# Data Analytics Methodology for Construction of Prognostic Indicators: A Use case of fouling in heat exchangers

Moncef Soualhi [1], Ahmed Ragab [2], Khanh T. P. Nguyen [1], Kamal Medjaher [1], Hakim Ghezzaz [2], Mouloud Amazouz [2]

[1] *Toulouse University, INPT-ENIT, 47 Av. d'Azereix, 65000 Tarbes, France.*
*moncef.soualhi; thi-phuong-khanh.nguyen; kamal.medjaher@enit.fr*

[2] *CanmetENERGY, Natural Resources Canada,1615 Boulevard Lionel-Boulet, QC J3X 1P7, Varennes, Canada.*
*ahmed.ragab; hakim.ghezzaz; mouloud.amazouz@canada.ca*

## ABSTRACT

Data-driven prognostics and health management play an important role in the predictive maintenance of complex processes. They offer promising perspectives for achieving high reliability, availability, maintainability and production safety. However, in complex process industries, due to the high level of system interconnectivity, sensor faults and heterogeneous data sources, the construction of representative prognostic indicators remains one of the major challenges. Therefore, special procedures for preparing the system's historical data are needed to construct such indicators, which are then used by machine learning and deep learning algorithms to build accurate dynamic prognostic models. These models can help achieve cost-effective predictive maintenance. For this purpose, this paper presents a data analytics methodology that automatically constructs reliable prognostic indicators representing system degradation. The methodology consists of data preprocessing and processing algorithms that extract, isolate and select the appropriate indicators that reveal the system health state from nominal to a specified critical level. The performance of the proposed methodology is investigated using real data collected from complex equipment of a pulp and paper mill located in Canada. In this use case, the system degradation is an evolution of fouling in a black liquor heat exchanger.

## 1. INTRODUCTION

In the last decade, data-driven prognostics and health management (PHM) of industrial systems have shown significant advances for building predictive maintenance strategies and offer promising perspectives for high reliability, availability, maintainability and production safety. Indeed, prognostics of system failure provides more convenient support for decision-making and more specifically for scheduling maintenance activities to reduce downtime and repair costs. One of the key factors to build a robust prognostic model is the deployment of efficient data analytic tools. In general, prognostics is performed through two phases. The first one is an offline phase which consists of collecting and analyzing raw data by preprocessing and preparation algorithms to build the prognostic model. Then, in the second phase (online phase), the constructed model is used to evaluate the future system conditions using a new set of observations collected from the monitored system. Hence, for the construction of accurate prognostic models, it is necessary to ensure that the available database is fully exploitable and that the pertinent processing methods are efficiently performed. This involves numerous challenges in data preprocessing and processing for the construction of effective health indicators for diagnostic and prognostic purposes. These challenges are highlighted below.

1. **Data preprocessing:** In general, condition monitoring of industrial systems starts with analyzing their architectures to identify the critical parts susceptible to failure and then sensor placement to collect monitoring data for the assessment. However, in most case studies, the collected data present the first obstacle for monitoring. It can be data contaminated with different sources of noises, missing values, redundant observations that lead to false alarms and poor quality results. There are several causes for these previously mentioned situations such as sensor failure, sensor overload, sensor aging, external impacts caused by the operator, etc. To cope with these situations, there exist techniques in literature for transforming the raw data that is contaminated with errors into exploitable information for the next steps of the monitoring process. For instance, the work presented in (Ahmed, Pai, Sriram, & Bhat, 2018) dealt with handling outliers from vibration signal of gears using several filtering techniques, it used

Empirical Mode Decomposition (EMD), Neighcoeff Coefficient (NC), Principle Component Analysis (PCA) and Wavelet Denoising (WD) techniques to remove the signal noises. In (Ratolojanahary et al., 2019), the authors used the Multivariate Imputations by Chained Equations (MICE) technique for imputing missing values of water quality dataset while the work in (He & Wang, 2007) used the Dynamic Time Warping (DTW) with low band-pass filter technique for imputing missing values of semiconductor component data. In the context of face recognition, the work presented in (Zhang, Sconyers, Orchard, Patrick, & Vachtsevanos, 2010) used non-linear interpolation of image time series data to reconstruct images with missing pixels.

Based on this synthesis, is noticeable that multiple data preprocessing techniques exist and each of them is application dependent (Sukumar, Robert, & Yuvaraj, 2016). They represent a prerequisite step for each application to provide clean data that enhance the efficiency of the monitoring procedures such as health indicators construction for fault detection, diagnostics and prognostics. However, with the variability of applications and their acquisition systems, the preprocessing data layer is always presenting new challenges and issues for researchers and practitioners.

2. **Data processing for health indicators construction:** This step of condition monitoring is one of the most important tasks for an efficient health assessment of an industrial system. In fact, fault detection, diagnostics and failure prognostics not only depend on data preprocessing but also depend on the extracted information from the raw data in case of no errors, and depend on the pre-processed raw data if it is contaminated with errors. The extracted information represents a health indicator that reveals the system's health state (whether healthy or faulty).. This system's health state groups two types of indicators: diagnostics and prognostics. The diagnostic indicator is a function that creates separated patterns indicating a specified condition of the system (healthy, fault type 1, fault type 2, etc.). One can cite the published works (Lamraoui, El Badaoui, & Guillet, 2015; Rai & Mohanty, 2007; Sait & Sharaf-Eldeen, 2011) where the authors extracted time, frequency and time-frequency features to build health indicators to detect and diagnose machining tool wear, bearing, gear and Li-ion battery faults, respectively, while the authors in (Zarei, Tajeddini, & Karimi, 2014) exploited multidimensional space features with a fusion layer to construct the diagnostic indicators to monitor bearings. The work proposed in (Oh et al., 2020) used a residual based features technique on heat transfer data to build a heat exchanger fault indicator of pulp mill manufacturing. Another work for characterizing and diagnosing faults of pulp mill plants is presented in (Guelpa & Verda, 2020) to exploit the mass flow rate and the temperature of both sides of the heat exchanger to extract multiple time features and fused them to build the health indicator. Besides, the prognostic indicator is a function that represents the system health state from nominal (healthy) to critical state (failure). The work in (Atamuradov, Medjaher, Camci, Dersin, & Zerhouni, 2018) uses the time domain features with an adaptive fusion algorithm to build a prognostics health indicator for railway systems. In (Saidi, Ali, Bechhoefer, & Benbouzid, 2017), the authors used spectral kurtosis to represent bearing degradation evolution while in (A. Soualhi, Medjaher, & Zerhouni, 2014) both time and frequency features were used to construct the bearing degradation indicator. An other research is proposed in (Nguyen & Medjaher, 2020) used combination of mathematics equations to build the health indicators combined with a genetic algorithm for automatic selection of the most representative ones. Also, there are many research works that exploit artificial intelligence techniques such as machine learning algorithms to build prognostic indicators. One can cite (Sun, Lee, & Lu, 2016) that used a neuro-fuzzy system for fusing the sensor measurements acquired from two different types of sensors (velocity, GPS) to build bridges health indicators. A fusion approach used to construct the health indicator is pointed out in (Ng & Srinivasan, 2010). The authors in (Gebraeel, Lawley, Liu, & Parmeshwaran, 2004) proposed to use an Artificial Neural Network (ANN) for the construction of bearings indicators. Another technique that combines features extraction and machine learning is proposed in (Ragab, Yacout, Ouali, & Osman, 2019) here the authors used a machine learning technique called the logical analysis of data (LAD) and extracted the time-frequency features to build a prognostic indicator using the vibration signals collected from the bearings that are subjected to multiple failure modes.

   Considering the above studies, one can see that the effectiveness of the constructed health indicators in both diagnostics and prognostics depends on the processing algorithms and techniques that can handle the different monitoring data types to provide efficient results for the modeling step.

In the light of the data preprocessing and processing challenges in PHM domain, this paper aims to address practical issues presented in pulp and paper manufacturing. We consider a black-liquor evaporator (heat exchanger), a critical system in the Kraft pulp mills (Steinhagen, Müller-Steinhagen, & Maani, 1993). The main role of this system is to increase the concentration of the black-liquor by evaporating the water through a high level of heat transfer. Concentration of black liquor is needed to improve the combustion in the Kraft recovery boiler (Naqvi, Yan, & Dahlquist, 2010) and convert the generated vapor from the boiler into other forms of energy such as electrical supply source (Pettersson & Harvey, 2010). In practice, the rate of heat transfer is significantly affected by the fouling phenomena, which exists in the heat exchanger tubes and increases over time (Ardsomang, Hines, & Upadhyaya, 2013; Markowski, Trafczynski, & Urbaniec, 2013) and consequently impacts solids concentration of the treated black liquor. As a solution to slow down the fouling evolution, operators try to better control operating parameters affecting the fouling such as used steam temperature and solids concentration of the treated black liquor. However, when the fouling level reaches a certain value, the mill shutdown the black liquor concentration system and do the cleaning. Therefore, an accurate predictive monitoring algorithm is extremely needed to effectively forecast the fouling and to properly plan the shutdowns for cleaning. This is to better manage the black liquor inventory and to avoid the shutdown of the entire mill. For this purpose, one needs to use the historical data of fouling evolution, from the beginning to its critical level to build an effective predictive model for fouling prognostics.

To our humble knowledge, there are fewer research works that addressed the data preprocessing and processing of the pulp and mill industry for prognostics. The first study in this context was conducted by the University of Tennessee in the USA through a simplified heat exchanger test bench (Ardsomang et al., 2013) and that study did not perform preprocessing for the raw data collected. In this paper, a new challenge in addition to missing values and noises in raw data is pointed out. This challenge is attributed to the fact that the historical observations collected along a period of time (months) that represent multiple fouling evolution scenarios have no information on the maintenance actions allowing us to isolate each trajectory of fouling evolution and build a predictive model for prognostics.

In this regard, this study proposes a data analytics procedure for raw data preprocessing and processing to construct efficient fouling prognostics indicators. The main contributions of this work are summarized as follows:

- Combination of data fusion techniques with non-linear interpolation for handling signal noises and missing values.
- Proposition of an efficient algorithm to automatically isolate and separate prognostic indicators.

The proposed methodology is presented and structured as follows. Section 2 introduces the global methodology for data preprocessing and processing to build the fouling indicators. Then, Section 3, highlights the performance of the proposed methodology through a real data carried out from a pulp and mill plant. Finally, the conclusion and perspectives will be presented in Section 4.

## 2. Proposed methodology: Industrial raw data preprocessing and processing

This section aims to present the overall steps of the proposed data analytics methodology for preprocessing and processing raw data towards the construction of prognostics indicators in the process industry. The steps are illustrated through an application to fouling prognostics of heat exchangers in the pulp and paper mills. In detail, this methodology starts with denoising and imputing missing values by a combination of a data fusion method, an auto-encoder (AE) network, and a non-linear interpolation technique. The next step consists of exploiting the physical characteristics of the targeted systems to construct the fouling prognostics indicator using the available historical data. The fouling indicator construction is a challenging task in the pulp and paper industry due to the missing information about the maintenance intervention time to isolate its different evolution trends. Therefore, an algorithm is proposed in this work to automatically separate these trajectories and provide exploitable information for training a prognostic model. The overall flow chart of the proposed methodology is shown in Figure 1.

### 2.1. Handling noises and imputing missing values

The first step of the methodology is based on preprocessing the raw data to impute the missing values. This step is divided into two tasks: dimensionality reduction and interpolation.

The dimensionality reduction is ensured by an AE network to denoise and fuse the raw data to remove outliers. The main objective of this fusion is to reduce the effort for imputing the missing values. As mentioned previously, the heat exchanger depends on multiple inputs from other systems, and accordingly, multiple measurements are used for monitoring. These measurements present a high level of noise caused by the dynamic behavior of the systems connected to the heat exchanger as well as the controllers that try to instantaneously correct the output of each system to reduce the fouling level. Hence, in the beginning, the missing values of each variable are temporarily replaced by its mean value to denoise data and reduce their dimensionality. The size of the reduction is defined according to the number of neurons in the hidden layer. In our case study, it is set to one. The performance of the encoding task is evaluated by the mean square error (MSE) metric, where the error is defined as the difference between the input data and the reconstructed data that is obtained through the decoding parameters. The data are fused through the hidden layer encoding parameters to denoise data and reduce their dimensionality. The size of the reduction is defined according to the number of neurons in the hidden layer. In our case study, it is set to 1. The performance of the encoding task is evaluated by the mean square error (MSE) metric, where the error is the difference between the input data and the reconstructed ones obtained through the decod-

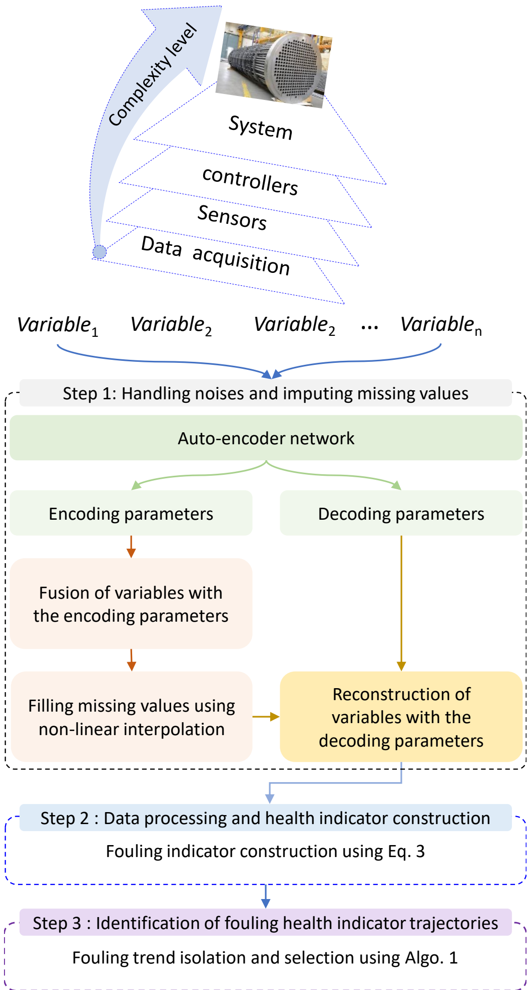


Figure 1. Flow chart of the proposed methodology.

ing parameters.

The second task relies on exploiting the fused data and replacing the previously added mean values with ones that capture the variability of the system instead of their average estimation. For this purpose, a non-linear interpolation technique is used to replace these values to better represent the non-linearity of the variables resulting from denoising and dimensionality reduction processes. Therefore, after handling the missing values, the new reduced array is used to reconstruct the original data through the decoding parameters and be ready for the processing step to build the fouling health indicator.

## 2.2. Data processing and health indicator construction

After preprocessing the raw data, the new observations are injected into feature extraction algorithm to build health indicator. This health indicators, in the prognostics framework, allows revealing the system's health state from its healthy condition to the critical state of failure. In the case of heat exchangers, the indicator is a fouling evolution level in the heat tubes. In practice, the construction of the fouling indicator strictly depends on the different inputs from multiple systems connected to the heat exchanger. Among these different parameters used in the literature are the inlet and outlet temperatures and flow rates that are exploited to estimate the overall heat transfer coefficient that is affected by the fouling phenomena (Ardsomang et al., 2013). Therefore, the variation in the overall heat transfer coefficient is a good indicator of fouling. The overall heat transfer coefficient can be calculated using the following heat exchanger design equation (Equation 1).

$$Q = U \times A \times \Delta T_{lm} \tag{1}$$

with

$$\Delta T_{lm} = \frac{(T_2 - t_2) - (T_1 - t_1)}{\ln[(T_2 - t_2)/(T_1 - t_1)]} \tag{2}$$

where $Q$ is the heat transfer rate between two fluids (hot and cold) in the heat exchanger, $U$ is the overall heat transfer coefficient, $A$ is the heat transfer surface area, and $\Delta T_{lm}$ is the logarithmic mean temperature difference ($t_1$, $t_2$), calculated from the inlet and outlet temperatures ($T_1$, $T_2$) of both fluids. Based on this heat exchanger design equation, a fouling health indicator (HI) representing the resistance to heat transfer can be defined as follows:

$$HI = \frac{\Delta T}{W} \tag{3}$$

$W$ represents the flow rate of the evaporated water, which is strongly correlated with the rate of heat transfer ($Q$) and $\Delta T$ represents the difference between the steam temperature and that of concentrated black liquor.

## 2.3. Identification of fouling health indicator trajectories

The processing layer output provides fouling level over time including multiple fouling scenarios from the healthy state to critical threshold due to the impact of maintenance actions and re-fouling mechanism after cleaning times. Indeed, when the fouling reaches a critical level, a cleaning action is performed to restore the heat tube to a healthy state that has a low level of fouling. Regarding the maintenance intervention of heat exchangers in the pulp and paper mills, most of the acquisition systems and their corresponding software do not provide the necessary information about the intervention time. Hence, in order to build an accurate prognostic model, it is necessary on one hand, to isolate each fouling trajectory to better observe the system behavior, on the other hand, to structure the input data in an appropriate format for prediction models. For this purpose, Algorithm 1 is proposed to deal with this challenge. This algorithm aims to automatically isolate the trajectories that represent separate fouling evolution trajectories. The isolation is based on a minimum threshold criterion and the difference in the fouling level at each time instant. When a cleaning action is performed, the fouling level is supposed to be at its lower value and the cleaning time will be used to set the beginning of a new trajectory. Hence, the procedure consists of calculating the difference between the fouling level at the time instant $(t)$ and the one at the time instant $(t+1)$. When this difference is lower than a threshold fixed by the expert, it means that the tubes are cleaned then the next observations are stored in a new trajectory.

**Algorithm 1** Fouling health indicator isolation and selection

**Initialization:**
Load the fouling indicator $HI$ obtained by Equation 3
Calculate the length of the fouling indicator as save in $len$
Set the initial threshold of fouling starting as $thr$
Create empty cell $Cycles$
**for** $i = 1$ *to* $len - 1$ **do**
  Calculate the difference between $(HI_{i+1}, HI_i)$ and save in $Diff$
**end for**
Find in $Diff$ the index where the values are lower than $thr$ and save in $Idx$.
Calculate the length of $Idx$ and save in $Idx.len$
**for** $j = 1$ *to* $Idx.len$ **do**
  Load the health indicator values from $HI_{Idx.len(j)}$ to $HI_{Idx.len(j+1)}$ and save in $Cycles_j$
**end for**
Evaluate the monotonicity of each cycle and select the ones greater than a defined threshold

The monotonicity is calculated using the following formula:

$$\text{monotonicity} = \frac{1}{\text{M}} \sum_{\text{j}=1}^{\text{M}} \left| \sum_{k=1}^{N_j-1} \frac{sgn\left(HI_j\left(k+1\right) - HI_j\left(k\right)\right)}{N_j - 1} \right| \tag{4}$$

where $HI_j$ represents the health indicator of the $jth$ heat exchanger, $M$ is the number of heat exchanger monitored, and $N_j$ is the number of observations collected from the $jth$ heat exchanger. In this paper $M = 1$.

After isolating the fouling trajectories, one needs to select the most informative ones. Because, in addition to the missing information about the time of maintenance, it happens that in some cases maintenance is carried while the fouling has not yet reached its critical level. In those cases, the selection of the fouling trajectories can be performed using some monotonicity metrics.

## 3. Case study

This section presents the case study used to validate the proposed methodology on the construction of fouling prognostics indicators of a complex piece of equipment in a pulp and paper mill in Canada. It gives an overview of the monitoring parameters and also shows the result obtained from each step of the proposed methodology.

### 3.1. The heat exchanger in Pulp and Paper Mills

The case study presented here is a system from a chemical pulping process, in which pulps are produced by cooking (digesting) the raw materials (wood chips), using the Kraft (sulfate) and sulfite processes. This system is a black liquor heat exchanger used to evaporate water from weak black liquor received from the pulp washing system to obtain highly concentrated liquor. This concentrated liquor is then burned in the chemical recovery section (recovery boiler) to generate steam that feeds a co-generation system. The fouling is evolved in the heat exchanger tubes due to dissolved solid deposits and increases over time resulting in a low-energy conversion performance. Figure 2 shows a simplified schematic of the closed-loop energy conversion process. Various sensors are placed on the systems to track and monitor the fouling parameters. A brief summary of these monitoring parameters is presented in Table 2. The data were collected during eight months of operation at a sampling frequency of 15 minutes under different operating conditions.

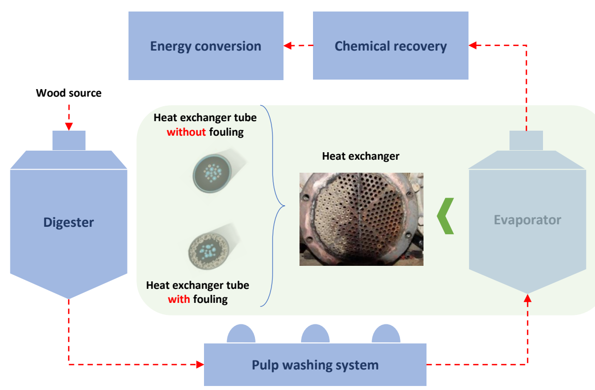


Figure 2. Simplified scheme of the case study.

### 3.2. Results and discussion

First, the indces of the missing values of the collected raw data were identified. Figure 3 shows the raw data with an illustration of the missing values in two different ranges of time.

In this figure, one can see that the raw data is contaminated with a high level of noise that can be attributed to the use of sensors with low quality noise rejection, in addition to frequent changes in controller setpoints or operating conditions. Moreover, there are several missing values that are mainly caused by unplanned maintenance actions made by the operators. In some reported cases, operators make unnecessary regulatory actions that need to halt the acquisition system.

Table 1. Summary of monitoring measurements.

| Variable | Description | Unit |
|---|---|---|
| V1 | Flow of liquor feeding the concentrator | $l/s$ |
| V2 | B.P.R. of liquor feeding the concentrator | $K$ |
| V3 | Concentration of liquor feeding the concentrator | $\%$ |
| V4 | B.P.R. of liquor leaving the concentrator | $K$ |
| V5 | Concentration of liquor leaving the concentrator | $\%$ |
| V6 | B.P.R. of liquor from concentrator's flash tank | $K$ |
| V7 | Solid concentration of liquor from flash tank | $\%$ |
| V8 | Temperature of vapor from concentrator | $K$ |
| V9 | Pressure of fresh steam sent to the concentrator | $kg/ms^2$ |
| V10 | Temperature differential steam-liquor | $K$ |
| V11 | Temperature of liquor from concentrator | $K$ |
| V12 | Temperature of liquor sent to storage tank | $K$ |
| V13 | Temperature of fresh steam | $K$ |
| V14 | Temperature of vapor from concentrator | $K$ |
| V15 | Liquor flow from concentrator's flash tank | $l/s$ |
| V16 | Liquor level in the concentrator | $\%$ |
| V17 | Flow of fresh steam to the concentrator | $kg/s$ |
| V18 | Pressure of concentrator's flash tank | $kg/ms^2$ |

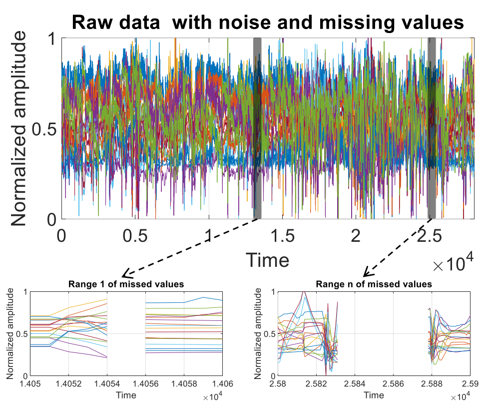


Figure 3. Raw data with noise and missing values.

To remedy this situation first, the data are injected into the AE network with the corresponding parameters summarized in Table 2 to reduce the dimensionality of these variables into one representative pattern. This is to facilitate the interpolation activity for imputing the missing values. Due to the existence of high level noisy signals, non-linear interpolation techniques are used to take into account these noises and outliers as principal information and generate the missing values of data. Hence, once the data are reduced using the encoding parameters of the network, the non-linear interpolation based on moving median method is applied to the fused data to replace the indices of missing values with new observations. Then, after filling in the missing values, it is now possible to reconstruct the original data with the filled missing values while keeping the maximum information on the non-linearity of the system. For the purpose of illustration, Figure 4 shows the filled missing values using the fused data and the reconstructed ones.

From Figure 4 it can be clearly seen that the non-linearity is well represented with the fused data. Also, thanks to the decoding parameters of the trained AE network, the original data are reconstructed with filled data with an MSE equal to $10^{-}2$ compared to the one obtained when using the principal component analysis (PCA) which is equal to $0.45$. As the missed values are now handled, the fouling indicator is build using Equation 3 through the variables v1,v15, v12 and v13. The obtained output from Equation 3 is injected into Algorithm 1 to automatically isolate the different trends representing the fouling evolution times.

Table 2. Parameter tuning of the AE model.

| Parameters | Attributed values |
|---|---|
| Training algorithm | Gradient descent momentum (GDM) |
| Encoder function | Logistic sigmoid function (Logsig) |
| Decoder function | Linear transfer function (Purelin) |
| Loss function | Mean absolute error (MAE) |
| Number of epochs | 500 |
| Hidden layer size | 1 |

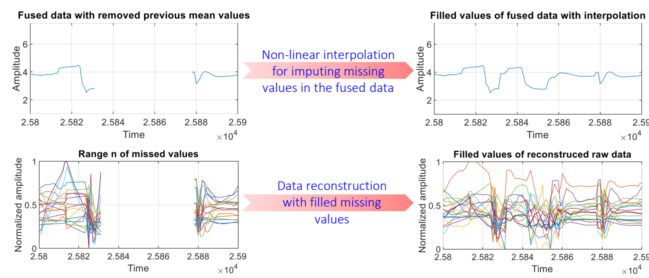


Figure 4. Filled missing values of fused and original data.

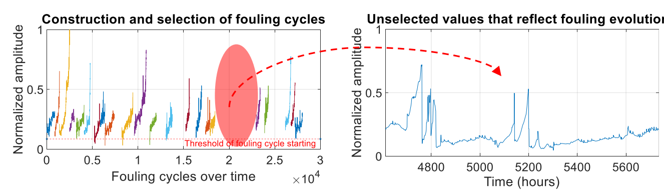


Figure 5. Identification of fouling trajectories.

From Figure 5, one can see that among the obtained trends, it happens that some of the observations represent only the primary part of fouling trajectory due to unplanned maintenance action for example. For this purpose, the last part of Algorithm 1 aims at calculating the monotonicity of each isolated trajectory in range of [0, 1] and then select the trends that have high monotonic tendency e.g. greater than $0.5$. For illustration, in the right side of Figure 5, an ensemble of observation does not show representative evolution of fouling with a monotonicity equal to $0.3$ and thus not selected as fouling indicator in this range of time.

The black-liquor heat exchanger is one of the most critical systems in the pulp and mill processes. Its main role is to increase the concentration of the black-liquor by evaporating the water through a high-level of heat transfer. The concentration of black-liquor is required to improve the combustion in the Kraft recovery boiler. The generated vapor from this boiler is converted into other forms of energy such as the electrical energy that is used as a supply source for the mill. In practice, the rate of heat transfer is significantly affected by the fouling phenomena, which exist in heat-exchanger tubes and increases over time. Hence, to ensure a high level of conversion operation and to avoid unnecessary maintenance actions, one needs to implement efficient monitoring of the fouling evolution and an accurate prognostics of the Time-to-Clean (TTC) of the heat-exchanger tubes.

It is worth mentioning that the methodology proposed in this paper was successfully applied to select the fouling trajectories that were then used to build an efficient prognostic tool to predict the time-to-clean of a black liquor heat exchanger in a pulp mill in Canada (M. Soualhi et al., 2021). This tool was used to provide the process operator with precise information about the future fouling evolution and to take the necessary preventive actions such as cleaning the heat tubes to reduce the fouling level and accordingly minimize the production downtime and its related costs.

## 4. Conclusion

This paper presents a data analytics methodology for the automatic construction of prognostic indicators. It uses special preprocessing procedures that allow transforming raw data with errors into exploitable information that reveals the different evolution trends. It starts by removing noise and reducing the dimensionality using an auto-encoder network. The reduced data are then used to fill the missing values through a non-linear interpolation technique. After that, the original data are reconstructed using the decoding parameters of the auto-encoder and then used to construct the prognostic indicator. This indicator is injected into an algorithm that automatically separates the different evolution trends and selects the appropriate ones. The proposed methodology is highlighted by a real application in the pulp and paper industry; the fouling prognostic in a heat exchanger. The obtained fouling trajectories are used to train machine learning algorithms to build an accurate fouling prognostic tool that helps the mill's operator take the right preventive maintenance actions.

## Acknowledgment

The project has been 65% cofinanced by the European Regional Development Fund (ERDF) through the Interreg V-A Spain France Andorra programme (POCTEFA 2014-2020). POCTEFA aims to reinforce the economic and social integration of the French-Spanish-Andorran border. Its support is focused on developing economic, social and environmental cross-border activities through joint strategies favoring sustainable territorial development. Also, the authors wish to thank the Natural Resources Canada's OERD (Office for Energy Research and Development) program for its financial support.

## References

Ahmed, R., Pai, P. S., Sriram, N., & Bhat, V. (2018). Comparison of wavelet based denoising schemes for gear condition monitoring: An artificial neural network based approach. In *Iop conference series: Materials science and engineering* (Vol. 310, p. 012010).

Ardsomang, T., Hines, J. W., & Upadhyaya, B. R. (2013). Heat exchanger fouling and estimation of remaining useful life. In *Annual conference of prognostics and health management society* (pp. 1–9).

Atamuradov, V., Medjaher, K., Camci, F., Dersin, P., & Zerhouni, N. (2018). Railway point machine prognostics based on feature fusion and health state assessment. *IEEE Transactions on Instrumentation and Measurement*, *68*(8), 2691–2704.

Gebraeel, N., Lawley, M., Liu, R., & Parmeshwaran, V. (2004). Residual life predictions from vibration-based degradation signals: a neural network approach. *IEEE Transactions on industrial electronics*, *51*(3), 694–700.

Guelpa, E., & Verda, V. (2020). Automatic fouling detection in district heating substations: Methodology and tests. *Applied Energy*, *258*, 114059.

He, Q. P., & Wang, J. (2007). Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE transactions on semiconductor manufacturing*, *20*(4), 345–354.

Lamraoui, M., El Badaoui, M., & Guillet, F. (2015). Chatter detection in cnc milling processes based on wiener-svm approach and using only motor current signals. In *Vibration engineering and technology of machinery* (pp. 567–578). Springer.

Markowski, M., Trafczynski, M., & Urbaniec, K. (2013). Identification of the influence of fouling on the heat recovery in a network of shell and tube heat exchangers. *Applied energy*, *102*, 755–764.

Naqvi, M., Yan, J., & Dahlquist, E. (2010). Black liquor gasification integrated in pulp and paper mills: A critical review. *Bioresource technology*, *101*(21), 8001–8015.

Ng, Y. S., & Srinivasan, R. (2010). Multi-agent based collaborative fault detection and identification in chemical processes. *Engineering Applications of Artificial Intelligence*, *23*(6), 934–949.

Nguyen, K. T., & Medjaher, K. (2020). An automated health indicator construction methodology for prognostics based on multi-criteria optimization. *ISA transactions*.

Oh, H., Lee, W.-Y., Won, J., Kim, M., Choi, Y.-Y., & Han, S.-B. (2020). Residual-based fault diagnosis for thermal management systems of proton exchange membrane fuel cells. *Applied Energy*, *277*, 115568.

Pettersson, K., & Harvey, S. (2010). Co2 emission balances for different black liquor gasification biorefinery concepts for production of electricity or second-generation liquid biofuels. *Energy*, *35*(2), 1101–1106.

Ragab, A., Yacout, S., Ouali, M.-S., & Osman, H. (2019). Prognostics of multiple failure modes in rotating machinery using a pattern-based classifier and cumulative incidence functions. *Journal of Intelligent Manufacturing*, *30*(1), 255–274.

Rai, V., & Mohanty, A. (2007). Bearing fault diagnosis using fft of intrinsic mode functions in hilbert–huang transform. *Mechanical Systems and Signal Processing*, *21*(6), 2607–2615.

Ratolojanahary, R., Ngouna, R. H., Medjaher, K., Junca-Bourié, J., Dauriac, F., & Sebilo, M. (2019). Model selection to improve multiple imputation for handling high rate missingness in a water quality dataset. *Expert Systems with Applications*, *131*, 299–307.

Saidi, L., Ali, J. B., Bechhoefer, E., & Benbouzid, M. (2017). Wind turbine high-speed shaft bearings health prognosis through a spectral kurtosis-derived indices and svr. *Applied Acoustics*, *120*, 1–8.

Sait, A. S., & Sharaf-Eldeen, Y. I. (2011). A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics. In *Rotating machinery, structural health monitoring, shock and vibration, volume 5* (pp. 307–324). Springer.

Soualhi, A., Medjaher, K., & Zerhouni, N. (2014). Bearing health monitoring based on hilbert–huang transform, support vector machine, and regression. *IEEE Transactions on instrumentation and measurement*, *64*(1), 52–62.

Soualhi, M., El Koujok, M., Nguyen, K. T., Medjaher, K., Ragab, A., Ghezzaz, H., … Ouali, M.-S. (2021). Adaptive prognostics in a controlled energy conversion process based on long-and short-term predictors. *Applied Energy*, *283*, 116049.

Steinhagen, R., Müller-Steinhagen, H., & Maani, K. (1993). Problems and costs due to heat exchanger fouling in new zealand industries. *Heat transfer engineering*, *14*(1), 19–30.

Sukumar, P., Robert, L., & Yuvaraj, S. (2016). Review on modern data preprocessing techniques in web usage mining (wum). In *2016 international conference on computation system and information technology for sustainable solutions (csitss)* (pp. 64–69).

Sun, D., Lee, V. C., & Lu, Y. (2016). An intelligent data fusion framework for structural health monitoring. In *2016 ieee 11th conference on industrial electronics and applications (iciea)* (pp. 49–54).

Zarei, J., Tajeddini, M. A., & Karimi, H. R. (2014). Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics*, *24*(2), 151–157.

Zhang, B., Sconyers, C., Orchard, M., Patrick, R., & Vachtsevanos, G. (2010). Fault progression modeling: An application to bearing diagnosis and prognosis. In *Proceedings of the 2010 american control conference* (pp. 6993–6998).