Energy 220 (2021) 119733

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

# Operational guidelines for emissions control using cross-correlation analysis of waste-to-energy process data

Cansu Birgen*, Elisa Magnanelli, Per Carlsson, Michaël Becidan

*SINTEF Energy Research, 7034, Trondheim, Norway*

ARTICLE INFO

*Article history:*
Received 28 September 2020
Received in revised form
14 December 2020
Accepted 26 December 2020
Available online 29 December 2020

*Keywords:*
Waste-to-energy
Hampel filtering
Process data
Cross-correlation
Nonlinear
Time lag

ABSTRACT

The aim of this study was to develop a data analysis method which could provide operational insights and guidelines waste-to-energy (WtE) plant operators. A method to filter outliers with changing properties was combined with a cross-correlation analysis method that can capture nonlinearity and quantify time lags between variables. The method was applied to a dataset obtained from a commercial WtE plant. The method was able to detect already established correlations such as the influence of combustion conditions on $NO_x$ and CO emissions, which both had positive correlation with $O_2$ concentration in the flue gas, while the effect of combustion conditions was unnoticeable for HCl emissions. Furthermore, the method could detect that $NO_x$ and $SO_2$ emissions exhibited positive correlations with the furnace temperature. Time lags provided additional information about the sensor locations and plant dynamics. This methodology can be used especially when process data is available while good process models are not immediately accessible for determining non-obvious process phenomena not only for WtE sector but also for process industry in general.



## 1. Introduction

Heat and/or power can be produced from municipal solid waste (MSW) in Waste-to-Energy (WtE) plants. WtE technology helps to improve resource and energy efficiency by recovering the energy from waste fractions that cannot be material recycled [1]. The number of WtE plants in Europe reached 512 in 2016 and is expected to increase hitting an accumulated capacity of 216 PJ [2]. A similar trend is observed for Norway with a 36% increase in total WtE capacity since 2010 with 17 plants in operation [3]. As evident from the progress in the field, applications of WtE will become more widespread making it a necessity to generate knowledge to advance this technology.

WtE plants are confronted with strict regulations in terms of emissions in addition to efficiency and reliability demands. Therefore, flue gas cleaning is crucial to achieve emission targets as well as cost reduction deriving both from corrosion induced maintenance and from the use of consumables such as chemicals and filters [4,5]. Emissions in a WtE plant are influenced by i) the chemical composition of MSW, ii) physical properties of MSW (moisture content, physical shape and geometry etc.) and iii) operating conditions (incineration system, gas flow rates, waste feeding rate etc.). After initial sorting, WtE plants need to incinerate whatever is delivered and it is therefore difficult to apply emission reduction measures related to the waste properties. As a result, controlling operating conditions of the combustion unit to minimize emissions has been central in numerous WtE studies and has shown great potential [6]. In this study we focus on unveiling relationships between gas emissions and combustion conditions by unlocking the potential of process data.

WtE plants are typically monitored by hundreds of sensors continuously measuring physical quantities. These high-dimensional and large datasets hold a great potential for knowledge mining. However, the data often need to be pre-processed before utilization. Presence of noise and outliers is one of the most common issues [7]. Noise can be defined as any point that is not the true data/signal whereas outlier is a broader term covering noise as well as failures, faults and natural variations in the system, and is typically characterized as being significantly different from the rest of the data [8]. The frequency and duration of noise and outliers might vary for each sensor and/variable in the plant requiring individual treatment. Conventional filtering methods such as moving-average filters are widely used for noise removal. However, they might not perform equally well in case of outliers

* Corresponding author.
*E-mail address:* cansu.birgen@sintef.no (C. Birgen).

https://doi.org/10.1016/j.energy.2020.119733

[9]. Therefore, it is crucial to implement a filtering technique that can efficiently eliminate both noise and outliers in case of varying characteristics. In this study, we address the problem of noise and outliers with changing characteristics by applying a Hampel filter [10] that has shown to be effective in a variety of applications [9,11].

Time lags are observed in process data due to inherent process properties, feedback control systems, time dependent noises, etc. [7]. Nonlinearity in data correlations is another issue that needs to be considered when analysing dependencies [7]. Cross-correlation analysis is used to quantify correlations as well as time lags between variables. However, this widely used method can only identify linear relationship between variables as it is based on Pearson's correlation coefficient [12]. Despite its limitations, Pearson's correlation was used in a MSW combustion emissions study [13] with a small dataset size (5 data points of each variable) and a corrosion kinetics study for MSW combustion [14]. These studies did not have any focus on the method applicability even though their data showed nonlinear trend. Similarly, the study of Akimoto et al., 2005 investigated the effects of human activities through cross-correlation of urban atmospheric concentrations of $N_2O$, $CO_2$, $CH_4$ as well as meteorological parameters without considering the nonlinearity of data or other aspects such as missing data points in their dataset [15]. Cross-correlation analysis has also been applied in various renewable energy systems studies, for example, determining if nearby wind farms are providing ancillary information by using time lags between measurements obtained from cross-correlation analysis [16], or in an inter-annual and seasonal analysis of the cross-correlation between wind and wave resources [17]. However, these studies did not explain why this specific method was used and its advantages over other methods. Moreover, these studies contained limited information about the uses and limits of the method's application e.g., linear correlation since neither of them dealt with development of the method. It is also important to mention the misuse of terms; for example, two studies that dealt with variability in power systems used the term cross-correlation [18,19], even though the cited source clearly names the method correlation analysis [20] and there is no presence of time lags. Therefore, it is necessary to develop methods that can address time lag and nonlinear properties of data while considering the properties of the dataset.

The aim of this study was to develop a cross-correlation analysis method that can capture the time lags and nonlinearity between combustion conditions and gas emissions in a WtE plant.

## 2. Methodology

The methodology is briefly described in Fig. 1. First, relevant process variables are selected, next each variable is filtered, later a cross-correlation analysis is performed to estimate the correlations as well as time lags between each variable pair. This section contains details of each of these steps starting with a description of the Returkraft WtE plant located in Kristiansand, Norway.

### 2.1. WtE plant and dataset

The Returkraft plant is schematically described in Fig. 2 and a more in-depth description can be found in our previous work [21]. MSW is fed to the system from the waste bunker via a feeder. Then it is incinerated using primary and secondary air to produce flue gas. The hot flue gas is then sent to a heat exchanger where steam is produced.

Twelve variables were selected for the analysis, each containing

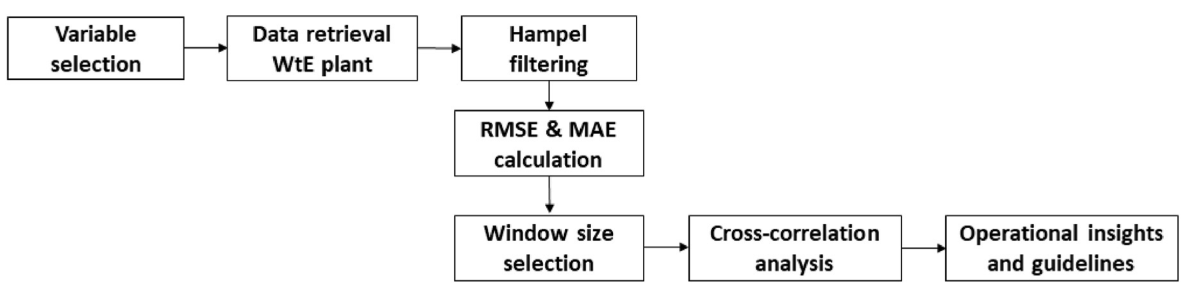


**Fig. 1.** Overview of the methodology followed in this study.

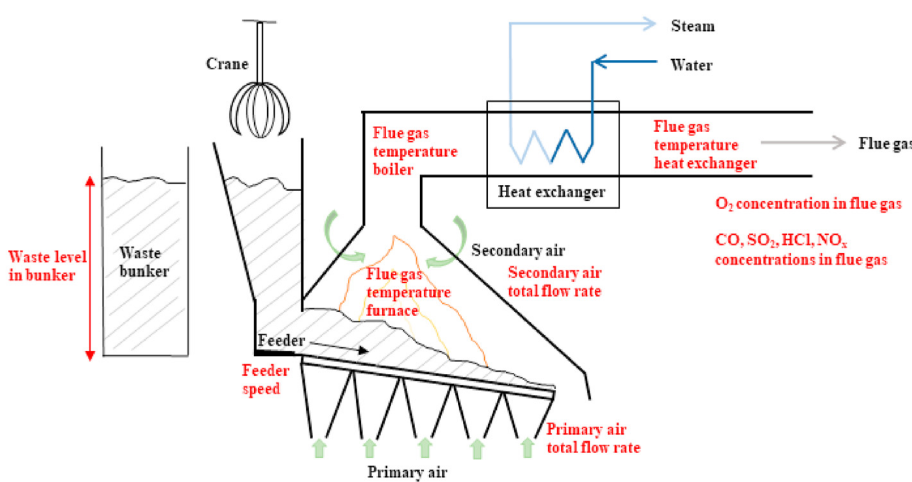


**Fig. 2.** Simplified illustration of the WtE plant considered in this study.

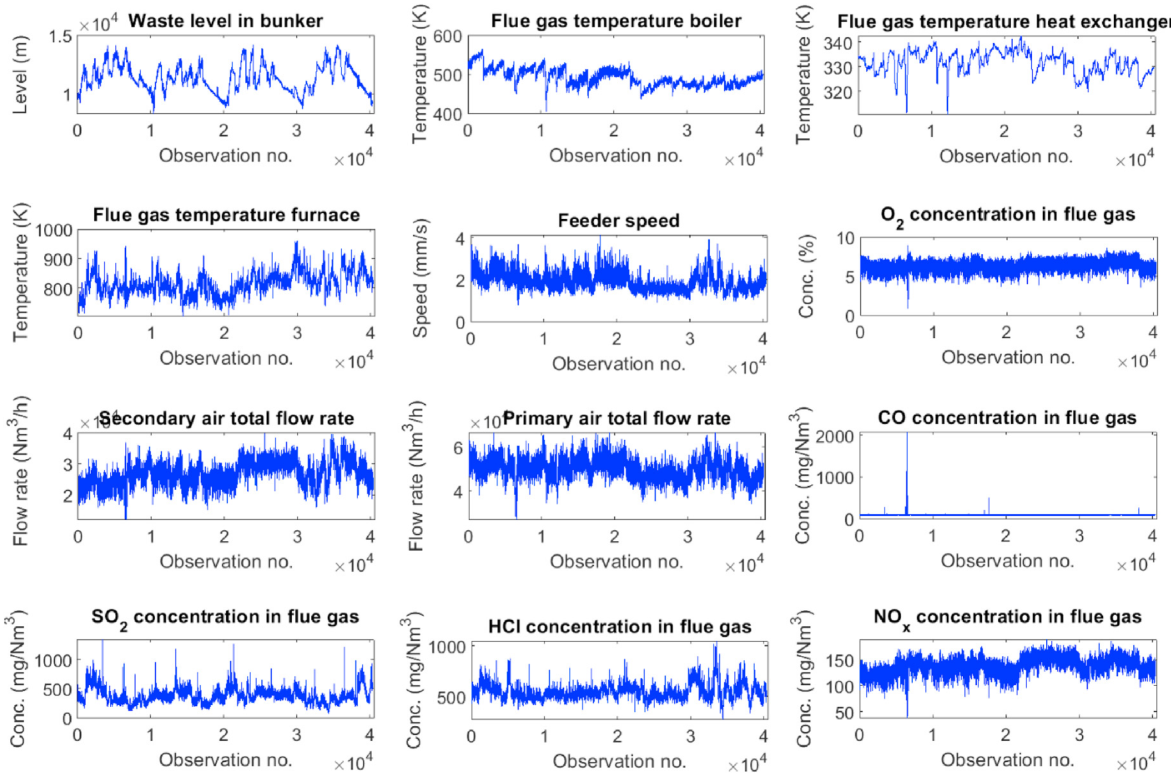


**Fig. 3.** Raw data of the variables used in this study.

40320 observations recorded every minute covering 4 weeks of plant operation (February 03, 2020 to March 01, 2020). The variables were selected according to domain knowledge and considering the suggestions from the experts working in WtE plant operation. Fig. 3 shows raw data plots of the variables: waste level in bunker, flue gas temperatures in boiler, heat exchanger and furnace, feeder speed, primary and secondary air total flow rates, $O_2$, CO, $SO_2$, HCl and $NO_x$ concentrations in flue gas.

### 2.2. Hampel filter

Outlier is a broader term covering noise as well. Therefore, all points removed by filtering are referred to as outlier throughout the study. The Hampel filter is calculated using the median value of the data sequence $x = x_1, x_2, \ldots, x_n$ and the mean absolute (MAD) from the median [10]. Two parameters that the Hampel filter works on are the predefined threshold, T, and the local median value of the chosen window size $(2k + 1)$ with k being the number of neighbors on either side of a data point $x_s$, as illustrated in Fig. 4. Local median, $m_i$ and local MAD, $MAD_i$ are shown in Equations (1) and (2), respectively, adapted from Ref. [11].

$$m_i = \mathrm{median}(x_{i-k}, x_{i-k+1}, \ldots, x_i, \ldots, x_{i+k-1}, x_{i+k}) \quad \text{Eq. 1}$$

$$MAD_i = \sigma_i / \kappa \quad \text{Eq. 2}$$

where $\sigma_i$ is the local standard deviation as shown in Equation (3) that was adapted from Ref. [11] and $\kappa$ is the scaling factor chosen as 1.4826 so that the standard deviation of the normal distribution of the data covers 50% of the standard normal cumulative distribution function [22]. The value of $\sigma_i$ becomes 0 when more than half of the data has the same value as $x_i$ because of crudely quantized data, and the rest of the data sequence is identified as outliers regardless of their distance from $m_i$ [23].

$$\sigma_i = \kappa\ \mathrm{median}(|x_{i-k} - m_i|, \ldots, |x_{i+k} - m_i|) \quad \text{Eq. 3}$$

According to the Hampel filter algorithm, a data point is detected as an outlier when Equation (4) holds, adapted from Ref. [11].

$$|x_{i-k} - m_i| > T\,\sigma_I \quad \text{Eq. 4}$$

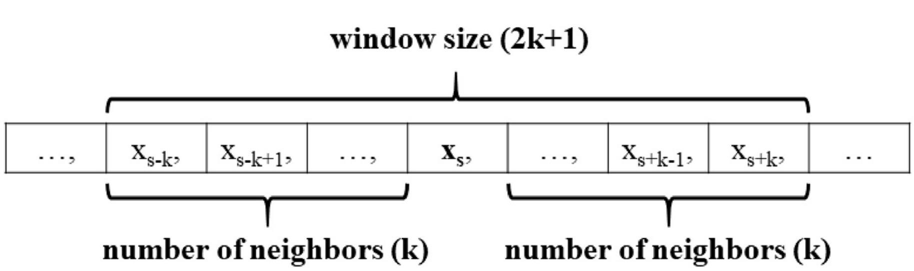


**Fig. 4.** Illustration of the window size for Hampel filter.

Outliers are replaced by the median value of the corresponding window. Threshold value, T is chosen as 3 in this study.

The built-in function hampel from MATLAB® Signal Processing Toolbox was used to apply the Hampel filter. First, the local MAD of each variable was calculated using Equation (2) with a window size of 2k+1. Then, the local MAD value was used together with the threshold to detect outliers according to Equation (4). Outliers were then removed.

Near the sequence endpoints, the function truncates the corresponding window to calculate $m_i$ and $\sigma_i$. Equations (5) and (6) are used when i < k+1, and Equations (7) and (8) are used when i < n-k by the Hampel filtering algorithm, adapted from Ref. [9].

$$m_i = \mathrm{median}(x_1, x_2, x_3, \ldots, x_i, \ldots, x_{i+k-2}, x_{i+k-1}, x_{i+k}) \qquad \text{Eq. 5}$$

$$\sigma_i = \kappa\ \mathrm{median}(|x_1 - m_1|, \ldots, |x_{i+k} - m_i|) \qquad \text{Eq. 6}$$

$$m_i = \mathrm{median}(x_{i-k}, x_{i-k+1}, x_{i-k+2}, \ldots, x_i, \ldots, x_{n-2}, x_{n-1}, x_n) \qquad \text{Eq. 7}$$

$$\sigma_i = \kappa\ \mathrm{median}(|x_{i-k} - m_i|, \ldots, |x_n - m_n|) \qquad \text{Eq. 8}$$

The procedure is then repeated for the entire dataset and all variables.

### 2.3. Hampel filter performance measures

Window size selection is of critical importance for an effective outlier removal. Too large window size can cause outlier points to remain undetected, while a too small window size could cause the removal of significant datapoints. Therefore, the Hampel filter was applied for a window size range of 1–180 that corresponds to 1 min - 180 min as the data is recorded every minute. This selection is based on the fact that it takes approximately 3 h for MSW in the waste bunker to be incinerated, and flue gas to be analysed by the last sensors as shown in the simplified process figure, Fig. 2.

For the predefined window size range, the Hampel filter was applied and the resulting performance measures were calculated to find the optimum window size. Two performance measures were considered: the root-mean-square error (RMSE) is more widely used; however, it can be less appropriate than the mean absolute error (MAE) in the presence of large outliers characterized as high spikes standing out in a given data sequence [11]. For window size of 2k+1, RMSE and MAE equations are given in Equations (9) and (10), respectively [9].

$$\mathrm{RMSE} = \sqrt{}(\mathrm{mean}(\mathrm{sum}((x_{i+k} - y_{i+k})^2, \ldots(x_{i+k} - y_{i+k})^{\ 2}))) \qquad \text{Eq. 9}$$

$$\mathrm{MAE} = \mathrm{mean}(\mathrm{sum}(|x_{i+k} - y_{i+k}|, \ldots\ |x_{i+k} - y_{i+k}|)) \qquad \text{Eq. 10}$$

where $x_i$ is the ith observation and $y_i$ is the filtered data point of the same variable at the same location. For each window size and variable, RMSE and MAE are calculated for the entire data sequence of the individual variable.

Upper and lower limits i.e. confidence interval (CI) for each outlier detection were estimated using Equation (11), adapted from Ref. [24].

$$\mathrm{CI} = m_i \pm 3\ \sigma_i \qquad \text{Eq. 11}$$

### 2.4. Cross-correlation analysis

Spearman's and Kendall tau's rank correlation measures are widely used in correlation analysis due to their robust performance in presence of outliers and their ability to detect nonlinear dependencies in contrast to the most used Pearson's correlation [25]. Kendall's tau is more robust to outliers than Spearman's correlation. However, Hampel filtering was applied prior to correlation analysis; therefore, being insensitive to outliers was a less critical quality. Therefore, the analysis in this study was based on Spearman's correlation.

Spearman's rank correlation for bivariate data (x,y) with sequences of $x = x_1, x_2, \ldots, x_n$ and $y = y_1, y_2, \ldots, y_n$ is calculated as in Equation (12) adapted from Ref. [25].

$$s_s(x,y) = \mathrm{sum}((s_{x,1} - \hat{s}_x)(s_{y,1} - \hat{s}_y), \ldots, (s_{x,n} - \hat{s}_x)(s_{y,n} - \hat{s}_y)) \qquad \text{Eq. 12}$$

where $s_{x,n}$ and $s_{y,n}$ are ranks of the nth numerical value of variables x and y. $\hat{s}_x$ - $\hat{s}_y$ are mean values of $s_x$ and $s_y$, and $\sigma_s(x,y)$ is the Spearman's rank correlation for x and y.

MATLAB® built-in function, xcorr for cross-correlation analysis was used. For a variable pair of x and y, xcorr calculates the raw correlation between x and shifted (lagged) copies of y as a function of the lag. For a maximum lag k, xcorr returns a cross-correlation sequence in range of -k to k. The normalization option of the xcorr function was used so that autocorrelations at zero lag equal 1.

xcorr calculates raw correlations in the same way as Pearson's correlation coefficient but without extracting the variable means. Cross-correlations were calculated for each pair of 12 variables with a maximum lag of 180 min. This was done by implementing ranks of the variables instead of the raw variables and extracted their mean so that xcorr returns Spearman's rank correlations that can capture nonlinear dependencies.

## 3. Results and discussion

This section provides the results of the Hampel filtering first, followed by the correlation, and the cross-correlation analysis results.

### 3.1. Hampel filtering of variables

RMSE and MAE values give indication of the filter performance and were used to find the optimal window sizes for each variable in this study as illustrated in Fig. 5.

Optimal window sizes were selected based on the peak points in RMSE and MAE values, then separately applied to the variables to remove the outlier using the Hampel filter. A higher error value suggests a better removal of outliers that have a large prominence from the rest of the data sequence. RMSE and MAE values were calculated for Hampel filtering with window size range of 1–180 for all 12 variables as shown in Fig. 5. The influence of window size on filter performance measure values was prominent for all variables. Local variations in RMSE and MAE values were also noticeable. For all the variables except CO and $O_2$ concentrations in the flue gas, RMSE and MAE showed a parallel trend with respect to the window size. The reason for this difference can be that this variable has outliers characterized as higher spikes with shorter duration in Fig. 3 compared to the rest of the variables.

Highest RMSE and MAE for CO concentration in flue gas were calculated as 10.714 and 0.348 mg/Nm$^3$. The window sizes corresponding to these values were 15 and 2. Hampel filtering with these window sizes were applied to CO concentration in flue gas as shown in Fig. 6. Both window sizes of 2 and 15 could successfully remove the large outliers seen as high spikes (Fig. 6). However, the Hampel filter with small window size identified 2834 outliers compared to 295 for window size of 15.

For some variables, several local maxima occurred for both

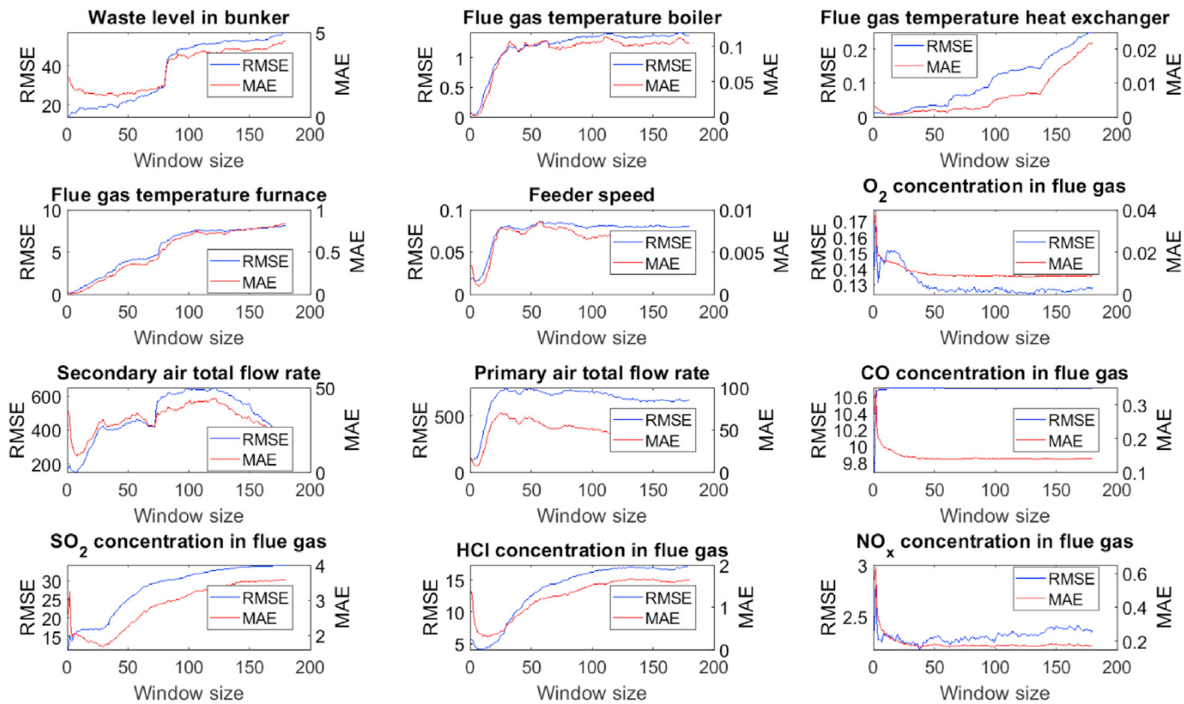


**Fig. 5.** RMSE (left y-axis) and MAE (right y-axis) values for Hampel filter applied with varying window sizes.

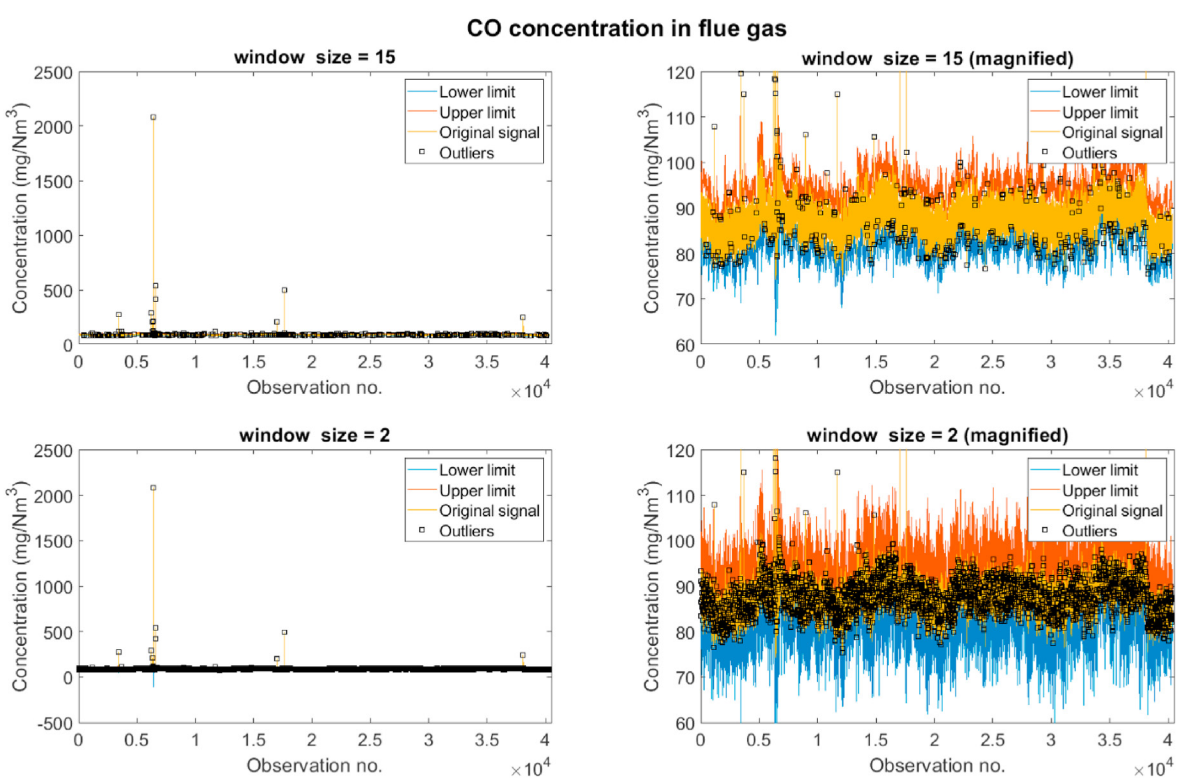


**Fig. 6.** Hampel filtering of CO concentration in flue gas for window sizes of 2 (bottom) and 15 (top).

RMSE and MAE values. One of them was waste level in bunker that was filtered with Hampel filter with different window sizes as shown in Fig. 7. RMSE and MAE values for waste level in bunker exhibited two apparent peaks at window sizes of 82 and 180 (Fig. 5). Hampel filtering was applied at these sizes and detected 395 and 269 outliers respectively out of 40320 data points.

Effect of window size is illustrated in Figs. 6 and 7 (two variables shown). There is a trade-off between removal of outliers and preserving the useful and meaningful process information. Therefore, the window sizes were selected by consulting the error values

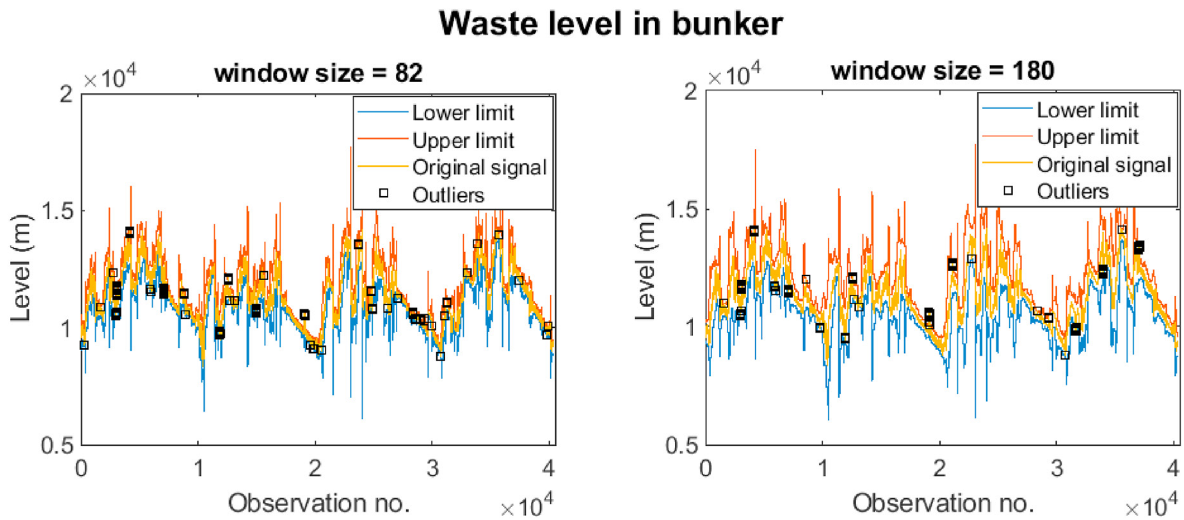


**Fig. 7.** Hampel filtering of waste level in bunker for window sizes of 82 (left) and 180 (right).

given in Fig. 5. Selected window sizes for Hampel filtering and the respective number of outliers detected by the filter are given in Table 1.

As it can be seen in Table 1, for flue gas temperatures, feeder speed and total air flow rates, optimal window sizes were in range of 29–32, meaning that the outliers are best detected in a time window of 29–32 min for these variables. The optimal time window for outlier detection was 84 min for waste level in bunker, 12–15 min for $O_2$ and CO concentrations in flue gas, and 53–54 min for $SO_2$, HCl and $NO_x$ concentrations in flue gas. This shows that it is important to tailor-make the data filtering procedure when we address a dataset where outliers have varying characteristics and occur with different time windows. Filtered data is used in cross-correlation analysis for each variable, as shown in section 3.2.

### 3.2. Correlation analysis of variables

Spearman's correlation was calculated for all variables without considering time lags to obtain an overview of all correlations. The correlation coefficient has a value between +1 and −1, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation [26–28]. In the literature, it is possible to find different interpretations and rankings of correlation coefficients in terms of statistical significance. Yet many researchers agree that when the absolute value of the correlation coefficient is smaller than 0.1, the correlation is considered insignificant, and a correlation is considered strong for values greater than 0.9 [26–28].

To avoid misinterpretation and overreading, the correlations were discussed with experts in WtE operation and the proposed explanations incorporated their perspectives and experiences in addition to the domain knowledge supported by the information and observations from scientific articles. The resulting correlations are visualized with the heatmap in Fig. 8.

The first line in Fig. 8 shows the correlations between waste level in the bunker and the other considered variables. Waste is delivered to the plant during working hours of weekdays. Therefore, waste level and waste residence time in the bunker varies over time. This influences the level of mixing of the waste in the bunker thus, its homogeneity. Variation in waste properties has a strong influence on combustion [21]. Waste level in bunker had positive correlations with secondary air total flow rate, $O_2$ (unreacted, surplus in flue gas) and CO (not converted to full combustion product $CO_2$), which might indicate incomplete combustion supported by a negative correlation with flue gas temperature in boiler (−0.217). Therefore, according to the correlation results, waste homogeneity can influence combustion efficiency.

Flue gas temperature boiler had a positive correlation (0.403) with flue gas temperature heat exchanger. This was expected, since these two temperatures are measured at the inlet and outlet of a heat exchanger unit. On the other hand, the correlation with flue gas temperature furnace was negative (−0.340), as well as the correlation with secondary air total flow rate (−0.280). The high correlations between feeder speed and primary air total flow rate were a result of the combustion control system structure. Indeed,

**Table 1**
Window sizes for Hampel filtering and number of outliers filtered for each variable.

| Variable | Window size | Number of outliers detected |
|---|---|---|
| Waste level in bunker | 84 | 285 |
| Flue gas temperature boiler | 32 | 405 |
| Flue gas temperature heat exchanger | 31 | 192 |
| Flue gas temperature furnace | 33 | 341 |
| Feeder speed | 32 | 437 |
| $O_2$ concentration in flue gas | 12 | 487 |
| Secondary air total flow rate | 29 | 234 |
| Primary air total flow rate | 30 | 370 |
| CO concentration in flue gas | 15 | 395 |
| $SO_2$ concentration in flue gas | 54 | 523 |
| HCl concentration in flue gas | 54 | 361 |
| $NO_x$ concentration in flue gas | 53 | 261 |

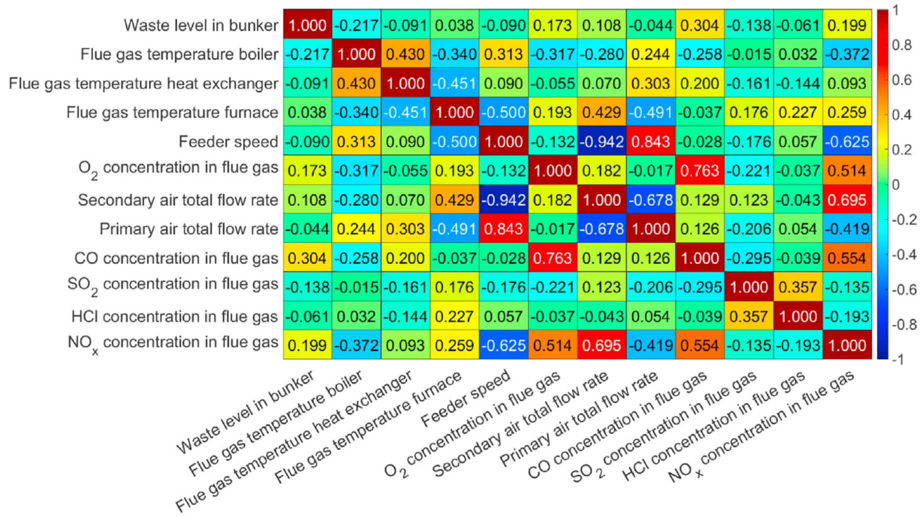


**Fig. 8.** Spearman's rank correlations between each pair of 12 variables.

both parameters are controlled by the control system to maintain a constant steam production. Flue gas temperature boiler showed negative correlations with $O_2$ and CO concentrations in flue gas as a result of the good combustion status evident from its negative correlation with flue gas temperature in the furnace as well.

Feeder speed had a negative correlation with flue gas temperature in the furnace (−0.500); high feeder speed might cause a thicker waste bed resulting in changed combustion conditions and thus a lower flue gas temperature in the furnace. $O_2$ and CO concentrations in flue gas give the most reliable information about combustion status. This is confirmed by the strong correlation between them (0.763). Feeder speed showed strong correlations with other manipulated variables as explained above.

### 3.3. Cross-correlation analysis of variables

After obtaining the correlations between all variables, a cross-correlation analysis between emissions and the other considered variables was performed. Figs. 9–12 show the cross-correlations for emissions ($NO_x$, CO, HCl and $SO_2$ concentrations in flue gas).

$NO_x$ emissions are a product of complete oxidation of nitrogen that has three major generation mechanisms: fuel, thermal and prompt $NO_x$ mechanisms [4]. In WtE plants, they typically originate from fuel mechanism while some thermal production can also occur. Correlation of $NO_x$ with waste level in bunker ranged from 0.171 to 0.203 for lags of −180 and 120 min, respectively. However, this correlation did not vary greatly with changing lags (Fig. 9). The

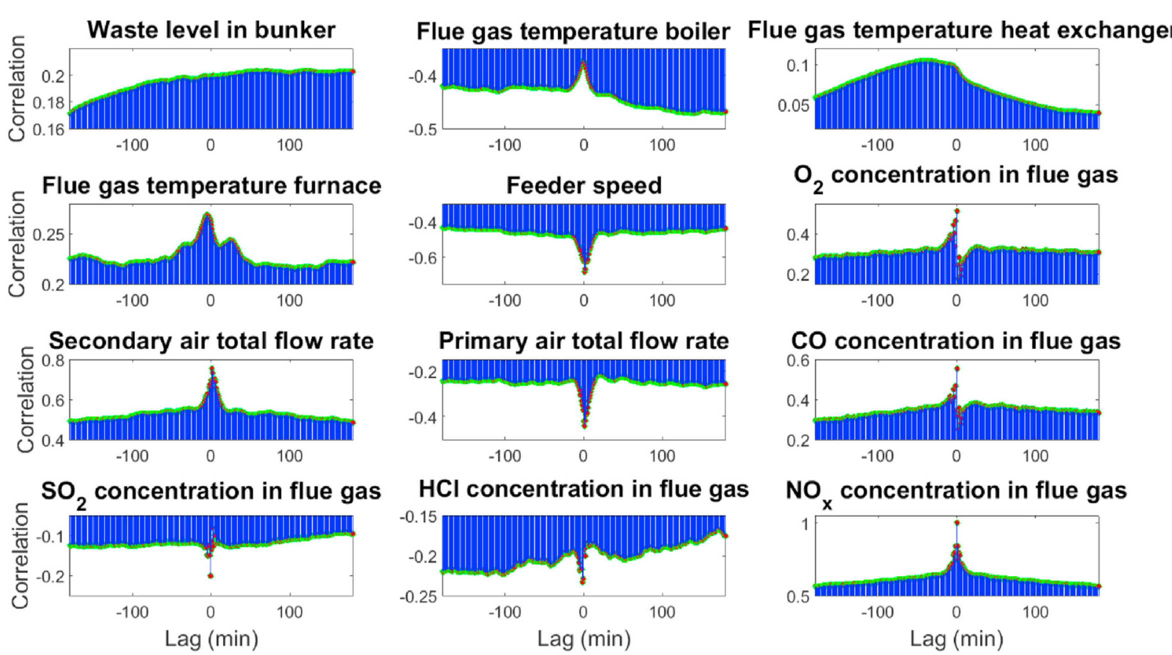


**Fig. 9.** Cross-correlations between $NO_x$ and other variables.

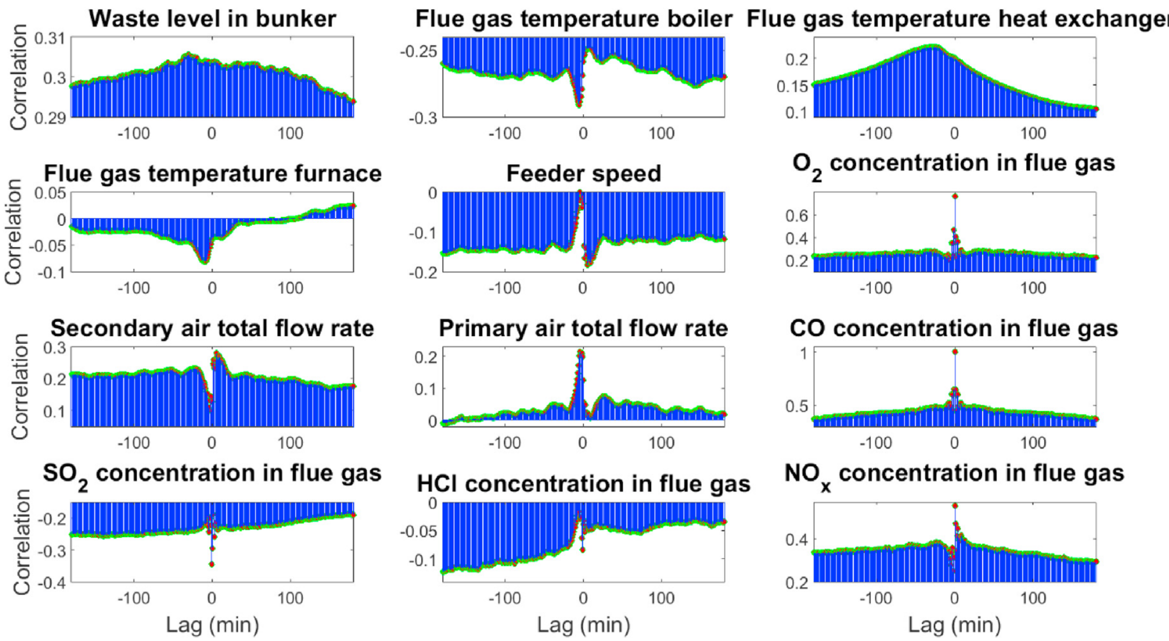


**Fig. 10.** Cross-correlations between CO and other variables.

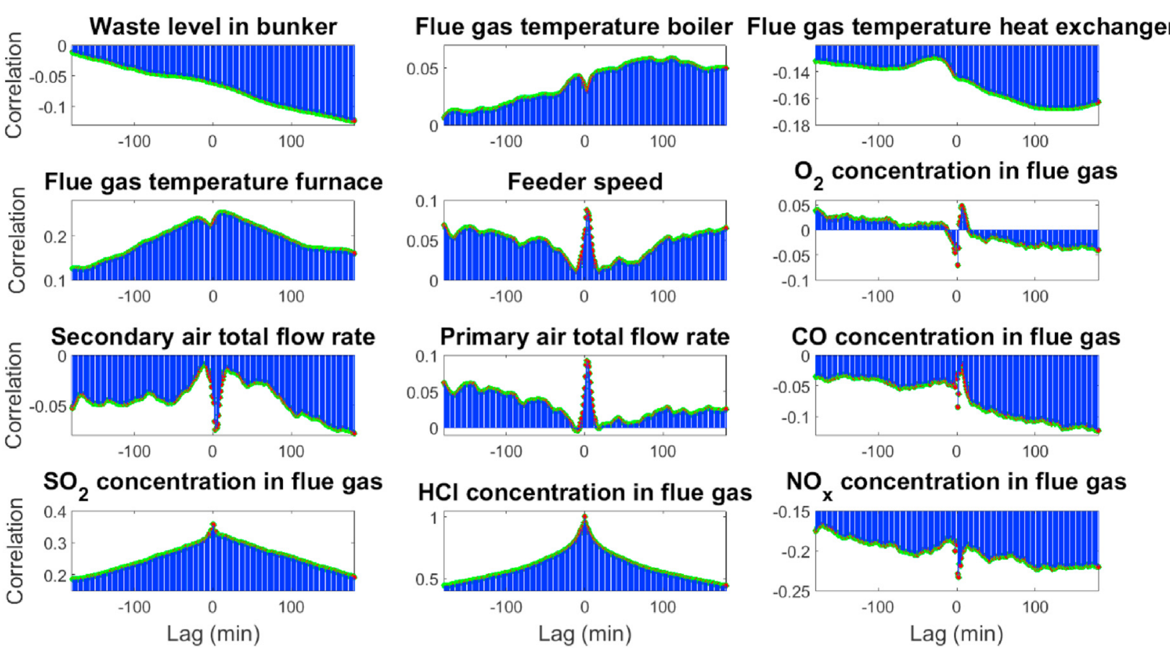


**Fig. 11.** Cross-correlations between HCl and other variables.

same applies to flue gas temperature downstream heat exchanger. $NO_x$ correlation with flue gas temperature in furnace was positive with highest value at −5 min lag, as the flue gas takes some time to travel from the furnace to the $NO_x$ sensor. Positive correlation can be due both to thermal $NO_x$ formation at higher temperatures and to higher heating value of nitrogen containing waste fractions. Strong correlations of $NO_x$ with manipulated variables (i.e. primary and secondary air and feeder speed) were highest at time lag of 1 that can be due to the control system action triggered by the $O_2$ measurement. Positive correlations of $NO_x$ with CO and $O_2$ suggest that $NO_x$ formation in this WtE plant is related to combustion conditions. Negative and weak correlations with $SO_2$ and HCl can be due to the elemental composition of different waste fractions, and the time lags (−1, −2 min) can be explained by difference in reaction kinetics as well as the measurement.

Correlation of CO with waste level in bunker was highest (0.306) at time lag of −30, but it did not vary significantly with lag as for $NO_x$ (Fig. 10). CO and flue gas temperature in boiler were negativelycorrelated, as high temperature promote CO oxidation. Correlation was the highest at lag −5, which could be explained by the time that flue gas takes from the boiler to the CO sensor location. The lag was longer than that of $NO_x$ even though the gas sensors are in the same location. CO had weaker correlations with manipulated variables at different lags.

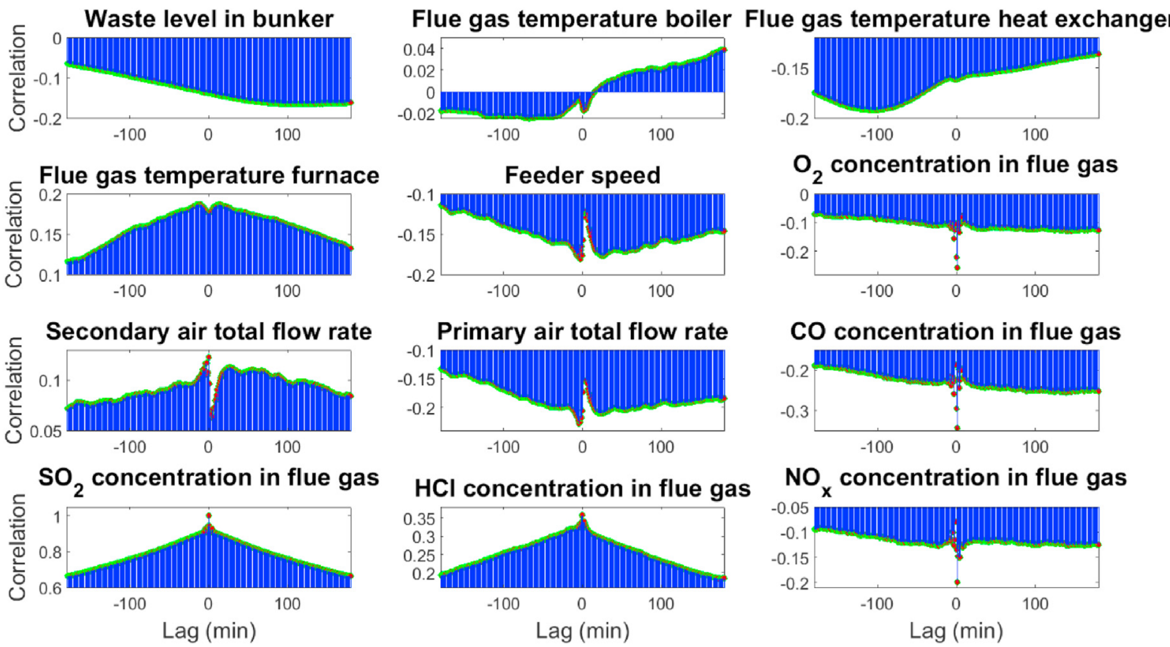


**Fig. 12.** Cross-correlations between $SO_2$ and other variables.

HCl concentration in flue gas showed a weak positive correlation with flue gas temperature in furnace with the highest correlation of 0.252 at time lag of 11; however, its value did not vary greatly at similar lags (Fig. 11). HCl had significant correlations with two other variables; $NO_x$ in flue gas that was explained above and $SO_2$ in flue gas that had its highest value of 0.357 at zero lag. The latter could be explained by the elemental composition of the waste fractions whose combustion yield both HCl and $SO_2$. It should be noted that HCl forms due to Cl-containing waste fractions such as plastics, which typically have a high heating value.

Correlations of $SO_2$ were similar to that of HCl's for flue gas temperature in furnace and manipulated variables (Fig. 12). A known S-rich waste fraction processed at the considered plant is sludge, which can cause unstable combustion. This can result in incomplete combustion that gives negative correlations with CO and $O_2$ concentrations in flue gas [29]. As for $NO_x$ concentration in flue gas, $SO_2$ concentration increases with increasing flue gas temperature in the furnace, which is in line with the findings of a lab-scale MSW incineration study without any sorbent addition performed at the same temperature range [30]. Similarly, a coal combustion study has shown the increase of $NO_x$ with added $SO_2$, where the NO emissions was found to be influenced by the S/N ratio in the flame, which was in turn controlled by the coal composition, pyrolysis behaviour and physical properties [31].

## 4. Conclusions

In this study, a new method was developed to filter data from outliers that have changing properties, and to perform cross-correlation analysis that can capture nonlinear relationships between variables as well as quantify time lags. The method was applied to a real time-series process dataset with 12 variables obtained from a commercial WtE plant. RMSE and MAE values were used as filter performance indicators when selecting optimal window size for Hampel filtering of each variable. This study revealed the importance of [1] tailor-made data filtering to address the outliers' varying characteristics and [2] the selection of appropriate time windows as outliers can be different for each variable in the dataset. The cross-correlation analysis results unveiled the relationships between combustion conditions and gas emissions together with time lags between them. The main contributions of this study can be stated as below:

1. The Hampel filtering of the high-dimensional real WtE process plant data was performed for the first time with the extension of optimal window size selection by using filter performance indicators. This provides an advantage over existing approach in which the specific outlier characteristics are not considered for each variable in a large dataset, thus overlooking the trade-off between outlier removal and restoring the useful data.
2. A cross-correlation analysis method that can capture the nonlinear correlations together with time lags was developed. The method was applied to a real WtE plant dataset to test its applicability, and the correlations could explain causations to a great extent; therefore, they can be used to provide operational insights.
3. The method developed in this study can especially be useful in situations where there is large amount of process data available while good process models are not immediately accessible.

The operational insights and guidelines obtained by this study can be summarized as follows:

i. $NO_x$ and CO emissions were highly influenced by the combustion conditions (e.g. combustion air flow rate), and their positive correlation with $O_2$ concentration in the flue gas suggested their link to incomplete combustion.
ii. $O_2$ concentration in the flue gas increased with the level in the waste bunker.
iii. CO concentration increased as the level in the waste bunker increased. This may be due to poor combustion conditions due to inhomogeneity of the waste. Indeed, a higher waste level in bunker can make it difficult to obtain a proper mixing of the waste.
iv. HCl emissions were not linked to the combustion conditions. Its correlation with $SO_2$ suggest the presence of waste

fractions whose combustion yield high levels of both HCl and $SO_2$.

v. NOx and $SO_2$ concentrations increased with increasing temperature in the furnace, which can be overcome by temperature control in the furnace.

vi. Estimated time lags gave information about the sensor locations as well as plant dynamics in response to the control system structure and actions, which can be fine-tuned by using the time lags.

All in all, the method developed in this study gave correlations some which were established knowledge in the WtE field, demonstrating its credibility in a real case study that can motivate its application to datasets from different plants containing a different set of variables to confirm or discover correlations and unveil dependencies.

## Credit author statement

Cansu Birgen: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. Elisa Magnanelli: Conceptualization, Writing – review & editing. Per Carlsson: Conceptualization, Funding acquisition, Writing – review & editing. Michael Becidan: Conceptualization, Funding acquisition, Project administration, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work is part of the Waste-to-Energy 2030 project co-funded by industry and public partners and the Research Council of Norway under the EnergiX program (WtE 2030–280949). The authors would like to thank Jostein Mosby and Returkraft AS.

## References

[1] Malinauskaite J, Jouhara H, Czajczyńska D, Stanchev P, Katsou E, Rostkowski P, et al. Municipal solid waste management and waste-to-energy in the context of a circular economy and energy recycling in Europe. Energy 2017;141:2013–44.
[2] Scarlat N, Fahl F, Dallemand J-F. Status and opportunities for energy recovery from municipal solid waste in Europe. Waste and Biomass Valorization 2019;10(9):2425–44.
[3] Becidan M, Wang L, Fossum M, Midtbust H, Stuen J, Bakken J, et al. Norwegian waste-to-energy (WtE) in 2030: challenges and opportunities. Chemical Engineering Transactions 2015;43:2401–6.
[4] Tabasová A, Kropáč J, Kermes V, Nemet A, Stehlík P. Waste-to-energy technologies: impact on environment. Energy 2012;44(1):146–55.
[5] De Greef J, Villani K, Goethals J, Van Belle H, Van Caneghem J, Vandecasteele C. Optimising energy recovery and use of chemicals, resources and materials in modern waste-to-energy plants. Waste management 2013;33(11):2416–24.
[6] Makarichi L, Jutidamrongphan W, Techato K-a. The evolution of waste-to-energy incineration: a review. Renew Sustain Energy Rev 2018;91:812–21.
[7] Ge Z, Song Z, Gao F. Review of recent research on data-based process monitoring. Ind Eng Chem Res 2013;52(10):3543–62.
[8] Salgado CM, Azevedo C, Proença H, Vieira SM. Noise versus outliers. Secondary analysis of electronic health records. Springer; 2016. p. 163–83.
[9] Bhowmik S, Jelfs B, Arjunan SP, Kumar DK, editors. Outlier removal in facial surface electromyography through Hampel filtering technique. IEEE Life Sciences Conference (LSC); 2017: IEEE; 2017.
[10] Pearson RK, Neuvo Y, Astola J, Gabbouj M, editors. The class of generalized hampel filters. European Signal Processing Conference (EUSIPCO); 2015: IEEE; 2015.
[11] Pearson RK, Neuvo Y, Astola J, Gabbouj M. Generalized hampel filters. EURASIP Journal on Advances in Signal Processing 2016;2016(1):1–18.
[12] Sheugh L, Alizadeh SH, editors. A note on pearson correlation coefficient as a metric of similarity in recommender system. AI & Robotics (IRANOPEN); 2015: IEEE; 2015.
[13] Zhang D-Q, He P-J, Shao L-M. Potential gases emissions from the combustion of municipal solid waste by bio-drying. J Hazard Mater 2009;168(2–3):1497–503.
[14] Wenga T, Chen G, Ma W, Yan B. Study on corrosion kinetics of 310H in different simulated MSW combustion environment. The influence of SO2 and H2O on NaCl assisted corrosion. Corrosion Sci 2019;154:254–67.
[15] Akimoto F, Matsunami A, Kamata Y, Kodama I, Kitagawa K, Arai N, et al. Cross-correlation analysis of atmospheric trace concentrations of N2O, CH4 and CO2 determined by continuous gas-chromatographic monitoring. Energy 2005;30(2–4):299–311.
[16] Hong Y-Y, Satriani TRA. Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network. Energy 2020;209:118441.
[17] Gaughan E, Fitzgerald B. An assessment of the potential for Co-located offshore wind and wave farms in Ireland. Energy 2020:117526.
[18] Roy S. The maximum likelihood optima for an economic load dispatch in presence of demand and generation variability. Energy 2018;147:915–23.
[19] Roy S. A technical perspective on variability costs: dependence on power variability and cross-correlations. Energy 2020:117350.
[20] Gut A. An intermediate course in probability. 2 ed. New York: Springer-Verlag; 2009.
[21] Birgen C, Magnanelli E, Carlsson P, Skreiberg Ø, Mosby J, Becidan M. Machine learning based modelling for lower heating value prediction of municipal solid waste. Fuel.283:118906.
[22] Pearson RK. Outliers in process modeling and identification. IEEE Trans Contr Syst Technol 2002;10(1):55–63.
[23] Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J Exp Soc Psychol 2013;49(4):764–6.
[24] Suomela J. Median filtering is equivalent to sorting. 2014. arXiv preprint arXiv:14061717.
[25] Yu H, Khan F, Garaniya V. A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes. AIChE J 2016;62(5):1494–513.
[26] Taylor R. Interpretation of the correlation coefficient: a basic review. J Diagn Med Sonogr 1990;6(1):35–9.
[27] Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018;126(5):1763–8.
[28] Xiao C, Ye J, Esteves RM, Rong C. Using Spearman's correlation coefficients for exploratory data analysis on big dataset. Concurrency Comput Pract Ex 2016;28(14):3866–78.
[29] Daniels SL. Products of incomplete combustion ($O_x$, $CO_x$, $HO_x$, $NO_x$, $SO_x$, $RO_x$, $MO_x$ and $PO_x$). J Hazard Mater 1989;22(2):161–73.
[30] Tang Y, Ma X, Lai Z, Zhou D, Lin H, Chen Y. $NO_x$ and $SO_2$ emissions from municipal solid waste (MSW) combustion in $CO_2/O_2$ atmosphere. Energy 2012;40(1):300–6.
[31] Hampartsoumian E, Nimmo W, Gibbs B. Nitrogen sulphur interactions in coal flames. Fuel 2001;80(7):887–97.