Causal Analytics in IIoT – AI That Knows What Causes What, and When

Authors:
Dr. John Galloway
Founder
Au Sable
johnjgalloway01@gmail.com

Pieter van Schalkwyk
Chief Executive Officer
XMPro
pieter.vanschalkwyk@xmpro.com

Dann Marian
Engineering and Projects Manager
dann.marian@iinet.net.au

INTRODUCTION

Finding the ***"because"*** behind certain business or operations events has always been a key part of any engineering, maintenance or operations manager's job in industrial businesses. "The First Stage compressor failed ***because***…" or "the supply tank ran dry ***because***…" are common phrases in maintenance and operations departments in industrial businesses. Finding the "because" traditionally relies on experienced engineers that can interpret event, contextual and temporal data to deduce the likelihood of specific factors causing others in either a negative or positive way.

Knowing the real root causes of events is critical to resolving problems rather than continuously dealing with the symptoms. It resulted in popular, formalized approaches such as "Root Cause Analysis," or RCA as it is generally known. The challenge is that there are often multiple causal factors for these events, and finding the one "root cause" may not always be possible. Understanding other causal factors that may influence the outcome of industrial processes and the behavior of equipment need to be considered.

Au Sable, in collaboration with XMPro, developed an algorithmic, artificial intelligence-based, approach for "Reliable Causal Analytics"™ (rCA) in industrial IoT applications. This article demonstrates:

- It is possible to perform Reliable Causal Analytics using industrial IoT data and Artificial Intelligence (AI) to determine causality of business and operational events such as equipment failure or operational issues
- How Reliable Causal Analytics provides data-driven decision support for traditional Root Cause Analysis approaches
- The approach to embed this causal analytics methodology in IoT Process Management software to be able to perform this in a repeatable and automated manner.

rCA is the result of many years of research and application of causal analytics in real-world scenarios. Through this, Au Sable developed rCA that enables cause and effect relationships to be identified from sensor-driven data and made known to the analyst (e.g. wear on part #105 has causally impacted the performance of device #65 with a causal coefficient of 0.86), as well as correlation relationships in the data.

This means:

- the risk of making false decisions about what were, or will be predictively, the causal drivers of an effect is reduced, and
- the potential for costly or disastrous mistakes is thereby reduced.

This article provides background on traditional Root Cause Analysis and the evolution of Causal Analytics. It demonstrates how to automate the analytics to scale with an IoT Process Management platform and how it is applied in an industrial application. It provides a practical example of Reliable Causal

Analytics (rCA) applied to a floating production storage and offloading (FPSO)[1] vessel for an Oil & Gas company.

Crude oil, gas and water from the reservoir are separated on board the FPSO. Oil is stored on the facility in six pairs of tanks, before export to trading tankers. The vessel is designed to store 1.4 million barrels of oil and processes approximately 170,000 barrels of oil per day (bopd).

Example Floating Production Storage and Offloading Vessel

Au Sable's rCA has functioned in intelligence, defense and anti-terrorism applications for many years. The solution described in this article is the combination of advanced IoT Process Management software from XMPro and the rCA AI software from Au Sable.

Root Cause Analysis Based on Correlation Doesn't Work in the IoT Era

Industrial RCA Background

Formal Root Cause Analysis for industrial applications started with the Total Quality Management (TQM)[2] movement advocated by Deming in Japan in the late 1980's and early 1990's.

Paul Wilson et al[3] described the root cause analysis process for Quality Management in detail during the TQM era. "Root cause analysis is a method of problem-solving used for identifying the root causes of faults or problems. A factor is considered a root cause if removal thereof from the problem-fault-sequence prevents the final undesirable outcome from recurring; whereas a causal factor is one that affects an event's outcome, but is not a root cause. Though removing a causal factor can benefit an outcome, it does not prevent its recurrence with certainty."

Even though root cause analysis formally originated in TQM, it finds many applications in industrial environments:[4]

- Safety-based Root Cause Analysis arose from the fields of accident analysis and occupational safety and health.

[1] Floating production storage and offloading https://en.wikipedia.org/wiki/Floating_production_storage_and_offloading

[2] http://asq.org/learn-about-quality/root-cause-analysis/overview/roots-of-root-cause.html

[3] Wilson, Paul F.; Dell, Larry D.; Anderson, Gaylord F. (1993). Root Cause Analysis: A Tool for Total Quality Management. Milwaukee, Wisconsin: ASQ Quality Press. pp. 8–17. ISBN 0-87389-163-5.

[4] Adapted from https://en.wikipedia.org/wiki/Root_cause_analysis (classification)

- Production-based Root Cause Analysis has roots in the field of quality control for industrial manufacturing.
- Process-based Root Cause Analysis, a follow-on to production-based RCA, broadens the scope of RCA to include business processes.
- Failure-based Root Cause Analysis originates in the practice of failure analysis as employed in engineering and maintenance.
- Systems-based Root Cause Analysis has emerged as an amalgam of the preceding schools, incorporating elements from other fields such as change management, risk management and systems analysis.

Root Cause Analysis became popular as an approach to methodically identify and correct the root causes of events instead of addressing symptomatic results of these events. The objective of root cause analysis is to prevent problem recurrence. Some popular root cause analysis techniques include “Five Whys” and Cause and Effect (Fishbone) diagrams. These techniques rely on human interpretation of event information and data and require experienced practitioners to conduct the analysis. It is often limited to a few critical production assets as the manual process is time-consuming and laborious. Wilson’s distinction between root causes and other causal factors provides some guidance on the application of causal analytics in an IoT context for this article. Traditional techniques focused only on finding the root causes through manual review. Modern techniques such as rCA described in this article, combined with IoT data and advances in AI, enable engineers to not only assess root causes, but also find other causal factors. These causal factors may not lead to equipment or process failure but may still impact equipment or process performance.

Recent advances in cloud computing and AI provide the necessary infrastructure to analyze event data for IoT and other sources at massive scale. This means analysts can have a more expansive view of causal events rather than a reductionist view where the scope of an analysis is limited to what a human can process.

Motivation for Data-Driven, Reliable Causal Analytics

There are three main reasons to find a reliable, data-driven approach to finding root causes and causal factors for equipment failure and operational performance in industrial environments:

- Aging workforce and a large number of experienced engineers retiring soon
- Complexity of equipment, making it harder to troubleshoot
- Inaccuracy of Root Cause Analysis

Retiring Workforce

With a retiring workforce in many industrial sectors, the experience needed to conduct meaningful RCAs is decreasing. As much of the traditional approaches rely on observational analysis, the number of experienced engineers that can provide reliable analysis is fast reducing.

According to a January 2017 assessment by the US Department of Energy, 25% of US employees in electric and natural gas utilities will be ready to retire within 5 years[5]. The US Department of Labor also estimates that the average age of industry employees is now over 50 and up to half of the current energy industry workforce will retire within 5-10 years.[6]

Manual RCA requires the combination of a rigorous methodology, fault analysis technology and experience to evaluate the possible causes of business events such as equipment failure, quality problems or safety incidents. Much of the expertise needed will be lost with the retiring workforce. A data-driven, algorithmic approach provides a viable replacement for the experience of people to determine causal relationships between business events.

Complexity of Industrial Equipment

As industrial equipment becomes increasingly sophisticated [7] and more complex, the ability to perform diagnostics becomes increasingly more difficult. As equipment becomes more complex and sophisticated, the number or combinations and permutations of potential causal factors for certain events increases exponentially. It follows a similar pattern to Metcalfe's law[8] for telecommunication devices that states "the effect of a telecommunications network is proportional to the square of the number of connected users of the system (n^2)".

Metcalfe's law, now also used in economics and business management, provides some quantification of the impact of the increasing complexity of equipment to troubleshoot potential causal relationships between operational events.

Inaccuracy of Root Cause Analysis

Root Cause Analysis gained popularity in industrial and other sectors such as healthcare. One of the main challenges that emerged centers around the fact that it requires facilitation and analysis by people who can process only limited amounts of information. People are also susceptible to opinions and organizational influences such as politics. Peerally [9] et al describe the problem with Root Cause Analysis with these 8 main challenges:

- The unhealthy quest for "the" root cause
- Questionable quality of RCA investigations
- Political hijack

[5] U.S. Department of Energy, Quadrennial Energy Review (QER) Task Force report second installment titled "Transforming the Nation's Electricity System." Chapter V: Electricity Workforce of the 21st-Century: Changing Needs and New Opportunities. January 2017. Retrieved from https://energy.gov/epsa/initiatives/quadrennial-energy-review-qer

[6] U.S. Department of Labor Employment and Training Administration "Industry Profile – Energy." Retrieved from https://www.doleta.gov/brg/indprof/energy_profile.cfm

[7] Challenges To Complex Equipment Manufacturers: Managing Complexity, Delivering Flexibility, and Providing Optimal Service http://www.oracle.com/us/solutions/046249.pdf

[8] Metcalfe's law https://en.wikipedia.org/wiki/Metcalfe%27s_law

[9] The problem with root cause analysis http://qualitysafety.bmj.com/content/26/5/417

- Poorly designed or implemented risk controls
- Poorly functioning feedback loops
- Disaggregated analysis focused on single organizations and incidents
- Confusion about blame
- The problem of many hands

Many of these are as a result of the subjective nature of the people doing analysis and can be addressed with a more objective, data-driven approach. People can't process all the potential data sources of event and contextual information. Modern advances in data, stream and event processing address some of that challenge and AI provides a means to make sense of the data at scale. It removes the reliance on the subjective nature of human analysis and opens the opportunity to analyze fact-based information at scale to derive insights.

The unhealthy quest for "the" root cause further describes a challenge that can be better addressed with an algorithmic approach to Root Cause Analysis. Peerally states that "the first problem with Root Cause Analysis is its name. By implying—even inadvertently—that a single root cause (or a small number of causes) can be found, the term 'root cause analysis' promotes a flawed reductionist view."

An algorithmic approach often provides more potential causal factors, their relationship to each other and the strength (causal coefficient) of the relationships. It offers additional insights into events and often finds causation that may be counterintuitive to the views of the people that do it manually. An algorithmic approach also provides repeatability and scale. It will analyze the IoT and contextual data in a consistent way that is independent of the person performing the analysis.

USING CAUSAL ANALYTICS TO PERFORM ALGORITHMIC ROOT CAUSE ANALYSIS

Correlation is Not Causation

In this era of big data, it is commonly said that data analytics is a prime driver of value to enterprises[10]. This is true, but only if the analytics performed across the data are well grounded methodologically and perform well and efficiently to derive the value.

Big data creates big and complex data volumes. This is of limited value however, if it is not accompanied by the best available analytics to enable the most valuable, accurate and reliable decisions to occur[11]. Hence, there is an increasing requirement for the analytics component in industrial IoT solutions to be fast, reliable and accurate to identify the problems and opportunities and

[10] How does business analytics contribute to business value? https://onlinelibrary.wiley.com/doi/pdf/10.1111/isj.12101

[11] The Age of Analytics: Competing in a data-driven world https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx

ensure that such problems are addressed correctly and urgently.

Correlation of events and systems is often a starting point for problem-solving in industrial environments but "correlation is not causation"[12]. Correlation helps to point the way, helps indicate what might be candidate causative or driving factors for some particular effect yet keeping in mind that correlation is simply a measure of association not causation.

Introductory statistics courses tell us that it is not possible to *prove* causation unless one conducts an experiment whereby treatment and control groups are randomized.

This is totally correct but is just not feasible to conduct an experiment in 99% of real-world situations. Algorithmic methods have a *probabilistic* and contributory approach – spurred on by big data's need for empirically-based data-driven decisions – to answering questions about what caused what or will. For example, a causal coefficient of 0.83 of X as a causative influence on Y, does not mean that X is necessarily the *sole* cause of Y (there may be multiple causes) nor does it *always* cause Y. X is identified however as a *contributory* cause of Y. Similarly, smoking is a contributory cause of lung cancer; it is not the sole cause, nor does it always cause the effect.

Correlations can be misleading. Valuable results and insights are often found, but the correlation methods upon which decisions are made mean that risks are inherent and could lead to mistaken or sub-optimal decision-making and outcomes.

The chief analytics tool of most industrial IoT analytics vendors is correlation. Most sensor-driven data (IoT and machine-generated logs) is analyzed using a proven but older form of statistical methods (even when operating within a machine learning framework). Correlational methods are the dominant form of analytics.

Some examples from IoT vendor publications and websites demonstrate this approach:

1. Cisco (Attaining IoT Value): …enable the company's customers to perform real-time data correlation and, as a result, quickly react to irregularities[13]
2. Huawei ('The IoT's Potential for Transformation'): …enables correlation-based process and productivity improvements.[14]

---

[12] Correlation does not imply causation https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation

[13] Attaining IoT Value: How To Move from Connecting Things to Capturing Insights https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/iot-data-analytics-white-paper.PDF

[14] The IoT's Potential for Transformation http://e.huawei.com/en-sa/publications/global/ict_insights/201703141505/focus/201703141643

3. ThingWorx's capabilities make it possible for users to correlate data, deriving powerful insights…[15]
4. Siemens PLM: …quantitative statistical relationships to real-life usage, called customer correlation[16]
5. Industrial Internet Consortium: …common issue in IIoT systems is correlating data between multiple sensors and process control states[17]

Correlational methods are established as powerful aids to decision-making as is witnessed in the rise of platforms that provide the capability. Correlations often vary such that at a given time one entity and another may be positively related and at other times only weakly related or not at all. There is no fact-based causal coefficient that describes the strength of potential causal relationships.

The lack of stability in correlations indicates complexity in the relationships and the presence of a dynamical system (common in IIoT). This results in variability according to the system state and nonlinearity in system behavior. It means that traditional statistical methods, correlation included, have limitations for obtaining precise analytics and improved decision making about performance in IIoT.

The result is that an observed correlation over time may or may not be coincidental; or, the observed correlation (and any implied causation) may be the result of one or more third-party variables (hidden confounders), e.g. another variable that influences two events that are seemingly correlated. An example of this may be ice cream sales and boating accidents that are correlated, but both are affected by summer temperatures, and so a causal inference would be spurious. In this example summer temperature is causal, but one may incorrectly infer causation that an increase in ice cream sales leads to boating accidents due to the high correlation factor. More humorous examples of these erroneous correlations can be found at Spurious Correlations.

Mathematically-based causal analytics attempts to improve on correlation for causality identification.

The Evolution of Causal Analytics

Causality for Real-World Applications

It is well accepted that causation cannot be *proven* statistically unless one conducts an experiment with randomization to control

[15] A survey of IoT cloud platforms https://www.sciencedirect.com/science/article/pii/S2314728816300149

[16] Customer Correlation Durability Methodology https://www.plm.automation.siemens.com/en/products/lms/engineering/customer-correlation.shtml

[17] Industrial Analytics: The Engine Driving the IIoT Revolution https://www.iiconsortium.org/pdf/Industrial_Analytics-the_engine_driving_IIoT_revolution_20170321_FINAL.pdf

for spurious relationships [18] [19], which is simply not practicable in most real-world situations. The position of the authors is that correlational methods have served well and are proven to provide useful insights, but are nonetheless prone to producing spurious relationships and hence mistaken decisions.[20] [21]

As noted earlier, causality research has been undertaken to develop different probabilistic methods and approaches for identifying cause and effect relationships in non-experimental or 'observational' data.

Causal analytics evolved over the past few decades from academic studies to practical solutions such as rCA. A stumbling block historically in reaching this goal has been to devise causal algorithms that produce reliable and accurate results for commercial and government application.

ADVANCES IN CAUSAL ANALYTICS AND THE DEVELOPMENT OF RELIABLE CAUSAL ANALYTICS (RCA)

In the 1980s, mathematical advances by Judea Pearl[22] from UCLA showed that causal relationships can be represented from data in terms of probabilities and led him later to declare that "causality has been mathematized". The mathematization was perhaps a little premature, but Pearl's outstanding work led him to be awarded in 2012 the industry's equivalent of the Nobel Prize, the Turing award, for advances in both machine learning and causality.

Problems remained however, e.g. how to identify a causal relationship when unknown delays occur between cause and effect. And, what are termed hidden confounders, were difficult to identify and control for. Earlier, the work of Weiner (1950s) laid the basis for several information-theoretic measures of causality (and for well-known data compression algorithms).

A landmark innovation was that of Clive Granger [23], awarded a Nobel prize for developing a test of causality: X is said to cause Y, if the past values of X contain information that helps predict future values of Y, above and beyond the information contained in past values of Y – graphically:

[18] https://us.sagepub.com/sites/default/files/upm-binaries/14289_BachmanChapter5.pdf

[19] http://www.statisticssolutions.com/establishing-cause-and-effect/

[20] https://hbr.org/2015/06/beware-spurious-correlations

[21] https://en.wikipedia.org/wiki/Spurious_relationship

[22] Judea Pearl https://en.wikipedia.org/wiki/Judea_Pearl

[23] Clive Granger https://en.wikipedia.org/wiki/Clive_Granger

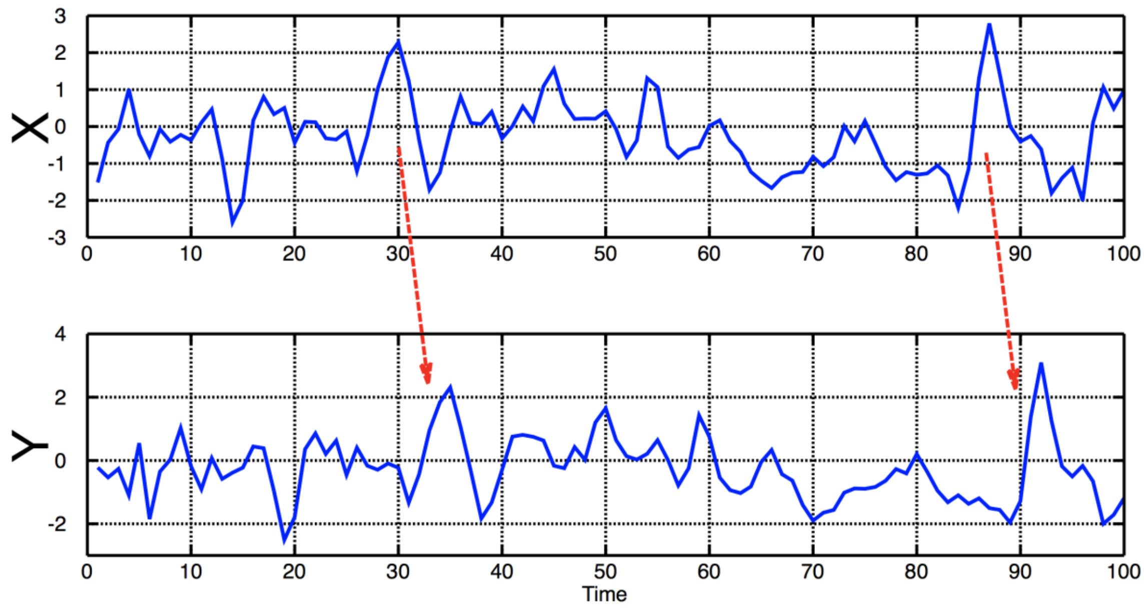

Figure 1: Granger causality test

Researchers extended this framework, e.g. to allow for analysis of multiple time series generated by nonlinear models, for lagging the cause and effect variables and for causal graphical models for better handling of latent variables.

Transfer Entropy [24] [25] (TE) is a later implementation of the principle that causes must precede and predict their effects. TE improves on Granger in that it directly caters for nonlinear interactions and helps minimize problems of noisy data. TE is a model-free and non-parametric measure of directed information flow from one variable to another.

The application of TE to empirical analytics has been substantial in areas of biomedicine and climate science. However, further developments were needed to help overcome shortcomings related to unreliability and a lack of accuracy [26]. Au Sable's work on improving the reliability of TE, combined with other Au Sable proprietary algorithms, have led to the development of an algorithmic approach to causal analytics that can process IoT event data and provide reliable results. This means that Causal Analytics can now be applied to real-world scenarios with non-experimental data.

[24] Transfer Entropy https://en.wikipedia.org/wiki/Transfer_entropy

[25] Transfer entropy between multivariate time series https://www.sciencedirect.com/science/article/pii/S1007570416305020

[26] Progress in Root Cause and Fault Propagation Analysis of Large-Scale Industrial Processes https://www.hindawi.com/journals/jcse/2012/478373/

These real-world applications involve methods that take into account the complexity of systems (thereby including analytics of system machine and log data). The inter-dependencies and dimensionality of many IIoT system devices mean that identifying their behavior (causal and otherwise) can be extremely difficult depending on the magnitude and nature of the couplings. One variable may be found to be a driver of another, but not alone. The multiple influences that have an impact on a particular variable must be teased out, such as the timings, state-dependencies and multi-dimensionality of other influences that impact an 'effect' of interest, such as a decrease in pressure or rise in temperature. These are identified as part of the rCA process for IIoT.

This approach has led to an area of causal research from a dynamical systems perspective. A dynamical system is one in which a function describes the time dependence of a point in a geometrical space.[27] [28] [29] [30] A dynamical systems course at Harvard states that the methods have a focus on the behavior of systems described by ordinary differential equations. Application areas "…are diverse and multidisciplinary, ranging over areas of applied science and engineering, including biology, chemistry, physics, finance, and industrial applied mathematics."[31] This is a fairly recent set of developments and especially with respect to incorporating AI and machine learning where these algorithms can be applied to IIoT data at scale.

Automating rCA in Industrial IoT Applications

Although the rCA approach can be employed on an ad-hoc basis by an analyst, the real benefits come from automating the rCA AI analysis as part of an IoT process. The rCA function can be executed based on trigger events such as data changes or exceptions. The rCA software and algorithms are embedded in the functions library of the XMPro IoT Process platform for IIoT applications.

---

[27] https://en.wikipedia.org/wiki/Dynamical_system

[28] https://en.wikipedia.org/wiki/Dynamical_systems_theory

[29] https://mathinsight.org/dynamical_system_idea

[30] http://math.huji.ac.il/~mhochman/research-expo.html

[31] https://scholar.harvard.edu/siams/am-147-nonlinear-dynamical-systems

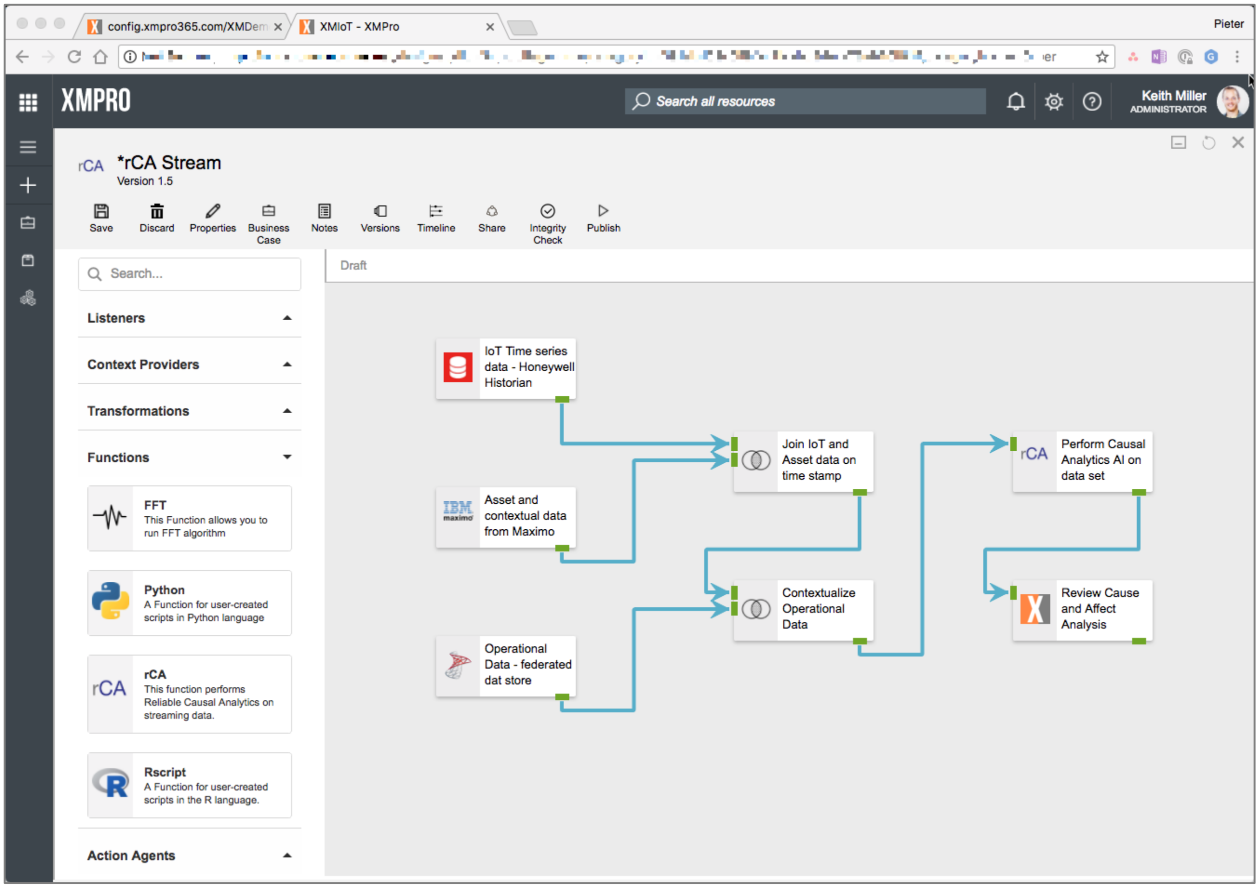

Figure 2: XMPro IoT Process Stream for rCA

In this example, event data is ingested from their Honeywell® historian and contextualized with asset data from their IBM Maximo® EAM system. Further context is provided from operational data stores. This information is passed to the rCA Causal Analytics AI function that creates the causal coefficient matrix and other outputs described later in the article.

This automated, process-based approach ensures repeatability, consistency and that it can be done at scale for a large number of assets in a process stream. The automated process can process and analyze much larger volumes of IoT data than human RCA analysts. In the FPSO example different analyses are automated at different time intervals such as daily for high impact equipment and weekly or monthly for other areas. This is configurable by the end users and ad hoc analysis can also be performed.

Customer Example: rCA in Oil and Gas Processing

Background to the Application of rCA in Oil & Gas

The example demonstrates how rCA can enable an FPSO to optimize production and productivity as well as predict and avoid incidents which threaten health, safety, environment, community and financial outcomes. The initial field study project was aimed at three key objectives:

Efficient Operations and Maintenance

This project will drive down the costs of reduced or lost production caused by unplanned failures. Other operational efficiency gains will be achieved by reducing the risks of environmental impact caused by operational failure and the risks to personnel safety caused by breaches of operational standards. Furthermore, the costs of asset maintenance will be reduced and the capability of diagnosing asset health in remote and challenging operating environments is increased.

Safety and Social License to Operate

Equipment failure and/or an unsafe work environment can potentially result in harm to humans or the environment, ultimately increasing operational risk and impacting an organization's social license to operate.

This solution will assist in providing a safe production environment. In addition, through to the inbuilt predictive analytics, further eliminate operational risks which could impact the social license to operate if undetected and left uninvestigated and unaddressed.

Enabling Effective Collaboration

Traditionally there exists a significant divide between the operational technology (OT) in heavy asset sectors like Oil & Gas and the information technology (IT) arena. Not only are they typically separated by physical, geographical and network constraints, they are also generally isolated philosophically.

The innovative solution and integrated application suite enables interoperability of data feeds from sensors and devices, with the associated referential information from the asset registry and maintenance framework. It combines data from both IT and OT and this new information provides insights that can be shared collaboratively between OT, IT and Operations. It makes new levels of operational excellence, collaboration and sustained productivity improvements possible.

The project mirrored an upstream oil and gas process flow including value-added services at each stage of the supply chain leveraging real-time IoT big data, machine learning and artificial intelligence.

Going beyond the obvious elements that cause an interruption to production, rCA is used to find root causes and inter-dependency which may be overlooked or not realized with current technology. This will enable the operations team onboard the FPSO to keep it in production without interruption for long periods and, when down, to be repaired and brought on-stream faster.

Most importantly, these improvements reduce the risk of events that impact the safety of all personnel on the FPSO and protect the environment on the vessel and in the geographic vicinity.

Project Background

The FPSO plant had experienced occasional periods of operational instability. These were largely unexplained, yet some significant and costly problems resulted. It was particularly challenging to identify the *actual* cause(s) of the problems. Routine correlational methods of analysis, such as traditional RCA, had been applied but provided the operators with only limited assistance.

The project was conducted in 2 phases. In the initial phase the FPSO operator wanted to validate the algorithms through an initial manual analysis before automating the process in phase 2.

Au Sable's rCA system used sensor-driven and machine data covering a defined period where these events occurred as the basis for the analysis. The rCA analysis discovered cause-effect relationships that were not previously known by the engineers and that helped to identify and address the real root causes of the problem. The results of the rCA analysis in phase 1 provided new insights into causal relationships that were previously not considered in the human analysis process. This is of significant value and benefit to the operations and engineering teams of the Oil & Gas customer operating the FPSO and it provided the support to automate the process in phase 2.

The example data shown is that of the phase 1 analysis that provided the insights and confidence in the output to support the decision to automate the process for a larger data set and additional equipment.

rCA on the FPSO

During phase 1 the analysis was done manually to validate the model and establish the workflow for the automated steps in phase 2.

The analysis approach in phase 2 consists of three main process steps:

- Ingest data: Real-time feeds of operational data are collected from intelligent equipment (e.g. submersible pumps, etc.) and from the Honeywell Historian through XMPro's Listener integration connectors that stream the data to the analysis step;
- Perform analysis: The streaming data from the previous step is passed to the rCA algorithms in the XMPro rCA Functions connector where the analysis is performed; and
- XMPro Action Agents provide reports and actions on the results of the analysis that identify real causes and not just symptoms of production outages.

Data from sensors at locations across the plant operations were mapped to locations on a process flow diagram (PFD) and are shown as red circles in Figure 3. This provides a familiar visual reference for the engineers of the physical process and the data from the different sources.

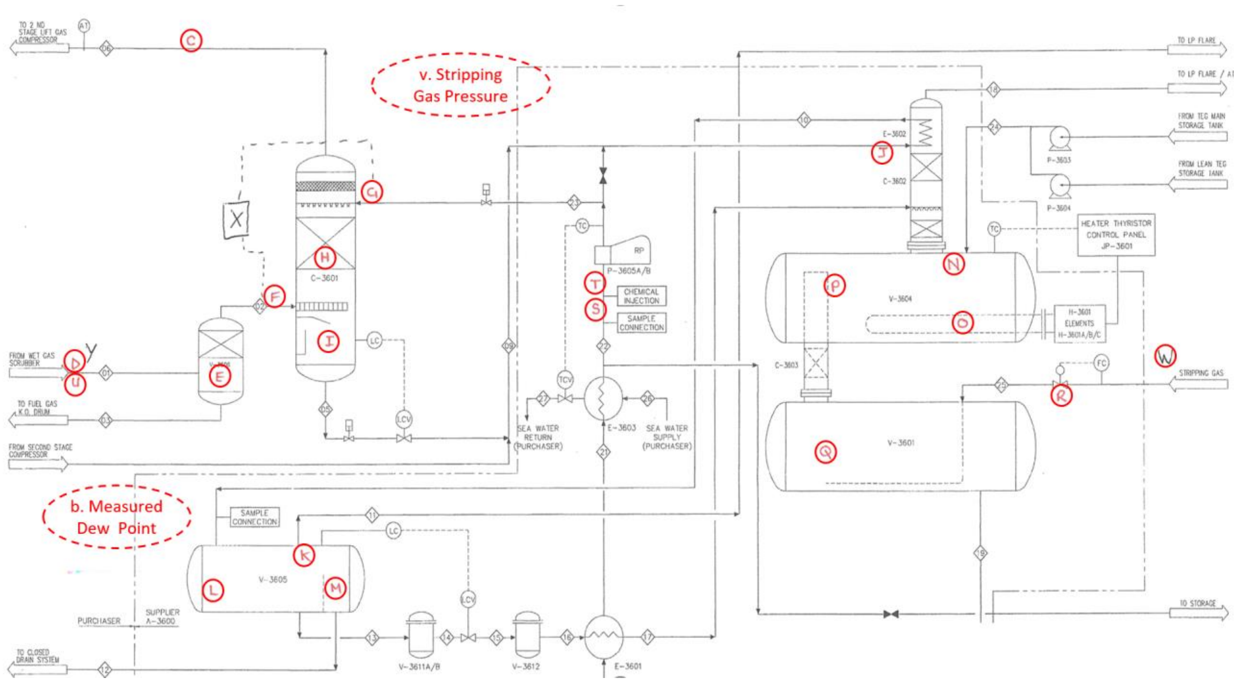

Figure 3: Locations of sensors (red circles) on a process diagram of plant operations

The FPSO plant engineers provided one month's data at one-minute intervals from sensors at the above locations. The rCA AI algorithm processed the data to identify a limited number of cause and effect relationships between the equipment or devices in Figure 3 where there is a high causation coefficient. This is derived from a proprietary causal effects matrix. The causal effects matrix tends to be sparse with a much small number than in a correlation matrix, typically about 10-15% for the same data.

The output from the causal effects matrix that provided invaluable insight for the engineers is a graph (Figure 4) that ranks causal relationship based on causal coefficient and the confidence level in the causal relationship. It identifies those relationships with high causality and high confidence at a glance and engineers can use this information to map it back to the physical process.

It is easy to spot events that have a high causal coefficient with high confidence levels and focus on the meaning and impact of these insights.

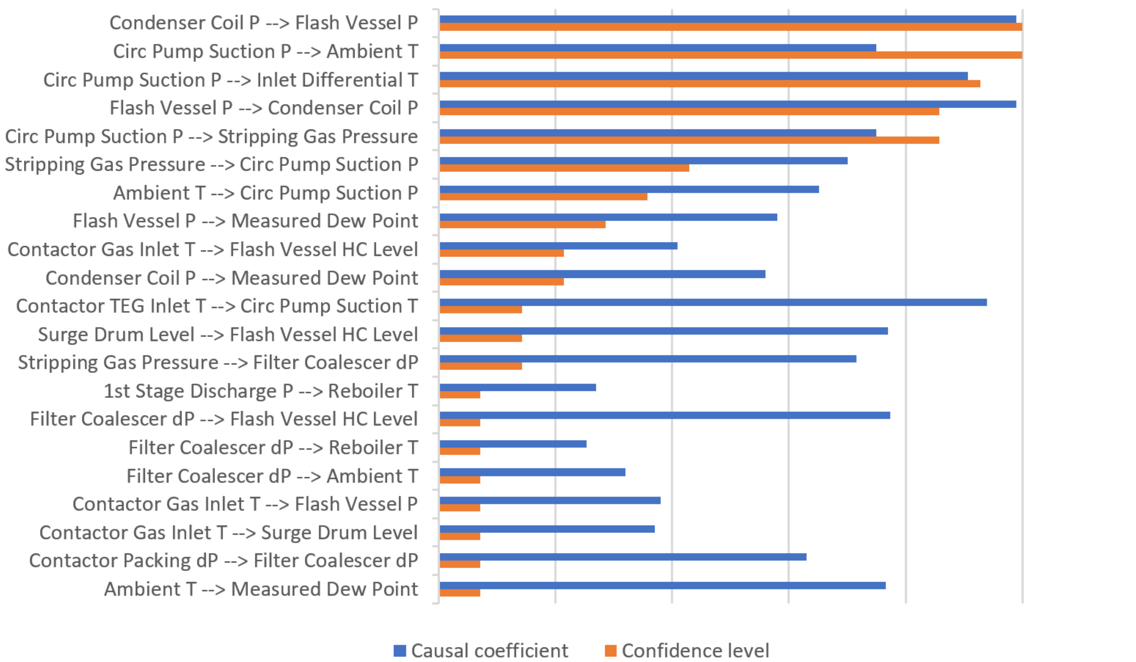

Figure 4: Chart of the top cause and effect relationships

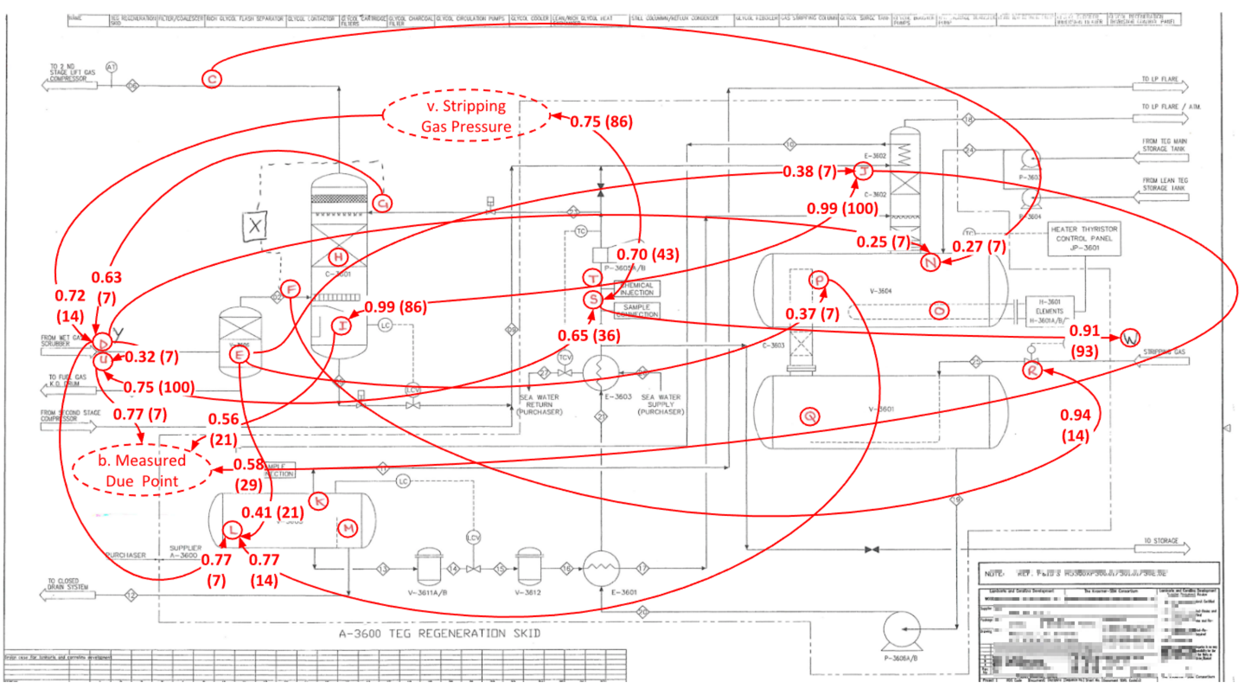

Figure 5: Overlay of causal relationships on a process flow diagram – showing causal coefficient values plus confidence levels in brackets

An enhanced PFD diagram, Figure 5, shows the top casual relationships between the equipment/devices overlaid on the process flow diagram. This approach connects the analytical model with the physical process model for the engineers who can interpret the results.

This diagram shows, for example the causal relationship and confidence level, in brackets, of the condenser coil pressure and the flash vessel pressure (nodes I and J) that correspond to the chart in Figure 4.

The plant engineers were most interested in examining the main relationships, those with the strongest measure casual coefficients and higher levels of confidence.

Findings and Observations from the Phase 1 Analysis

The chart in Figure 4 provided the most insight to the FPSO operators. The initial objective to validate the rCA algorithm was accomplished with physical evidence to support the outputs of the rCA analysis. The following three examples describe some of that validation process.

The high casual coefficient of the condenser coil pressure and the flash vessel pressure made sense from an engineering perspective as it is part of the design. It was expected to have a high causal relationship and it did. This proved that the rCA algorithm performed as expected.

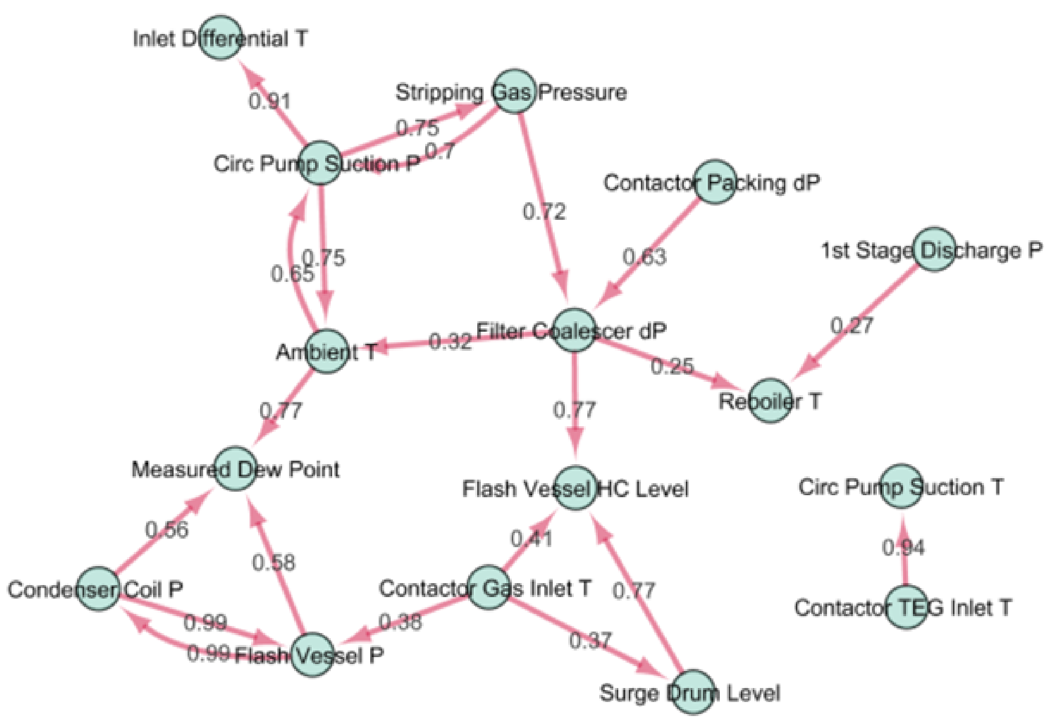

Figure 6: The top cause and effect relationships shown as a directed graph

Figure 6 provides a causal coefficient view in a traditional graph that removes the contextual bias that an engineer may have. This view enabled the engineers to see causal relationships without the physical process relationships. It triangulated some of their findings and observations from the chart and process flow diagram views.

The causal relationship between the condenser coil pressure and the dewpoint (middle of the chart) was not obvious prior to the analysis and it turned out that it had an impact on operations. It led to further investigations by the engineering team.

The third example relates to the causal relationship at the bottom of the chart. It has a high causal coefficient, but a low confidence level for ambient temperature to measure dewpoint. In a physical, isolated system the 2 should not be related from an engineering perspective. One of the engineers found the cause after investigation. The dewpoint measurement sensor is exposed to full sun when the vessel sits in a certain orientation, which affected the dewpoint measurement and led it to track the ambient temperature as a measure of exposure to sunlight. This in turn affected the efficiency of operations as the rate of lift-gas being dehydrated was affected.

The three examples provided enough evidence to proceed with phase 2 to automate the process for scheduled time intervals and to include additional data points. One of the main benefits of the automated process is that the FPSO engineers can perform the root cause analysis without the assistance of a data scientist to manually perform the analysis. This is an important requirement for the FPSO operator to deploy rCA at scale. The nature of causal analytics requires a complete re-analysis if any of the causes found in previous analyses were addressed or changed. There are often unintended consequences of making changes that only show up in a new analysis. These analyses need to be performed in a consistent manner which is facilitated by the automation of the process. Phase 2 is underway at the time of writing this article and the benefits of the automation process will described in a future article by the authors.

The use case illustrated that an additional dimension of understanding can be added to the analysis and troubleshooting of Industrial IoT problems. The results are reported in the above formats and able to be extended to additional desired forms of output.

CONCLUSION

Most IoT platform vendors use traditional statistical methods based on correlation and regression for their analytics functionality. These are proven to work but also have limitations that restrict their applicability to quickly and accurately identify causative factors that did, or will, cause some effect. This means the decision making may be sub-optimal or even inaccurate.

rCA adds a different and complementary dimension of IoT data analysis and decision making. rCA is a breakthrough technology that brings new methods for the application of cause and effect analytics to real-world industrial problems. By adding a new and powerful dimension of analytics, it enables users to refine their decision-making, reduce risks, improve safety and reliability, and reduce costs and equipment downtime.

The culmination of this "Reliable Causal Analytics" approach is the ability for the solution to enable operators to more completely understand:

- What caused what (and when)
- What will cause what (and when)

- What could/should be done to avoid…
- Which processes could/should be changed to improve going forward
- Continuous improvement of processes

Future work and research is planned to extend rCA to other use cases and IIoT applications. Interested parties are invited to submit use cases for consideration.

➢ Return to IIC Journal of Innovation landing page for more articles and past editions.