Causal Inference for Root Cause Analysis in Safety-Critical Engineering Systems

Evangelia Perivolaki[1], Steve King[2], Irene Moulitsas[3]

[1,2,3]*Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK*

[1]*Rolls-Royce plc, Derby, DE24 8BJ, UK*

ABSTRACT

Root-cause analysis (RCA) in safety-critical engineering relies heavily on correlation-based methods and expert judgment. These approaches are useful but limited: they can identify which signals are associated with a fault, but not why that fault occurred or which factors genuinely caused it. This research explores how causal inference techniques, combined with domain engineering knowledge, can produce more reliable, explainable, and reusable diagnostic tools for Prognostics and Health Management (PHM). The work is grounded in a structured literature review and early prototyping and will be validated on a real industrial dataset from a large civil aerospace engine programme.

1. MOTIVATION AND PROBLEM STATEMENT

Complex engineering systems such as gas-turbine engines generate large volumes of operational data through health monitoring systems, flight data recorders, and maintenance records. Despite this data richness, fault investigations tend to follow the same basic workflow: engineers look for signals that correlate with the problem, draw on accumulated experience to form hypotheses, and apply judgement to narrow down a cause. This process works, but it is slow, difficult to audit, and heavily dependent on who is doing the investigation.

The fundamental issue is that correlation is not causation. A signal that correlates strongly with a fault may be a symptom, a cause, or simply a feature that happens to co-vary due to shared operating conditions. Machine learning methods are good at detecting and predicting anomalies, but they do not explain the mechanisms that generate them. In regulated industries, this matters: decisions about safety, maintenance, and design need to be justified, traceable, and physically plausible, not just statistically associated.

A concrete example of this gap can be seen in aircraft engine performance analysis. When specific fuel consumption (SFC) varies between nominally identical engines, or shifts abruptly during a test, engineers need to understand which manufacturing or assembly factors are responsible. Current practice uses correlation analysis (e.g., Pearson coefficients, t-tests) to screen candidate features. This identifies associations but cannot confirm whether a given factor is a cause, a downstream effect, or a confounded indicator. The result is that investigations remain expert-intensive and their conclusions are difficult to reuse systematically.

This research investigates whether modern causal discovery methods, constrained and guided by engineering knowledge, can address this gap. The aim is not to replace engineering expertise but to formalise and augment it by producing causal graphs that are both data-driven and physically consistent, with explicit uncertainty quantification and the ability to accumulate knowledge across investigations.

2. RESEARCH QUESTIONS

The research is organised around four questions, each addressing a distinct aspect of the problem:

RQ1. How can engineering knowledge from system architecture, physical models, and failure analysis documents be encoded as constraints to guide causal discovery from observational data, rather than letting algorithms search unconstrained?

RQ2. How can we quantify and communicate uncertainty in the resulting causal graphs in a way that is meaningful for engineers making safety-critical decisions?

RQ3. Does a causally grounded investigation workflow outperform correlation-based methods in practice-in terms of accuracy, speed, or interpretability-and under what conditions?

RQ4. Can the knowledge produced by one investigation-the inferred causal structure, the supporting evidence, and the uncertainties-be stored and reused to accelerate future investigations on related systems?

3. PROPOSED APPROACH

The core idea is to combine two things that are currently kept separate: engineering knowledge about how a system works, and statistical causal discovery applied to its operational data. Neither on its own is sufficient. Pure data-driven causal discovery can produce graphs that are statistically plausible but physically impossible-for example, suggesting that a downstream symptom causes an upstream fault. Engineering knowledge alone, codified in diagrams and failure mode analyses, is incomplete and may miss causal pathways that only become apparent in data.

The proposed approach uses engineering knowledge to define constraints on the causal search-ruling out physically impossible causal directions and giving preference to known pathways-while applying an ensemble of improved and existing causal discovery algorithms (such as PC, GES, and DirectLiNGAM) to identify non-obvious causal relationships from data. The outputs of the algorithms are reconciled with the engineering constraints to produce a causal graph in which each edge has a confidence score reflecting how consistently it is supported across algorithms and how well it aligns with prior knowledge.

Uncertainty quantification is a central part of the design. There are two distinct types of uncertainty to address: uncertainty about the graph structure itself (which edges are present, in which direction) and uncertainty about the size of causal effects. Both need to be expressed in a form that is interpretable to an engineer, not just to a statistician. How to do this well is one of the open questions the research needs to answer.

A further element of the design is a knowledge base that stores the results of each investigation in a structured way, so that findings from one fault case can inform future analyses. The idea is that if a causal pathway has been identified and validated in one investigation, it should not have to be rediscovered from scratch the next time a similar fault occurs.

Several methodological questions remain open at this stage and addressing them will likely require developing new or adapted techniques rather than simply applying existing methods. Key challenges include: determining how to optimally weight and reconcile outputs from different causal discovery algorithms when they disagree; extending standard causal discovery approaches to handle feedback loops and non-stationary dynamics, which are common in complex engineering systems but typically ignored by current methods; and designing uncertainty outputs that are genuinely interpretable and actionable for practising engineers rather than remaining theoretical constructs.

4. WORK COMPLETED SO FAR

4.1. Literature Review

A structured review of the causal discovery and causal inference literature has been completed, covering constraint-based methods (PC, FCI), score-based methods (GES, KCRL), and functional causal models (DirectLiNGAM, NO-TEARS), as well as applications to engineering and health monitoring domains. The review produced several findings that directly shape the research design.

Modirrousta et al. (2024) showed that integrating physical process knowledge as soft constraints substantially improves causal graph recovery in industrial time-series settings, compared to unconstrained baselines. This provides support for the core design decision to use engineering knowledge as a prior rather than letting algorithms search freely.

Fu et al. (2022) demonstrated that encoding prior domain knowledge as hard and soft constraints within a score-based causal search (KCRL) recovers more accurate causal structures than unconstrained methods like DirectLiNGAM and GES. This is directly relevant to the proposed approach, which encodes engineering knowledge (from failure mode and effects analyses and system architecture documents) as constraints on the causal search space.

Petersen et al. (2025) evaluated the PC algorithm against expert-defined causal structures in a clinical study and found that fully automated causal discovery can produce biased results when used without domain expertise. Their conclusion that algorithms should augment rather than replace expert knowledge reinforces the hybrid design of this research and clarifies what RQ1 is really asking: not whether to use engineering priors, but how to formalise and weight them well.

A recurring limitation across the literature is that most causal discovery methods assume acyclicity and are designed for cross-sectional or simplified longitudinal data. Aerospace health monitoring data typically involves feedback loops, time-varying dynamics, and non-stationary operating conditions-all of which violate these assumptions. This is a known open problem and shapes the longer-term direction of the research.

4.2. Industrial Validation Case

A validation case has been identified and scoped with the industrial partner, who operates a large civil aerospace engine programme. The problem is one of genuine industrial interest: why does specific fuel consumption vary between engines built to the same specification, and what causes abrupt performance shifts observed during testbed measurements?

The available dataset is a combination of manufacturing records (geometric measurements, component clearances, nozzle guide vane areas, and build configurations) and comprehensive testbed performance data across multiple engine variants and test sites. Current analysis relies on Pearson correlation and t-tests to identify candidate features. These identify associations but cannot determine whether a manufacturing feature is a genuine cause of SFC variation or simply a correlated indicator. This is exactly the distinction the proposed causal approach is designed to make.

This case is valuable for the research for several reasons. It provides a real, multi-source industrial dataset for validation beyond simulated experiments. The problem structure-manufacturing variation influencing engine thermodynamics and ultimately observable performance-is well matched to the causal inference approach being developed. And the existence of the current correlation-based workflow gives a natural baseline against which to evaluate any improvement (RQ3).

5. EXPECTED CONTRIBUTIONS

This research aims to produce the following contributions:

1 A method for hybrid causal discovery that combines engineering knowledge constraints with data-driven causal search, producing physically consistent causal graphs with associated confidence scores. This addresses a gap in the literature, where causal reasoning is rarely used and engineering priors are not formally integrated into graph discovery.

2 An uncertainty quantification framework for causal graphs in engineering contexts, designed to communicate confidence in causal claims in a way that supports decision-making and regulatory documentation.

3 A structured knowledge base for storing and reusing causal investigation outcomes, enabling cumulative learning across fault cases rather than treating each investigation in isolation.

4 Empirical validation on real industrial datasets, benchmarked against the current correlation-based workflow, providing evidence on whether and under what conditions causal methods offer practical advantages for PHM engineers.

6. PROPOSED RESEARCH PLAN

The PhD is a six-year part-time programme. The research is currently in the first year, with the following phases planned:

Year 1 – Literature review and scoping. Systematic review of causal inference and PHM literature to establish theoretical foundations and identify methodological gaps. In parallel, industrial use cases are being scoped with the research sponsor to ensure the framework is tested on problems of genuine practical relevance.

Years 2–3 – Framework design and prototype development. The methodology will be formalised and implemented. Small-scale validation on the identified industrial use cases will begin using historical data.

Years 4–5 – Full validation and comparative evaluation. The framework will be applied to complete industrial datasets and benchmarked against existing correlation-based approaches. This phase will also address the more advanced components of the research, including real-time reasoning and counterfactual analysis.

Year 6 – Synthesis and write-up. The full framework will be consolidated, findings across use cases synthesised, and the thesis written up. This phase will also document the limitations of the approach and directions for future work.

This plan will inevitably be refined as the research develops. At this stage, the priority is to complete the literature review and demonstrate the core approach on real industrial data.

REFERENCES

Chen, J., et al. (2024). Physics as the Inductive Bias for Causal Discovery. Preprint.

Fu, W., Pan, Y., Zhang, H., & Xiong, L. (2022). KCRL: A Prior Knowledge Based Causal Discovery Framework with Reinforcement Learning. AAAI, 36(8), 8055–8063.

Modirrousta, M., Memarian, A., & Huang, B. (2024). Causal discovery in industrial systems via physics-guided variational attention and probabilistic interventions.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference. Cambridge University Press.

Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of Causal Inference. MIT Press.

Petersen, M. L., et al. (2025). Can algorithms replace expert knowledge for causal inference? American Journal of Epidemiology, 194(5), 1399–1411.

Shimizu, S., et al. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian SEM. JMLR, 12, 1225–1248.

Spirtes, P., Glymour, C., & Scheines, R. (2001). Causation, Prediction, and Search. MIT Press.

Zheng, X., et al. (2018). DAGs with NO TEARS: Continuous optimization for structure learning. NeurIPS.