Check for updates

Received: 15 December 2023 | Accepted: 12 January 2024

DOI: 10.1002/maco.202314240

ARTICLE

Materials and Corrosion

Causal discovery to understand hot corrosion

Akhil Varghese | **Miguel Arana-Catania** | **Stefano Mori** | **Adriana Encinas-Oropesa** | **Joy Sumner**

Cranfield University, Cranfield, UK

Correspondence
Miguel Arana-Catania, Cranfield University, Cranfield MK43 0AL, UK.
Email: miguel.aranacatania@cranfield.ac.uk

Funding information
None

Abstract

Gas turbine superalloys experience hot corrosion, driven by factors including corrosive deposit flux, temperature, gas composition, and component material. The full mechanism still needs clarification and research often focuses on laboratory work. As such, there is interest in causal discovery to confirm the significance of factors and identify potential missing causal relationships or codependencies between these factors. The causal discovery algorithm fast causal inference (FCI) has been trialled on a small set of laboratory data, with the outputs evaluated for their significance to corrosion propagation, and compared to existing mechanistic understanding. FCI identified salt deposition flux as the most influential corrosion variable for this limited data set. However, HCl was the second most influential for pitting regions, compared to temperature for more uniformly corroding regions. Thus, FCI generated causal links aligned with literature from a randomised corrosion data set, while also identifying the presence of two different degradation modes in operation.

KEYWORDS

causal discovery method, causal inference, FCI algorithm, gas turbine superalloys, hot corrosion, Kernel-based conditional independence test

1 | INTRODUCTION

Gas turbines are used to generate power and provide thrust.[1] In 2023 between 26% and 37% of UK electricity was generated from natural gas combustion, with gas turbines now shifting towards zero emission fuels such as H_2 or NH_3, to meet UN goals. To increase efficiency, gas turbines operate in a combined cycle with steam turbines, and are required to operate at higher temperatures and pressures, generating a more challenging environment for the materials used for their manufacture. Furthermore, contaminants such as sulphates, halides, and chlorides, contained in the fuel, together with salt impurities from the air, can create a highly corrosive environment.[2] Deposition of such contaminants on to the gas turbine's component blades and vanes gives rise to a corrosion mechanism called 'hot corrosion', which adversely affects the service life of the gas turbines. Depending on the operating temperature conditions inside the gas turbine, Type-I or Type-II hot corrosion can occur.[3] Type-I and Type-II hot corrosion are temperature-dependent and deposit-induced corrosion mechanisms that occur across the approximate temperature ranges of 850–950°C and 650–800°C, respectively.[2] Melting temperatures of the salt contaminants influence both these types of accelerated corrosion mechanisms, but factors such as a high partial pressure of SO_3 are crucial for Type-II hot corrosion to occur.[2] While it is known that factors like operating temperatures, melting temperatures

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Authors. *Materials and Corrosion* published by Wiley-VCH GmbH.

of salt contaminants, the deposition rate of the flux on gas turbine surface, partial pressures of gas contaminants, and material and gas compositions lead to hot corrosion in gas turbines, the exact variation of the underlying mechanism with changing parameters needs clarification.

Moreover, the degree of co-dependence and independence between these factors will help in understanding the influence that each factor has on hot corrosion. Hence, Causal Discovery[4,5] has been introduced to understand the causal relationships between the different corrosion factors from a statistical approach. These data-driven, causal relationships will help in understanding the underlying causal relationships of each corrosion variable, and thus clarifying the physical corrosion mechanism. Such findings from this early-stage study can also help in designing future hot corrosion experiments, both in terms of the variables and their values, and can become the foundation upon which a predictive maintenance model can be developed.

1.1 | Hot corrosion

Hot corrosion is a type of material degradation which is induced by the deposition of contaminants contained in the exhaust stream of a gas turbine.[6] These deposits are usually formed by alkali compounds that can increase the corrosion rates if in molten state.[3] Some of the contaminants come from air intake (such as sodium, potassium, and calcium), while others come from fuel. The amount of deposition influences the corrosion rate of the materials in such environments.[7] Based on the operating temperatures and the type of contaminant attack, hot corrosion can be further divided into Type-I and Type-II hot corrosion:

1.1.1 | Type-I hot corrosion (high-temperature hot corrosion [HTHC])

Type-I hot corrosion is also called HTHC because it occurs within the temperature range of 850–950°C.[2] Various compounds can form, for example, sodium from the ingested air and sulphur from the fuel could combine to form sodium sulphate (Na_2SO_4), during the combustion cycle of the gas turbine.[8,9] The sodium sulphate could then condense onto the colder blading, and form a liquid (molten) salt, which initiates Type-I hot corrosion. In the presence of further impurities, such as NaCl from the industrial or marine atmosphere combines or K from sea droplets,[9] the formation of mixtures with lower melting points could occur which extend the temperature range of Type-I hot corrosion.[10]

Using a fluxing reaction, these eutectic mixtures dissolve the protective layer of the superalloy and attack the base material. These mixtures shorten the incubation period. Type-I hot corrosion is characterised by internal sulfidation, depletion of protective layers formed by chromium or aluminium, and leads to severe metal loss from the base material, impacting the gas turbine life. Sulfidation can be seen in Figure 1b,d.

1.1.2 | Type-II hot corrosion (low-temperature hot corrosion [LTHC])

Type-II hot corrosion, also called LTHC, occurs at temperatures between 650°C and 800°C.[3] As in the case of Type-I hot corrosion, the formation of a molten deposit is also involved in the mechanism controlling the corrosion behaviour. However, the mechanism occurs at temperatures below the melting temperature of many pure salts. In Type-II hot corrosion the Na_2SO_4 forms mixtures with metallic inorganic compounds,[2] which lowers the combined melting temperature and initiates the corrosion process. These inorganic metal compounds are formed by SO_3 present in the combustion gas and the metal,[10] making Type-II hot corrosion not only a function of the temperature, but also a function of the partial pressure of the SO_3 gas.[3] Type-II hot corrosion is characterised by pitting with localised failure as seen in Figure 1a.[2] The cause of pitting initiation is currently under consideration and has been linked to a wide range of factors including grain boundaries, precipitates, gas environment and the salt deposit.

1.2 | Gas turbine materials

To increase efficiency, higher turbine operating temperatures and pressures are required. This has driven the formulation and selection of materials to be used, which need to have[10]:

- high mechanical strength at temperatures close to the melting point
- high creep resistance
- high corrosion resistance

For these reasons 'superalloys' with better mechanical properties than conventional alloys have been developed. These alloys are nickel-, iron-nickel and cobalt-based,[12] with other elements in solution, such as chromium and aluminium, which preferentially oxidise to form a thin protective oxide layer to provide resistance to corrosion and oxidation.[12]

To ensure resistance to hot corrosion, a minimum amount of chromium is needed, for instance, 22% chromium is present in Co-base Haynes 188 superalloy,

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

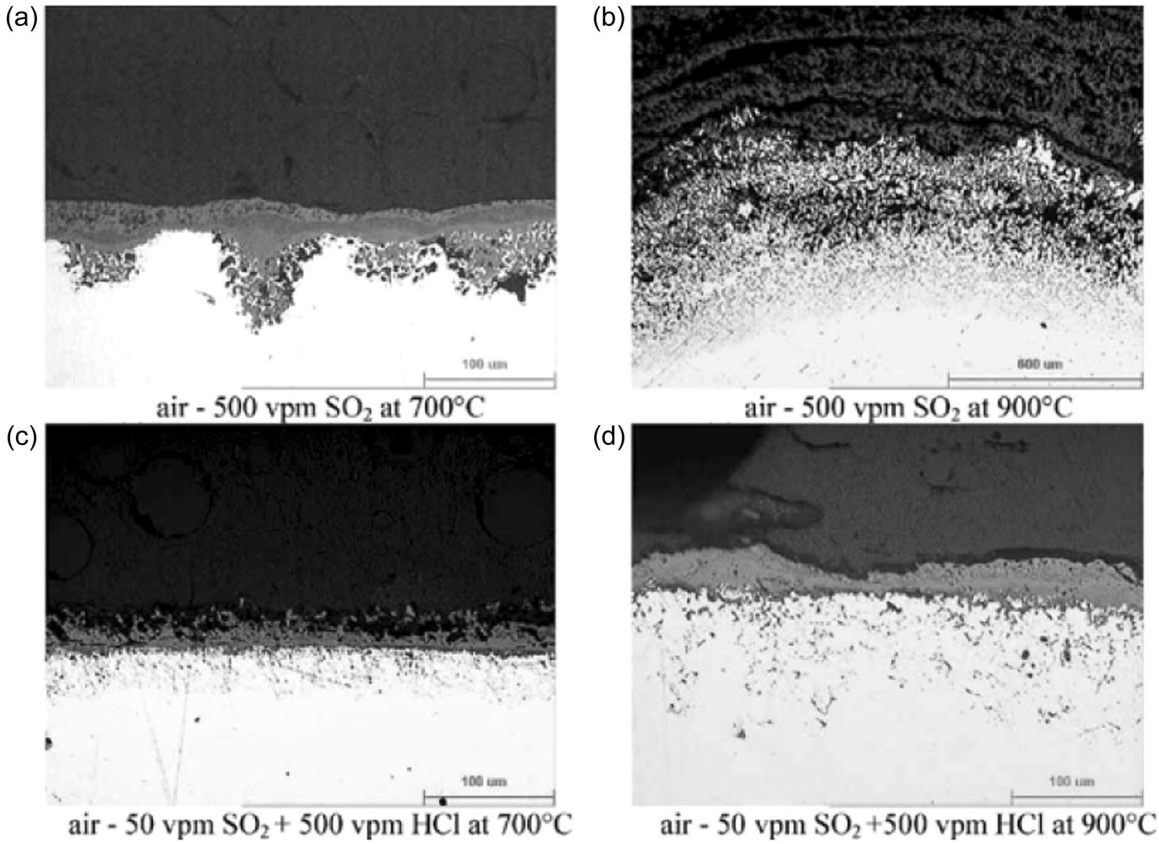


FIGURE 1 Optical micrograph of an uncoated SC[2] after 500 h exposure in different conditions (a) air + 500 vpm SO2 at 700°C; (b) air + 500 vpm SO2 at 900°C; (c) air + 50 vpm SO2 + 500vpm HCl at 700°C; (d) air + 50 vpm SO2 + 500vpm HCl at 900°C. Adapted from[11]

which provides high fatigue strength and strong resistance to hot corrosion.[13] In comparison, even though nickel-base Haynes 214 contains 16% chromium, it provides high oxidation resistance to temperatures above 900°C due to the presence of 4.5% aluminium.[12]

Microstructures can also be controlled with commercially used single crystal superalloys like CMSX-11C and SC-16 providing increased resistance to hot corrosion due to the presence of more than 12% chromium.[6]

1.2.1 | Protective coatings

Despite the more recent advancements, most of the superalloys are not able to provide the desired lifetime in all conditions within gas turbines. Thus, for specific areas of the turbines, coatings are required.[6,12] Aluminide diffusion coatings, overlay coatings and thermal barrier coatings are the most used types of coatings for gas turbine applications.[14] Platinum-modified aluminide (Pt-Al) coatings like RT-22 and CN-91 are widely used[15] due to their great resistance to Type-I and Type-II hot corrosion.[16] On the other hand, overlay coatings provide excellent oxidation and hot corrosion resistance due to their ability to form

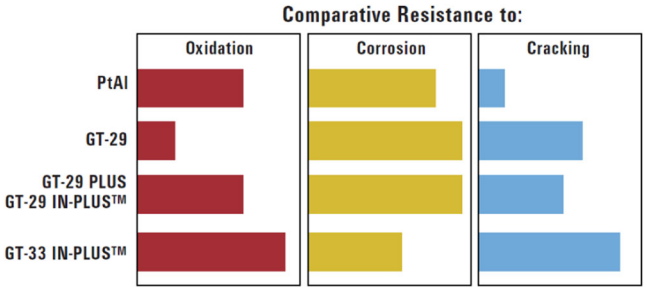


FIGURE 2 Comparison of resistance performance between platinum-aluminide, and overlay coatings.[8] [Color figure can be viewed at wileyonlinelibrary.com]

alumina and chromia scales.[17] Figure 2 shows the degradation resistance of a platinum-aluminide diffusion coating versus three overlay coatings.[8]

1.2.2 | Hot corrosion summary

Hot corrosion is dependent on different factors including the chemistry of the corroding material, temperature, partial pressure of contaminants and chemistry of the contaminants. For this reason, it is not straightforward to

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

understand the connection between the different parameters and the rates of corrosion. Understanding the cause-effect relationship between the corrosion factors discussed can help in analysing materials' degradation in gas turbines. The discovery of cause-effect relationships can be done through the use of causal discovery techniques, which can investigate the links between these corrosion factors.

1.3 | Causal discovery

Causal analysis techniques have been prominently used in the fields of engineering, medicine, and economics.[4] Manufacturing, however, has yet to fully embrace causal discovery methods when compared to the previous fields, and thus there is limited work on hot corrosion using these techniques.[18]

One example of the use of causal discovery methods was applied to detect strong relationships in degradation data.[19] A neural network was used to assess the degradation state of some equipment. The causal discovery method FCI was used to create a model of the relationship between the variables. The variables responsible for the degradation state found using FCI, were then fed into long short-term memory (LSTM) neural networks for the subsequent assessment. In this manner, the data-driven model was trained, and its interpretability was improved.

In a second example,[20] the method was applied to the design process of energy-efficient buildings. The authors applied the causal discovery algorithm greedy equivalence search (GES) to the variables that potentially affect building design. Creating such a causal framework is expected to allow designers, developers, and construction workers to inspect and continuously improve their own designs and construction methodology.

In the final example reported here, causal discovery algorithms were used on an Alzheimer's disease data set.[21] The study compared between two causal discovery algorithms and an existing standard graph on Alzheimer's disease, which was formed using literature and prior experience. FCI and Fast Greedy Equivalence Search were the methods used to form causal graphs, with initial causal graphs formed purely based on observational datasets that is, without prior subject knowledge. Subsequently, 'background knowledge' was added to the algorithms and the changes on average accuracy, recall and precision were compared with the former results. These causal graphs were later validated based on the existing standard graph and the discovered graphs were found to be very close.

This research aims to understand the degree of influence, independence and co-dependence of several hot corrosion variables causing material degradation in a gas turbine setting using causal discovery techniques.

2 | METHODS

Section 2.1 explains the corrosion data set along with its seven variables. Section 2.1.1 illustrates the arrangements and the necessary fine-tuning made to the data set to use as input to the causal discovery algorithms. Section 2.2 explains the significance levels and the assumptions made for analysing the causal graphs during the discussion. Sections 2.2–2.5 explain the causal discovery algorithm and everything required for its implementation.

2.1 | Corrosion data

The corrosion data set used in this work was formed by Dr Adriana Encinas-Oropesa in 2005, during her PhD thesis in collaboration with the Advanced Long Life Turbine Coating Systems project (ALLBATROS).[11] To produce the corrosion data set, experiments were conducted on a single crystal CMSX-4. Three different metallic protective coatings, RT22, CN91 and LCO22 were applied to the CMSX-4 base alloy. The uncoated CMSX-4 material and the three coated versions of the CMSX-4 alloy were treated as four different materials during the application of the causal discovery algorithm. The data set consists of two different operating temperatures, 700°C and 900°C, with varying levels of gas compositions and deposit chemistries and fluxes.

After an exposure time of 1000 h, the material loss data were collected. This data corresponded to three different flux deposition rates of 0.5, 1.5 and 5 $\mu g/cm^2/h$. The gas composition consisted of a constant 300 vppm (volumetric parts per million) of SO_2 along with varying amounts of HCl (0 or 100 vppm). Material loss values (due to hot corrosion) corresponding to each deposition rate, temperature, material, and amount of gas composition were tabulated.

2.1.1 | Data preprocessing

To assess the extent of hot corrosion, pre- and postexposure sample dimensions were compared resulting in 30 values. The data set formed gives material loss as a function of the cumulative probability of each amount of metal loss. The probability indicates the likelihood of having a certain amount of metal loss or more. Thus, the most extensive damage occurs with low probability. For simplicity, the cumulatively distributed data set was truncated to three values: highest, median, and lowest material loss. The highest and the lowest material loss (HML and LML) are used to understand the factors that are dominantly influencing the hot corrosion at the opposite ends of the material loss spectrum, while the median material loss

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(MML) helps in identifying the typical corrosion factors leading to hot corrosion from an overall perspective.

'Amount of salt', 'temperature', 'SO_2', 'HCl', 'time of exposure', 'material', and 'material loss' were the seven variables tabulated. Although the variables could take continuous values, in the experiments they were fixed to specific values. Thus all are considered categorial variables for the causal discovery algorithm, except for the case of the material loss which took different, continuous values.

2.2 | Statistical independence tests

The cause-effect relationship between the corrosion variables present in the data set are analysed from a statistical point of view. The null hypothesis considered is a lack of any causal relationship. As such, results are presented with their significance level α, which represents the probability of incorrectly rejecting the null hypothesis when it is true. The confidence level (CL) follows the relation CL = 1-α.[22,23]

Causal discovery methods use conditional independence tests (CIT) to identify causal links between the variables of the data set and attempt to eliminate the spurious correlations within those variables.[24] The independence between the set of nodes X and Y, conditional to a set of nodes Z, is written as $X \perp Y \mid Z$.[4] The independence between variables can be inferred locally from the CIT, but also globally from the causal structures present in the graph and how they are connected. Analysing all the causal paths connecting any variables, it can be inferred what is their causal relationship. A criterion commonly used in this regard is the d-separation of variables.[4]

Causal discovery algorithms use various statistical tests to assess the independence between variables, such as the χ^2 test (nonparametric test measuring the goodness-of-fit between expected and observed frequencies which works well with large discrete categorical samples),[25,26] Fisher-Z test (used for partial and zero correlation, this parametric test assumes that the variables are normally distributed and works mainly on large sample sizes of continuous variables),[26,27] G-squared test (nonparametric test that highlights relationships between categorical variables with more than two levels),[25] or the Kernel-based conditional independence test (KCI). In this work, it was used the latter.

The KCI test is a nonparametric test that can be derived from the kernel matrices of the variables under consideration, which characterise the similarity of the samples of those variables.[28,29] These kernel functions recognise nonlinear relationships between data points. This test can be applied to discrete or continuous variables. Figure 3 shows the evolution of the accuracy of the test as a function of the number of samples for the data set analysed in the original publication of the method.

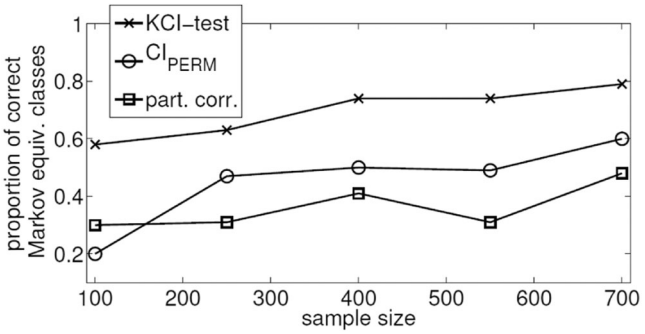


FIGURE 3 Accuracy of different conditional independence tests to infer the correct Markov equivalence class as a function of the sample size.[28]

2.3 | Causal links, structures and graphs

The different causal nature of the relationships between the variables can be represented with different types of causal links,[30] as illustrated in Table 1. The relationship depends on the variables measured and on possible confounding variables (those unmeasured variables which influence the underlying causal mechanism).[31]

After applying the CIT to the variables from the data set, the result of the causal discovery algorithm is a graphical representation of the causal links between the variables called a causal graph.[30] Chains, forks, and colliders (also called v-structures) are the three building blocks used in these causal graphical models to illustrate the cause-effect relationship between the variables.

Figure 4a shows the chain structure wherein $X \to Y \to Z$ forms a chain where X causes Y and Y causes Z, therefore the conditional independence can be written as $X \perp Z \mid Y$.[34] Figure 4b shows the fork structure $Y \leftarrow X \to Z$ wherein the node X forms a directed edge towards Y and Z making it the common ancestor for Y and Z.[34] Since there is only one path between Y and Z through X based on the d-separation criterion, conditional independence can be written as $Y \perp Z \mid X$. Figure 4c shows a collider $X \to Z \leftarrow Y$, wherein the descendant Z has two common ancestors X and Y. Even though there is one path between X and Y, the presence of a collision node Z makes X and Y conditionally dependent given the collision node Z, that is $X \not\perp Y \mid Z$. Figure 4d shows a collider structure with extra descendant W where $X \not\perp Y \mid Z$ and $X \not\perp Y \mid W$.

A causal directed acyclic graph (DAG) consists of a set of random variables with edges between them which never form a directed cycle within the graph.[4] Generally, causal discovery algorithms do not allow identifying the causal

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Different types of causal links and their respective denotations ignoring any selection bias.[31–33]

Causal link		Description
A → B (directed)		A is the cause of B
A—B (undirected)		Undetermined. A can cause B and B can cause A
A ↔ B (bidirected)		A and B do not cause each other but have a latent common cause
		Confounding variables between A and B
A o → B (partially directed)	A → B	A causes B ('o' turns into tail end)
	A ↔ B	There exists a confounder between A and B ('o' turns into arrow end)
A o—o B (undirected with 'o' ends)	A → B	A causes B
	A ← B	B causes A
	A ↔ B	A and B do not cause each other. Confounder between A and B
		This last option can be also combined with the previous two

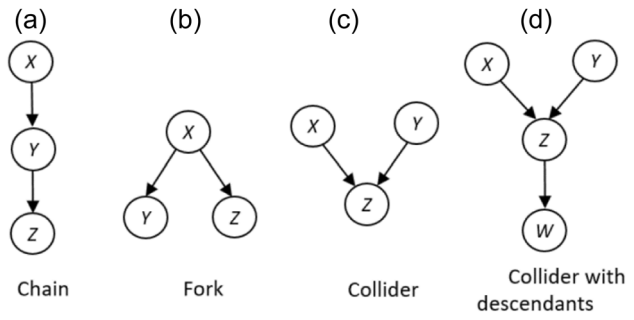


FIGURE 4 Building blocks of a Causal Graph.[30]

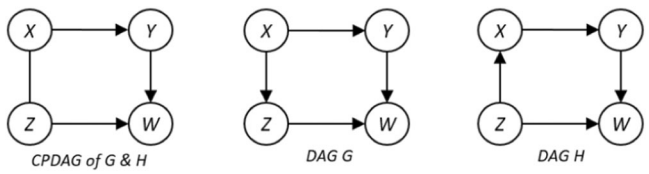


FIGURE 5 Example of Markov equivalence class.[30]

graph, but the Markov equivalence class (MEC) of the graph. If two DAGs are Markov equivalent, they have the same skeleton and set of colliders as well as the same (conditional) independencies,[24] which is what the algorithms can usually identify. Once the class is identified, using interventions in the variables it is then possible to identify the actual causal DAG.[35]

The example taken from Hasan et al.[30] in Figure 5 shows a completed partially DAG (CPDAG) of G and H which represents the union of the Markov equivalent DAGs G and H. The undirected edge between X and Z in the CPDAG suggests that it might contain $X \rightarrow Z$ (shown in the DAG G) or $Z \rightarrow X$ (shown in the DAG H).

Table 2, shows the different edges that can be observed in each type of causal graph. Each causal discovery algorithm produces as an output a different type of causal graph.

2.4 | Causal discovery algorithms

The two main categories of causal discovery algorithms are constraint-based and score-based methods. The first checks the graph structure against the independence constraints imposed by the data. In the second method, possible graphs are scored for their ability to fit the data. In the latter, the space of DAGs is searched to find the graph that maximises the score. This last method is especially useful when dealing with a large number of variables since the combinatorial space of possible graphs grows exponentially. In this work, since a small number of variables are studied it is used a constraint-based method.

The constraint-based algorithms use CITs to investigate the type of edges between the variables or their absence.[5] One of the earliest and most common of these algorithms is the PC (Peter & Clark) Algorithm.[40] It uses CITs to understand the underlying causal mechanism of the causal structures. It assumes independent and identically distributed (i.i.d) samples and absence of confounding variables.

Here it uses the fast causal inference (FCI) algorithm.[40] It is a variant of the PC algorithm that provides asymptomatically correct results while considering the presence of confounding variables in a data set with i.i.d samples.[5] The output causal graph of the FCI algorithm is a partial ancestral graph (PAG), including the presence of directed, undirected, partially directed and bi-directed edges.[30]

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Causal Graphs and their types of edges.[30,36–39]

	Directed (→)	Undirected (—)	Bi-directed (↔)	Partially directed (o →)
Directed acyclic graph (DAG)	X			
Partially DAG (PDAG)	X	X		
Completed PDAG (CPDAG)	X	X		
Maximal ancestral graph (MAG)	X	X	X	
Partial ancestral graph (PAG)	X	X	X	X

TABLE 3 Categories of variables present in the datasets for highest (HML), median (MML), and lowest (LML) material loss.

Mat. Loss/Data set	Materials				Temperature (°C)		Amount of salt (μg/cm²/h)			HCl (ppm)	
HML	1	2	3	4	700	900	0.5	1.5	5.0	0	100
MML	1	2	3	4	700	900	0.5	1.5	5.0	0	100
LML	1	2	3	4	700	900	0.5	1.5	5.0	0	100

2.5 | FCI algorithm implementation

The causal-learn[1] package[41] was used for implementing the FCI algorithm. The data set used as the input of the FCI algorithm included the variables shown in Table 3.

From the initial seven variables available, mentioned in Section 2.1.1, SO_2 and Time of Exposure were removed from the data set because of their constant values. The rest of the variables were used as input to the algorithm. The four materials used in the experiments were assigned numbers from 1 to 4, respectively. The algorithm was applied with the default parameter settings, CIT = KCI and no background knowledge. Since the data set consisted of only five variables, no background knowledge was introduced into the causal algorithm to avoid selection- or expert-bias. Therefore, the causal graphs formed were purely based on the observational data. A range of 1%–99% significance level, with a 1% step increment, was implemented to observe how the causal links would form with high and low confidence levels.

3 | RESULTS

This section presents the causal graphs obtained by implementing the FCI algorithm on the material loss datasets. The descriptions of the algorithms, graphs and types of links can be found in Sections 2.3 and 2.4.

Based on the degree of material loss, the data set was divided into three parts: HML, MML and LML (see Table 3). Varying significance levels from 1% to 99% were considered in the FCI algorithm. The results are differentiated according to HML, MML and MML and the significance level of each graph.

Table 4 shows the causal graphs obtained by applying KCI in the FCI algorithm on the HML data set. The 'o' termination in the causal link formed between amount of salt and material loss at $\alpha = 1\%$, illustrates that it can be either an arrowhead (>) or a tail end of a directed edge (see Table 1). Hence, the direction of the causal relationship is not clearly depicted at $\alpha = 1\%$ between the two variables.

At $\alpha = 7\%$, partially directed edges 'o→' from amount of salt and HCl to material loss were formed. This means that if 'o' becomes an arrowhead, it forms a bidirected edge. Table 1 shows that the bidirected edge indicates that there is an unmeasured confounder present between the two variables. On the other hand, if 'o' becomes a tail end, it confirms that the amount of salt and HCl causes material loss. Similarly, temperature and material formed partially directed edges with material loss at $\alpha = 8\%$ and $\alpha = 60\%$ for HML conditions, respectively.

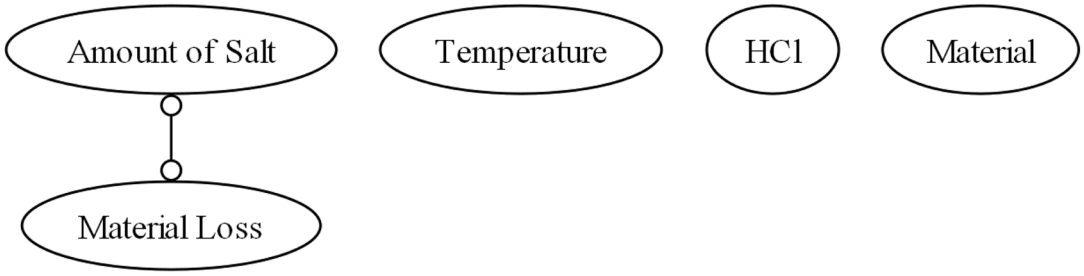
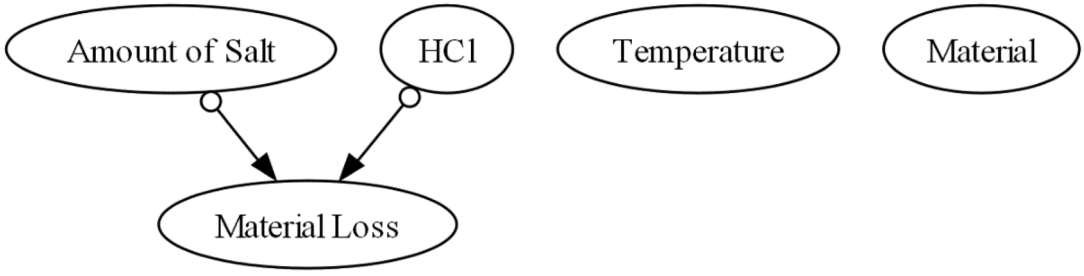
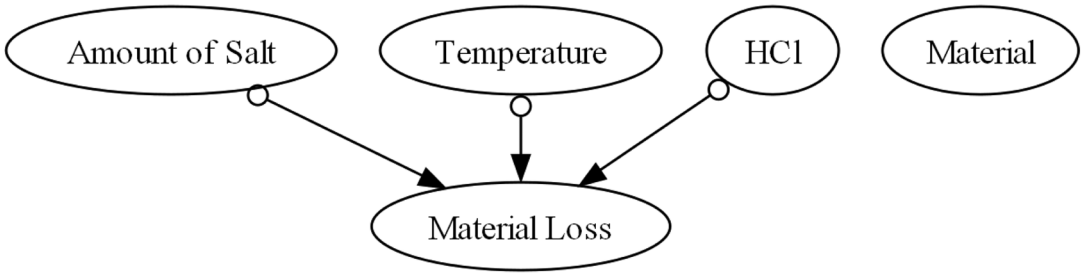
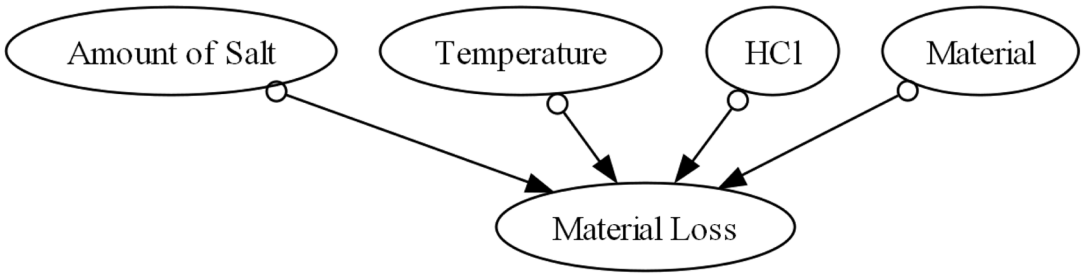
Table 5 shows the results for the MML data set. The undirected edges with 'o' ends which formed between amount of salt and material loss at $\alpha = 1\%$, get converted to a partially directed edge from the amount of salt to material loss, along with a partially directed edge from temperature to material loss at $\alpha = 9\%$. Gradually, partially directed edges were formed from HCl and material to material loss at $\alpha = 20\%$ and $\alpha = 60\%$, respectively.

The causal graphs for the LML data set are presented in Table 6. At $\alpha = 1\%$ for the LML data set, an undirected edge with 'o' ends was formed between the amount of salt and material loss. With the increase in α to 20%, 50%

[1]https://github.com/py-why/causal-learn.

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Causal graphs using Kernel-based conditional independence test (KCI) in fast causal inference (FCI) algorithm with increasing significance levels for highest material loss (HML) data set.

α	Causal Graphs
0.01	
0.07	
0.08	
0.6	

and 70%, partially directed edges were formed from amount of salt and temperature, HCl and material to material loss, respectively.

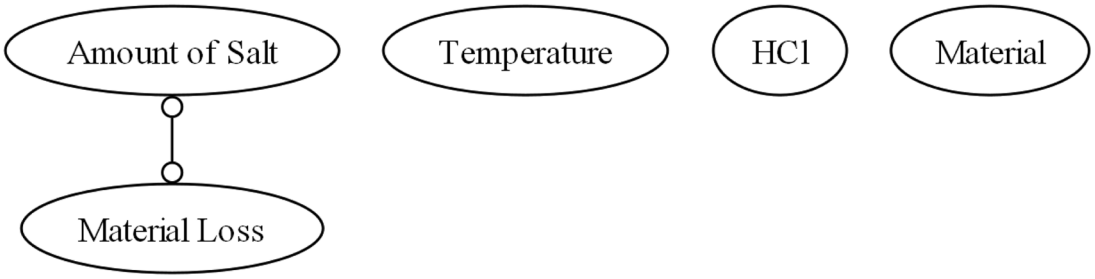
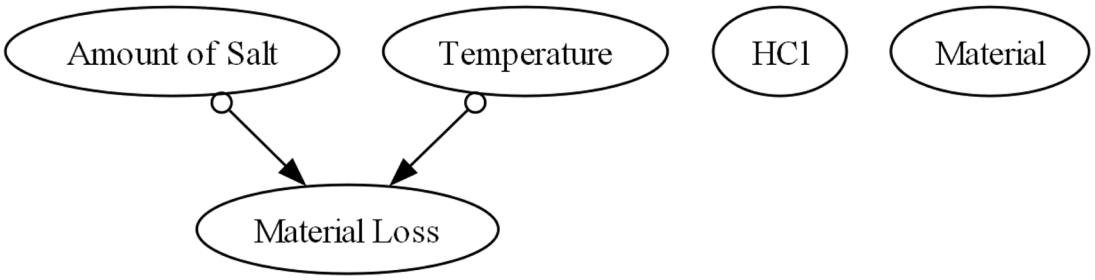
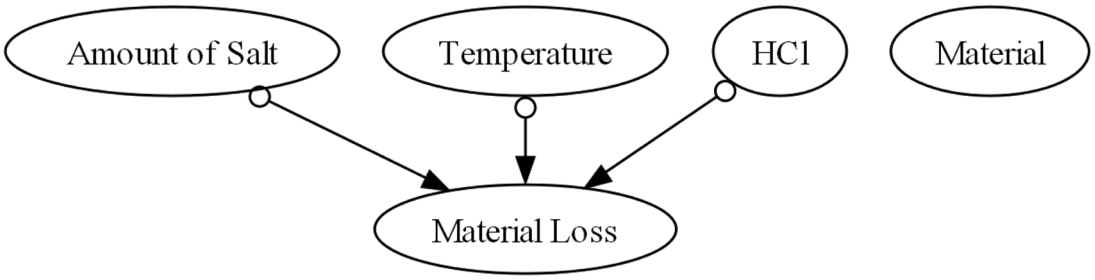
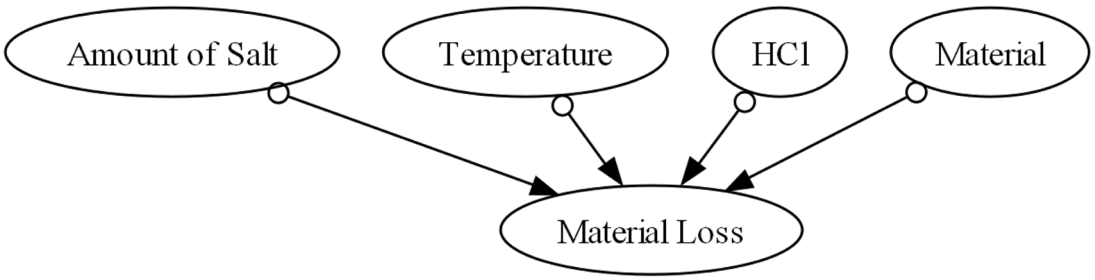
4 | DISCUSSION

In this section, it is discussed the nature of the current corrosion data set and its drawbacks, as well as the causal graphs formed by the FCI algorithm and their key findings.

The type of data set that is used as an input to the algorithm is one of the major factors determining the result. The current corrosion data set was designed to maintain control over the corrosion variables to understand the effects on the material loss. The causal graphs in Tables 4–6, show that all the corrosion variables are directed towards material loss with increasing significance values. This means that the variables were controlled to observe the amount of material loss in the base material. However, for a pure causal inference study, randomisation in the variables would have offered a greater benefit.[40] For instance, if all the variables were uncontrolled and randomised, observing causal links within the variables could have been a possibility. This would have helped in

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Causal graphs using Kernel-based conditional independence test (KCI) in fast causal inference (FCI) algorithm with increasing significance levels for median material loss (MML) data set.

α	Causal Graphs
0.01	
0.09	
0.2	
0.6	

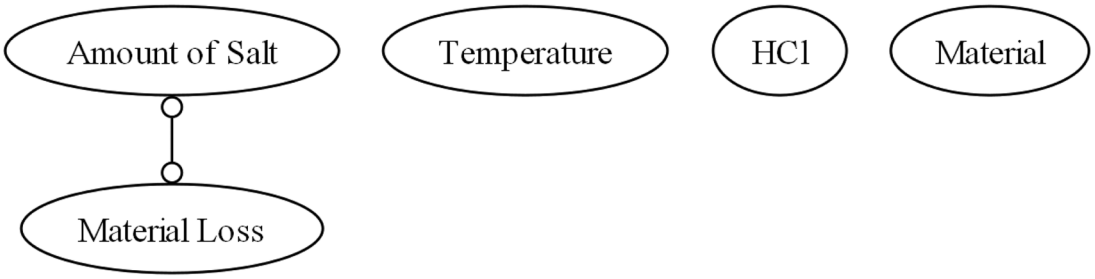
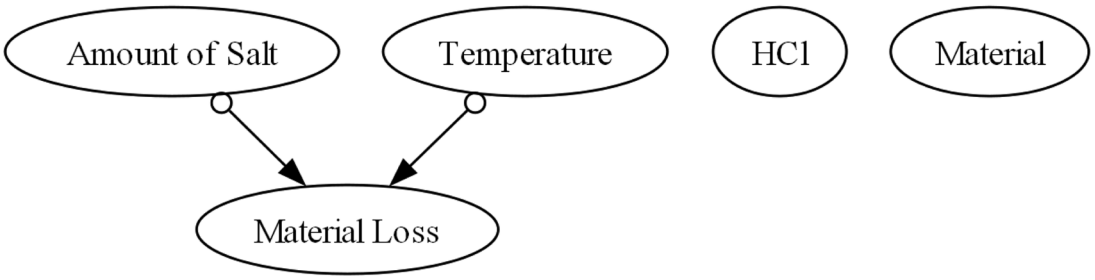
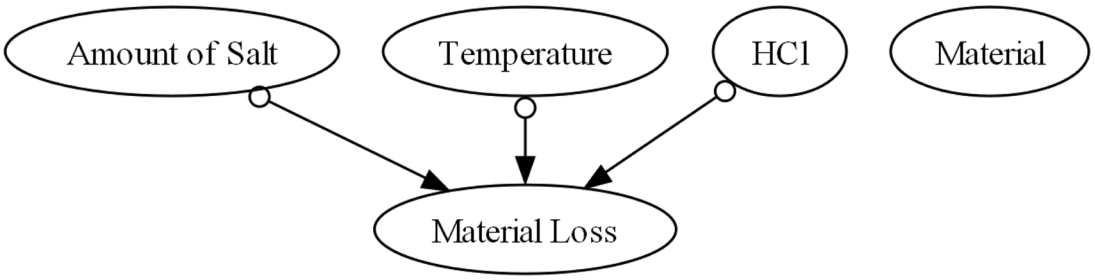
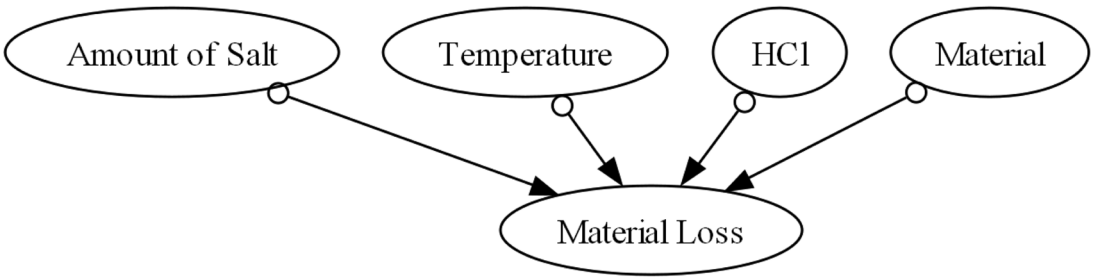
understanding the cause-effect relationship between the variables and not just through the material loss. The degree of influence that each variable has on material loss with varying significance levels can be observed in Tables 4–6. However, no claims can be made about the presence of causal relationships between 'amount of salt', 'temperature', 'HCl', and 'material' using the current data set.

As discussed in Sections 1.1.1 and 1.1.2, the salt deposition on the material surface is a significant factor in causing hot corrosion and, indeed, hot corrosion is also defined as a deposit-induced accelerated form of corrosion.[3] The salt deposits initiate the breakdown of the protective oxide layer of the substrate which is the starting point of hot corrosion.[2] Therefore, even at smaller significance values, the amount of salt makes a directed or an undirected causal link with material loss for all the material loss levels (HML, MML, and LML).

Operating temperature facilitates a corrosive environment that enables hot corrosion.[3] Moreover, the lower melting point of the eutectic mixtures formed by different salts accelerates the corrosion process,[2] hence proving that temperature is another factor that has a significant influence on hot corrosion. Indeed, for MML and LML temperature becomes the second significant

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Causal graphs using Kernel-based conditional independence test (KCI) in fast causal inference (FCI) algorithm with increasing significance levels for lowest material loss (LML) data set.

α	Causal Graphs
0.01	
0.2	
0.5	
0.7	

variable that influences hot corrosion in the given data set, at smaller significance values. For MML and LML HCl in the gas only appears to form a causal link with material loss at higher significance levels. Chloride-contaminated oxide layers are known to accelerate the corrosion rate.[42] HCl is a crucial factor that participates in increasing hot corrosion rates. More experimental research is required to clarify the influence of HCl. However, this analysis shows its significance for driving extreme metal loss such as pitting.

Table 4, formed using the HML data set, shows that at 1% significance level, the outcome from FCI shows an undirected causal link between the amount of salt and material loss with 'o' at either end. This means that at such a conservative significance value, FCI is only able to infer the existence of a link between the two variables without pointing out the cause-effect relationship. The undirected causal link can also point towards the probability of incorrectly accepting the null hypothesis that is, Type-II error. The 'o' ends can be perceived as an arrowhead or a tail end of a directed edge because the algorithm is too conservative in assigning them to form a meaningful causal link. In case both 'o' ends become arrowheads, then it means

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that there is a confounding variable that has not been considered in the data set.

In Table 4, with a significance value of 7%, the three partially directed (o→) causal links confirm that there is a possibility of amount of salt, HCl, and temperature being factors causing material loss, or there exist unmeasured confounders between them that are influencing the behaviour of those variables. The FCI algorithm is not able to clarify the causal relationship between these variables and shows the possibility that confounder variables are present. HCl forms the same partially directed edge with material loss and a similar argument can be made for this causal link as well. Under the specific conditions of the test, it seems that HCl is linked to HML, so it seems as HCl is one of the driving forces involved in the formation of pitting. This provides a very useful insight into the mechanism of hot corrosion, as the cause of pit initiation is still under debate. From $\alpha = 7\%$, temperature also forms a partially directed causal link with material loss. This suggests that for the HML data set, HCl and temperature are the most influential factors advancing material loss, after amount of salt. There is also a possibility that unmeasured confounders exist which influence all the said variables. Therefore, datasets with a larger sample size and greater number of variables can further help in understanding the presence of these confounders.

Table 5 shows that the FCI algorithm is not confident enough in forming causal links between the variables up to 9% significance value, for the MML data set. At this value, both amount of salt and temperature form partially directed causal links pointing towards material loss. The amount of salt and temperature are observed to be the dominant factors accelerating the material loss because the salt deposits initiate the breaking of the protective layer of the base material and high operating temperature facilitates such corrosion mechanisms. There is a possibility that the inclusion of variables like partial pressure would have addressed the presence of the unmeasured confounder at LTHC (700°C). This is because LTHC is a function of temperature and partial pressure, as discussed in Section 1.1.2.

For the LML condition in Table 6, the undirected causal link with 'o' ends formed between amount of salt and material loss at 1% significance value suggests that at such a low level of material loss, amount of salt might have just started to break down the protective layer to cause the material loss. Or there is an unmeasured variable that exists which is influencing both variables.

Results beyond the 10% significance level are not discussed since they are not statistically significant.

The causal links shown in the previous tables suggest the possible presence of confounding variables. This opens up the avenue to design experiments that also cater to other variables that were not included in this data set. The FCI algorithm indicates that for the experimental conditions represented in this study's limited data set, amount of salt has the maximum influence on material loss.

5 | CONCLUSIONS

Understanding the underlying causal mechanisms between the factors leading to hot corrosion in gas turbines can highlight the inner workings of this process. Implementing causal discovery methods can help in recognising the causal relationships between these corrosion factors. These methods were applied to three different datasets that were divided based on the degree of material loss observed on the materials tested: highest, median, and lowest material loss. The causal discovery algorithm FCI was applied to produce the causal graphs that illustrate the causal relationships between the corrosion variables present in these three datasets. A wide range of significance levels was analysed, to showcase the confidence level with which the causal relationships were formed between the corrosion variables.

After analysing the causal graphs for the given range of significance levels, it was observed that the number of causal links decreased in the order of highest to lowest material loss. As the degree of material loss decreased to the median range, only two causal links that is, amount of salt and temperature to material loss were formed within the set range of significance levels. Eventually, only a single undirected causal link was formed between amount of salt and material loss for LML conditions. This showcased that there is uncertainty in claiming the real causal relationship between the two variables at low degrees of material loss. Causal graphs produced using the FCI algorithm suggested the possible existence of unmeasured confounding variables.

From this nascent-stage study of causality in hot corrosion, it can be concluded that amount of salt and temperature were the typical factors causing material loss from an overall perspective. However, HCl also proved to be a dominant factor for the HML data set. As this data set can include pitting regions this can give insight into factors driving pit formation rather than more average metal loss.

The FCI algorithm proved beneficial in understanding the causal relationships, but a randomised and uncontrolled type of corrosion data set can further help in future causal research.

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Since this study is at a nascent phase of its research timeline, several avenues for future research can be delved into to improve the understanding and application of causal discovery techniques. Following are the recommendations that can potentially improve the quality of further studies:

1. A randomised data set which can help in performing a more comprehensive causal discovery study should be produced.
2. Including variables such as partial pressures of the gaseous contaminants, crystalline structures, and varying gas compositions can improve the depth of the data set.
3. The sample sizes must be increased to generate stronger causal links with higher confidence levels.
4. If a time-dependent data set is produced, causal discovery algorithms for time series data can be implemented such as Granger Causality-based algorithms.[30]

These recommendations can help in designing the next phase of the causal discovery study of hot corrosion. Expanding upon this study can further help in developing predictive maintenance and material degradation detection models.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Miguel Arana-Catania http://orcid.org/0000-0003-4277-0292
Stefano Mori http://orcid.org/0000-0001-5463-5691
Adriana Encinas-Oropesa http://orcid.org/0000-0002-0306-4718

REFERENCES

[1] R. T. C. Harman, *Applications for Gas Turbine Engines*, Macmillan Education UK, London **1981**, pp. 18.
[2] N. Eliaz, G. Shemesh, R. M. Latanision, *Eng. Fail. Anal.* **2002**, *9*, 31.
[3] F. Pettit, *Oxid. Met.* **2011**, *76*, 1.
[4] J. Pearl, *Causality*, Cambridge University Press, **2009**.
[5] C. Glymour, K. Zhang, P. Spirtes, *Front. Genet.* **2019**, *10*, 524.
[6] N. Rao, Materials for Gas Turbines - An Overview, in *Advances in Gas Turbine Technology*, InTech. **2011**. https://doi.org/10.5772/20730
[7] J. Sumner, A. Encinas-Oropesa, N. J. Simms, J. R. Nicholls, *Oxid. Met.* **2013**, *80*, 553.
[8] P. W. Schilke, A. D. Foster, J. J. Pepe, *Advanced Gas Turbine Materials and Coatings Volume 3569, Part 2 of GER/General Electric Company*, General Electric Company, **1991**.
[9] S. Mori, M. Mendil, J. Wells, N. Chapman, N. Simms, R. Wells, J. Sumner, *Energy Syst.* **2023**. https://doi.org/10.1007/s12667-023-00612-0
[10] P. C. Gasson, *Aeronaut. J.* **2008**, *112*, 291.
[11] A. Encinas-Oropesa, *PhD Thesis*, Cranfield University, UK **2005**.
[12] M. J. Donachie, S. J. Donachie, *Superalloys: A Technical Guide*, ASM International, **2002**.
[13] G. A. Kool, *J. Ther. Spray Technol.* **1996**, *5*, 31.
[14] G. Goward, *Surf. Coat. Technol.* **1998**, *108*, 73.
[15] N. Vialas, D. Monceau, *Oxid. Met.* **2006**, *66*, 155.
[16] W. T. Wu, A. Rahmel, M. Schorr, *Oxid. Met.* **1984**, *22*, 59.
[17] M. Pomeroy, *Mater. Des.* **2005**, *26*, 223.
[18] M. Vuković, S. Thalmann, *J. Manufact. Mater. Process.* **2022**, *6*, 10.
[19] C. Feng, X. Liu, S. Bi, J. Kang, Degradation State Assessment Modeling Using Causality Discovery. in *2022 Prognostics and Health Management Conference (PHM-2022 London)*. IEEE London, United Kingdom **2022**, 545. https://doi.org/10.1109/PHM2022-London52454.2022.00102
[20] X. Chen, J. Abualdenien, M. M. Singh, A. Borrmann, P. Geyer, *Energy Build.* **2022**, *277*, 112583.
[21] X. Shen, S. Ma, P. Vemuri, G. Simon, *Sci. Rep.* **2020**, *10*, 2975.
[22] C. Stephen, *Loftus in Basic Statistics with R*, (Ed. S. C. Loftus), Academic Press, **2022**.
[23] A. Banerjee, U. Chitnis, S. Jadhav, J. Bhawalkar, S. Chaudhury, *Ind. Psychiatry J.* **2009**, *18*, 127.
[24] J. Peters, D. Janzing, B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*, The MIT Press, Cambridge, USA **2017**.
[25] N. Cressie, T. R. C. Read, *International Statistical Review/Revue Internationale de Statistique* **1989**, *57*, 19.
[26] A. Agresti, *Categorical Data Analysis*, 792, John Wiley & Sons, New York **2012**.
[27] J. D. Ramsey, *arXiv preprint arXiv:1401.5031* **2014**.
[28] K. Zhang, J. Peters, D. Janzing, B. Schölkopf, *arXiv preprint arXiv:1202.3775* **2012**.
[29] J. Yang, Y. Liu, Y. Liu, W. Sun, *Ann. Appl. Statist.* **2021**, *15*, 194.
[30] U. Hasan, E. Hossain, M. O. Gani, *arXiv preprint arXiv:2303.15027* **2023**.
[31] A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, J. Gama, *Wiley Interdiscipl. Rev. Data Mining Knowl. Discov.* **2022**, *12*, e1449.
[32] J. Zhang, *Artif. Intell.* **2008**, *172*, 1873.
[33] J. Zhang, *J. Mach. Learn. Res.* **2008**, *9*, 1437.
[34] J. Pearl, M. Glymour, N. P. Jewell, *Causal Inference in Statistics: A Primer*, John Wiley & Sons, New York **2016**.
[35] Y. He, J. Jia, B. Yu, *J. Mach. Learn. Res.* **2015**, *16*, 2589.
[36] T. Richardson, P. Spirtes, *Ann. Statis.* **2002**, *30*, 962.
[37] E. Perkovic, Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), in Proceedings of Machine Learning Research* **2020**, *124*, 530. https://proceedings.mlr.press/v124/perkovic20a.html
[38] J. Tian, *arXiv preprint arXiv:1207.1428* **2012**.
[39] Z. Hu, R. Evans, In *Conference on Uncertainty in Artificial Intelligence*. PMLR. **2020**, 739.

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[40] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction, and Search*, Adaptive Computation and Machine Learning. CogNet **2000**.
[41] Y. Zheng, B. Huang, W. Chen, J. Ramsey, M. Gong, R. Cai, S. Shimizu, P. Spirtes, K. Zhang, *arXiv preprint arXiv:2307.16405* **2023**.
[42] P. Viklund, R. Pettersson, *Oxid. Met.* **2011**, *76*, 111.

How to cite this article: A. Varghese, M. Arana-Catania, S. Mori, A. Encinas-Oropesa, J. Sumner, *Mater. Corros.* **2024**, 1–13.
https://doi.org/10.1002/maco.202314240

15214176, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/maco.202314240 by Test, Wiley Online Library on [14/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



https://dspace.org/

Causal discovery to understand hot corrosion

Varghese, Akhil; Arana-Catania, Miguel; Mori, Stefano; Encinas-Oropesa, Adriana; Sumner, Joy

2024-12-01
Wiley
https://doi.org/10.1002/maco.202314240
Varghese A, Arana-Catania M, Mori S, et al., (2024) Causal discovery to understand hot corrosion, Materials and Corrosion. Volume 75, Issue 12, December 2024, pp. 1639-1651

Downloaded from a DSpace repository. https://dspace.org/