. The third industrial revolution – the digital revolution – superseded its predecessors and manifested around the 1970s as a result of the development of information technology and electronics and the automation of manufacturing processes, which enabled more efficient production [6][7]. Currently, a new paradigm, one in which the fusion of several technologies is occurring not only in terms of automating products and connectivity but also knowledge, is emerging [7]. This is the fourth industrial revolution, and many terms have been used to refer to this observation, such as “Industrie 4.0” (the Netherlands), “Industry 4.0” (Germany), and “Smart manufacturing” (United States) [8]. Countless engineering innovations have occurred since the first industrial revolution, but perhaps the most dramatic transitions have occurred in the last 50 years [9]. These changes have affected how assets are maintained [10]. Figure 1 presents an overview of industrial revolutions and their trend towards growing complexity.

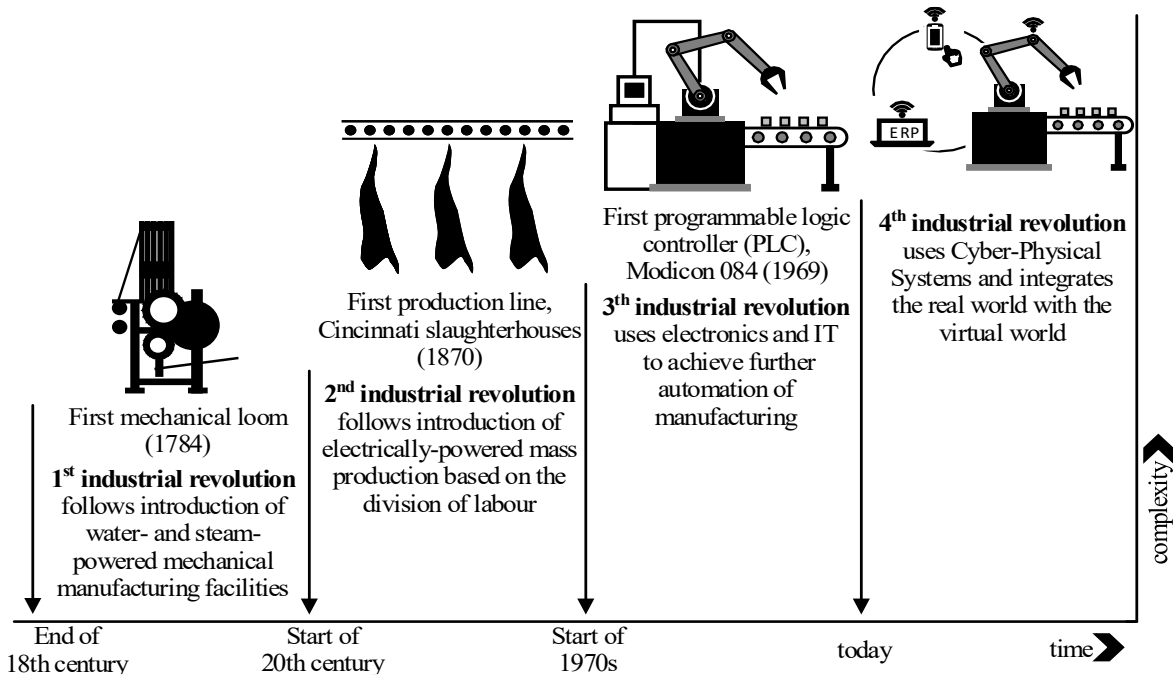


Figure 1. The four industrial revolutions and their increasing complexity. Reproduced from: W. Tiddens (2018), and Kagermann et al. (2013) [6] [8].

Maintenance strategies have progressed with the revolutions. Prior to the Second World War, maintenance strategies could be summarized as only regarding as repair work [9]. Machinery was outfitted with relatively basic instrumentation and control systems and operated until it broke down [9][11]. Production downtime demands were not excessively severe, and it was acceptable to maintain on a breakdown basis, which is a reactive maintenance policy [9][12]. This policy has the advantage that, without intervention, the service lifetime of an asset is fully utilized when a failure occurs. Failure is a state in which the intended function of a part or system can no longer be fulfilled [12]. Consequently, no remaining useful life (RUL) is wasted by replacing or restoring components before actual failure [12].

In the 1950s, increasingly competitive markets and rising labour costs led to an intolerance for downtime. At the same time, systems were increasingly operated at higher speeds, and wear rates thus increased [9][11]. Production demanded reliability improvements, and the concept of planned or preventive maintenance was introduced. Preventive maintenance is a strategy based on restoring or replacing assets at fixed intervals or following designated machine usage checkpoints, regardless of their condition or need for maintenance [9][12]. This policy may increase the reliability of a system or its ability to perform the required function for a specified period under specified operating conditions [12]. The downside of this policy is that replacement or repairs occur before the component or system fails. To reduce the risk of unplanned downtime, components are replaced far before the ends of their RULs, which results in wasted component lifetime and a potential increase in total operational costs [12][13].

During the third industrial revolution (the digital revolution), plants and systems became increasingly complex as a result of the integration of electronics and information technology (e.g. Programmable Logic Controllers (PLCs) and robotics) [7][8]. To keep pace with the demands of a competitive marketplace, intolerance for downtime increased, which resulted in an increase in maintenance costs [9]. During this period, various (sometimes computerized) systemized maintenance management concepts were developed and employed with the aim of increasing reliability at lower costs [10]. These systemized management concepts included condition monitoring, computerized maintenance management systems, and reliability-centred maintenance. With the development of these concepts came a new awareness of failure processes and technologies, whereby an increased understanding of asset and component health was obtained [9]. Combining computerized maintenance management systems with preventive maintenance strategies such as condition monitoring allowed for more optimized maintenance and logistic planning. The combination of these computerized maintenance management systems with preventive maintenance strategies allowed for more accurate and optimized maintenance and logistic planning. To optimise preventative maintenance, the need for such maintenance should be determined based on the actual condition of an asset. This requires collecting and processing (real-time) data on the condition of the asset in question [8]. However, practically, this is not always feasible (e.g. due to sensor and computing limitations).

In recent years, the computing capacity to share, store, and work in real-time with massive amounts of data is developing. This in turn has enabled the development of cyber-physical systems (CPSs) [7]. Ragnathan Rajkumar et al. defined CPSs as physical and engineered systems whose operations are monitored, coordinated, controlled, and integrated by computing and communications technology [7]. The deployment of CPSs in manufacturing industries has enabled the development of networks of systems and assets with the ability to perform more efficiently, collaboratively, and resiliently through the management of interconnected systems [8][14]. However, to maintain such complex systems, there exists a need for rapid and precise decision-making with regard to repair and maintenance facilitated by the real-time monitoring of system conditions [15].

To cope with the fourth industrial revolution, preventive maintenance and the quality management methods that were previously controlled by humans are being transformed into predictive maintenance through the development of various information technologies, such as model-based prognostics by big data and artificial intelligence [15]. Cyber-physical systems permit the infrastructure required for implementing predictive maintenance (PdM), which refers to policies that trigger maintenance activities through predicting failure [8]. Predictive maintenance, is possible today due to advanced digital technologies [8][15]. Such maintenance may be based on the actual conditions of equipment, which are tracked in real time, rather than the average of the estimated RUL statistics. The ultimate goal of PdM is to perform maintenance on a just-in-time basis to prevent failure, which is when maintenance activity is the most cost-effective and before the equipment reaches its specified performance threshold. With regard to cost effectiveness, the costs associated with maintenance and support can account for 60–75% of the total lifecycle costs of a manufacturing system [8]. Therefore, maintenance actions should only be conducted when required, where predictive maintenance can contribute to just-in-time decision-making.

1.2 Details of the SUPREME project

System degradation, which can lead to a functional failure (FF), is a dynamic process that is governed by changes in both a system and its environment. In order to save on maintenance costs (by not make replacements too early) and increase system availability (by not make replacements too late), the challenge is to achieve just-in-time maintenance [16]. This challenge led to the following general research question investigated in the SUPREME project: “How can advanced sensing technology and the modelling of degradation and failure processes be used to develop a PdM concept for production systems?”

The approach adopted to answer this question consists of four steps: First, the system and process in question must be studied and relevant parameters identified; on the basis of these parameters, appropriate sensors should be used to collect the relevant data. The second step is efficiently collecting the required data. The third step is model development, which involves utilizing the physics of failure to predict the critical failures in the system being investigated. The collected data on usage, loads, and environmental conditions will be used as input for the models. The final step is to combine the collected data with the model and demonstrate the integral concept on a system of the HIsarna pilot plant. Based on the advanced sensing and modelling of critical parts of the system in question, one can make predictions concerning component degradation. This information can then be used to optimize maintenance processes. In other words, advanced sensing and modelling can warn of potential failures before they happen. If notified in time, the vulnerable asset can be corrected before it creates any unwanted effects (e.g. a stop in production). The information provided by such an approach concerning component degradation allows process optimization in the areas of planning, logistics, and resource allocation [16]. The work performed in this research is part of the first step.

1.2.1 The case study

The HIsarna pilot plant located in IJmuiden is part of the sustainability initiatives undertaken by Tata Steel and the plant uses technology that reduces energy use and emissions during iron-making. Maintenance has been identified as a hurdle to the rollout of this technology. An overview of the HIsarna plant is presented in Figure 2. As can be seen from this schematic overview, HIsarna consists of various interactive facilities (labelled 1–14), including the raw material preparation plant (RMPP), also called the coal preparation plant (CPP). The RMPP processes coal on demand for the HIsarna iron-making process. This HIsarna RMPP is used as a case study for the SUPREME project because investigating this facility would involve sharing less sensitive information than would discussing other plants involved in the HIsarna process. The reduced risk of sharing confidential information leads to fewer objections on the company's part to the publication of this research. A more detailed description of the HIsarna process and the relevant operations can be found in Appendix 1.1.

Currently, the HIsarna pilot plant is transitioning from being operated as a research and development facility to self-sustaining operations. As a self-sustaining pilot production plant, it will pursue identifying the potential capabilities and sustainable operational conditions of the HIsarna technology. An analysis determined that the RMPP currently uses about 13% of its operating level when compared to a fully operational RMPP, meaning an operating level of 24 hours, seven days a week. Should usage be increased, the level of wear and tear would also increase, which would lead to more intensive maintenance and costs. Maintenance is of critical importance in keeping the system's availability, reliability, and costs at acceptable levels.



- | | | |
|---|---|--------------|
| 1. Alternative raw materials storage silo | 8. Smelting cyclone | 15. Offices |
| 2. Off-gas duct | 9. Smelting reduction vessel | 16. Workshop |
| 3. Gas cooler | 10. Fore hearth | |
| 4. Coal and lime storage silos | 11. Control room | |
| 5. Cooling towers | 12. Coal grinding, drying and screening | |
| 6. Bag filter | 13. Ore drying and screening | |
| 7. Secondary dedusting | 14. Raw materials storage | |

Figure 2. Schematic overview of the HIsarna pilot plant in IJmuiden, the Netherlands. Reprinted from Tata Steel (2019) [63].

1.3 Complexity of systems

In Section 1.1, it was observed that industrial revolutions are a result of systems becoming increasingly complex due to technological advances. Systems can become so sophisticated that their behaviour becomes impossible to understand even for experts. Designers are not able to consider all potential system states, and operators are not able to safely and effectively cope with both expected and unexpected situations. These technological advances have led to systems whose management is beyond human intellectual capabilities, which has resulted in challenges in maintaining complex systems and managing the risks associated with them [17].

The HIsarna RMPP is identified as an immature multilevel coupled complex system that is characterised by interactive, dynamic, and non-linear complexities [17]. This RMPP consists of multiple sub-systems, which makes it a multilevel system. Interactive complexity indicates that dependencies exist among system components. The term dynamic refers to changes that occur over time (e.g. system degradation), and an immature facility such as the RMPP is continuously subjected to plant design and operational modifications. Non-linear means that input and output -- cause and effect -- are not obvious or directly related.

An accident is an unforeseen and unplanned event or circumstance [19]. In engineering, component failure accidents have received much more attention compared to component interaction accidents [17][20]. However, interaction accidents are becoming increasingly common as the complexity of system designs increases [17][20]. According to K. Marais et al. (2007), systems with unknown interactive complexity and tight coupling will undergo accidents that cannot be predicted or prevented [20]. Simultaneously, the development time provided to test systems and designs to understand all their potential behaviours and formulate management strategies has been so reduced that it is often a luxury. Limited development time and other factors make it challenging to identify priorities and appropriate trade-offs when investing resources. By developing a tool that is addressing the ever-increasing complexity of systems, it may allow to predict and even prevent accidents.

1.4 Context of this research

1.4.1 Project description

As noted in Section 1.2, this work is part of the first step of the SUPREME project and involves studying the process of the HIsarna RMPP system and identifying relevant parameters. This section describes how the problem for the research question originated.

From a maintenance and process perspective, it is beneficial to identify the critical components and parameters that influence product quality, production uptime, and plant safety. The availability of such knowledge can be used to improve the availability, reliability, and cost effectiveness of complex technical systems such as an RMPP production facility. When the influence and interactions of parameters are known, a strategy can be developed by which to ensure that the final product meets specifications. However, as system components degrade and operating conditions change, the operational window available may no longer be sufficient to develop products that fulfil the specified criteria. The degradation of a single component can set off a chain reaction that can negatively impact other processes involved in producing a final product. This can result in a spiral of products and processes that do not meet specifications. This reaction continues until the system is restored in such a manner that it can resume its intended function, or the specifications are adapted to new acceptance levels.

As noted previously, the challenge in terms of saving on maintenance costs and increasing system availability is to achieve just-in-time maintenance by predicting component degradation. It is particularly difficult to identify critical components for PdM when an RMPP has a complex and interactive nature and is not yet fully matured and continuously operating. Consequently, comprehensive knowledge of the operational and reliability aspects of the HIsarna RMPP is not yet available.

When considering functional failures within an RMPP, the dominant failure parameters are expected to be the inputs and outputs of the RMPP system, such as temperature, pressure, mass flow rates, composition of the materials used, and valve settings [21][22]. These parameters and operating conditions are expected to have the greatest influence on the overall performance of the RMPP in that they are involved in maintaining its functional requirements; therefore, they dictate the maintenance actions required. Beyond monitoring these operational failure parameters, which may be difficult to measure directly and potentially impossible to measure continuously, an alternative strategy is to understand and monitor how the operation of the RMPP as a system influences these parameters. It is often easier and more economical to monitor these influencing operational parameters than specific asset failure mechanism parameters.

Developing an understanding between operation and a parameter of interest requires a considerable effort in terms of both time and cost, but is not prohibitive. It would be possible to initiate an extensive testing and operating campaign followed by an analysis of trends and data analytic investigations. Sensors can be developed to continuously monitor specific failure parameters; alternatively, the physical interactions of the system can be modelled.

When adopting a modelling approach, the development of an adequate modelling approach based on thermodynamics may require many years and relevant expertise to develop adequately. As such, this approach might not be an option for maintenance personnel. The challenges in developing such an approach lie in a combination of the actual state of knowledge on the subject and the complexities involved in the concurrent physical processes that occur (e.g. evaporation, fractures, potential changes in chemical composition, and turbulent flows). Thus, if developing a system-specific model is beyond the capabilities of the available workforce, measurement techniques are unavailable, and sensor placement is cost-prohibitive, how can information concerning the influencing parameters be obtained?

A potential solution is to utilize an existing process model, such as the HIsarna Excel simulation (HEXS) model. Therefore, this thesis adopts the following strategy: It makes use of an existing thermodynamic process model of an RMPP and investigating how it can be used for PdM. The reasoning underlying this strategy is that within the process model the influence and interactions of input parameters in relation to the output can be studied. Subsequently, the input factors can be compared by ranking their relative influence on the response value, whereby the dominant parameters can be identified. Here, the input refers to factors that change the operating conditions, while the output is a factor related to a potential FF of a system. Based on this information, one can make substantiated choices in terms of the allocation of resources.

The HIs melt Excel simulation (HEXS) model

The HIsarna plant originated from technology and information gathered and developed in IJmuiden and acquired from Rio Tinto, which was located in Kwinana, Australia. As part of the acquisition of the Rio Tinto's HIs melt intellectual property rights, the HIsarna plant also obtained the HEXS process model of the RMPP used by Rio Tinto.

Due to the fact that the HIsarna RMPP is based on HIs melt RMPP design, the HEXS models simulates the process behaviour of a facility similar to that of HIsarna. However, it has not been validated against HIsarna's operations. Should the HEXS model be found reasonably reflecting the real-world process of HIsarna's RMPP, it could provide insights into the parameters that influence the actual system and sub-systems of the RMPP.

Furthermore, no documentation and only limited information are available concerning the HEXS model. This makes it challenging to utilize the model for the SUPREME project, as the model's functioning, objective, and operation are not known. In addition, to practically apply the model, it would be necessary to understand both the commonalities and the differences between model and physical plant. The scope of this dissertation can be determined by identifying the commonalities among the research aspirations, the model, and the actual plant, as the research can then be aligned to these commonalities.

1.4.2 Aim and research questions

Ensuring the efficient operation of complex systems and saving on maintenance costs require strategies that can cope with today's complexities and allow for just-in-time maintenance, such as PdM. To apply PdM, it is first necessary to identify the relevant parameters and conditions influencing the operation of a system. A potential method for identifying the relationship between operation conditions and a parameter of interest is utilizing the existing undocumented HEXS model.

The current work therefore **aims** to obtain an understanding of the use and adoption of the HEXS model for PdM and, based on these observations, develop tools to support the practical application of PdM. The focus is on the identification of dominant model input factors, as one of the purposes of this thesis is to gain insight into a particular complex system, but with the possibilities for further research into critical conditions in mind (i.e. potential functional operating failure conditions). This research is guided by the following **main research question**:

How can a process model be utilized for predictive maintenance?

The practical application is demonstrated by means of a proof of concept (PoC), which supports the method and assists in answering the main question. The case study for the PoC will be performed on the RMPP. The **sub-questions** formulated to provide guidance for this work are as follows:

1. How can the HEXS process model be validated in relation to the physical response behaviour of the RMPP?
2. How accurately can the metamodel describe the behaviour of a selected response of the RMPP?
3. Do the responses (mill outlet temperature) of the HEXS model and metamodel agree when given a similar input?
4. For similar input factors, which dominant factor(s) in the HEXS model and metamodel are in agreement and dominantly influence the mill outlet temperature?
5. Which dominant factor(s) in the HEXS model and metamodel dominantly influence the mill outlet temperature?
6. Which HEXS model input factor(s) and operating conditions can lead to a potential functional failure condition in RMPP operations?

1.5 Research approach

The research conducted in this work was based on a global Lean Six Sigma project charter and project plan that was generated at the beginning of the research. The project charter provides a concise one-page overview of the project, including its overall goal (the main research question), expected results and benefits, limitations, and scope, and is displayed in Appendix 1.2. This charter was used to identify any limitations regarding the scope and goals of the project. These limitations or risks could be minimized should they be identified early and addressed in the research strategy as emphasized in sub-section 1.5.4. Therefore, this one-page overview was used in the development of the global strategy, the scope and the activities conducted for this research, as discussed below.

1.5.1 Global strategy

As noted previously, technical advances in terms of systems have led to complex processes whose management is beyond human intellectual capabilities [17]. These complex processes can be made more manageable when influential interactive process parameters, as well as their accompanying controlling assets, are identified.

The HEXS process model may represent a solution for identifying these parameters and assets, as it may allow for the identification of the factors that are more influential and contribute more significantly to variability to the response of the model. Specifically, it may allow the identification of the influence of an input parameter in relation to the model output. Such an examination is referred to as a **sensitivity analysis (SA)**. Performing an SA on a virtual (modelled) process would allow for an investigation of the factors that affect that process' role and function and make it possible to obtain a greater understanding of the physical process on which the models are based. For example, an unexpected strong dependence of the output upon an input expected to be irrelevant might create awareness of an unexpected aspect of the process [23].

However, to utilize the HEXS process model for PdM, it is first examined which mechanism or asset could potentially cause a failure in a process. Failure is considered a state in which the intended function of a part or system can no longer be fulfilled [12]. Furthermore, it must be determined which elements are dominant in influencing the process that could lead to a potential failure. When known, these elements (e.g. phenomenon, assets, or settings) can be monitored, and, with some computing, the conditions of the system can be estimated.

Combining SA and a process model can be useful, as doing so can allow for the identification of interactions among factors and the subsequent ranking of the most influential input factors and their associated assets. Knowledge of the process variability of the plant and the dominant variables and assets supports decision-making in activities such as allocating resources and prioritizing monitoring conditions and/or assets. In PdM, the combination of SA and the process model can be used as a tool to identify process conditions assets that are most critical or most influential regarding to a potential operational functional failure (FF).

The applicability of such a combined approach is explained by the following example and is illustrated in Figure 3. First, a response by the RMPP that can lead to a functional operating failure of the process is determined. This response is given on the y-axis in Figure 3. The process maintains its functionality when the response is operated above the threshold value, which is the area coloured in green in Figure 3. However, when the response is operating below the threshold, the process is operating in a potentially critical operating domain, which is coloured in red in Figure 3. The functioning of the process is likely to fail when the input factor is below the threshold. The plant is expected to operate without a functional process failure when the input factor is operated within the domain of S_0 till S_1 , but a potential operating failure can occur when the factor is operated between S_1 till S_2 . However, the factor(s) that are dominant in influencing the response and under which operation condition they can lead to a potential FF condition are unknown.

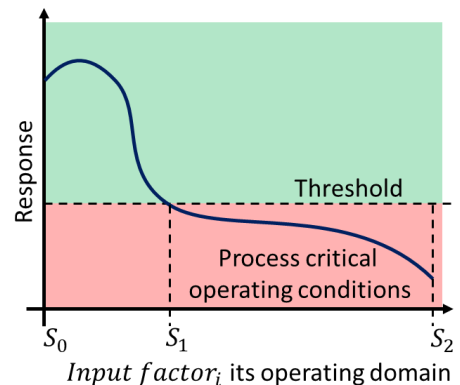


Figure 3. Schematic depicting functional operating conditions and input operating conditions that can potentially lead to a functional failure of the process.

However, researchers caution that special care should be taken when drawing conclusions based on an SA when the process being analysed is poorly known [23][24][25]. Therefore, the functioning and the process of the RMPP and the HEXS model are investigated extensively and included in the scope of the research, as defined in Section 1.4.2. Therefore, the first step in this work is to examine the actual HIsarna RMPP and the HEXS model, specifically with regard to their communalities and difference between them. The aim is to determine the system is functioning and to identify potential limitations and opportunities to utilize the existing HEXS model in conjunction with the research objective.

As discussed previously, Rio Tinto's HEXS model simulates a process behaviour that is to some extent comparable to that of the RMPP system used at HIsarna. However, it has not been validated against HIsarna's operations. Should the HEXS model be found to reasonably reflect the real-world HIsarna RMPP process, it can provide insight into the operations of the actual system. The identified and monitored communalities, such as monitored parameters in the HEXS model and RMPP sensors, will be utilized to **validate** the HEXS model behaviour in relation to the operation of HIsarna's RMPP.

The method presented in this work can already be applied in the design phase, where the assumed process condition can be used to approximate the limitations of the system and identify the factors that can potentially control it. However, if the entire operation space is not fully known, a prognosis with a large uncertainty is obtained. The identified dominant factor(s) and the associated asset(s) offer an opportunity to optimize both design and operation, as they can be used to predict unfavourable operation scenarios. (Unfavourable operation scenarios are those conditions that can lead to functional failure of a system should no measures be taken.) By taking countermeasures such as developing and applying operation and maintenance strategies, these unfavourable conditions can be prevented or can lead to alternative operating strategies. However, this work is limited to identifying the factors and/or assets with dominant influences on the outputs and **potential functional failure conditions**.

1.5.2 Project activities

In its essence, the research strategy adopted in this thesis consists of three main steps: validation, the SA, and determining potential functional failure conditions and were stated in bold text in sub-section 1.5.1. To realize the strategy for the PoC and answer the research question investigated in this work, the following **project activities** are defined and categorized for each step.

(1) Validation:

- Identify the main components of the HEXS process model and how they operate.
 - When identified, to which extent are the components of the HEXS model similar to the functioning and operating of the RMPP?
- Identify communalities between the monitored response pattern parameters in the HEXS model and the RMPP.
- Develop a method to validate the HEXS model in relation to the RMPP by means of a metamodel.
 - Which data is available, and in which form should it be presented to develop a metamodel?
- Develop an experimental approach to investigate the relative difference between a selected response of the HEXS model in comparison to the RMPP for various operational conditions. (*Relative difference gives the scale of difference between the comparative ratio of two factors; absolute difference gives the real magnitude of the observed relationship between the two factors.*)

(2) Sensitivity analysis:

- Develop an experimental approach for characterizing parameter sensitivities under different RMPP operation conditions.
 - Which methods are applicable for determining the response and its sensitivity to varying the input parameters of the HEXS model and the metamodel?

(3) Potential functional failure conditions:

- Determine a potential operational functional failure condition(s).

1.5.3 Scope

The real-world processes of HIsarna's RMPP and the existing process model are considered within this research. There are more than 300 process HEXS model parameters to consider, but these are reduced to one response value, which is used to determine the applicability of the method in question. The response parameter studied for the SA conducted in this research is the MOT, as this parameter is related to a potential functional failure condition of the RMPP; a further benefit is that it is monitored in both the HEXS model and at the RMPP. As the MOT is monitored in both systems, there is a possibility that the model could be validated by comparing the regression trend of the recorded plant data to that of the model. However, providing detailed information concerning the regression supervised machine learning (SML) algorithms and options with regard to optimising the features is not within the scope of this study.

Information concerning the operation behaviour of the physical plant in terms of physical relationships among mill outlet temperature (MOT) is used to validate the process model. When attempting to understand a complex model or system, an SA, an understanding of physics-based relationships, and experience are essential. To identify the potential dominant input parameter(s) influencing the MOT, this work is scoped to study the influence of various plant operation input scenarios in relation to the MOT response.

1.5.4 Limitations

Halfway through the initial research, the research strategy was drastically altered to what is presented in this work. Due to the change in strategy some limitations were already identified, these are stated below. The HEXS process model does not incorporate all the details of complex natural phenomena, and the results might therefore not include the desired level of detail. For example, the model does not include the fracturing of coal due to grinding. This mechanism might be of considerable importance when studying the degradation rate of a component and/or system. In addition, the model output cannot directly be compared to physical plant data due to unit differences, which are not directly convertible (for example % <--> °C).

Limited resources are available concerning topics such as the experience and knowledge of the company's workers, the composition of its workforce, the measurements used, and the computational power available; in addition, limited documentation is available concerning the existing plant and the HEXS process model, which could lead to inaccurate information and incorrect conclusions. No literature has been found that utilizes (thermodynamic-related) steady state process models for (predictive) maintenance. The relatively limited understanding of the workings of the applied HEXS model of the acquired HEXS model and the lack of validation may inhibit its adaption because its benefits are not clear. It is not known whether or to what extent the HEXS model can be validated and whether and how it can be used in an SA.

1.6 Definitions and terms

To ensure clarity some commonly used terms in relation to the content of this research are defined below.

The term **model** refers to a numerical procedure that simulates the behaviour of a process by solving a set of mathematical equations over a spatial domain.

Input factors, or input variables, are any element that can be changed before a model is executed. Examples of input factors include the variables that appear in a model equation, the starting or initial states (baseline settings), and the boundary conditions (limits).

The **Output**, also called the response, is a variable that is the result of a model. It is obtained after the model is executed with inputs to the process, which can be either controllable or uncontrollable parameters.

Physics-based models are essentially bottom-up approaches that enable estimation of future events. To provide accurate representations of reality, these models require a proper understanding of a system's physics and behaviour. The main advantages of these methods are that they can be related to the physical properties of a system, such as changes in flow rates, temperature, pressure, and material properties, and that they are not dependent on historical events in order to predict a future response. The HEXS model is a physics-based process model of an RMPP that also integrates some statistical empirical relations, such as the Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam (IAPWS 1997).

A **data-driven model** (or **regression model**) uses historical data in combination with statistical tools to estimate relationships among variables so that a future event can be forecasted. It has the main advantage that no direct theoretical knowledge about the system or process in question is required should sufficient historical data on how it behaves be available. Using such data, it is possible to learn from experience, meaning that experience can be used to investigate the physical behaviour of systems. However, historical data of the required quality is not always available. In such circumstances, the method, loses forecasting accuracy for future events that are not within the historical design space, as is the case for new or immature systems in which the operating space has not yet been explored.

A **process model** is a mathematical description of a change that systems undergo from one state to another, called a process, and the series of states through which a system passes during the process, called the path of the process. To describe a process completely, one should specify the initial and final states of the process, as well as the path it follows and the interactions it engages in [26]. Such a system is summarized in the form of mathematical process model. A process model is thus a set of operations, including the flow of execution "paths". These enable the establishment of virtual experiments to examine the effect of a given input on an output. A process model can be physics-based or data-driven; alternatively, it can take a hybrid form that combines the two.

1.7 Thesis outline

To accomplish the stated research goal, a research outline has been developed, which is illustrated in Figure 4. The figure graphically presents the structure of this work, which is divided into six chapters, and depicts the interactions among the various sub-steps involved in this research. The first chapter is an introduction to the research topic, the motivation, and the research objectives, as well as the accompanying research questions and strategy. In the second chapter, an investigation is performed into the available resources (the HEXS model and the RMPP), as well as their functioning and operating. Chapter 3 discusses the experimental design and the methodology. Chapter 4 applies the methodology developed for constructing the metamodel. As noted previously in the project activities sub-section 1.5.2, the research consists of three main disciplines. The experiments are executed and the results of the disciplines, i.e. validation, sensitivity analysis and potential functional failure conditions, are presented in three sections in Chapter 5, followed by three sections for each discipline in which the results are discussed. Lastly, in chapter 6, a conclusion is formulated and recommendation for future work are presented.

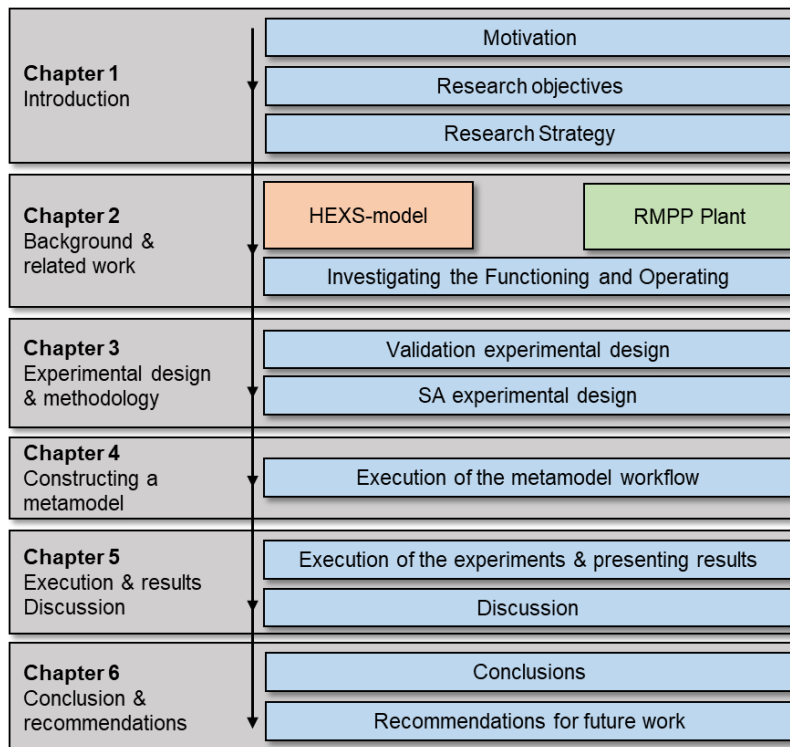


Figure 4. Outline of the thesis and the main steps in each chapter.

2 Background and related work

As stated previously, researchers caution that drawing conclusions based on an SA should be done with care when a process is poorly known [23][24][25]. To obtain an understanding of the processes involved in the RMPP, this chapter describes background information to reduce the risk of a potential false conclusion during the SA.

This chapter first establishes a generic overview of the RMPP, its process and functions. Secondly, a detailed analysis is given of the sub-system involved within the RMPP. A similar procedure is performed for HEXS model. Then the communalities between the actual RMPP operation and the HEXS modelled behaviour are investigated. Whereby a research scope is formed for that part that is common between the two.

2.1 The HIsarna raw material preparation plant

An RMPP, also called a coal preparation plant (CPP), is located within the HIsarna plant; this RMPP began production with logged data on 24 February 2018. Since then, there have been ongoing operational and design modifications and improvements, such as of the feed rate control through the replacement of the vibratory feeder with a screw conveyor. The RMPP removes undesirable material from run-of-mine (ROM) materials by means of a separation process that can differentiate between physical surface properties and impurities. Run-of-mine materials such as chunks of coal or charcoal are ground down to fine powder within the milling system, and they are then pneumatically conveyed by the preheated air/fuel gas mixture through a pipe network for storage in silos before utilization in the HIsarna iron-making process. Filtered, ground, and dried raw coal is used as a material reductant to reduce iron oxide to iron, lower the melting point of the iron, and increase heat (thus providing energy for the process), and coal can be utilized for controlling slag foaming within the HIsarna iron-making process. The RMPP was built to not be dependent on external coal preparation suppliers. However, the RMPP is used for experimentation with different materials types, such as various coal compositions and charcoal, which are thus used in the HIsarna plant. Nonetheless, the RMPP is not designed to operate with charcoal; this is thus an off-specifications condition based on the initial design of this facility.

The main functions of the coal preparation plant are as follows:

- filtering raw coal into an acceptable contamination content (sufficient purification),
- drying the coal to an acceptable moisture content and distribution,
- screening and crushing the coal to an acceptable particle size and distribution, and
- producing a sufficient mass rate of reproducible conditioned coal.

Coal filtering is required for equipment operating conditions, and impurities in the coal used will limit the quality of the iron, which should ultimately become a high-quality steel. The raw material received from coal suppliers can have undesirable inclusions, such as stones, concrete, and bags. A defined and repeatable coal particle size distribution (PSD) is specified. The coal is dried to reduce its moisture content below 2%. The PSD and moisture content must be within specifications to ensure process stability. The required capacity of coal drying is defined as follows: a minimum wet production of 1,500 kg/hr, a normal wet production of 6,000 kg/hr, and a maximum dry production of 10,000 kg/hr [27].

2.2 Functional description of the RMPP

The first step of the project activities as described in Section 1.5.2 is to identify the main components of the HEXS process model and to determine how they operate and to what extent they are similar to the functioning and operating of the HIsarna RMPP. As stated previously, researchers caution that drawing conclusions based on an SA should be done with care when a process is poorly known [23][24][25]. As the researchers were unfamiliar with both the RMPP and the HEXS model and how they function, both were thus thoroughly investigated, with the aim being to reduce the potential of drawing false conclusions.

Section 2.1 describes the main aspects of the RMPP process. To obtain a more detailed understanding of the RMPP in the section, a global description is provided of the materials used, the nature of the production process, and the various subsystems of the RMPP. Some subsystems are included in the HEXS model, which is discussed in Section 2.3 and Appendix 2. The information presented in this section was obtained through discussions with operators and process technologists and consulting design specifications, checklists, and plant guides. No references are made to these documents, however, as they include sensitive or confidential information.

2.2.1 Production process

The product process of the RMPP is explained by means of a schematic, which is presented in Figure 5. The numbers between the brackets (“{ }”) correspond to particular subsystems of the RMPP. The RMPP performs the continuous grinding and drying of raw coal into dried granular coal until the two silos of the granular coal injection (GCI){18} plant of the HIsarna pilot facility are filled. If the GCI silos are filled, the RMPP interrupts its process; it will then resume production when the coal level falls below a threshold value. Upstream from the RMPP, the raw coal charging (RCC){1-3} plant supplies material to the raw coal supply silo (RCSS){4} used as a buffer between both plants. The RMPP {4-15} plant supplies granular coal to the granular coal preparation (GCP) plant in the granular coal supply silo (GCSS) {17}. The GCP plant pneumatically transports the granular coal to the two silos located in the GCI {18} plant. The granular coal is ultimately consumed in the HIsarna process.

2.2.2 {1} Raw coal storage yard

Raw coal is stored in an external yard {1}, which means that it is exposed to environmental conditions. The raw coal is fed into storage bunker {2} multiple times a day by means of a shovel machine when HIsarna is operating. When the shovel machine unloads the raw coal into a storage bunker, it undergoes various filtering processes before it is fed into the feed bin {4} and subsequently into the impact dryer mill {5}, which is a subsystem of the RMPP.

2.2.3 {2–3} Raw coal charging plant

The raw coal from the storage bunker {2} is transported by a belt conveyer to the screening plant {3}; it features a vibrating grizzly screen, which should reject media larger than 50 mm. The media, which can be up to 50 mm in size and may feature long objects, is transported upward by another belt conveyer equipped with an over-belt ferromagnetism and a metal detection system. The magnetic over-belt is used to detect ferrous pieces; it removes ferrous materials from the raw coal and feeds it into a container. Downstream of the magnet, a metal detector detects metallic materials. The detector is used to control the position of the two-way diverter flap, which serves as a reject valve should metal be detected and diverts it into the container. The remaining media is fed into the raw coal supply silo {4}, also called the feed bin, which acts as a buffer of raw coal for the RMPP plant. The feed bin is equipped with level indicators, which indicate when the supplied media reaches a certain level or height within the feed bin. These indicators also show whether conveyors should feed the bin or not. In addition, weight sensors are also present on the feed bin; they provide an estimate of the amount of coal mass stored in the bin. To avoid fresh air aspiration into the RMPP circuit, the raw coal level within the feed bin must be maintained above a certain height, which is measured by the lowest level indicator

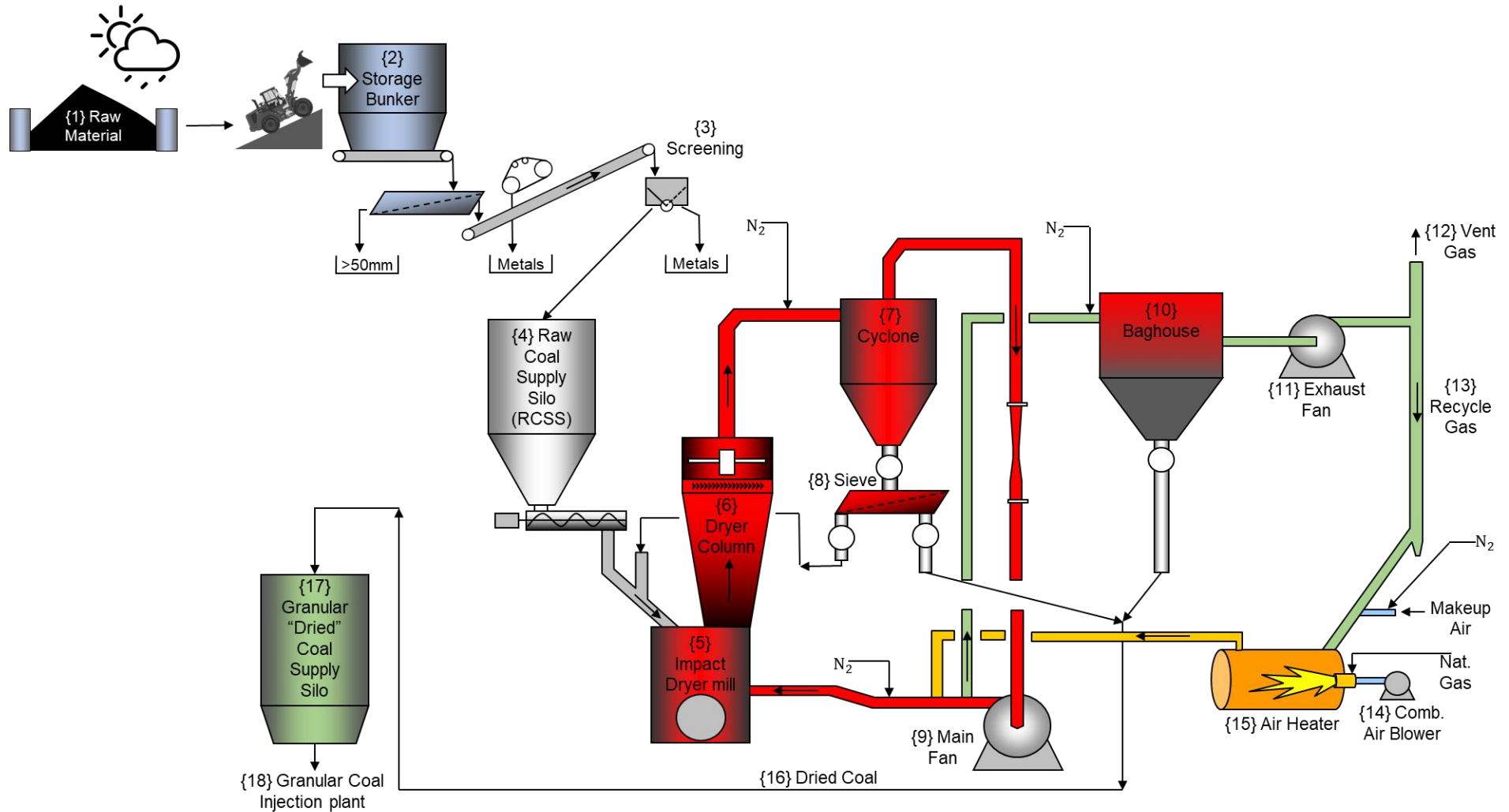


Figure 5 Schematic representation of the RMPP process. In the left corner is the raw coal storage yard {1}, followed by the raw coal charging plant {2–3}, which filters the coal and supplies it to the raw coal supply silo {4}. The raw coal is conveyed towards the impact dryer mill {5}, where it is ground, and mixed with the drying medium (drying gas); this is part of the main drying gas circulation loop marked in red {5–9}. The coal is transported by the heated air through the chevrons of the drying column {6} towards the spinner separator, followed by the cyclone {7}. Within the cyclone, the medium is separated into three directions: Large particles fall through the sieve {8}, and, if rejected, they are returned towards the impact dryer mill. Otherwise, they are collected in the granular coal supply silo (GCSS) {17}, and the finer particles are pneumatically transported towards the main fan {9}, where they are either transferred to the impact dryer mill or flow into drying gas recirculation {10–13}, which is marked in green towards the baghouse {10}. In the baghouse, fine particles are collected in the filter sleeves. Periodically, a purge of nitrogen breaks the cake of fine particles, which by gravity fall towards a rotary valve and are pneumatically transported towards the GCSS {17}. The process air at the outlet of the baghouse can either flow into the environment as vent gas {12} or be reused as recycled gas {13} should it be heated by the air heater {15} before re-entering the main drying gas circulation loop.

2.2.4 {4-9} Coal preparation plant: Main drying gas circulation

The following section describes part of the main drying gas circulation (MAIN). This circulation is colored in red in Figure 5. Depending on the coal demand from the GCI plant {18}, the feed bin {4} is emptied by gravity into the screw conveyer, which conveys the raw coal into a pipe section. This pipe section mixes two streams before they enter the impact dryer mill {5}. The raw coal from the screw conveyor can be mixed with another stream of process gas and coal which has been rejected by the sieve {8}. An overview of the mixing chamber and the impact dryer mill (mill) is presented in Figure 6.

The raw and rejected coal is fed into the impact dryer mill, where it falls by gravity onto the breaker plate and the rotating hammer, which subsequently mills the coal. The impact dryer mill chamber has two inlets where the raw and rejected coal is mixed with heated gases from the air heater, also known as the burner. The surface area of the particle size and volume of the chunks of coal are reduced by the pulverizing action, by which drying is accelerated in comparison to unbroken particles. In other words, the inherent moisture content of the coal is reduced by evaporation due to the heated drying gas. The mill is a particular part of the RMPP that is examined more extensively and thermodynamically modeled in Appendix 3, to obtain an understanding of the drying mechanism involved in the RMPP.

The upward flow carries solid coal particles from the mill chamber through a chevron filter of the dryer column {6} towards the spinner separator, where inherent moisture is continuous, and which is mounted on top of the mill. Based on their settings, the spinner separator and the chevron influence the coal particle size that will be allowed to enter to the cyclone collector {7}. Oversized particles of coal should fall back onto the grinding components for further size reduction. To control its rotational speed, the spinner separator is driven by a frequency-controlled motor, which allows adjustment of the particle-size distribution of the granular coal output. Lower speeds (lower frequencies) result in particle size distributions that on average consist of smaller particles [28]. Higher frequencies result in more revolutions of the spinner and increase the average particle size.

The granular coal is carried by the drying gas from the mill through the spinner separator, where inherent moisture continues to be evaporated, before passing into the cyclone collector {7}. The mill outlet temperature in between sections {6} and {7} is measured to ensure that enough heated drying gas is added for the evaporation of the moisture. This temperature controls the air heater capacity, which heats the drying gas. Within the cyclone collector, most of the granular coal is separated from the drying gas by the centrifugal effect and is fed into a rotary valve arranged at the bottom, which is connected to a vibrating sieve {8}. The differential pressure (dp) is measured at the cyclone, the dp is used as a blockage indicator. If the dp increases, excessive material will build up within the cyclone or at the sieve. The granular coal that is sieved through a vibrating granular coal screening machine (the sieve) will then be fed into the GCSS {17}. Sieved coal particles larger than 2.5 mm are returned to the mill inlet, where they can be mixed with raw materials before they are reinserted into the mill {5}. Dried particles smaller than 2.5 mm are transported by means of gravity to the granular coal supply silo {17}.

Downstream, in the cyclone collector, the drying gas is sucked into the main fan (MFAN) {9}. One part of the drying gas flow is sucked into the main drying gas circulation through the main circuit. The other part is fed towards the baghouse filter. The function of the main fan is to supply heated gases to the drying column {6} for drying moisturized “wet” coal and pneumatically conveying the material through the drying column. The volume flow rate of the main drying gas is controlled through the control damper arranged upstream the main fan. Based on practical experience, it is known that the drying gas flow rate and the spinner separator speed are key factors in controlling the particle size distribution of the coal product [28].

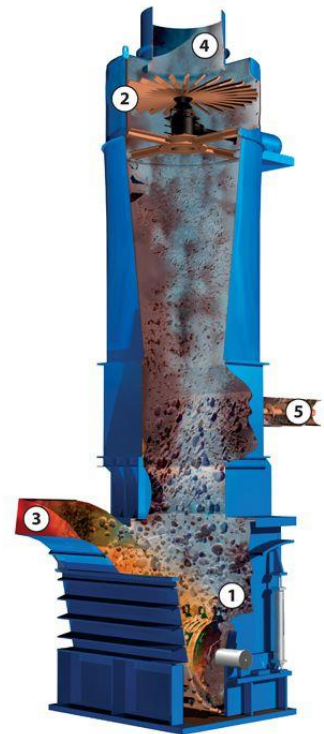


Figure 6. Standard Williams impact dryer mill, without the chevrons that are included in the HIsarna RMPP before the spinner separator. 1) Fluid bed drying and grinding chamber with the rotating hammer mill; 2) adjustable spinner separator (with adjustable speed and number of blades); 3) drying gas; 4) outlet of the impact dryer mill; and 5) input of from the feed bin and returned media from the sieve. Reprinted from “Standard System” by Williams, 2019 [64]

2.2.5 {10–15} Coal preparation plant: Drying gas recirculation

While the previous section described the various components of the main drying gas circulation (MAIN) loop, the focus of this section shifts to describing the systems involved in the recirculation of drying gas (RECIRC)".

The drying gas recirculation process is the second loop and is coloured in green in Figure 5. Its main function is to remove the moisture from the drying gas and to collect the fine coal particles that have not been collected at the bottom of the cyclone collector. Fine coal particles that exit the top of the cyclone and are not transferred towards the granular coal supply silo or the mill are conveyed towards the baghouse filter {10}. They are recirculated towards the baghouse filter by means of main fan {9}. The baghouse separates the fine coal particles from the exhaust drying gas through a filter, which rejects particles that are too large to pass through. Over time, the rejected particles form a cake on the external surface of the filter sleeves, which is periodically back-flushed to remove the trapped materials. This back-flushing is performed by an automatically controlled procedure of purging nitrogen through injection pulse jet valves every few milliseconds. During the nitrogen pulse, the filter sleeves swell, which enables the agglomerates of fine coal particles to fall off, which is required to be able to continue filtering. These fine particles are then fed through a rotary valve into the GCSS {17}.

The automatic back-flushing cleaning sequence starts cyclically if the differential pressure between the dirty gas inlet duct and clean gas outlet duct of the baghouse is within an acceptable range. If the differential pressure becomes too high, the sequence will start earlier to maintain the efficiency of collecting particles at the filter sleeves. Finally, the drying gas flows through the baghouse filter outlet duct towards the exhaust fan {12}.

One part of the exhaust gas is recycled through the air heater to reduce the heat power required to dry the ground coal and evaporate water. The mixture of the waste combustion gases with the clean process gas contains a low oxygen content, which allows the RMPP process to become a self-inert process. The unrecycled part of the clean process gas is exhausted through the stack (**vent gas damper**) and makes it possible of eliminating moisture from the clean process gas. This is required because, should the system not be allowed to eliminate moisture from the RMPP process gas, the process gas can become (super)saturated with water vapour. If the system's process air is saturated, no moisture can be evaporated from the coal [29]. At this condition, any drop in the temperature of the drying process air will cause condensation of any water vapour in the system.

The drying gas is the second loop involving drying gas; it is intended to remove any moisture and the fine coal particles. As the recycled drying gas is pre-heated, less energy is required to heat up the make-up air, the amount of recycle gas can be controlled by the position of the **recycle gas damper**. The recycled drying gas {13} and make-up air is mixed with the waste combustion gases inside the air heater {15}. The air heater is the energy source that heats up the drying gas; the heater is placed in the recycling line of the exhaust circuit. The recycled drying gas {13} flow is heated up inside the air heater before entering the main fan {9} and is split afterwards, with a small volume rate flowing to the baghouse {10} and a large volume rate to the mill {5}. The air heater heats up the drying gas by mixing it with waste gases produced by the combustion of natural gas. As the waste gases contain a small amount of O₂ content, they contribute to the self-inert process of the RMPP plant.

2.2.6 Inert conditions and oxygen content

At nominal capacity, the entire RMPP plant should operate under inert conditions, which are achieved by mixing the recycling drying gas with the combustion waste gases from the air heater and sealing the circuits from environmental air intake. The reason why the RMPP is considered a self-inert system during operation is that the O₂ content in the drying gas is usually below 7–8%. Even though the entire RMPP circuit operates under inert conditions, the inert process provides the option of adding pressurized nitrogen to the drying gas circulation at several locations. Since the use of enormous quantities of inert gas such as nitrogen can be expensive, it is important to limit the consumption of inert gas; to do so, effective sealing of the system from the external environment is required.

An inert environment can reduce the air-to-gas ratio, which makes flame propagation in the coal dust cloud impossible. This reduces the risk of a coal dust explosion within the RMPP. The O₂ content of the drying gas is continuously monitored by two redundant gas analysers integrated into the recirculation loop at the baghouse filter outlet of the RMPP. Should one of the two gas analysers detect an O₂ content of more than 7% in volume, it will set off an alarm at the HIsarna control room. According to Paul Wurth, the original equipment manufacturer (OEM) of the drying column and the mill, 12% is used for standard coal, but several types of coal can be used in the RMPP. Therefore, it has been decided to use a stricter standard of 7% to trigger the alarm. If the level reaches 9%, all processes are immediately stopped, and nitrogen is automatically released throughout the whole plant until O₂ levels fall below the threshold. (The O₂ threshold levels are defined according to the VDI-2263 standard.) At an O₂ volume of 7%, it can be chosen at the human-machine interface (HMI) to either automatically or manually reduce the levels; the HMI also allows an operator to select the purging location where the pressurized nitrogen will be added to the drying gas. There are several locations where nitrogen can be independently added to the RMPP drying gas circulation (as long as the temperature inside the circuit is higher than 60 °C): before the mill inlet from the combined main and recirculation loop, the inlet of the cyclone collector, the baghouse inlet, and before the air heater.

The oxygen content of the process gas used in drying gas circulation in the RMPP is mainly influenced by the following factors:

- the flow rate of combustion waste gases coming out of the air heater,
- the amount of steam extracted from the raw coal through water evaporation,
- the amount of leaking air entering the RMPP circuit, and
- the makeup air flow.

2.2.7 Water spray

During normal operation of the RMPP at nominal capacity, a water spray system is generally on stand-by. However, when the raw coal moisture is low (for example, during a summer period), the heat output power requested of the air heater may be at the minimum ratio. Therefore, it becomes useful to spray some water on the fed raw coal to increase the air heater heat output and avoid overheating the RMPP plant. Furthermore, the water spray system could be also used when the RMPP plant has difficulties in maintaining a low oxygen content in the drying gas.

The water spray can also be used to rapidly reduce the mill outlet temperature before the maximum temperature is reached, which will lead to an interruption of the operation of the RMPP. However, the water spray system is not considered in this work, as it was not utilized any of these conditions or functional operating conditions during the period of time in which this dissertation was written.

2.2.8 Makeup air flow rate

The makeup air adds fresh environmental air to the RMPP; the makeup air pump is located between the air heater {15} and the recycle damper {13}. It has several functions, but its main function is reducing the dew point temperature of the drying gas. As long as the drying circulation gas temperature is higher than the dew point temperature, no condensation occurs. Therefore, keeping the dew point as low as possible minimizes the risks of water condensation in the RMPP, and a stable dew point temperature also indirectly impacts the residual moisture of the granular coal. Therefore, the purpose of the makeup air flow rate is to achieve a compromise between a dew point that is as low as possible and the need to maintain inert conditions inside the CPP.

2.2.9 Operation of the RMPP

The RMPP operating runs of February and March 2019 are considered within this work. However, it should be noted that operating procedure of the RMPP is subject to continuous modification. Therefore, the behaviour of the plant in February 2019 might significantly differ from the operational behaviour exhibited during March. The modifications that were made to the plant during the period in which this research was conducted are discussed below. The data and information of these two operating runs are utilized in Section 0 to construct a metamodel.

In the run of February 2019 (campaign E2.16), the RMPP operated at intervals of 12 hours. At the end of each 12-hour interval, the RMPP shut down and the sieve {8} and the inlet {4-5} of the VF-822 coal feeder to the mill were inspected and, if necessary, cleaned. The hammers of the hammer mill were inspected every 72 hours, which was at every sixth 12-hour interval. If the two silos of the GCIP {18} became saturated before 12 hours had elapsed, the RMPP was taken out of service and the sieve and feeder were inspected. When the coal provision of the GCIP was reduced, the RMPP could continue its operation. The first period of 12 hours was conducted with a coal inlet feed of 75%, a spinner separator set at 60 Hertz, and the main flow rate of the main drying gas circulation set at 17.5k Nm³/h. During the second interval, the inlet feed was increased to 80%, and the setpoints of the spinner separator and the main flow rate remained equal to the previous settings. At the start of the third period (36 hrs), the coal infeed was increased by 5%, as a result of which it reached a level of 85%.

In the run of March 2019 (campaign E2.17), several RMPP design adjustments were made in comparison to the run of E2.16. In this run, nitrogen purges were only used at the inlet before the burner – previously, multiple locations were used to purge nitrogen, which was done in no specific order but rather based on “operator judgment”. The coal inlet feeder (VF-822) from the feeder bin to the mill was perhaps the vibration feeder was replaced with the screw conveyor. During a practical experiment, it was observed that an inlet of the “old” vibration gutter had a setpoint of 80%. This is equivalent to when the “new” worm wheel is set at a feed rate of 40%. At the end of each 12-hour interval, the RMPP was shut down and the sieve {8} and the inlet {4-5} of the VF-822 coal feeder to the mill was inspected and cleaned if necessary.

During the run of March 2019 (E2.17), several operating changes were made in comparison to E2.16; however, the RMPP continued to be operate for 12-hour intervals. At 12-hour intervals, once again the RMPP was shut down and the sieve {8} and the inlet {4-5} of the VF-822 coal feeder to the mill were inspected and cleaned if necessary. The hammers of the hammer mill were inspected every 72 hours, was at every sixth 12-hour interval. If the two silos of the GCIP {18} were saturated before a period of 12 hours was reached, the RMPP was taken out of service and the sieve and feeder were inspected. When the coal provision of the GCIP was reduced, the RMPP could continue its operation. The first period of 12 hours was with a coal inlet feed of 40%, a spinner separator set at 30 Hertz, and the main flow rate of the main drying gas circulation set at 18.5 kNm³/h. At the second interval reached after 24 hours, the inlet feed was increased to 45%, and the setpoints of the spinner separator were reduced by 5% at which the first period had ended, and the setting of the main flow rate remained equal to the settings used in in the first period. At the start of the third period (36 hr), the coal infeed was increased by 5% (reaching a level of 50%); the spinner separator frequency thereof was reduced by 5% by the end setting of the second, and the main flow rate remained equal to the settings utilized in period 2.

2.3 The HIs melt Excel simulation model (the HEXS model)

During the design stage of the HIs melt CPP located at Rio Tinto in Kwinana, Australia, which occurred in (1995-2005), Dr. Rodney James Dry developed the HEXS model. The model was originally developed to provide insight into the operation of HIs melt CPP and was intended to quantify and investigate the capabilities of process and operation parameters such as production rates. The model mathematically describes energy and mass transfers among the various systems and subsystems of the CPP. It uses mass balances and energy flows to calculate the drying capacity that can be attained with a particular amount of energy at certain media compositions and flow rates. Therefore, the model can be utilized to determine process conditions to operate the RMPP, such as material composition, flow rates, temperatures, and other settings.

Version 2.3 of the HEXS model is used as a starting point. This is an undocumented model, meaning that there currently exists no externally available documentation and no comments are included within the model's code. Therefore, this dissertation takes the initiative in presenting the details of this model, how it operates, and how it can be applied.

2.3.1 The HEXS model

This section executes the first project activity as described in sub-section 1.5.2 “Identify the main components of the HEXS process model and how they operate.”. In addition, this section also presents the operational limitations of the model and the HIsarna RMPP. The first project activity is required to understand the HEXS model, with the aim to find an application to utilize the model for PdM.

HEXS V2.3 is an Excel-based model which has five tabs: Gas_DB, Solids_DB, Model, Case, and Mill. In the Gas_DB tab, the chemical composition of gasses, such as environmental air, natural gas, and combustion air, can be adjusted. The composition of several types of coal is stored in the Solids_DB tab. The Model tab stores various calculation “modules”, such as a chemical burner reaction, air humidity calculations, power consumption, material phase state (solid, liquid, gas), material balances, and energy balances throughout the various components of the RMPP. The tab Case is a saved worksheet of the Mill tab; HEXS allows one to save up to 20 unique cases. An example of an HEXS calculation is given in Appendix 3, in which the drying mechanism of surface moisture for a sub-system of the HEXS model, the mill, is analytical calculated. The values utilized in the calculations of the energy and mass balances are derived from the actual HIsarna RMPP operations.

The dashboard of the HEXS model is named Mill; it is presented in Figure 7. This figure describes in detail the HEXS model in relation to the HMI of the RMPP, which is detailed in the Appendix 2. This tab is interactive and allows parameters to be modified; it also provides the output response or “the results” once the model solver button has been clicked. The presented solution is dependent on the solution mode, the coal type, the gas type, the initial conditions of the values of the various parameters, and whether the condenser is present or not. (In this work, the condenser is neglected, as there is not one installed in the physical plant; thus, it is set as not present.)

The model has several types of parameters, which are denoted with colours. User-adjustable parameters are called “inputs” and are mainly in red text, whereas “solvers” are marked in green and use a convergent algorithm to determine one of the solution modes. The “responses” given based on the given inputs are marked in blue, while the “errors” have a pink colour and indicate the performance of the solver and the model. Lastly, parameter(s) highlighted in yellow, which are named “alerts”, are flags that are raised when the boundaries of the model or process conditions are reached. If such an alert is present, it can be manually removed by adjusting the input parameters in trial and error fashion. Table 1 presents a list of possible errors and limitation criteria that should be met for an alert to be raised. These limits are compared with the alerts given in the HMI of the physical coal preparation plant. The HEXS model has more limits with regard to operating conditions, which can result in more alarms compared to the operational system. For example, burner temperature limit allowed by the HEXS model is $<750^{\circ}\text{C}$, that is significantly higher than that at the plant $<550^{\circ}\text{C}$. The model assumes that various conditions, such as the oxygen content of the system, are monitored; at the real-life plant, these conditions are not or only monitored at one location.

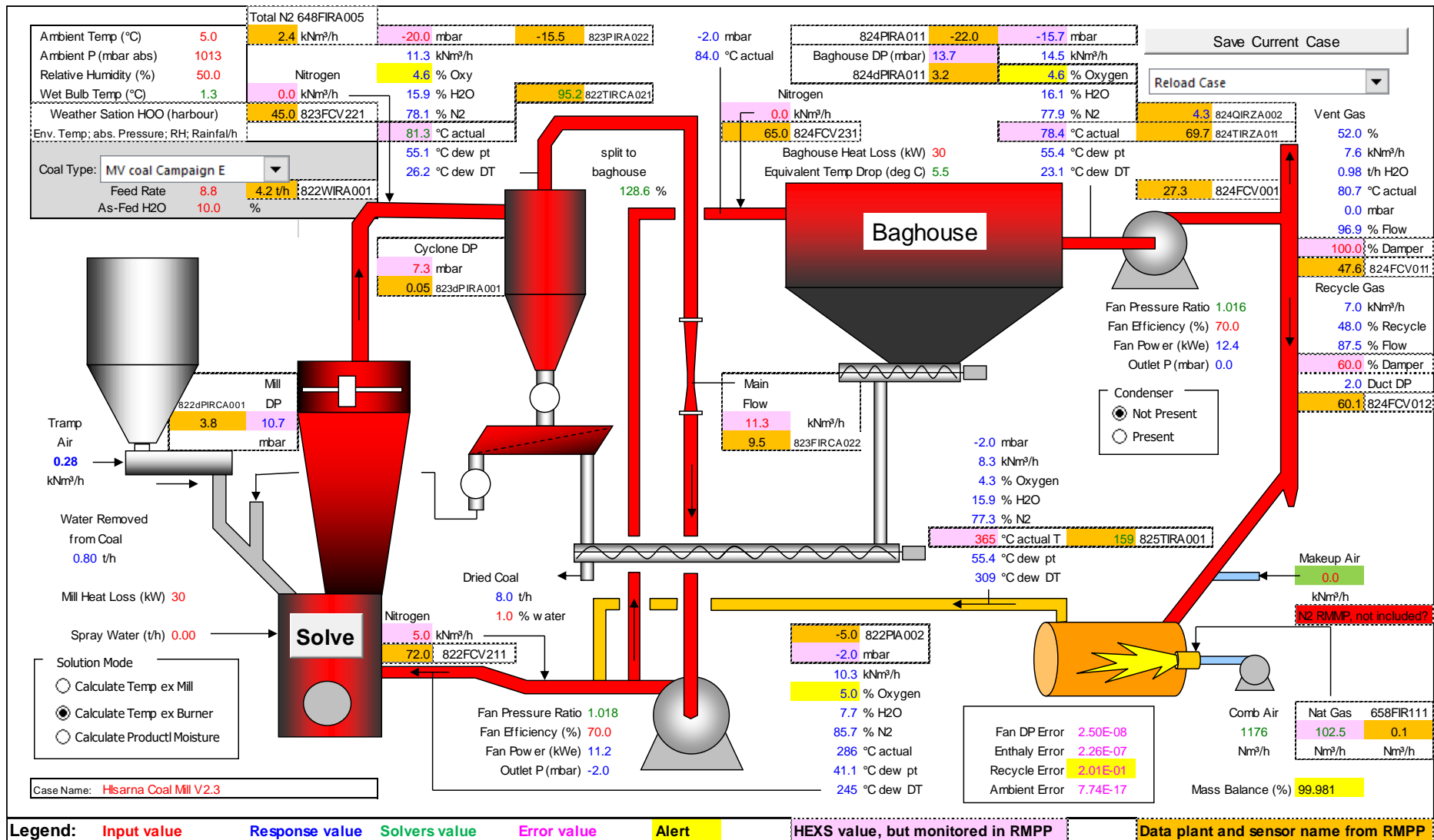


Figure 7. Dashboard of the HEXS model, tab Mill. This dashboard differs from the normally used dashboard in that, within this figure, a comparison is made between physically monitored parameters and the parameters in HEXS. Parameters originally in HEXS are highlighted in pink, and the monitored parameters from the human machine interface of the RMPP are highlighted in orange. The only exception is the data provided by the HOO weather station; this data is not included in the RMPP HMI, rather, it is obtained from the Hoogovenhaven harbour located in Velsen-Noord.

Table 1. An alert will be triggered when any of these limits are exceeded the thresholds for specific location in the HEXS model or the HMI of the physical RMPP; dp is the abbreviation for differential pressure.

Limits of operation	HEXS model	Physical RMPP
The medium drying temperature is below the dew point temperature (Baghouse outlet, burner outlet, mill inlet, and mill outlet)	< 0	
The vent gas total pressure	<-0.001	
Main fan differential pressure error and the ambient error	> 0.00,001	
Mass balance not within limit	If < 99.99 & > 100.01%	
Oxygen level not within limit at main fan outlet	If < 7 & > 11 %	
Oxygen level not within limit at mill outlet and baghouse outlet	If < 7 & > 10.5 %	> 9% O ₂ (only monitored at the BH outlet)
Burner temperature	> 750 °C	< 200 & > 550 °C
Dry bulb temperature at the mill outlet		< 75 & > 110 °C
Differential pressure (dp) between inlet dryer column and mill outlet		dp < -80 mbar
Cyclone dp between outlet mill and inlet main fan		dp > 30 mbar

Version 2.3 of the HEXS model has three solutions modes: calculate the temperature at the exit of the mill, calculate the temperature at the exit of the burner, and calculate the product moisture. Only one mode can be selected at a time. If a mode is selected, the corresponding solver parameters will be transformed from an input parameter into a “solver”. When the Solve button is clicked, the solved parameters numerically converge to produce minimal error, and be within the limitations if possible, as defined in Table 1. The generalized reduced gradient nonlinear solver has a configuration limit of a maximum of 100 iterations, and the maximum duration per iteration permissible is 100 seconds. In extreme cases it can take up to 166 minutes to solve for one configuration case of input factors.

The original HEX model was designed such that it could solve one scenario at a time. By manually adjusting the “input” parameters, the model can simulate one operation condition. When the solution is found, it is up to the user to review the result and determine whether boundaries are reached or operational limits are exceeded as presented in Table 1. By manually manipulating the inputs, a scenario that is within the boundaries can be created, and the alerts will consequently be removed. The limitation of only being able to solve one scenario at a time was addressed in version 4 of the HEXS model, which can automatically run multiple scenarios. With version 4 a larger variety of operation conditions can be explored with minimum manual labour and it allows to save up 1,048,576 unique cases instead of only 20 in version 2.3. This expanded functionality has the drawback that alerts will not be addressed in the simulation; however, they can be evaluated at a later stage.

In the Mill dashboard, materials can be selected by clicking on the triangular symbol above the feeder bin. Coal types can be chosen from a drop-down list; coals of different chemical compositions and featuring different combinations of elements are available. An outline of the adjustable input parameters used in this work is presented in Appendix 6.2.

2.3.2 Evaluation, limitations, and comparison of HEXS with the HIsarna RMPP

This section briefly discusses how the model was validated at Rio Tinto and evaluates how the HIsarna HEXS can be validated within this work and the possible limitations of its application in this work. Validating the model and investigating its limitations are desirable because doing so will make it possible to confirm whether the model can be used for practical applications. Furthermore, doing so will make it possible to answer to partially answer the main research question “How the model can be used for PdM?”

The HEXS model was originally designed for the plant at Rio Tinto. It was slightly adjusted since it was originally developed for use at Rio Tinto with the aim that it could be applied to the CPP at HIsarna in IJmuiden, the Netherlands. At the Rio Tinto plant, CPP operating information was used to validate and adjust the model until it fit, at which point the model was valid to some extent. In an interview with Dr. Dry, he was asked how well the HEXS model fits the physical plant. He was unaware of the degree to which the model was aligned with the CPP. Instead, he briefly advised the interviewer on how to adjust the HEXS model to better fit the HIsarna CPP design. The model can be tuned by adjusting the input “tuning parameters” of mill heat loss, baghouse heat loss, tramp air, main fan efficiency, and exhaust fan efficiency. A similar trial was performed by a technologist at HIsarna. Data from the plant is used as an input for HEXS. It was expected that inputting a certain combination of known physical input parameters and a known HIsarna CPP response would lead to a given HEXS parameter producing a similar output by adjusting the tuning parameters. It was found, however, that, when the plant data was used as the input, the model was unable to present results that demonstrated adequate agreement with the expected response.

The reason for this outcome is believed to be associated with the various design changes that the CPP at HIsarna has undergone and the resulting differences between it and the CPP of Rio Tinto. In addition, the HIsarna CPP is continuously modified with the aim of increasing its performance. These physical design changes or differences have not been incorporated within the HEXS model, which is in all probability the reason why previous attempts to validate the model have failed. The validation procedure that was applied is as follows: First, data was collected from the physical plant after it had been steadily operating for a few hours. Various data sets have been implemented in the HIsarna HEXS model, and the model has been tuned repeatedly to fit it to the CPP. The resulting outputs of the model do not agree with that of the data from the plant. Another reason could be that not all parameters are physically monitored, and fictional values are therefore used in HEXS as a parameter input. The monitored parameters do not have the same units and are not monitored at the same location and therefore cannot be used to obtain a one-to-one fit.

The model is a steady-state process while the plant's state is transient, meaning that any latency in a response of a sub-system that influences another system is not modelled in HEXS. It has been found that if the coal preparation process is not operating and that the equipment temperature is close to the ambient temperature, when the CPP is turned on for operation it will not be heated up immediately to dry the coal. There exists a delay between the burner turning on and the plant circuit heating up. These types of delays are also found in the data used for validating. In this work, validation is performed by comparing the information in a relative sense, meaning that the difference in gradient sign will be compared instead of the gradient magnitude. It is expected that, despite the fact that the model and the CPP do not exhibit a one-to-one fit, validation can be performed using a relative method. The main assets involved and required for drying are included in HEXS, and the model's circulation loops are equal to that of the RMPP.

Other difference between HEXS and the CPP can be identified in Table 1, which presents the limits of operation of both HEXS and the RMPP. In HEXS, the difference between the actual temperature and the dew point temperature at various locations is compared. If the difference is less than zero, condensation will occur, and no liquid can be removed from the coal. As this is unwanted in a drying process, an alert is given should this scenario occurs. In the RMPP, the temperatures of the burner, outlet mill, baghouse inlet, and outlet are monitored; they should operate within a range that has been determined based on previous experience. Note, this limit as presented in Table 1 can be utilized as potential functional failure conditions of not being able to dry coal in the RMPP operations. In this work this limit in the first row of the Table 1 will be examined in Section 5.5 for factors that can lead to this potential functional failure condition.

Comparing the RMPP's flow rates, temperatures, composition analyses, and pressures with those of HEXS is challenging. In HEXS, these parameters are not influenced by transitions from one sub-system to another. Instead, the model takes the steady energy and mass balance of the whole. In practice, geometrical changes caused by pipes, transitions, contact boundaries, and locations of sensors do influence the behaviour being monitored. In addition, the units used are not equal. For example, the flow rates of mediums within the RMPP are regulated with regard to opening of valves, which is expressed as a percentage. In HEXS, the unit is the volume rate of the individual gas composition per unit time. The greatest difficulty is that the amounts and flow rate of all media are not monitored. Therefore, the mass entering and leaving the system is not known. Although, ideally, all other parameters are monitored, there will never be an equal one-to-one

fit result between the HEXS and the RMPP operations; due to the inherent randomness of a real physical process. However, when comparing the results of the RMPP and the HEXS model in a relative sense, an equal result is not required, which is explained by the following example: If all conditions stay similar and only one parameter is adapted in this illustration, the burner temperature is increased. It is expected that the overall temperature will increase in the system. A similar trend is expected to occur when the burner temperature is increased of the RMPP. This relative comparing can only be performed as long as no automatic regulation system interferes, which may nullify the response. Thus, the limits for the RMPP in Table 1 should not be exceeded. These limits set in the RMPP are continuously adapted as HIsarna's workers obtain more experience of operating the plant as presented in Table 1. The presents list of certain conditions of Table 1 that are monitored in the RMPP and are utilized within this work; these will be further discussed in the validation procedure.

2.4 Introduction to drying

In Section 2.3, it was stated that the HEXS process model is a heat and mass transfer model. The purpose of the HEXS model is to estimate values for process parameters in which wet coal can be dried to a specific degree by the RMPP. This section provides a brief review of the drying of solids and the physical mechanisms that are involved in extracting moisture from solids by convective drying. Drying is one of the functions of the CPP. Other RMPP functions, such as the screening and grinding of coal, are not discussed because these mechanisms are not included in the HEXS process model; however, these mechanisms affect the drying process.

2.4.1 Introduction to the convective drying of porous solids

Drying is described as the removal of moisture from solids, gases, or liquids. The drying of solids involves separating a volatile substance (moisture) from a solid material by vaporization. Vaporization is the transitional phase change “conversion” of a substance from the liquid or solid phase into the gaseous (vapour phase) by the application of heat [30]. Heat can be applied by means of convection (direct drying), conduction (contact or indirect drying), radiation, electromagnetic fields, or a combination of these methods.

The characteristics of drying or vaporization depend on how moisture is held within a material. The storage of moisture is schematically represented in Figure 8. The liquid loosely adhering to the surface of the material, which is called the “unbound” moisture, is marked in dark blue [31]. This unbound moisture can freely move and exchange liquid surface moisture when it, for example, encounters a neighbouring solid. This surface liquid is the first to be removed by vaporisation (stage I in Figure 8). Once this liquid has been

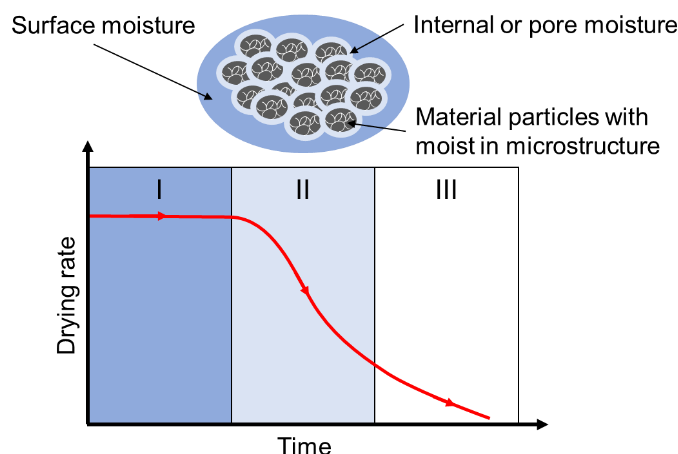


Figure 8. At the top of the figure, a solid particle with moisture is presented. The dark blue outer layer is the unbounded surface moisture, the light blue is the bound internal or pore moisture, and the black is the bounded moisture within the solid (its microstructure). Below the particle, a simple form of a drying characteristics graph of a wet solid is given, divided into drying sections (I–III): The drying rate of the moisture content is presented on the y-axis as $[\text{kg}(\text{water})/\text{kg}(\text{dry solid})]$, and the x-axis represents the drying time. The dark blue is the unbounded surface moisture that will evaporate first (stage I); the bounded water within the capillaries and the pores, which is coloured in light blue, will subsequently be removed in stage II; and, finally, the crystal water which is bonded into the crystal structure of the material will be removed in stage III. Adapted from: *Drying of Solids Materials*, E. Tsotsas, et al. (2005) [33] and *Drying and Evaporation*, Gunt Hamburg (2009) [32].

removed, drying of the moisture contained within the capillaries and pores of the material begins [32]. The liquid vapour interface (the drying front) moves inside the porous solid. The dried surface of the solid thermally insulates the moist inner portions of the solid [31].

Simultaneously, the dried regions impede the transport of the vapour to the bulk of the drying medium [31]. The capillaries and pore moisture is marked in light blue; this is the “bound” moisture, which is present as liquid solution within the solid or even trapped in the microstructure of the solids [31][33]. The drying rate decreases due to the need to overcome capillary forces and diffusion resistance (stage II) and tends to zero as all of the unbounded and bounded moisture is removed from the material (stage III) [32]. The moisture that is bonded into the microstructure of the material can only be removed by intense heating in combination with a low drying rate, as shown in stage III [32]. The liquid moisture to be removed is usually water, but it could also be a solvent, such as ethanol, methanol, or a mixture of such solvents.

In the context of this work, the term “drying” refers to the removal of liquid surface water (H₂O) from a product by vaporization by convection. In the context of this thesis, the solid material considered is coal, which is a hygroscopic substance, meaning that it has the ability to absorb moisture from its surroundings.

In general, drying is a heat and mass transfer process in which moisture is separated by evaporation from a solid, semi-solid, or liquid [34]. During vaporization, the liquid is transferred into a vapour phase by the application of heat. This heat is added by the dried process air to the wet solid (coal), and the moisture of the coal turns into vapour, as a result of which it separates from the coal and mixes with the unsaturated process gas (the drying medium is air). In turn, the concentration of moisture in the air increases until the driving force is insufficient, or the air is saturated. Mass transfer refers to the movement of a chemical species (the moisture) from a high concentration (wet coal) region toward a lower concentration one (the process air). As there exist two regions with different chemical compositions, this phenomenon is a mass transfer process.

At a macroscopic level, the coal and the process air are treated as bulk fluid in motion, as it is transported from one location to another. The primary driving force of fluid flow is the pressure differential, which, for mass transfer or diffusion, is the difference in concentration. According to Cengel and Ghajar (2011), a fluid in motion enhances mass transfer considerably faster than a stationary fluid [35] by replacing a high moisture concentration near the surface with a lower concentration.

In summary, the process of drying, heat is required to evaporate moisture from the coal, and a flow of air is required to carry away the evaporated moisture. There are two basic mechanisms involved in the drying process: the migration of moisture from the interior of an individual coal particle to the surface and the evaporation of moisture from the surface to the surrounding air [36]. The rate of drying is determined by the moisture content of the coal, the temperature of the coal, the drying medium temperature, the relative humidity, and the velocity of the air in contact with the coal. The capacity of air to remove moisture is primary dependent upon its initial temperature and humidity: the greater the temperature and the lower the humidity, the greater the moisture removal capacity of the air. The relationship between drying and the required conditions can be explained using Mollier and psychrometric charts, an example for a drying calculation of the mill that utilizes these charts is given in Appendix 3.

2.5 Discussion of the commonalities

In this section the commonalities between the HEXS model and the HIsarna CPP in relation to the research aspirations are discussed. There are both differences and similarities between the HEXS model and the HIsarna CPP. The hammer mill and drying column at HIsarna, IJmuiden, is produced by the same OEM as the type that was used in the RMPP for HIsarna at Rio Tinto. This also holds for the HEXS model, as it was originally developed for the RMPP at HIsarna, and, due to the similarities between the model and the HIsarna CPP, it is currently being investigated for utilization at HIsarna. There is a potential to create added value for the SUPREME project and HIsarna. Therefore, it is necessary to identify operating and monitoring commonalities. By investigating the resources at hand – the HEXS model, the RMPP, and its data in relation to the aspirations of the research topic – an area of agreement can be identified, which is shown in Figure 9. The area of agreement in Figure 9 is found to be the mill outlet temperature (MOT). Each of the three areas is discussed below, being model possibilities, research aspirations and data possibilities.

Section 2.1 discussed the RMPP's main process functions. These include processing coal at a sufficient rate of coal, which is screened, dried, and crushed in accordance with the specifications. Unfortunately, the HEXS model is a steady-state process and does not evaluate screening and crushing of the coal in the RMPP. This would be of interest, as the screening classification of particle sizes and crushing of coal influence the operating conditions, which in turn influence the degradation rate of the equipment used. A direct link between operating conditions and the degradation rate of an asset is yet to be identified. Using the available resources, operating conditions for drying can be studied by means of model simulation and utilizing plant data.

Failure is considered to be reaching a state in which a part or a system cannot fulfil its intended function(s) [12]. Functionality is maintained when the specifications are met “that is when a part or system **can** fulfil its intended function”. If the specification is not met, the system fails, and it is referred to as a functional failure. The condition that led to the failure is referred to as the potential functional failure condition (PFFC). One of the main functions of the RMPP is the drying of coal to specification. If the RMPP is not able to dry coal to specification, it will not preserve its drying functionality. As the functionality in this instance is defined with reference to the specification, the system will fail to preserve its drying functionality. The failure condition that can occur is when the outlet temperature of the mill drops below the dew point temperature, at which point drying stops and the evaporated liquid can once again infiltrate the coal. This condition is thus identified as being critical to the process: The mill outlet temperature should be above the dew point temperature. If the difference between the two is equal to or less than zero, then a potential functional failure can occur.

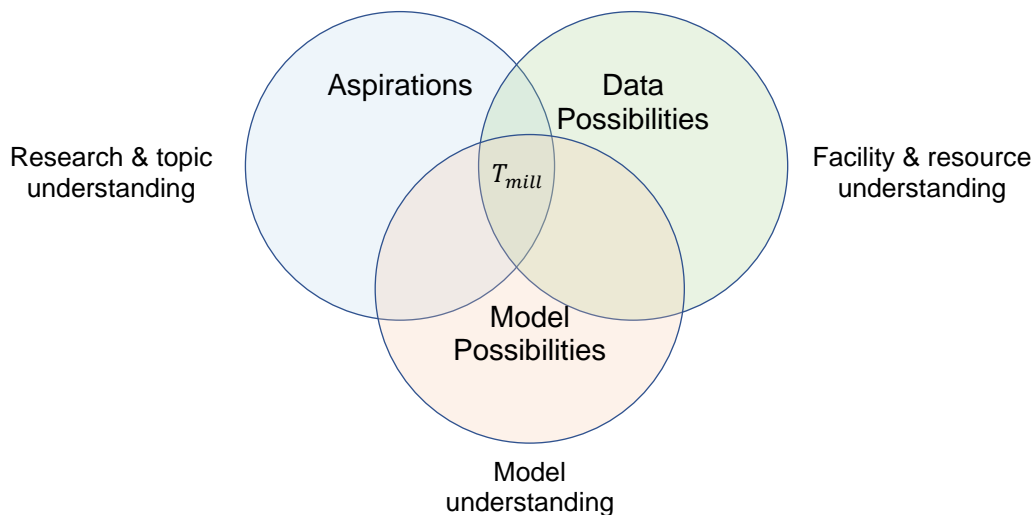


Figure 9. Common ground: the mill outlet temperature.

The mill outlet temperature is identified as the commonality by investigating the functioning and operating of the HEXS model and the RMPP plant in relation to aspirations of this PdM research topic. The mill outlet temperature is related to the potential functional failure condition of not being able to dry coal to specifications. As this temperature is a potential candidate for a functional failure condition, it fits within the scope of the aim of this research. In addition, the mill outlet temperature is monitored in both the physical plant and the model. Therefore, the actual RMPP response behaviour of the mill outlet temperature can be compared to the response behaviour presented in the HEXS model. Thus, a relative comparison between the HEXS model and the plant can be performed. The MOT was found to be the common ground, and thus the MOT was chosen as the response value for further investigation processes. The MOT is used as the response in the sensitivity and regression procedures. Furthermore, the difference between the mill outlet (drying medium) temperature and the dew point temperature is used to determine potential functional failure conditions.

3 Experimental design and methodology

In this chapter, the experimental approach adopted in this research and the design thereof, as well as the relevant theory, are detailed. Each method and theory are bundled as a workflow, which is utilized in the execution of the experiments. First, the validation strategy is established, followed by a section that focuses on theory and workflow. Second, the strategy adopted for the SA is discussed; thereafter, an SA method is selected, the theory underlying the selected method is discussed, and the presented information is then bundled into the workflow procedure.

3.1 Validation strategy

In the global strategy (as discussed in sub-section 1.5.1), it was noted that an SA can be utilized to investigate the influence of input factors on a selected response. If an SA can indicate the relative importance of each factor in the HEXS model in terms of contributing to a selected response, it does not necessarily mean that the results obtained reflect reality. This is because the HEXS model was developed to simulate the process behaviour of a comparable HIsarna RMPP system, namely that of the Rio Tinto RMPP. However, it has not been validated against HIsarna's operations. The SA results can provide realistic insights into dominant factors controlling the system and sub-systems should the HEXS model be found to reasonably reflect the real-world process. In the absence of a validation proof, it does not seem appropriate to perform any further analysis with the HEXS model, as it is unknown whether its resulting output would be of any value or could be applied to the HIsarna plant (i.e. the model may not accurately represent the behaviour of the HIsarna plant). Below an answer is presented to the first sub-question, namely "How can the HEXS process model be validated in relation to the physical response behaviour of the RMPP?"

In Section 2.3, it was stated that the HEXS model applies energy and mass balance in its calculations. Initially, the validation approach would be to derive an energy and mass balance of a small sub-section of the actual RMPP (e.g. the mill). Once established, the derived energy and mass balances could be compared with the energy and mass balances of the HEXS model for validation. Therefore, an extensive thermodynamic analysis was performed on the HEXS model, as presented in Appendix 3. Through this analysis, an understanding was obtained of the inner workings of the HEXS model through establishing the mass and energy balances to be used for validation. However, obtaining the real-world values for these would involve monitoring elements that are currently not monitored, such as the material flow rates. The sensors required are prohibitively expensive and largely unavailable, which resulted in the choice to adopt the data-driven alternative validation method.

To be able to quantitatively compare the HEXS model to the RMPP operations, a relationship that represents the physical behaviour of the actual plant is required. As there are multiple sensors that monitor and record the plant's behaviour, it is possible to formulate a data-driven relationship based on historic sensor data. Regression algorithms are techniques that can be utilized to estimate such a relationship among the input variables. They can be used to fit the initial sensor data to express a physical process or system. These relations are also known as surrogates or metamodels [37]. It is challenging to select an appropriate regression algorithm, but, as is discussed in the following section, machine learning represents a solution to this problem.

When a metamodel that has learned properties from a sensor-generated dataset is established, it can be used to make predictions. These predictions represent the behaviour of the physical plant and are produced from a data-driven metamodel. Predicting the behaviour of RMPP makes it possible to compare the behaviour of the HEXS model outputs with the outputs of the metamodel.

In order to assess the goodness-of-fit for the HEXS model with regard to the physical operation behaviour of the RMPP, a validation was performed. In this work, validation is performed by asking the following (these questions are further explained and answered in Section 5.2):

- (1). Do both models' responses for the mill outlet temperature move in a similar direction for a factor's given operating domain?
- (2). Do both models' responses for the mill outlet temperature operate as expected (based on the expectations identified as a result of an extensive RMPP investigation)?
- (3). Are both models able to identify similar dominant influencing factors on the mill output temperature in the SA?

3.1.1 Regression strategy

A regression analysis is a collection of statistical modelling tools for estimating a data-driven relationship between one or more independent variables [38]. A regression analysis is used to model and explore relationships among variables that are related in a nondeterministic (stochastic) manner [38]. A nondeterministic manner refers to a system exhibiting different behaviours on different executions, even for the same initial inputs. Such nondeterministic behaviour also occurs in the operation of the physical RMPP. Although similar initial settings are used, nondeterministic behaviour will always lead to a variance in the output when the operation is repeated. This is due to the inherent randomness involved in nature. Therefore, the regression analysis finds an approximation of the solution, as opposed to the exact solution.

There exist numerous scientific methods for studying or analysing the relationships between two or more variables, and there are many types of regression algorithms. However, there is no precise procedure for selecting a regression algorithm [39], which makes it challenging to select an appropriate algorithm for any given data. The choice often depends on the data available on the dependent variable(s), the computational expense required for a particular analysis, and which type of regression model that provides the best fit. The best fit means that a specific regression algorithm will have the smallest error between the predicted value and the actual value in comparison to the predictions of other regression techniques. Machine learning (ML) represents a solution to this problem. Machine learning can be used to evaluate several algorithms against a stated criterion. By implementing a criterion (e.g. the best fit), ML can be used to identify the best-fitting model for the given data.

In this work, statistical tools are utilized to model and explore the relationships and interactions among variables of the various subsystems of the RMPP, specifically in relation to the response represented by dry-bulb temperature at the outlet of the mill. The expression created by the regression analysis can be regarded as a transfer function that produces an output for a given input. This metamodel fits the initial data to the actual response to express a physical process or systems [37]. The exact inner workings of the simulation code are not assumed to be known (or even understood), as it is solely the input-output behaviour that is important [40]. (This approach is also known as black-box modelling.) There might be similarities in the trends and influences of the input factors in the two process models (the HEXS model and the metamodel), and some of the relationships within the HEXS model are understood (mill thermodynamic relations as presented in Appendix 3). These relations can assist in unravelling some of the causes and effects within the metamodel and its black box. The results of investigating the models might provide insight into the process and yield a method by which to predict which system is dominant in influencing another system. Such a method could be used as a tool with which to allocate resources.

Machine learning is a process that is used for learning a set of rules from a given input. In the case of this research, the rules considered are those that govern the behaviour of the plant. When ML algorithm is fed with labelled input data, the learning that occurs is referred to as supervised; in contrast, in unsupervised learning, the factors are unlabelled [41]. As the data is labelled, supervised machine learning (SML) is applied. The main steps in the process of regression-based SML, which are applied to the problem investigated in this research, are described in the workflow presented Figure 10. Each step in the figure is numbered as mm “metamodel” followed by the step number; the two blue-coloured boxes in the figure are elaborated on in Section 4.4.

The first step is defining the objectives; in this case, the goal is to identify a relationship based on the plant data that can describe the behaviour exhibited by the mill outlet in terms of its temperature when varying the input. Consequently, the required input data should be identified in the form in which it is retrieved and in its original resolution. The next step involves obtaining, preparing, and pre-processing the data (e.g. determining whether datapoints are missing from the dataset). These steps are followed by sampling the pre-processed data in two sets, a training set and test set. The next step is to train various regression algorithms for the selected response factor and to create a metamodel for each algorithm; each metamodel should then be evaluated with reference to a stated criterion (e.g. best fit) by evaluating the trained model against a test set. Finally, the error between the prediction and the real data must be determined, and the best algorithm should be selected for further use; alternatively, certain steps in the workflow must be repeated until the model satisfies the requirements. This workflow is executed in chapter 0.

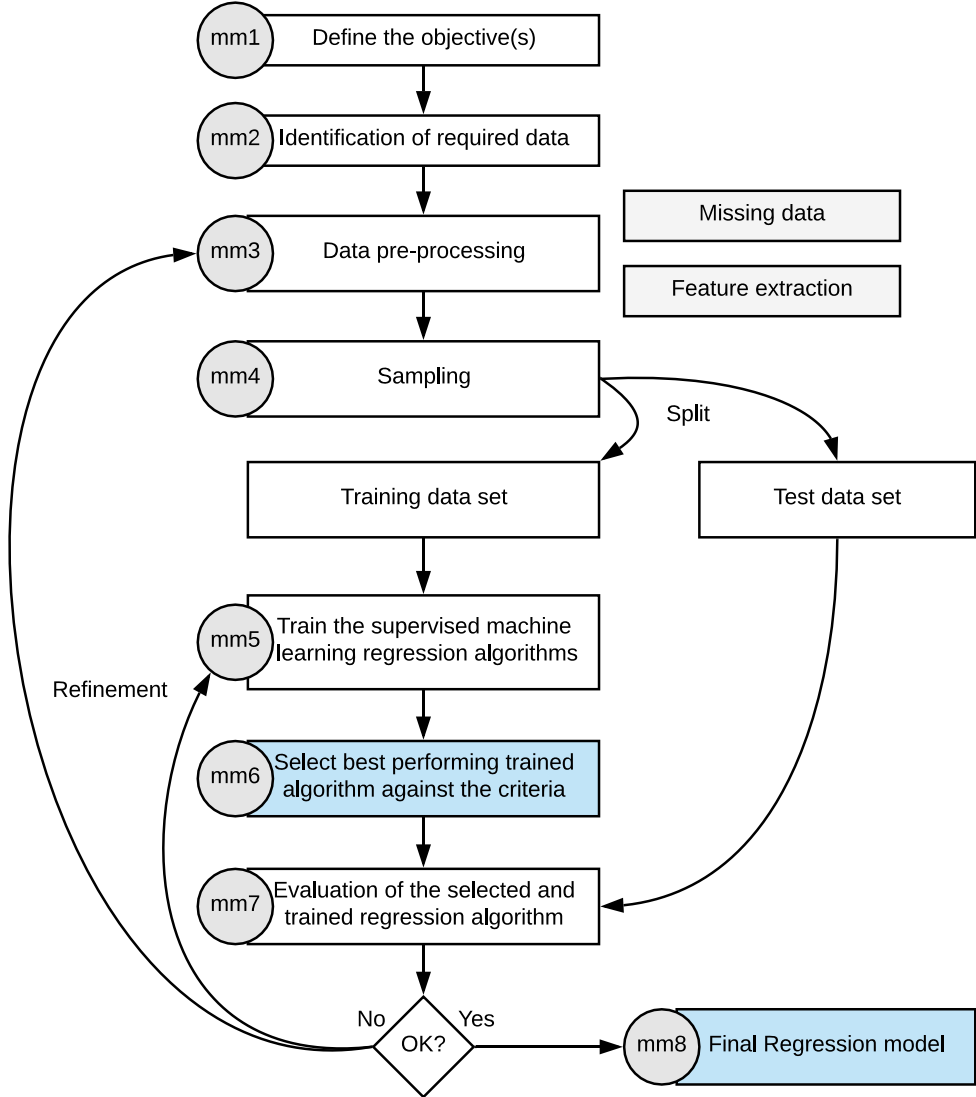


Figure 10. The main workflow for obtaining a regression model, also referred to as a metamodel, adapted from Kotsiantis, 2007 [41].

3.1.2 Evaluation criteria for the trained model

In the previous sections, the steps of the workflow were discussed and illustrated in Figure 10. In steps mm6 and mm7 of the workflow, the regression algorithm is evaluated, each algorithm is evaluated against the others in step mm6, and the best fit is evaluated against a test data set in step mm7. The evaluation criteria used for this purpose are discussed in this section.

Machine learning methods allow large sets of data to be used for the development. In this work, regression SML is applied, as the available training data for each sensor details the sensor type, the specific physical quantity detected, and the location. The desired output, “the mill outlet temperature”, is known precisely, as historical data is utilized to train regression algorithms. As the desired output is evaluable, it becomes possible to directly evaluate the quality of the model using quantifiable and objective metrics.

The root mean squared error (RMSE) is calculated for each trained regression algorithm. The RMSE evaluates how close the actual response data values are to the predicted values [42]. The RMSE is calculated according to equation 3.0, in which \hat{y}_i the predicted value (computed by the metamodel), and the observed value y_i (the actual response value from sensor data) for n , the number of observations. The RMSE is always 0 if the prediction perfectly fits the real response because all of the observed points lie on the predicted trend and therefore no difference exists among them. The RMSE will be always positive, and its units will match the units of the selected response. For the mill outlet temperature, the RMSE is expressed in °C.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.0)$$

The RMSE can quantify how well a metamodel corresponds to the observed data. However, the RMSE does not measure how well it can make a new prediction for a new set of input settings. For example, the error calculated by the RMSE can be reduced if the n number of parameters increases without improving the actual predictive quality of the metamodel. This phenomenon is defined as overfitting [42]; as the metamodel has already observed the data upon construction of itself and is optimised for these observations, it is said to over-fit the data.

To evaluate the predictive performance of a metamodel, a superior method is to test the model with unseen “untrained” data and evaluate how well the model generalises. The dataset on which the models are estimated is referred to as the training set, and the independent set that is used for validation is called the test set. The RMSE can also be calculated on the test set to obtain a quantified error.

The disadvantages of SML are that it requires an independent test set and additional experiments. However, the MATLAB 2018b regression learner tool utilized in this work computes a cross-validation root mean squared error (CV-RMSE). A CV-RMSE measures the quality of the metamodel its predictive performance without the need for this extra evolution and is less sensitive to overfitting than the RMSE. During cross-validation, the model shifts between training itself and validating a fold for a number of times [43]. A dataset is split into k-folds (a fold is a part of the dataset) used for validation. In cross-validation, the full data set is randomly split into a selected number of k-folds. For a k-1 folds, the model is trained, and its prediction power is evaluated against the remaining fold for model evaluation. The regression coefficients are adjusted accordingly to increase the model its predictive power.

Although a CV-RMSE was performed, the best-performing (i.e. which received the lowest CV-RMSE) metamodel’s predictive performance was assessed using the split-set validation method. The steps performed are illustrated in Figure 18 and will be discussed in Section 4.4. The complete dataset was divided according to a specific ratio of a training set and test set for validations. The ML algorithms are trained in no predetermined order using a fivefold cross-validation to minimize initial overfitting. The performance of the selected trained model (trained algorithm with the lowest CV-RMSE) is assessed in comparison to a test data set, where the trained model its RMSE is determined. The details of the execution of this process are presented in chapter 0 and the results are presented in chapter 0.

3.2 Sensitivity analysis

3.2.1 Sensitivity analysis strategy

Sensitivity analysis is a modelling technique used to evaluate how the outputs or outcome of a model may change when that model's input values are changed. Based on the study performed on the HEXS model in Chapter 2, and presented in Figure 5, it was shown that the RMPP system consists of several sub-systems or assets, such as the mill, cyclone, main fan, baghouse, exhaust fan, and burner. Each of these sub-systems has its own inputs and outputs. The outputs of one sub-system may be the inputs of another sub-system section. More than 300 parameters are associated with these RMPP systems thermodynamics of the HEXS modelled RMPP. These parameters include the quantity of an element, matter composition, flow rates, temperatures, pressures, heat losses, efficiencies, critical points, and valve positions.

The philosophy of SA is that if the relationships and relative levels of importance of parameters are understood and the model used reasonably reflects the real-world process in question, then SA can provide insights into the dominant controls of the real system and sub-systems. These insights can in turn be used to support decision-making by addressing questions such as what is(are) the critical factor(s) that lead to system failure (i.e. not drying the coal)? By understanding how a system operates, unwanted process conditions can be identified in advance. These can be accounted for during operation or can be addressed during maintenance procedures to ensure that the system functions as intended.

To study the influence of a selection of the abovementioned parameters on the mill outlet temperature, an SA method is first selected. Thereafter, a literature review is conducted on the selected SA, with the focus being on how it can and should be applied to examine the chosen response. These will be discussed in the following.

3.2.2 Selecting an SA method

Sensitivity analyses have been applied and studied by many authors and SAs have proven their value in many sectors [44][45]. Throughout the decades, new methods and adaptations of existing techniques have been developed. Such new methods and adaptations offer a wide range of potential solutions when attempting to achieve a particular objective. Some of the different objectives of SAs are factor importance and function, regions of sensitivity, factor interdependence, factor and model reduction, and uncertainty apportionment [44].

The objective of performing an SA analysis is to assist in identifying the importance and function of a factor. More specifically, an SA allows one to, through performing a limited number of simulations, identify, prioritize, and screen the factors that are more influential and contribute most significantly to the variability of the selected response (in this case, “the mill outlet temperature”). The objective determines the possible choice of SA methods. However, not all SA models are suitable for implementation in the HEXS model without major modifications to the model's code. To assist in identifying an appropriate SA method, additional requirements are presented below.

Four additional requirements included in the SA method selection criteria are the ability to cope with nonlinearity and interaction, the ability to increase in specific step sizes of the input, and the ability to fixate a factor. These are discussed in the following.

The response behaviour for a given input to the response “the mill outlet dry-bulb temperature” is unknown. The chosen SA method should be able to cope with nonlinearity and interactions, as, otherwise, it could lead to misleading results. These two requirements are added to the criteria, as not all SA methods are able to cope with nonlinearities and interactions [46].

The other two requirements are that the method should be able to increase the factor by a specific step size (levels) and that it should be able to fix factors at a specific value. The input factors used in an SA analysis can be in the form of a variety distributions or upper and lower bounds. These bounds are then sampled into evenly spaced levels p , which are the “step sizes”. For the RMPP, some of the upper and lower operating limits are known, but not their distributions. Therefore, the choice has been made to use evenly spaced levels instead of distributions for the SA analysis.

The RMPP investigations demonstrate that setting input factors at constant values without variation (factor fixating) is desirable, as some of the factors are variable or not utilized. For example, as described in sub-section 2.2.7, the water spray at the inlet of the mill was not being utilized in the actual RMPP. Neglecting a factor can be done in a model by switching a factor to the off position.

Based on the identified objective and additional requirements, a literature review was conducted to determine an appropriate SA method. Saltelli et al. (2008) presented a table that assists in the selection of SA method by desired characteristics, this table is presented in Table 2 [46]. In Table 2, the following SA techniques are compared: scatterplots, fractional factorial, elementary effects, variance-based, metamodeling, and Monte Carlo filtering. In addition, they are evaluated based on the following criteria: coping with nonlinearity, coping with interaction, samples taken from number of input factors, number of input factors, CPU time per run, cost of analysis, and setting for the SA (factor prioritization, factor fixing, variance cutting, and factor mapping).

Table 2 which is adapted from Saltelli et al. (2008), does not show how each technique performs with respect to detecting interactions or influences. The literature indicates that it is challenging to compare the performance of each method [4][5][49]. For example, Nguyen et al. (2015) compared the performance of nine SA methods, including elementary effects (EE) [49]. The authors compared the performance of these methods based on their ranking of factor importance, interaction among variables, variables' sensitivity indices, and computational expense. The authors found that certain methods failed to predict interactions and/or rank the importance of an input factor. However, the authors could not determine if this was due to the techniques themselves or the sampling strategy used in their methodology. That is because methods rely on random sampling strategies; each set of initial input factors per method is unique (for generating the factor combinations where the model response function is evaluated). As each initial input condition per method is unique, it results in variance in output and thus makes comparison challenging. Of the nine techniques, the authors compared three to determine whether they could detect interactions among other input parameters. These SA techniques were a Sobol, a Fourier amplitude sensitivity test, and EE. The authors indicated that, of these three techniques, only EE can detect interactions among other parameters.

As it is not clear which of the methods performs the best when identifying interactions or influences, the comparison presented in Table 2 was used to decide upon a method. From this table, the EE and fractional factorial techniques were identified as the most appropriate methods for the criteria presented previously. However, as EE has lower computational costs than the fractional factorial, therefore EE method was selected for the SA in this work.

The EE method is also referred to as the Morris method. In comparison to the other screening methods mentioned above, this method features sensitivity measures that tend to be qualitative. Therefore, its results can be utilized to rank the input factors in order of importance; however, by itself, the method does not attempt to quantify by how much more important a given factor is than another. To obtain an exact percentage of the total output variance that each factor (or group of factors) accounts for in influencing the response, a quantitative method would be useful, such as the cumulative percentage used in the Pareto principle; this method is further examined in the sub-section 3.2.5 "Extensions of the Morris method".

	Sensitivity Analysis Techniques						
Characteristics of each SA method	β_i^2	Scatterplots	Fraction Factorial	Elementary Effects	Variance-based	Meta modelling	Monte Carlo filtering
Coping with nonlinearity	No	Yes	Yes	Yes	Yes	Yes	Yes
Coping with interactions	No	Yes	Yes	Yes	Yes	Yes	Yes
Samples taken from	Distributions	Distributions	Levels	Levels	Distributions	Distributions	Distributions
Number of input factors	< 100	< 10	< 100	20–100	< 20	20–100	< 20
CPU time per run	1 min–1 h	<1 h	< 10 h	< 1 h	< 1 min	< 1 h	< 1h
Cost of analysis	500–1,000	1,000	$k: 2k$ (FF)	$r(k + 1)$	$N(k + 2)$	100–1,000	500–2,000
Settings for SA	FP	FM	FF	FF	FP, FF, VC	FP, FM	FM

Table 2. Table adapted from Saltelli et al. (2008); SA selection by desired characteristics. Utilizing this table, an appropriate SA method was identified based on the desired characteristics requirements. The authors state that these indications are not to be taken as prescriptive and that more than one method might be applicable to the same context. In choosing among the various methods based upon the cost of a single simulation (row: CPU per run), the authors assumed a single computing unit (no parallel computing) and a maximum computing time of between two and three days (bearing in mind a typical weekend-long simulation). k : number of factors; typically, $N \approx 500/1,000$; r is generally set to $r \approx (4:10)$ and corresponds to the number of the trajectories; FP: factor prioritization; FF: factor fixing; VC: variance cutting; and FM: factor mapping [46]

3.2.3 Local and global sensitivity

The Elementary-Effects method was introduced by Morris et al. (1991) and is derived from the local sensitivity technique one-factor-at-a-time (OFAT) screening methods [50]. Sensitivity analysis techniques are commonly divided into global and local methods [46][23][51]. To facilitate the reader's understanding of the fundamentals of the EE method, the concepts of local and global sensitivity analyses are discussed first, followed by details on EE and extensions that enhance the EE technique.

Local sensitivity analysis (LSA) belongs to the class of OFAT sampling methods [45]. This method examines the behaviour of one parameter with regards to the output by varying one model input while keeping all other input factors fixed at their base line values. To explain local sensitivity, the following general formulation is presented: $y = f(\mathbf{x})$. The function f is established, which is the generic black-box model, representing a function that maps the input factors onto the output. The formula has multiple input factors, $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$, that belong to the input variability space and have scalar outputs y . The LSA method, in its simplest form, yields the partial derivative of a model with respect to one of its input factors [23]. For example, the “sensitivity” of a model’s response y to variable x_i is described as the “rate of change (gradient)” $\partial y / \partial x_i$ of the response y in the direction of increasing values of the factor x_i .

The LSA has two notable weaknesses: If f is nonlinear with respect to x_i , then its partial derivative will change depending on where in the range of x_i it is being monitored. The second deficiency is that if there are interactions between model inputs, then the $\partial y / \partial x_i$ ratio will change depending on the values of the remaining input factors as well [23]. It should be noted that the first partial derivative is only a valid measure of sensitivity when the model is linear, in which case $\partial y / \partial x_i$, “the gradient”, will remain constant for any \mathbf{x} [23]. Therefore, OFAT methods cannot effectively explore a multidimensional space [23].

Detecting non linearities with a partial derivative OFAT method can lead to defining a factor erroneously, thus creating a type I or type II error. Errors in an SA can be divided into three types [46]: a type I error occurs when a noninfluential factor is falsely defined as important, a type II error occurs when a significantly important factor is classified as noninfluential, and a type III error is a framing error that occurs when the “right” answers are sought for the wrong questions [46]. An SA method is not resilient against type III errors.

Nonlinearities or interactions are a weakness of the LSA technique. A global SA, however, can be used to overcome these weaknesses [45][23]. It can evaluate output variation due to one input variable by varying an entire set of parameters over their whole input space simultaneously. This technique is called all-factors-at-a-time (AFAT), as it uses sample intervals in which a whole input set is varied simultaneously (as opposed to varying one factor at a time). Therefore, the sensitivity to each factor considers the direct influence of the varied parameter, as well as the joint influence due to interactions. Thus, global methods can be used to examine the interaction of variables and provide a sensitive measure for nonlinear models. Global SA methods are more likely to be resilient to type II errors than local derivative-based methods [46]. The drawback of AFAT methods is that they may require more extensive sampling and are therefore more computationally expensive when compared to OFAT.

While local SA methods utilize OFAT sampling, global SA uses either OFAT, AFAT, or a combination of these strategies. The above discussion indicated that LSA methods are unprotected against producing misleading results when the model investigated is of unknown linearity or should inputs interact. As the response behaviour of the HEXS model is not known and given that it can behave in either a linear or non-linear fashion, global methods are considered more reliable, as they examine interactions and non-linearities. As noted previously, researchers caution that special care is required when drawing conclusions from an SA when the process itself is poorly known [23][24][25]. Therefore, the functioning and the process of the RMPP and the HEXS model are investigated extensively in Chapter 2 and Appendix 2 and 3.

3.2.4 Elementary effects (Morris method)

The EE method was derived from a local sensitivity method. The EE method is considered to be a global method, as it can be used as an indication of correlation and/or non-linear relationships between parameters [45][44].

The Morris method involves calculating the influence for each model input x_i on the output variable of interest y , of the variability input space $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$. The design variability space D is restricted to the k^{th} dimension and assumed to vary across p_i evenly spaced levels. The EE method computes for a given input value of \mathbf{x} the elementary effect of the i th input factor, that is, between two points of the trajectory r where the variable i is modified, as shown in equation 3.1. Equation 3.1 uses the outputs of EE as inputs, as shown in equation 3.2, where e_i is a vector of zeros but with its i -th component equal to ± 1 [45].

$$\mu_i = \frac{1}{r} \sum_{t=1}^r EE_{it} \quad (3.1)$$

$$EE_i(\mathbf{x}) = \frac{y(x_1, x_2, \dots, x_{i-1}, x_i + \Delta_i, x_{i+1}, \dots, x_k) - y(\mathbf{x})}{\Delta_i} = \frac{y(\mathbf{x} + e_i \Delta_i) - y(\mathbf{x})}{\Delta_i} \quad (3.2)$$

Following the conventional practise in SA, factors are normalized to the dimensionless range [0,1] as $x'_i = (x_i - x_{min}) / (x_{max} - x_{min})$ [45][52]. The x_{min} and x_{max} are respectively the minimum and maximum of the input variable x_i jointly. The domain of vector \mathbf{x} , is then a unit hypercube H^k . For each of the normalized [0,1] input variables, only discretized values are considered using the evenly spaced grid p_i , with steps, as presented in the following equation.

$$x_i \in \left\{0, \frac{1}{p_i - 1}, \frac{2}{p_i - 1}, \dots, 1\right\} \quad (3.3)$$

The sampling strategy originally used by Morris builds r trajectories in the input space or hypercube H^k , with each trajectory being composed of $(k + 1)$ points. In other words, r is defined as a sequence of $(k + 1)$ points in the defined hypercube H^k . In the sampling strategy, the first position of a trajectory r is randomly selected over a uniform grid p_i , and the subsequent k points are obtained by moving OFAT of a fixed, pre-defined amount Δ_i . The step Δ_i allows each trajectory for one EE per parameter to be evaluated. Based on these trajectories, Morris could study the correlation between the input parameters, which makes this a global SA method. The step increment Δ_i for one of the factors used to determine its sensitivity equals one such spacing and is defined as[24]:

$$\Delta_i = \frac{1}{(p_i - 1)} \quad (3.4)$$

Each trajectory r allows one EE per factor to be evaluated. To obtain r elementary effect for each variable, a screening plan is built from r orientations. At each step in a trajectory, the coordinate to be modified is also randomly selected. Thus, trajectories differ in their starting points and the order of their modified coordinates. The set of r different random trajectories (with index t) defined in the hypercube of the input factors is selected. The set provides r estimates of EE_{it} of elementary effects related to each input variable i at the expense of the total number of simulations, which is $r(k + 1)$. Taking one random orientation at a time, the adjacent rows of the screening plan and the corresponding function values from t can be inserted into equation 3.2 to obtain k (one for each input variable). When a sample of r elementary effects has been collected for each of the factors, the means μ and sample standard deviation σ can be computed as presented in equations 3.1 and 3.5:

$$\sigma_i = \sqrt{\frac{1}{r - 1} \sum_{t=1}^r (EE_{it} - \mu_i)^2} \quad (3.5)$$

For each parameter, the mean value and standard deviation of distribution of the EE method are calculated. The larger the mean value of the EE, the higher the level of influence of the response parameter. Meghoo et al. (2019) stated that Morris method can be considered to indirectly examine the relationship between the input factor and the response value [24]. Conversely, standard deviation can be considered the standard error of the regression analysis that estimates how jointly the fitted model represents the relationship between the input and the response. Standard deviation assesses the extent of the non-linear and interaction effects of each of the factors. If the standard deviation or standard error is large, it either suggests that the obtained fit is insufficient to forecast the correct response value and that a higher order or non-linear model is required [24] or that a high correlation is present between the factor under consideration and one or more of the other parameters in the factor space [25][24]. Without physical knowledge of the process or system under consideration, making a clear distinction between the two options (non-linear or interaction between parameters) is impossible [24][25]. Not being able to make a clear distinction between the two, is one of the reasons why researchers caution that conclusions drawn based on an SA should be considered with special care when the process itself is poorly known [23][24][25].

The conventional Morris method results in the mean and standard deviation of the EE method. These can then be represented on a screening plot, as shown in Figure 11. The screen plot illustrates three factors, A, B, and C [24]. In the screening plot, factors A and B have the same standard deviation, and parameters B and C have the same mean value. Hence, parameter A has less influence and thus less importance in comparison to parameters B and C. Factors B and C have equal influence with respect to each other. Lastly, factor C has the largest standard deviation in comparison to B and C. The facts that B and C have equal mean values, but C has a larger standard deviation suggests that either the relationship between C and the response value is non-linear or that C is correlated with one or more parameters.

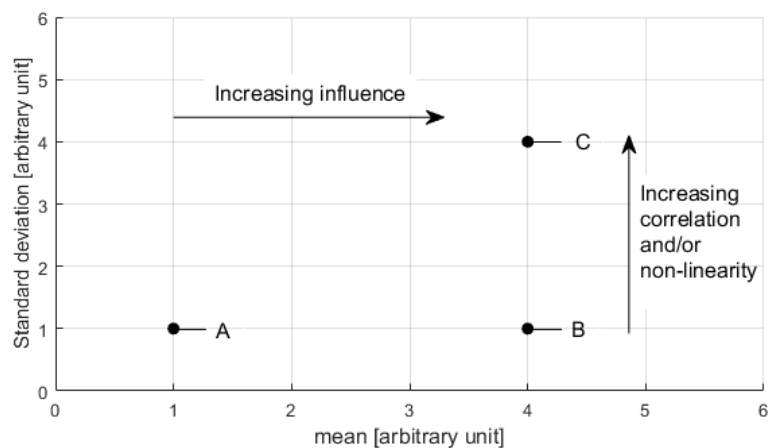


Figure 11 Example of a screening plot for the mean and standard deviation of factors A, B, and C, reprinted from Meghoo et al. 2019 [24].

Note that, about the last effect, “C”, no distinction can be made between the non-linear or interaction effects of the factor. A solution to this is to graphically plot the individual behaviour of the factor i that was identified as effect “C” for its entire variability domain and manually determine whether or not this trend behaves in a non-linear fashion.

In summary, by plotting the conventional statistical indicators (mean and standard deviation), the Morris method identifies the input that can be considered to cause one or more of the following effects:

- A. negligible factors (low mean, low standard deviation),
- B. linear and additive factors (high mean, low standard deviation), and
- C. non-linear factors or factors involved in interaction with other input factors (high mean, high standard deviation).

3.2.5 Extensions of the Morris method

The conventional Morris method generates a random sample of EEs. Campolongo et al. (2007) noted that this random sampling could lead to a non-optimal coverage of the input space [52]. In general, the higher the number of r trajectories generated, the higher the probability that an overall spread will be obtained over the input space, known as “space-filling”. Thus, the larger the number of scenarios, the more likely the space will be filled, and the higher the accuracy of the model that is fitted, such as in determining the standard deviation in Morris method.

Forrester et al. (2007) proposed a method for evaluating the space-filling of the design matrix based on the maximum distance, where d_p is the minimum distance between each possible pairs of points in the design matrix X [25]. In equation 3.6, the rectangular distance is obtained for $p = 1$, whereas the Euclidean distance is obtained for $p = 2$. As the conventional Morris method establishes random trajectories and therefore a random design matrix, “the sampling plan”, the space-filling of that plan is not guaranteed. To evaluate the space-filling of the design matrix, two random design plans “ $X1$ and $X2$ ” are created. Each of the plans is evaluated on the maximin space-filling criterion, as established in equation 3.6. A distinction will be made between the “good” screening plan and the “bad” one. The one that exhibits “good” performance (i.e. that with the highest maximin values), is considered to exhibit the most uniform space-filling of the input space [46] and is thus selected for the Morris analysis.

$$d_p(\mathbf{x}^{(i_1)}, \mathbf{x}^{(i_2)}) = \left(\sum_{j=1}^k |x_j^{(i_1)} - x_j^{(i_2)}|^p \right)^{\frac{1}{p}} \quad (3.6)$$

For the conventional Morris method, Campolongo et al. (2007) noted that, when using equation 3.1, different signs will cancel each other out when the response surface is non-linear and contains both a positive and negative element [13]. The calculated mean of equation 3.1 will result in a low mean value even if the factors have dominant influences on the response. In some instances, this could lead to a type II error, meaning that considerably influential input factors will not be identified [3][13]. Therefore, the authors proposed using the average of the absolute EEs as an alternative, as shown in equation 3.7, but at the cost of losing information concerning the sign of the EEs [2]:

$$\mu^* = \frac{1}{r} \sum_{t=1}^r |EE_{it}| \quad (3.7)$$

The alternative proposed by Campolongo et al. (2007) can be presented in the form of conventional bar plot, where μ^* alone can be used to assess the importance of each factor in the model [3]. Figure 12 shows a combination of the bar plot suggested by Campolongo et al. and the Pareto principle. In Figure 12, each of the factors' μ^* that is, the absolute mean of the elementary effect, the μ^* value is presented on the left vertical axis in the figure. To indicate which factor has the most influence on the response, the results are presented in descending order based on the absolute mean value μ^* for each factor. The individual factors are represented in descending order as bars in descending order. This bar plot alone can be used to assess the importance of each of the factors that is included in a model. Highly influential factors are on the left side of the plot, and they have taller bars when compared to those of less influential factors.

However, the bar plot does not indicate how much influence a factor has on a response when compared to the to the other factors. A means of solving this issue is to utilize a cumulative percentage, whereby the relative influence of a factor can be compared in respect to the others; this can be achieved utilizing a Pareto plot.

In the Pareto plot presented in Figure 12, the cumulative percentage line starts at the first (tallest) bar and extends to the last bar. This line assists in determining the influence of each factor, with each factor being assessed given the added contribution. The cumulative percentage is also displayed for each bar under the chart as “Cum %” and this value is displayed for all of the bars. After a specific influence threshold, set at around 90%, Minitab combines the remaining factors into a group called “Other”. The Other category is displayed as the last bar, even though it has a greater total influence than the previous factors.

The line and the cumulative percentages assist in highlighting the most important factors among the large set thereof by utilizing the Pareto principle. The Pareto principle, commonly known as the “80/20 rule”, states that approximately 80% of all effects come from roughly 20% of causes [53]. By drawing the cumulative line on the bar plot, the 80% threshold can be used to make a distinction between the “vital few” and the “useful many” factors estimated as having the most influence on the response parameter. The small number of factors that disproportionately contribute to influencing the response are referred to as the “vital few”. In this example figure, the largest source of influence is factor X1. It has an absolute mean of 52,518, which accounts for a “Cum %” of 30.1% of all influence in comparison to all the other factors. In this graph, 83.7% of all the influence is estimated to stem from the first four factors: X1, X10, X5, and X6. Over 95% of all the influence is from the first seven factors (counting from left to the right). By utilizing the Pareto principle, an exact percentage of the total output variance that each factor contributes can be obtained.

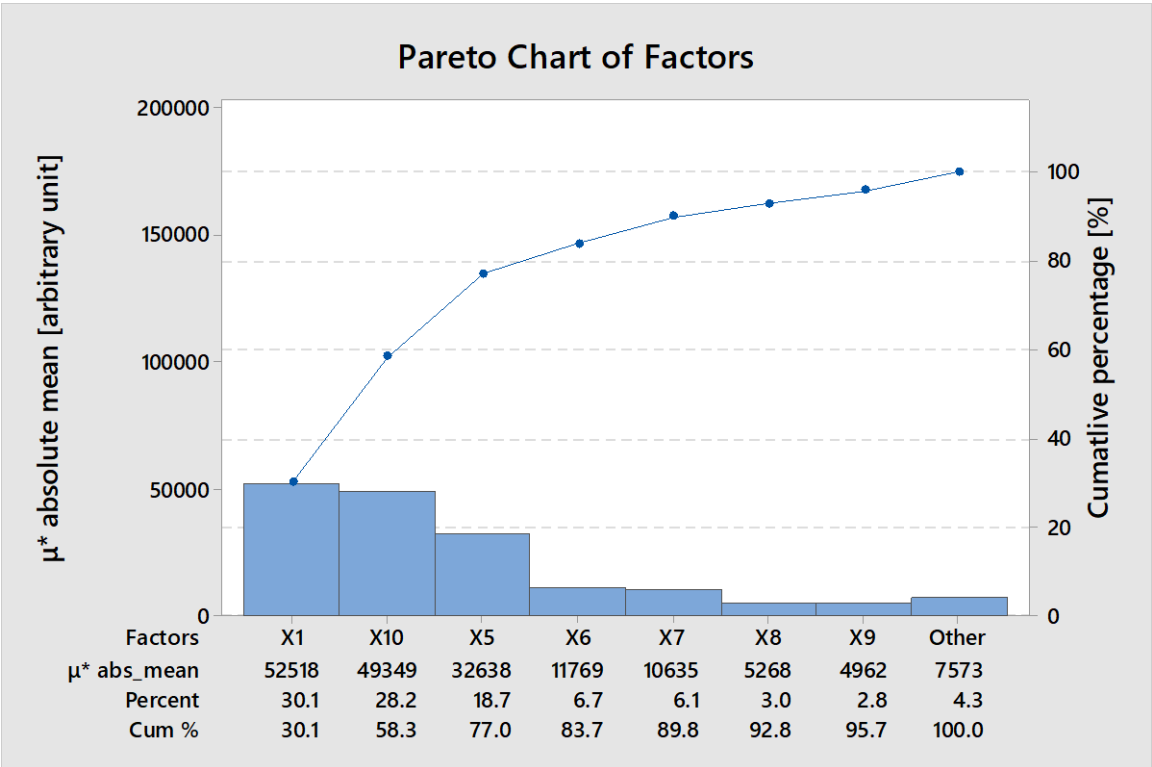


Figure 12. Illustrative example of a Minitab Pareto bar plot with the absolute mean μ^* taken from the mean elementary effects for 10 arbitrary input factors. The factors are ranked in descending order of influence on the output.

3.2.6 Applying the theory to extend the Morris method

The previous section presented several theoretical methods that can be used to improve the conventional Morris method. These methods are put together in this section in schematic order.

The method starts with creating a screening plan, which needs four inputs factors: r , p , x_i , and k . The Morris method is based on the construction of r trajectories in the input space, and, typically, values of between 10 and 50 are used for r [52]. In order for the Morris screening to work correctly, it is suggested that p be an even number and that it should have a level of 16 or 32, while the number of replications r should be at least 20 or more [54]. The x_i elementary effect step is set to 1.

Based on the settings, two screen plans are created with the r , p , and k values. The screen plans are then evaluated with respect to each other on the space-filling criterion. The plan with the highest maximin value is considered to have most uniform space-filling and is thus selected for further processing. This screening plan, which is still normalized, is inserted into the model with the input space and each of the physical ranges of each of the factors. The screen plan that is normalized is transformed from $[0,1]$ to the actual operation ranges; consequently, the model is compiled until all of the $r(k + 1)$ simulations have been performed. The screen plan, levels, and actual input range are inserted into the Morris as the outputs. The absolute mean and standard deviations are calculated, and a screening plot is created, the plotted factors need to be visually evaluated and categorized into effects of A, B, or C. The absolute standard deviation of the EE is calculated, from which a pareto bar plot is created to determine if any factors have been cancelled out. This process is schematically shown in Figure 13. In the figure, k is the number of design variables, p is the number of discrete levels along each dimension, x_i is the number of EE steps, r is the number of random orientations, and y is the response.

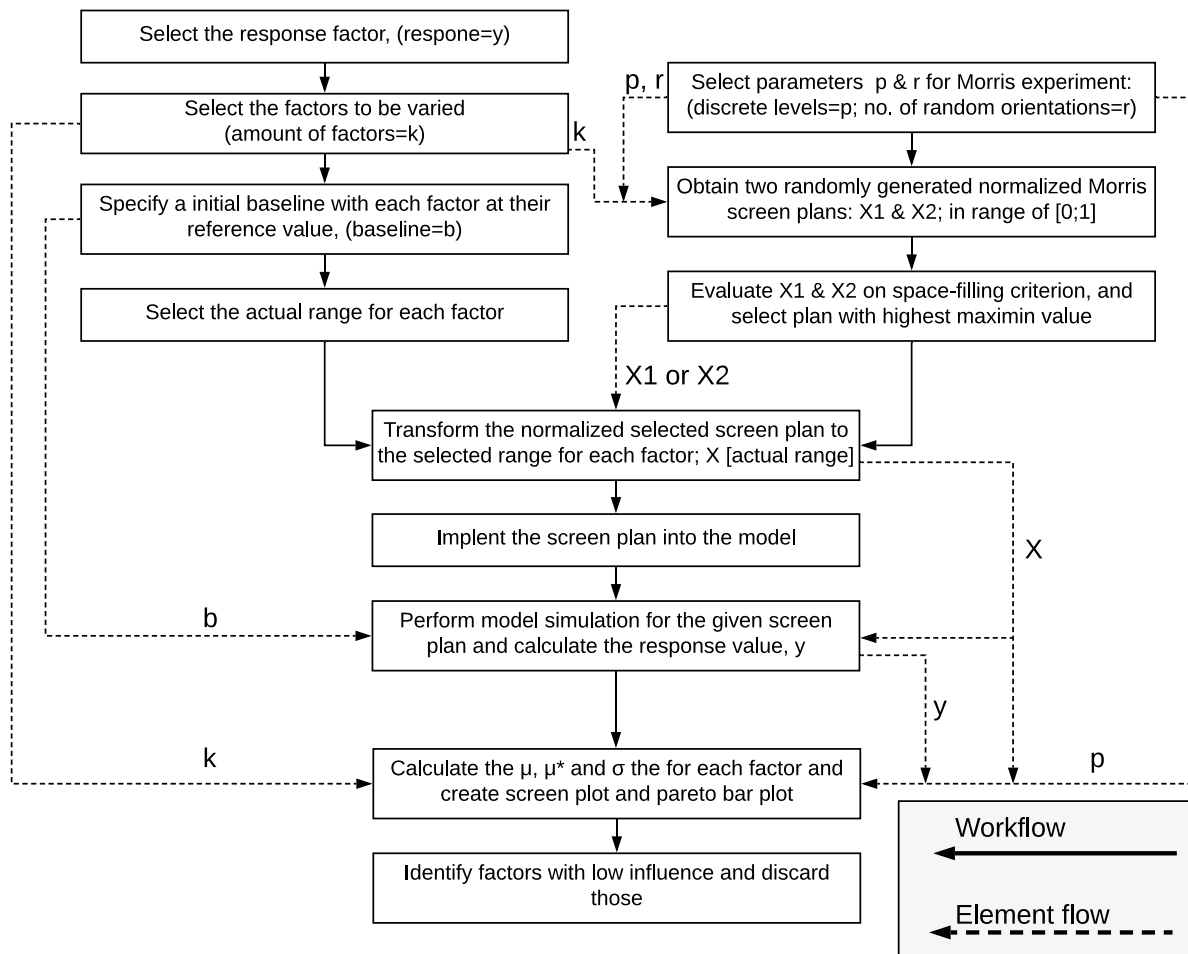


Figure 13. Sensitivity analysis strategy. The process or workflow is denoted by a solid arrow line; the element flow of some of the elements utilized by the extended Morris method is marked with a dashed arrow line. Note that “b”, the baseline element, is only required in the HEXS model to reset the initial state of the model.

4 Constructing a metamodel for the RMPP

In sub-section 3.1.1, the workflow to establish a regression supervised ML model was discussed and illustrated in Figure 10. The steps of this workflow are executed in this chapter to construct a metamodel of the mill outlet temperature of the HIsarna RMPP.

4.1 Step mm1: Metamodel objectives

The first step of the workflow presented in Figure 10 is to define the objective(s) to be used by the trained regression algorithm (the metamodel). There are four objectives: (I) obtain an expression that represents the behaviour of the mill outlet dry bulb temperature of the HIsarna RMPP, (II) validate the obtained expression with respect to the actual response of the RMPP, (III) validate and compare the HEXS model with respect to the HIsarna RMPP, and (IV) determine which variable(s) have a dominant influence on response variable. The first objective is discussed in this chapter, and the other three are discussed in chapter 0.

4.2 Step mm2: Identification of the required data

The second step mm2 is identifying and obtaining the data. The challenge in this step is that the sensors in the HIsarna RMPP monitor behaviour in different quantities and units than does the HEXS model. For example, in the plant, the valve position controls the flowrate and is measured as a percentage. In HEXS, the flowrate is controlled by changing its value expressed in m³/h. Therefore, developing a metamodel from data presenting from data that is presenting the behaviour of the mill outlet temperature that can be compared with output of the HEXS model is challenging and cannot not be done without conversion errors. Both of these modelling systems (HEXS and the metamodel) have their own conceptual modelling space, and each features a different inherent level of abstraction. The abstraction level is that factors are monitored at different locations and/or by physical unit factor(s) in the HEXS model, whereas they are monitored by sensors at the RMPP. However, not all abstraction levels are so different; thus, the metamodeling was not too concerned with identifying a suitable abstraction level space. In addition, as it obtains process data from the plant database and the data provided by weather stations, the HEXS model considers ambient physical quantities.

The sensor data was utilized when the measured parameter could be mapped or related to a parameter within the HEXS model and could be measured continuously at fixed intervals. The sensors that measured continuously and were used for the regression algorithms are identified in Table 3, which presents the parameters used in the HMI of the RMPP plant and those used in HEXS, along with their operational limits. A few additional sensors outside of the HIsarna plant were utilized and included in the analysis. Sensors utilized from the weather station are those, the ambient dry-bulb temperature, the ambient absolute pressure, the ambient relative humidity, and the amount of rainfall. These first three parameters are used in the HEXS model. Precipitation or rainfall was added for the following reason: At the time of writing, no sensors or devices measure the moisture content of the coal at fixed intervals. As the coal is stored outside and therefore exposed to the environment, it is expected that the coal's moisture content will increase due to rain. When a batch of coal with increased moisture content due to rain is processed in the RMPP, an increase in the burner temperature required for drying is expected. The response value, which is the mill outlet temperature, is dependent on the burner capacity. Therefore, rainfall is expected to have an influence on the system and the response.

Other parameters that are included in the regression analysis and are used in the HMI are not directly used in the HEXS model but are expected to influence the response parameters and were for that reason added to the experimental set of parameters. These parameters are the power and current of the M-822 mill motor; the current, speed, and torque of the spinner separator in the dryer column; and the power of the main K-823 fan motor. The measurement of the level of the coal supply bin refers to an increase in the layer of coal between the screw conveyor and the ambient environment, when a higher level indicates an increase in layer. This layer forms a seal, which acts as a barrier preventing the exchange of gasses between the ambient environment and the internal environment of the system. It is expected that, when this layer is reduced and the level indicator is thus lowered, an increase in the exchange of gases will occur. There is a dust content measurement after the bag filter. An increase in dust content will result in an increase in the resistance against a material flowing through the filter. Such an increase in resistance will cause the main fan to

increase its power and thus the heat and pressure in the system to maintain the volume rate required to pneumatically convey the media. These additional sets of parameters are all expected to influence the behaviour of the RMPP, and all of them can influence the mill outlet temperature and are therefore included in the analysis.

Table 3 displays a list of parameters used in the regression analysis and descriptions. Furthermore, in the first column of Table 3 symbolic names are reported, M (M is for meta) followed by a number. Symbolic names marked with * are parameters that have equivalent variable factors in the HEXS model; these factors are utilized to compare results, as stated in objective III. A detailed comparison of the sensors and the factors included in the HEXS model can be found in Appendix 2.

The data used for the regression model is from the month of March 2019. During this month, the RMPP had been operating for about 18 days, which was the longest continuous operation prior to July 2019. In between the runs in March, February, and July, physical design changes were made, such as the installation of the two-way rotation feeder at the mill inlet in March (previously, it was a vibrational feeder). These and other adjustments may therefore have influenced the data set from which the regression model was generated. Therefore, it might be the case that the two forecast models created from each data set would not give similar responses when giving identical input. Therefore, only the dataset from March 2019 was utilized. A more detailed discussion between the run of February and March is reported in sub-section 2.2.9.

An Anderson-Darling test was performed to determine whether the obtained data time series was normally distributed [55]. It was observed that not one of the factors had a data set that was normally distributed. As the data was not normally distributed, none of observations fell within ± 1 standard deviation from the mean, and thus they do not account for about 68% of all the observations of the same factor type. Nonetheless, the minimum, maximum, and average values with their standard deviations are reported in Table 3. The SML conducted for this research, utilizes regression algorithms to construct the metamodel. These regression algorithm did not assume any type of underlying distribution, such as a normal distribution [56]. This is one of the reasons why SML is applied more frequently, as it does not require the investigation of distributional assumptions before it can be utilized.

Table 3. Sensor data of the RMPP and weather stations during the run of the 7th to 25th March 2019, presented with two decimal places. Symbolic names with “*” are parameters that have equivalent HEXS model variable factors.

Symbol	TagName Plant	Description	Unit	Minimum value	Maximum value	Mean value	Stdev (mean)
M1*	Dry-bulb temperature	Ambient dry-bulb temperature at Hoogoven harbour Tata	°C	3.10	11.90	7.38	1.31
M2*	Meteo barometer	Barometer absolute pressure at Hoogoven harbour Tata	mBar	992.30	1,035.90	1,015	13.28
M3*	Relative humidity	Ambient Relative Humidity at Hoogoven harbour Tata	%	57.00	100.00	82.13	9.81
M4*	822WIRA001.PV	Total Weight V-822	kg	0.00	5,384.20	3,141	550.61
M5	Meteo precipitation	Precipitation pumping station 3 at Ijmuiden.	mm/hr	0.00	4.00	0.00	0.05
M6	825GI004.PV	Burner power BU-825	%	-1.90	84.90	49.74	26.17
M7*	823FCV221.PV	Nitrogen injection control valve cyclone inlet	%	0.00	20.00	0.20	1.85
M8*	822PdIRCA001.PV	Differential pressure mill drying column DY-822	mBar	-1.22	36.63	16.89	12.13
M9*	823PdIRA001.PV	Differential pressure measurement cyclone	mBar	-1.26	17.91	5.34	5.14
M10	648FIRA005.PV	Flow measurement coal system inertisation nitrogen supply	m3/h	0.87	8,474.15	3,385.38	1,644
M11*	822FCV211.PV	Nitrogen injection control valve mill hot gas inlet	%	0.00	34.73	0.20	1.87
M12	648PIRZA005.PV	Pressure measurement 3 BarG nitrogen supply	BarG	0.00	3.51	2.16	0.86
M13*	824FCV231.PV	Nitrogen injection control valve bag filter inlet	%	0.00	25.00	0.25	2.26
M14	648TT001.PV	5D24 area 18 bar nitrogen supply temperature	°C	3.42	22.40	12.51	1.86
M15	822LIRCA001.PV	Level measurement coal supply bin V-822	%	-0.03	73.13	17.36	8.30
M16*	824FCV011.PV	Exhaust stack damper (vent gas)	%	10.00	100.00	46.51	17.41
M17*	824FCV012.PV	Recycle gas damper	%	0.00	80.00	41.03	22.83
M18	822M001.Current.PV	Mill motor M-822	A	-0.04	173.03	53.18	29.84
M19*	825TIRA001.PV	Temperature measurement burner hot gas outlet	°C	9.11	447.79	251.15	147.31
M20*	823FCV022.PV	System flow control valve inlet main fan	%	0.00	25.00	11.61	8.53
M21	822M001.Power_PV	Mill motor M-822	kW	-0.07	290.36	29.28	17.32
M22	822M021.Current.PV	Spinner separator S-822	A	0.00	89.99	23.64	12.24
M23	822M021.Speed_PV	Spinner separator S-822	Hz	0.00	30.65	15.32	8.53
M24	822M021.Torque.PV	Spinner separator S-822	%	-0.01	4.86	1.98	1.08
M25	822PIRA002.PV	Under pressure measurement mill inlet	mBar	-15.77	28.80	-4.11	3.51
M26	823FIRCA022.PV	Differential pressure flow measurement cyclone outlet	Nm3/h	-0.47	27.62	12.58	7.54
M27	823M001.Power.PV	Main K-823fan motor	kW	-0.04	227.47	126.39	70.39
M28	823PIRA022.PV	Under pressure measurement system flow measurement	mBar	-66.03	28.56	34.33	23.96
M29	823TIRA022.PV	System flow temperature measurement	°C	6.97	95.63	65.55	23.22
M30	824FCV001.PV	Exhaust fan inlet damper	%	8.75	50.00	23.64	9.43
M31	824FCV244.PV	N2 Injection burner valve	%	-0.06	103.49	66.23	27.09
M32	824M001.Power.PV	K-824 exhaust fan motor	kW	0.00	64.31	12.60	7.11
M33	824MIA001.PV	Dust content measurement after bagfilter	mg/m ³	4.95	100.03	47.30	35.61
M34	824PdIRA011.PV	Pressure difference measurement baghouse	mBar	-0.37	22.39	8.32	5.04
M35	824PIRA011.PV	Under pressure measurement Exhaust fan Inlet"	mBar	-40.86	27.46	16.42	10.47
M36	824PIRA012.PV	Under pressure measurement recirculation	mBar	-2.53	28.45	5.51	3.46
M37	824PIRA201.PV	Pressure measurement cleaning air supply.	BarG	0.00	11.23	6.88	2.38
M38	824QIRCA001.CO	CO measurement 1	%	0.00	22.10	8.30	4.49
M39	824QIRCA001.O2	CO measurement 1	%	0.00	22.10	8.30	4.49
M40	824QIRZA002.O2	Measured O ₂ value	%	0.00	21.54	8.04	4.34
M41	824TIRA001.PV	Temperature measurement bag filter inlet	°C	8.08	107.82	78.60	27.14
M42	824TIRA011.PV	Temperature measurement bag filter outlet	°C	8.59	92.56	69.90	21.21
M43	658FIR111.PV	Flow measurement natural gas supply gas BU-825 heater	m3/h	0.03	138.89	68.94	50.51
M44	658PIRA004.PV	Pressure measurement natural gas supply.	mBar G	-2.97	343.53	267.87	66.26
Output	822TIRCA021.PV	Temperature measurement outlet S-822 spinner	°C	8.13	124.88	73.74	22.82

4.2.1 Retrieved forms of the data elements

The data for the regression analyses, as presented in Table 3, was aggregated from the real-time relational database of the Hlsarna plant. The data was labelled with its specific locations and sensor type in relation to the piping and instrumentation diagrams (P&IDs) of the RMPP, and descriptions of the monitored events or changes, including dates and time stamps, are recorded. However, the predictive power of algorithms created by regression analyses relies heavily on the data used to train the models, including the volume of the data set, the sampling interval of the time series, and the number of overlapping periods of the same length [57]. One of the first decisions was how the data would be retrieved from the database. Two aspects were examined: the interval resolution and the form in which the data was retrieved.

In this work, the interval retrieval resolution is defined as the granularity required to represent the time difference between two successive boundaries in a time series. The retrieval resolution cannot be lower than the recording resolution of the sensor. Note that, from this point on, when the term resolution is used, it refers to retrieval resolution.

Data can be cyclic or averaged; cyclic data is that which is not modified by any computation (e.g. averaging). An averaged retrieval returns a value that is the average of all data points between two successive retrieval times. The difference between the two will be explained by utilizing the illustration presented in Figure 14, which shows how values are returned for cyclic retrieval. The dataset contains 11 data points, ranging from P1 to P11. The data is retrieved with a start time T_0 and an end time of T_2 . The resolution has been set in such a way that the data is returned for each of the three boundaries marked with the dashed lines. The green dots P2, P6, and P10 in the figure represent the data that is returned for cyclic retrieval. At T_0 , P2 is returned because it falls directly on the cycle boundary for the given resolution. At T_1 , P6 is returned because it is the last data point before the cycle boundary, and, at T_2 , P10 is returned for the same reason. If the retrieval mode average is utilized, at T_1 , the average value for P3, P4, P5, and P6 is calculated, and this average value is then returned. The same holds true for T_2 , where the average value returned is the average of data points P7 to P10. How these two data retrieval forms – cyclic or averaged – affect the obtained time series is discussed in the following section.

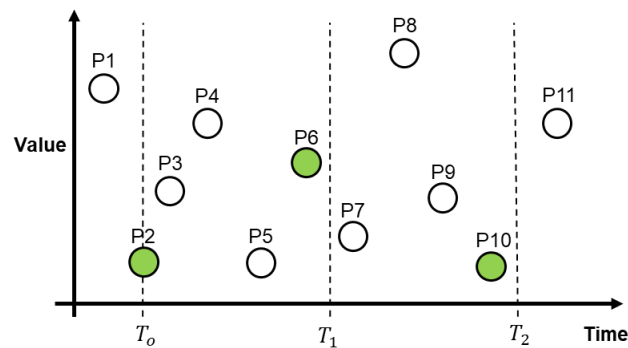


Figure 14. Schematic of the cyclic retrieval form.

4.2.2 Selecting the data retrieval form and resolution

The forecasting algorithm's ability to predict a response that is close to the actual response is dependent on the resolution and the number of factor elements used. If the resolution increases, the amount of computational time and effort required to perform a process increases, along with the time and effort required to create such an algorithm. As limited time and computational power was available, a compromise between resolution and computational effort is required. The data sequence, as presented in Figure 15 was used to determine which resolution could be selected for the SML. The data is a 48-hour cyclic time series of the RMPP mill outlet temperature sensor "822TIRCA021.PV", which was obtained from campaign E2.16. In campaign E2.16, the plant operated from 29 January 2019 to 08 February 2019. The data as presented below is from the first day of February. This particular section was selected because the plant had already been in operation for approximately two days, so start-up disturbances were no longer present. From the sequence, three periodic maintenance activities can be observed, as discussed in Chapter 2.2.9. It can be observed that this time series starts with unstable temperatures due to a process disturbance, which was followed by a scheduled inspection; then, a more stable temperature period is obtained, followed by a second inspection. Near the end, a more time-consuming inspection was performed, where, in addition to the inspection and cleaning of the sieve and feeder, the mill was opened to allow the mill hammers to be inspected.

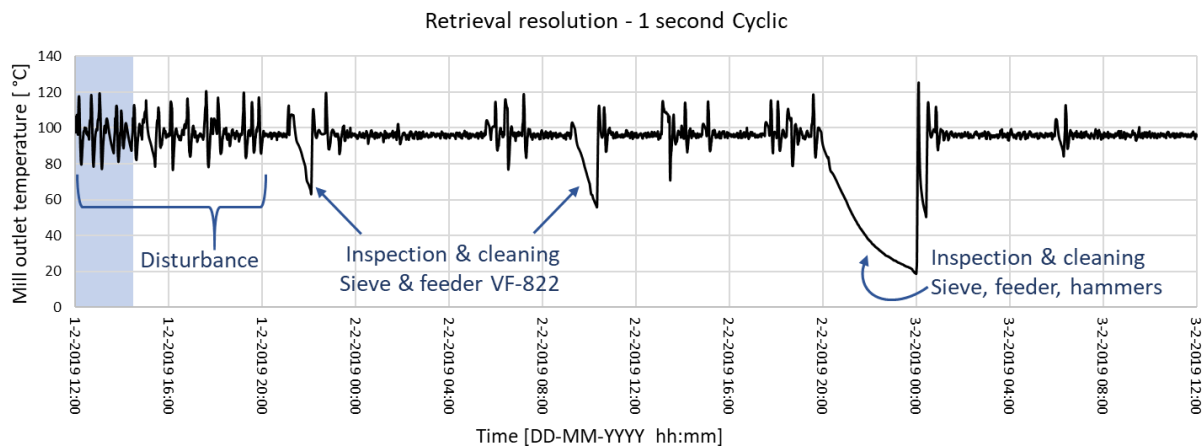


Figure 15. Operation data of the mill outlet dry-bulb temperature sensor (822TIRCA021.PV) obtained from the HIsarna RMPP during the run in February 2019 (Campaign E2.16). The retrieval mode was 1-second cyclic.

Although the plant had been operating for two days, the mill outlet temperature exhibits some uncontrolled disturbance at the beginning of the sequence. Due to this disturbance, the rate at which the mill outlet temperature changes are relatively rapid, and therefore this part is interesting to analyse with regard to determining the retrieval resolution. The particular part of the disturbance was studied more extensively to determine whether lower resolutions can follow the rapid change, this section is marked with a blue area.

In Figure 16, the previous blue area (of Figure 15) is presented in another scale, which allows more details to graphically present. The figure consists of four different cyclic retrieval resolutions: 1, 30, 120, and 300 seconds. The term accuracy refers to between the observed value and an absolute standard (the "truth"). The standard resolution for the 822TIRCA021.PV sensor is recorded at a 1-second frequency and is, for this example, the "truth". A section of the graph in Figure 16 has been enlarged; it can be seen from this enlarged area that 1-second and 30-second resolutions are in excellent agreement. The green line, represents samples being taken every 120 seconds, shows that the green line cannot always follow the rapid changes that can be recorded using finer 30- and 1-second resolutions. In the enlarged section, the 120-second resolution has a difference of 2.8 °C between the black and the green line, but, overall, this sequence can reflect the major temperature changes. The 300-second resolution, which is coloured blue, has multiple large deviations in which it is not able to follow the rate in change of the temperature outlet of the mill ("the truth"). For example, during the decrease in temperature that occurs between time stamps 13:00 and 13:30 hours, the blue line reaches a lower temperature of approximate 86.2 °C, while the green (120 s), red (30 s), and black (1 s) lines all reach exactly 77.5 °C. This is a difference of 8.7 °C, and, at the increasing slope, it has a difference of 10.7 °C. One can notice from the graph and the discussion that the accuracy declines from the absolute standard of "the truth" should the retrieval resolution be reduced.

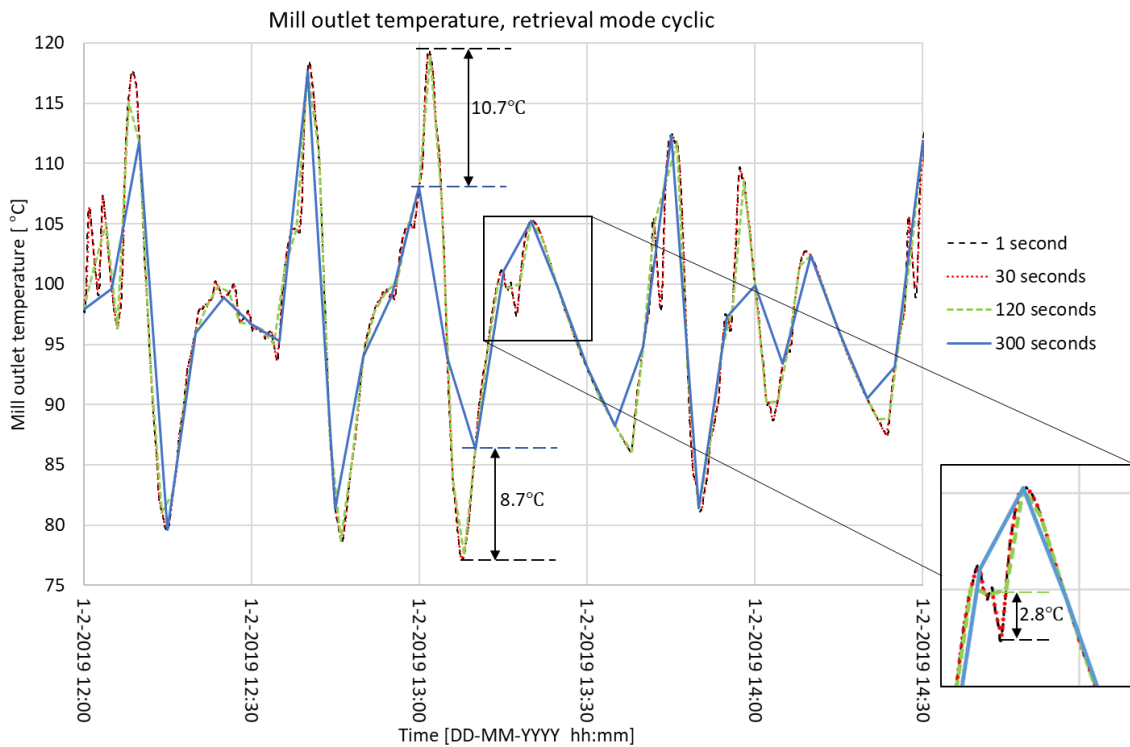


Figure 16. Enlargement of the blue marked disturbance section for the mill outlet dry-bulb temperature (822TIRCA021.PV) obtained from the HIsarna RMPP during the run of February 2019. Retrieval mode is cyclic, and resolutions are 1, 30, 120, and 300 seconds.

One of the objectives with regard to the data was that it should include extremes, such as the highest and lowest values of the mill outlet temperature. The extremes were used to set the range for the SA: the greater the extreme, the greater the operational range that can be used in the SA. Therefore, the form (cyclic or averaged) with the largest range was selected as the form in which the data would be retrieved from the database.

The minimum and maximum temperatures for each “form” were obtained from the dataset of 1 February from 12:00 hr until 14:30 hr, as shown in Figure 16. In Figure 17, the 1s cyclic, 60s cyclic, and 60s average data sets are presented. It is noted that the 60 s averaged data experienced a time lag and smoothed over some fluctuations in comparison to the 1 s cyclic. The smoothing of the 60 s average reduces the entire range (the difference between the minimum and maximum temperatures): the smaller this difference, the smaller the range that will be captured within an interval. The difference between the maximum and minimum results in the extreme range. For the 60s average, this extreme range is less than that of the two cyclic data sets. Therefore, the average form gives the smallest range. As the cyclic retrieval form is more representative of the true physical modalities of “the truth (1 s retrieval resolution)” and has the largest operating range, the cyclic form was used to retrieve data for other factors from the database. An analysis of the complete dataset produced the same results as the analysis of the 2.5 hours.

The performed analysis limited to examining the resolution and form for the response (output) factor. It can be observed from Figure 17 that differences between the cyclic and average values are minimal. However, as there exist 44 input factors that are utilized for the metamodel, it is expected that the number of errors would increase when the average form is applied instead of the cyclic form. Although not examined, the error of the averaged form can potentially lead to a larger prediction deviation than the cyclic form when compared to the “the truth”.

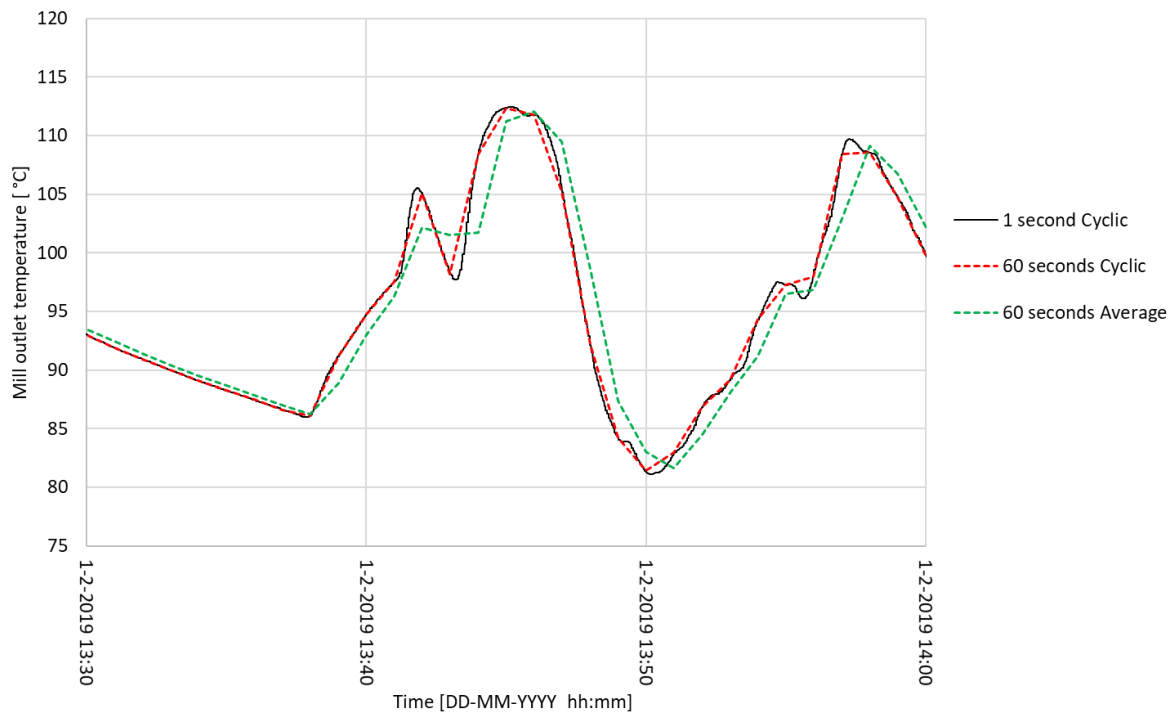


Figure 17. 1 s cyclic, 60 s cyclic, and 60 s average mill outlet temperature of the HIsarna RMPP. Note that the 60 s averaged (green dashed) experiences a time lag and smooths some fluctuations in comparison to the 1s cyclic.

Based on the reasons provided above, the data utilized for the regression analysis was cyclic in form, referring to data that has not been modified by any form of computation such as averaging. The stored data for the given time period is based on the specified cyclic retrieval resolution. However, it may not correctly reflect the stored data of 1 s resolution, as important process values (gaps, spikes, etc.) might fall between the cycle boundaries, which are at intervals of 60 seconds. This behaviour is schematically illustrated in Figure 14, in which the green data points are retrieved and the white-coloured data points are not retrieved, which leaves gaps within the dataset.

Table 4 provides information about the dataset that is used within the experiment, including the start and stop date and times, the interval used, and the number of datapoints within each factor’s dataset. Intervals of 60 seconds was selected because it is able to represent most of the extremes with less deviation than larger intervals. Furthermore, the size of the dataset is reduced by utilizing a 60s interval resolution, which is less computational expensive in comparison to the finer 1 s resolution.

Under the conditions presented in Table 3 and the number of parameters given in Table 4, the Lenovo P50 computer with 16 GB RAM and an i7-6700HQ CPU at 2.6GHz can perform a regression analysis in MATLAB 2018b. If a finer interval, such as 30 seconds or 1 second, was used with 45 factors, the regression toolbox of MATLAB displayed an error message indicating that it could not execute the analysis, as not enough RAM was available.

Table 4. Overview of the data utilized for the regression algorithm.

Dataset	Start and stop date and time	Retrieval resolution	Number of datapoints per factor (observations)
March 2019	07 March 2019; 12:00 hr / 25 March 2019; 12:00 hr	Each 60 seconds, form cyclic	25,920

4.3 Step mm3: Data pre-processing

In the data pre-processing step, the data was inspected with regard to continuity to determine whether it was within the defined operation limits. This means that the dataset was inspected for outliers and to determine whether that the resolution interval of the data time series was less than or equal to 60 seconds at a fixed rate. If a datapoint was not available or was “missing”, this interval was replaced by the average between the previous and the consecutively data point. All data as presented in Table 3, except for the rainfall, was monitored at frequencies of at least once every second. Rainfall was measured in units of length per unit time, and the utilized weather station data was reported in millimetres per hour. As it deviated from the resolution of the other sensors, the rainfall data could be sampled by taking a distribution of its unit length across that hour. The distribution could for example, be skew distribution, as rainfall tends to fall in a skew pattern when observed at interval of minutes [58]. This approach to sampling using a skew distribution is not further discussed in this work, but this is a topic that is recommended for future research.

All of the parameters presented in Table 3 were extracted at intervals of 1 s, with the exception of the rainfall. During the data cleaning, no outliers were detected. Thus, all factors were within the operational range. In total, 39 points were missing; of the total 1,140,568 data points, those missing points account for ~0.027% of missing data. To account for the missing data and ensure continuity, the missing data was replaced with interpreted values: the averages of the previous and subsequent data points. All other data was used without modification, and no additional measures were taken regarding potentially “missing” or misclassified data. This approach was adopted to preserve the application of real-world data.

4.4 Step mm4 and mm5: Sampling and training the regression algorithms

It has been noted that regression models can be optimized to avoid both under-fitting and over-fitting of training data using cross-validation. Such optimization aids in developing a model that that increases in predictive accuracy when assessed against the test set. In this section, data is sampled in both a test and a training set; the regression model is chosen based on the CV-RMSE values.

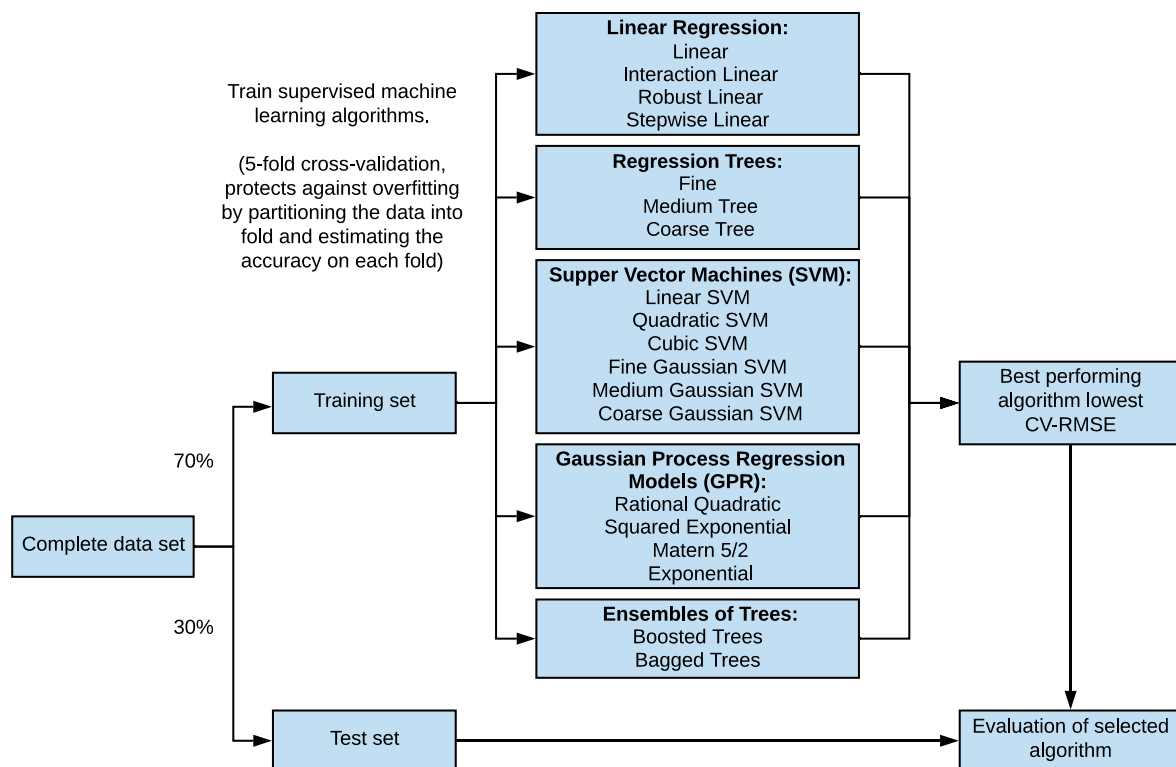


Figure 18. Diagram of regression algorithms. The complete data set is split into training and test sets. The supervised ML algorithms were trained on the training set, and the best performer (i.e. that with the lowest CV-RMSE) was selected and applied to the test for validation.

A typical sampling split used for validation is 70/30, where 70% of the data is used for training the regression algorithms and 30% is held out for the test set, as illustrated in Figure 18. The data set utilized is from the run of 7th to the 25th of March 2019, which includes 44 factors in a cyclic form with a retrieval resolution interval of 60 s. That interval resolution resulted in a total of 25,922 observations per factor. The selected response is the dry-bulb temperature of the mill outlet. In random order, each of 19 regression algorithms was trained on the training set using a fivefold cross-validation. The algorithms that were trained utilizing the MATLAB 2018b regression learner are presented in Figure 18 [59].

4.5 Step mm6: Select the regression algorithm

The CV-RMSE score is the RMSE of all observations, with each observation being counted when it was in a held-out fold. After training all of the regression algorithms, the best-performing algorithm (as determined by the lowest CV-RMSE value) was used for further evaluation. In Table 5, the resulting scores for each trained algorithm are presented. The results from the SML are also summarized in Table 5. It can be noted that the GPR rational quadratic has the lowest CV-RMSE (0.99 °C). The last steps mm7 and mm8 of the workflow (Figure 10)– the evaluation of the goodness-of-fit – is discussed in sub-section 5.1. In that section the trained metamodel is evaluated on its goodness-of-fit; that is, the closeness of its trained response was compared to the actual response.

Table 5. The CV-RMSE results for the 19 regression algorithms used the SML conducted for this research.

Class of regression algorithm	Species of class	CV-RMSE
Gaussian process regression (GPR)	Rational quadratic GPR	0.99
Gaussian process regression (GPR)	Matern 5/2 GPR	1.00
Gaussian process regression (GPR)	Exponential GPR	1.22
Gaussian process regression (GPR)	Squared exponential GPR	1.27
Ensemble	Bagged Trees	1.64
Support vector machine (SVM)	Quadratic SVM	2.11
Support vector machine (SVM)	Medium Gaussian SVM	2.23
Tree	Fine tree	2.23
Tree	Medium tree	2.28
Stepwise linear regression	Stepwise linear	2.35
Tree	Course tree	2.62
Support vector machine (SVM)	Cubic SVM	2.99
Support vector machine (SVM)	Course Gaussian SVM	3.77
Linear regression (LR)	Linear	3.99
Support vector machine (SVM)	Linear SVM	4.24
Linear regression (LR)	Interactions linear	4.29
Ensemble	Boosted trees	4.32
Support vector machine (SVM)	Fine Gaussian SVM	5.29
Linear regression (LR)	Robust linear	5.88

5 Execution, results and discussion

This chapter reports the goodness of fit for the observed sensors' response against the predicted metamodel; a validation of the HEXS model against the metamodel; the sensitivity of the HEXS model; a sensitivity analysis of the metamodel; and a PoC of identified potential failure conditions for functional operation.

For clarity, the execution is described together with the results of the performed experiment and supplemented with a discussion of these results. Finally, the chapter closes with a general discussion and discusses some of the identified limitation of the method(s). For completeness, note that a list of pre-identified limitations is presented in sub-section 1.5.4.

5.1 Goodness of fit, comparison of the observed versus the predicted

In Section 4.1, four metamodel objectives were presented, the first two are discussed in this section. The first objective was (I) to obtain an expression that represented the behaviour of the HIsarna RMPP, and the second was (II) to validate the obtained expression in respect of the RMPP's actual response. Ultimately, the metamodel should be used in to validate the HEXS model. In addition, discussing the goodness of fit provides an answer the second sub-question: How accurately can the metamodel describe the behaviour of a selected response of the RMPP? The accurateness will be evaluated in terms of the predictive power of the metamodel and quantified using its RMSE.

This study achieved the first objective by constructing a metamodel with a regression algorithm, GPR Rational Quadratic, which has the lowest CV-RMSE in comparison to the other trained algorithms. This metamodel is utilized in the second objective by comparing the difference—or “error”—between the predicted value (y_{fit}) and actual value (y_{real}) of the MOT in a regression analysis. If the metamodel exhibited abnormal behaviour (y_{fit}) in a specific part of the design space or high inaccuracy for some initial settings, this could indicate that the metamodel is partly based on erroneous results (outliers). Metamodels can aid in identifying these outliers. This is a starting point for investigating areas for improvement when retraining the metamodel, such as removing irrelevant and redundant features [41]. Furthermore, metamodels may contradict the intuition of process behaviour. This can be a motivation for further inspecting the metamodel, its black box, or the phenomena of interest.

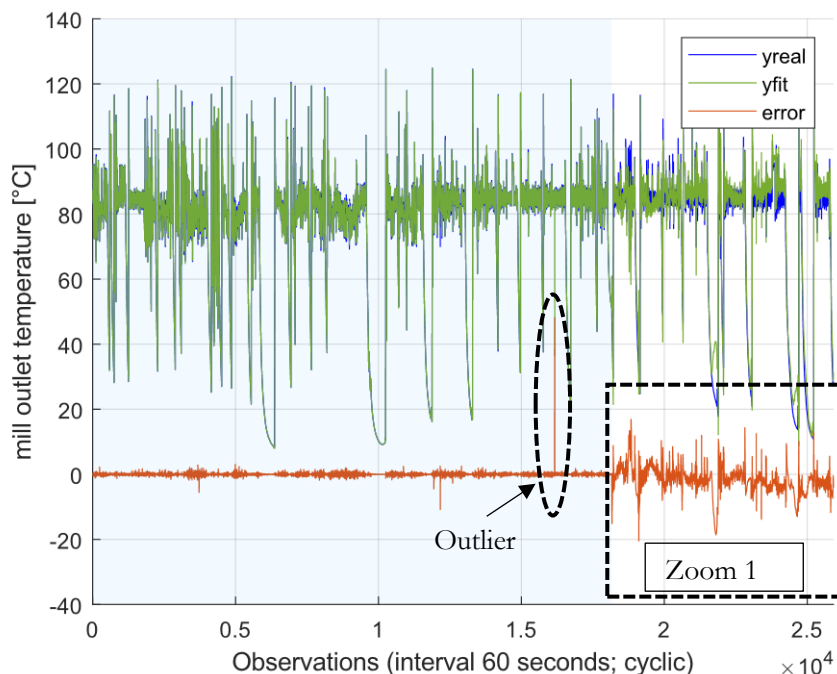


Figure 19. The GPR Rational Quadratic algorithm trained with 70% of the complete dataset (this 70% is the light blue area in the graph); the metamodel's prediction has an RMSE of 2.45 °C.

Figure 19 plots the actual values (y_{real}) and predicted data (y_{fit}): the goodness of fit is graphically evaluated through observing the difference between the predicted (y_{fit}) and actual (y_{real}) values; this difference is calculated using equation 5.1 and plotted in red.

$$error = y_{real} - y_{fit} \tag{5.1}$$

The predicted data were forecasted using the GPR Rational Quadratic algorithm and trained with 70% of the March dataset. This first 70% (18,144 observations) is marked within the blue area in Figure 19. As shown in Figure 19, the trained algorithm overlaps with the actual data closely within the blue area. Except for one outlier within the trained region, this large outlier of approximately 45 °C is encircled by the dashed black ellipse; overall, the entire region (trained and untrained) has an RMSE of 2.45 °C. However, outside the trained blue area, a sudden increase in deviations occurs. This deviation is the transition from the trained 70% of the data to the untrained 30%, where the error increases. The untrained 30%, as marked by the dashed box, is enlarged in Figure 20 for further analysis.

Figure 20 presents the absolute error of the prediction for the untrained 30%, with a maximum error of approximately 21 °C. Overall, this particular deep dive has an RMSE of 4.41 °C, which is larger than the overall RMSE of 2.45 °C reported in Figure 20. To put this into context, it is compared with the difference of the minimum of 8.13 °C and maximum mill outlet temperature of 124.88 °C in the entire dataset—“the mill outlet temperature’s operating domain.” The difference ($124.88 - 8.13 = 116.75 = 100\%$), for 4.41 °C it reflects an error of approximately 3.8% over the entire MOT operating domain. On average, this seems to be a reasonable prediction, although one outlier surpassed the 20 °C, which is approximately 17.5% of the entire MOT operating domain.

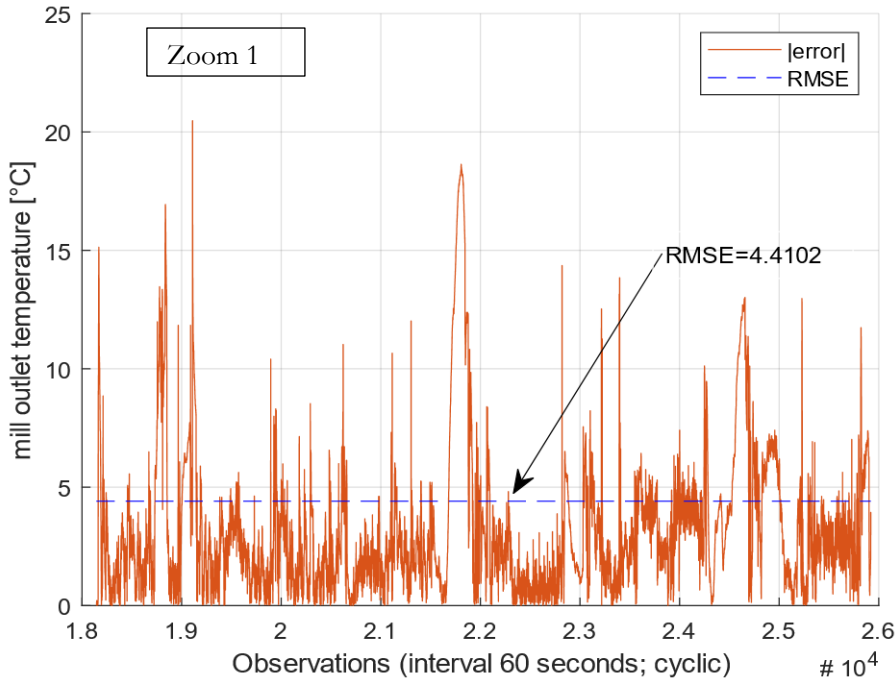


Figure 20. Zoomed-in absolute error and RMSE; this part has not been trained by the regression model. In this local part, the metamodel’s prediction has an RMSE of 4.41 °C.

When the algorithm is retrained using the full dataset, it is called the metamodel, which is presented Figure 21. This study observed from Figure 21, that the model trained with 70% of the dataset had its RMSE of 2.45 °C reduced towards 0.52 °C; this equals an error of approximately 0.4% of the entire operating domain. Furthermore, the frequency and extremes of the outliers are noted to be reduced. In the 70% trained model, the largest outlier is detected as approximately 45 °C (38.6% of the domain). Compared with the 100% trained model, it has a maximum outlier of approximately 18 °C (15.4% of the domain). Therefore, as expected, the 100% trained model seemed to have increased predictive power compared with the partially trained metamodel.

The objective of validating the obtained metamodel expression in respect to the actual response of the RMPP was examined. Because an RMSE of 0.4% is relatively small compared with its operating domain, the 100%-trained metamodel is expected to simulate the representative behaviour of the physical RMPP’s mill outlet temperature within the operating range of its training dataset. It is for that reason expected that this metamodel would reasonably predict the MOT of the HIsarna RMPP, as long as its predictions are within the historical operating domain. The metamodel could therefore, be used for the metamodel objectives (III) and (IV) as stated in sub-section 4.1.

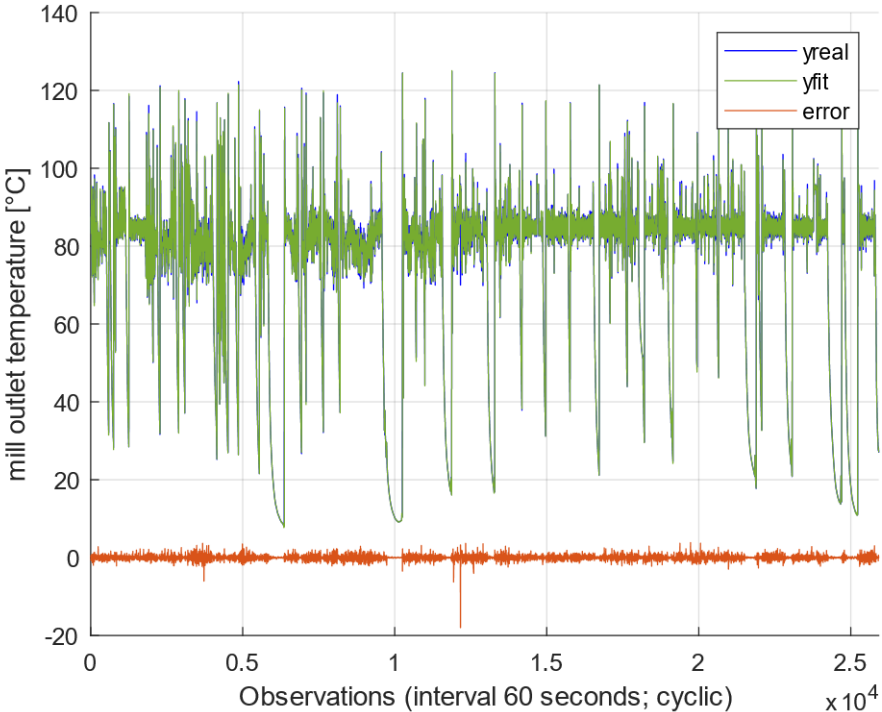


Figure 21. The GPR Rational Quadratic algorithm trained with 100% of the complete March dataset. The metamodel’s prediction has an RMSE of 0.52 °C.

5.2 Validation of the HEXS model

In the previous section, the predictive power of the obtained metamodel was evaluated for untrained and trained datasets. The GPR Rational Quadratic algorithm 100% trained model was observed to have an RMSE of 0.52 °C and an error of approximately 0.4% of the entire operating domain. Therefore, the metamodel was found to reasonable to predict the MOT of the HIsarna RMPP.

The major goal, as stated in objective III in Section 4.1, is to validate the HEXS model to accumulate evidence to demonstrate that the physics-driven algorithms in the code are implemented correctly and functioning as intended. Should the HEXS model be found to reasonably reflect the real-world process of the HIsarna RMPP, then it could provide insights into the parameters that influence the actual system and sub-systems of the HIsarna RMPP. The intended function of the HEXS model is to estimate the process behaviour of an RMPP, and the HEXS model is known to be modelled differently from the RMPP used at HIsarna. Through this validation and comparison, the aim is to increase awareness of and insights into the differences between the HEXS model and behaviour of the HIsarna RMPP expressed by the metamodel. With the absence of a proof of verification, performing further analysis with the HEXS model is unappealing because whether its resulting output is of any application value to the HIsarna plant (representing the behaviour of the HIsarna plant) is unknown.

In this section, the physics-based HEXS model is compared with the data-driven metamodel; however, it is limited to examining the mill outlet dry-bulb temperature response within a specified input range. This validation consists of the three validations described in Section 3.1, which are repeated here for the reader's benefit:

- (1). Do both model responses for the mill outlet temperature move in a similar direction for a factor's given operating domain?
- (2). Do both model responses for the mill outlet temperature operate as expected (expectations from extensive RMPP investigation)?
- (3). Are both models able to identify similar dominant influencing factors in the SA that vary the mill outlet temperature?

5.2.1 Initialization of the validation experiment for (1) and (2)

The first step is to identify factors that existed for both models. The criteria for identifying these factors is as follows. The factors should be monitored at similar locations in the process and have equal physical quantities and units. From the 20 changeable input factors in the HEXS model and 44 factors in the metamodel, seven factors met the criteria and are listed in Table 6 with their symbolic names, descriptions, units, and minimums and maximums. The factors of most interest in the validation are M19 and H19, the burner outlet temperatures, because they are driving forces for the mill outlet temperature having a higher value than the ambient temperature. It is expected that if the burner temperature is varied, the mill outlet temperature will vary in the equivalent way if all other factors are fixed and thus not change; for example, if the burner temperature rises, it is expected that mill outlet temperature is expected to also increase.

Meta Symbol	HEXS Symbol	Description	Unit	Minimum	Maximum
M1	H1	Ambient Dry-Bulb Temperature	°C	3.98	11.02
M2	H2	Ambient Absolute Pressure	mBar	996.66	1031.54
M3	H3	Ambient Relative Humidity	%	61.30	95.70
M9	H9	Cyclone, Differential Pressure	mBar	0.66	15.99
M16	H16	Vent Gas Damper position, behind Exhaust Fan	%	19.00	91.00
M17	H17	Recycle Gas Damper position, behind Exhaust Fan	%	8.00	72.00
M19	H19	Burner Outlet Temperature	°C	200.00	403.92

Table 6. Operating range utilized for the HEXS and metamodel factors for the validation experiment.

For each of these factors, graphical representations are given in Figure 22 and Figure 23. In these figures, the graph starts at the minimum value reported in Table 6, after which it increases by 100 equal steps until it reaches the reported maximum. These minimums and maximums are derived from the HIsarna RMPP monitored operating space during the run in March 2019, as described in Appendix 5.2 and 6.2. The range is selected to be 80% of the full monitored space of the RMPP. The first and last 10% of the monitored data is neglected because these upper and lower bounds could cause disturbances in the metamodel's predictions. In general, this study noted that the frequencies of the monitored upper and lower bounds occurred less often, and the prediction accuracy of the metamodel increased when these upper and lower 10% bounds were neglected. The lower 10% bound was not used for the burner outlet temperature, which was set to a minimum of 200 °C because this is the lowest temperature setting used for drying according to the HIsarna RMPP operators. Both model responses for each factor are plotted in similar graphs, using a blue solid line with O symbols for y_{meta} and a blue solid line with asterisks for y_{HEXS} .

From these responses, as presented by the blue solid lines in Figure 22 and Figure 23, a linear regression dashed red trendline was derived with a linear regression equation: $y = \theta x + b$, where θ is the slope of the line, x is the actual observed value, and b is the y -intercept. Linear regression was used as a rough correlation measure between the meta and HEXS models for each selected response factor. The correlation was studied by examining the modelled and linear responses in a manner that the two models move together, both in the direction and closeness of their movement. A positive correlation ($+\theta$) is found when the linear regression line increased from the lower left corner to the upper right corner; a negative correlation ($-\theta$) is found when the linear regression line decrease from upper left corner to the lower right corner; and a neutral correlation ($\theta = 0$) is found when the linear regression line moved from left to right horizontally. The closer the two models moved together, the stronger the validity relationship.

Because the HEXS model is not tuned to simulating the behaviour of the HIsarna RMPP, and the metamodel has some deviations as well, the two were expected to not be an accurate fit. Therefore, only the sign of the slope $\pm\theta$ or neutral was considered for comparing the two models. This method evaluates the overall global gradient of the specified domain and indicates the direction of “the trend line.” The gradients θ for each factor are shown in Table 7. If both slopes of the linear regression have a positive–positive, negative–negative, or neutral–neutral correlation, then they would move together in the similar direction. Otherwise, they would not be in agreement with each other. This slope sign provides a relative scale of difference between the comparative ratio of two factors, which is a method of validating whether the model outputs are in agreement or not.

5.2.2 (1) Results: Are the signs of the linear regression equation similar?

In a manual comparison of seven factors, Table 7 shows that four out of the seven had similar slope signs. These were factors 1, 16, 17, and 19. Three out of the seven did not have similar slope signs, which were 2, 3, and 9. When the blue solid lines were manually inspected, this study observed factor M2's trendline to be significantly different compared with that of H2. When H2 almost moved horizontally across its domain, M2 decreased. Input factors M3 and H3 had responses that were nearly neutral (horizontal lines), although they both had very minor steepness. When H3 was slightly positive (0.001146), M3 was slightly negative (-0.042717). The metamodel M9's response had a positive parabolic curvature that opened upwards. If, for M9, a linear regression had been established for the first half of the parabola, it would have obtained a negative slope sign similar to that of H9. However, because the second half of the parabola was slightly steeper, the gradient coefficient became positive. H9 had a negative slope of -0.000003, which can be regarded as a horizontal line. Notably, the maximum variance in MOT for a factor over the specified domain was less than 2.1°C as could be calculated from the M9 linear regression equation, which is relatively small with regard to the maximum operating range of the MOT as stated in 5.1.

Table 7. Slope coefficients θ of the linear regression line obtained from the metamodel and HEXS model responses.

	Factors						
	1	2	3	9	16	17	19
Meta (M)	0.236810	-0.185150	-0.042717	0.134900	-0.242030	0.014258	0.120430
HEXS (H)	0.623830	0.000024	0.001146	-0.000003	-0.851900	1.337000	0.404700
Comparison	Positive– Positive	Negative– Positive	Negative– Positive	Positive– Negative	Negative– Negative	Positive– Positive	Positive– Positive

For factors 16 and 19, the two models moved together in quite a comparable manner. For factors M16 and H16, a similar curvature was noted in the trendlines, which were flattened at the end of the operating domain. Whereas M16 and H16 followed curvatures, M19 and H19 followed strict linear lines; as can be observed, the linear regression line lay almost perfectly on top of the modelled response. For some factors, differences in the gradients were minimal. However, for three out of the seven factors, these trends are used in another behaviour comparison. Expected behaviour, based on obtained knowledge about the RMPP operation, was that when they are varied, the MOT trend line would move in a particular direction.

5.2.3 (2) Results: Are the response behaviours as expected?

In this section three input factors modelled trendline is compared with expected behaviour, that is described in sub-section 2.2.4 and sub-section 2.2.5, these expectations are for convenience briefly repeated here. The expectation about the burner temperature was correct. For both models, when burner temperature increased, so did the mill outlet temperature if all other input factors were fixed. This behaviour was expected because the burner is the main heat source in the RMPP that increase the drying medium temperature and thereby the MOT. For the two other factors 16 (valve position of the vent gas damper) and 17 (valve position of the recycle gas damper), it was expected that they would operate in a certain manner as illustrated in Figure 23. The vent gas of factor 16 is releasing process gas from the system to the ambient air, whereby the water vapor concentration within the system does not become saturated. When the vent gas damper position is opened, less process gas could be recycled for its heat, and thus it is expected that when it opened further, the mill outlet temperature would decrease. For both models M16 and H16, the trend was observed that the mill outlet temperature decreased when the position of the vent gas damper is opened (increase in percentage). Vice versa, when the recycle gas damper is opened, the process gas is already heated by the system and could be reused. When the recycle gas damper position is opened more, it is expected that the mill outlet temperature will increase because more heated air recycled, and all other input factors are unchanged. This behaviour was observed in M17 and H17, although the effect from varying the position of the M17 recycle damper position was minimal on the mill outlet temperature for the given domain. Nonetheless, for both models all three factors their trendline behaviour was as expected.

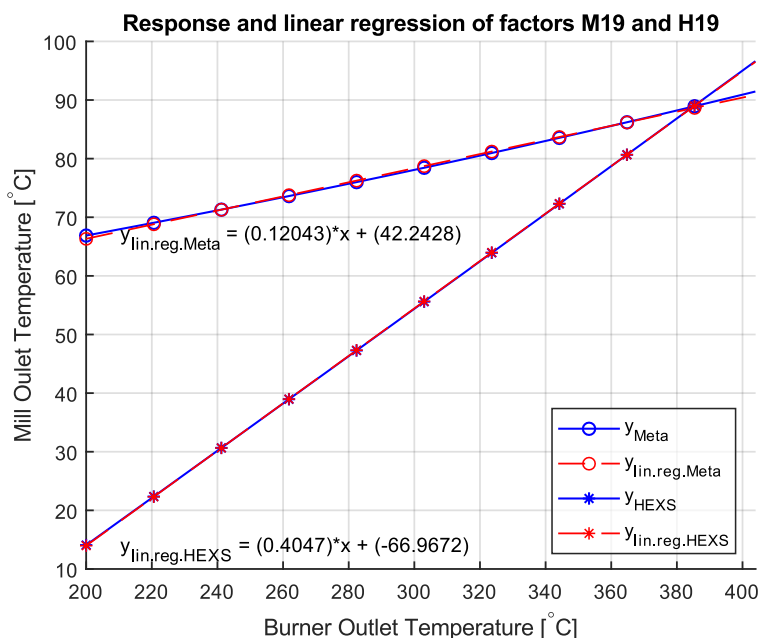


Figure 22. The response (solid blue line) and linear regression trendline (dashed red line) with their mathematical expressions were given 100 steps to get from the minimum to the maximum, as stated in Table 8. This was performed for factors M19 and H19 from the metamodel and HEXS model.

A part of the third sub-question can now be answered. The agreement of the SA results for both models will be evaluated later in this work (in sub-section 5.2.4). The third sub-question is: Are the responses (mill outlet temperature) of the HEXS model and metamodel in agreement when given a similar input?

Both models were executed for an equal operation input range, as presented in Table 9. However, the metamodel requires more baseline settings than that of the HEXS model because it utilizes more input factors. By contrast, the HEXS model has physical expressions for some of these factors. During this first part of the validation experiments, all three expected behaviours were in agreement between the HEXS model and metamodel, indicating that the HEXS model is valid. From comparing the linear regression coefficient signs, four out of the seven did provide a similar sign but three did not. Thus, a minor amount of disagreement existed between the two models. Therefore, no evident conclusion can be drawn that the seven input conditions of the HEXS model and metamodel are in agreement with each other. To answer the sub-question, there exists agreements exist between the two, but this does not apply to all conditions and factors.

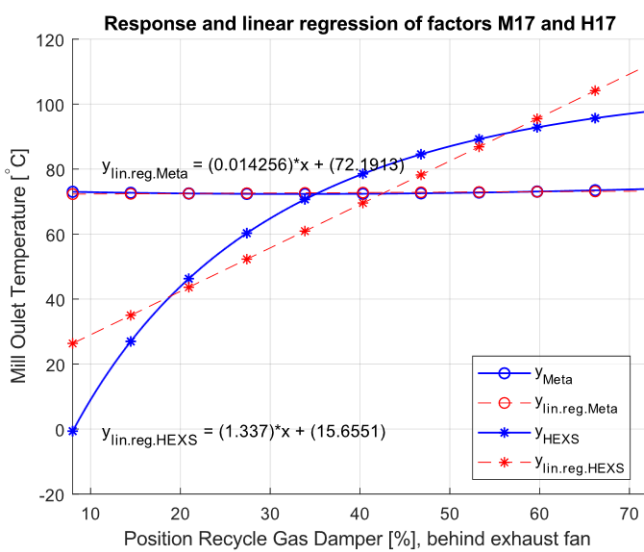
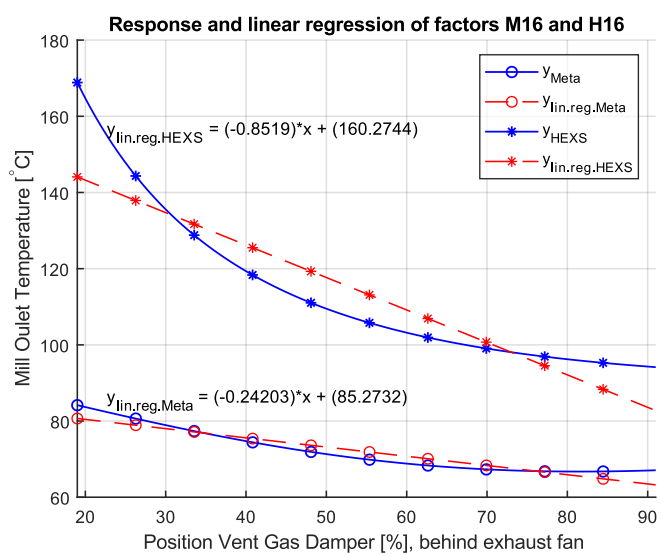
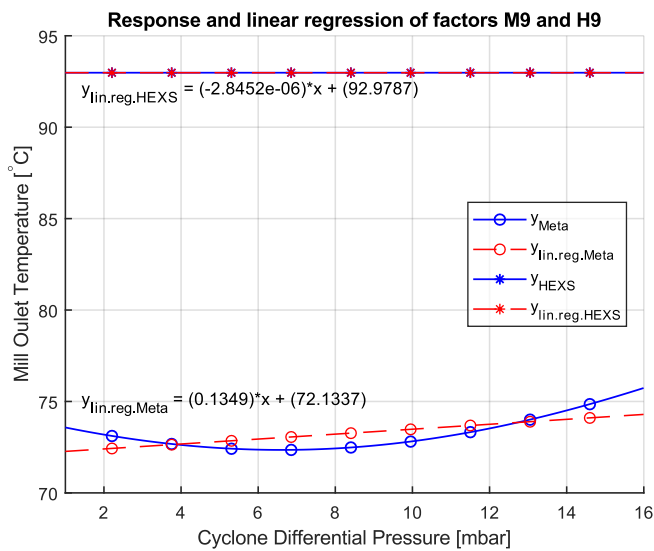
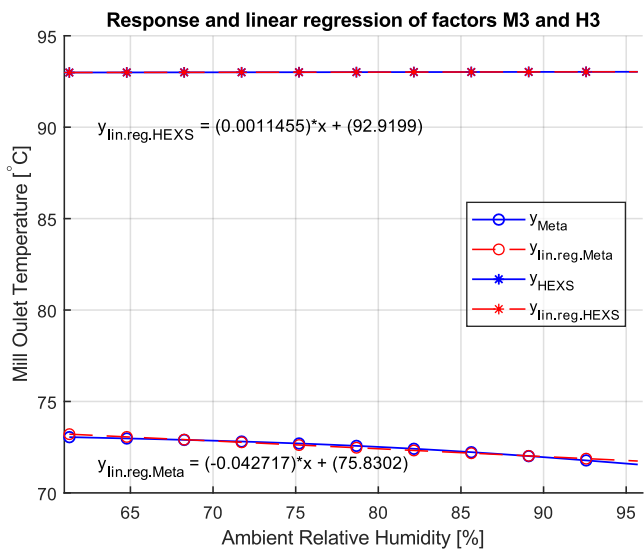
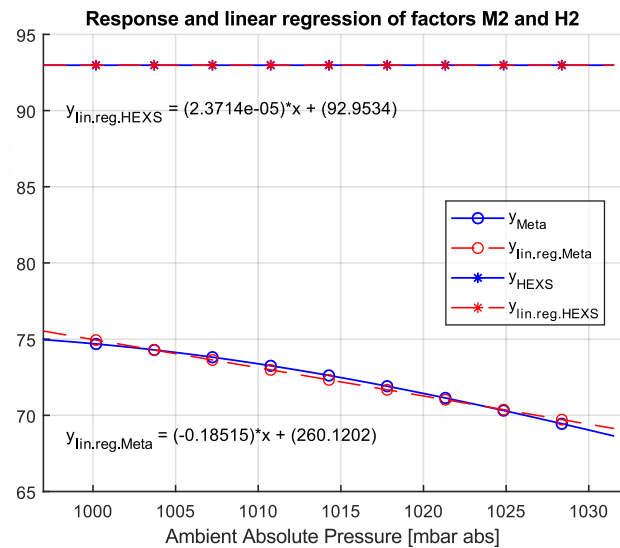
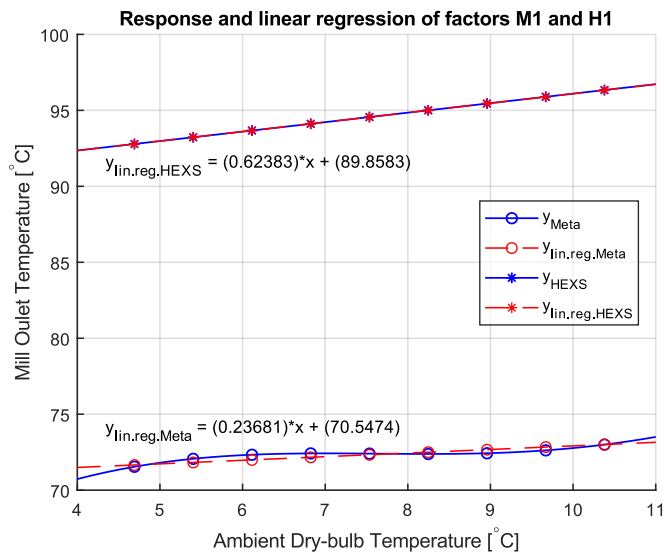


Figure 23. Response (solid blue line) and linear regression trendlines (dashed red line) with their mathematical expressions for 100 steps to get from the minimum to the maximum, as stated in Table 10. This was performed for factors from the metamodel and HEXS model: M1, H1, M2, H2, M3, H3, M9, H9, M16, H16, M17, and H17.

5.2.4 (3) Results: Are the SA results similar for the HEXS and metamodel?

In this section, sensitivity analysis results are compared with the input factors that can be varied in both the HEXS and metamodel. These are comparable factors that are monitored at similar locations in the process and have equal physical quantity and units. These factors are marked with * in the previously presented tables, and are summarized with the SA results in Table 11 for the HEXS model and Table 12 for the metamodel.

The SA results have an arbitrary unit for the μ^* absolute means, μ mean and σ standard deviations, hereby the results from the tables cannot directly be compared with each other. For example, if the μ^* values are taken for the first row of H1* in Table 11 which is 1599 and compared with M1* Table 12 which is 1.543. It can be noted from this comparison, that the order of magnitude is different. To be able to compare the SA results of both models, they can be normalized or the pareto plot can be utilized; in pareto only the μ^* are normalized as cumulative percentage. The (absolute) means and standard deviation SA results can be compared by first normalizing the results to relative percentages. For each of the tables, the (absolute) means and standard deviation of the elementary effects are calculated to the relative percentage as: $x_{relative} = \frac{x_i - x_{min}}{x_{max} - x_{min}} * 100\%$ where x_i is each factor's absolute mean or standard deviation.

A Pareto plot is set to 95%. The Pareto plot's factors are ranked in decreasing order of importance according to μ^* , which is a measure of the overall factor of importance. The Pareto bar plots are presented in Figure 24 and Figure 25, where the sensitivity measures μ^* and cumulative percentage are given for five of the seven factors for the HEXS model and for six of the seven factors for the metamodel. The labels indicate the symbolic names of the factors until approximately 95% of the cumulative percentage. The remaining factors are grouped in the last bar of the chart as the "other" category.

In the two Pareto plots and tables, only the comparable input factors that occur in both the HEXS and metamodel are presented. In the SA, the input factors are varied in the OFAT manner, and their influence is studied on the response. Therefore, only the changeable input factors are compared in the two plots. Each symbolic name given in these tables and plots corresponds with the similar factor of the models; for example, H19 = M19, H1 = M1, etc. For both models H19 and M19, "the burner outlet temperature" it is observed that they have the highest absolute mean values and therefore are most influential on the dry-bulb temperature of the mill outlet. The HEXS model factor H19, it is estimated that it contributes 69.3% of the cumulative percentage, and for M19, 50.9%. None of the other factors have a comparable ranking or order in the Pareto plots. In Appendix 7, the actual values for the absolute mean, mean, and standard deviation are given for the HEXS and meta SA, which are utilized for computing the Pareto charts.

The abovementioned discussion provided the answer to sub-question four: Which dominant factor(s) in the HEXS model and metamodel are in agreement for similar input factors that dominantly influence the mill outlet temperature response? From the seven similar factors, only one of the factors – H19 and M19 – is in agreement for the SA; the other factors are not. The results of Table 11 and Table 12 are graphically represented in a screen plot, this screen plot is included in the Appendix 9.

Table 11. Sensitivity analysis results for the HEXS model for seven comparable inputs factors, response dry-bulb temperature at the mill outlet, and their relative percentages.

HEXS-Model						
Symbol	Factors	μ^*	μ	σ	$\mu^* [\%]$	$\sigma [\%]$
H1*	Ambient Temperature (°C)	1599	-807	5627	0%	0%
H2*	Ambient Pressure (mbar abs)	2145	-1993	6431	1%	0%
H3*	Ambient Relative Humidity (%)	2541	703	10345	2%	1%
H9*	Differential Pressure, Cyclone (mbar)	5268	1372	20068	7%	4%
H16*	Vent Gas, Damper, behind Exhaust Fan (%)	2887	-1936	7648	3%	1%
H17*	Recycle Gas, Damper, behind Exhaust Fan (%)	8840	4605	45749	14%	12%
H19*	Outlet burner temperature (°C)	52518	-51097	339099	100%	100%

Table 12. Sensitivity analysis results for the metamodel for seven comparable inputs factors, response dry-bulb temperature at the mill outlet, and their relative percentages.

Meta-model						
Symbol	Factors	μ^*	μ	σ	$\mu^* [\%]$	$\sigma [\%]$
M1*	Ambient dry-bulb temperature at Hoogoven harbour Tata (°C)	1.543	0.031	1.784	18%	15%
M2*	Barometer absolute pressure at Hoogoven harbour Tata (mBar)	0.652	-0.135	1.003	2%	4%
M3*	Relative humidity at Hoogoven harbour Tata (%)	0.557	-0.027	0.692	0%	0%
M9*	Differential pressure measurement cyclone (mBar)	0.566	-0.006	0.698	0%	0%
M16*	Exhaust stack, Vent Gas damper (%)	1.773	0.136	2.156	22%	21%
M17*	Recycle Gas damper, behind Exhaust Fan (%)	0.744	-0.180	0.818	3%	2%
M19*	Temperature measurement Burner hot gas outlet (°C)	6.049	0.108	7.808	100%	100%

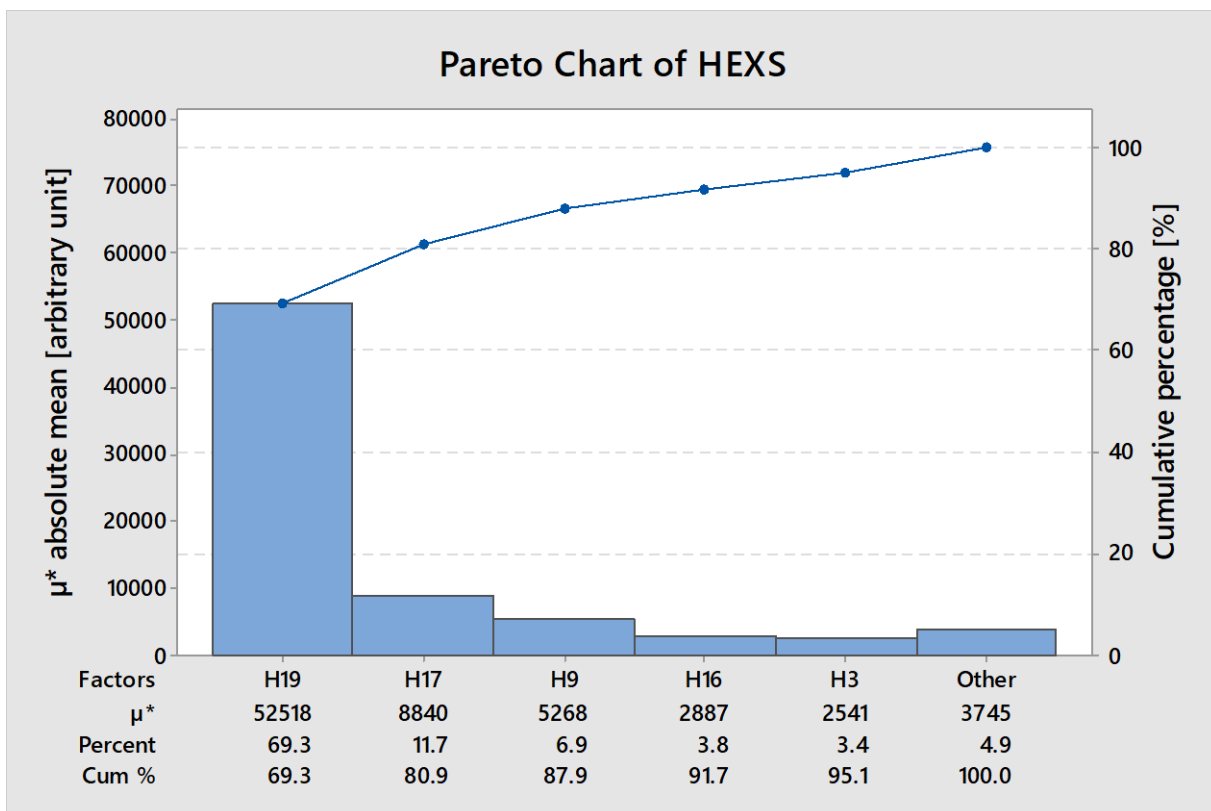


Figure 24. The Pareto bar plot of the absolute mean μ^* taken from the mean elementary effects for seven input factors utilized in the HEXS model. The factors are ranked in descending order of influence to the output “mill outlet dry-bulb temperature”.

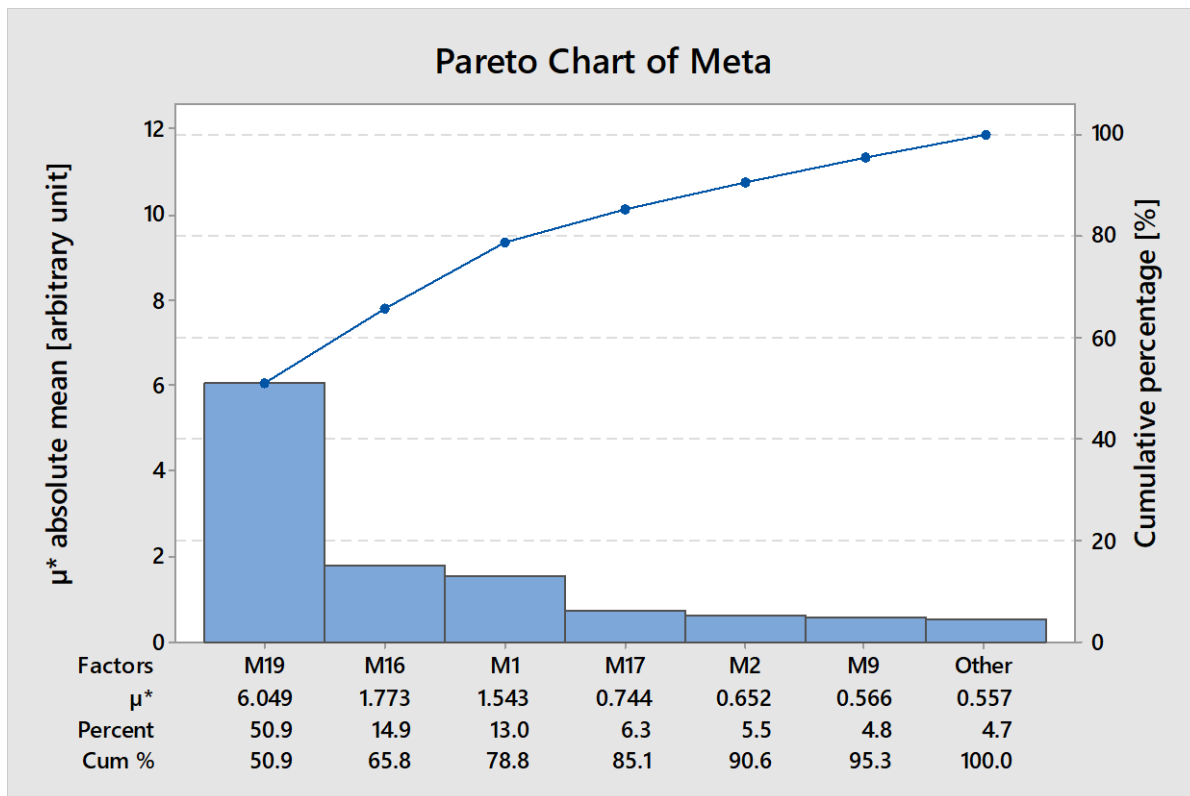


Figure 25. The Pareto bar plot of the absolute mean μ^* taken from the mean elementary effects for seven input factors utilized in the metamodel. The factors are ranked in descending order of influence to the output “mill outlet dry-bulb temperature.”

5.3 Metamodel sensitivity analysis

In the previous sub-section 5.2.4 only the SA results for comparable factors that are both in the HEXS and the metamodel are examined. In this section, the SA results for all input factors of the metamodel are examined, whereby an answer is provided for sub-question four: Which dominant factor(s) in the metamodel dominantly influence the mill outlet temperature?

The extended Morris method was applied to the metamodel (GPR Rational Quadratic, run in March 2019), with the response factor selected as the mill outlet dry-bulb temperature: a sample size $r = 200$. Each of the 44 input factors followed a uniform distribution between its extreme values, reported in Table 2 as presented in the Appendix 5.2 as the minimum and maximum. The minimum and maximum are derived from the actual operation conditions of the RMPP of the run of March 2019, more detail about these maximums and minimums is provided in Appendix 5.2. In the design space, each factor was varied across 32 levels ($p = 32$). A total number of $r(k + 1) = 9,000$ model evaluations were performed, where k is the number of input factors.

The results of the experiment are shown in the log-log screen plot in Figure 24, where the sensitivity measures μ^* and σ are plotted for 44 meta input factors. A Pareto bar plot is plotted in Figure 25. Figure 24 is presented as a log-log plot in which both the horizontal and vertical axes use logarithmic scales. By implementing this, more factors can be displayed in one figure. Although a log-log plot is enabled, some overlapping symbol names remain. Appendix 8.2 reports some zoomed-in plots that do not have these overlapping name tags, and thus each factor can be distinguished.

In the Pareto bar plot, the factors are ranked in decreasing order of importance according to μ^* , which is a measure of the overall factor of importance; this is shown in Figure 25. The plot is presented in Figure 24, where the sensitivity measures μ^* and cumulative percentage are given for 11 out of the 44 factors. Labels indicate the symbolic names of the factors and continue to approximately 70% of the cumulative percentage.

The remaining factors are grouped in the last bar of the chart and are named “other.” The actual description of the symbolic names can be found in Table 3, but can also be observed in in the Appendix 5.2, where the plot and table can be held next to each other for clarity. However, notably for this case, the order of importance that would have been obtained using σ instead of μ^* is highly similar. The results of the experiment show that each input parameter with a high value for the estimated absolute mean μ^* also has a high value for the estimated standard deviation σ ; in other words, none of the factors has purely linear effects. This can also be observed from the log-log screen plot in Figure 24, where all the factor points lie around the diagonal, which could also be observed from the conventional screen plots as shown in Appendix 8.2.

Several conclusions can be drawn by examining Figure 24 and Figure 25. A single factor that stands out is M5: the participation of rainfall in mm/hr. Furthermore, in the first group consisting of M29, the “system flow temperature, before the main fan,” and M41, the “the bag house inlet temperature,” these three factors are clearly separated from the others, indicating a dominant influence on the mill outlet dry-bulb temperature. It can be noticed from the Pareto bar plot that M5 has an absolute mean of 26.15, which accounts for a cumulative percentage of 25.0%. M29 contributes 11.47%, and M41 contributes 10.9%. Those three together account for a cumulative percentage of 47.3%, which indicates their relative influence compared with the other input factors. This study expected that when these factors are modified within the given input space, they will most dominantly influence the response.

This group is followed by a second single factor: M19, “the burner outlet temperature,” which contributes approximately 5.8% to the total cumulative contribution of the absolute mean. A second group consists of five factors that, in decreasing order of influence, are M42, the “bag house outlet temperature”; M25, the “mill inlet pressure”; M11, the “nitrogen injection at the mill inlet”; M21, “the mill motor power”; and M44, “the pressure of the natural gas supply.” This second single factor and group has a total cumulative contribution of 19.4%. Together with group one, they have a total cumulative contribution of influence of approximately 66.7%.

A third group consists of M22, “the current of the spinner separator”; M7, “nitrogen injection at the inlet of the cyclone”; and M16, the “exhaust stack damper valve position.” Then, the other factors follow immediately: M13, M14, M1, and so on. The absolute mean values for the reaming factor decrease smoothly without any discontinuity, indicating that distinguishing another group of influencing factors from a group of non-influencing variables is very difficult.

In the Pareto bar plot, the first 11 factors are estimated to account for 70.8% of the influence on the response of “the mill outlet temperature.” These first 11 factors are the “vital few” that disproportionately contribute to influencing the response. However, one should take care that these results are only valid for the model with the given input space.

From the experiments reported in the Appendix 8.1, this study noted that when the r trajectories were varied to 50, 100, 200, and 400, and these were repeated three times for each trajectory through the extended Morris method, these four factors provided consistent results for all experiments. In all 12 experiments, these four factors in order were identified as having the highest absolute mean value. For factors with lower absolute mean values, it was observed that their order of influence was less consistent, and their inconsistencies in order increased as their absolute mean values fell. In general, these are the factors that are close to the lower left corner of the screen plot in Figure 24.

When factors with large absolute mean μ^* are identified, they can be more extensively examined in an OFAT analysis. A distinction can be made whether the factor is either nonlinear or involved in interaction with other input factors. Currently, which of these two effects is present is unknown. Computational effort can be reduced for factors with a small absolute mean value. They can be fixed at a baseline value without significant loss of information. However, one should first determine whether that factor has an interaction effect.

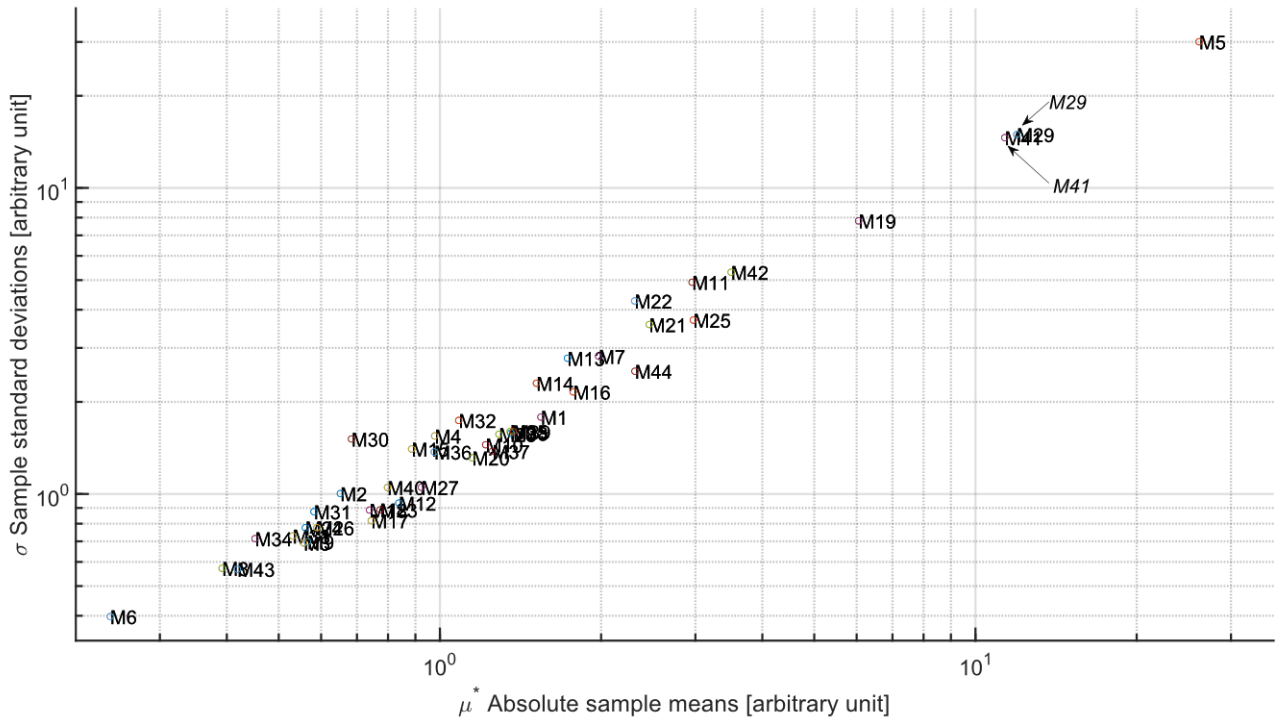


Figure 27. Graph displaying the extended Morris sensitivity measure μ^* , the absolute mean, and σ the standard deviation, for 44 metamodel input factors. The graph is displayed as a log-log plot to respond to skewness, as factors M5, M29, M41, and M19 have much larger values than the bulk data. Note that in the Appendix 8.2 shows that this plot does not contain factor effects of type “B” linear and additive factor (high absolute sample mean, low standard deviation) and would show in the right lower corner of the non-log-log plot. The corresponding description of each of the factors can be found in Table 3 (the table also presented in the Appendix 5.2, which therefore can be held next to this graph).

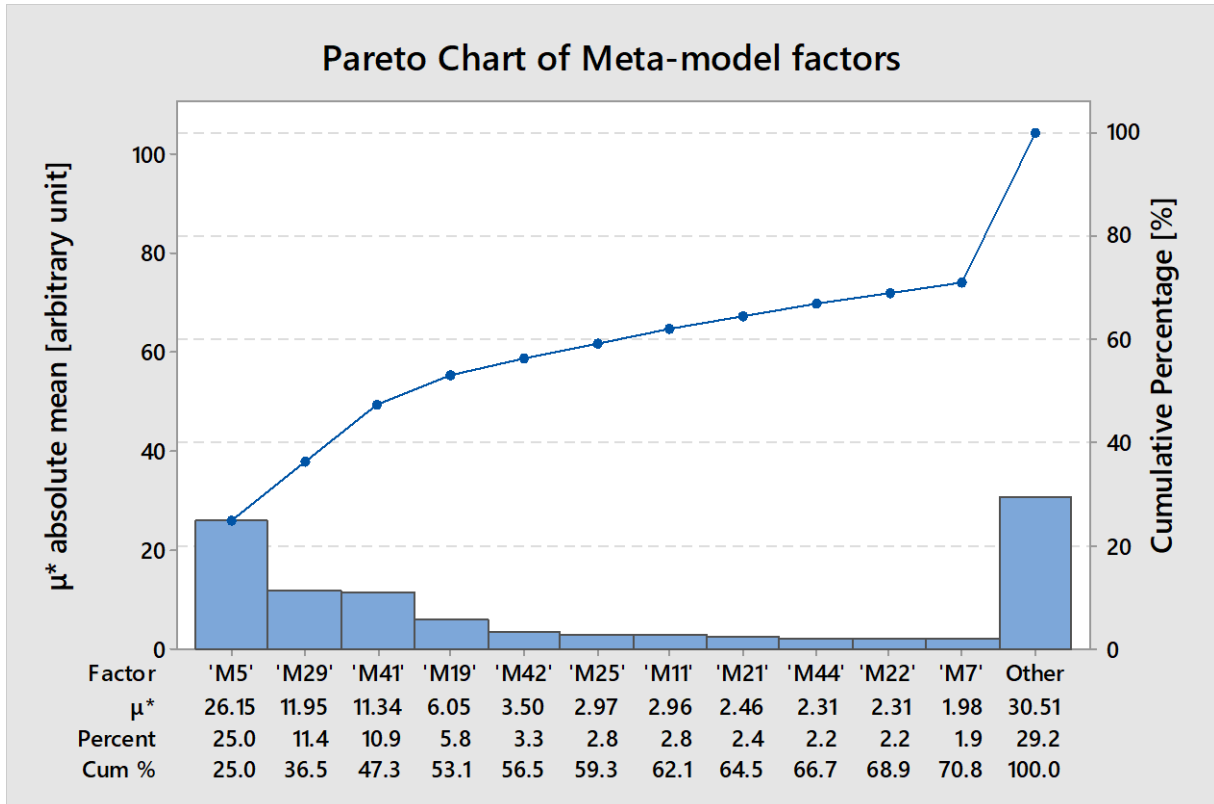


Figure 26. The Pareto bar plot of the absolute mean μ^* taken from the mean elementary effects for 44 input factors utilized in the metamodel. The factors are ranked in descending order of influence on the output “mill outlet dry-bulb temperature.”

5.4 HEXS model sensitivity analysis

In this section, an answer is provided to sub-question five: Which dominant factor(s) in the HEXS model dominantly influence the mill outlet temperature?

The extended Morris method was applied to the HEXS model, with the response factor selected as the mill outlet dry-bulb temperature and a sample size of $r = 50$. Each of the 20 input factors followed a uniform distribution between extreme values, as reported in Appendix 6.2 as the minimum and maximum. The minimum and maximum were derived from the actual operation conditions of the RMPP during the run in March 2019. In the design space, each factor was varied across 32 levels, and thus $p = 32$. A total of $r(k + 1) = 1,050$ model evaluations were performed; k is the number of input factors. The HEXS model must be returned to an initial baseline after each calculation, otherwise the previous result will influence the new simulations. Thus, each simulation was reset to an initial baseline. In Appendix 6.1 and 6.2, the factors, operating range, and baseline settings utilized in the SA are reported. The majority of these settings were derived from a steady state operation of the physical RMPP of the campaign in March 2019. For other settings that are not monitored by the plant and weather station, the initial values proposed by the HEXS model developer continue to be used, such as heat loss, fan efficiencies, material composition (gasses, coal), and tramp air.

The results of the experiment are shown as a log-log screen plot in Figure 28, where the sensitivity measures μ^* and σ are plotted for 20 HEXS input factors. A Pareto bar plot is presented in Figure 29. Figure 28 presents a log-log plot where both the horizontal and vertical axes use logarithmic scales, and through implementing this, all factors can be displayed in one figure. In a conventional plot that uses a linear grid, factors and name labels are displayed as overlapping each other, and reading and distinguishing multiple groups of factors are challenging. Multiple screen plots that zoom in on sections where overlapping occurs can be read in Appendix 7.2.

Factors are ranked in decreasing order of importance according to μ^* , which is a measure of the overall factor of importance. This is shown in the Pareto bar plot presented in Figure 29 where the sensitivity measures μ^* and cumulative percentage are given for 11 of the 20 factors, with labels indicating the symbolic names of the factors up to approximately 90% of the cumulative percentage. The remaining factors are grouped in the last bar of the chart as the “other” category. The actual description of the symbolic names can be found in the table presented in Appendix 6.2. However, notably in this case, the order of importance that would have been obtained using σ instead of μ^* is highly similar. Results of the experiment show that each input parameter with a high value for the estimated absolute mean μ^* also has a high value for the estimated standard deviation σ . In other words, none of the factors have purely linear effects. This can also be observed in the log-log screen plot in Figure 28, where all the factor points lie around the diagonal, which could also be observed in the conventional screen plots shown in Appendix 7.2.

A number of conclusions can be drawn by examining Figure 28 and Figure 29. The first group consists of H19, the “burner outlet temperature,” and H8, the “mill outlet pressure.” These are clearly separated from the other factors, indicating a dominant influence on the mill outlet dry-bulb temperature. It can be noticed from the Pareto bar plot that H19 has an absolute mean of 52,518, which accounts for a cumulative percentage of 28.4%. H8 contributes 26.7%, and together with H8 accounts for a cumulative percentage of 55.2%. This study expected that when these factors are modified within the given input space, these factors will most dominantly influence the response.

A second group consisting of four factors that are quite influential is as follows (in decreasing order of importance): H10, the “main fan efficiency”; H12, the “H₂O content remaining in the coal after drying”; H20, the “main process flow rate”; and H17, “valve position of the recycle gas damper, behind exhaust fan.” This second group had a total cumulative contribution of 23.8%, and together with group one, they had a total cumulative contribution of influence of approximately 79%.

A third group consisted of H4, “feed rate of the coal”; H9, “differential pressure in the cyclone”; and H5, the “H₂O content of coals as-fed.” Then, the other factors follow immediately: H7, H11, and so on. The absolute mean values for the remaining factor decrease smoothly, without discontinuity, indicating that

distinguishing another group of influencing factors from a group of non-influencing variables is very difficult.

In the Pareto bar plot, the first seven factors are estimated to account for 81.8% of the influence on the response “the mill outlet temperature.” These first seven factors are the “vital few” that disproportionately contribute to influencing the response. However, one should take care that these results are only valid for the model with the given input space.

When factors with large absolute mean μ^* are identified, they can be more extensively examined in an OFAT analysis, where the distinction can be made whether the factor is either nonlinear or involved in interaction with other input factors. Currently, which of the effects is present is unknown. Operatively, when factors have a small absolute mean value, they can be fixed at a baseline value without significant loss of information, allowing the computation effort to be decreased. However, one should first determine whether that factor has an interaction effect.

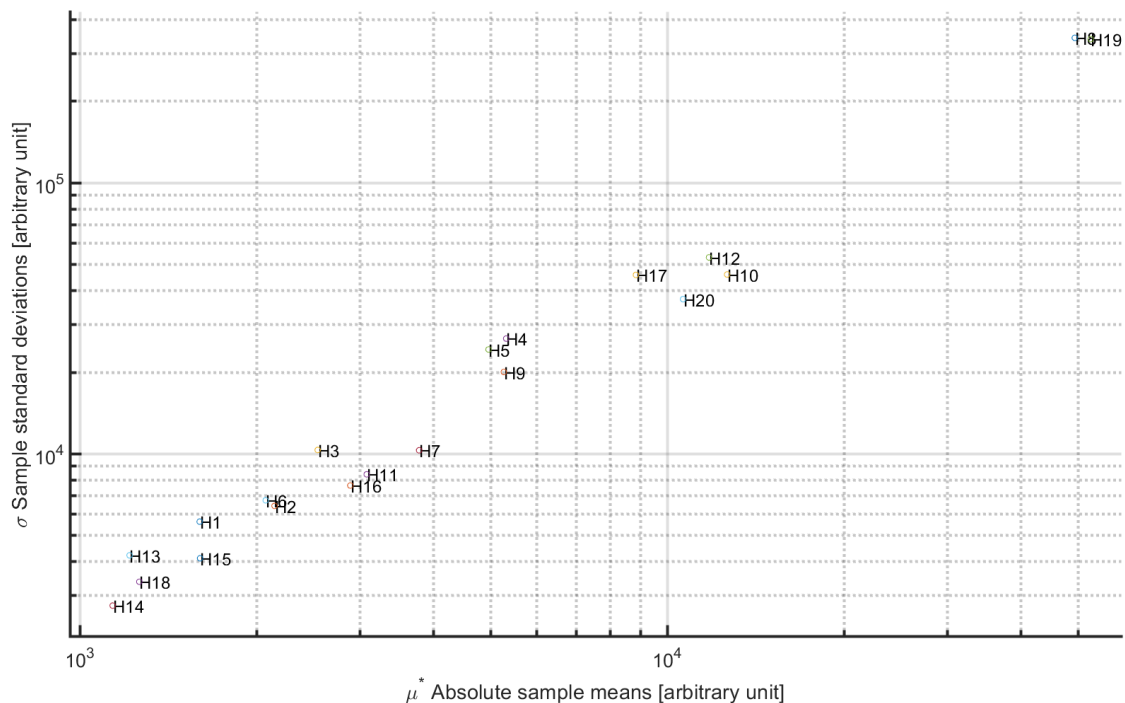


Figure 28. Graph displaying the extended Morris sensitivity measures μ^* , absolute mean, and σ the standard deviation, for 20 HEXS model input factors. The graph is displayed as a log-log plot to respond to skewness, because factors H8 and H19 have much larger values than the bulk data. Note that figure shown in the Appendix 7.2 shows that this plot does not contain factor effects of type “B” linear and additive factor (high absolute mean, low standard deviation) and would show in the lower-right corner of the non-log-log plot. The corresponding description of each of the factors can be found in the table presented in the appendix 6.2, can be held next to this graph for description of their symbolic names.

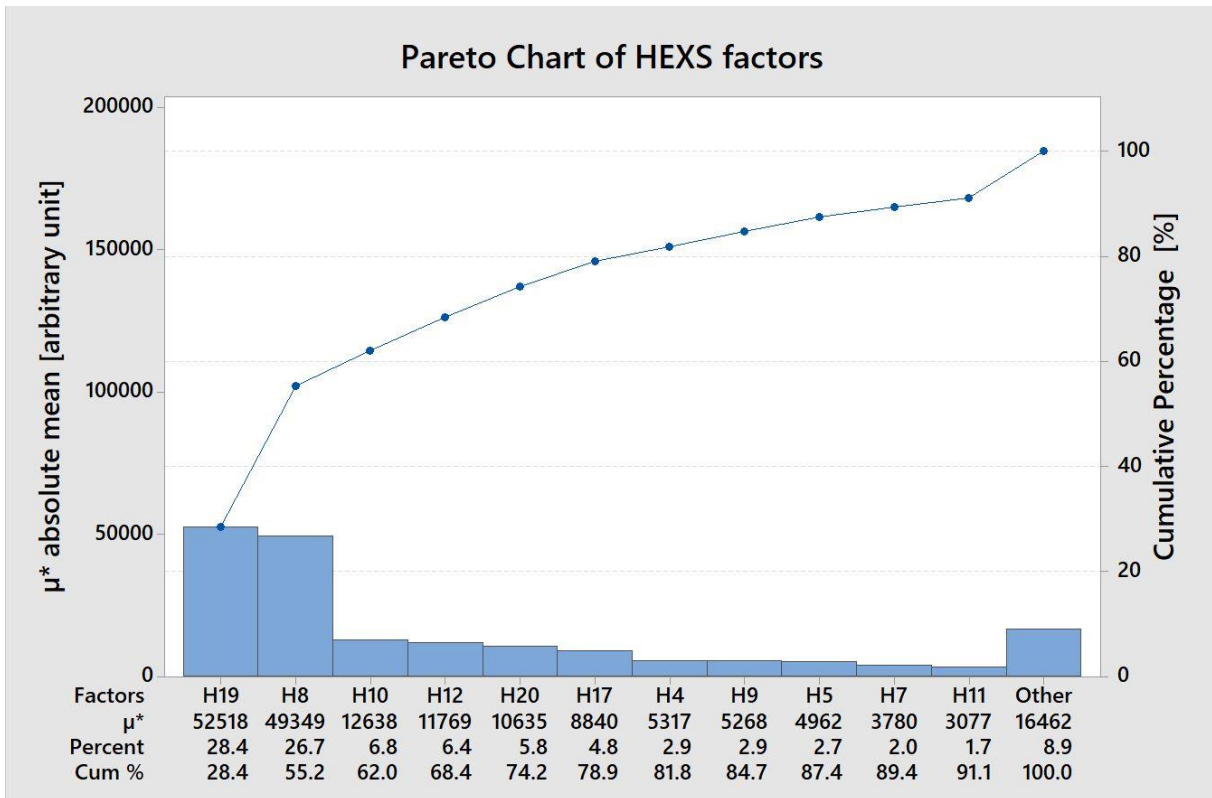


Figure 29. Pareto bar plot of the absolute mean μ^* taken from the mean elementary effects for 20 input factors utilized in the HEXS model. The factors are ranked in descending order of influence to the output “mill outlet dry-bulb temperature.”

5.5 Potential operational functional failure condition (PoC)

Sub-section 2.4 discussed how drying can stop, that is when the mill outlet temperature of the drying medium drops below the dew point temperature. When the drying stops, the evaporated liquid water from the coal can return into the coal. This can result in coal leaving the RMPP system with a high moisture content, which does not meet the specified requirements. Coal with too high moisture content can affect subsequent production steps in the HIsarna iron-making process.

Because reducing the moisture of coal at a continuous rate is one of the main functions of the RMPP, not being able to evaporate moisture from the coal is an unfavourable operation scenario. If no counter measures are taken, unfavourable operation conditions can lead to functional failure (FF) of the system. However, which operation conditions can lead to FF of the RMPP? One of many FF conditions is discussed in this section—the inability to dry coal—because drying coal is one of the main functions of the RMPP. First, this section provides a brief introduction of the physics involved in drying.

As previously discussed in Sub-section 2.4, the capacity of the drying medium to absorb and carry away moisture from the coal determines the drying rate, incubation time of the coal, and the maximum amount of moisture to be evaporated. Two elements are essential in this process: the drying medium temperature and its flowrate. In general, the higher the temperature of the drying medium, the greater its vapor-carrying capacity. For this example, maintaining a sufficient flowrate of the drying medium is assumed not to be a limiting factor in the drying process as long as the fans are functionally operating and there is no blockage or leakage within the piping. Under this assumption, the flowrate is neglected in this analysis.

In the SA, this study investigated which factors and accompanying assets influence the dry-bulb temperature of the mill outlet, because this is monitored in the RMPP. However, this does not provide all the necessary information for determining which factors can lead to FF. The other element required is the dew point temperature, also known as the saturation temperature. This is the point at which the drying medium is saturated with water vapor. If the difference in temperature, ΔDT , between the drying medium temperature T_{gas} and the dew point temperature T_{dp} drops to or below zero, then the drying comes to a hold, and this can lead to FF. This is expressed with the following expression for a control volume with a constant pressure:

$$\Delta DT = T_{gas} - T_{dp} \quad (5.2)$$

$$T_{dp} = T_{gas} - \left(\frac{100 - RH_{system}}{5} \right) \quad (5.3)$$

The dew point is related to the relative humidity (RH; in %), and the approximation for calculating the dew point temperature was proposed by M. G. Lawrence, as reported in equation 5.3 [60]. This is the RH within the RMPP system, not to be confused with the ambient RH presented earlier. The RH is not a measurement in the existing RMPP system, but it can be monitored within the HEXS model as the drying medium temperature.

When the drying medium temperature comes below the dew point, that is when the ΔDT becomes negative. Vaporized water will begin to condense on solid surfaces, such as those of coal particles. The moisture can be reabsorbed by the coal, which increases its moisture content. This phenomenon, where the medium and dew point temperatures cross each other, is investigated with OFAT analysis in the HEXS model for its input factors. The responses given for each of the factors and its operating range are screened for the condition, where ΔDT is lower than or equal to zero. By applying this condition for the operating range in which each factor is varied, a screening can be performed where factors and operating conditions that can lead to a potential FF are identified.

To provide an example, the H₂O content of fed coal is selected as the input factor. The fed raw coal has an initial H₂O content of 1% , which is increased to 25% in 100 steps. For each incremental step, the HEXS model is returned to its baseline values, as stated in Appendix 6.1. When the 100 steps for one factor are executed, the adapted “HEXS model V4” continues to the next factor in an OFAT fashion. Thus, multiple conditions for different input factors can be examined.

However, for this example, only one input factor is graphically reported; other conditions that can lead to potential FF of the inability to dry are summarized.

The objective is to dry the fed coal to a 1% moisture content. In the process, changes in the drying medium (solid grey line), dew point (solid blue line), and difference in temperature (solid red line) are observed, as shown in Figure 30. From this figure, this study noted that when the temperature of the drying medium (mill outlet temperature) crosses that of the dew point temperature, the ΔDT changes to zero; that is, at 15% H₂O. As the water from the coal evaporates, the moisture-carrying capacity of the drying medium becomes saturated by water vapor. When saturated, the RH is 100% and the dew point and medium temperature are equal. The driving force for drying comes to a hold.

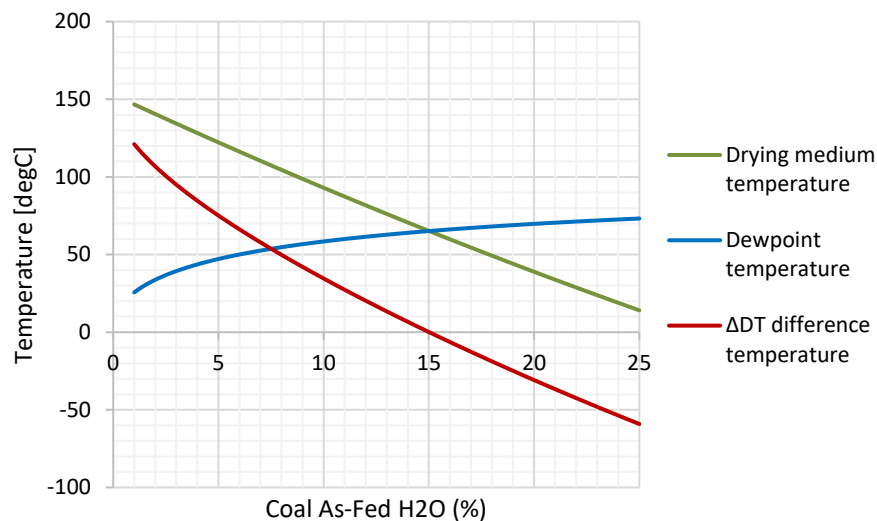


Figure 30. Functional failure caused by increasing H₂O content of the fed raw coal, temperature measured at the outlet of the mill. The drying medium temperature is the mill outlet temperature.

Here, the opportunity arises to utilize the SA and identify which factors are dominantly influencing the ΔDT . The identified factors can, for example, be used to counteract or delay this potential FF by adjusting the RMPP operating settings of the dominant factors accordingly. However, this proposed method of identifying factors for process adjustments to delay or prevent unwanted behaviour is left for future research. Additional insights are presented in the Appendix 10.

Other process conditions (scenarios) were examined with the HEXS model V4 that resulted in a mill outlet temperature less than or equal to the dew point, and then the driving force for drying stops. These potential FFs of the inability to dry according to the given baseline settings are:

- A. if the fed coal contains more than 15% moisture H₂O
- B. if the mill heat loss is greater than 507 kW,
- C. if the valve position of the recycle damper is opened less than 26%,
- D. when the burner outlet temperature is lower than 309 °C,
- E. when the spray water is more than 0.3 metric tons per hour, and
- F. when the mill inlet feed rate of the coal is larger than 8.6 metric tons per hour.

If the HEXS model reflects the behaviour of the HIsarna RMPP, which it currently does not, then these operating conditions with the baseline setting can lead to FF. For the given example, this will mean that when the conditions reach that state, the RMPP will be unable to dry at the mill outlet. The current RMPP only monitors conditions B and C. For these conditions, alarm can be raised, according to which the process should be adjusted to prevent FF. Furthermore, if how to prevent the FF is unknown, the SA process can be performed, which will allow influential factors to be identified for that condition. Focusing on the factors that can be controlled allows a potential FF to be counteracted through adjusting the settings of the influential factor. A brief example is reported in Appendix 10.3, which presents how the SA may be utilized to counteract a potential FF.

5.5.1 Special note for a potential safety hazard

Through examining inert conditions in the HEXS model's behaviour for the input space, as stated in Appendix 6, this study observed that a potential operational condition may exceed a safety threshold at another location in the plant than where it is currently monitored in the actual RMPP. More details about the inert operations of the RMPP is reported in Section 2.2.6.

At nominal capacity, the whole RMPP plant should operate under inert conditions; that is, when the dioxygen (O_2) content in the drying gas is usually below 7% (Hi alarm at 7%, and a HiHi alarm at 9%). An inert environment can eliminate the oxygen through a medium ratio (process gas and fine particles of coal), which makes flame propagation in the coal dust cloud unlikely. Reducing the oxygen through a medium ratio lowers the risk of coal dust explosions within the RMPP. However, when the O_2 levels exceed the threshold in combination with the coal dust as a fuel and heat source for combustion, an explosion can occur, which can cause damage.

By varying the input factors in the HEXS model, the O_2 levels can be monitored at the burner outlet, mill inlet, cyclone outlet, and baghouse outlet. This allows an examination of the operating conditions and positions in the system at which O_2 levels exceeds the threshold limit. The location might be of special interest because O_2 levels are currently only monitored behind the exhaust fan in the actual RMPP. In the performed experiment, this study observed that the O_2 level between the main fan and mill was 10.38% (**exceeding the HiHi safety threshold of a maximum 9% of O_2**), whereas the level at the currently monitored position behind the baghouse was 8.39%, which is below the HiHi limit. This difference is 1.24 higher than that at the currently monitored position. Although not detailed in this study, various simulated scenarios exist in which the O_2 levels behind the main fan exceed that of behind the baghouse monitoring location.

By obtaining information at which location(s) concentrations are expected to be highest for certain operating conditions, it allows to allocate the sensor position more delicately. Through positioning the sensor i.e. where O_2 concentrations are expected to be highest. Therefore, the abovementioned discussion demonstrated that the HEXS model can be utilized to allocate resources.

Notably, although this potential safety hazard is recognized, the HIsarna RMPP is equipped with pressure relief valves. When a coal dust explosion occurs, internal pressure rises. With a pressure relief valve, this pressure is relieved through opening the valve, which reduces the risk of damage to the system.

5.6 General discussions

When attempting to extract meaningful information for maintenance or process configurations from an existing process model, data and deployed sensors of a production plant, one may be faced with an overwhelming number of data points. This study offers a systematic approach for extracting meaningful information from complex systems, as well as identifying operating conditions, factors, and assets that can cause potential FF conditions in a complex production system. The proposed strategy was assessed by means of a PoC on a commercial RMPP at HIsarna. However, the workflow strategies for the SA (Figure 10) and metamodel (Figure 13) are flexible in terms of being adopted into alternative systems.

If the relationships and relative importance of parameters are understood, the model can reasonably reflect the real-world process. The SA philosophy utilized in this study can provide insights into the dominant factors and controls of the actual system and subsystems. In this section, the three general steps of this study are discussed, namely validation, sensitivity analysis and potential operational FF conditions (PoC).

5.6.1 Constructing the metamodel

In the validation experiment reported in Section 5.2, supervised machine learning was utilized for validating the HEXS model. SML algorithms were constructed based on monitored RMPP data, which were utilized to obtain a metamodel that described the behaviour of the mill outlet temperature.

The regression algorithm used for constructing the metamodel was evaluated against 19 other types of algorithm, as discussed in Section 4.4. This study observed that the GPR Rational Quadratic equation had a CV-RMSE of 0.99 °C and was the algorithm with the lowest value when trained on 70% of the dataset. Although each of the individual regression algorithm methods may have its own restrictions, most are not bound by them. The models can be pre-processed and tuned using techniques for minimizing the impact of model assumptions and reducing the forecasting error. Therefore, it is feasible that if other algorithms are tuned, they may perform equally or superior to the GPR Rational Quadratic regression algorithm.

Furthermore, the construction of the metamodel was limited to available data. Although more than 1e6 observations were incorporated among 44 input factors, the data extracted only represented an 18-day period within a single RMPP operation campaign. A dataset over a more extensive time period may result in a different metamodel as well as allow a model that can predict a larger variety of conditions more accurately. As observed in Figure 19 (Section 5.1), accuracy reduced when forecasting the behaviour of unseen conditions.

5.6.2 Validation

The proposed validation procedure was not ideal; as was expected, the response behaviour of both models was not always aligned. An alternative method is to derive energy and mass balances of a small subsection of the actual RMPP, such as the mill. Once established, these could be compared with the energy and mass balances of the HEXS model. However, obtaining these balances would require currently unmonitored elements, such as material flow rates. The required sensors were prohibitively expensive and unavailable, which resulted in the alternative validation method presented in this study.

5.6.3 Sensitivity analysis

As described in section 3.2.6, the extended Morris sensitivity analysis was applied to the HEXS model and metamodel to examine the relative influence of an input parameter on the output “the mill outlet dry-bulb temperature.”

An alternative method of executing the Morris method routine was presented by implementing the space-filling criteria for evaluating a design space proposed by Forrester et al. (2007). In this proposed method, two random generated design spaces were constructed and evaluated in terms of uniform space-filling. In general, as the number of design spaces generated increases, the probability also increases that an overall spread is obtained over the input space “space-filling.” Thus, the larger the number of scenarios, the more likely the space is to be filled, and thus, the better the accuracy of the model that is fitted.

The design space with the highest maximin values was considered to have the most uniform space-filling and was utilized for the SA. However, one minor concern related to this application in the extended Morris sensitivity analysis should be addressed, which is that only two designs were evaluated; it could be the case that both had poor space-filling. The probability of poor space-filling reduces when more than two design spaces are evaluated. Therefore, this study recommends evaluating more than two designs spaces in terms of uniform space-filling, but these additional evaluations come at the cost of considerable computational effort.

Furthermore, two alternatives were given for presenting the extended Morris method’s results. The conventional screen plots use a linear grid, such as in Figure 11 (Section 3.2.4). The alternative reported in Figure 28 (Section 5.3) utilized a double logarithmic axis rather than linear. This allows a larger number of variables to be displayed with less overlapping labels in one plot, and enables distinguishing multiple groups of factors, as discussed in Section 5.3. Moreover, considering the double logarithm of the output assists in splitting the input parameters in groups for discussing their importance; however, this is at the cost of the loss of negative signed values due to the logarithm. Notably, investigating using different screen-plot scales is not required for performing additional simulations; the computed results from the SA were utilized, which only required postprocessing task of adapting the linear axes to the logarithmic scale.

Second, a Pareto plot was utilized with the absolute mean; the cumulative percentage allowed quantification of the relative influence of a factor compared with the other input factors, thereby allowing a distinction to be made between the “vital few” and the “useful many.” However, the Pareto plot did not include the standard deviations, and thus no classification could be made about its effect (A, B, or C) as defined in section 3.2.4; this can be performed with the screen plot.

As described in Section 5.3, the SA determined which factors have the highest mean absolute mean values because they are expected to have the greatest influence on the mill outlet response. The metamodel SA demonstrated that when 44 inputs were considered, four factors accounted for a cumulative percentage of 53.1% of the influence on the response; ranked in order of descending influence, these are:

- M5, ambient rainfall (mm/hr);
- M29 “systems flow temperature °C”; M41 “baghouse inlet temperature in °C”; and M19 °C “the burner outlet temperature °C.”

The SA with the HEXS model reported in Section 5.4 was executed with 20 input factors, resulting in the four factors with the highest absolute mean values having a cumulative percentage of 68.4%. The four factors, in descending order of influence, are:

- H19 the “burner outlet temperature in °C”;
- H8 the “mill outlet pressure in mbar”; H10 the “main fan efficiency in %”; and H12 the “H₂O content remaining in the coal after drying.”

From the SA analysis for both models, this study observed that the four factors with the highest absolute mean values differed between models. This could be because each model had a separate set of input factors. Only those factors that are monitored at a similar location in the process, have an equal physical quantity, and have an equal unit can be compared. A comparable factor is the burner outlet temperature, which was recognized in both models as one of the four factors with a dominant influence on the mill outlet

temperature. Furthermore, in the validation experiment (Section 5.2.4), only comparable factors of the HEXS- and metamodel were applied in a SA. In this experiment, both models identified the burner temperature as the most dominant factor.

The values of the sensitivity analysis using the extended elementary effects method were presented. In any case, for a given RMPP being simulated with a specific modelling tool, no general sensitivity could be derived, and the results depended on the input parameters and their range with fixed factor values and varying factor and/or values. Therefore, the SA must be executed for each particular situation, in relation to the modelling objective, with a careful choice of variation intervals. Thus, the SA results are only valid for this particular scenario.

5.6.4 Potential operational functional failure condition (PoC)

FF have demonstrated use with respect to operational and safety related failure scenarios. The ability to detect potential FF operating conditions without a physical system is encouraging simulating these behaviours in the design phase of a plant. When the operating windows are known, they can be used to enhance the system control logic and to monitor particular conditions. The control logic can be used to adjust operating conditions so that unwanted process conditions that potentially could lead to an FF can be prevented. If conditions cannot be monitored continuously (for example, by a sensor), they could be included in a maintenance plan in which they are frequently inspected.

It is noted in sub-section 5.5 that when $T_{gas} \leq T_{dew}$ drying ceases, which can lead to a potential FF. From the SA in sub-section 5.6.3 it is observed that for comparable input factors the burner outlet temperature is most dominant in influencing the MOT (T_{gas}). In sub-section 5.5 and Appendix 10, it is noted that burner outlet temperature can lead to a potential FF condition when it is lower than 309 °C; at this burner temperature the MOT is below de dew point temperature. The burner temperature is identified as dominant influencing factor by the SA before it was identified as a potential FF failure condition. This example illustrates that pre-identified influential factors by SA can cause a potential FF condition.

In sub-section 5.5.1, it is discussed that too high O₂ levels in the drying circuit can lead to an explosion and therefore lead to a potential FF of the system. Furthermore, it is demonstrated that HEXS model can simulate operation conditions such as O₂ levels at multiple locations in the plant. As O₂ levels behaviour are modelled at multiple sub-system locations, each location can now be compared with respect to each other. Therefore, this comparison can assist in allocating resources (i.e. sensors) to those locations at which O₂ levels are expected to be the highest.

5.7 Remarks and observations

As noted in chapter 1 of this work, the number of variables involved in a complex system is vast and they can operate according to nonlinear interactive trends, so a system can supersede human understanding. Utilizing process models (physics-based or data-driven) with a sensitivity analysis can allow researchers to obtain valuable insights into the influence of input parameter variation in relation to the output in a complex system, thereby allowing potential FF conditions to be identified and prevented.

Furthermore, during the execution of this study, it was observed that advantages and disadvantages exist for supervised machine learning algorithms compared with physics-based models, and these are addressed next. SML can incorporate large quantities of disparate data into a unified automatic generated metamodel, whereas in the HEXS physics-based model, multiple algorithms are manually constructed. However, where the HEXS model derives multiple outputs at once (due to the set of equations), the learned algorithm can only produce one output. There is limited knowledge required about the physics to generate metamodels, and thus they can be constructed relatively quickly compared with physics-based models. This is because, for the physic-based model, proper understanding of the system physics and behaviour is required before it can be modelled. Due to this, the HEXS model has superior interpretability compared with the metamodel; this is because from the HEXS model, physical relations and quantities can be observed instead of regression coefficients that do not have a direct physical meaning. To add, for example, an additional factor to the computational space, a data-driven model can be constructed when data are available on the additional input factor. However, to incorporate a new feature within the HEXS model, the physical relation should be first known as well as how it influences the already incorporated relations. When known, it can be manually

incorporated, whereas the metamodel can autonomously learn this additional feature. Where the metamodel computes a response within a second, the HEXS model requires several seconds or even minutes. Another benefit for a physics-based model is that in general, nonhistorical (operational) data are required to construct a model, whereas for SML regression algorithms, historical data are required to establish a model. Thus, the data-driven model is somewhat limited in terms of forecasting because in its historical design space, as demonstrated in Figure 19 of Section 5.1, the prediction accuracy reduces when forecasting operational conditions that are not within the historical data. Table 13 summarizes these observations.

Table 13. Comparison between a physics-based and data-driven model. Note that these observations are broad generalizations that may change frequently according to circumstances.

	HEXS model (physics-based)	Metamodel (data-driven)
Time required to construct a model	III	II
Requires process knowledge	III	I
Interpretability	III	I
Easy to incorporate additional design feature(s)	I	III
Computational effort for a prediction	III	I
Requires historical data	I	III

I Low; II Medium; III High

Based on this study's findings and shortcomings, the following aspects are noted. Each of the aspects is preceded by an alphabetic letter that corresponds with a recommendation presented in sub-section 6.3.

- A. In the executed experiments, a relatively small dataset was utilized for constructing the metamodel. To construct a metamodel that can describe the behaviour of a larger operating space, a larger dataset is required. In the process, this study found that some dominant influencing responses were related to ambient weather conditions, such as the amount of rainfall as can be observed in section 5.3. As the variation in rainfall utilized in the metamodel is not a representative data set for all the weather seasons, the behaviour might not be correctly reflected in the metamodel.
- B. In the proposed extended Morris SA, as described in sub-section 3.2.5, a minor application-related concern is identified. In this proposed method, two randomly generated design spaces were constructed and evaluated in terms of uniform space-filling. The one space with the highest maximum values was considered to have the most uniform space-filling and was used for the SA. The concern is that only two designs were evaluated, and both could have poor space-filling. The probability of having poor space-filling can be reduced when a larger number of design spaces is evaluated. However, these evaluations will come at the cost of considerable computational effort.
- C. As discussed in section 5.5, drying ceases when the drying medium temperature is less than or equal to the dew point temperature, assuming that the drying medium flowrate is sufficient. Currently, within the HIsarna RMPP system at multiple locations the drying medium temperature is monitored. However, no measurement system or method exists to derive the dew point temperature. If the dew point is monitored, it can indicate if the RMPP is able to continue to evaporate moisture or the drying medium is saturated by water vapor.
- D. Note that during the research this study found that determining the proper sensors to implement in an RMPP was highly challenging. That is because any obstacle (i.e. a sensor) located in the drying RMPP medium circuit is exposed to the highly abrasive flowing coal media.
- E. The investigation in sub-section 5.5.1 revealed that the simulated O₂ concentration between main fan and mill is 23.7% higher than that at the actual monitored position, which is behind the baghouse. Therefore, the current position to measure O₂ levels in the actual RMPP might not be the optimal position, to indicate if the O₂ concentrations within the RMPP circuit exceed the threshold limit. Too high O₂ concentration in combination with coal dust as a fuel and heat source for combustion can lead to increased risk of an explosion, which can cause damage.

- F. In this work, it is demonstrated that the Pareto plots allow to distinct between the “vital few” and the “useful many” factors estimated of having the most influence on the response parameter. It is expected that this proposed distinction method by combining the Pareto plot with a SA might be of interest to other research fields.

6 Conclusions and recommendations

In this chapter, first the answers are given to the research sub-questions as stated in sub-section 1.4.2. Then an answer is given to the main research question, and lastly the recommendations are reported.

6.1 Answers to the research sub-questions

1. How can the HEXS process model be validated in relation to the physical response behaviour of the RMPP?

Validation can be performed by comparing a response behaviour of the HEXS model with the same response of the RMPP. For each response the difference between the HEXS response and the actual response is monitored. The smaller the difference, the more valid the HEXS model is considered to be. In this work the MOT is selected to be the response and operational data in combination with SML is utilized to construct the behaviour of the MOT of the RMPP. If the (meta)model constructed by SML from the operational data is sufficiently accurate then it can be utilized in the validation process.

2. How accurately can the metamodel describe the behaviour of a selected MOT response of the RMPP?

To validate the HEXS model's behaviour against the HIsarna RMPP, a metamodel was constructed of the mill outlet dry-bulb temperature. The accuracy of the metamodel is determined by the RMSE. The metamodel trained by 100% of the data set by a GPR Rational Quadratic algorithm, had an RMSE of 0.52 °C for predictions; that is, an error of approximately 0.4% of the entire operating domain. Because an RMSE of 0.4% is relatively small compared with its operating domain, it is expected that the metamodel can simulate representative behaviour of the mill outlet temperature of the physical RMPP within the operating range of its training dataset.

3. Do the responses (mill outlet temperature) of the HEXS model and metamodel agree when given a similar input?

Using the trained metamodel, three validation experiments were performed between the HEXS model and metamodel to check the agreement.

- I. In the first validation experiment described in Section 5.2.2, the signs of the linear regression equation gradient coefficients were compared. Four out of seven had similar signs, whereas the other three had opposite sign.
- II. For three of the seven factors, it was expected that the linear trend in the mill outlet temperature would move in a particular direction when the input was varied. These expectations were confirmed by both models' responses, because each of the three factors trendline moved in the expected direction.
- III. The third and last validation experiment was performed by comparing the SA results for both models in Section 5.2.4. Both models identified the burner outlet temperature (H19 and M19) as the most influential. However, for factors with a lower absolute mean than the burner outlet temperature, their rankings in descending order of influence to the response were not similar between models. Thus, this study concluded that both models could identify the most influential factor for the given input, but factors with lower absolute mean values did not have the same order of influence.

During the three validation experiments, the majority of the checked aspects of the HEXS model, metamodel, and expected behaviour were similar. This indicated that the HEXS model in general was valid; However, a minor amount of disagreement existed between the three validation experiments. Therefore, no evident conclusion can be drawn that for all input conditions the HEXS model represents the behaviour response of the MOT of the real RMPP. Therefore, without modification of the HEXS model, the SA results and potential failure cannot be utilized for direct practical applications. However, they still provide an indication of potential factors that influence the response.

4. For similar input factors, which dominant factor(s) in the HEXS model and metamodel are in agreement and dominantly influence the mill outlet temperature?

An extended Morris SA method is performed on the case study of a commercial RMPP. The conventional Morris method does not provide a normalization of ratings for the relative contribution of influence with respect to other compared factors. The proposed extended Morris method accounts for this by allowing the comparison of normalized values of different models, such as the SA results from the HEXS and data-driven model. Normalization is achieved by utilizing the cumulative percentages of a Pareto plot. The cumulative percentage is demonstrated in the Pareto plots in Section 5.2.4 Figure 24 and Figure 25, that allow to distinct between the “vital few” and the “useful many” factors estimated of having the most influence on the response parameter. In the SA of Section 5.2.4 the burner outlet temperature is identified to be the most influential factor to the MOT by both models, supporting the assumption of repetitive validity. However, the burner outlet temperature is the only one out of seven that is a dominantly influencing factor and is in agreement by influential rank in both models SA.

5. Which dominant factor(s) in the HEXS model and metamodel dominantly influence the mill outlet temperature?

The analysis of the extend Morris results identified the dominantly influencing HEXS model input factors to the MOT. The four most dominant factors, in descending order of influence, are:

- H19 the “burner outlet temperature in °C”;
- H8 the “mill outlet pressure in mbar”;
- H10 the “main fan efficiency in %”; and
- H12 the “H₂O content remaining in the coal after drying.”

The four factors of the metamodel model dominantly influencing the MOT, in descending order of influence, are:

- M5, ambient rainfall (mm/hr);
- M29 “systems flow temperature °C”;
- M41 “baghouse inlet temperature in °C”; and
- M19 °C “the burner outlet temperature °C.”

6. Which HEXS model input factor(s) and operating conditions can lead to a potential functional failure condition in RMPP operations?

The analysis demonstrated that the proposed methodology is capable of identifying potential FF conditions, these are reported in sub-section 5.5.

6.2 Answers to the main research question

The main question of this research study was as follows:

- How can a process model be utilized for predictive maintenance?

The presented work demonstrates that a process model can assist PdM by defining potential thermodynamic related failure conditions and identifying their related dominant influencing factors.

This research aimed to create a strategy to identify the influence of input operation condition on the systems output, that may be critical for systems operations. Based on the results discussed above, a (HEXS) process model can be utilized to determine potential FFs for RMPP operations. The dominant influencing factors linked to the FF can be identified using the proposed extended Morris method. By identifying potential FF conditions and influencing factors, it can be utilized for PdM to gain insights into potential FF conditions and the factors that dominantly influence the response. The process model can be exploited for other thermodynamic-related potential FFs of the RMPP. However, there is no process model that incorporates all the details of complex natural phenomena, and therefore the results from the analysis might not include the desired level of detail. As a result, the model response is limited in detail for the factors and mechanisms included in the model(s).

In summary, it was possible to identify the most dominant factor for the modelled mill outlet temperature, that is relevant for a potential FF condition. Both the HEXS model and metamodel identified the burner outlet temperature as being the most dominant for seven equal input factors. Furthermore, this study found that the process model could be utilized to identify multiple operation conditions that can lead to a potential FF operating condition of the RMPP.

6.3 Recommendations

Based on this study's findings and shortcomings, the following recommendations are given for potential future research or as implementations for monitoring potential failure conditions in an RMPP. Each of the aspects is preceded by an alphabetic letter that corresponds with the discussion that is presented in subsection 5.7.

- A. It is recommended to create a metamodel that considers the influence of all-weather conditions and therefore a representative dataset of all seasons should be included in the metamodels training data set.
- B. This study recommends evaluating more than two designs spaces in terms of uniform space-filling.
- C. Additional research is required if it is desired to implement a monitoring system that can identify whether drying is occurring or not in the actual RMPP.
- D. It is recommended to derive the dew point from physics or utilize indirect measurement systems in the RMPP.
- E. It is advised to examine if the current position is the optimal position for monitoring O₂ concentrations to adequately monitor conditions.
- F. It is proposed to investigate if the Pareto plot in combination with SA is more beneficial for higher order inputs in distinction the "vital few" in comparison to other distinction methods.

Bibliography

- [1] Robert Wilde, “A Beginner’s Guide to the Industrial Revolution,” 2018. [Online]. Available: <https://www.thoughtco.com/guide-to-the-industrial-revolution-1221914>. [Accessed: 06-Jun-2019].
- [2] M. Roser, “Human Development Index (HDI),” *Our World in Data*, 2019. [Online]. Available: <https://ourworldindata.org/human-development-index>. [Accessed: 09-Jun-2019].
- [3] M. Roser, “Life Expectancy,” *Our World in Data*, 2019. [Online]. Available: <https://ourworldindata.org/life-expectancy>. [Accessed: 09-Jun-2019].
- [4] M. Roser and H. Ritchie, “Technological Progress,” *Our World in Data*, 2019. [Online]. Available: <https://ourworldindata.org/technological-progress>. [Accessed: 09-Jun-2019].
- [5] M. Roser, “Working Hours,” *Our World in Data*, 2019. [Online]. Available: <https://ourworldindata.org/working-hours>. [Accessed: 09-Jun-2019].
- [6] W. Wahlster *et al.*, “Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0 Report, the Industrie 4.0 Working Group of Acatech,” no. April, 2013.
- [7] N. W. Gleason, *Higher Education in the Era of the Fourth Industrial Revolution*. Singapore: Palgrave Macmillan, 2018.
- [8] W. Tiddens, *Setting sail towards predictive maintenance: developing tools to conquer difficulties in the implementation of maintenance analytics*. Enschede: University of Twente, 2018.
- [9] P. Brown and M. Sondalini, “Asset Maintenance Management - The Path toward Defect Elimination - THE EVOLUTION OF MAINTENANCE PRACTICES,” *Lifetime Reliab. Solut.*, p. 3.
- [10] F. Salvado and P. Couto, “Asset management evolution and main developments,” no. June, pp. 1301–1302, 2017.
- [11] S. Takata *et al.*, “Maintenance: Changing role in life cycle management,” *CIRP Ann. - Manuf. Technol.*, vol. 53, no. 2, pp. 643–655, 2004.
- [12] T. Tinga, *Principles of Loads and Failure Mechanisms*. London Heidelberg New York Dordrecht: Springer, 2013.
- [13] S. Yao, Xiaodong Yao; Fernández-Gaucherand, Emmanuel; C. Fu, Michael; I. Marcus, “Optimal Preventive Maintenance Scheduling in Semiconductor Manufacturing,” *Underst. Fabless IC Technol.*, vol. 17, no. 3, pp. 345–356, 2004.
- [14] B. Bagheri, S. Yang, H. A. Kao, and J. Lee, “Cyber-physical Systems Architecture for Self-Aware Machines in Industry 4.0 Environment,” *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 1622–1627, 2015.
- [15] S. M. Lee, D. Lee, and Y. S. Kim, “The quality management ecosystem for predictive maintenance in the Industry 4.0 era,” *Int. J. Qual. Innov.*, vol. 5, no. 1, p. 4, 2019.
- [16] “Research Projects | Smart Sensing and Predictive Maintenance in Steel Manufacturing (SUPREME) | Department MS3.” [Online]. Available: <https://www.utwente.nl/en/et/ms3/research-chairs/dbm/research/currentprojects/SUPREME/#partners-funding>. [Accessed: 02-Oct-2018].
- [17] N. G. Leveson, *Engineering a Safer World - Systems Thinking Applied to Safety*, vol. 67, no. 6. Cambridge, Massachusetts, 2012.
- [18] “Tata Steel Health and Safety Policies | Sustainability.” [Online]. Available: <https://www.tatasteel.com/sustainability/sustainability-pillars/health-and-safety/>. [Accessed: 11-Sep-2019].
- [19] “Accident | Definition of Accident by Merriam-Webster.” [Online]. Available: <https://www.merriam-webster.com/dictionary/accident>. [Accessed: 26-Aug-2019].

- [20] K. Marais, N. Dulac, and N. Leveson, "Beyond normal accidents and high reliability organizations: The need for an," *Symp. A Q. J. Mod. Foreign Lit.*, pp. 1–16, 2007.
- [21] H. C. Meng and K. C. Ludema, "Wear models and predictive: their form and content," *Wear*, vol. 181–183, pp. 443–457, 1995.
- [22] H. Uuemõis, H. Kangur, and I. Veerus, "Wear in high-speed impact mills," *Int. J. Miner. Process.*, vol. 44–45, no. SPEC. ISS., pp. 301–313, 1996.
- [23] A. Saltelli *et al.*, "Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices," *Environ. Model. Softw. - Elsevier*, vol. 114, no. March 2018, pp. 29–39, 2019.
- [24] A. Meghoe, R. Loendersloot, and T. Tinga, "Rail wear and remaining life prediction using meta-models," *Int. J. Rail Transp.*, pp. 1–26, 2019.
- [25] A. I. J. Forrester, A. Sóbester, and A. J. Keane, *Engineering Design via Surrogate Modelling - A practical Guide*. Chichester: John Wiley and Sons, 2008.
- [26] Y. A. Cengel, *Thermodynamics: An Engineering Approach INTRODUCTION AND BASIC CONCEPTS*, 8th ed., no. 8th Edition. New York: McGraw-Hill Education, 2015.
- [27] M. Ogink, T. Moons, and M. Markesteijn, "Kolen maal en droog installatie - Tata Steel," IJmuiden, 2019.
- [28] N. Meisner, *Coal preparation MAIN v2 - Experimental results*. Hlsarna Tata Steel IJmuiden, 2017.
- [29] H. Torreman, *Inleiding procestechnische berekeningen*, 2nd ed. Overberg: Delta Pres B.V., 1986.
- [30] J. P. Rafferty, "Vaporization | phase change | Britannica.com," *Encyclopaedia Britannica*. [Online]. Available: <https://www.britannica.com/science/vaporization>. [Accessed: 29-Jun-2019].
- [31] E. Tsotsas, V. Gnielinski, and E.-U. Schi, *Drying of Solid Materials*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 2005.
- [32] A. Buchanan, "Basic Knowledge," *Philos. Phenomenol. Res.*, vol. 37, no. 1, p. 101, 1976.
- [33] C. W. ; M. A. S. Hall, *Handbook of Industrial Drying*, 3rd ed. Singapore: Taylor & Francis Group, LLC., 2006.
- [34] A. S. Mujumdar and S. Devahastin, *Handbook of Industrial Drying- Chapter One: Fundamental Principles of Drying*. 2006.
- [35] C. Y. A. . G. A. J., *Heat and Mass transfer*, 4th ed. New York: Mc Graw Hill education, 2011.
- [36] D. L. Proctor, "Grain storage techniques Evolution and trends in developing countries," *Food and Agriculture Organization of the United Nations AGRICULTURAL SERVICES BULLETIN No. 109*, Rome, 1994.
- [37] B. Iooss and P. Lemaître, "A review on global sensitivity analysis methods," *Springer*, 2015.
- [38] D. C. . R. G. C. Mongtgomery, *Applied Statistics and Probability for Engineers*, 6th ed. Hoboken: John Wiley & Sons, 2014.
- [39] P. R. Choudhury and M. G. Endres, "Machine Learning Complementarities with Regression Analysis, Selection, Evaluation and Interpretation of Machine Learning Methods for Exploratory Data Analysis and Pattern Detection," *Harvard Bus. Sch.*, 2019.
- [40] B. Alexander, B. Mikhail, B. Evgeny, and Y. Yuri, "Smoothing of Surrogate Models," Moscow, 2011.
- [41] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Supervised Machine Learning: A Review of Classification Techniques," *Inform. - Artif. Intell. Rev.*, vol. 31, pp. 249–268, 2007.
- [42] E. Stinstra, "The meta-model approach for simulation-based design optimization.," Tilburg

University, 2006.

- [43] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” University of Wisconsin-Madison, 2018.
- [44] H. V. Razavi, Saman; Gupta, “What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in Earth and Environmental systems models,” *J. Am. Water Resour. Assoc.*, vol. 5, no. 3, pp. 2–2, 2015.
- [45] D. G. Sanchez, B. Lacarrière, M. Musy, and B. Bourges, “Application of sensitivity analysis in building energy simulations : Combining first- and second-order elementary effects methods,” *Energy Build.*, vol. 68, pp. 741–750, 2014.
- [46] A. Saltelli *et al.*, *Global Sensitivity Analysis*. Chichester: John Wiley and Sons, The Primer, 2008.
- [47] Y. Gan *et al.*, “A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model,” *Environ. Model. Softw.*, vol. 51, pp. 269–285, 2014.
- [48] F. Weber, S. Theers, D. Surmann, U. Ligges, and C. Weihs, “Sensitivity Analysis of Ordinary Differential Equation Models,” pp. 1–33, 2018.
- [49] A. T. Nguyen and S. Reiter, “A performance comparison of sensitivity analysis methods for building energy models,” *Build. Simul.*, vol. 8, no. 6, pp. 651–664, 2015.
- [50] M. D. Morris and N. May, “Factorial Sampling Plans for Preliminary Computational Experiments,” vol. 33, no. 2, pp. 161–174, 2007.
- [51] F. Pianosi *et al.*, “Sensitivity analysis of environmental models: A systematic review with practical workflow,” *Environ. Model. Softw.*, vol. 79, pp. 214–232, 2016.
- [52] F. Campolongo, J. Cariboni, and A. Saltelli, “An effective screening design for sensitivity analysis of large models,” *Environ. Model. Softw.*, vol. 22, no. 10, pp. 1509–1518, 2007.
- [53] H. B. Harvey and S. T. Sotardi, “The Pareto Principle,” *J. Am. Coll. Radiol.*, vol. 15, no. 6, p. 931, 2018.
- [54] M. V. Ruano, J. Ribes, J. Ferrer, and G. Sin, “Application of the Morris method for screening the influential parameters of fuzzy controllers applied to wastewater treatment plants,” *Water Sci. Technol.*, vol. 63, no. 10, pp. 2199–2206, 2011.
- [55] M. A. Stephens, “The Anderson-Darling Statistic,” Stanford University, 1979.
- [56] J. Stone, “Most ML models assume an underlying data distribution for them to function well. Where can I learn about the probability distributions that each ML model assumes? - Quora,” 2018. [Online]. Available: <https://www.quora.com/Most-ML-models-assume-an-underlying-data-distribution-for-them-to-function-well-Where-can-I-learn-about-the-probability-distributions-that-each-ML-model-assumes>. [Accessed: 22-Aug-2019].
- [57] F. Bahrpeyma, M. Roantree, and A. McCarren, “Multi-resolution forecast aggregation for time series in agri datasets,” *CEUR Workshop Proc.*, vol. 2086, pp. 193–205, 2017.
- [58] One minute based Rainfall, “IEM :: Features Tagged: heavyrain,” *IOWA University*. [Online]. Available: <https://mesonet.agron.iastate.edu/onsite/features/tags/heavyrain.html>. [Accessed: 13-Aug-2019].
- [59] “Choose Regression Model Options - MATLAB & Simulink - MathWorks Benelux.” [Online]. Available: <https://nl.mathworks.com/help/stats/choose-regression-model-options.html>. [Accessed: 14-Aug-2019].
- [60] M. G. Lawrence, “The Relationship between relative humidity and Dewpoint Temperature in Moist Air,” *Am. Meteorol. Soc.*, no. February, p. 9, 2005.
- [61] I. C. Kemp, “Fundamentals of Energy Analysis of Dryers,” *Mod. Dry. Technol.*, vol. 4, pp. 1–45, 2014.

- [62] W. Wagner *et al.*, “The IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam,” *J. Eng. Gas Turbines Power*, vol. 122, no. 1, p. 150, 2002.
- [63] “How It Works | Tata Steel in Europe.” [Online]. Available: <https://www.tatasteeleurope.com/en/innovation/hisarna/how-it-works>. [Accessed: 11-Sep-2019].
- [64] “Impact Dryer Mills | Crusher Dryer | Williams Crusher.” [Online]. Available: <https://www.williamscrusher.com/products/impact-dryer-mills#278636-standard-system>. [Accessed: 27-Aug-2019].
- [65] “Superheated Steam -Steam Engineering Principles and Heat Transfer.” [Online]. Available: <https://beta.spiraxsarco.com/learn-about-steam/steam-engineering-principles-and-heat-transfer/superheated-steam#article-top>. [Accessed: 09-Jul-2019].

Appendices

The appendices can be found in the appendix report.



Appendices report

**Application of a process model to define potential
operational functional failure conditions**

**By
Paul Schepers**

Emmen, 19-9-2019

Table of Contents

Appendix 1. Background and related work	1
Appendix 1.1 The steel industry and HIsarna.....	1
Appendix 1.2 Project charter	2
Appendix 2. Comparison of HEXS-model factors and RMPP sensors.....	5
Appendix 3. Analytical drying calculation for the mill of the RMPP.....	9
Appendix 3.1 Vaporization of surface moisture of the coal in the RMPP	9
Appendix 3.1.1 Mass and energy balance	10
Appendix 3.2 Conclusions	16
Appendix 4. Interview questions.....	17
Appendix 5. Meta-model.....	19
Appendix 5.1 Meta-model input for Regression Learner.....	19
Appendix 5.2 Meta-model input table for Sensitivity analysis.....	21
Appendix 5.3 Meta-model data resolution and retrieval mode	23
Appendix 5.4 Validation of HEXS-model and Meta-model	24
Appendix 6. HEXS-model input settings	27
Appendix 6.1 HEXS-model baseline settings	27
Appendix 6.2 HEXS-model input table.....	28
Appendix 7. HEXS-model Sensitivity Analysis	29
Appendix 7.1 Estimated HEXS sensitivity measures	29
Appendix 7.2 The HEXS screen plots.....	30
Appendix 7.2.1 The log-log HEXS-model screen plot	33
Appendix 8. Meta-model Sensitivity Analysis	35
Appendix 8.1 Selecting r trajectories	35
Appendix 8.2 The Meta screen plots.....	38
Appendix 8.2.1 Absolute mean screen plot (with zooms)	39
Appendix 8.2.2 Log log screen plots (with zooms).....	41
Appendix 9. Sensitivity comparison, HEXS-model vs Meta-model.....	45
Appendix 10. Potential functional failure conditions determined with HEXS-model.....	47
Appendix 10.1 Input HEXS-model.....	47
Appendix 10.2 OFAT potential functional failure by ΔDT	48
Appendix 10.3 Counteracting potential function failure by utilizing Sensitivity Analysis, response ΔDT	
50	
Bibliography	53

Keywords: Steady state thermodynamic process model, complexity of systems, predictive maintenance, sensitivity analysis, extended Morris method, supervised machine learning.

Appendix 1. Background and related work

Appendix 1.1 The steel industry and HIsarna

Steel is an alloy consisting of mostly iron, carbon, and other elements. Because of its high tensile strength and formability and low cost, it is widely used in various applications such as infrastructure, buildings, art, machinery, tools, and transportation.

When iron is smelted from its ore, it initially contains more carbon than is desirable. To become steel, it must be reprocessed to reduce the carbon to the correct amount [1]. Additionally, metal oxides such as hematite (Fe_2O_3) occur in nature mostly as compounds (minerals). To obtain the material properties required in certain applications, these compounds are chemically separated into elemental forms (Fe and O).

Currently, the iron making process industry-wide in blast furnace steel making uses carbon, in the forms of coke and coal, as a chemical reductant in the blast furnace to reduce iron oxide into iron. This converts the carbon into carbon monoxide (CO) and carbon dioxide (CO_2) gases [2], with which oxides in metals can be removed (e.g. by forcing them to form gasses and float out) [3]. Low carbon and impurity levels must be achieved to be able to obtain high quality steels with the desired material properties.

Carbon dioxide is a natural by-product during the steel-making process (CO_2 is involved in many combustion reactions). The iron and steel industry contributes considerably to global CO_2 emissions, and the trend is that they will continue to do so [4]. In 2015, the iron and steel industry accounted for 6.7% (2.4 gigatons) of global anthropogenic carbon dioxide emissions, and Tata Steel Netherlands was responsible for 7.5% of the Netherlands' national industrial CO_2 emissions in 2017 [4][5][6].

These increasing quantities of carbon dioxide are affecting Earth's atmosphere and the pH levels of the oceans. Carbon dioxide is the most important greenhouse gas in terms of current radiative forcing, and due to this, it possesses a high global warming potential [7][8][9]. Climate change is already affecting a wide variety of flora and fauna and their distributions [10]. An increase in global warming may lead to an additional loss of flora and fauna, creating an associated risk to human wellbeing [11][12]. To boost the carbon investment and to ensure the technology transfer to more sustainable technologies, carbon taxes and policies have been created [4][13]. One of these policies is the United Nations Paris Agreement of 2015. This policy aims to limit the increase of the global mean temperature to less than 2 °C with respect to pre-industrial levels, and accelerate and intensify the actions and investments needed to achieve a sustainable low carbon future [10][14].

HIsarna, as a low carbon emission iron-making process is intended contribute towards such a sustainable future an alternative to using the traditional blast furnace, the HIsarna plant was constructed in 2010 and currently operates as a research facility. It uses a different type of iron-making process. A blast furnace requires a porous burden from which raw materials such as iron ore and metallurgical coal are pre-processed into pellets, sinter, or coke to be able to make liquid iron. HIsarna avoids the need for coke and pellets due to the directly reduced iron process. The iron ore and coal can be directly injected as powders and are directly liquefied [15] due to the utilization of reactor temperatures above the iron melting point throughout the vessel. The coal powder reacts with the molten ore to produce liquid iron, which is the main material in steel manufacturing. Therefore, it eliminates the energy loss of sintering and coking and thus results in more efficient iron ore melting [15][16]. This technology can reduce CO_2 emissions by 20%, and in conjunction with carbon capture, it can reduce CO_2 emissions by up to 80% when compared to conventional blast furnaces. Additionally, it removes several pre-processing steps and requires less stringent conditions regarding the quality of the raw materials used in comparison to conventional iron-making facilities [17]. In summary, HIsarna has the following expected key advantages over the currently used blast furnace process [18][19]:

- it saves 20% of energy consumption and it eliminates the steps of sintering and coking;
- it reduces carbon emission by 20% to 80% when combined with biomass and scrap, and uses carbon capture and storage;
- it releases less CO , NO_x , SO_x , and fine dust;
- it allows for more steel scrap to be recycled – up to 50%; and

Appendix 1.2 Project charter

In the beginning of the research, a global project charter and project plan were generated. The project charter template, as displayed in Figure 3, is structured as a one-page overview of the project description, goal, results, benefits, risks, and scope to keep a concise overview. This charter is used to identify risks within the given scope. These risks could be minimized if they are recognized early-on and dealt with in the research approach.

Project Description

From a maintenance and process perspective, it is beneficial to identify the critical components and parameters that influence product quality, production uptime, and plant safety. This improves the availability, reliability, and cost effectiveness of technical systems such as a Raw Material Preparation Plant (RMPP) production facility. When the influence and interactions of parameters are known, a strategy can be developed to ensure that the final product meets specifications. However, as system components degrade and change operating conditions, the operational window available may no longer produce products that fulfil the specified criteria. The degradation of a single component can set off a chain reaction, negatively impacting other processes involved in producing the final product. This results in a spiral of products and processes that do not meet specifications. This reaction continues until the system is restored so it can fulfil its function, or the specifications are adapted to new acceptance levels.

To save on maintenance costs (not make replacements too early) and increase systems availability (not make replacements too late), the challenge is to achieve just-in-time maintenance by predicting component degradation. It is particularly difficult to choose critical components for predictive maintenance if the plant has a complex interactive nature and is not yet fully matured and continuously operating. Consequently, total knowledge of the operational and reliability aspects of the plant systems are not yet available. A potential solution is to utilize existing process models such as the HEXS model. Therefore, the following strategy is generated: to make use of an existing thermodynamic model of RMPP and investigate how it can be used for predictive maintenance. The strategy is that the influence and interactions of input parameters in relation to the outputs can be studied within the process model. Subsequently, the input factors can be arranged by comparing the relative influence on the response value, whereby the dominant parameters are identified. Consequently, it is allowed to make substantiated choices to allocate resources.

Goal

The goal is the evaluation of how an existing process model can be used for predictive maintenance,

Results

The results include an increased understanding of complex non-linear relations, a proposed strategy/method that assist to identify the influence of input parameter in relation to an output parameter, a case study that uses the method to determine influences of process parameters with respect to a defined response parameter, an overview of the influence of process parameters on the response of interest, an overview and description of the RMPP at the HIsarna plant, an overview of the RMPP operation conditions, and machine settings, an overview and description of the HEXS process model, a regression model that can run various operating scenarios, identification of the limitations of the HEXS process model, and modelled potential operational failure conditions of the RMPP.

Benefits

Benefits might include information is gathered about the operation settings as well as insight into the functioning and operating of the RMPP. This information can be used to narrow the scope and choose critical components for predictive maintenance, to and improve the operation procedures of the physical RMPP. Model developing time can be saved if the HEXS process model turns out to be applicable within the SUPREME research. Documentation can be made of how the HEXS model is performing calculations.

In this work, differences and similarities between the physical plant and the HEXS model are collected. Without the spoiling of energy, materials and time for physical testing, digital simulations run various of RMPP operating scenarios. This work studies their influence on a response parameter to gain an understanding of the relationships between various plant operation scenarios and parameters of interest. The method can be used as a tool to understand the behaviour of the model and/or the physical plant, which can then facilitate plant maturing.

Limitations

The process model does not incorporate all the details of complex natural phenomena, and therefore the results might not include the desired level of detail. For example, the model does not include the fracturing of coal due to grinding. This mechanism might be of significant importance when studying the degradation rate of a component and/or system. The model output cannot directly be compared to physical plant data because of unit differences, which are not directly convertible (% <--> °C). Limited resources are available, such as experience, workforce, measurements, computational power, knowledge, and documentation about the existing plant and the HEXS process model, which could lead to inaccurate information and incorrect conclusions. No literature has been found that utilizes (thermodynamic-related) steady state process models for predictive maintenance. The lack of understanding and working of the discovered HEXS model and the lack of validation can inhibit its adaption because the benefits are not clear. It is not known if or to what extent the HEXS model can be validated, and how and if it can be used in a sensitivity analysis. Time shortages also occurred, due to the fact that halfway through the initial research the research strategy was drastically adapted to what is presented in this work.

Scope

The RMPP real-world process and the existing process model were considered within this research. The information about the operation behaviour of the physical plant as physical relationships is used to be able to validate the process model. A sensitivity analysis, physics-based relations, and experienced base knowledge plays an important role in understanding the complex model. There are more than 300 process model parameters to consider, but this is reduced to one response value to determine the applicability of the method. The response parameter studied for a sensitivity analysis will be the mill outlet temperature, because this parameter is related to be a potential functional failure condition of the RMPP, with the benefit that it is monitored in the real-world plant. Due to this, there exists the possibility of validating the model by comparing the trend of the recorded plant data to that of the model. However, detailed information about the algorithms and optimisation of the features is not within the scope of this study.

Appendix 2. Comparison of HEXS-model factors and RMPP sensors

In this figure the sensors of the physical RMPP are compared with the factors (inputs/outputs) of the HEXS-model. Due to the continuous modifications of the plant, sensors are removed or changed in locations, whereby not all changes are updated in the documentation or software. Therefore, during the months February and March a comprehensive study has been performed to obtain a list of sensors installed and available at other resources. The RMPP and weather stations sensors are listed here by consulting the PI&Ds, personal, Human Machine Interface, Programmable Logic Controllers, Query and Trend (database), and performing physical plant inspections.

The following Figure 1 provides a legend that is used in the Figure 3. The legend is explained by two examples, as illustrated in Figure 1 and Figure 2. With 1 being a adaptable input factor, 2 being a response value (output for a given equation), 3 being a solver value in which a solver uses a group of factors to convergence to an output within specifications, 4 the error of the model, 5 an Alert that is given when the boundaries of the model has been reached or if undesired physical response (e.g. to high O₂ concentration), 6 in pink the monitored parameters in the HEXS-model that is recorded from the plant, 7 the actual sensor that monitors behaviour of the plant with its tag name. The values given in the figures are values that can occur in the plant but are not values that have a significant meaning for an experiment in this work. Note that, for the ambient temperature, pressure and relative humidity the values are not measured by the plant, but are recorded at a local weather station, therefore these values are not marked with an orange label. The combination between the in the legend presented red input values (1) and the orange (7) parameters are utilized in creating the Meta-model if the factor is monitored by the plant or weather station.

Legend:	
1	Input value
2	Response value
3	Solvers value
4	Error value
5	Alert
6	HEXS value, but monitored in RMPP
7	Data plant and sensor name from RMPP

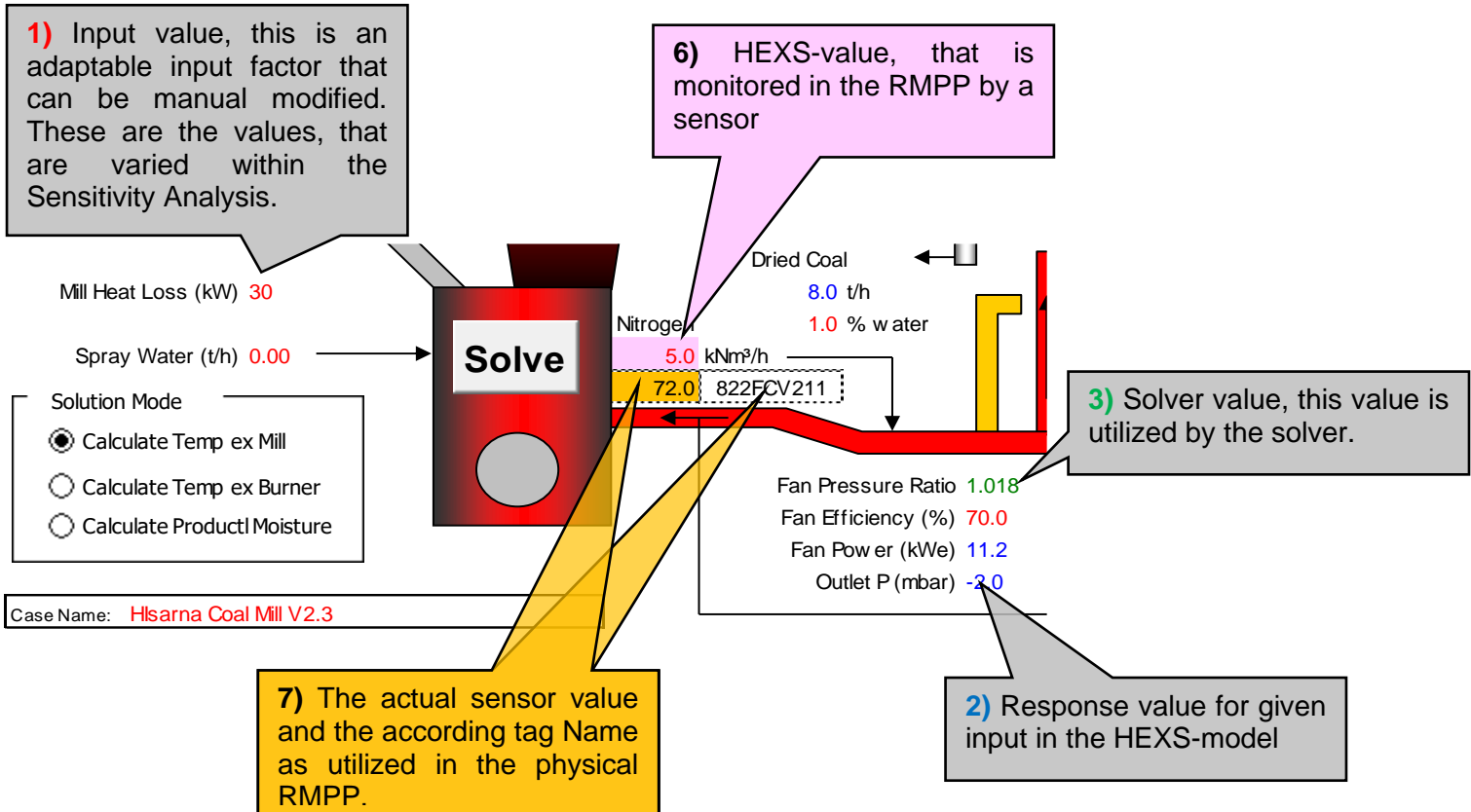


Figure 1 The legend used in Figure 3, this figure explains how the legend operates within Figure 3.

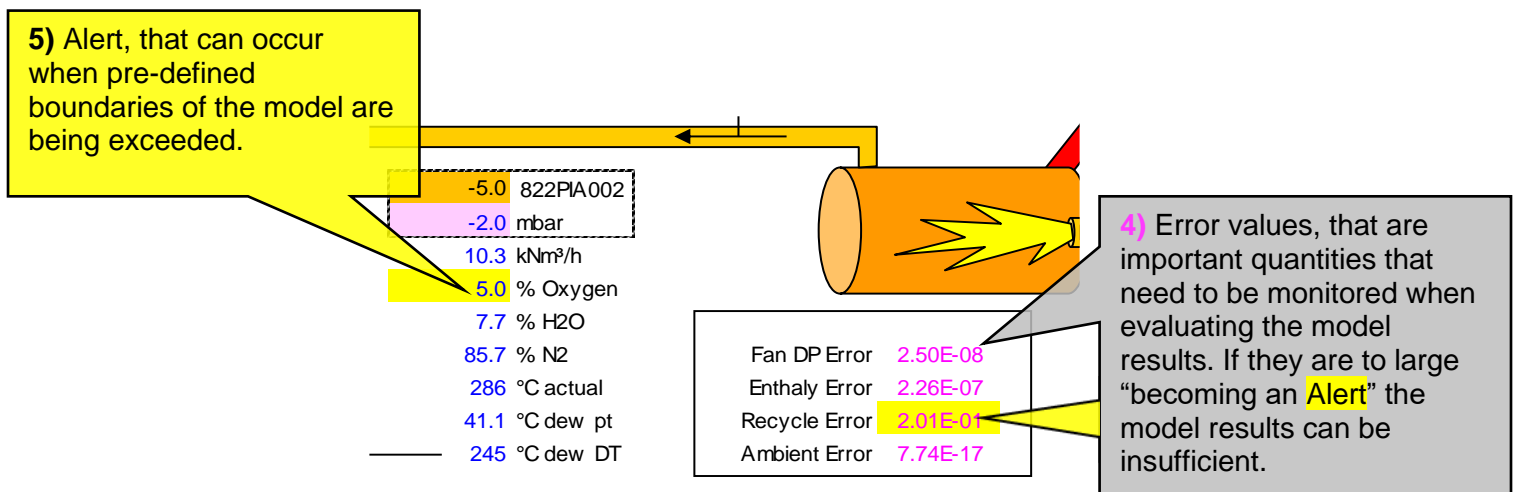


Figure 2 Continuing the explanation of the legend and how the legend operates within Figure 3

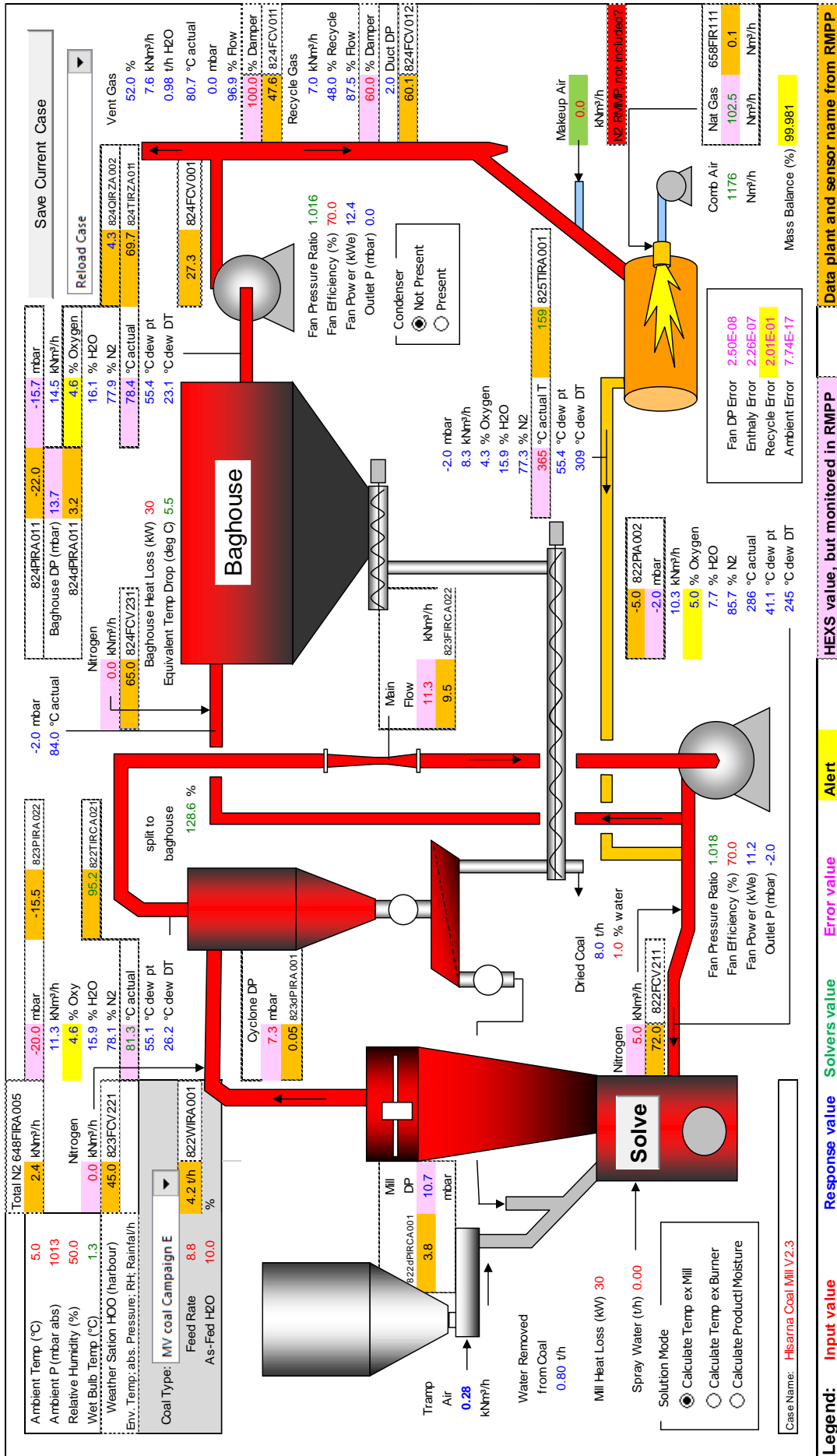


Figure 3 Schematic overview of the HEXS-model, in combination with an overlay of the physical sensors monitored at the Hsarna Raw Material Preparation Plant. Note, that the weather station (environmental condition) are not monitored at Hsarna, instead local Tata Steel weather stations are consulted.

Appendix 3. Analytical drying calculation for the mill of the RMPP

This chapter mimics the calculation of the mill systems of the HEXS model by an analytical calculation. With the aim to get insight in the model its workings.

Appendix 3.1 Vaporization of surface moisture of the coal in the RMPP

In summary, wet or moist raw coal is transported from the feed bin into the mill, where it falls on the grinding table, is crushed by hammer mills, and is mixed with the main drying gas circulation (the drying medium gas). The drying gas, blown from the bottom of the mill dries the moist coal mainly by forced convection and pneumatically conveys coal particles towards the top of the drying column and to other sections of the RMPP. During this pneumatic transport, the coal grains are heated and dried. The maximum accepted moisture content after drying is approximately 2% of the mass of coal. When the coal reaches this specified threshold, it is considered dried and is named a dried product. However, in practice the aim is set to achieve a moisture content below 1%, because this reduces handling problems such as coal clogging the pneumatic pipes transferring the coal to other systems.

This above-mentioned process is explained by the following example, which is derived from a part of the thermodynamics used in the HEXS model. By this, a section of the undocumented HEXS model is reconstructed and documented and in parallel the drying process is investigated and explained. The emphasis is placed on the mill and dryer column where wet coal is mixed with the drying gas, and simultaneously, the moisture content of the coal is reduced by vaporization. The objective of this example is to demonstrate drying a wet product containing 10% water mass fraction to a dry product with 2% water mass fraction.

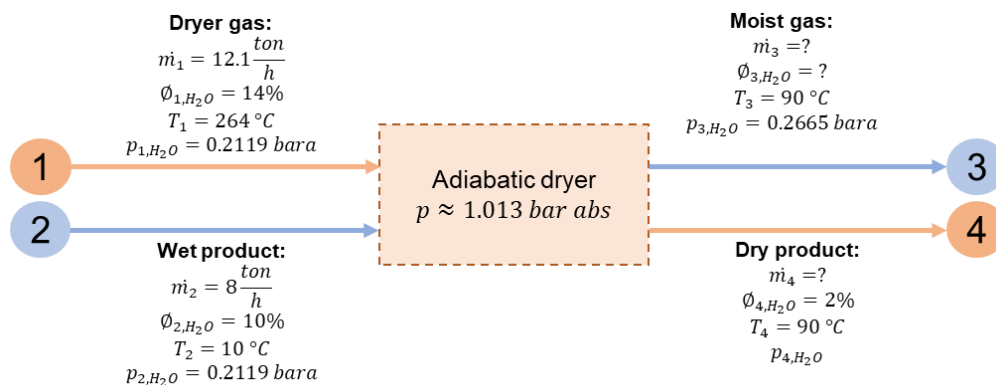


Figure 4 Schematic representation of the inlet streams of the mill and outlet stream of the drying column of the RMPP, here represent as o the adiabatic dryer control volume.

In Figure 4, there are two inlet streams at different states that are individual mixtures of elements. Stream 1 is the drying medium (vapour phase) and stream 2 is the wet coal (liquid and solid phase). The adiabatic dryer chamber is taken as the system in which the two streams are mixed. The control volume of the system is marked with dotted line since mass crosses the system boundary during the process. Depending upon the state of the individual streams, the mixing process can take place with or without vaporization of the moisture. The conditional values used in Figure 4 are derived from the actual RMPP operation for those that are being monitored, otherwise, they are estimated or calculated.

The following situations are investigated in this section.

1. Mass balance
 - o Is the ideal gas law applicable? (Mass and energy balance)
 - o Is the amount of dryer air required for the vaporization of surface moisture sufficient?
2. Energy balance
 - o Is the heated mixture (streams 3 and 4) above the dew point or the saturation line?
3. Is the analytic calculation in agreement with the results of the HEXS model?

The following assumptions are made for the mill and dryer column. These will be further discussed when deriving the mass and energy balances for the adiabatic dryer system.

- The process has steady uniform flow process, therefore there is no change with time at any point.
- The kinetic and potential energies are negligible, and thus are not considered.
- The drying medium is considered as an ideal gas, as it is at a high temperature and low pressure relative to its critical point values.
- The dryer is considered adiabatic, and thus there is no heat transfer from the dryer column to its surroundings.
- There are no work interactions involved.
- Non-reacting mixtures are treated as pure substances. Therefore, they can be used as a single compound rather than many individual constituents.

Appendix 3.1.1 Mass and energy balance

The relations used within the HEXS model are a combination of thermodynamic equilibrium relations and empirical relations. The basis is the mass and energy balances, which will be discussed in this section in respect to the assumptions made for modelling, as stated previously.

Mass balance

In physics, a control volume (CV) is often used to study flow through the CV region of open systems. The CV is a selected region in space and is stationary relative to a coordinate frame. The CV in Figure 4 is marked as a dotted orange line around the box. In a control volume, the conservation of mass principle can be express as the following: *the net mass transfer to or from a control volume during a time interval Δt is equal to the net change (increase or decrease) of the total mass within the control volume during Δt* [20]. That is:

$$\left[\begin{array}{c} \text{Total mass entering} \\ \text{the CV during } \Delta t \end{array} \right] - \left[\begin{array}{c} \text{Total mass leaving} \\ \text{the CV during } \Delta t \end{array} \right] = \left[\begin{array}{c} \text{Net change of mass} \\ \text{within the CV during } \Delta t \end{array} \right]$$

$$m_{in} - m_{out} = \Delta m_{CV} [kg] \quad (1)$$

In the CV, the mass balance m is the mass and \dot{m} the mass rate for sum of all species. When studying a steady-flow process, there is no interest in the amount of mass that flows into or out of a system over time. Instead, the amount of mass flowing per unit time is of significance, which is the mass flow rate \dot{m} . During a steady-flow process, the net quantity of mass or mass rate contained within a control volume is equal to the net quantity of mass or mass rate leaving the control volume and does not change with time, which turns the term $\Delta m_{CV} = 0$ in equation 1. This expression is presented as mass rate in equation 3 and is the sum of mass rates for each incoming and outgoing stream. The mass balance for a steady-flow system with multiple inlets and outlets is expressed in rate form as in equation 4. Considering a substance mixture composed of k components, the mass rate of the mixture \dot{m}_i is the sum of the mass rates of the individual components, called specie. Subscript i is the stream number, in this example, streams 1 and 2 of Figure 4 are denoted as “in” and refer to the inlet port and index “out” to the outlet port, streams 3 and 4. Addendum j runs over all “in” or “out” coming streams, j is used for specific species. Equation 5 can be used to determine the mass rate for a specific species denoted with subscript j , for example, $j = H_2O$ and ϕ is the percental quantity of that species within the total mass of a stream.

$$\sum_{in} \dot{m}_{i,in} = \sum_{out} \dot{m}_{i,out} [kg/s] \quad (3)$$

$$\dot{m}_i = \sum_{j=1}^k \dot{m}_{i,j} [kg/s] \quad (4)$$

$$\dot{m}_{i,j} = \frac{\phi_{i,j} \dot{m}_i}{100} [kg/s] \quad (5)$$

Energy balance

An energy balance describes the difference between the energy input and the energy output. The energy balance for a steady state flow process of an open-system is the quantity of energy contained within the system that will never change with respect to time. Therefore, the rate of energy accumulation is zero and can be expressed as the following:

$$\begin{aligned} \left[\begin{array}{l} \text{rate of energy} \\ \text{accumulation} = 0 \end{array} \right] &= \left[\begin{array}{l} \text{rate of heat flow} \\ \text{in or out system} \end{array} \right] + \left[\begin{array}{l} \text{work done on} \\ \text{or by the system} \end{array} \right] + \left[\begin{array}{l} \text{rate of energy} \\ \text{of a stream in} \end{array} \right] - \left[\begin{array}{l} \text{rate of energy} \\ \text{of a stream out} \end{array} \right] \\ 0 &= \dot{Q} + \dot{W} + \sum_{in} \dot{E}_{i,in} - \sum_{out} \dot{E}_{i,out} \end{aligned} \quad (6)$$

For a steady flow process, the rates at which heat \dot{Q} and work \dot{W} cross the control surface (CV) remain constant. For a system with no significant heat exchange with its surroundings, $\dot{Q} = 0$, known as an adiabatic system. The absence of any heat transfer can be due to perfect thermal insulation or the fact that the system and surrounding are at the same temperature. In the example, it is assumed that the drying process is adiabatic and that no work is done by or on the system, which reduces the equation to the following:

$$\sum_{in} \dot{E}_{i,in} - \sum_{out} \dot{E}_{i,out} = \Delta \dot{E}_{system} \left[\frac{kJ}{s} = kW \right] \quad (7)$$

Energy can exist in numerous forms such as chemical, nuclear kinetic, potential electric, and magnetic. Their sum constitutes the total energy E of a system. During a steady-flow process, the energy content in the CV remains constant, and thus the change in total energy of the control volume is zero ($\Delta \dot{E}_{system} = 0$). For simple steady state compressible systems with ideal flowing gas, as assumed for the dryer column, the system equation 7 can be reduced to the sum of change in flow, internal, kinetic, and potential energies, and be expresses as[20]:

$$\dot{E}_i = \dot{E}_{flow} + \dot{E}_{internal} + \dot{E}_{kinetic} + \dot{E}_{potential} \quad (8)$$

$$\dot{E}_i = \dot{m}_i \left(p_i \hat{v}_i + \hat{u}_i + \frac{V_i^2}{2} + gz_i \right) = \dot{m}_i \left(\hat{h}_i + \frac{V_i^2}{2} + gz_i \right) \quad [kW] \quad (9)$$

$$p_i \hat{v}_i = n_i R_i T \quad (10)$$

$$\hat{h}_i = p_i \hat{v}_i + \hat{u}_i = n_i R_i T + \hat{u}_i = C_{p,i} \dot{m}_i \quad \left[\frac{kJ}{kg} \right] \quad (11)$$

$$\hat{H}_i = \dot{m}_i (p_i \hat{v}_i + \hat{u}_i) = P_i \hat{V}_i + \hat{U}_i \quad [kJ] \quad (12)$$

The rate of energy \dot{E}_i transported by flowing media consists of flow energy, internal energy, macroscopic kinetic energy, and potential energy. Macroscopic forms of energy are those a system possesses as a whole with respect to some outside reference frame. It is assumed that all parts of the system move with the same velocity for a kinetic relation. Where p_i is the partial pressure of the gas [kPa], \hat{v}_i is the partial volume of the gas [kNm³], \hat{u} [kJ/kg] is the specific internal energy, V is the velocity g[m/s²] gravitational acceleration, and z [m] is the elevation of the system relative to some external reference point. The ideal gas equation of state, as expressed in equation 10, involves the specific gas constant R_i and four variables: absolute pressure p_i , specific volume \hat{v}_i , the number of moles of the gas n_i , and temperature T . Note that the combination of $p\hat{v} + \hat{u}$ is defined as the specific enthalpy \hat{h} [kJ/kg] and \hat{H}_i as the enthalpy. C_p is the specific heat capacity. The specific heat capacity can be recognized by lowercase letters and has the units in [1/kg], and nonspecific heat capacity is written with capitals and is multiplied by the mass. In the HEXS model, kinetic and potential energy are not included. Therefore, the energy balance reduces to $\dot{E}_i = \dot{m}_i \hat{h}_i$.

The drying medium is air (gas state), which is a composition of nitrogen, oxygen, water, carbon dioxide, and small amounts of some other gases. In the drying column system, the outlet temperature of the mixture is a higher temperature (90°C ≈ 363.2 K) and lower pressure (approximately 1 bara ≈ 0.101MPa)

relative to the critical point values of the individual gases. The critical point is not reached if the process temperature is larger than critical temperature and the process pressure is below the critical pressure. As can be observed from Table 1, all critical pressures have larger value than the process pressure of 1 atm. Therefore, the elements are not in a critical state. In such cases, a non-reacting gas mixture and its components can be treated as ideal gases with negligible error, and thus the ideal gas law can be used [20]. Each gas component in the mixture behaves as if it exists alone at the mixture temperature T_{mix} and mixture volume V_{mix} . This is the reason why in the schematic Figure 4, the output is drawn as a separate line instead of being one line in which the two streams are mixed. Each stream can be studied independently, which is beneficial when investigating the vaporization process.

Table 1. Substance properties, properties reproduced from Chapter 13, pages 690 and 700 [20].

Substance	Molar mass, M [kg/kmol]	Gas constant $R = R_{mix}/M$ [kJ/kg · K]	Critical point properties		
			Temperature [K]; [°C]	Pressure [MPa]	Volume [m ³ /kmol]
Air	28.97	0.2870	132.5; -140.65	3.77	0.0883
Nitrogen N_2	28.013	0.2968	126.2; -146.95	3.39	0.0899
Oxygen O_2	31.999	0.2598	154.8; -118.35	5.08	0.0780
Carbon dioxide CO_2	44.01	0.1889	304.2; 31.05	7.39	0.0943
Water H_2O	18.015	0.4615	647.1; 373.95	22.06	0.0560

Air in the atmosphere normally contains an amount of water vapour (or moisture) and is referred to as atmospheric air. In the situation of this example in Figure 4, the drying medium (stream 1) contains 14% water vapour. However, the drying medium has the maximum amount of moisture it can hold according to its pressure and temperature. If the air contains a maximum amount of moisture, it is called saturated air. In general, a higher air temperature can absorb a higher quantity of moisture. By utilizing the mass balance equation 5, the total mass rate of water in the drying medium can be determined.

From equation 5, it can be computed that for the two inlet streams the total mass rate is **12.9 ton/h**. Stream 1 has a mass rate of water at $\dot{m}_{1,H_2O,v} = 1,694 \text{ kg/h}$ and for stream 2 at the inlet the rate is $\dot{m}_{2,H_2O,l} = 800 \text{ kg/h}$. The objective is a wet product with 10% moisture that will be dried to 2%, therefore $\dot{m}_{2,H_2O,v} = 640 \text{ kg/h}$ is vaporized from coal to the drying medium and $\dot{m}_{4,H_2O,l} = 160 \text{ kg/h}$ of liquid water will remain in the dried product. If not explicitly explained, the subscripts *s*, *l*, or *v* give the state of the matter: *s* = solid, *l* = liquid, and *v* = vapor (gas).

In order to carry the vaporized moisture, the drying medium should be able to carry the total quantity of vaporized moisture, which is $\Phi_{3,H_2O}\dot{m}_3 = 2334 \text{ kg/h}$ from equation 15. The maximum moisture carrying capacity (MCC) of the drying medium “heated air” is determined by the following procedure. First, one should determine the Relative Humidity (RH) by the following equation, (page 727, equation 14-9 [20]):

$$RH = \frac{P_v \hat{v}_i / R_v T}{P_g \hat{v}_i / R_v T} = \frac{p_{\text{vapor}, H_2O}}{p_{\text{saturated vapor}, H_2O}} 100 = \frac{101.325 (P_{\text{atm}})}{5004.6 (P_{\text{sat}@264C})} 100 = 2.02 \text{ [%]} \quad (13)$$

in which P_v is the water vapour pressure, P_g is the saturation pressure of water at that temperature, \hat{v}_i is the specific volume, R_v is the gas constant (different for each gas) and T is the temperature. The saturated vapour pressure can be found by using saturated water temperature tables (page 904 [20]). For a temperature of 264°C , a saturation pressure of 5004.6 kPa is found. In the control volume, atmospheric pressure (surrounding the coal) is constant, but as temperature is increases, a phase transition occurs from liquid water to the gaseous state of water vapour. With the RH known – approximate 2.02% , and the dry bulb temperature of air known – 264°C , the psychometric chart of humid air can be used to determine the moisture carrying capacity [21]. The maximum moisture carrying capacity for this example is $\omega = 47 \text{ [kg water vapor/kg dry air]}$. This is obtained from Figure 5 by intersecting the indirect-heated convective dryer temperature of 264°C with the 2% of RH. The maximum amount of moisture that can be held by the drying medium in the system is expressed below in 14. At this point, the drying medium is saturated with moisture and is called saturated air. Any additional moisture in the saturated drying medium will condense.

$$\dot{m}_{\text{max}, \text{MCC}, \text{system}} = \Phi_{1, \text{dry air}} \dot{m}_1 * \omega = 0.75 * 12100 * 47 \approx 426\,525 \text{ [kg water vapor/h]} \quad (14)$$

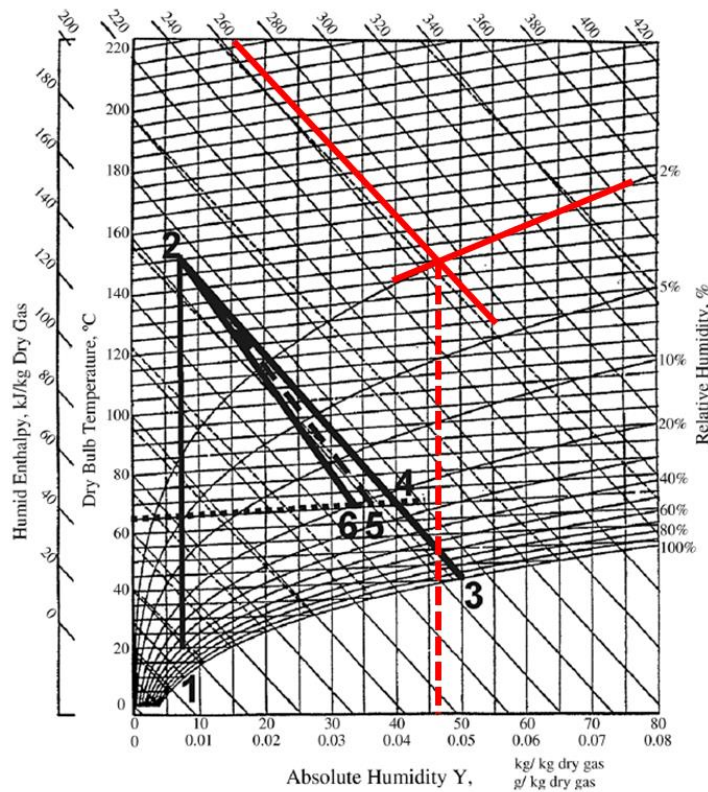


Figure 5. Mollier psychrometric chart for unsaturated air at normal temperature, 101.325 kPa . The drawn red operating lines for an indirect-heated convective dryer, 264°C is intersected with the 2% of RH to obtain $\omega = 47 \text{ [kg water vapor/kg dry air]}$. The black numbered lines 1 to 6 can be ignored, as these are hard copied in the image from the source and no unaffected chart have been found with the required temperature range and relative absolute humidity lines. [21]

The minimum required MCC required in the drying column is sum of the amount of water vapour

$$\phi_{1,H_2O}\dot{m}_1 + (\phi_{2,H_2O} - \phi_{4,H_2O})\dot{m}_2 = \dot{m}_{3,H_2O} = (0.14 * 121,00 + 0.08 * 8,000) = 2,334 \text{ [kg water vapor/h]} \quad (15)$$

It can be noted that the required MCC in the system is less than the maximum MCC ($2,334 < 427e3 \text{ [kg water vapor/h]}$); therefore, the drying medium supply is sufficient to carry the amount of water vapour if it is vaporized. When the moisture capacity of air becomes saturated by vapor, the Relative Humidity becomes 100% and the drying rates fall towards zero. However, there is an additional phenomenon when the drying rate comes to a hold.

The drying medium cools at a constant vapour pressure during the evaporation of water because the temperature of the heated air falls due to the heat required for vaporization of the water (page 731 [20]). Therefore, the vapour in the air undergoes a constant-pressure cooling process until it reaches the saturated vapour temperature, which is equal to the dew point temperature. At the dew point temperature, condensation begins. If the moist air is cooled below its dew point, a reverse process initiates and part of the water vapour in the air condenses and out the air stream as liquid. As a result, the temperature, humidity ratio and the water vapour pressure P_v of the drying air all decrease. This liquid can then be absorbed by the coal, which increases its moisture content. Therefore, to maintain a dry product, it is crucial that when the two inlet streams are mixed, the temperature of the outlet T_3 and the surface of T_4 remain above the dew point temperature determined from equation 16, or the saturation line. Due to the complexity of the model, the energy required for vaporization is not investigated, which in turn lead to temperature and pressure changes between the inlet and outlet. Alternatively, some of the physics will be discussed using the Mollier diagram as shown in Figure 6. This is based on the relationship between the content and water vapour content. The total heat content, or enthalpy, is plotted against the entropy.

$$T_{sat@p_{vapor,H_2O}} > T_{dp} \quad (16)$$

The dew point or saturation vapour temperature can be solved directly when the partial saturation pressure p is known. The equation to determine the saturation temperature is not presented in this work due to its extensive use of various tables. One could determine this temperature by using equation 55 from the “The Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam (IAPWS)”, which is the backward saturation-pressure equation for region 4 [22]. This and other formulations from IAPWS are used within the HEXS model.

To be able to determine the partial pressure for a species, the mole fraction of the species is required. At the RMPP facility, samples are taken from the various media for elemental analysis. In this analysis, the sample is analysed to determine the proportion of each element present. The compositions are used within the mass and energy balances in the HEXS model. Multiplying the mole fraction of the individual component x_i with the total pressure p_{total} , which is the pressure of the mixture results in the partial pressure p_i , is called Dalton’s law equation 17. From IAPWS equation 55, the saturation temperature is determined to be 64.4°C for a pressure $p_{1,H_2O} = 0.2119 \text{ bara}$. A benefit of this relationship is the requirement of a single input parameter, rather than the multiple inputs required by other expressions, to determine the dew point temperature.

$$p_i = x_i p_{total} \quad (17)$$

The mass flow rate of vapour in the drying medium increases by an amount equal to the rate of evaporation $(\phi_{2,H_2O} - \phi_{4,H_2O})\dot{m}_2$. This cools the drying medium. To determine if the drying medium carries sufficient heat to maintain vaporization to maintain drying, the streams presented in Figure 4 are further studied.

When stream 1 (vapour) and 2 (saturated liquid) of different states are mixed, the mixing process occurs with or without vaporization of the liquid. With condensation, the mixture is below the saturation line as defined by the green area in Figure 6. If the mixture remains above the saturation line, vaporization of the liquid to vapour occurs. State point h_1 is drawn with the inlet temperature measured at 264°C and a partial pressure for vapour as 0.21 bar in Figure 6. State point h_{mix} , which is the outlet and the combination of stream 1 and 2, is drawn at a pressure of 0.2665 bara and a temperature of 90°C . The result is that this mixture remains above the saturation line. Therefore, in this example, the process continues with the reduction of the surface moisture content of the coal until the saturation line is reached.

This example demonstrates the following influences on the drying capacity of the heated medium with respect to moisture removal from the wet product:

- the mass rate and the composition of the drying medium,
- the mass of water vapour already carried by the drying medium,
- the average drying medium temperature, and
- the mixture temperature in relation to the dew point temperature or the saturation line. (In the RMPP case study, this is equivalent to the mill outlet temperature).

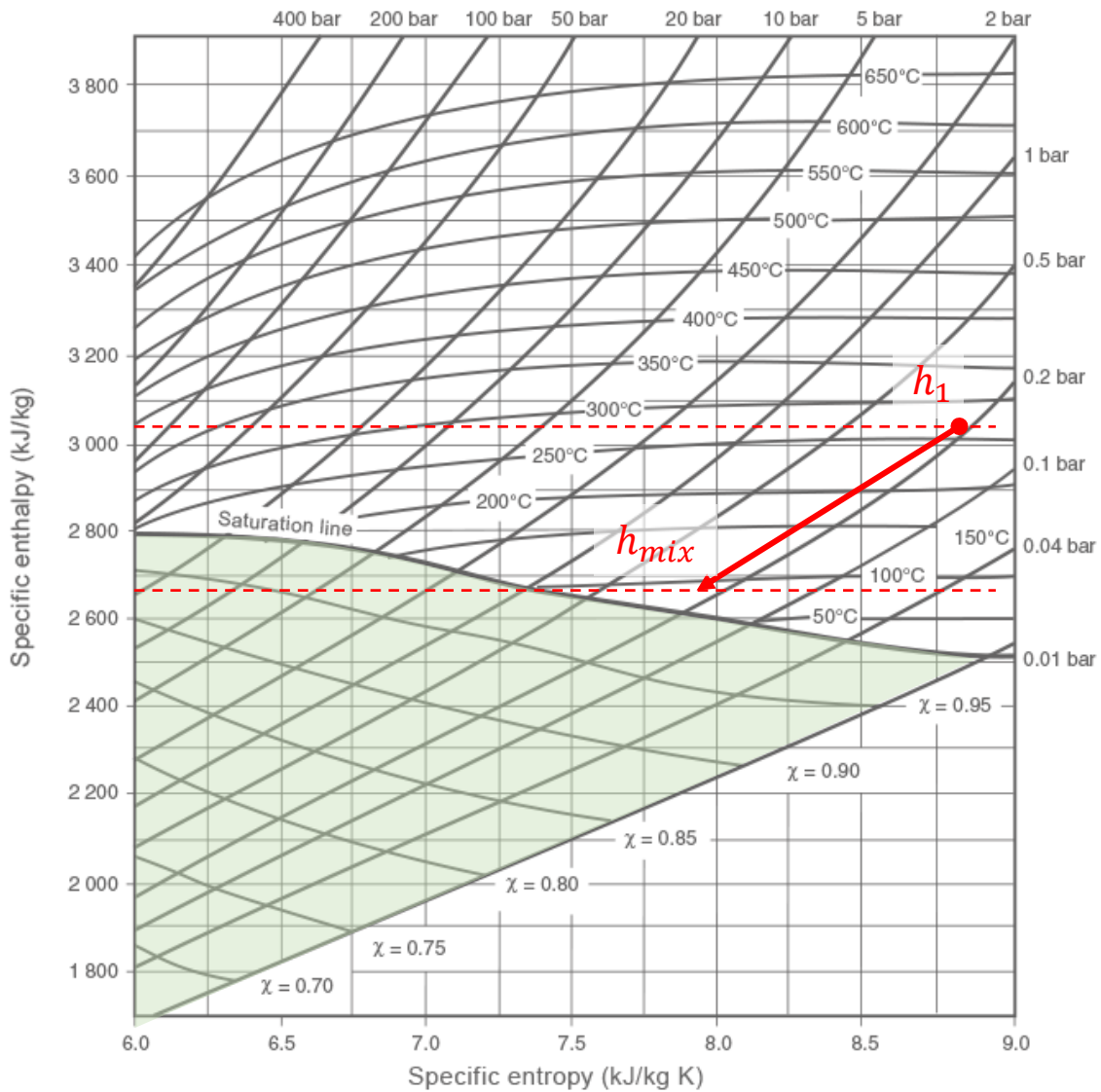


Figure 6. Enthalpy plotted against entropy, or the Mollier chart for steam. h_1 is the inlet of the example and h_{mix} the combination of the two inlet streams. A partial pressure of $p_{1,H_2O} = 0.2119 \text{ bar}$. The area in green is below the saturation line. If the h_{mix} is within this area, condensation will occur [23].

Appendix 3.2 Conclusions

The following situations were investigated in the upper sections:

1. Mass balance
 - Is the ideal gas law applicable? (Mass and energy balance)
 - Is the amount of dry air required for the vaporization of surface moisture sufficient?
 2. Energy balance
 - Is the heated mixture (streams 3 and 4) above the dew point or the saturation line?
 3. Is the analytic calculation in agreement with the results of the HEXS model?
-
1.
 - The ideal gas law is applicable and can be used in the way it is utilized in the HEXS-model.
 - For the given boundary conditions, the amount of dryer air for vaporization of the surface moisture with the current settings of the RMPP are sufficient.
 2.
 - The moisture carrying capacity required for the drying air is 2,334. That is less than calculated maximum MCC of $427e3$ [*kg water vapor/h*]. Therefore, the drying medium supply is sufficient to carry the amount of water vapour if it is vaporized. The ability to dry can be monitored in two ways: when the drying medium temperature is larger than the dew point temperature, and when the saturation vapour pressure (the drying medium) is less than the partial pressure of the moisture content surrounding the coal.
 3. The analytic calculation performed agrees with the results of the HEXS model.

Appendix 4. Interview questions

Once the experts are identified, they were individually interviewed in sessions lasting between 30 minutes to one hour. The experts include, scientists, researchers, process technologist, individuals with hands on experience with operation and working with the RMPP. During the interviews the following questions were asked at the level of subsystem level. Whereby each of the 22 changeable input factors of the HEXS-model are discussed. A more extensive interview is held with Dr. Rod Dry, the developer of the HEXS-model. This interview is recorded, and its recording and questions can be shared on request.

- Which of the factor are manually operated (adaptable in HMI) and which ones are stationary (controlled by the logic controls)?
- What is a factor its response on the mill outlet temperature?
- Which input factors do you think have the most influence of changing the mill outlet temperature and why?
 - Burner temperature
 - Feed rate of the coal
 - Moisture content of the coal
 - Particle size distribution of the coal
- Which parameters do limit the operation conditions with the RMPP?
 - O2 levels
 - CO2 levels
 - Burner capacity that heats the drying air
- Which factors do have the most dominant influence on the final moisture content of the coal?
 - Burner temperature
 - Incubation time
 - Particle size
- What are common operation levels for each of the factors, with which unit are the monitored?
- How is the trend data on the dashboard used?

Case	Independent variable	Controllable in HMI?
case 1	Ambient Temp (°C)	No
case 2	Ambient P (mbar abs)	No
case 3	Relative Humidity (%)	No
case 4	Feed Rate Coal (t/h)	Yes
case 5	Coal As-Fed H2O (%)	No
case 6	Mill Heat Loss (kW)	No
case 7	Spray Water (t/h)	Yes
case 8	Nitrogen In, Mill outlet (kNm ³ /h)	Yes, (multiple factors influencing this)
case 9	Pressure, Mill outlet (mbar)	Yes, (multiple factors influencing this)
case 10	Temperature, Mill outlet (°C)	Yes, (multiple factors influencing this)
case 11	Differential Pressure, Cyclone (mbar)	Yes, (multiple factors influencing this)
case 12	Main Fan efficiency (%)	No
case 13	Nitrogen In, Main Fan outlet (kNm ³ /h)	Yes, (multiple factors influencing this)
case 14	Water output/ product Moisture (%)	Yes
case 15	Nitrogen In, before Baghouse (kNm ³ /h)	Yes, (multiple factors influencing this)
case 16	Baghouse Heat Loss (kW)	No
case 17	Exhaust Fan efficiency (%)	No
case 18	Vent Gas, Damper, behind Exhaust Fan (%)	Maybe, control logic
case 19	Recycle Gas, Damper, behind Exhaust Fan (%)	Maybe, control logic
case 20	Makeup Air, before burner (kNm ³ /h)	Maybe, control logic
case 21	Burner temperature (°C)	Yes
case 22	Main Flow, cyclone to main fan (kNm ³ /h)	Yes

Appendix 5. Meta-model

Appendix 5.1 Meta-model input for Regression Learner

Input for the regression is data of March 2019, the data set is inspected if there is any missing data, if there is any missing data, they are filled with the average between the first neighbouring upper and lower bound of the same species. In total 39 points were missing, of the total 1140568 data points. These 39 points are filled with the average between the upper and lower bound, note that no two data points were directly after each other missing. The resolution of the sensors observations is chosen to be cyclic with an interval of 60 seconds. Other computational filtering or data cleaning is not performed. The Anderson-darling test has been performed to identify if one of the factors data timer-series is normal distributed. It is observed that not one of the factors has data that is normally distributed. As it is not normally distributed, none of observations fall within ± 1 standard deviation from the mean and thus they do not account for about 68% of all the observations of the same type. Nonetheless, in the table the minimum and maximum values with their standard deviations are reported. The tag name and the description can be utilized to find the specific asset within the Raw material Preparation Plant.

Symbol	Tag Name Plant	Description	Unit	Minimum value	Maximum value	Mean value	Stdev (mean)
M1*	Dry-bulb temperature	Ambient dry-bulb temperature at Hoogoven harbour Tata	°C	3.10	11.90	7.38	1.31
M2*	Meteo barometer	Barometer absolute pressure at Hoogoven harbour Tata	mBar	992.30	1035.90	1015.32	13.28
M3*	Relative Humidity	Relative humidity at Hoogoven harbour Tata	%	57.00	100.00	82.13	9.81
M4*	822WIRA001.PV	Total Weight V-822	kg	0.00	5384.20	3141.38	550.61
M5	Meteo precipitation	Precipitation pumping station 3 at Ijmuiden.	mm/hr	0.00	4.00	0.00	0.05
M6	825GI004.PV	Burner BU-825 Power	%	-1.90	84.90	49.74	26.17
M7*	823FCV221.PV	Nitrogen injection control valve cyclone inlet	%	0.00	20.00	0.20	1.85
M8*	822PdIRCA001.PV	Differential pressure mill drying column DY-822	mBar	-1.22	36.63	16.89	12.13
M9*	823PdIRA001.PV	Differential pressure measurement cyclone	mBar	-1.26	17.91	5.34	5.14
M10	648FIRA005.PV	Flow measurement coal system inertisation nitrogen supply	m3/h	0.87	8474.15	3385.38	1644.32
M11*	822FCV211.PV	Nitrogen injection control valve mill hot gas inlet	%	0.00	34.73	0.20	1.87
M12	648PIRZA005.PV	Pressure measurement 3 BarG nitrogen supply	BarG	0.00	3.51	2.16	0.86
M13*	824FCV231.PV	Nitrogen injection control valve bag filter inlet	%	0.00	25.00	0.25	2.26
M14	648TT001.PV	5D24 area 18bar nitrogen supply temperature	°C	3.42	22.40	12.51	1.86
M15	822LIRCA001.PV	Level measurement Coal supply bin V-822	%	-0.03	73.13	17.36	8.30
M16*	824FCV011.PV	Exhaust stack damper	%	10.00	100.00	46.51	17.41
M17*	824FCV012.PV	Recycle Gas damper	%	0.00	80.00	41.03	22.83
M18	822M001.Current_PV	Mill motor M-822	A	-0.04	173.03	53.18	29.84
M19*	825TIRA001.PV	Temperature measurement Burner hot gas outlet	°C	9.11	447.79	251.15	147.31
M20*	823FCV022.PV	System flow control valve inlet main fan	%	0.00	25.00	11.61	8.53
M21	822M001.Power_PV	Mill motor M-822	kW	-0.07	290.36	29.28	17.32
M22	822M021.Current_PV	Spinner separator S-822	A	0.00	89.99	23.64	12.24
M23	822M021.Speed_PV	Spinner separator S-822	Hz	0.00	30.65	15.32	8.53
M24	822M021.Torque_PV	Spinner separator S-822	%	-0.01	4.86	1.98	1.08
M25	822PIRA002.PV	Under pressure measurement mill inlet	mBar	-15.77	28.80	-4.11	3.51
M26	823FIRCA022.PV	Differential pressure flow measurement cyclone outlet	Nm3/h	-0.47	27.62	12.58	7.54
M27	823M001.Power_PV	Main fan motor K-823	kW	-0.04	227.47	126.39	70.39
M28	823PIRA022.PV	Under pressure measurement system flow measurement	mBar	-66.03	28.56	-34.33	23.96
M29	823TIRA022.PV	System flow temperature measurement	°C	6.97	95.63	65.55	23.22
M30	824FCV001.PV	Exhaust fan inlet damper	%	8.75	50.00	23.64	9.43
M31	824FCV244.PV	N2 Injection burner Valve	%	-0.06	103.49	66.23	27.09
M32	824M001.Power_PV	Exhaust fan motor K-824	kW	0.00	64.31	12.60	7.11
M33	824MIA001.PV	Dust content measurement after bag filter	mg/m3	4.95	100.03	47.30	35.61
M34	824PdIRA011.PV	Pressure difference measurement baghouse	mBar	-0.37	22.39	8.32	5.04
M35	824PIRA011.PV	Under pressure measurement Exhaust fan Inlet"	mBar	-40.86	27.46	-16.42	10.47
M36	824PIRA012.PV	Under pressure measurement recirculation	mBar	-2.53	28.45	5.51	3.46
M37	824PIRA201.PV	Pressure measurement cleaning air supply.	BarG	0.00	11.23	6.88	2.38
M38	824QIRCA001.CO	CO measurement 1	%	0.00	22.10	8.30	4.49
M39	824QIRCA001.O2	CO measurement 1	%	0.00	22.10	8.30	4.49
M40	824QIRZA002.O2	Measured O2 value	%	0.00	21.54	8.04	4.34
M41	824TIRA001.PV	Temperature measurement bag filter inlet	°C	8.08	107.82	78.60	27.14
M42	824TIRA011.PV	Temperature measurement bag filter outlet	°C	8.59	92.56	69.90	21.21
M43	658FIR111.PV	Flow measurement natural gas supply Gas heater BU-825	m3/h	0.03	138.89	68.94	50.51
M44	658PIRA004.PV	Pressure measurement natural gas supply.	mBarG	-2.97	343.53	267.87	66.26
Response	822TIRCA021.PV	Temperature measurement outlet spinner S-822	°C	8.13	124.88	73.74	22.82

Appendix 5.2 Meta-model input table for Sensitivity analysis

The symbol names denoted as M with *, is a factor that is also in HEXS-model as an adaptable input factor. Other, meta-factors are factors utilized that cannot be controlled without modification in the HEXS-model but can be monitored. Therefore, also their influence can be investigated.

These, minimum and maximum are later transferred to an 80% range to be used in the SA analysis, in order to eliminate some of the disturbance, as most of the disturbance occurs at the under and upper bound, these are removed by removing the lower 10% and the upper 10% of the data.

Table 2. 80% of operating range from cyclic retrieved 60s data HIsarna RMPP, March 2019.

Symbol	Tag Name RMPP	Factors	Unit	Minimum	Maximum
M1*	Dry-bulb temperature	Ambient dry-bulb temperature at Hoogoven harbour Tata	°C	3.98	11.02
M2*	Meteo barometer	Barometer absolute pressure at Hoogoven harbour Tata	mBar	996.66	1031.54
M3*	Relative Humidity	Relative humidity at Hoogoven harbour Tata	%	61.30	95.70
M4*	822WIRA001.PV	Total Weight V-822	kg	538.42	4845.78
M5	Meteo precipitation	Precipitation pumping station 3 at Ijmuiden.	mm/hr	0.00	3.60
M6	825GI004.PV	Burner BU-825 Power	%	6.78	76.22
M7*	823FCV221.PV	Nitrogen injection control valve cyclone inlet	%	0.00	18.00
M8*	822PdIRCA001.PV	Differential pressure mill drying column DY-822	mBar	2.57	32.85
M9*	823PdIRA001.PV	Differential pressure measurement cyclone	mBar	0.66	15.99
M10	648FIRA005.PV	Flow measurement coal system inertisation nitrogen supply	m3/h	848.20	7626.82
M11*	822FCV211.PV	Nitrogen injection control valve mill hot gas inlet	%	0.00	31.25
M12	648PIRZA005.PV	Pressure measurement 3 BarG nitrogen supply	BarG	0.00	3.15
M13*	824FCV231.PV	Nitrogen injection control valve bag filter inlet	%	0.00	22.50
M14	648TT001.PV	5D24 area 18bar nitrogen supply temperature	°C	5.32	20.50
M15	822LIRCA001.PV	Level measurement Coal supply bin V-822	%	7.29	65.82
M16*	824FCV011.PV	Exhaust stack damper	%	19.00	91.00
M17*	824FCV012.PV	Recycle Gas damper	%	8.00	72.00
M18	822M001.Current_PV	Mill motor M-822	A	17.27	155.73
M19*	825TIRA001.PV	Temperature measurement Burner hot gas outlet	°C	52.98	403.92
M20*	823FCV022.PV	System flow control valve inlet main fan	%	2.50	22.50
M21	822M001.Power_PV	Mill motor M-822	kW	0.00	261.32
M22	822M021.Current_PV	Spinner separator S-822	A	0.00	80.99
M23	822M021.Speed_PV	Spinner separator S-822	Hz	0.00	27.58
M24	822M021.Torque_PV	Spinner separator S-822	%	0.00	4.38
M25	822PIRA002.PV	Under pressure measurement mill inlet	mBar	-11.32	24.34
M26	823FIRCA022.PV	Differential pressure flow measurement cyclone outlet	Nm3/h	2.34	24.81
M27	823M001.Power_PV	Main fan motor K-823	kW	0.00	204.72
M28	823PIRA022.PV	Under pressure measurement system flow measurement	mBar	-56.57	19.10
M29	823TIRA022.PV	System flow temperature measurement	°C	15.83	86.76
M30	824FCV001.PV	Exhaust fan inlet damper	%	12.88	45.88
M31	824FCV244.PV	N2 Injection burner Valve	%	0.00	93.13
M32	824M001.Power_PV	Exhaust fan motor K-824	kW	0.00	57.88
M33	824MIA001.PV	Dust content measurement after bag filter	mg/m3	14.46	90.52
M34	824PdIRA011.PV	Pressure difference measurement baghouse	mBar	1.90	20.11
M35	824PIRA011.PV	Under pressure measurement Exhaust fan Inlet"	mBar	-34.02	20.63
M36	824PIRA012.PV	Under pressure measurement recirculation	mBar	0.57	25.36
M37	824PIRA201.PV	Pressure measurement cleaning air supply.	BarG	1.13	10.10
M38	824QIRCA001.CO	CO measurement 1	%	2.21	19.89
M39	824QIRCA001.O2	O2 measurement 1	%	2.21	19.89
M40	824QIRZA002.O2	Measured O2 value	%	2.15	10.00
M41	824TIRA001.PV	Temperature measurement bag filter inlet	°C	18.05	97.85
M42	824TIRA011.PV	Temperature measurement bag filter outlet	°C	16.99	84.16
M43	658FIR111.PV	Flow measurement natural gas supply Gas heater	m3/h	13.91	125.00
M44	658PIRA004.PV	Pressure measurement natural gas supply.	mBarG	31.68	308.88

Appendix 5.3 Meta-model data resolution and retrieval mode

Presented a range of HIsarna Raw Material Preparation Plant data Figure 7, followed by a deep dive in Figure 8 of a particular part of that Figure 7. It can be observed that data retrieval cyclic with an interval of each 60 seconds (red dashed line) follows the 1 second cyclic “the truth” (black solid line) more closely in comparison to the average 60s retrieval mode (green dashed line). This particular part “red boxed section” in Figure 7 is selected as the rate of change in the mill outlet temperature is quite rapid, whereby a coarser resolution (60 seconds) might not be able to follow a finer resolution (1 second).

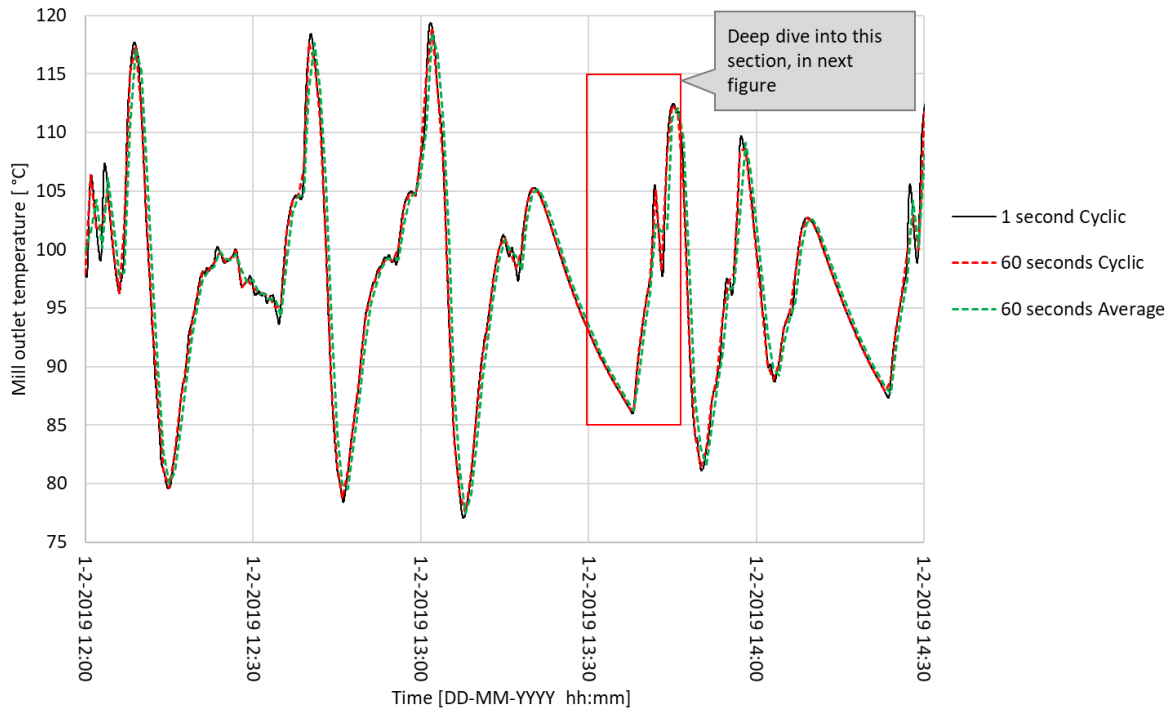


Figure 7 Mill outlet dry-bulb temperature for different retrieval modus, run in February 2019.

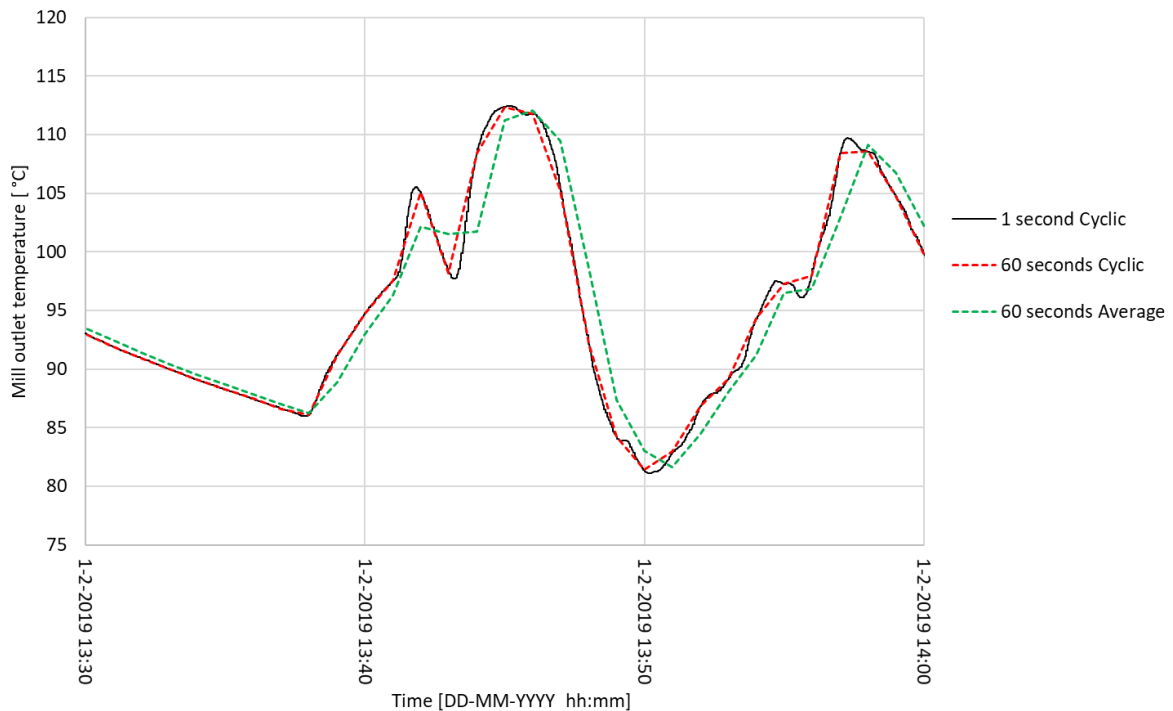


Figure 8 Deep dive (closer look), mill outlet dry-bulb temperature for different retrieval modus, run in February 2019.

Appendix 5.4 Validation of HEXS-model and Meta-model

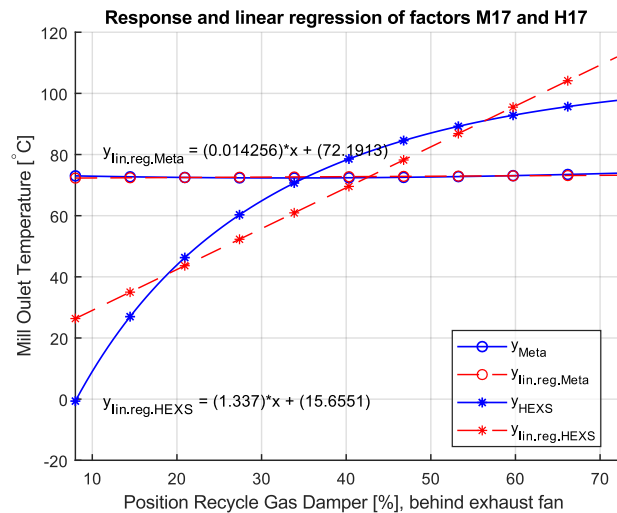
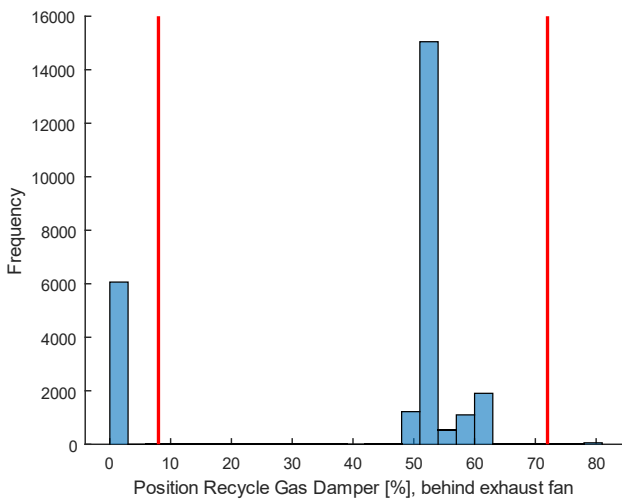
Histograms, the red lines in the histogram are the minimum and maximum bound as stated in Table 3. The histogram is presenting the full monitored operating space of the run of March 2019 for the HIsarna RMPP. From the full space a Meta-model has been created.

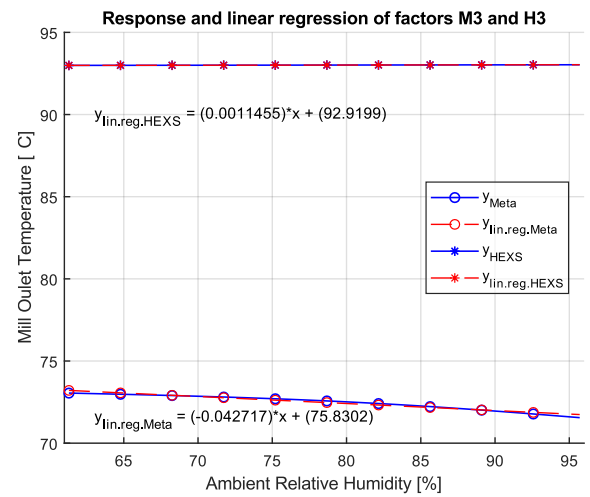
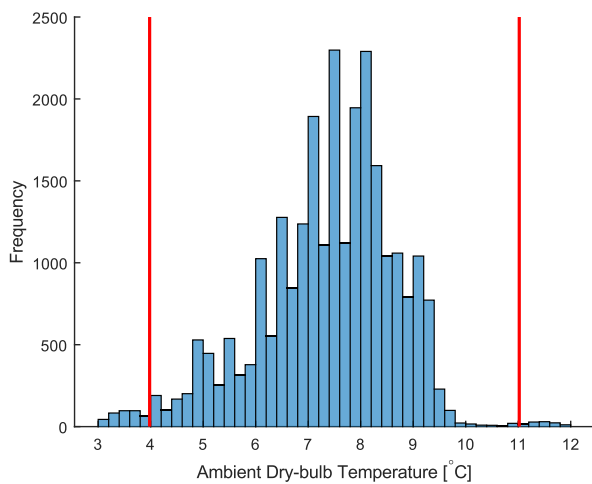
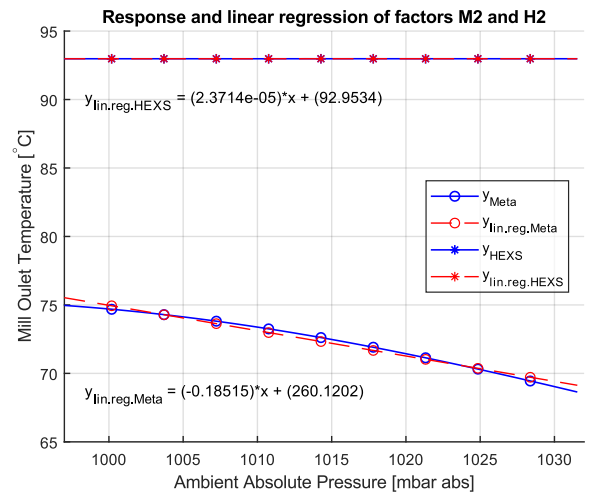
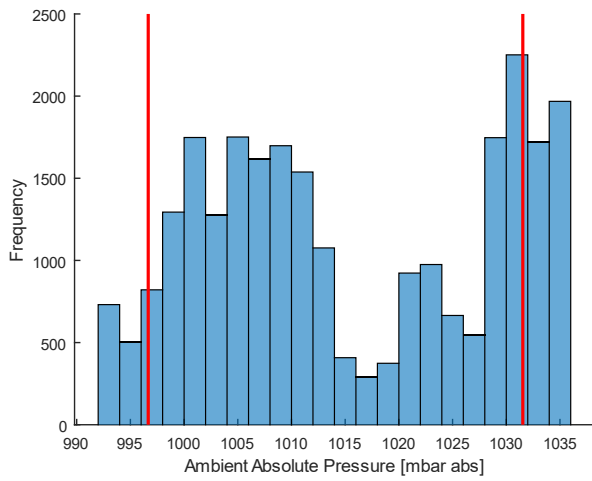
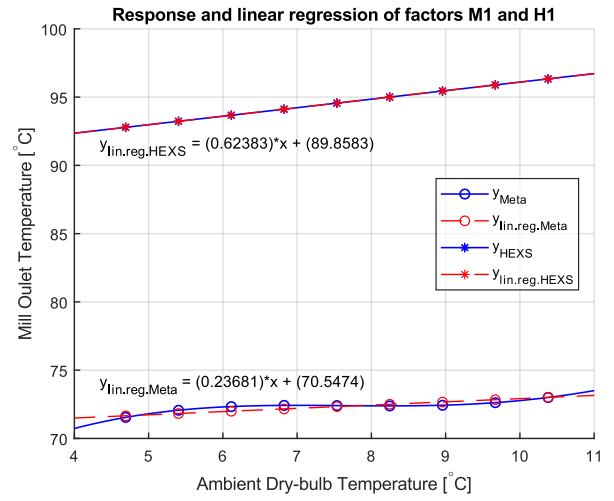
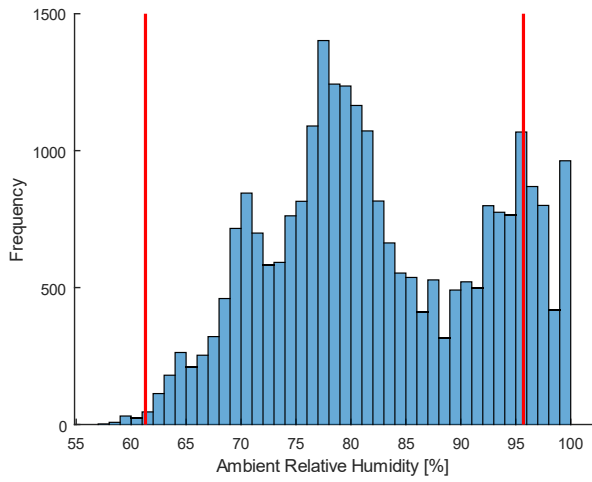
For each of these factors a graphical representation is calculated by the blue lines for the Meta and HEXS model as reported in the Figures, the graph starts at the minimum as reported in Table 3, where it is increased by 100 equal steps till it reaches the reported maximum. The redline is the linear regression fit that is created for each of the models, the mathematical expression for these linear relations are presented within the plots.

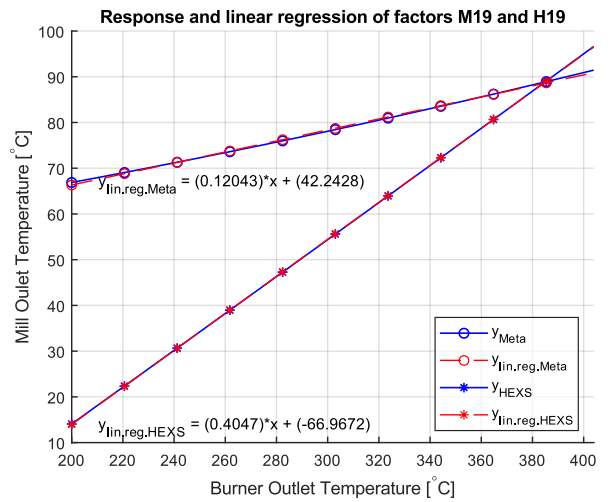
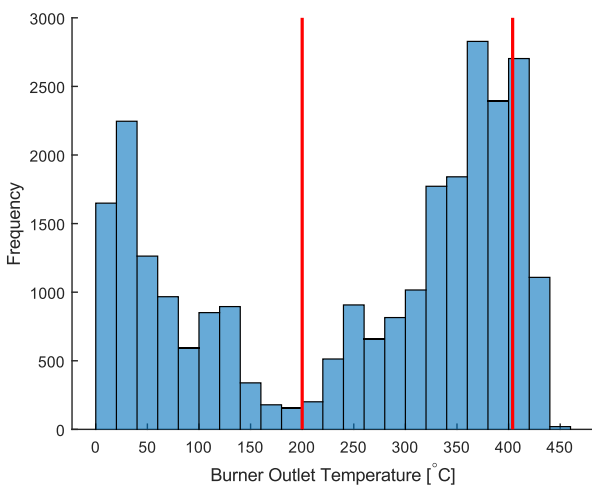
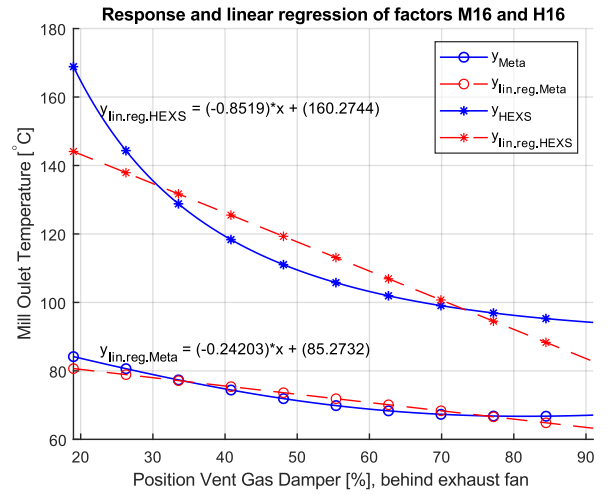
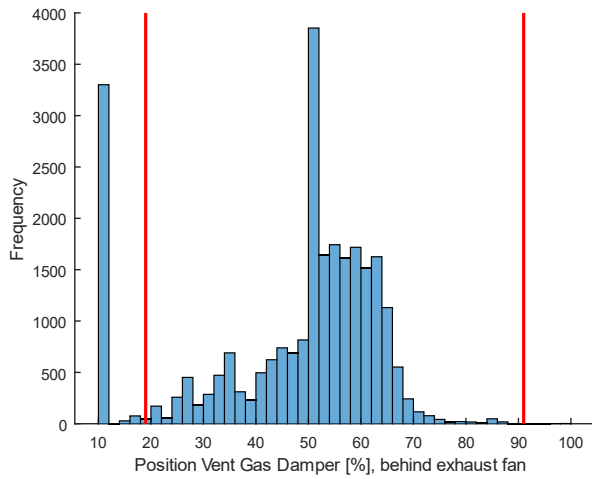
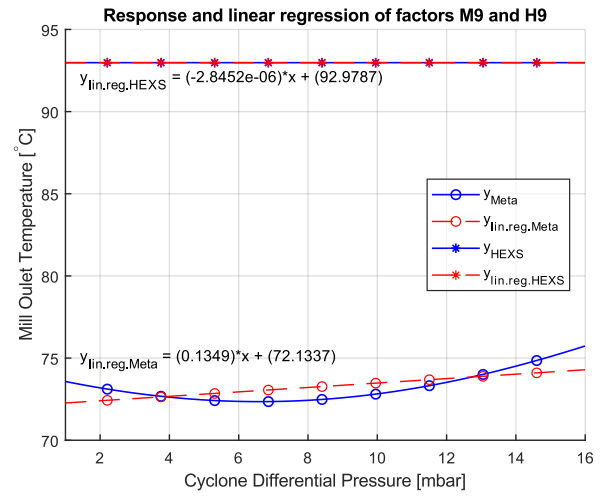
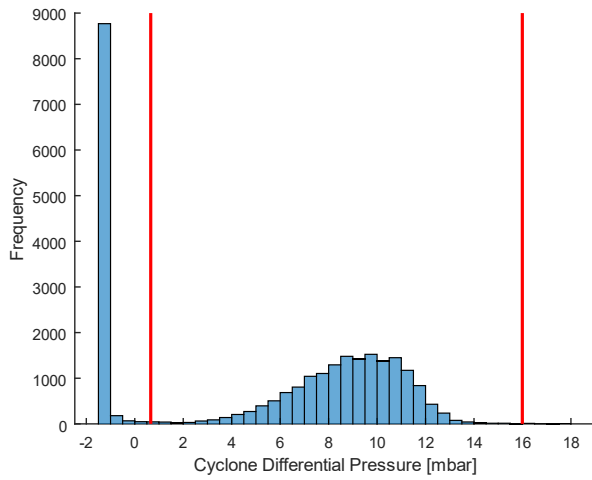
The response (solid blue line) and linear regression trendline (dashed red line) with its mathematical expression for 100 steps to get from the minimum to the maximum as stated in Table 3. This is performed for factors from the Meta-model and the HEXS-model: M1, H1, M2, H2, M3, H3, M9, H9, M16, H16, M19 and H19.

Table 3 Operating range utilized for the HEXS and metamodel factors for the validation experiment.

Meta Symbol	HEXS Symbol	Description	Unit	Minimum	Maximum
M1	H1	Ambient Dry-Bulb Temperature	°C	3.98	11.02
M2	H2	Ambient Absolute Pressure	mBar	996.66	1031.54
M3	H3	Ambient Relative Humidity	%	61.30	95.70
M9	H9	Cyclone, Differential Pressure	mBar	0.66	15.99
M16	H16	Vent Gas Damper position, behind Exhaust Fan	%	19.00	91.00
M17	H17	Recycle Gas Damper position, behind Exhaust Fan	%	8.00	72.00
M19	H19	Burner Outlet Temperature	°C	200.00	403.92







Appendix 6.2 HEXS-model input table

The minimum and maximum as stated in the Table 4 are derived from the operating settings used at the physical Raw Material Preparation Plant, as stated in the for the meta model. The data used is extracted from the Human Machine Interface (HMI), for the run of March 2019. The settings stated in the HMI are the set-points for the operation of the RMPP. These set-points are monitored in the plant by sensors, if deviated from the set-point an electronic control system stabilise the system towards the set-points by adapting settings from sub-systems of the RMPP. Whereby, a control loop is established with the aim of obtaining a steady state operation of the RMPP. For the factors of which no HMI set-points data existed, the initial values where used asset by the HEXS-model developer. However, it should be noted that these set-points can deviate time to time as the plant is not matured fully and is adapted to changes, as are the initial set-points.

The ranges used 80% of full ranges as obtained from the monitored sensor data. The range is used in the SA, to eliminate some of the disturbance, as most of the disturbance occurs at the under and upper bound, these are removed by removing the lower 10% and the upper 10% of the data. This 80% is required as accuracy of the obtained regression increases if the extreme boundaries are not reached. Otherwise, physical monitored outliers can reduce the quality of the meta-model its prediction. In the following two situations it is deviated from this rule. If a process or system can be not present (such as input being turned off, or no rain fall) or if a control limit is reached (valve is open or closed 0 or 100%). This range will be 80% of the entire dataset; and is calculated by the follow rule, i is the sensors, and j is the factor:

$$\begin{aligned} \text{minRange}_j &= \text{maximum}_i + (\text{maximum}_i - \text{minimum}_i) * 0.1 \\ \text{maxRange}_j &= \text{maximum}_i - (\text{maximum}_i - \text{minimum}_i) * 0.1 \end{aligned}$$

The minRange, is in table stated as the minimum, the maxRange is stated as the maximum for each of the factors j . It is named range, as this is the range of the input space utilized in the Sensitivity Analysis.

Table 4 Input factors used for the extended Morris Sensitivity Analysis. The symbol names denoted as H with *, is a factor that is also in Meta-model as an adaptable input factor.

Symbol	Factors	Minimum	Maximum	Baseline
H1*	'Ambient Temperature (°C)'	3.98	11.02	5.00
H2*	'Ambient Pressure (mbar abs)'	996.66	1031.54	1013.00
H3*	'Relative Humidity (%)'	61.30	95.70	50.00
H4	'Feed Rate Coal (t/h)'	0.54	4.85	6.00
H5	'Coal As-Fed H2O (%)'	4.00	12.00	10.00
H6	'Mill Heat Loss (kW)'	5.00	60.00	30.00
H7	'Nitrogen In, Mill outlet (kNm ³ /h)'	0.00	3.00	0.01
H8	'Pressure, Mill outlet (mbar)'	-50.00	10.00	-6.00
H9*	'Differential Pressure, Cyclone (mbar)'	0.66	15.99	7.30
H10	'Main Fan efficiency (%)'	1.00	95.00	70.00
H11	'Nitrogen In, Main Fan outlet (kNm ³ /h)'	0.00	3.00	1.20
H12	'Water output, H2O content remain in coal' (%)'	0.80	3.00	1.00
H13	'Nitrogen In, before Bag house (kNm ³ /h)'	0.00	3.00	0.01
H14	'Baghouse Heat Loss (kW)'	5.00	60.00	30.00
H15	'Exhaust Fan efficiency (%)'	1.00	95.00	70.00
H16*	'Vent Gas, Damper, behind Exhaust Fan (%)'	19.00	91.00	70.00
H17*	'Recycle Gas, Damper, behind Exhaust Fan (%)'	8.00	72.00	50.00
H18	'Makeup Air, before burner (kNm ³ /h)'	0.00	3.00	0.01
H19*	'Burner temperature (°C)'	200.00	403.92	395.00
H20*	'Main Flow, cyclone to main fan (kNm ³ /h)'	2.50	22.50	18.00

Appendix 7. HEXS-model Sensitivity Analysis

Appendix 7.1 Estimated HEXS sensitivity measures

The following table presents the estimated sensitivity measures that are obtained from the extended Morris sensitivity analysis for the HEXS-model ($r=50$, $p=32$), response factor is the mill outlet dry-bulb temperature. Estimated measures are the factors: absolute mean μ^* , the mean μ , and the standard deviation σ .

Table 5 Estimated sensitivity measures for the HEXS-model. The measures are estimated using $r=50$, $p=32$.

Symbol	Factors	μ^*	μ	σ
H1*	'Ambient Temperature (°C)'	1599	-807	5627
H2*	'Ambient Pressure (mbar abs)'	2145	-1993	6431
H3*	'Relative Humidity (%)'	2541	703	10345
H4	'Feed Rate Coal (t/h)'	5317	-3562	26653
H5	'Coal As-Fed H2O (%)'	4962	-2981	24293
H6	'Mill Heat Loss (kW)'	2072	723	6743
H7	'Nitrogen In, Mill outlet (kNm ³ /h)'	3780	3233	10316
H8	'Pressure, Mill outlet (mbar)'	49349	48222	342948
H9*	'Differential Pressure, Cyclone (mbar)'	5268	1372	20068
H10	'Main Fan efficiency (%)'	12638	5007	45936
H11	'Nitrogen In, Main Fan outlet (kNm ³ /h)'	3077	-797	8407
H12	'Water output, H2O content remain in coal' (%)'	11769	-9782	53086
H13	'Nitrogen In, before Baghouse (kNm ³ /h)'	1215	-985	4226
H14	'Baghouse Heat Loss (kW)'	1137	-710	2758
H15	'Exhaust Fan efficiency (%)'	1602	177	4121
H16*	'Vent Gas, Damper, behind Exhaust Fan (%)'	2887	-1936	7648
H17*	'Recycle Gas, Damper, behind Exhaust Fan (%)'	8840	4605	45749
H18	'Makeup Air, before burner (kNm ³ /h)'	1263	124	3377
H19*	'Burner temperature (°C)'	52518	-51097	339099
H20	'Main Flow, cyclone to main fan (kNm ³ /h)'	10635	9389	37261

Appendix 7.2 The HEXS screen plots

In the following four plots are presented from the extended Morris sensitivity analysis for the HEXS-model($r=50$, $p=32$), the response factor is the mill outlet dry-bulb temperature. In the Figure 10 the sample mean is displayed these are also displayed in Figure 11, however in a red particular section is zoomed in. As discussed in the sensitivity analysis section, it is known that when utilizing the sample mean instead of absolute mean factors could lead to occurrence that factors cancel each other out by opposite signs. The absolute mean μ^* is presented in Figure 13 and in log-log scale displayed in Figure 16. As can be noted from Figure 10, Figure 11 and Figure 13 it is challenging to show all factors in one plot when; some factors overlap each other; and for factors that have a much large standard deviation and/or absolute mean than the bulk of data are displayed at the edges of the graph. To be able to display all the factors in one plot a log-log is plotted as is shown in Figure 16, where both the horizontal and vertical axes use logarithmic scales. Another benefit is that enables to distinguish multiple groups of factors. It should be noted, that a log-log plot cannot be performed for plots with mean values, as these screen plots can consists out of negative numbers as shown in Figure 10 and Figure 11; a logarithm of negative number does not exist “it is undefined”, once should then use complex numbers instead, which is left for further research. These similar results hold for the metamodel, but these are not displayed here as these plots hold even more factors than the HEXS-model.

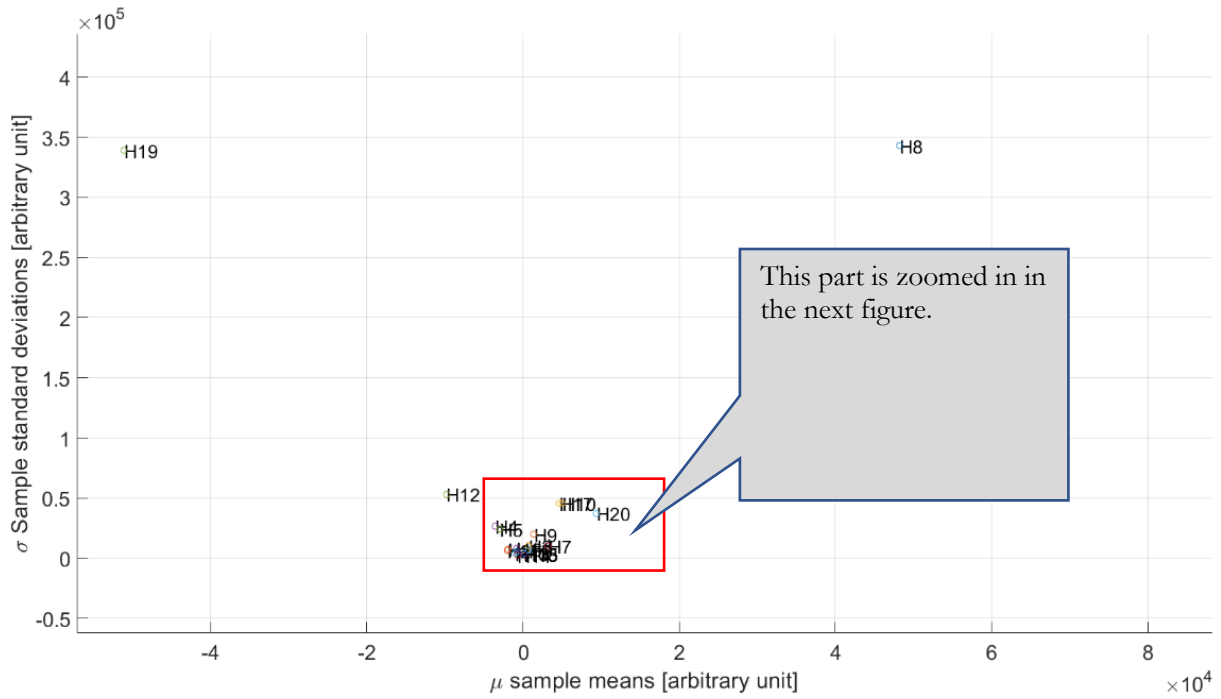


Figure 10 The graph displaying the conventional Morris sensitivity measures μ the mean and σ the standard deviation, for 20 HEXS-model input factors. The corresponding description of each of the factors can be found in Table 6.

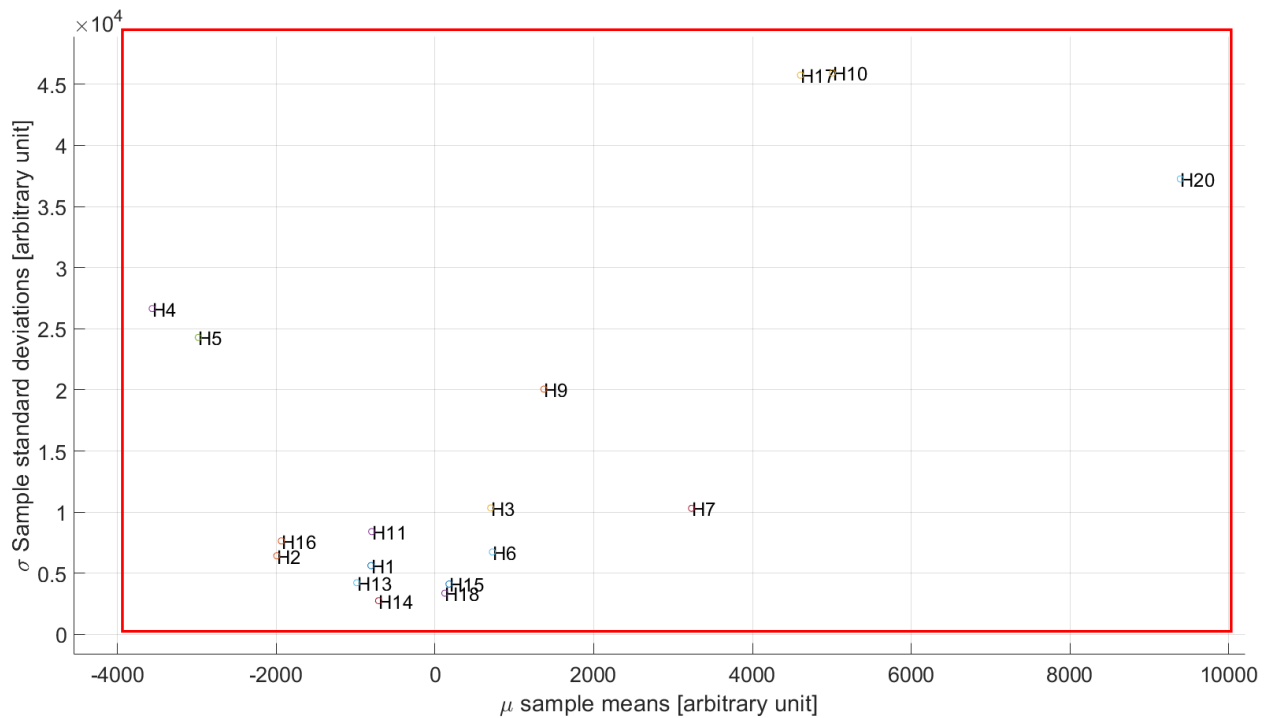


Figure 11 The ZOOMED graph displaying the conventional Morris sensitivity measures μ the mean and σ the standard deviation, for 20 HEXS-model input factors. The graph is zoomed for the section of the previous graph Figure 12. The corresponding description of each of the factors can be found in Table 7.

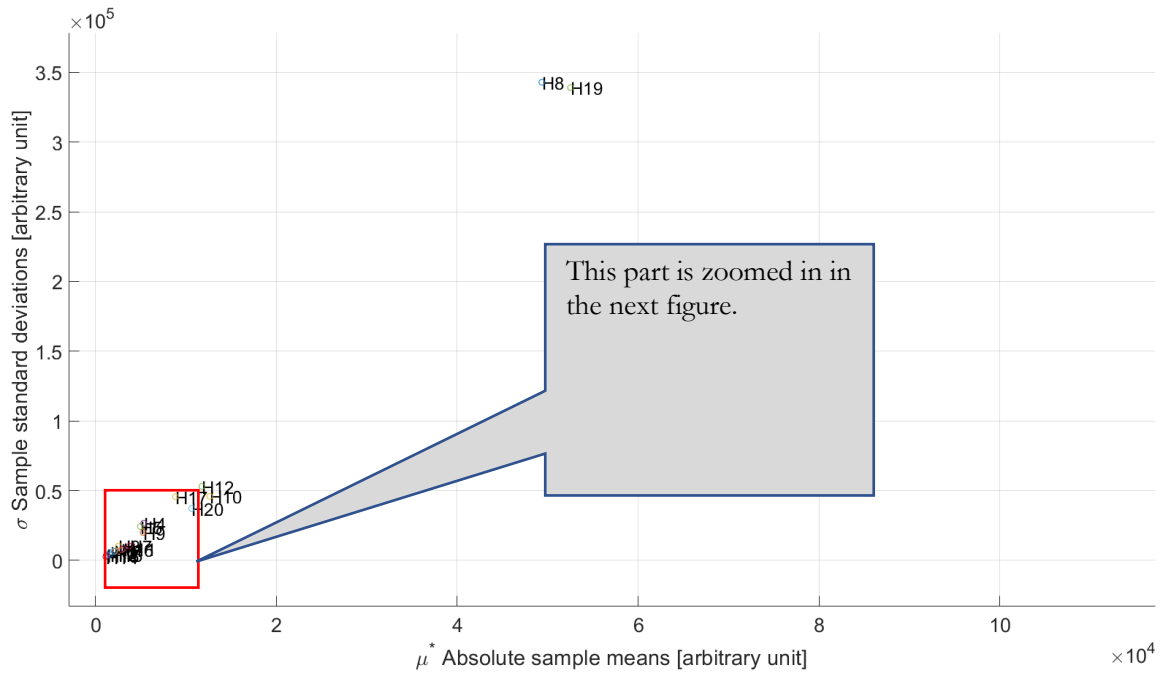


Figure 13 The graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 20 HEXS-model input factors. The graph is shows that there exists skewness, as factors H8 and H19 have much larger values than the bulk data. The corresponding description of each of the factors can be found in Table 8.

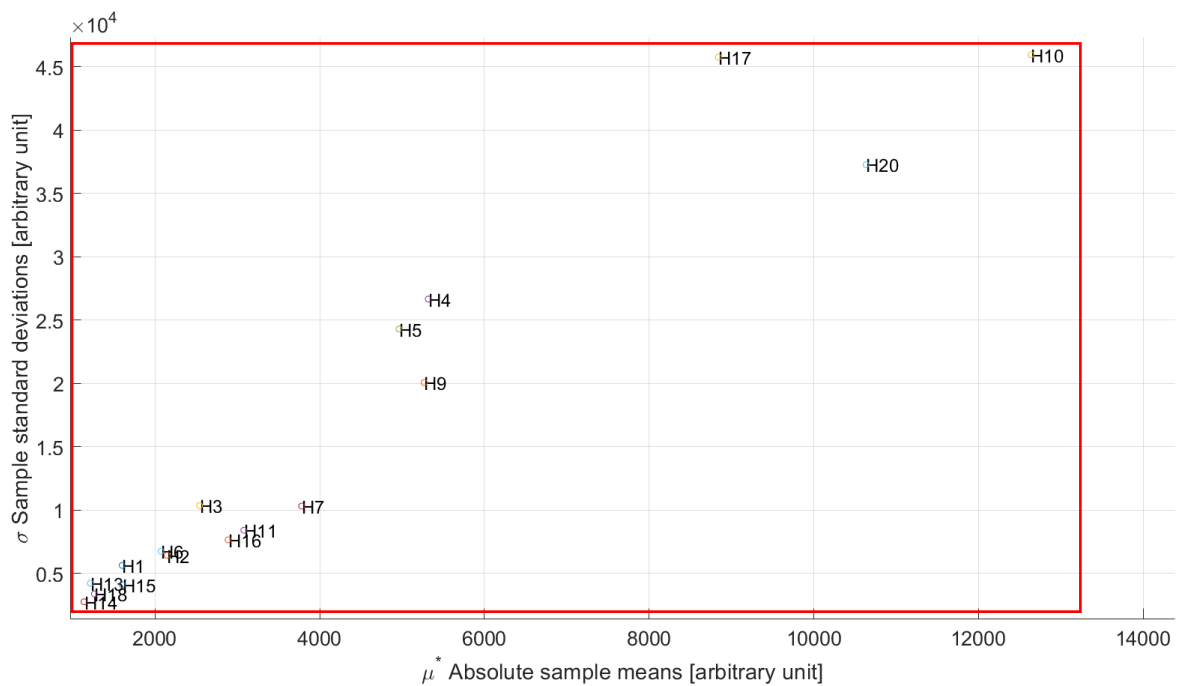


Figure 14 The ZOOMED graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 20 HEXS-model input factors. The graph is zoomed for the section of the previous graph Figure 15. The corresponding description of each of the factors can be found in Table 9.

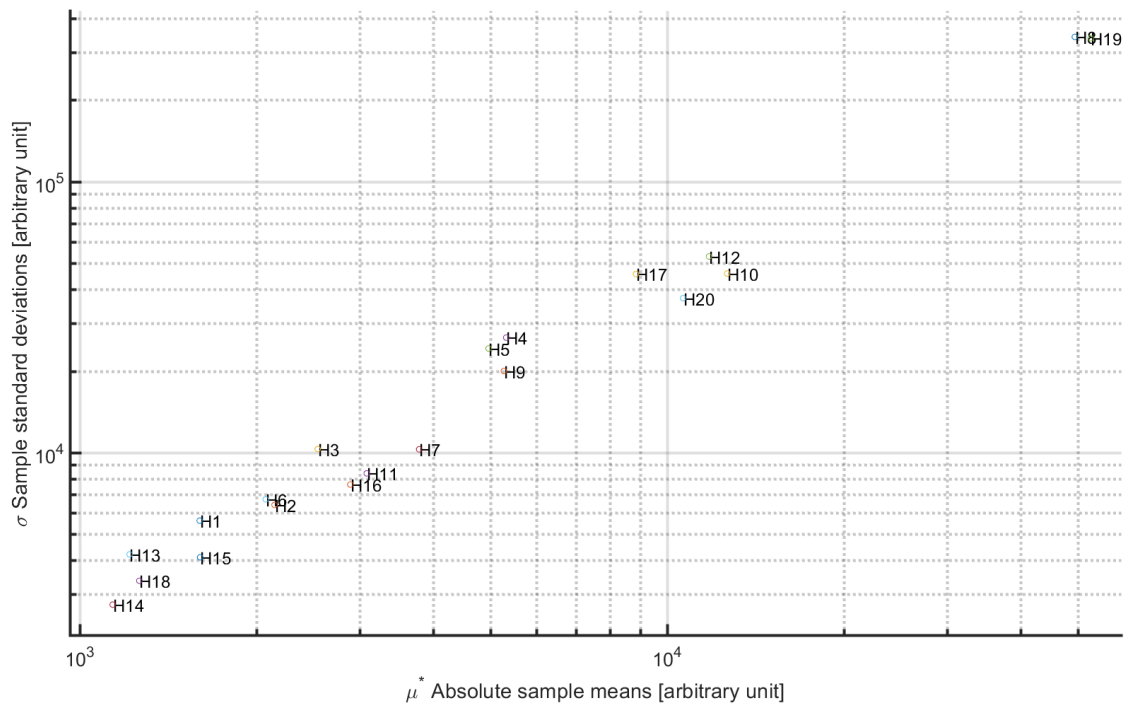


Figure 16 The log-log graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 20 HEXS-model input factors. The graph is displayed as log-log plot to respond to skewness, as factors H8 and H19 have much larger values than the bulk data. The corresponding description of each of the factors can be found in Table 10.

Appendix 8. Meta-model Sensitivity Analysis

Appendix 8.1 Selecting r trajectories

In the section it is investigated how similar or varied the set of observed absolute means are for different trajectories r being 50, 100, 200 and 400. Each of the trajectories is computed three times, and their results are evaluated by the pareto bar plot, making a total of 12 experiments. By this, it is investigated if the number of trajectories chosen in the extended Morris sensitivity analysis for the Meta-model has influence on the output. This is, researched as the number of input factors k is 44 for the Meta-model in comparison to $k = 20$ for the HEXS-model is different. It is known that the computational space explored by using similar levels p and trajectories r for the Meta-model would be less, in comparison to the HEXS-model. If the amount of simulation is linear scalable to obtain equal space exploring, then r would be approximated to be 113 for the Meta Sensitivity Analysis. This is calculated by the following formula, where subscript M is for Meta, and H for HEXS.

$$r_M = r_H \left(\frac{(k_M + 1)}{(k_H + 1)} \right) = 50 \left(\frac{45}{21} \right) \sim 113$$

To determine if $r = 50$ or that 113 is appropriate to use for SA, the variability of the pareto bar plot is examined for different trajectories of r as discussed. The setup is that for each of the four trajectories, 3 pareto bar plots are computed, whereby the spread of the 7 most influential factors is examined further. When these in total 12 pareto bar plots are computed, for each plot the first 7 factors with the highest absolute mean value are identified. How, often a specific factor is identified as 1 out of 7, with 1 being the most influential and thus having the highest value for the absolute mean. The data from the 12 pareto bar plots as shown in Figure 18, and the first most influential factors are summarized in Table 11 “the count matrix” positions #. From Table 11 it is observed that for all trajectories and is repeating M5, M29, M41 and M19 have the largest absolute mean value. From position 5 onward, the spread increases of which factor has the largest absolute mean value. If, each factor would have a unique absolute mean value for all the experiments, a diagonal matrix would be obtained. However, this is not the case; as can be observed from position 5#; as position 5# M42 has 9 out of 12 experiments a position of 5#; M11 has two times this position and M25 only ones. This spread of factors across the position space # increases when more positions are included.

		How often is a factor at position #											
		Description	Unit	Symbol	1#	2#	3#	4#	5#	6#	7#	Sum	
Factors		Precipitation pumping station 3 at Ijmuiden.	mm/hr	M5	12							12	
		System flow temperature measurement	°C	M29		12						12	
		Temperature measurement bagfilter inlet	°C	M41			12					12	
		Temperature measurement Burner hot gas outlet	°C	M19				12				12	
		Temperature measurement bagfilter outlet	°C	M42					9	3		12	
		Nitrogen injection control valve mill hot gas inlet	%	M11					2	5	4	11	
		Underpressure measurement mill inlet	mBar	M25					1	4	6	11	
		Mill motor M-822	kW	M21								1	1
		Pressure measurement natural gas supply.	mBarG	M44								1	1

Table 11 Counting Matrix, counts how often a factor occurs as one of the 7 most influential factors or positions #.

As stated, when all the factors have a unique position, and are not influenced by the number of trajectories selected in the SA; it would result in a diagonal line on the count matrix. Meaning that there is only one factor for each position, instead of having a spread. As can be seen from the count matrix, this is not the case, meaning that there is some position spread for each factor for a given number of trajectories. To quantify this spread for each numbers of trajectory, the following expression is obtained for a reference value. The total number of positions for this scenario is selected to be 7, for ideal diagonal case it will occupy 7 positions “space-filling”. This is set to be the reference filling $7 \times 7 = 49$. For each trajectory there has been created three Pareto Plots. The more varied the results from the SA, the larger the occupied space would be. An example is presented in Table 12, where 12 positions are held by 8 factors, making an occupied space of 196%, meaning that the variability of the absolute mean is larger for the SA results.

$$Reference = 7 * 7 = 100\%$$

$$Occupied\ space = \frac{Space\ filling * number\ of\ factors}{Reference} * 100\%$$

Table 12 Table obtained from 3 experiments from the Pareto Bar plot with trajectories r being set to 50. The calculated occupied space is 196%.

		Position #								
		1#	2#	3#	4#	5#	6#	7#		
r=50	Space filling	12	M5	3						
	factors	8	M29		3					
	Occupied space	196%	M41			3				
			M19				3			
			M42					2	1	
			M25					1	1	1
			M11						1	1
			M44							1

From Figure 17 it can be observed that occupied space for a factor in the count matrix is larger for Sensitivity Analysis performed with a smaller number of trajectories than that with a larger number of r . Note that occupied space for trajectories 200 and 400 have a similar occupied space. It is observed although it is not presented here in detail, that the occupied space increases when more positions are included; whereby the variability of the SA results increases. From this experiment it can be concluded although less trajectories are being used, the most influential factors or similar to those experiments that utilized a higher value of r . Thus, it is likely that the factor with the highest absolute mean are recognized. However, the order whereby the factors are ranked are sensitive for variance depending on the trajectory r value chosen. In the performed simulation the variance is similar for a r of 200 and 400; a selection of $r = 200$ is chosen for the Sensitivity Analysis.

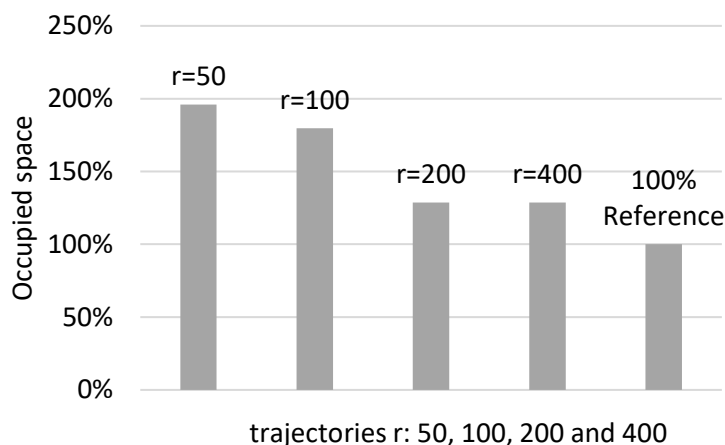


Figure 17 Occupied space, for various trajectories for the Sensitivity Analysis

Repeated 3 times for similar r

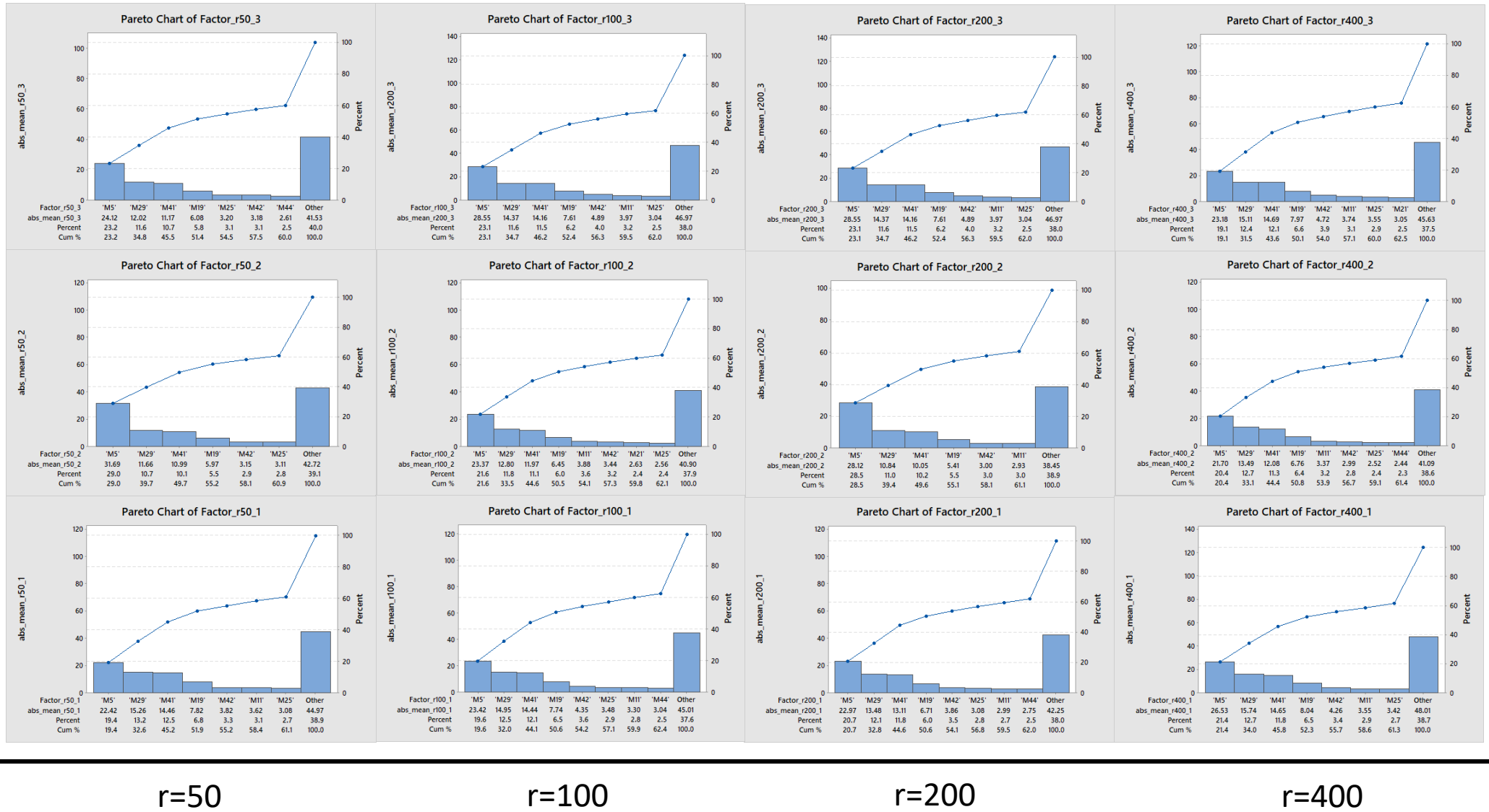


Figure 18. Pareto bar plot created for $r=50, 100, 200$ and 400 . Each of results is repeated three times. The entire data set used is the from run of March 2019, with the regression technique model being GPR Rational Quadratic.

In the investigation of identifying the number of trajectories r for the meta-model required as input for extended Morris Method; the consistency of the order of influence is evaluated; by varying the r trajectories being 50, 100, 200 and 400 and repeating this for three times for each trajectory. During this examination it is demonstrated that for all 12 experiments the four stated factors and, in their order, have consistent results. These factors were identified by all experiments, in their order and having the highest absolute mean value. For factors with lower absolute mean values, it is observed that their order of influence is less consistent; as their absolute mean values reduces their inconsistencies in order increases.

Appendix 8.2 The Meta screen plots

Conventional screen plot of extended Morris method for Meta-models, input parameters from Table 13.

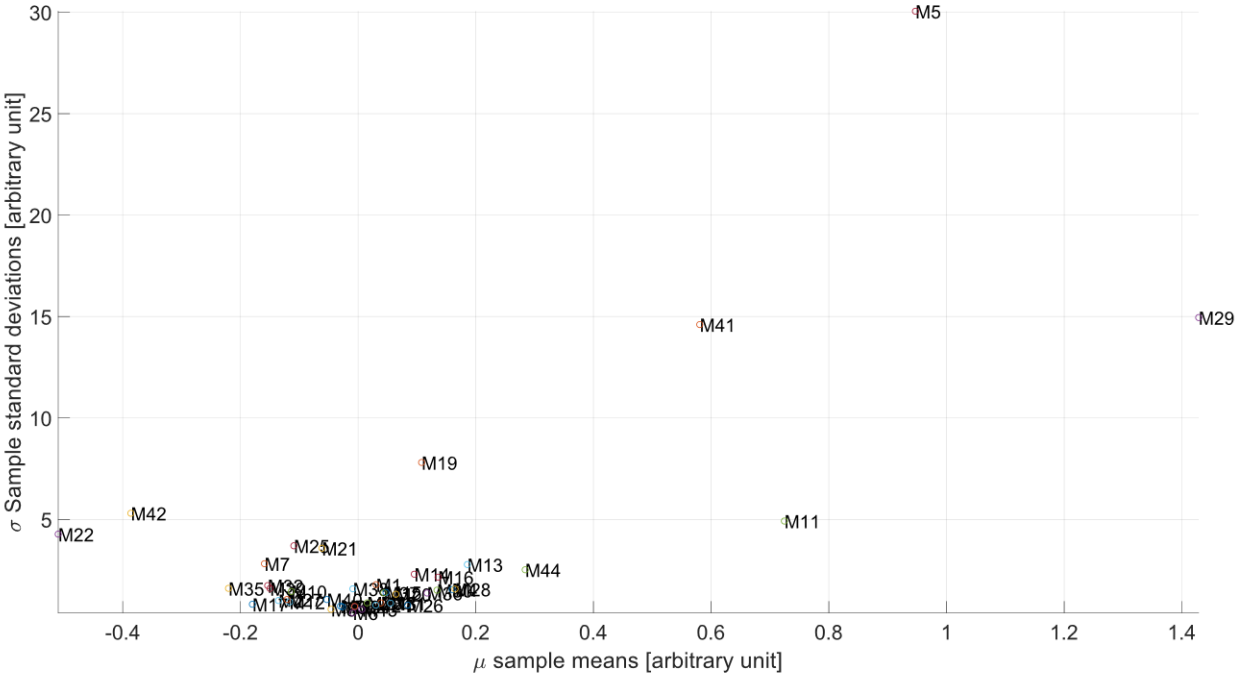


Figure 19. The graph displaying the conventional Morris sensitivity measures μ the mean and σ the standard deviation, for 44 metamodel input factors. The corresponding description and input range of each of the factors can be found in Table 2.

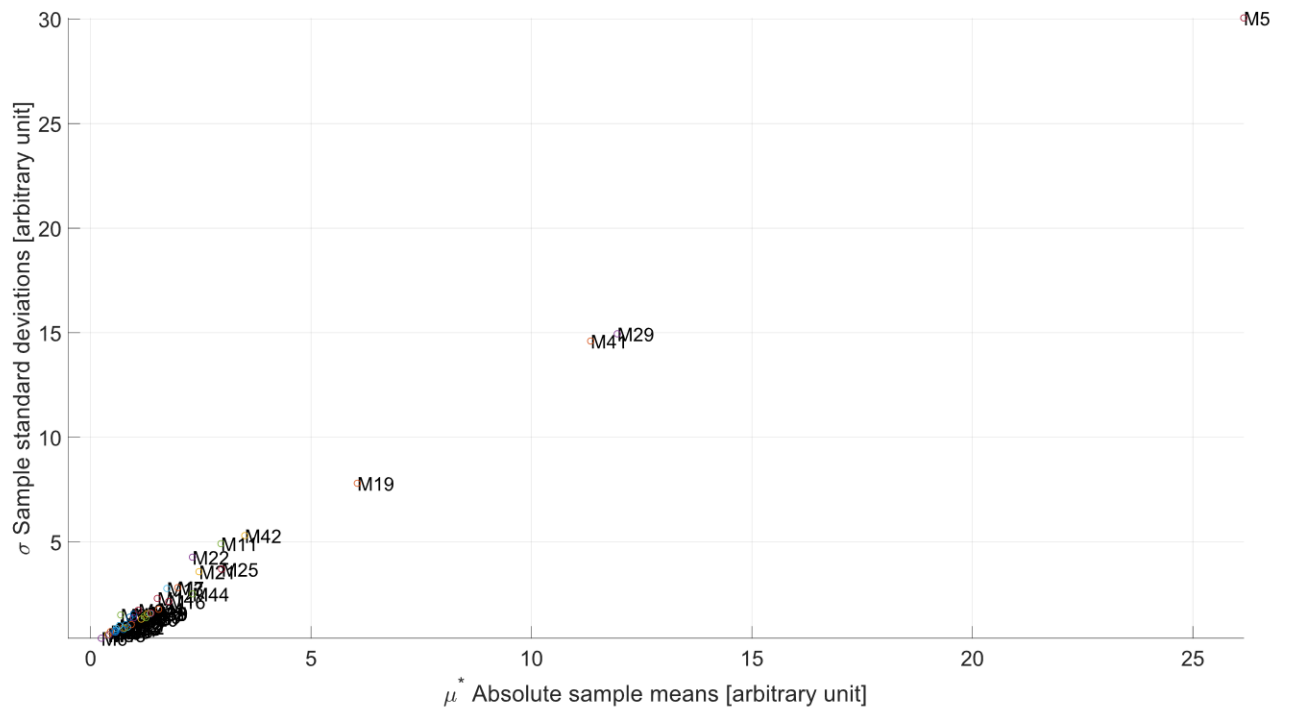


Figure 20 The graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 44 metamodel input factors. The corresponding description and operating range of each of the factors can be found in Table 2.

Zoom 1

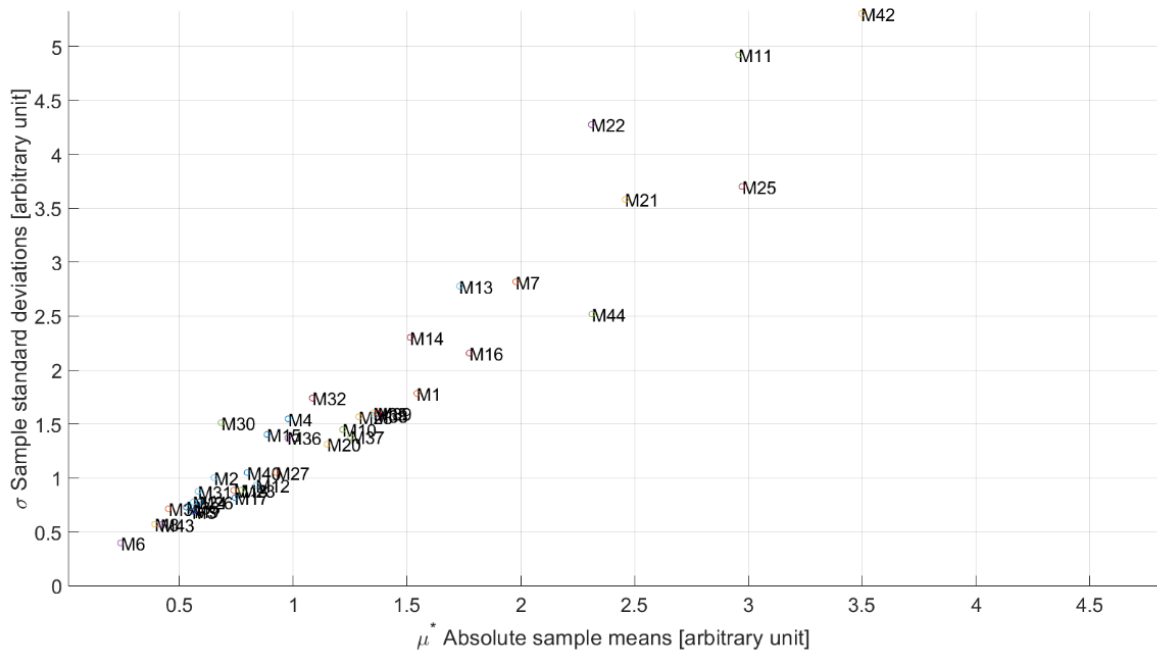


Figure 21 The Enlarged graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 44 metamodel input factors. The corresponding description and operating range of each of the factors can be found in Table 2.

Zoom 2

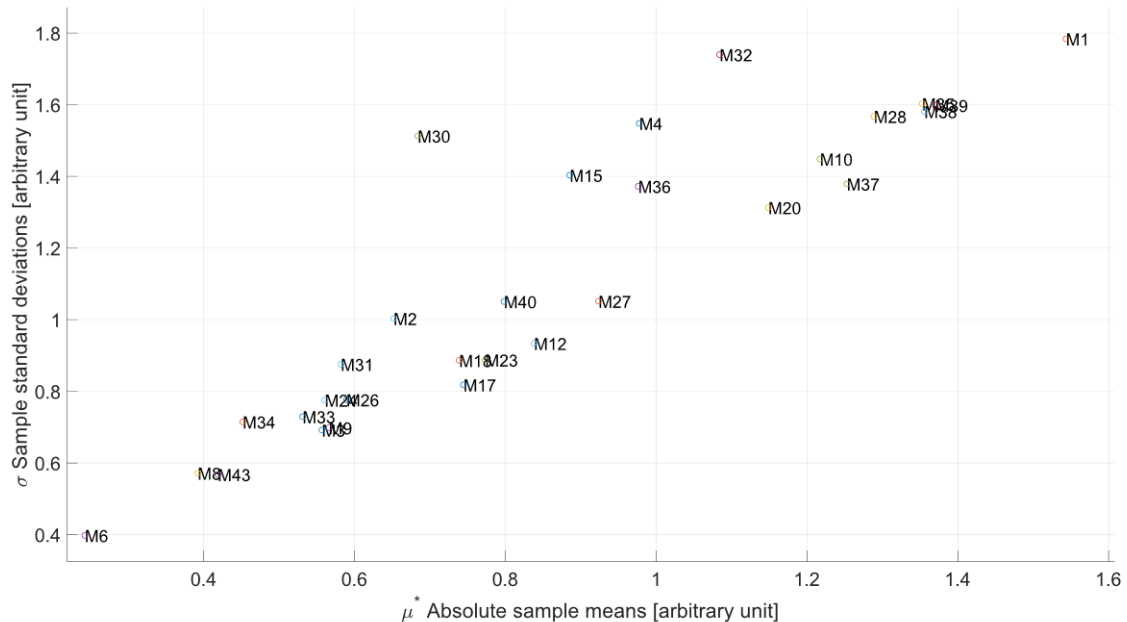


Figure 22 The even more enlarged graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 44 metamodel input factors. The corresponding description and operating range of each of the factors can be found in Table 2.

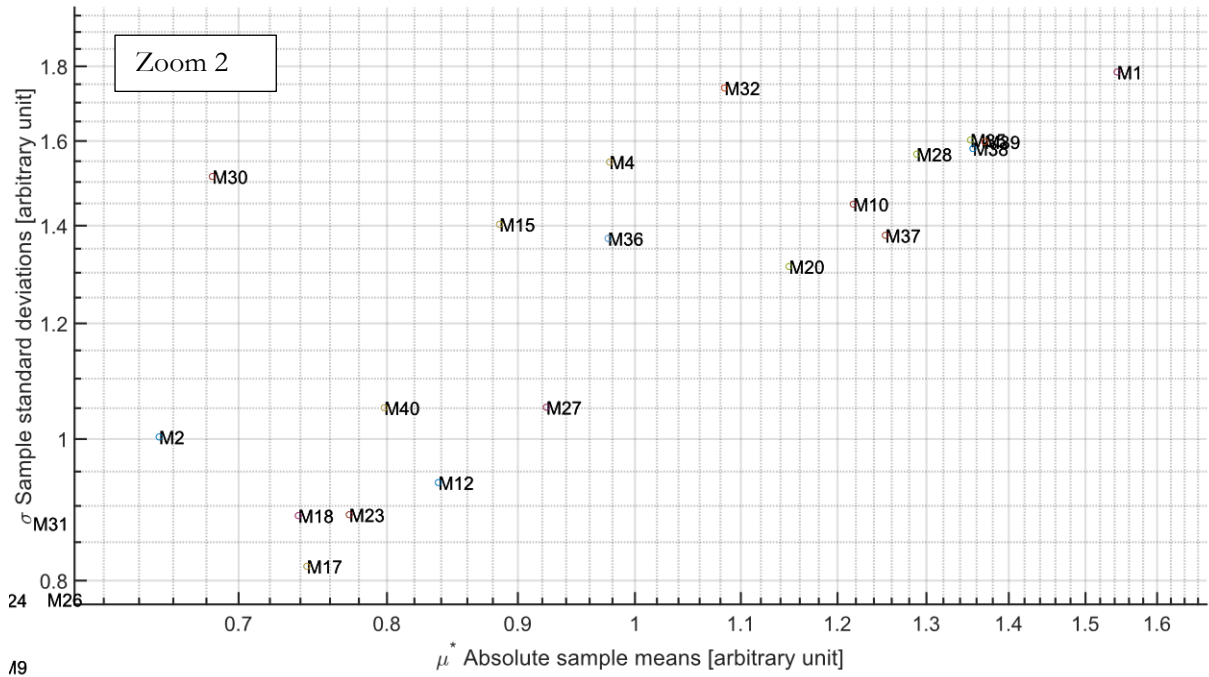


Figure 25 Zoom 2: The log log graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 44 metamodel input factors. The corresponding description and operating range of each of the factors can be found in T'a

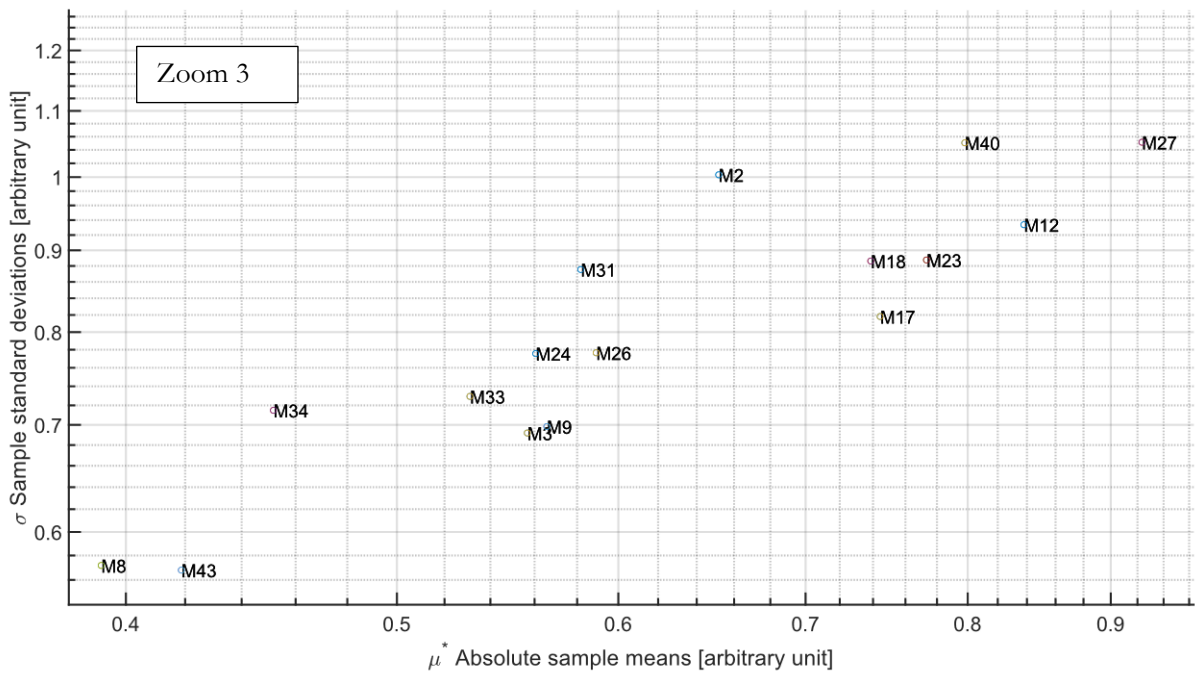


Figure 26. Zoom 3: The log log graph displaying the extended Morris sensitivity measures μ^* the absolute mean and σ the standard deviation, for 44 metamodel input factors. The corresponding description and operating range of each of the factors can be found in T'

Table 14 Numerical SA results of the metamodel.

Symbol	Factors	μ^*	μ	σ
M1*	Ambient dry-bulb temperature at Hoogoven harbour Tata	1.543	0.031	1.784
M2*	Barometer absolute pressure at Hoogoven harbour Tata	0.652	-0.135	1.003
M3*	Relative humidity at Hoogoven harbour Tata	0.557	-0.027	0.692
M4	Total Weight V-822	0.977	0.160	1.548
M5	Precipitation pumping station 3 at Ijmuiden.	26.155	0.947	30.052
M6	Burner BU-825 Power	0.242	-0.009	0.398
M7	Nitrogen injection control valve cyclone inlet	1.978	-0.159	2.819
M8	Differential pressure mill drying column DY-822	0.392	-0.046	0.572
M9*	Differential pressure measurement cyclone	0.566	-0.006	0.698
M10	Flow measurement coal system inertisation nitrogen supply	1.217	-0.114	1.448
M11	Nitrogen injection control valve mill hot gas inlet	2.957	0.724	4.921
M12	Pressure measurement 3 BarG nitrogen supply	0.838	-0.117	0.934
M13	Nitrogen injection control valve bag filter inlet	1.730	0.186	2.778
M14	5D24 area 18bar nitrogen supply temperature	1.514	0.096	2.302
M15	Level measurement Coal supply bin V-822	0.885	0.047	1.403
M16*	Exhaust stack damper	1.773	0.136	2.156
M17*	Recycle Gas damper	0.744	-0.180	0.818
M18	Mill motor M-822	0.739	0.039	0.886
M19*	Temperature measurement Burner hot gas outlet	6.049	0.108	7.808
M20	System flow control valve inlet main fan	1.149	0.064	1.313
M21	Mill motor M-822	2.459	-0.061	3.580
M22	Spinner separator S-822	2.310	-0.510	4.275
M23	Spinner separator S-822	0.773	0.015	0.888
M24	Spinner separator S-822	0.561	0.030	0.776
M25	Under pressure measurement mill inlet	2.972	-0.109	3.701
M26	Differential pressure flow measurement cyclone outlet	0.589	0.084	0.777
M27	Main fan motor K-823	0.923	-0.121	1.052
M28	Under pressure measurement system flow measurement	1.289	0.165	1.567
M29	System flow temperature measurement	11.945	1.428	14.952
M30	Exhaust fan inlet damper	0.684	0.134	1.513
M31	N2 Injection burner Valve	0.582	0.055	0.876
M32	Exhaust fan motor K-824	1.084	-0.154	1.740
M33	Dust content measurement after bag filter	0.531	-0.030	0.729
M34	Pressure difference measurement baghouse	0.452	-0.006	0.715
M35	Under pressure measurement Exhaust fan Inlet"	1.353	-0.220	1.603
M36	Under pressure measurement recirculation	0.976	0.117	1.372
M37	Pressure measurement cleaning air supply.	1.253	0.044	1.379
M38	CO measurement 1	1.356	-0.009	1.581
M39	O2 measurement 1	1.370	-0.150	1.598
M40	Measured O2 value	0.798	-0.053	1.051
M41	Temperature measurement bag filter inlet	11.342	0.581	14.606
M42	Temperature measurement bag filter outlet	3.497	-0.386	5.306
M43	Flow measurement natural gas supply Gas heater BU-825	0.419	0.006	0.568
M44	Pressure measurement natural gas supply.	2.312	0.284	2.518

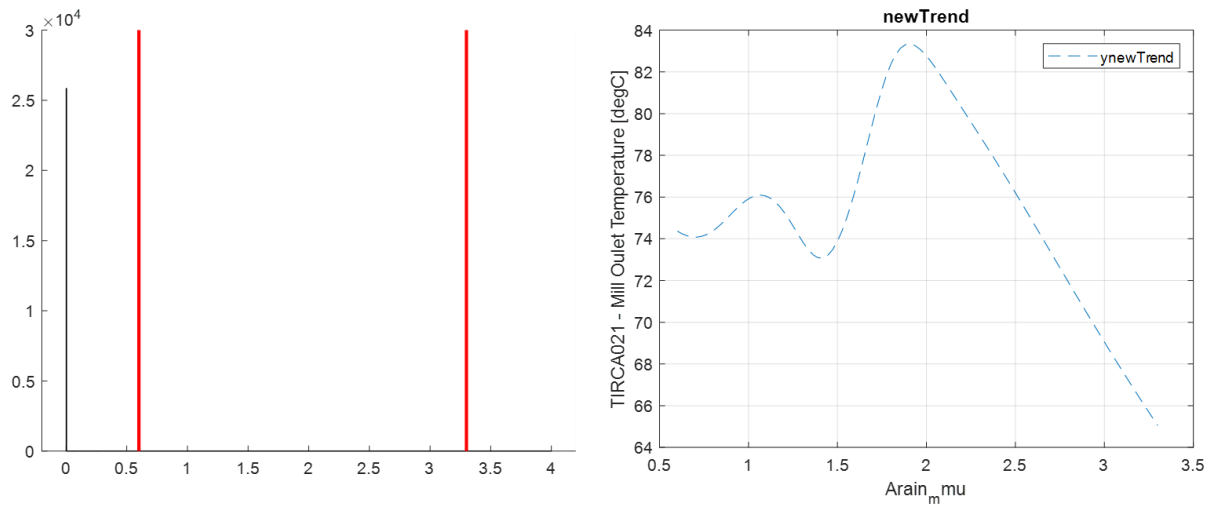


Figure 27 Deep dive for the participation (rainfall) of the metamodel factor M5. On the left histogram of the historical data, on the right the prediction response for the metamodel.

The rain fall measured as mm per hour, the only factor that is measured in another resolution that is utilized as an input factor for the meta-model. As can be observed, this predicted value with the meta-model presents a highly non-linear behaviour for the mill outlet temperature for lower amounts of rain fall. This is classified as an effect C, for the Sensitivity Analysis. On the left side, a histogram is presented, the vertical red-lines presenting the operating domain for the predictions. Furthermore, it can be observed, that during the period is was most often dry, as in the histogram there is steep peak at zero.

Appendix 9. Sensitivity comparison, HEXS-model vs Meta-model

The previous results are filtered for seven factors. The seven selected factors of each method are those factors who are monitored at a similar location in the process, have an equal physical quantity and unit. Their sensitivity results are compared, by first normalizing there results to percentage. By normalizing, there results can be compared. For each of the tables the absolute mean and standard deviation of the elementary effects are calculate to the relative percentage as: $x_{relative} = \frac{x_i - x_{min}}{x_{max} - x_{min}} * 100\%$. Where x is either the absolute mean or the standard deviation.

Table 15. HEXS-model results extended Morris method, ranked by highest absolute mean first.

HEXS-Model						
Symbol	Factors	μ^*	μ	σ	μ^* [%]	σ [%]
H1*	Ambient Temp (°C)	1599	-807	5627	0%	0%
H2*	Ambient P (mbar abs)	2145	-1993	6431	1%	0%
H3*	Relative Humidity (%)	2541	703	10345	2%	1%
H9*	Differential Pressure, Cyclone (mbar)	5268	1372	20068	7%	4%
H16*	Vent Gas, Damper, behind Exhaust Fan (%)	2887	-1936	7648	3%	1%
H17*	Recycle Gas, Damper, behind Exhaust Fan (%)	8840	4605	45749	14%	12%
H19*	Burner temperature (°C)	52518	51097	339099	100%	100%

Table 16. Meta-model results extended Morris method

Meta-model						
Symbol	Factors	μ^*	μ	σ	μ^* [%]	σ [%]
M1*	Ambient dry-bulb temperature at Hoogoven harbour Tata (°C)	1.543	0.031	1.784	18%	15%
M2*	Barometer absolute pressure at Hoogoven harbour Tata (mBar)	0.652	-0.135	1.003	2%	4%
M3*	Relative humidity at Hoogoven harbour Tata (%)	0.557	-0.027	0.692	0%	0%
M9*	Differential pressure measurement cyclone (mBar)	0.566	-0.006	0.698	0%	0%
M16*	Exhaust stack, Vent Gas damper (%)	1.773	0.136	2.156	22%	21%
M17*	Recycle Gas damper, behind Exhaust Fan (%)	0.744	-0.180	0.818	3%	2%
M19*	Temperature measurement Burner hot gas outlet (°C)	6.049	0.108	7.808	100%	100%

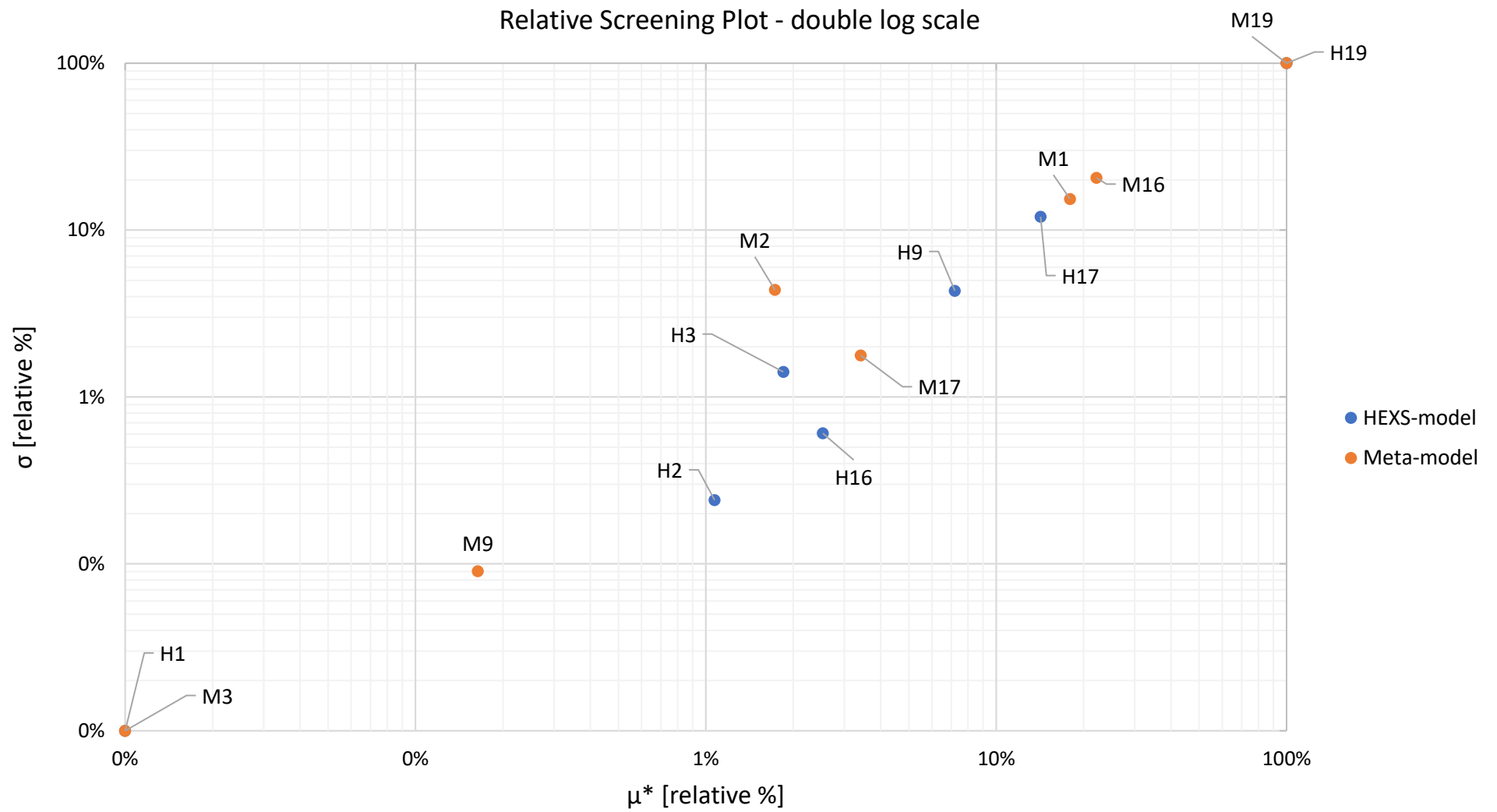


Figure 28. Sensitivity analysis results of the HEXS-model and the metamodel, for seven comparable input factors for each model.

Appendix 10. Potential functional failure conditions determined with HEXS-model

Appendix 10.1 Input HEXS-model

The table as reported are the settings utilized for the SA analysis.

Table 17 Input values HEXS-model for finding potential functional failure conditions.

Case	Independent variable	Minimum	Maximum	Baseline
H1	Ambient Temp (°C)	-25	45	5
H2	Ambient P (mbar abs)	900	1100	1013
H3	Relative Humidity (%)	20	106	50
H4	Feed Rate Coal (t/h)	0	10	6
H5	Coal As-Fed H ₂ O (%)	1	25	10
H6	Mill Heat Loss (kW)	5	60	30
H7	Nitrogen In, Mill outlet (kNm ³ /h)	0	3	0.01
H8	Pressure, Mill outlet (mbar)	-50	50	-6.00
H9	Differential Pressure, Cyclone (mbar)	-5	20	7.3
H10	Main Fan efficiency (%)	1	95	70
H11	Nitrogen In, Main Fan outlet (kNm ³ /h)	0	3	1.2
H12	Water output (%)	0.01	10	1
H13	Nitrogen In, before Baghouse (kNm ³ /h)	0	3	0.01
H14	Baghouse Heat Loss (kW)	5	60	30
H15	Exhaust Fan efficiency (%)	1	95	70
H16	Vent Gas, Damper, behind Exhaust Fan (%)	0	100	70
H17	Recycle Gas, Damper, behind Exhaust Fan (%)	0	100	50
H18	Makeup Air, before burner (kNm ³ /h)	0	10	0.01
H19	Burner temperature (°C)	100	550	395
H20	Main Flow, cyclone to main fan (kNm ³ /h)	6	25	18

Appendix 10.2 OFAT potential functional failure by ΔDT

For each factor the initial value is set to each minimum and increased to its maximum in 100 steps, as reported in Table 17. For each incremental step the HEXS-model is returned to its baseline values as stated in Appendix 6.1; where its objective is to dry the fed coal to 1% moisture content with an outlet burner temperature of 395 °C. In the process the change in the drying medium (solid grey line), dewpoint (solid blue line) and difference in temperature (solid red line) is observed as presented in each of the figures. When ΔDT crosses zero, a potential functional failure can occur as then the driving force for drying comes to a hold.

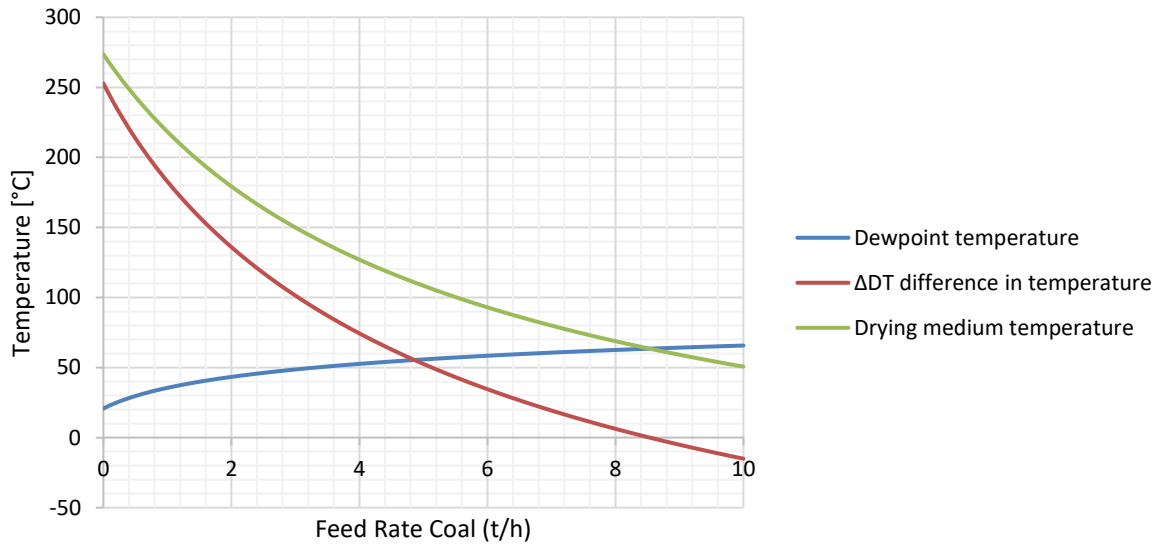


Figure 29. Functional failure caused by increasing feed rate of the coal, temperature measured at the outlet of the mill. The drying medium temperature is the mill outlet temperature.

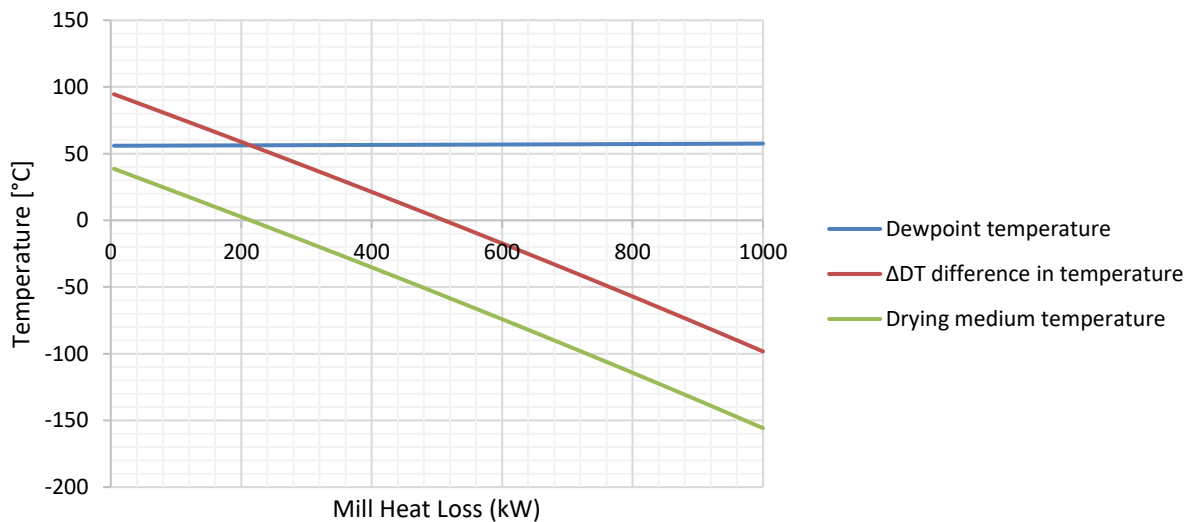


Figure 30 Functional failure caused by increasing mill heat loss, temperature measured at the outlet of the mill. The drying medium temperature is the mill outlet temperature.

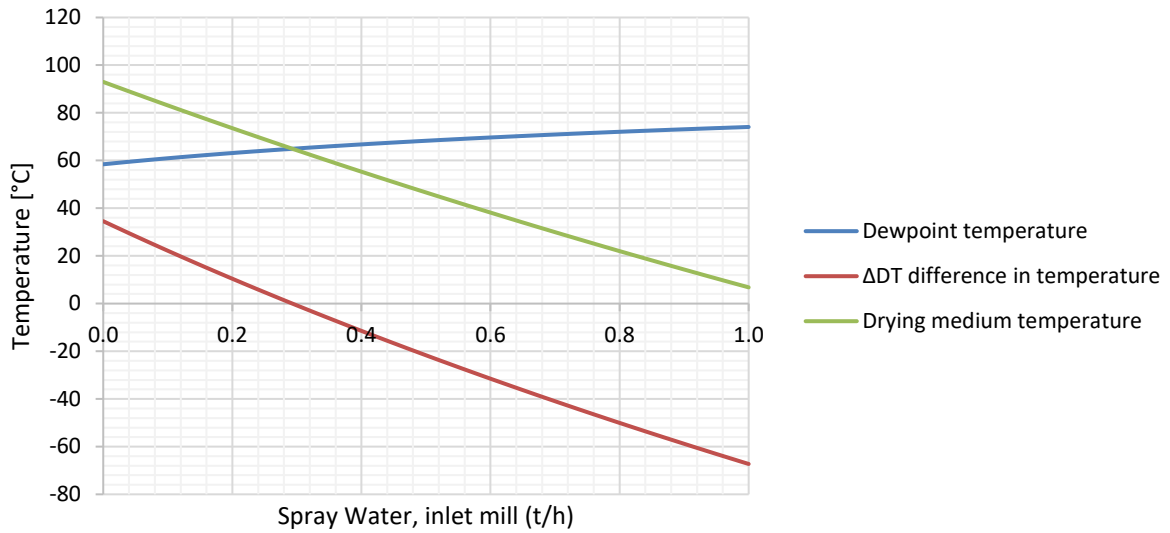


Figure 33. Functional failure caused by increasing spray water, temperature measured at the outlet of the mill. The drying medium temperature is the mill outlet temperature.

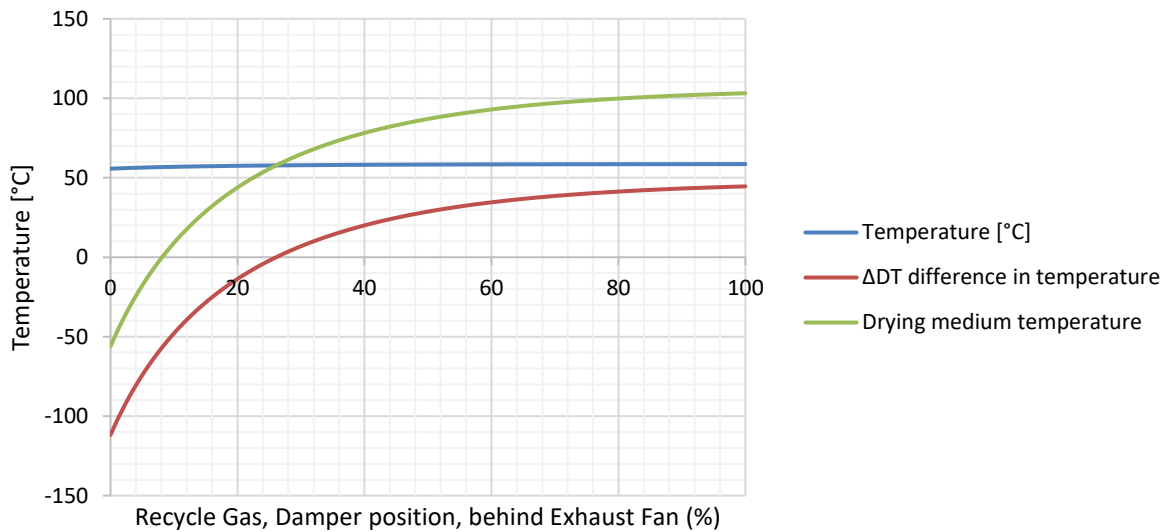


Figure 31. Functional failure caused by increasing the opening position of the recycle gas damper, temperature measured at the outlet of the mill. The drying medium temperature is the mill outlet temperature.

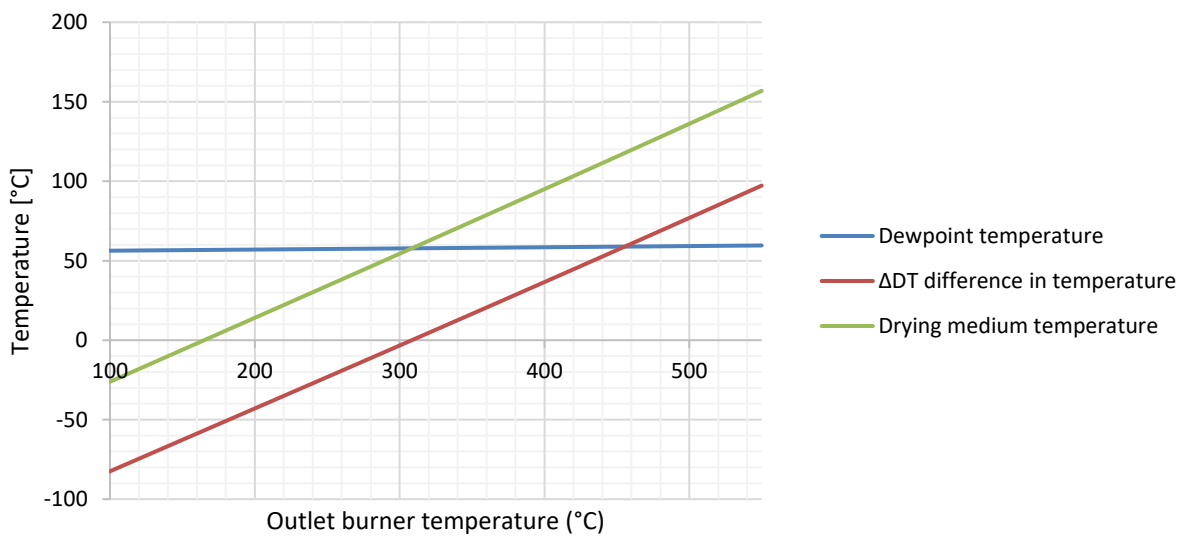


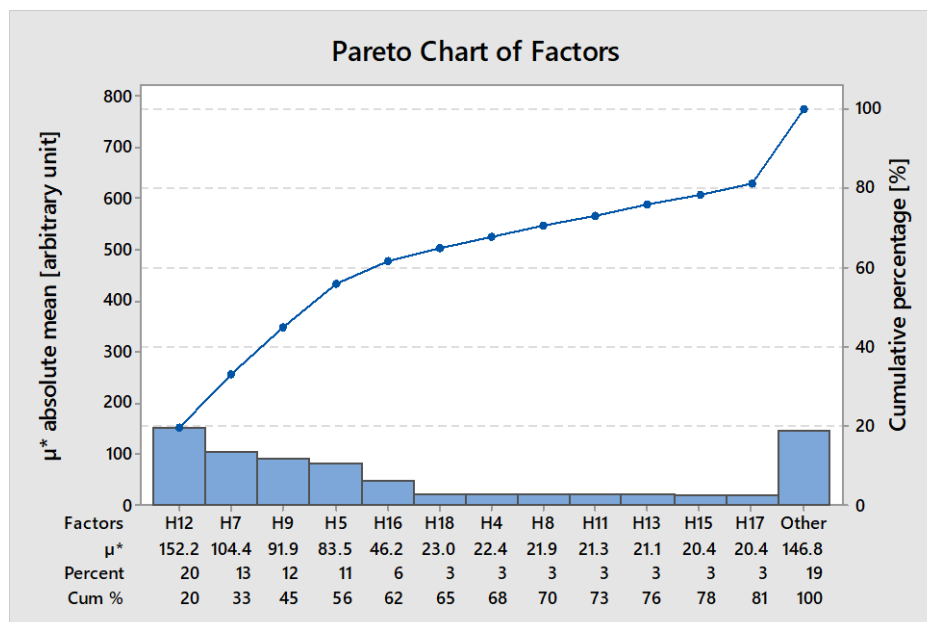
Figure 32. Functional failure caused by increasing burner outlet temperature, temperature measured at the outlet of the mill. The drying medium temperature is the mill outlet temperature.

Appendix 10.3 Counteracting potential function failure by utilizing Sensitivity Analysis, response ΔDT

In the following results are presented from the extended Morris sensitivity analysis for the HEXS-model ($r=50$, $p=32$), the response factor ΔDT , being the in the mill the difference in temperature between the medium temperature and the dewpoint temperature. With the input factors as stated in Table 17. By utilizing the most dominant influencing but controllable factors of the SA for the response ΔDT . The dominant controllable factor can be a candidate to adjust its operating setting such, so that an operational functional failure can be prevented.

Results SA:

Response DT: dew DT (difference between Tg-Tdew) [degC]				
Symbol	Factors	μ^*	μ	σ
H12	'Water output (%) left in coal'	152	124	960
H7	'Nitrogen In, Mill outlet (kNm ³ /h)'	104	69	600
H9	'Differential Pressure, Cyclone (mbar)'	92	35	444
H5	'Coal As-Fed H2O (%)'	83	47	450
H16	'Vent Gas, Damper, behind Exhaust Fan (%)'	46	-42	184
H18	'Makeup Air, before burner (kNm ³ /h)'	23	-14	42
H4	'Feed Rate Coal (t/h)'	22	-17	47
H8	'Pressure, Mill outlet (mbar)'	22	-17	43
H11	'Nitrogen In, Main Fan outlet (kNm ³ /h)'	21	-16	38
H13	'Nitrogen In, before Baghouse (kNm ³ /h)'	21	-16	43
H15	'Exhaust Fan efficiency (%)'	20	-15	35
H17	'Recycle Gas, Damper, behind Exhaust Fan (%)'	20	-16	37
H6	'Mill Heat Loss (kW)'	20	-15	39
H10	'Main Fan efficiency (%)'	20	-16	36
H20	'Main Flow, cyclone to main fan (kNm ³ /h)'	20	-15	37
H2	'Ambient P (mbar abs)'	20	-8	38
H1	'Ambient Temp (°C)'	18	-12	27
H14	'Baghouse Heat Loss (kW)'	17	-11	27
H3	'Relative Humidity (%)'	16	-10	25
H19	'Burner temperature (°C)'	16	-10	27



It is observed from the SA results with the pareto plat that the remaining water content in the coal when dried has the most dominant influence on the ΔDT . As is expected from theory explained. The second in inline is purge nitrogen in at the mill outlet. The effect of nitrogen on the ΔDT is studied by performing a HEXS-model simulation. If the nitrogen is increased from 0 to being 3 kNm³/h in HEXS-model, it is observed that the ΔDT “the difference between the two temperature” increases. Meaning that it can be potential solution to increase ΔDT if the RMPP system does not dry, but at the cost of using nitrogen. The results seem counter intuitive as the medium temperature is increased by adding nitrogen that has ambient temperature of 5°C. When two mass streams are mixed, one being hot the other being cold, it is expected that the mixture will be somewhere between the hot and the cold temperature. Not being hotter than before, as this graph suggest. The exact cause is unknown, but it has the potential to be a programming logic error, or it might be unknown phenomenon. Therefore, it is advised examine this further.

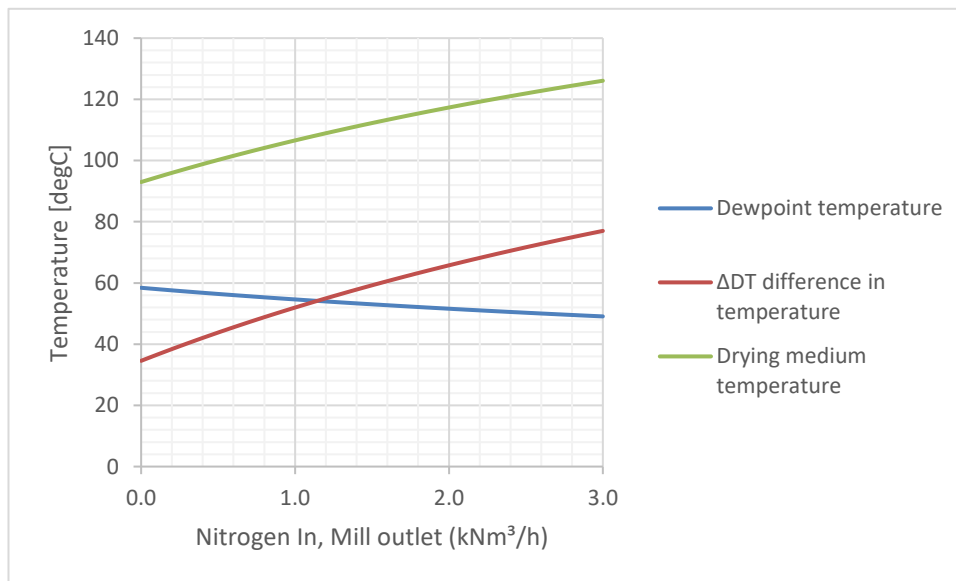


Figure 34 HEXS-model, temperature change at the mill outlet due to varying the nitrogen inlet at the mill.

The SA results for ΔDT are presented as a potential dominant factor that can be utilized to delay or prevent this unwanted functional failure. The four factors with the largest absolute mean are: the % of moisture content after drying the coal, the flowrate of nitrogen at the mil outlet and the differential pressure at the cyclone, the moisture content of the coal as fed.

Bibliography

- [1] “Steel production. *.” [Online]. Available: <https://en.wikipedia.org/wiki/Steel>. [Accessed: 15-Nov-2018].
- [2] Tata Steel Europe, “Energy and climate change | Tata Steel in Europe.” [Online]. Available: <https://www.tatasteeleurope.com/en/sustainability/environment/energy-and-climate-change>. [Accessed: 15-Nov-2018].
- [3] S. Seetharaman, *Fundamentals of metallurgy*. Cambridge: Woodhead Publishing and Maney Publishing, 2005.
- [4] V. Shatokha, “Environmental Sustainability of the Iron and Steel Industry: Towards Reaching the Climate Goals,” 2016.
- [5] J. G. J. . Olivier, G. ; Janssens-Maenhout, and J. A. H. W. . Peters, “Trends in global CO2 emissions - 2014 report,” *PBL Netherlands Environ. Assess. Agency*, pp. 10–13, 2014.
- [6] “Tata Steel: Grotere CO2-reductie mogelijk met nieuwe technologie | NU - Het laatste nieuws het eerst op NU.nl,” *NU.nl*, 2018. [Online]. Available: <https://www.nu.nl/economie/5449871/tata-steel-grotere-co2-reductie-mogelijk-met-nieuwe-technologie.html>. [Accessed: 15-Nov-2018].
- [7] “CO2, Mass Extinction of Species, and Climate Change.” [Online]. Available: <https://www.laprogressive.com/co2-mass-extinction-species-climate-change/>. [Accessed: 20-Nov-2018].
- [8] T. Ming, R. de_Richter, S. Shen, and S. Caillol, “Fighting global warming by greenhouse gas removal: destroying atmospheric nitrous oxide thanks to synergies between two breakthrough technologies,” *Environ. Sci. Pollut. Res.*, vol. 23, no. 7, pp. 6119–6138, 2016.
- [9] L. Cao and K. Caldeira, “Atmospheric CO2 stabilization and ocean acidification,” *Geophys. Res. Lett.*, vol. 35, no. 19, pp. 1–5, 2008.
- [10] R. Warren, J. Price, J. VanDerWal, S. Cornelius, and H. Sohl, “The implications of the United Nations Paris Agreement on climate change for globally significant biodiversity areas,” *Clim. Change*, vol. 147, no. 3–4, pp. 395–409, Mar. 2018.
- [11] J. Hansen *et al.*, “Target atmospheric CO2: Where should humanity aim?,” *Open Atmos. Sci. J.*, vol. 2, pp. 217–231, Apr. 2008.
- [12] N. Bostrom and M. M. Cirkovic, *Global Catastrophic Risks*. Oxford: Oxford University Press, 2011.
- [13] M. Kundak, L. Lazic, and J. Crnko, “Co 2 Emissions in the Steel Industry,” *Metallurgija*, vol. 48, no. 3, pp. 193–197, 2009.
- [14] “What is the Paris Agreement? | UNFCCC.” [Online]. Available: <https://unfccc.int/process-and-meetings/the-paris-agreement/what-is-the-paris-agreement>. [Accessed: 20-Nov-2018].
- [15] N. Lucas and D. R. di Valdalbero, “European Steel - The Wind of Change,” *Publ. Off. Eur. Union*, pp. 1–39, 2018.
- [16] E. van der Oest, “CO2 - CATO - Tata Steel Nederland,” *Tata Steel Nederland*. [Online]. Available: <https://www.co2-cato.org/cato/locations/regions/other/tata-steel-nederland>. [Accessed: 15-Nov-2018].
- [17] “Sustainable in every sense | Tata Steel in Europe.” [Online]. Available: <https://www.tatasteeleurope.com/en/innovation/hisarna/sustainable-in-every-sense>. [Accessed: 15-Nov-2018].
- [18] Ramon Burns, “Hisarna process for ironmaking -,” 2018. [Online]. Available: <https://www.steel-360.com/technology-next/hisarna-process-for-ironmaking>. [Accessed: 15-May-2019].
- [19] “Today towards tomorrow- October 2018 .pdf.” corporate communication and public affairs- Tata Steel IJmuidem, Ijmuiden, 2018.
- [20] Y. A. Cengel, *Thermodynamics : An Engineering Approach INTRODUCTION AND BASIC CONCEPTS*, 8th ed., no. 8th Edition. New York: McGraw-Hill Education, 2015.
- [21] I. C. Kemp, “Fundamentals of Energy Analysis of Dryers,” *Mod. Dry. Technol.*, vol. 4, pp. 1–45, 2014.
- [22] W. Wagner *et al.*, “The IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam,” *J. Eng. Gas Turbines Power*, vol. 122, no. 1, p. 150, 2002.

- [23] "Superheated Steam -Steam Engineering Principles and Heat Transfer." [Online]. Available: <https://beta.spiraxsarco.com/learn-about-steam/steam-engineering-principles-and-heat-transfer/superheated-steam#article-top>. [Accessed: 09-Jul-2019].

