Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-14 (2019) 243–248

Controller Maintenance Prioritisation with LoopRank for a Milling and Flotation Plant

John T. McCoy*,, Lidia Auret*,****

** Department of Process Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa. (e-mail:jtmccoy@sun.ac.za); ** Stone Three Digital, 24 Gardner Williams Ave, Paardevlei, Somerset West, 7130, South Africa (email: john.mccoy@stonethree.com)*

Abstract: This paper describes the use of the LoopRank algorithm to determine the importance of control loops in large operating plants. Defining control loop importance is an important step towards the prioritization of control loop maintenance efforts, building on extensive existing research into controller performance monitoring, which can identify poorly-performing control loops. LoopRank uses a connectivity matrix, which can be extracted from operating data using techniques such as partial correlation or Granger causality. The algorithm was applied to synthetic data from a simulated tank network, and to industrial data from a minerals processing plant. For the simulated data, LoopRank with Granger causality provided no significant improvement over LoopRank with the relatively simple partial correlation results. However, for the industrial data, LoopRank with partial correlation was not able to correctly assign importance to variables which are critical to advanced process control, such as mill load, power and speed. LoopRank with Granger causality was able to identify these variables, suggesting that incorporating lagged information in the connectivity matrix provides better importance rankings than using simpler methods.

Keywords: Controller performance monitoring, Controller maintenance prioritisation, Minerals processing

1 INTRODUCTION

Significant effort has been made in the control literature to monitor controller performance, using a number of benchmarks and indices, as described in several comprehensive review papers (Bauer et al., 2016; Jelali, 2006; Shardt et al., 2012). However, relatively few publications present detail on the correction of poor performance, because of the uncertainty regarding which interventions are available, which intervention to select, and when to intervene. Another challenge is the difficulty of quantifying the economic impact of controller performance improvements in a large system.

The review papers mentioned above highlighted the need for methods to prioritise the loops which have been identified as poorly performing. This would allow controller performance monitoring efforts to be translated into maintenance actions.

Fig. 1 shows a proposed workflow for controller performance monitoring and maintenance prioritisation. Prioritisation can be based on measures of control loop importance, and/or the urgency of required maintenance.

Impact on economic performance would be the most useful measure of control loop importance, but defining economic performance functions (EPFs) and relating controller performance to EPFs can require extensive plant knowledge and models (Bauer et al., 2007; Olivier and Craig, 2017), and is difficult for industrial data (Siwella, 2017).

A number of papers have proposed a strong link between control loop importance and measures of loop connectivity, correlation or interaction (Farenzena et al., 2009; Farenzena and Trierweiler, 2009; Garg and Tangirala, 2014; Rahman and Shoukat Choudhury, 2011; Reis, 2014). The most promising of these approaches is LoopRank (Farenzena and Trierweiler, 2009), although the technique has received very little attention in the literature.

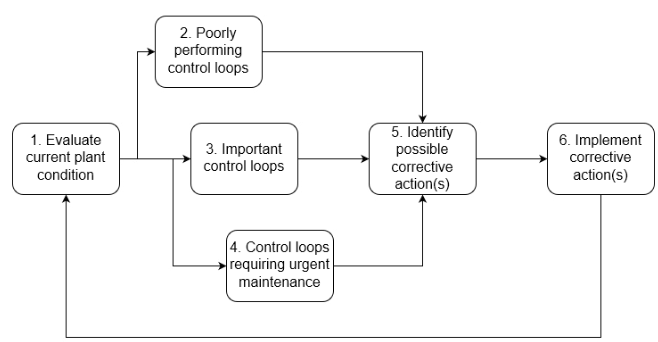

Fig. 1: Workflow for controller performance monitoring and maintenance prioritisation

Apart from recent work on the regulatability of process plants in the presence of faults (Olivier and Craig, 2017), the control literature does not seem to have developed methods to determine remaining useful life, as is prevalent in reliability engineering and condition-based maintenance.

This paper describes the LoopRank algorithm, and the application of the technique to a simulated tank network and an industrial milling and flotation plant.

2 BACKGROUND

The LoopRank algorithm was proposed as a modification of the method used to rank web pages by Google, the PageRank

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.09.194

algorithm (Bryan and Leise, 2006), specifically for ranking the importance of control loops on the basis of interactions (Farenzena and Trierweiler, 2009).

2.1 PageRank

The PageRank algorithm determines the importance of a number of web pages, with links between the pages. The importance score of the i-th page, x_i, is expressed as the weighted sum of the importance scores of the pages which link to x_i; this set of pages is denoted L_i. The importance score of each page x_j linking to x_i is divided by n_j, the number of links from x_j to other pages. It is assumed that pages do not link to themselves (or if they do, these links are excluded). Each score is then as shown in equation (1).

$$x_i = \sum_{j \in L_i} \frac{x_j}{n_j} \quad (1)$$

For a system of n web pages, the importance scores for each page can be expressed in matrix form, as shown in equation (2). The vector of scores is denoted $\boldsymbol{x}$, and $\boldsymbol{A}$ is an $n \times n$ matrix containing the connections between each page, referred to as the link matrix. The diagonal elements are zero (since pages do not link to themselves) and the elements in each column sum to one (since the scores are weighted by the number of links leaving each page).

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} \quad (2)$$

This set of equations can be solved by eigendecomposition: the eigenvector associated with the eigenvalue $\lambda = 1$ is the solution to equation (2), since eigenvectors satisfy $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. As discussed in (Bryan and Leise, 2006), link matrices with columns summing to one, and zeros on the diagonal, are known as "column stochastic" matrices, and will always have an eigenvalue equal to one.

Multiple eigenvalues equal to one can occur if the link matrix, $\boldsymbol{A}$, describes multiple subnetworks of pages, which do not have interconnections; there are then multiple solutions to equation (2), and a unique ranking cannot be found.

A single eigenvalue equal to one, with a unique solution to equation (2), and therefore a unique ranking, can be ensured by adjusting the link matrix, $\boldsymbol{A}$, as described in (Bryan and Leise, 2006). The link matrix, $\boldsymbol{A}$, is replaced by the matrix $\boldsymbol{M}$, calculated in equation (3), in which m is a number between 0 and 1 (the default value is 0.15), and $\boldsymbol{S}$ is an $n \times n$ matrix where each element is ${}^{1}/_{n}$.

$$\boldsymbol{M} = (1 - m)\boldsymbol{A} + m\boldsymbol{S} \quad (3)$$

In the case where the matrix $\boldsymbol{A}$ describes a set of disconnected subnetworks, replacing $\boldsymbol{A}$ by the matrix $\boldsymbol{M}$ guarantees a unique solution, and allows comparisons of the importance scores of pages within and between disconnected networks.

If there are pages which do not link back to any other page in the network (known as "dangling pages"), then both $\boldsymbol{A}$ and $\boldsymbol{M}$ will be "column substochastic", and will have a positive eigenvalue close to one; the associated eigenvector still provides a solution to the ranking problem for this case.

2.2 LoopRank

LoopRank, as described in (Farenzena and Trierweiler, 2009), replaces the link matrix, $\boldsymbol{A}$, with a connectivity matrix determined from operating data (specifically, the controlled variables in each control loop) using partial correlation. The underlying assumption is that the partial correlation connectivity matrix provides a similar form of input to the method for control loops as the link matrix does for web pages: that the importance of a control loop is related to the number and strength of the connections with other loops.

The output from partial correlation analysis, an $n \times n$ matrix for a system of n control loops, will be a symmetric matrix with ones on the diagonal; for the LoopRank algorithm, the diagonals are set to zero, and each column normalised to sum to one.

The replacement of the connectivity matrix by the matrix $\boldsymbol{M}$ was not described in (Farenzena and Trierweiler, 2009), but this modification of the algorithm was included here (using the default value for m of 0.15) for two reasons.

First, it is likely that any system of control loops will have at least one control loop which does not directly link to other loops in the system (for example, a loop controlling flow to a downstream process).

Second, it is possible that control loop importance scores may be required for disconnected subsystems of control loops (for example, parallel sections of the plant), to prioritise maintenance. The replacement of the connectivity matrix by the matrix $\boldsymbol{M}$ ensures a unique solution, which allows direct comparison of the importance of control loops, even when they are in subsystems with no direct connections.

3 METHODS

In this work, LoopRank has been implemented in MATLAB, as described above. The code has been made available (github.com/ProcessMonitoringStellenboschUniversity/IFAC-LoopRank).

LoopRank is applied to two datasets (described in more detail in section 4), using connectivity matrices derived from partial correlation, as in (Farenzena and Trierweiler, 2009), and from Granger causality (Granger, 1988; Li et al., 2016). Granger causality is included to investigate whether connectivity matrices derived from causal methods (which include time lag information) are more useful for estimating control loop importance than those derived from simple partial correlation.

For the synthetic dataset, which is a simulated network of tanks with level controllers, link matrices were also manually derived, to provide a ground truth for comparison with the results from the connectivity matrices.

The link matrix was calculated using the physical connections between the tanks, with the weighting determined from the ratio of the steady state flow rates out of the tank. This is expected to give a physically realistic indication of the importance of the control loops, as the greater the flow rate from one tank to the next, the greater the potential impact of changes in that flow rate on the downstream tank.

3.1 Partial correlation

Given a data matrix $\boldsymbol{X}$, with m observations of n variables, partial correlation analysis determines the sample linear correlation between each pair of variables in $\boldsymbol{X}$, controlling for the correlations with the other variables. Partial correlation was calculated using MATLAB's built in "partialcorr" function, and the default settings for the function.

3.2 Granger causality

Granger causality determines the causal relationships between variables, based on the intuition that if including past values of variable x in an autoregressive model of variable y improves the predictions of the future values of variable y, compared to the pure autoregressive model of y, then variable x is said to have a causal influence on variable y (Granger, 1988).

In this study, Granger causality is implemented as described in (Wakefield et al., 2018); in particular, the specification of the maximum autoregressive model order as a multiple of a representative process time constant, and the use of early stopping of the model order search after a certain number of increasing values of the Akaike Information Criterion (AIC). Full details are provided in (Wakefield et al., 2018), and a description of the method and open source version of the Granger causality code available in (Lindner et al., 2017a, 2017b).

4 DATASETS

4.1 Simulated tank network

To provide a ground truth for tests of the LoopRank algorithm, synthetic data were generated for simulated tank networks, with differing numbers of tanks and connectivities. As shown in Fig. 2, the tank networks consist of two banks of tanks. Each tank has level control (with outlet flow as manipulated variable), and some tanks have underflow streams (with flow rate determined by tank level and an underflow constant proportional to tank volume). In the first bank, outlet flows are connected to subsequent tanks, and underflow streams are connected to the second bank. In the second bank, outlet flows can leave the system or connect to the subsequent tank, and underflows (where present) are recycled to Tank 1. The locations and connections of outlet flows and underflows in the second bank are determined randomly when creating each simulation.

Sensor noise and feed flow variations are included, and level control is implemented as simple proportional integral control, with tuning by direct synthesis (Chen and Seborg, 2002). See github.com/ProcessMonitoringStellenboschUniversity/IFAC-LoopRank for code and further details on the simulation.

4.2 Industrial milling and flotation circuit

Operating data for the primary and secondary milling and flotation circuit sections of a minerals processing plant (schematically shown in Fig. 3) were studied. The data included measurements of the set point, process value and manipulated value for the control loops in the plant, for two days at one minute intervals.

The process values for the various control loops were extracted from the data set. Variables with a missing value fraction higher than 5% were excluded from the analysis. Simple mean replacement was applied to variables with missing values in less than 5% of the observations; missing values were replaced with the mean of the observed values, with the addition of uniformly-distributed noise in the range of ±5% of the mean value.

To ensure that duplicate variables were not included, a check for linear dependencies between the variables was performed, and only linearly independent variables retained for analysis.

The resulting data was standardised (by mean subtraction and division by the standard deviation) before further analysis.

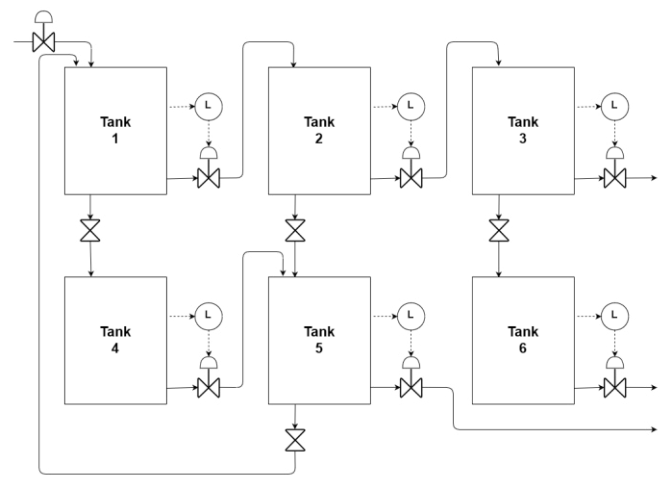

Fig. 2: Tank network with six tanks, level-controlled outflows, and underflows.

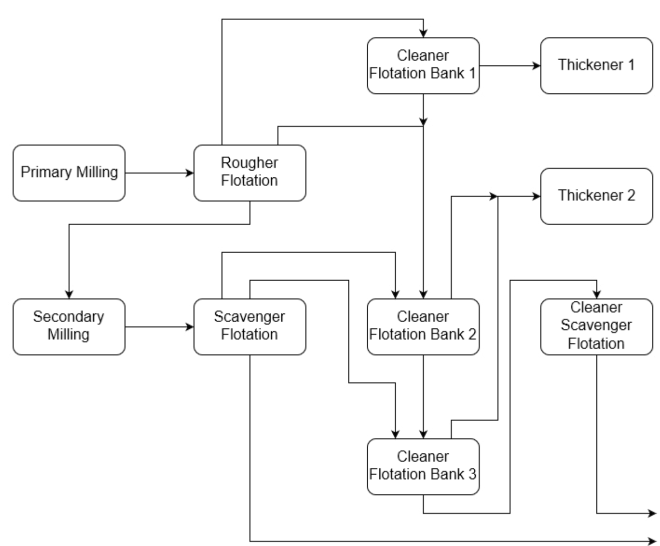

Fig. 3: Overview of the milling and flotation circuit portion of a minerals processing plant. Note that not all internal recycles are shown, for figure clarity.

5 RESULTS

5.1 Simulated data

In order to investigate the accuracy of the importance scores derived from partial correlation and Granger causality, a tank network was simulated multiple times. For each simulation, the root mean square error (RMSE) of the importance scores from partial correlation and Granger causality was calculated, using the importance scores calculated from known connections, with the flow ratio as weighting, as the ground truth.

The network was simulated with between 5 and 15 tanks; for each network size, sixteen replicates were performed, setting the random number generator seed value from 0 to 15. The results are shown in Fig. 4.

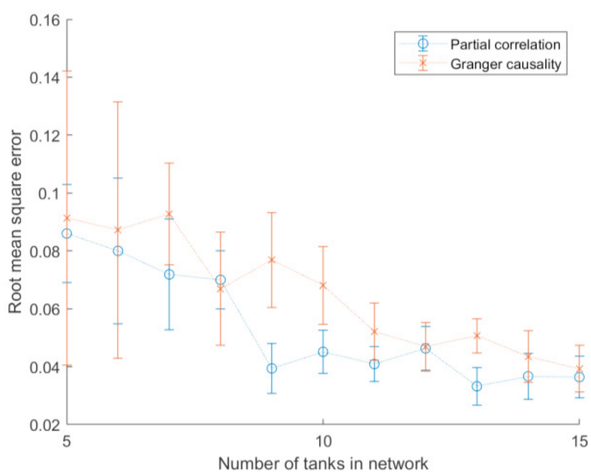

Fig. 4: Comparison of RMSE for partial correlation and Granger causality as generators of connectivity matrices for LoopRank algorithm. One standard deviation shown in error bars, based on 16 replicates for each network size. RMSE calculated using link matrix derived from known connections, with weightings from steady state flow rates.

For most network sizes, the RMSE values for partial correlation and Granger causality are almost indistinguishable (based on the means and one standard deviation); where the RMSE values differ by more than one standard deviation, the RMSE for scores for partial correlation is lower than that for Granger causality.

The results in Fig. 4 suggest that using a causal method which includes timelag information (such as Granger causality) does not significantly improve the estimation of control loop importance when using the LoopRank algorithm, compared to using a simple method like partial correlation, for networks of control loops with simple controllers. However, these conclusions should be tested for significantly larger systems, or networks with advanced control systems.

5.2 Industrial data

After preprocessing the industrial data the dataset consisted of 201 controlled variables. Although Granger causality is a relatively efficient method, the full dataset required several hours to complete the model order search, and converged on a model order equal to the maximum model order (approximately 150 minutes). This was not only time consuming, but also represented nonsensical results: most causal connections in a chemical plant will occur in significantly less time than 150 minutes.

To speed up the analysis and look for more sensible results, the data was separated into three subsets, based on plant sections: primary milling, secondary milling, and flotation. The primary milling dataset consisted of 32 controlled variables, the secondary milling dataset of 29 controlled variables, and the flotation dataset of 137 controlled variables.

The results of control loop importance analysis with LoopRank, using partial correlation and Granger causality to construct the connectivity matrices, are shown for the three plant sections in Fig. 5, Fig. 6 and Fig. 7. The discussion focusses on the results using partial correlation to generate the connectivity matrix, since the simulation study suggested that partial correlation may be preferable to Granger causality. However, the results using Granger causality are shown for completeness.

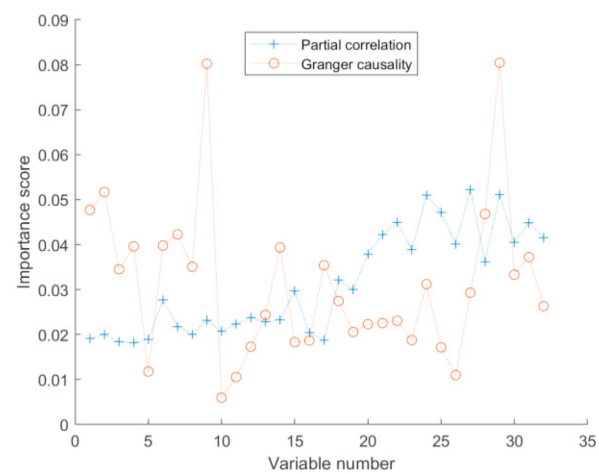

Fig. 5: Control loop importance scores for primary milling section

Based on the LoopRank results for the primary milling section, using partial correlation for the connectivity matrix, the most important control loops are the bearing oil pump pressure control loops (variables 18-27, but particularly 27 and 24) and one of the two sump density loops (variable 29, although variables 28, 30 and 31 are also relatively important, corresponding to sump density, level and dilution water flow). Mill water inlet (variable 32) also seems relatively important for this network of loops. The mill load, power and speed (variables 4-7) seem relatively unimportant, as do the cyclone pressures (variables 1 and 2), bearing oiling cooling (variable 3) and bearing oil flows (variables 8-17).

Interestingly, LoopRank with Granger causality highlights variable 29 (one of two sump density readings), which was ranked second by LoopRank with partial correlation, along with variable 9 (one of several bearing oil flows) and variables 1 and 2 (cyclone pressures), which were not highly ranked by LoopRank with partial correlation. These differences are likely due to the inclusion of time-lagged information in the Granger causality connectivity matrix, particularly where advanced process control systems may be present.

The results for LoopRank with partial correlation for the secondary milling section are shown in Fig. 6. Variables 21-24 (the last of ten bearing oil pressure loops) are the most important, followed by variables 17-20 (another four bearing oil pressure loops). The sump density, flow and level control loops (variables 26-28) also appear to be relatively important. Once again, the cyclone pressures (variables 1 and 2) are relatively unimportant.

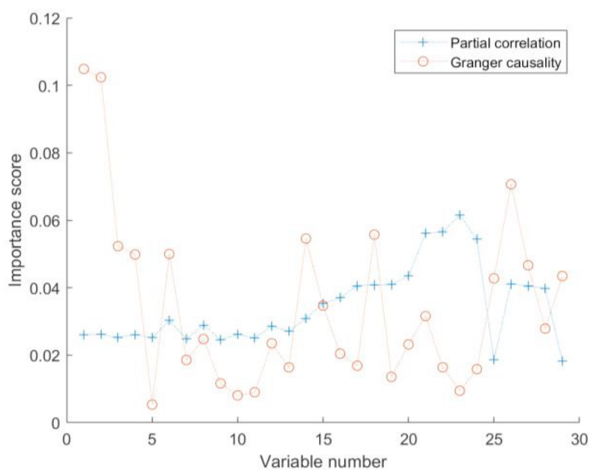

Fig. 6: Control loop importance scores for secondary milling section

These results are very similar to the most important loops for the primary milling section (shown in Fig. 5), although for secondary milling the sump dilution water flow (variable 29) is the least important loop. The sump screen water loop is also unimportant in this section.

Just as for the primary milling section, LoopRank with Granger causality suggests that the cyclone pressures (variables 1 and 2) and sump density (variable 26) are the most important loops in the secondary milling section. Counterintuitively, mill speed (variable 5) is ranked as the least important; this may be because safety control systems governing mill speed hide or complicate some of the relationships between mill speed and other variables, and so the data-based methods do not detect this loop as highly-connected or important.

Control engineers on the milling plant studied here indicated that LoopRank with partial correlation did not accurately capture the importance of loops which are critical to advanced process control (APC) systems, including mill load, power and speed, and cyclone pressures. However, a number of these variables were identified by LoopRank with Granger causality, suggesting that including timelagged information (as in causal techniques) may account for the connections present in APC systems.

For the flotation section, LoopRank with partial correlation ranks the loops with a relatively steady increase in importance with increasing variable number, presumably reflecting the physical layout of the plant. Although plant engineers did not provide any specific feedback on these results, some discussion and interpretation is provided below.

The most important loops are the sump density, flow and level control loops (variables 134-136) and the sump pump speed loop (variable 131), although the sump dilution water loop is not important. The blower pressure controllers (variables 132 and 133) are almost as important as the sump loops, presumably because of interactions with the operation of many of the flotation cells and their loops. Some of the level controllers in the scavenger 1 and 2 banks (particularly variables 128 and 129, and 112-113 and 115) are also highly ranked, presumably because these banks are heavily integrated with the rest of the system by recycle streams.

LoopRank with Granger causality shows the same general trend as LoopRank with partial correlation, with the sump density (variable 134) and sump pump speed (variable 131) being ranked as most important.

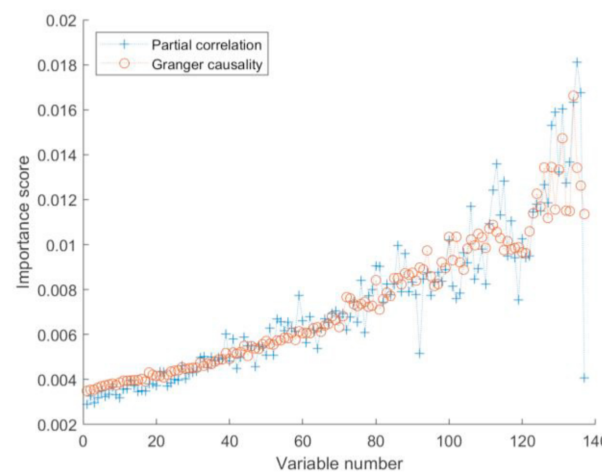

Fig. 7: Control loop importance scores for flotation section

6 CONCLUSIONS

This paper described the implementation of the LoopRank algorithm for determining the importance of control loops in a network (Farenzena and Trierweiler, 2009); interactions and connectivity are used to determine importance by the algorithm, which is based on Google's algorithm to rank web pages, PageRank (Bryan and Leise, 2006).

The algorithm was applied to synthetic data from a simulated network of tanks with level control. For this simulated system, with simple controllers, no significant improvement in the accuracy of the LoopRank algorithm occurs when using a causal technique (such as Granger causality), compared to using a simple method like partial correlation. Partial correlation, rather than Granger causality, is appealing since it is a simple and efficient method which is easy to implement (or available in most code environments) and has no hyperparameters; while Granger causality is more complex to implement and more computationally demanding to run, and the challenges of hyperparameter selection can lead to spurious and missing connections, as discussed in (Wakefield et al., 2018).

The LoopRank algorithm was also applied to industrial data, portioned into primary milling, secondary milling and flotation sections. For all three sections, LoopRank with partial correlation identified common control loops as the most important; these included loops around the sumps in all three sections, and the blowers in the flotation section. The cyclone pressures, and mill power, speed and load, were not highly ranked by LoopRank with partial correlation for both milling sections, although LoopRank with Granger causality identified these variables as important. These variables are critical to advanced process control systems, and so are expected to have a high importance.

LoopRank with partial correlation appears to be a promising technique for determining the importance of control loops in a network of many interacting basic control loops, and thus allow for the prioritisation of controller maintenance, in conjunction with existing controller performance monitoring techniques. However, further development of techniques to account for advanced process control systems may be required to make LoopRank more widely applicable.

7 ACKNOWLEDGEMENTS

The financial support of this research by Anglo American Platinum is acknowledged.

8 REFERENCES

Bauer, M., Craig, I.K., Tolsma, E., de Beer, H., 2007. A Profit Index for Assessing the Benefits of Process Control. Ind. Eng. Chem. Res. 46, 5614–5623. https://doi.org/10.1021/ie0614273

Bauer, M., Horch, A., Xie, L., Jelali, M., Thornhill, N., 2016. The current state of control loop performance monitoring – A survey of application in industry. J. Process Control 38, 1–10. https://doi.org/10.1016/J.JPROCONT.2015.11.002

Bryan, K., Leise, T., 2006. The $25,000,000,000 Eigenvector: The Linear Algebra behind Google. SIAM Rev. 48, 569–581. https://doi.org/10.1137/050623280

Chen, D., Seborg, D.E., 2002. PI/PID controller design based on direct synthesis and disturbance rejection. Ind. Eng. Chem. Res. 41, 4807–4822. https://doi.org/10.1021/ie010756m

Farenzena, M., Trierweiler, J.O., 2009. LoopRank: A Novel Tool to Evaluate Loop Connectivity. IFAC Proc. Vol. 42, 964–969. https://doi.org/10.3182/20090712-4-TR-2008.00158

Farenzena, M., Trierweiler, J.O., Shah, S.L., 2009. Variability matrix: A novel tool to prioritize loop maintenance, in: IFAC Proceedings Volumes. Elsevier, pp. 673–678. https://doi.org/10.3182/20090712-4-TR-2008.00109

Garg, A., Tangirala, A.K., 2014. Interaction assessment in multivariable control systems through causality analysis. IFAC Proc. Vol. 47, 585–592. https://doi.org/10.3182/20140313-3-IN-3024.00217

Granger, C.W.J., 1988. Some recent development in a concept of causality. J. Econom. 39, 199–211. https://doi.org/10.1016/0304-4076(88)90045-0

Jelali, M., 2006. An overview of control performance assessment technology and industrial applications. Control Eng. Pract. 14, 441–466. https://doi.org/10.1016/j.conengprac.2005.11.005

Li, G., Qin, S.J., Yuan, T., 2016. Data-driven root cause diagnosis of faults in process industries. Chemom. Intell. Lab. Syst. 159, 1–11. https://doi.org/10.1016/j.chemolab.2016.09.006

Lindner, B., Auret, L., Bauer, M., 2017a. Investigating the Impact of Perturbations in Chemical Processes on Data-Based Causality Analysis. Part 1: Defining Desired Performance of Causality Analysis Techniques. IFAC-PapersOnLine 50, 3269–3274. https://doi.org/10.1016/J.IFACOL.2017.08.463

Lindner, B., Auret, L., Bauer, M., 2017b. Investigating the Impact of Perturbations in Chemical Processes on Data-Based Causality Analysis. Part 2: Testing Granger Causality and Transfer Entropy. IFAC-PapersOnLine 50, 3275–3280. https://doi.org/10.1016/J.IFACOL.2017.08.620

Olivier, L.E., Craig, I.K., 2017. Should I shut down my processing plant? An analysis in the presence of faults. J. Process Control 56, 35–47. https://doi.org/10.1016/j.jprocont.2017.05.005

Rahman, A., Shoukat Choudhury, M.A.A., 2011. Detection of control loop interactions and prioritization of control loop maintenance. Control Eng. Pract. 19, 723–731. https://doi.org/10.1016/J.CONENGPRAC.2011.03.007

Reis, R.T.F., 2014. Multivariable controller performance monitoring. MSc Chemical Engineering, University of Coimbra, accessed at https://estudogeral.sib.uc.pt/bitstream/10316/40268/1/Multivariable controller performance monitoring.pdf.

Shardt, Y., Zhao, Y., Qi, F., Lee, K., Yu, X., Huang, B., Shah, S., 2012. Determining the state of a process control system: Current trends and future challenges. Can. J. Chem. Eng. 90, 217–245. https://doi.org/10.1002/cjce.20653

Siwella, M., 2017. Process monitoring with economic performance functions: Feasibility assessment for milling circuits. MEng thesis, Stellenbosch University, accessed at http://hdl.handle.net/10019.1/102992.

Wakefield, B.J., Lindner, B.S., McCoy, J.T., Auret, L., 2018. Monitoring of a simulated milling circuit: Fault diagnosis and economic impact. Miner. Eng. 120, 132–151. https://doi.org/10.1016/j.mineng.2018.02.007