Application of an Unvalidated Process Model to Define Operational Functional Failures

M.Schwarz

Dynamics Based Maintenance, University of Twente, The Netherlands.E-mail: m.m.schwarz@utwente.nl

P. Schepers

Dynamics Based Maintenance, University of Twente, The Netherlands.

J. Van Boggelen

HIsarna Pilot Plant, Tata Steel, The Netherlands.E-mail: johan.van.boggelen@tatasteeleurope.com.nl

R. Loendersloot

Dynamics Based Maintenance, University of Twente, The Netherlands.E-mail: r.loendersloot@utwente.nl

T. Tinga

Dynamics Based Maintenance, University of Twente, The Netherlands.E-mail: t.tinga@utwente.nl

Comprehensive transient models (CTMs) are not readily available for complex industrial processes. In contrast, fundamentals-based process models (FbPMs) often are readily available and data-driven models (DDMs) can be readily developed. Generally, FbPMs have enough accuracy and safety margin to size equipment for steady-state operations but in contrast to CTMs, are not accurate enough to predict the unique operational responses required for applications, such as the definition of system functional failures in predictive maintenance (PdM). However, in the absence of more accurate models, FbPMs may be valid to indicate response trends or determine operational windows, with respect to safety and functionality. The case study is a Raw Material Preparation Plant, used to screen, grind and dry coal for an iron-making process. Following DDM construction through supervised machine learning from operational data, the validity of an available FbPM against operations is investigated through: (1) comparison of FbPM and DDM regression responses (2) consideration of physical phenomena and (3) comparison of sensitivity analysis results. Following validation, the definition and detection of functional failures in the plant as obtained from the FbPM will be used as the first step towards system PdM.

Keywords: Functional Failure, Supervised Machine Learning, Sensitivity Analysis, Predictive Maintenance, Data Driven Model, Fundamentals-Based Process Model, Coal, Morris Method.

1. Introduction

The premise of predictive maintenance (PdM) is the desire to optimize maintenance through the prediction of failures. A failure can be defined as the physical breakage of a component or, more generally, as the inability of the system to maintain functionality.System degradation, which can lead to a functional failure (FF), is a dynamic process that is governed by changes in both the system and its environment. In complex systems and operations, the definition of FF conditions is not trivial. An FF consists of a set of system operating conditions that define a state when the system may not maintain functionality. If an FF occurs, functionality may be restored or reassured through the adjustment of operational settings or the completion of maintenance; ideally, FFs can be predicted and thus, an early step in the pursuit of PdM is defining system FFs.

FFs are defined through physical measurements, modeling process responses, or a combination of theses. Best practices recycle proven designs, resulting in increasingly complicated systems that are beyond the unaided human ability to accurately model the, often transient, operational responses. In contrast to these unavailable and potentially undevelopable comprehensive transient models (CTMs), fundamentals-based process models (FbPMs) are often available and data-driven models (DDMs) can be readily developed.

A process model is a mathematical description of a change that systems undergo from one state to another and the series of states through which a system passes during the process; these are the process and the path of the process, respectively. A process model can be data-driven (e.g. DDM), physics-based (e.g. FbPM), or it

Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference
Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio

ISBN: 978-981-14-8593-0; doi:10.3850/978-981-14-8593-0

can be a hybrid form combining the two concepts. In this work, a DDM refers to a regression model constructed from historical operational data which is combined with statistical tools to estimate relationships among variables to forecast a future event. FbPMs are essentially bottom-up approaches that enable estimation of system process states or future events based on first principles; they are not dependent on historical events to predict a future response. The term FbPM is utilized instead of physics-based process models to discriminate FbPMs from physics-based predictive models within PdM. The latter focus on prediction of the remaining useful life (RUL) through a physics-of-failure approach. Process models, either FbPMs or DDMs, may provide the operational inputs required for physics-based RUL models when monitoring the necessary input parameters is unfeasible.

These generalized FbPMs have enough accuracy and safety margin to size equipment for steady-state operations but in contrast to CTMs, are not accurate enough to predict the unique operational responses required for the definition of FFs for PdM. However, in the absence of more accurate models, FbPMs may be valid to indicate response trends or determine operational windows (OWs), with respect to safety and functionality. DDMs developed from process data may predict unique operational responses but are limited to predicting responses of measurable parameters, typically within limited OWs.

To customize a FbPM for a specific installation, multi-disciplinary expertise is required that is often not readily available. Due to the gap between the required expertise for FbPMs and relative ease of constructing DDMs, DDMs may be preferred over FbPM. This is the incentive for the work presented in this paper, aiming to improve system specific DDMs predictive capabilities through incorporating additional relevant parameters by focused monitoring. However, which parameters are relevant may be identified through FbPM investigations, as FbPMs incorporate unmeasurable parameters but also permit exploration of fundamentally defined OWs, rather than OWs restricted by the training data used to construct the DDM. Therefore to identify the relevant parameters, sensitivity analysis (SA) of the FbPM can be utilized.

As researchers are cautious for drawing conclusions based on a SA of poorly understood processes (Sóbester et al. (2014); Meghoe et al. (2019); Saltelli et al. (2019)), the validity of the FbPM against operations should be investigated prior to utilization of the FbPM on a system. Such a validation will in this paper be accomplished through comparison of model responses with monitoring data, consideration of physical phenomena and comparison of model SA results. The main contribution of this paper is demonstrating the value of a relatively simple and unvalidated process model in detecting and predicting functional failures in process industry plants, namely a Raw Material Preparation Plant (RMPP) which functions to screen, grind and dry coal for an iron-making process.

2. Case Study

The analyzed RMPP is part of the HIsarna Pilot Plant (HIsarna) at Tata Steel in IJmuiden, The Netherlands. The HIsarna RMPP process is depicted in Figure 1 with the sub-processes numbered corresponding to the descriptions below. The RMPP accomplishes its functions by removing undesirable materials from run-of-mine (ROM) materials (e.g. raw coal) by means of a separation process. The complex process begins at the atmospheric raw coal storage (1). ROM materials are manually transported to temporary storage and subsequently, mechanically and magnetically screened (2). The screened coal is then ground within the impact dryer mill (3). After grinding, as it is transported through the drying column (4) towards the spinner separator (5) and into the cyclone (6), it is concurrently being dried by the preheated air/fuel gas mixture. At the cyclone it continues into one of three process paths depending on the particle size; propelled by the main fan (7), this composes the main drying gas circulation loop, depicted in red (3-7). Exiting the cyclone, the coal either returns to the mill (3), collects in the baghouse (8) as it flows through the drying gas recirculation loop (8-10) or is pneumatically conveyed for temporary silo storage (13) before utilization in the HIsarna iron-making process. Prior to storage, further drying/moisture absorption can occur during transportation. Process air at the baghouse outlet either flows into the stack as vent gas (9) or is reincorporated as recycled gas (10) propelled by the combustion blower (11) to the air heater (12) before re-entering the main drying gas circulation loop at an increased temperature. At the baghouse exit, process air flows are controllable through two valves: the vent gas damper and the recycled gas damper, at (9) and (10), respectively. Further details on the process are found elsewhere (Schepers (2019)).

HIsarna operations are research focused to develop a novel iron-making process that can improve steel production sustainability performance, Meijer et al. (2019); this results in continuous improvement through adjustment of operating procedures and asset modifications. Production at the RMPP began with logged data in early 2018. However, due to continuous improvement initiatives, there are limited periods of extended operation and the data is inconsistent. Thus, the process is considered immature. There are interacting process paths that consist of multiple subsystems, as depicted in Figure 2. The interactive complexity of these paths and experience indicate

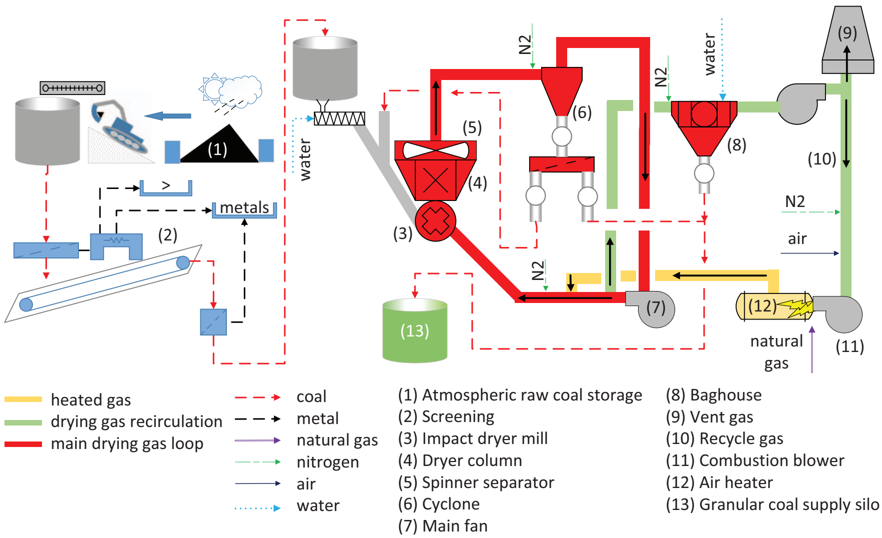

Fig. 1. Schematic of the HIsarna Raw Material Preparation Plant process.

dependencies exist. The system is dynamic due to the statistical nature of all grinding operations; the dominant sources are the inherent randomness in the ROM material composition and grinding response, system degradations and ongoing modifications. This results in a non-linear system where the inputs and outputs, or the cause and effects, are not obvious or directly related. In conclusion, the HIsarna RMPP is an immature multilevel coupled complex system that is characterized by interactive, dynamic, and non-linear complexities.

As iron-making process stability and product quality depend on the processed coal meeting specifications, reliable RMPP functionality is operationally critical. Considering the complex interactions and operational philosophy, developing a CTM to investigate FFs is impractical for this evolving system. Alternatively, FFs can be investigated through FbPMs or DDMs that relate operational conditions to the maintenance driver of functionality degradation.

In absence of direct relationships between maintenance and functionality, established indirect relationships are considered. Systemic wear is a leading mechanism in RMPP functionality deterioration and a maintenance driver. The parameters that influence systemic wear also influence the milling efficiency, Uuemõis et al. (1996). As milling efficiency relates the system's ability to meet size and drying specifications, the relationship between maintenance and operational functionality is established. The parameters expected to dominate system degradation are the inputs and outputs of the RMPP interacting process paths. Examples of these parameters are temperature, pressure, mass flow rates, and composition of the ROM materials, processed materials, and gas-mixtures (Uuemõis et al. (1996)). In this case study, the inability to dry the wet raw coal to specification is a FF state.

3. Functional Failure - Drying

In this work, *drying* refers to the removal of liquid surface water from coal by vaporization due to convection. The rate of drying is influenced by the properties of the coal and the drying medium, e.g. moisture content, structure, composition, temperature, relative humidity and the velocity of the drying medium. The capacity of air to remove moisture is primary dependent upon its initial temperature and humidity; that is, higher temperatures and lower humidity levels increase the air's moisture removal capacity.

Process gas temperatures are monitored at 5 locations at the HIsarna RMPP. The mill outlet temperature (MOT), measured between (5) and (6) in Figure 2, is the primary continuous operational indicator that enough heated drying gas is present for the evaporation of the moisture. As coal is hygroscopic and can absorb moisture from its surroundings, it is important to maintain the process temperature above the system dew point temperature (T_{DEW}); this ensures drying and prevents moisture reabsorption. The RMPP is intended to operate above, but as near, the system dew point as practicable. Fundamentally, the failure condition occurs when the MOT is less than T_{DEW}; at this point moisture condenses on surfaces below T_{DEW} and the evaporated liquid can re-infiltrate the coal. Thus, it is critical that the mill and subsequent systems operate above T_{DEW}. Monitoring T_{DEW}, and all other associated inputs and outputs, directly is impractical due to the harsh operating conditions. Therefore, process models are beneficial. A FbPM for a RMPP and the DDMs developed from the RMPP operational data are presented in Section 4 and Section 5, respectively.

4. Unvalidated Fundamentals-based Process Model

The HEXS model is an in-house simulation tool that is a steady-state heat and mass transfer process model of RMPP operations. It is a FbPM for RMPP drying that estimates process parameters values related to a specific degree of drying of wet coal by simulating RMPP process conditions. The model integrates physics-based fundamentals and statistical empirical relationships in a spreadsheet environment utilizing automated iterative calculations that incorporate look-up tables. Other RMPP functions, such as the screening and grinding of coal, are not included in the HEXS process model. However, as these mechanisms affect the drying process, the steady-state HEXS model cannot be considered a CTM. It is also not considered a DDM or a hybrid process model, as the statistical relationships utilized are from standardized industrial processes and not derived directly from operational data, e.g. Wagner et al. (2000).

The HIsarna RMPP is constructed from a modified design of a RMPP installation that previously operated in Kwinana, Australia as part of Rio Tinto's HIsmelt Process Plant (HIsmelt). The HIsarna RMPP incorporates many design improvements but also utilizes many of the same original equipment manufacturers as those at the Kwinana RMPP. Consequently, the HIsarna RMPP design is not the same as the Kwinana RMPP, but the two plants have many comparable and/or representative qualities including the general configuration of the process paths as depicted in Figure 2. This situation is common in manufacturing and may be at the level of redundant equipment within a single facility or entire plants, as in the case study.

When Tata Steel acquired the patents from Rio Tinto, they also received the thermodynamic HEXS model without any model specific documentation. Limited access to the HEXS model developer and former HIsmelt staff indicate that the available version of the HEXS model is satisfactorily tuned to operational data from the HIsmelt RMPP. That is, the HEXS model reasonably represents HIsmelt RMPP operations. The developer also indicated that the model is tunable to HIsarna RMPP operations, or any equivalent plant, by adjusting selected parameters. As opposed to DDMs constructed from operational data, the HEXS FbPM can be dimensioned to the modeled RMPP. Attempts by HIsarna staff to tune the HEXS model with HIsarna RMPP process data were unsatisfactory. Specifically, when HIsarna RMPP data is used as the input, the HEXS model is unable to return results that adequately demonstrate agreement with expected responses, despite tuning attempts. The inability to tune could mean the model is (1) not valid for HIsarna RMPP due to design and operational modifications or (2) valid but untunable due to a lack of specific tuning expertise. Moreover, the plant is operating in a transient manner, while the model is steady-state and thus, latency in sub-system responses is not modeled. While this does not necessarily invalidate the HEXS model to operation, it may indicate additional data cleaning is required to remove transient periods prior to comparison. The model may be valid as not all parameters are physically monitored, and estimated values are used in the HEXS model as input parameters. The parameters that are monitored do not have the same units and are not monitored at the same locations. With questions of validity, data quality, and conditional model mismatch, when comparing similar inputs and outputs, a one-to-one fit of the HEXS model to HIsarna RMPP data would be more coincidental, rather than expected. Despite this, validation can be investigated using a relative method.

Validation in a relative manner requires identifying commonalities between the HEXS model and the HIsarna RMPP. The MOT is continuously monitored at the HIsarna RMPP, simulated in the HEXS model and relates to the potential FF associated with inadequate coal drying. Considering the MOT as the model output allows comparison of plant responses to FbPM predictions through a relative comparison as presented in Section 6 and 7. Furthermore, the difference between the MOT (drying medium) and T_{DEW} is considered with respect to FFs in Section 8.

The direct comparison of input parameters from the HIsarna RMPP with those of HEXS is challenging, as some modalities do not have the same units and/or comparable ranges, but represent the same physical quality. In the HEXS model, drying ability is ensured when the medium drying temperature is above the T_{DEW} at four locations throughout the modeled RMPP. In the HIsarna RMPP, there are only two comparable locations where applicable temperatures are monitored, and the limits of acceptable/explorable OWs differ by as much as 250 ° C. In HEXS, influences due to transitions from one sub-system to another are only accounted for with efficiency parameters. In practice, geometrical changes caused by pipes, transitions, contact boundaries, and locations of sensors influence the behavior being monitored. Within the HEXS model, there are over 300 variables that can be tuned or applied as input; a comparison of these variables with those that are controllable or can be monitored at the HIsarna RMPP, identified 20 input factors. The approach may result in neglecting the truly dominant parameters but if they cannot be controlled or monitored, practically, they are process noise to the controllable parameter and those that can be monitored. As the model uses iterations to converge to a the given output, prior to determining the model outputs presented throughout this case study, the remainder of the over 300 variables are reinitialized each time.

5. Data-Driven Model

The DDM is constructed by means of supervised machine learning (SML) and trained with historical HIsarna RMPP operating data. The DDM construction and the validation activities are similar to those recommended by Kotsiantis (2007) including data identification, pre-processing, training set definition, algorithm selection, and evaluation against a test set. Forty-four relevant inputs are available for the SML that potentially relate the MOT, T_{DEW} and operational settings. The modalities considered include temperatures, pressures, flow rates, masses, concentrations, valve positions, torque, and power. The data set spans approximately 18 days and is retrieved in a cyclic form with a 60 second resolution interval resulting in a total of 25,922 observations per factor. The representative accuracy of data is not significantly affected by the selected retrieval method and interval while accounting for available computing resources; this is based on investigations comparing 1, 30, 60, 120 and 300 second intervals retrieved either in cyclic or average mode. The representative accuracy improved at lower retrieval intervals (higher data resolution) but not to an appreciable level that justifies the computational expense.

Minor data pre-processing was required to account for data gaps. Approximately 0.027% of the data is missing and substituted by linearly interpolated data based on previous and consecutive data points. The 70/30 approach is utilized for training data definition; 70% of the data is utilized as the training set and 30% of the data is the test set to check the validity of the model predictions.

To identify a suitable regression algorithm for the DDM, in random order, nineteen regression algorithms are trained by k-fold cross-validation (CV) SML and evaluated on their root mean square error (CV-RMSE). A fivefold cross-validation is selected and the CV-RMSE is computed by the MATLAB 2018b regression learner tool. The predictive performance of the models are assessed by validating against the historically recorded MOT response. This best performing DDM is defined by the lowest CV-RMSE of the algorithms. Gaussian Process Regression Models (GPR) perform better than Support Vector Machines (SVM), Linear Regression (LR), Regression Trees and Ensembles of Trees. The GPR Rational Quadratic algorithm is selected as the best performing algorithm and the goodness of fit (GoF) is checked against the test data set.

The GoF between the selected best performing model and the measured MOT is evaluated by the absolute error (difference) between the predicted (y_{fit}) and actual (y_{real}) values. The y_{fit} are forecasted using the GPR Rational Quadratic algorithm and trained with 70% of the data set containing 18,144 observations. The GoF is good within the training data set, except for a single outlier of approximately 45 ° C or 38.6% of the predictive range. Considering the model predictive capabilities within the 30% test data set, a sudden increase in deviations occurs at the transition from the trained 70% to the untrained 30%. Within the test data set, the RMSE is 4.41 ° C (3.8%) and the maximum absolute error associated with a single outlier is approximately 21 ° C, 17.5% of the entire MOT operating domain. The errors are reasonable for the processes considered.

Retraining the model with 100% of the data set (25,922 observations) reduces the RMSE to 0.52 ° C, or an error of approximately 0.4% over the MOT range. The frequency and extremes of the outliers are also reduced; the maximum outlier is approximately 18 ° C (15.4%) compared to approximately 45 ° C (38.6%) in the 70% trained model. Therefore, as expected, the 100% trained model appears to have increased predictive power as compared to the partially trained DDM. Additionally, as an RMSE of 0.4% is relatively small compared with its operating domain, the 100%-trained DDM is expected to simulate the representative behavior of the physical HIsarna RMPP MOT, as long as the RMPP is within the operating range of its training data set. The minimum and maximum values utilized in the DDM analysis are derived from the HIsarna RMPP data set; however, it is found that utilizing data within 80% of the monitored range improved results. The prediction accuracy of the DDM increases when data within the upper and lower 10% are neglected, as within the dataset, observations in the monitored upper and lower bounds are less frequent. Nevertheless, the DDM trained by 100% of the data is used in subsequent FbPM validation.

6. Linear Regression and Physical Phenomena

As part of the validation, Linear Regression (LR) of selected response factors can be utilized to roughly correlate the DDM and HEXS FbPM. By comparing simplified LRs fit to model responses, correlations can be considered. LR is in the form of $y = ax + b$, where a is the slope, x is the observed value, b is the y-intercept representing the model determined output. This method evaluates the overall global gradient of the specified domain and indicates the direction of the trend line. Based on the expected mismatch of outputs described in Section 4, only the sign of the slope is considered to compare the two model responses: positive (P), negative (N), or neutral (O). The responses are considered in agreement if the slopes of the LRs are $P-P$, $N-N$, or $O-O$; however, results must be cautiously considered as agreement may occur by chance, e.g. one of the compared slopes is parabolic.

From the 20 changeable input factors in the HEXS model and 44 factors in the DDM, seven common factors are identified that are monitored/modeled in a similar location, and have equal physical quantities and units. An Anderson-Darling test, Anderson and Darling (1952), on the time series data demonstrates the individual operational data factors are not normally distributed. Therefore, to compare the LR response of each model, these seven individual factors are varied between the minimum value and maximum value in 100 equal steps, and the resultant model response is fit with a LR trendline. The resultant slopes are shown in Table 1. The same minimum and maximum values, based on operational data, are utilized in the regressions for both models.

Table 1. Slopes of linear regressions

Factor	DDM	HEXS	Compare
1	0.237	0.624	P-P
2	-0.185	0	N-P
3	-0.043	0.001	N-O
4	0.135	0	P-O
5	-0.242	-0.852	N-N
6	0.014	1.34	P-P
7	0.120	0.405	P-P

Table 1 shows that four of the seven slopes are in agreement. More detailed analysis of Table 1 and the responses determines that the validity of the HEXS model to represent the HIsarna RMPP as determined by LR is non-conclusive. The factors that have the greatest effect on the MOT are intelligible, based on knowledge of the physical phenomena within the system, as they relate to heat input (burner temperature) and heat extractions (amount of recycled gases as indicated by two valve positions). For both the DDM and HEXS model, when the burner temperature (Factor 7) increased, the MOT increases while all other input factors are fixed. This is expected because the burner is the main heat source in the RMPP. When the vent gas damper position (Factor 5) is opened, less process gas is recycled for heat, and thus it is expected that the valve is opened further, the MOT decreases. When the recycle gas damper (Factor 6) is opened, the already heated process gas is reused and further opening should result in the MOT increasing. These trends are confirmed for both valves (factors 5 and 6); however, varying the position of the recycle damper position has a minimal effect on the MOT within the given domain.

For the three factors not in agreement, Factors 2, 3 and 4, the slopes of the HEXS model LR are approximately zero, particularly when compared to the other factors. The disagreement of these three factors does not support, nor disprove the HEXS model validity, but may indicate inclusion of these parameters is unnecessary in the DDM. It also indicates additional factors in the HEXS model may be dominating factors and the influence is not identified by limiting the investigation to seven corresponding factors. These findings are supported by detailed response investigations (not presented). As HEXS model validity is non-conclusive following examination of the LR and physical phenomena, validation attempts continue through comparing sensitivity analyses for both models and further examination of physical phenomena in Section 7.

7. Sensitivity Analysis

Three aspects of SA are considered for validation to examine the relative contribution of inputs: (1) SA of 20 HEXS model inputs (2) SA of 44 DDM inputs, and (3) comparison of SA results for the seven common factors (as identified in Section 6). Analysis of similarities and differences, in the ranking and magnitude of the relative contribution, of common inputs by each model, provides an indication of validity and allows further consideration of physical phenomena. Based on case study characteristics, elementary effects (EE) are selected for the SA, Morris (1991). The Morris method uses a randomized one-factor-at-a-time design and data analysis is based on the resulting EE that allows isolation of the changes in an output that are solely due to a particular input, Morris (1991). The Morris method can be used to indirectly examine the relationship between input and response factors, Meghoe et al. (2019). Details on the method are covered elsewhere, Morris (1991); Campolongo et al. (2007); Sóbester et al. (2014); Saltelli et al. (2019); Meghoe et al. (2019); Schepers (2019).

For the HEXS model, the number of runs $r(k+1)$ is equal to 1050 with $r=50$ and $p=32$; it is 9000 for the DDM with $r=200$ and $p=32$, where the design space is a k-dimensional p-grid, k is the number of inputs, p is the number of levels for each input and r is the number of trajectories constructed. The estimated measures considered are the absolute mean, μ^*, the mean, μ, and the standard deviation, σ of the EEs. For simplicity, equally spaced steps are utilized because each parameter has a different distribution and some of the distributions are unknown. The levels are varied between the maximum and minimum values, based either on operational data, or HEXS model constraints. Since the intent is to compare the results, the ranges for the seven comparable factors identified in Table 1, are equal and based on operational settings.

As the EE sensitivity measure is a relatively qualitative global SA, it is difficult to rank the relative importance of each factor. Therefore, to compare results from the DDM and HEXS model; therefore, to compare, the results require normalization, which is done in two ways:

$$x^*_{\text{prop}} = x_i \left(\sum_{i=1}^{k} x_i \right)^{-1} \cdot 100\% \qquad (1)$$

$$x_{\text{relative}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \cdot 100\%q \qquad (2)$$

Eq. 1 considers the proportion of the contribution, x^*_{prop}, relative to the total contribution of all compared factors for a model, where x is either the absolute mean or the standard deviation. Eq. 2 accounts for the order of magnitude mismatch between models by determining the relative contribution within the considered range of contributions. These comparisons require corresponding OWs. The relative absolute mean, μ^*, and the corresponding normalizations for the DDM and HEXS model are presented in Table 2. The comparison by standard deviation is also possible but as factor interaction is not discussed, results are omitted. The resultant screening plot for the normalized μ^*_{relative} and $\sigma^*_{\text{relative}}$ for the DDM and HEXS models is shown in Figure 7.

Table 2. DDM and HEXS: relative absolute mean

Factor	DDM			HEXS		
	μ^*	μ^*_{prop}	μ^*_{relative}	μ^*	μ^*_{prop}	μ^*_{relative}
1	1.543	13	18	1599	13	0
2	0.652	5	2	2145	3	1
3	0.557	5	0	2541	3	2
4	0.566	5	0	5268	7	7
5	1.773	15	22	2887	4	3
6	0.744	5	3	8840	12	14
7	6.049	51	100	52518	69	100

Figure 7 shows the relative contribution for each model is not equivalent by factor, e.g. D5 $\neq$ H5, unless an artifact of figure construction, e.g. D7 = H7. For each factor 7, the D and H correspond to the DDM and HEXS model, respectively. Considering Table 2, the greatest agreement is in Factor 7; it demonstrates a dominant contribution in both model responses, with 69% and 51% for the HEXS model and DDM, respectively. Of the 20 HEXS model inputs, factor 7 accounts for 28% of the total contribution. Of the 44 DDM model inputs, factor 7 accounts for 5.8% of the total contribution. As the other factors do not contribute significantly to the overall response, particularly to the DDM, comparison is potentially meaningless when considering contributions. However, for the demonstration of validation procedures, comparison is continued. In Figure 7, only factor 7 has a direct correlation for comparable ranking due to diagram construction; however, the values and the ordering of the other factors are generally consistent when considering the physical phenomena. Investigations revealed, the mismatch in the magnitude of contribution of Factor 1 between the models is due to an artificial correlation as a consequence of the limited OW of the DDM. This artifact is also a viable explanation of the mismatch in total contribution of factor 7. Within the DDM, the SA indicates that rainfall and external temperatures dominate responses. The correlation between ambient temperatures is due to corresponding rainfall rather than a fully representative response; however, there is an operational conclusion to the analysis, that improving the atmospheric storage of the ROM is beneficial to process control and consistently meeting product specification. That is, a more consistent initial condition of the ROM moisture results in a more consistent drying process. While not a profound interpretation of physical phenomena, this conclusion is now fully supported by the data analysis and a solid foundation for the business case of directed system improvements.

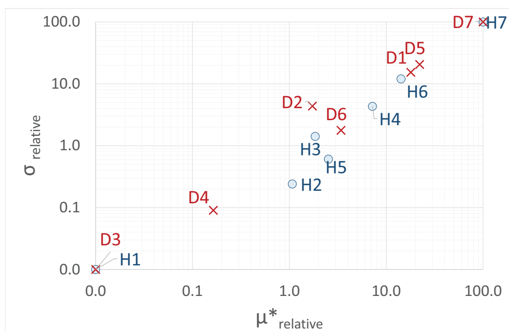

Fig. 2. Screening plot relative absolute mean and relative standard deviation for seven comparable factors

8. From Validation to Functional Failures

The premise of the investigations is, should the FbPM be found to reasonably reflect the real-world process of the HIsarna RMPP, that it can provide insights into the parameters that influence the actual system and sub-systems of the HIsarna RMPP, such as defining FFs. During validation, agreements and disagreements between the HEXS model and the DDM are identified. The agreements are inline with expectations based on physical phenomena and the disagreements are explainable considering the limited OW of the process data and model construction; thus, indicating the validity of the HEXS model to reasonably represent conditions. As there are disagreements, no evident conclusion can be made that the FbPM model and DDM are in agreement for all OWs. However, based on FbPM historical development and usage, the level of confidence achieved is acceptable; the approach provides a viable methodology that accommodates available resources to investigate FFs.

As stated in Section 3, FFs can be related to operation through the T_{DEW}. To utilize the models to investigate FFs, a SA can be repeated on the HEXS model with the output set to the difference in temperature, ΔT, between the drying medium temperature and T_{DEW}. Across the HEXS defined OWs, conditions when $\Delta T \leq 0$ are then determined for each input factor. This identifies the set of conditions that can lead to the potential FF. The most influential factors are prioritized due to the use of SA. Of the most influential, the controllable factors are candidates for operational adjustments to prevent or respond to FFs.

If FbPM validation confirms representation of operational behavior, automated controls and alarms for FFs can be directly incorporated into transient operations. When a generalized representation is established and there is not a one-to-one representation, as in the case study, the results can be used either with additional validation or to improving data-driven approaches through targeted monitoring. These solutions are a compromise that accommodates available resources and permits continued caution regarding SA conclusions. Either through non-resource intensive spot measurements or highly targeted sensor installations, implementation of either strategy is an economical solution to explore and improve operations while pursuing PdM.

9. Conclusion

The work demonstrates how an unvalidated FbPM can be utilized to define the dominant influencing factors and functional failure conditions in a complex system. The FbPM can be validated through comparison with a DDM constructed from process data, consideration of physical phenomena, and sensitivity analysis. It can assist predictive maintenance pursuits by defining the potential failure conditions in larger operating windows than DDMs permit. The work demonstrates how tools and information typically available in an industrial settings can be developed into actionable information for operational improvements. It presents an economical solution to several commonly encountered problems, rather than an ideal solution, such as comprehensive transient models may permit.

Acknowledgement

Executed within the ‘Smart Industries’ research program, project number 15467, financed by the Netherlands Organisation for Scientific Research (NWO), in partnership with Tata Steel, Semiotic Labs, IJssel Technologie, and M2i.

References

Anderson, T. W. and D. A. Darling (1952). Asymptotic Theory of Certain ”Goodness of Fit” Criteria Based on Stochastic Processes. *Ann. Math. Statist. 23*(2), 193–212.

Campolongo, F., J. Cariboni, and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software 22*(10), 1509–1518.

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica (Slovenia) 31*(3), 249–268.

Meghoe, A., R. Loendersloot, and T. Tinga (2019, 6). Rail wear and remaining life prediction using meta-models. *International Journal of Rail Transportation 00*(00), 1–26.

Meijer, K., C. Zeilstra, H. Hage, P. Broersen, and J. van Boggelen (2019). Various roads to CO2 reduction with HIsarna technology: The HIsarna process. METEC & 4th ESTAD 2019, European Steel Technology and Application Days: Düsseldorf, Germany.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics 33*(2), 161–174.

Saltelli, A., K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu (2019, 4). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software 114*(January), 29–39.

Schepers, P. (2019). Application of a process model to define potential operational functional failure conditions. Master’s thesis, University of Twente.

Sóbester, A., A. I. J. Forrester, D. J. J. Toal, E. Tresidder, and S. Tucker (2014, 3). Engineering design applications of surrogate-assisted optimization techniques. *Optimization and Engineering 15*(1), 243–265.

Uuemõis, H., H. Kangur, and I. Veerus (1996). Wear in high-speed impact mills. *International Journal of Mineral Processing 44-45*(SPEC. ISS.), 301–313.

Wagner, W., J. R. Cooper, A. Dittmann, J. Kijima, H.-J. Kretzschmar, A. Kruse, R. Mares˘, K. Oguchi, H. Sato, I. Sto¨cker, O. S˘ifner, Y. Takaishi, I. Tanishita, J. Tru¨benbach, and T. Willkommen (2000, 1). The IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam . *Journal of Engineering for Gas Turbines and Power 122*(1), 150–184.