INL/CON-23-71470-Revision-0

Changing the World's Energy Future

Anomaly Detection and Identification Using a Leave-One-Variable-Out Method

July 2023

Jacob A Farber, Ahmad Y Al Rashdan, Randall D Reese

INL/CON-23-71470-Revision-0

Anomaly Detection and Identification Using a Leave-One-Variable-Out Method

Jacob A Farber, Ahmad Y Al Rashdan, Randall D Reese

July 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Anomaly Detection and Identification Using a Leave-One-Variable-Out Method

Jacob A. Farber[*], **Ahmad Y. Al Rashdan**, **Randall Reese**

Idaho National Laboratory, Idaho Falls, Idaho

ABSTRACT

At nuclear power plants (NPPs), anomaly detection and identification (i.e., determining the causes of anomalies) are important tasks for ensuring the safe and efficient operation of NPPs. These tasks are currently labor-intensive and costly, and are made more difficult by the size and complexity of NPP systems. An alternative approach to conducting these tasks is to automate them, such as via the reconstruction-based contribution method, which is a well-researched unsupervised machine learning method that uses a data-driven model of anomaly-free behavior to detect events and then identify each variable's contributions to those events. The present effort developed a novel contribution approach that utilized a leave-one-variable-out (LOVO) model, with which each variable is predicted using all the other variables. The novelty lay in transforming this model into a reconstruction model and modifying the identification algorithm to work with the new reconstruction model. To evaluate this method in a controlled environment, a synthetic dataset based on spring-mass-damper (SMD) systems (commonly found in mechanical engineering references) was used, with known anomalies introduced into the system. The proposed method successfully detected the anomalies and afforded insights into their causes, thus enabling the appropriate identifications to be made.

Keywords: anomaly detection, anomaly identification, root cause analysis, nuclear power

1. INTRODUCTION

The standard approach to detecting process anomalies in nuclear power plants (NPPs) is primarily reactive in nature. This means plant operators do not search within time-series process data for subtle signs of anomalies, but rather wait until anomalies become significant enough to generate alarms. And once an anomalous condition is detected, the next step is to identify the cause of the anomaly (sometimes called root cause analysis) so that the condition can be corrected. Current approaches to identification are labor-intensive and costly, requiring careful manual study of the process variables, inspections and/or diagnostic testing, and evaluation of the results. More recently, NPPs have been investing in proactive approaches based on automated anomaly detection tools that could enable subtle signs of anomalies to be detected prior to those anomalies escalating into unexpected equipment failures. The present work focuses on automated approaches to both anomaly detection and identification.

For successfully detecting and identifying anomalies, a significant amount of research has been invested in reconstruction-based contribution methods. While these methods can be implemented as either unsupervised or supervised machine learning methods, this effort focuses on unsupervised machine learning. Reconstruction-based contribution methods try to recreate the anomaly-free data that would lead to a reduced anomaly score. To accomplish this, the data that exceed a user-defined threshold are modified by

[*]Jacob.Farber@inl.gov

subtracting amounts from some or all of the variables, then the modified data are run through the same normal model until the modified data's anomaly score falls below the threshold. This process is referred to as a reconstruction-based approach because it uses a reconstruction model (i.e., it takes measurement data as input, masks or compresses some of the information, and then tries to recover the input data from the reduced data) to reconstruct normal operating data (i.e., data with a reduced anomaly score).

The most popular reconstruction-based contribution approach uses dynamic principal component analysis (DPCA), which compresses the data to a specified latent size before attempting to recover them [1–4]. However, one challenge in implementing the DPCA method is the selection of the latent size, as an incorrect choice could result in poor performance.

In the present effort, a novel contribution-based method was developed based on a leave-one-variable-out (LOVO) model, with which each variable is predicted using all the other variables. The novelty of this work lay in transforming this model into a reconstruction model and modifying the identification algorithm to work with the new reconstruction model. In addition, this model entirely bypasses the need for latent size selection, thus removing the method's sensitivity to this hyperparameter.

To assess the results, it was necessary to have labeled time-series data containing both normal and anomalous behavior for a range of different anomaly types. This was achieved by generating synthetic data based on spring-mass-damper (SMD) systems, commonly found in mechanical engineering reference texts.

This paper is structured as follows. Section 2 details the LOVO contribution method, Section 3 discusses the synthetic data used to assess the method, Section 4 presents the results of applying the method to the synthetic data, and Section 5 discusses the conclusions reached.

2. METHODS

The LOVO anomaly detection and identification approach was implemented by creating a reconstruction model of the process parameters. This model takes measurement data as input, masks key information when making predictions, and then tries to recover the input data from the masked data. The model then creates an error term that can be used for anomaly detection. When the error term exceeds the predetermined threshold, the model identifies which sensors are causing the anomaly by assigning anomalous magnitudes to each sensor. This section discusses these steps in greater detail.

Time-series data can be modeled as a sequence of measurement vectors $x_t \in R^m$, each consisting of m sensor measurements at sample times $t \in \{1, 2, \ldots, n\}$, where n is the number of time steps in the data.

To set up the model, the data samples are separated into input and output matrices. This is done for each variable, in rotation. The input and output matrices are defined for each variable i in the vector x_t. The matrix definitions are simplified using the notations $x_{i,t}$ and $x_{-i,t}$, which represent variable i of x_t and all variables in x_t except for variable i, respectively. Then the input and output matrices for variable i are defined as:

$$X_i = \begin{bmatrix} x_{-i,1}^T \\ \vdots \\ x_{-i,n}^T \end{bmatrix}, \tag{1}$$

$$Y_i = \begin{bmatrix} x_{i,1}^T \\ \vdots \\ x_{i,n}^T \end{bmatrix}, \tag{2}$$

where X_i is the input matrix, Y_i is the output matrix, and x^T indicates the transpose of x. Using these matrices, each row of X_i is the input information used to predict each row of Y_i.

Once the input and output matrices are created, they can be used to train a linear regression model for each variable i. The regression model is of the following form:

$$\hat{x}_{i,t} = A_i x_{-i,t} + B_i. \tag{3}$$

Once all the models are solved for, the goal is to then transform the set of models into a single reconstruction model. Using matrix algebra, all these models for individual variables can be combined into a single model of the following form:

$$\hat{x}_t = Ax_t + B, \tag{4}$$

where A is formed by stacking the A_i terms, but with zeros padded along the diagonal to ensure that variable i is not included in that row, and B is formed by stacking the B_i terms.

In terms of anomaly detection, the model can be used to generate an anomaly score, which is a scalar value that quantifies the degree to which the augmented data are abnormal. When the score exceeds a certain threshold, the sample is considered anomalous. To determine the score, the prediction error is calculated as the difference between the measured and the estimated values:

$$e_t = x_t - \hat{x}_t = (I - A)x_t - B, \tag{5}$$

and the score is calculated as the weighted sum of the squared error:

$$\phi(x_t) = e_t^T \hat{\Sigma}^{-1} e_t, \tag{6}$$

where $\hat{\Sigma}$ is a diagonal matrix containing the estimated variances of the error vector that normalizes the error statistics.

The model is then used to generate reconstruction-based contributions that identify the extent to which each variable contributes to an elevated anomaly score. When an anomaly score exceeds the threshold, it means the model no longer does a good job of predicting the original data (i.e., the patterns learned from normal process behavior no longer hold true). As such, the idea is to modify the data until they can be better predicted by the model (i.e., until they better follow the patterns learned from normal process behavior). This analysis is accomplished by calculating a fault direction matrix and magnitude vector that modify the original data until the anomaly score falls below the threshold [3,4].

The reconstruction-based contribution problem can be framed as an optimization problem in order to determine the fault direction and magnitude that reduce the anomaly score back below the threshold. The fault direction and magnitude are used to define the modified measurement vector $x_t - \Xi_t f_t$, where Ξ_t is the fault direction matrix and is comprised of columns of the identity matrix, and f_t is the fault magnitude. By temporarily assuming a known fault direction matrix Ξ_t, the fault magnitude is calculated as the solution to the following optimization problem:

$$\arg\min_{f_t} \phi(x_t - \Xi_t f_t). \tag{7}$$

In other words, for a given fault direction matrix, the solution to this optimization problem finds the corresponding magnitude that modifies the data to align as closely as possible with the patterns learned from normal process behavior. The function $\phi(\cdot)$ is a matrix quadratic function and thus has a unique minimum value that can be calculated by setting its first derivative equal to zero [3]. The contributions of each variable can be calculated as $\Xi_t f_t$.

The reconstruction-based contribution method can be implemented as either an unsupervised or a supervised algorithm. If it were implemented as a supervised algorithm, known fault directions corresponding to known anomalies would already exist, and each direction would be used to solve the optimization problem until a solution was found that resulted in an anomaly score that came in below the threshold (i.e., a viable solution). These directions do not exist with unsupervised anomaly identification; instead, the idea is to find the simplest solution (i.e., the one with the fewest number of compromised variables) that results in such an anomaly score [4].

This unsupervised approach is implemented via combinatorial optimization in which all the simplest directions (i.e., single-variable directions) are tested first. If none of these directions work, then all directions involving two variables are tested. This process continues until a viable solution is found.

3. DATA

To demonstrate the proposed approach, this work used synthetic data based on an SMD system commonly seen in mechanical engineering references. The basic building blocks of this system are springs, masses, dampers, sensors, and actuators: the masses respond to forces, the springs apply restorative forces to the mass, the dampers apply damping forces to the mass, the actuators apply forces directly to masses, and the sensors measure the position of the mass. This simulator was used in previous efforts, and additional details on the equations of motion and the differential equations can be found in [5].

The present work used a system (see Figure 1) with three masses (labeled "m") connected in series, with the two end masses also being connected to fixed reference points (commonly called grounds). All connections used linear spring and damper components, meaning their forces were linearly proportional to their relative displacements and velocities, respectively. In addition, position sensors (labeled "s") and actuators that applied forces (labeled "f") were placed on each mass. Within this system, seven anomaly types were selected (highlighted in red in Figure 1), consisting of both process and sensor anomalies. Here, a process anomaly is defined as an anomaly that changes the physics of the system (e.g., pump wear or failure). Process anomalies in the SMD system occur on the spring and damper components of a process connection point between either two masses or a mass and a ground, and are labeled by their connection (e.g., a p12 process anomaly occurs between m1 and m2). A sensor anomaly is defined as an anomaly that changes the measurement data (e.g., temperature sensor bias or a failure that alters the data received). Sensor anomalies in the SMD system occur in the data from position sensors, and are labeled by their sensor name (e.g., an s0 sensor anomaly). Sample time-series data are shown in Figure 2, with the anomalous periods shaded in grey.

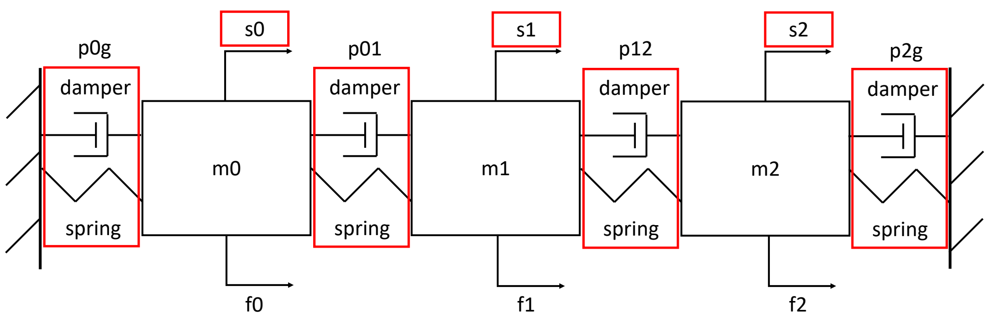


Figure 1: Sketch of the three-mass SMD system.

There are several reasons why the SMD system was selected for use in this research. First, though it is conceptually simple and features just a few basic components, those components can be combined to generate high-order systems with coupled variables. Second, the system is easily scalable to include many sensors and actuators. These two characteristics are important for emulating the large-scale, high-order systems in NPPs. Third, the system allows for straightforward incorporation of sensor anomalies (by directly modifying sensor measurements) and process anomalies (by modifying system parameters [in this case, the spring stiffness and damping coefficients]).

4. RESULTS

The LOVO anomaly detection and identification approach was applied to the SMD dataset. To assess the results, both the anomaly scores and contributions were plotted over the durations of a few different types of anomalies. In the contribution plots below, the top subplot shows the anomaly scores over the anomaly windows (plus a small buffer on both sides), and the bottom subplots provides stack plots showing the contribution magnitudes. Anomalies are detected through anomaly score increases during the anomaly, and are identified by matching their cumulative contributions over the duration of the anomaly to some intuitive understanding of the system behavior.

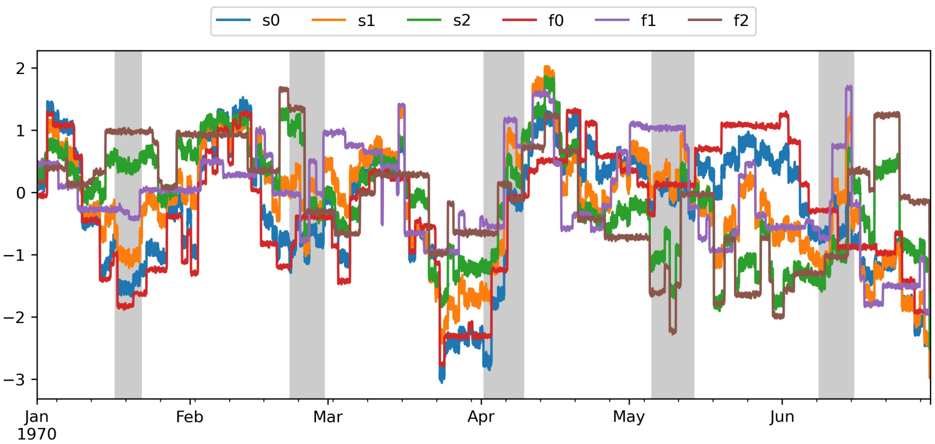


Figure 2: Time-series data for the SMD sensors over a period of a few months.

In looking at two different examples, Figure 3a first shows the results for a p12 process anomaly. The anomaly detection algorithm shows a clear increase in the anomaly score. The anomaly itself was inserted as a ramp change in the process parameter, which is reflected in the shape of the scores. From the identification algorithm, the contributions come from f1, s2, and f2. The fact that the results prioritize variables for both m1 and m2 strongly indicates that the anomaly occurs in the connection between those two masses. And given that the only anomaly based on this connection is a p12 process anomaly, these contributions make intuitive sense given an understanding of the system.

Secondly, the results for a s1 sensor anomaly are shown in Figure 3b. As with the p12 process anomaly, this s1 sensor anomaly was inserted as a ramp offset in the sensor value, and this is reflected in the shape of the scores. As such, this event is clearly detected. The contributions show up almost exclusively in the s1 variable, which aligns with our intuitive understanding of sensor anomalies. Because sensor anomalies are defined as anomalies in a single data stream, the contribution-based approach is well suited for detecting and identifying these types of anomalies, with minimal interpretation needed.

5. DISCUSSION AND CONCLUSIONS

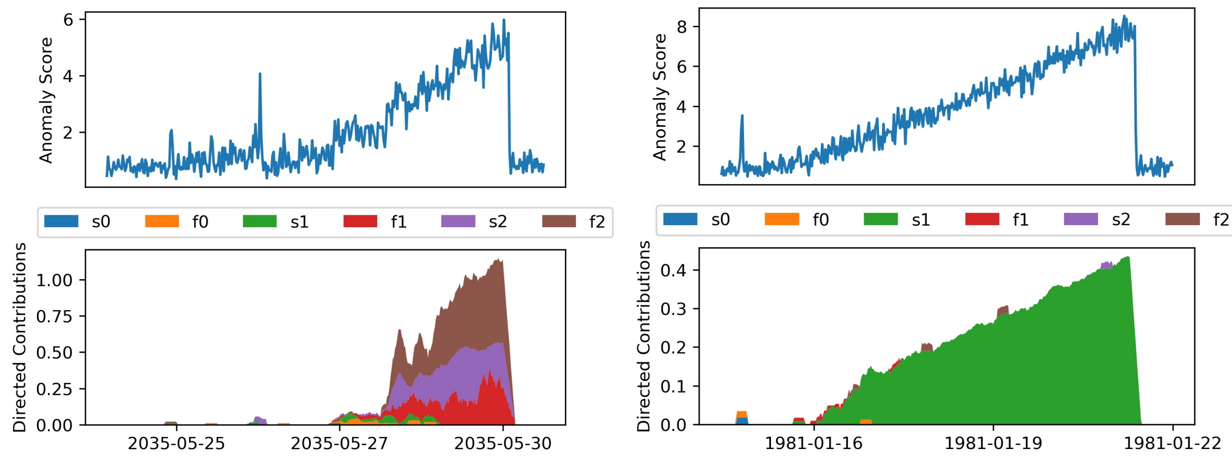
This work focused on the development of the LOVO anomaly detection and identification approach. It included a discussion of the system model and how it was utilized for both detection and identification purposes. Ultimately, it was applied to a synthetic dataset with known anomalies so that the results could be assessed.

Looking at the results, we see that the approach was able to both detect and identify anomalies, even without ever having encountered comparable anomalies. This is because the contributions aligned with an intuitive understanding of the system.

Areas of future work include testing the approach on more anomalies to obtain more statistically significant results. In addition, it will be important to benchmark the method against the DPCA method.

ACKNOWLEDGEMENTS

The authors wish to thank the U.S. Department of Energy Light Water Reactor Sustainability (LWRS) program for funding this effort. They also wish to thank Florida Power and Light Company—which is part of NextEra Energy, Inc.—for collaborating on this effort.



(a) p12 process anomaly. **(b) s1 sensor anomaly.**

Figure 3: Anomaly scores and contribution plots for two anomaly types.

DISCLAIMER

Idaho National Laboratory is a multi-program laboratory operated by Battelle Energy Alliance, LLC for the U.S. Department of Energy under contract no. DE-AC07-05ID14517. This work of authorship was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof. The document number issued by Idaho National Laboratory for this paper is INL/CON-23-71470.

REFERENCES

[1] Dunia, R., and S. J. Qin. 1998. "Subspace approach to multidimensional fault identification and reconstruction." *AIChE Journal* **44**(8), pp. 1813-1831. https://doi.org/10.1002/aic.690440812.

[2] Yue, H. Henry, and S. Joe Qin. 2001. "Reconstruction-Based Fault Identification Using a Combined Index." *Industrial & Engineering Chemistry Research* **40**(20), pp. 4403-4414. https://doi.org/10.1021/ie000141.

[3] Alcala, C., and S. Qin. 2009. "Reconstruction-based contribution for process monitoring." *Automatica* **45**(7), pp. 1593-1600. https://doi.org/10.1016/j.automatica.2009.02.027.

[4] Li, G., S. J. Qin, and T. Chai. 2014. "Multi-directional reconstruction based contributions for root-cause diagnosis of dynamic processes." *2014 American Control Conference*, pp. 3500-3505. https://doi.org/10.1109/ACC.2014.6859002.

[5] Farber, J., et al. 2021. "Process Anomaly Detection for Sparsely Labeled Events in Nuclear Power Plants." INL/EXT-21-64303, Idaho National Laboratory. https://lwrs.inl.gov/Advanced\%20IIC\%20System\%20Technologies/ProcessAnomalyDetectionSparselyLabeled.pdf.