DBSCAN for improved oscillations diagnosis in mineral processing plants

E.C. Nienaber[a], B.S. Lindner[a]*, L. Auret[a], J.W.D. Groenewald[b]

[a] *Stone Three Digital, 24 Gardner Williams Avenue, Paardevlei, Somerset West, South Africa*

[b] *Anglo American Platinum Limited, Johannesburg, South Africa*

**Corresponding author: brian.lindner@stonethree.com*

ABSTRACT

Mineral processing plants have complicated control strategies, with different control layers, multiple interacting loops, and feedforward disturbance rejection. The complicated control strategies mean that oscillations can arise and propagate to multiple loops and degrade the control performance. Oscillation detection techniques, such as the spectral envelope method, can be applied to detect oscillations within an observation window. However, the start- and end-points of the oscillations within the observation window cannot be identified with current oscillation detection techniques. The start- and end-points would be useful to label the oscillatory intervals, and then to use these labels to identify process conditions associated with the oscillation.

In this paper, a pattern recognition technique, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), was used to label oscillation intervals. An oscillation diagnosis workflow combining spectral envelope and DBSCAN was demonstrated on an industrial case study of oscillations in a flotation circuit concentrate sump. The workflow successfully detected the oscillation, labelled the oscillatory intervals, and enabled accurate root cause analysis. The impact of time series features, namely the signal power ratio (PR) and the oscillation frequency, on the performance and optimal parameter selection of DBSCAN was also investigated. Low oscillation frequencies and low PRs negatively affected the performance of DBSCAN. Additionally, the results indicated that higher oscillation frequency was related to higher optimal values for the lag parameter.

Keywords: Oscillation diagnosis, process monitoring, data mining, clustering, flotation control

1. Introduction

Oscillations can arise in multiple control loops and degrade the control performance in mineral processing plants. Oscillation diagnosis techniques have been developed to detect the presence of oscillations to aid root cause analysis. The detection ability of these techniques is robust (Dambros et al., 2019). The spectral envelope method, demonstrated by Jiang et al. (2007) for plant-wide oscillation diagnosis, provides an oscillation contribution index, which gives an indication of the relative contribution of each oscillating variable to the oscillation. Additionally, the spectral envelope method provides an oscillation test statistic for each variable, to test the hypothesis of whether that variable is oscillating at the frequency under consideration. However, none of the available techniques provide an indication of the oscillation interval, when the oscillatory conditions started and ended (Dambros et al., 2019). Highlighting these intervals can enable the rapid and focused investigation of the process conditions associated with the oscillation. Identifying oscillation intervals would also allow calculation of the percentage of time a single variable spent in an oscillating state, to aid tracking and reporting of control and process health.

Pattern recognition techniques can be applied to time series data to label intervals of data with similar features, such as similar oscillation frequencies. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a data mining approach designed for pattern recognition in spatial data (Ester et al., 1996).

This paper demonstrates the use of the oscillation diagnosis workflow incorporating the spectral envelope and DBSCAN for root cause analysis in an industrial case study. The oscillation intervals are used to diagnose the process conditions associated with the oscillations. This paper also investigates the use of DBSCAN to identify oscillation intervals in multiple simulated oscillatory signals to test the performance of DBSCAN, and to investigate the relationship between optimal parameters and features of the oscillatory time series.

2. Methods used for oscillation diagnosis

The spectral envelope method for oscillation detection, DBSCAN for oscillation diagnosis, and the workflow combining the two methods, are described in this section.

2.1. Spectral envelope

The spectral envelope technique uses the cross power spectral density to identify peak power frequencies in time series. The detailed methodology is presented in Jiang et al. (2007). A multivariate data matrix ($X \in R^{m \times n}$, where *m* is the number of variables, and *n* is the number of samples) is analysed using the power spectral density matrix $\boldsymbol{P_x}$. Calculate the spectral envelope, ($\lambda(\omega)$, where ω represents frequency) by computing the eigenvalues of $\boldsymbol{P_x}$. Peaks of the spectral envelope at different frequency locations represent oscillations in the input data. An important hyperparameter for the spectral envelope is the significance threshold for peaks in the spectral envelope. See Duan et al. (2014) for details.

Thornhill (2003) lists the ideal requirements for an oscillation diagnosis technique: (1) be able to detect one or more oscillations; (2) assess the oscillations magnitudes for the identification of critical oscillations; and (3) evaluate the periods of oscillations for the plant-wide analysis. The spectral envelope technique satisfies requirements (1) and (2), but fails requirement (3). It cannot give an indication of when the oscillation started or ended.

2.2. DBSCAN

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering algorithm that groups data points according to sample density, as the algorithm loops through each data point $x \in X$ (X is a given feature set with dimension $X \in R^{m \times n}$). At each iteration, a viable seed point, x_p, is selected based on data point density within the neighbourhood C_{x_p} of x_p ($C_{x_p} \subset X$). A candidate cluster C_{x_p} is formed by grouping points x_j according to its proximity to $x_p - D(x_j, x_p) < \epsilon$ (where $\epsilon \in R^+$). C_{x_p} is only considered a candidate cluster if $N_{x_p} > minPts$ (where $\left|C_{x_p}\right| = N_{x_p}$ and $N_{x_p}, minPts \in Z$). Neighbouring data points, x_j, are added to until border points are reached ($\left|C_{x_j}\right| < minPts$). When iterations are completed, each x will belong either to a cluster $C = \{1,2,\ldots\}$ or will be classified as noise. $|C|$ can be an arbitrary number which can change depending on changes in X, ϵ or $minPts$. Further details are provided in Ester et al. (1996) and Schubert et al. (2017).

The parameters important for optimal clustering of oscillations using DBSCAN are: ϵ, the maximum distance between two samples for one to be considered as in the neighbourhood of the other; $minPts$, the number of samples in a neighbourhood for a point to be considered as a core point, including the point itself; and *n_lags*, the number of lagged timeseries vectors to include (to a feature set). The feature set passed to the DBSCAN algorithm contains the original timeseries vector in the first column, and in subsequent columns lagged versions of this timeseries vector. Each additional column is lagged by 1 sample (with lags in the range of 1 to *n_lags*).

2.3. Oscillation diagnosis workflow

A workflow combining spectral envelope and DBSCAN for oscillation diagnosis is presented in Figure 1. The input data in the data window ($X \in R^{m \times n}$) is analysed using spectral envelope to detect the presence of oscillations. For each oscillation identified, the algorithm provides a list of oscillating variables, as well as the oscillation contribution index (OCI). The OCI provides initial diagnostic information, as an indication of which variables' oscillations are the most severe. DBSCAN classification is then performed on each oscillating variable to identify their oscillating intervals. These intervals are then used to identify process conditions associated with the oscillations. The oscillation intervals can also be used to calculate the percentage of time each variable spent in the oscillating state, to track process and controller health.

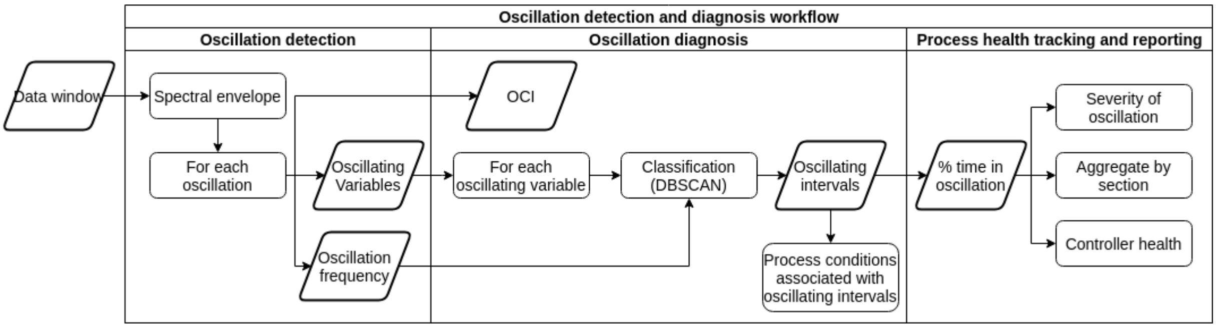

Figure 1: Oscillation diagnosis workflow

3. Oscillation diagnosis workflow demonstrated on oscillations in a concentrate sump

This section demonstrates the use of the oscillation diagnosis workflow incorporating the spectral envelope and DBSCAN for root cause analysis in a case study of oscillations in the concentrate sump of an industrial flotation circuit.

In this flotation circuit, the concentrate from two flotation cells in series flow into the concentrate sump. The first flotation cell is fed from a buffer tank just before the flotation circuit. The variables of interest for this case study were those associated with the flow of slurry from the buffer tank (Tank 1), the two flotation cells (Cell 1 and Cell 2), and the concentrate sump (Tank 2). This is a total of 10 variables, refer to Table 1 for a list of these input features. Following the workflow in Figure 1, a 7-hour observation window was chosen, which is long enough to capture normal operational variation, as well as abnormal operation caused by low and high frequency oscillations. The measurements were available at 10s sampling time in the process historian. Subsampling was not performed, so that the highest possible resolution could be captured. The input data for the spectral envelope method was therefore a matrix $X \in R^{10\times2520}$.

The spectral envelope method detected two oscillations, with frequencies of 7. 021 kHz (period of 2.4 minutes) and 0.000873 Hz (period of 20 minutes). This is shown by the red markers in in Figure 2. The other peaks present on the spectral envelope may represent harmonics of the frequencies at the detected peaks. Alternatively, the peaks may indicate missed detection of independent oscillations.

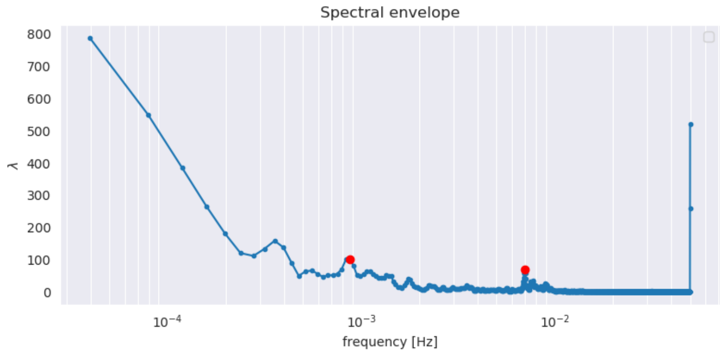

Figure 2: Spectral envelope, $\lambda(\omega)$, for concentrate sump variables

The results of the hypothesis test, and the OCI are shown in Table 1. The threshold statistic follows a χ^2 distribution with two degrees of freedom. This means that for a 99.9% confidence interval, a value larger than 13.8 indicates that the variable is oscillating at the frequency under consideration.

Table 1: OCI and statistic for χ^2 test for detected oscillations oscillation

OCI for Variable Name	0.87 Hz (19 min)	7.02 Hz (2.4 min)
pump_speed_tank1	0.05	0.01
pump_flow_tank1	0.06	0.00
pump_density_tank1	0.44	0.01
valve_pos_cell1	0.17	0.00
level_cell1	1.03	0.02
valve_pos_cell2	0.32	0.02
level_cell2	1.20	0.03
pump_flow_tank2	0.19	1.24
level_tank2	0.14	0.00
pump_current_tank2	0.12	1.17

Stat for Variable Name	0.87 Hz (19 min)	7.02 Hz (2.4 min)
pump_speed_tank1	5.3	0.8
pump_flow_tank1	5.2	0.4
pump_density_tank1	23.3	1.4
valve_pos_cell1	116.7	0.0
level_cell1	5306.0	0.5
valve_pos_cell2	97.1	0.3
level_cell2	3998.2	1.2
pump_flow_tank2	5.0	30771.2
level_tank2	2.4	0.1
pump_current_tank2	5.3	23484.8

The levels of cell 1 and cell 2 showed large OCI values for the 20-minute oscillation, indicating large contributions to this spectral envelope peak. Inspection of their time series' showed that regular repeating spikes in the levels, due to flow disruptions to the cells, caused this peak. No clear performance degradation was linked to this behaviour, so this event was not investigated further.

The peak in the spectral envelope at 7.02 kHz (2.4 minutes) represented a persistent recurring oscillation (see Figure 2). The hypothesis test in Table 1 indicated that only pump_flow_tank2 and pump_current_tank2 were oscillating at the 7.02 kHz frequency. The OCI shows that pump_flow_tank2 and pump_current_tank2's time series had the highest relative contribution to the oscillation. This pump transports the concentrate produced in this section. Variation in this important product stream represents serious performance degradation.

Figure 3 shows the time series of the oscillating variable for the observation window. The figure shows a few oscillation intervals, but the oscillation does not persist for the entire window. The spectral envelope method can only detect the presence of an oscillation within the observation window.

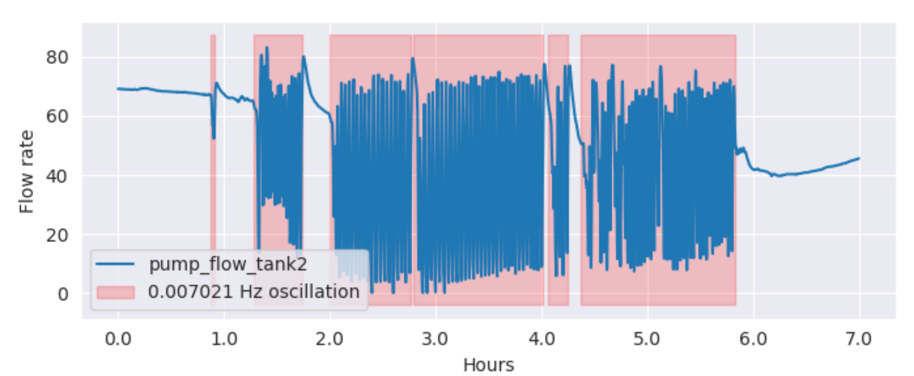

Figure 3: Oscillation intervals identified for oscillating variable using DBSCAN. Repeating peaks 2.4 minutes apart, which confirms the spectral envelope result of 2.4 min (see Figure 4 for zoomed-out view).

The next step in the workflow in Figure 1 is to use DBSCAN to find clusters in the time series of the oscillating variable that show the oscillation intervals. Figure 4 highlights the resulting oscillation intervals found using DBSCAN. The results showed that the variable was oscillating for 59.43% of the 7-hour window. Further investigation showed that this same oscillation arose multiple times over the course of a few days. DBSCAN was used to find the oscillating intervals over a longer observation window. Figure 3 shows the resulting accurate identification of the oscillating intervals.

Visual inspection shows that DBSCAN accurately classified the oscillation intervals. In this example thresholding on the gradient of the time series would give comparable results. However, in practice a threshold would need to be automatically selected, and some method to merge segmented events would be required. Figure 4 shows the same signal for a longer observation window, where the signal's mean and variance changes in normal operation, even without the presence of oscillations. Simple thresholding on this signal would not automatically identify the oscillating intervals. A frequency based method such as wavelet transforms could also be used to solve this problem, however, these techniques would require additional user input to select the appropriate waveform. The automated pattern matching of DBSCAN is generalisable to different signals and dynamic operation, requiring minimal user intervention to identify oscillating intervals.

Further investigation showed that this same oscillation arose multiple times over the course of a few days. DBSCAN was used to find the oscillating intervals over a longer observation window. Figure 4 shows the resulting accurate identification of the oscillating intervals.

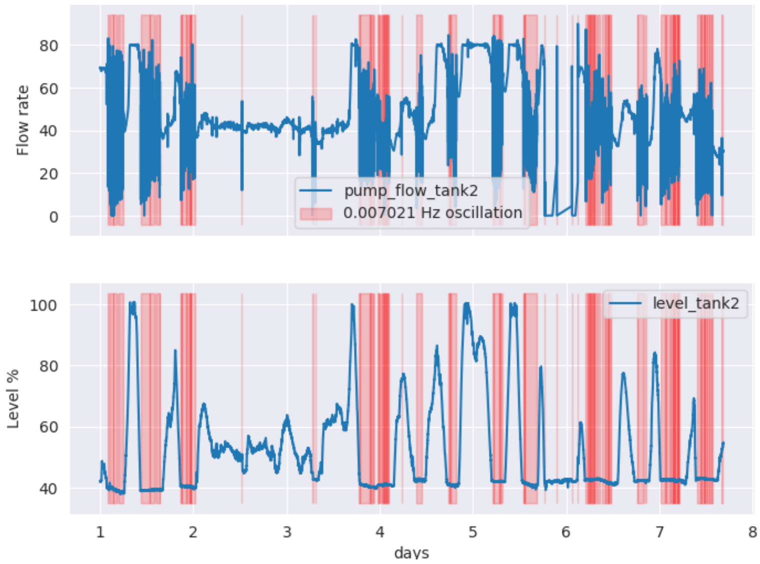

Figure 4: Oscillating intervals for the pump flow rate oscillation, with tank level superimposed.

These oscillating intervals can highlight process conditions during these oscillations, for root cause analysis. Figure 4 shows the time series of the tank level with the oscillating intervals highlighted. The plot shows that the oscillating intervals corresponded with intervals where the tank level was below 45%. Site personnel confirmed that the oscillating pump flow rate was consistent with the pump sucking in air, indicating accurate root cause analysis results. This means that the low level may have caused the pump to suck in air. The lower limit of the tank level could be set higher to avoid this in future. The combination of spectral envelope and DBSCAN, following the workflow presented in Figure 1, was used to successfully identify the process conditions that lead up to the repeated oscillations.

4. Impact of time series features on DBSCAN performance and optimal parameter selection

The previous section demonstrated the effective use of DBSCAN in an oscillation diagnosis strategy. However, DBSCAN's performance is sensitive to features of the signal being analysed, as well as optimal parameter selection. This section investigates the use of DBSCAN to identify oscillation intervals in multiple oscillatory signals to determine: the performance of DBSCAN; and the relationship between optimal parameters and features of the oscillatory time series.

Twenty examples of oscillatory signals from measured variables in an industrial flotation circuit were analysed using DBSCAN. Examples of the signals used are shown in Figure 5. The oscillation intervals in these examples were hand-labelled to define the ground truth. The properties of these examples that were investigated were the Power Ratio (PR – see Appendix A), and the oscillation frequency. These properties are post-classification results to summarise DBSCAN performance.

4.1. Impact of time series features on oscillation interval identification performance

Figure 5 shows examples of the test signals which exhibit the best and worst selectivity scores (for the DBSCAN oscillation interval identification performance). With the examples shown, good performance was typically observed where the PR was higher. When the normal signal was smooth compared to the oscillatory part of the signal, DBSCAN correctly identified the oscillation interval. Additionally, DBSCAN performed better at labelling the examples with higher frequency oscillations.

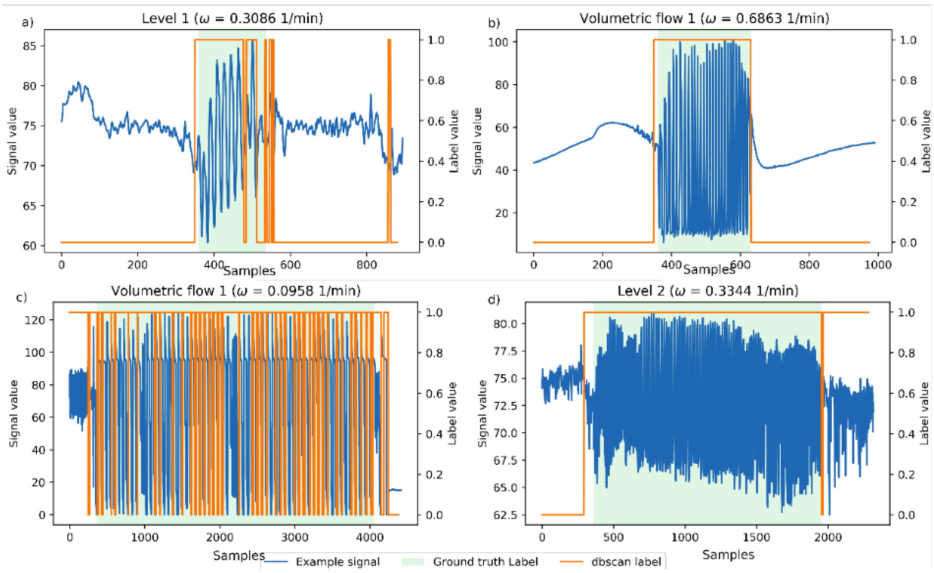

Figure 5: Sample of test signals showing best (a and b) and worst (c and d) DBSCAN oscillation interval identification performance. 'dbscan label' signal: 0 indicates no oscillation, 1 indicates oscillation.

Figure 6 shows the impact of PR on the performance of the DBSCAN method. Note that error bars are included in Figure 6 and Figure 7 to show the standard deviation for repeated experiments. The random search employed in the parameter optimization strategy introduced the variance observed in the performance, as well as the optimal parameter selection. In Figure 6, higher precision, recall and selectivity are generally achieved for higher sample PR values. Higher PRs mean that the range of the signal during oscillatory intervals was much higher than that of the normal signal. Therefore, DBSCAN was able to distinguish oscillatory points from normal points. This confirms findings in the discussion of Figure 5, where the trends showing good performance had higher PR scores. These results prove the robustness of DBSCAN to diverse types of signals.

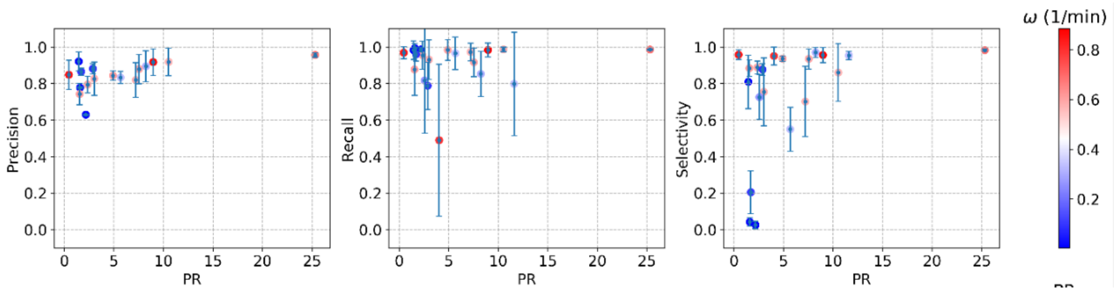

Figure 6: Impact of PR on DBSCAN interval identification performance. Higher PR and frequency show improved selectivity. Error bars show the standard deviation for repeated experiments.

The impact of oscillation frequency on DBSCAN interval identification performance is also captured in Figure 6. Higher frequency oscillations show improved selectivity, but no impact on precision and recall. Considering only precision and recall, it seems that DBSCAN performed consistently well, showing high values for both precision and recall. However, the poor selectivity indicates that DBSCAN was not always able to discriminate between oscillatory and non-oscillatory points, labelling both as oscillatory (i.e. increasing false positives). The examples in Figure 5.c) and d) demonstrate poor selectivity, where normal and oscillatory points were labelled as oscillatory.

4.2. Impact of oscillation frequency on optimal lag parameter selection

Figure 7 plots the optimal lag parameter against the oscillation frequency. The figure shows that at higher frequencies the optimal lag parameter is lower. This indicates that when the oscillation peaks are closer to each other, the time span covered by the number of lags can be smaller to still capture enough of the oscillatory trend for DBSCAN to separate the oscillatory data from other periods of data. At lower frequencies, where the oscillation peaks are further apart, the number of lags has to be increased to capture the oscillatory trend.

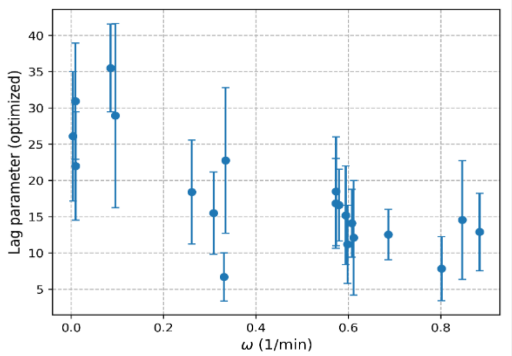

Figure 7: Relationship between optimal lag parameter (n_lags) and the frequency of the observed oscillation. Error bars show the standard deviation for repeated experiments.

5. Conclusions

The use of spectral envelope and DBSCAN for oscillation diagnosis was demonstrated on an industrial case study involving oscillations in a concentrate sump of a flotation circuit. The workflow successfully detected the oscillation, labelled the oscillatory intervals, and enabled correct accurate cause analysis.

The case study of the oscillation in the concentrate sump consisted of only one oscillation, and one oscillatory variable. The techniques are applicable to cases where multiple simultaneous oscillations are present in multiple signals. Future work can demonstrate and test such cases. This would also allow the process health tracking feature of the workflow to be demonstrated.

The impact of time series features, such as the signal to noise ratio, and the oscillation frequency, on the performance and optimal parameter selection of DBSCAN was investigated. The performance of DBSCAN was shown to be affected by the low oscillation frequencies and the low PRs. Additionally, higher oscillation frequency was shown to be related to higher optimal values for the lag parameter.

Future work can focus on the impact of other signal features not considered in this paper on the performance and optimal parameter selection of DBSCAN. For example, the nonlinearity index of the signals can be calculated, using the metric presented by Thornhill (2005), and the impact of the nonlinearity of a signal on DBSCAN can be determined. Section 4 demonstrated a link between the optimal lag and the oscillation frequency. Future work can focus on how this link can be used to inform the optimal parameter selection in advance.

6. References

Dambros, J.W.V., Trierweiler, J.O., Farenzena, M., 2019. Oscillation detection in process industries – Part I: Review of the detection methods. J. Process Control 78, 108–123. https://doi.org/10.1016/j.jprocont.2019.04.002
Duan, P., Chen, T., Shah, S.L., Yang, F., 2014. Methods for root cause diagnosis of plant-wide oscillations. AIChE J. 60, 2019–2034.
Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96. AAAI Press, Portland, Oregon, pp. 226–231.
Jiang, H., Shoukat Choudhury, M.A.A., Shah, S.L., 2007. Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method. J. Process Control 17, 143–155.
Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. ACM Trans. Database Syst. TODS 42, 19:1–19:21. https://doi.org/10.1145/3068335
Thornhill, N.F., 2005. Finding the source of nonlinearity in a process with plant-wide oscillation. IEEE Trans. Control Syst. Technol. 13, 434–443.
Thornhill, N.F., Cox, J.W., Paulonis, M.A., 2003. Diagnosis of plant-wide oscillation through data-driven analysis and process understanding. Control Eng. Pract., Award winning applications-2002 IFAC World Congress 11, 1481–1490. https://doi.org/10.1016/S0967-0661(03)00108-4

Appendix: Feature and performance calculations

Power ratio (PR) is calculated by taking the ratio of oscillating and non-oscillating signal segment's power (according to Equation 1).

$$PR = P_{osc}/P_{non-osc} \qquad \text{(Eq. 1)}$$

Where:

- P_{osc} is the power of the oscillating signal segment.
- $P_{non-osc}$ is the power of the non-oscillating signal segment.

Average power P in the time domain is approximated with Equation 2.

$$P = \frac{1}{N}\sum_{n=1}^{N}(x(n) - E(x(n)))^2 \qquad \text{(Eq. 2)}$$

Where:

- $x(n)$ is a signals amplitude at time n
- $E(x(n))$ is approximated as the mean of $x(n)$.

The performance of the oscillation interval identification was defined as precision, recall, and selectivity. These performance metrics are described in Table 2. All of these calculations were possible because the ground truth oscillatory intervals were known for each of the example signals.

Table 2: Interval identification performance metrics. T=True, F=false, P=Positive, N=Negative.

Metric	**Calculation**	**Description**
Precision	$\frac{TP}{TP+FP}$	Fraction of oscillatory points identified correctly labelled as oscillatory
Recall	$\frac{TP}{TP+FN}$	Fraction of actual oscillation points correctly labelled as oscillatory
Selectivity	$\frac{TN}{TN+FP}$	Fraction of non-oscillatory points correctly labelled as non-oscillatory