Computers and Concrete, *Vol. 31, No. 4 (2023) 277-292*
https://doi.org/10.12989/cac.2023.31.4.277

Causality, causal discovery, causal inference and counterfactuals in Civil Engineering: Causal machine learning and case studies for knowledge discovery

M.Z. Naser*[1,2] and Arash Teymori Gharah Tapeh[1]

[1]School of Civil & Environmental Engineering and Earth Sciences (SCEEES), Clemson University, USA
[2]Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson University, USA

(*Received November 9, 2022, Revised December 19, 2022, Accepted December 19, 2022*)

Abstract. Much of our experiments are designed to uncover the cause(s) and effect(s) behind a phenomenon (i.e., data generating mechanism) we happen to be interested in. Uncovering such relationships allows us to identify the true workings of a phenomenon and, most importantly, to realize and articulate a model to explore the phenomenon on hand and/or allow us to predict it accurately. Fundamentally, such models are likely to be derived via a causal approach (as opposed to an observational or empirical mean). In this approach, causal discovery is required to create a causal model, which can then be applied to infer the influence of interventions, and answer any hypothetical questions (i.e., in the form of What ifs? Etc.) that commonly used prediction- and statistical-based models may not be able to address. From this lens, this paper builds a case for causal discovery and causal inference and contrasts that against common machine learning approaches - all from a civil and structural engineering perspective. More specifically, this paper outlines the key principles of causality and the most commonly used algorithms and packages for causal discovery and causal inference. Finally, this paper also presents a series of examples and case studies of how causal concepts can be adopted for our domain.

Keywords: causal discovery; causal inference; civil engineering; machine learning

1. Introduction

Seeking causal knowledge is a foundational pursuit with branching philosophical, epistemological, and ontological ties [1]. Causal knowledge accurately describes how a phenomenon, Y, comes to be by answering key questions such as, what causes Y? How, when, and why does Y occur? Etc. (Bunge 1979). Arriving at precious answers to the above questions can be ambitious, and thus, causal knowledge can be difficult to pinpoint or measure (Schölkopf 2019).

In the civil and environmental engineering domain, we often design experiments in which a set of parameters is identified or assumed to be responsible for a given phenomenon. For example, say we are trying to understand the concept of flexural capacity in beams. First, we would fabricate a Beam A with certain geometrical and material features (i.e., W16×36, Grade A992). Then, we add boundary conditions (i.e., simply supported) to load this beam in a manner that enables us to capture its flexural response. We load the beam[2] and report that this beam fails once the level of the applied moment reaches 361.6 kN.m.

At this point, a link is then obtained by associating the reported moment at failure to the geometrical and material features as well as the loading configuration of Beam A. This link draws from the following observation: Applying a bending moment of 361.6 kN.m to Beam A has caused it to fail. A series of questions may arise: 1) What has caused the failure of Beam A? And 2) Was failure triggered due to the geometric configurations of W16×36? Or, perhaps due to the material properties of Grade A992? What about the effect of boundary conditions? Or Loading configuration?

The above are examples of causal questions. Another set of questions that may also arise includes: Other things constant [*Ceteris Paribus*] 3) Would Beam A have failed if it was not for the presence of the bending moment? 4) Would this beam fail at 361.6 kN.m if it had been a W18×40? Or had it been made from Grade A36? These are counterfactual questions that also belong to the causal family. Answering the above two sets of questions requires a causal investigation.

Fortunately, our domain knowledge can aid, if not substitute, the need for a thorough causal investigation. For example, we know that in the case of W-shaped steel beams under bending, the geometric features are lumped into the plastic modulus, Z, and that the material features are represented by the yield strength of structural steel, f_y. Both are also tied via Eq. (1), which represents a multiplication form to estimate the moment capacity (i.e., resistance) of a given W-shaped steel beam.

$$Resistance = Z \times f_y \tag{1}$$

*Corresponding author, Assistant Professor
E-mail: mznaser@clemson.edu
Website: www.mznaser.edu

[1] A cohesive look at causality from the lens of philosophy, epistemology, and ontology, can be found in (Michotte 2017, Salmon 2003).
[2] Say with one point load at mid-span.

http://www.techno-press.org/?journal=cac&subpage=8

ISSN: 1598-8198 (Print), 1598-818X (Online)

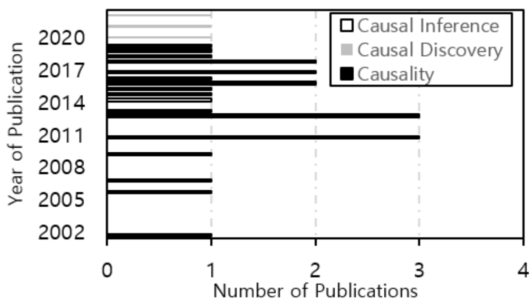


Fig. 1 Results of a scientometrics analysis on [causality], [causal discovery], and [causal inference] in structural engineering spanning 2002-2022

Eq. (1) presents the simple fact that the flexural resistance of a W-shaped steel beam is a function of Z and f_y. This equation also conveys that both Z and f_y contribute to the moment capacity in a complimentary (equal) manner. More importantly, this equation allows us to answer all the above four questions intuitively.

For instance:

1) What has caused the failure of Beam A? *Ans. The presence of a bending moment that exceeds the level of moment capacity.*

2) Was the beam's failure triggered due to the geometric configurations of W16×36? Or due to the material properties of Grade A992? *Ans. Building on Eq. 1, both Z and f_y have contributed equally.*

3) *Ceteris Paribus*, could Beam A has failed if it was not for the presence of the bending moment? *Ans. No. Since if it was not for such loading, the capacity of the beam would not have been exceeded.*

4) *Ceteris Paribus*, Would this beam fail at 361.6 kN.m if it had been a W18×40? Or had it been made from Grade A36? *Ans. No. Since W18×40 has a larger plastic modulus than W16×36, and Grade 36 has lower yield strength than Grade A992*

The identification of failure, as outlined above, may be perceived as a trivial problem[3]. This simplicity stems from the low-dimensional space of parameters involved in the phenomenon of flexural failure in W-shaped steel beams. However, the subject of causality exponentially grows with the addition of more dimensions to the space of parameters (say, flexural failure of tapered beams, beams subjected to high levels of shear-&-moment, presence of geometric imperfections and instability, etc.). The same is equally valid, if not more critical, for phenomena we do not have a working model (or, more appropriately, an/set of equation(s)) or those we lack domain knowledge of. Overcoming these critical limitations is one of the key motivations behind this work. Incorporating causality to address such challenges can be of merit.

A complimentary motivation stems from the fact that there is a very limited body of work that explores causality from a civil engineering perspective. From this lens, Fig. 1 presents the number of available studies that contain the terms [causality], [causal discovery], and [causal inference] in our domain over the last two decades. This figure was obtained through a scientometrics analysis from the open-source scholarly database, *Dimensions* (Dimensions 2021, Thelwall 2018). As one can see, this analysis returns 27 papers, therefore averaging 1.35 papers per year. The bulk of such works belongs to civil engineering sub-domains with a substantial human/social component, such as transportation and construction management (Beyzatlar *et al.* 2014, Gibb *et al.* 2014, Tong and Yu 2018). As one can see, the number of studies on causality from a civil engineering perspective is minute.

In our pursuit of advancing knowledge, we seek to identify, or possibly retrieve, the underlying data generating process (DGP)[4] responsible for the observations that result from the phenomenon we hope to understand (Huntington-Klein 2021). As such, we devise experiments. Such experiments are designed to test and explore hypotheses. A hypothesis targets a specific direction to uncover the DGP behind a given phenomenon (i.e., the cause(s) leading to the so-called effect) (Chambliss and Schutt 2013).

Thus, this paper hopes to showcase the importance of causality from a civil and structural engineering point of view. This paper reviews key concepts pertaining to causality, causal discovery, and causal inference. Then, this paper presents recent advancements in automated algorithms and causal machine learning. To reinforce the covered concepts, a series of civil engineering case studies are delivered. More specifically, two case studies pertaining to causal discovery to obtain the causal structure for a known DGP and an unknown DGP are presented. In addition, one case study aimed at causally inferring the merit of adopting a new structural system is also presented. Finally, this paper ends with a discussion on how commonly used machine learning can benefit from incorporating causal principles.

2. Causality

2.1 Correlation does not imply causation

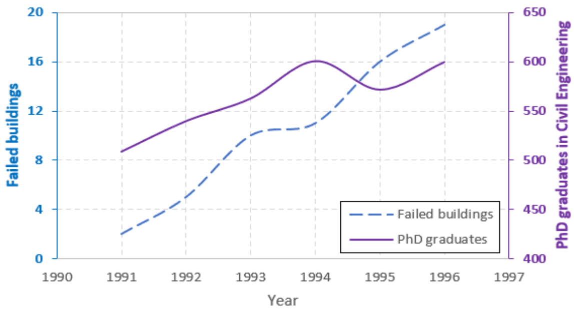
Causality is the science of cause and effect, specifically, causality seeks to identify how effects (or events) come to be (or are caused) by their causes (or triggers). More often, causality is tied to correlation and, in many cases, to spurious correlations. However, correlation alone does *not* imply causation[5].

A prime example of spurious correlation can be seen in Fig. 2(a) and Fig. 2(b), which depict a historical analysis of

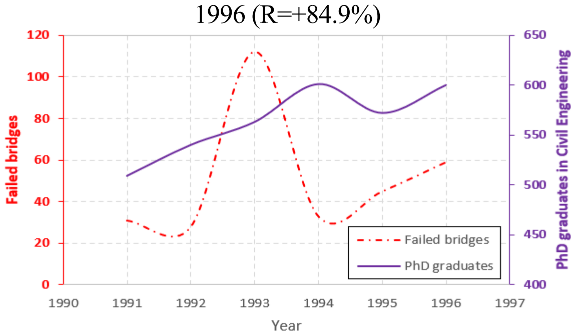
[3] A new question can also be formulated: 5) How could we have prevented failure in Beam A? *Ans. Failure can be prevented by switching Beam A with a larger section, and/or using a higher Grade of structural steel*. On a more fundamental and philosophical notion, *failure could have been prevented by reducing the level of bending moment, or possibly eliminating the existence of such bending moment.* The two former solutions stem from our engineering knowledge, while the latter stems from a causal perspective to our engineering understanding of the flexural failure problem.

[4] A DGP is often defined as a mechanism (or simply, a process) responsible for generating the data.

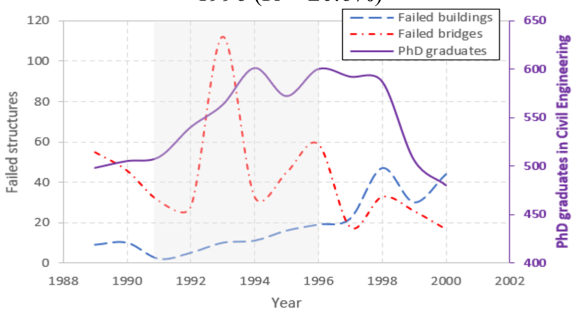
[5] However, causation may imply some form of correlation (Bunge 1979).



(a) Correlation between failed buildings in the US and the number of received PhDs in civil engineering within 1991-1996 (R=+84.9%)



(b) Correlation between failed bridges in the US and the number of received PhDs in civil engineering within 1991-1996 (R=+20.6%)

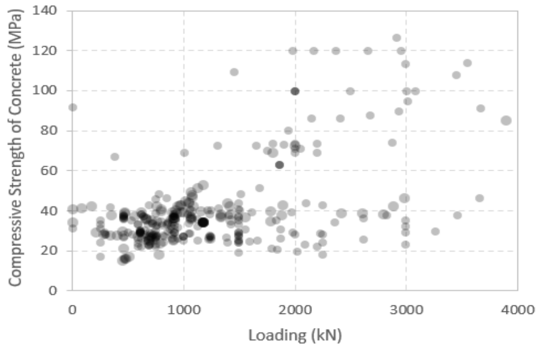


(c) Correlation between failed structures in the US and the number of received PhDs in civil engineering between 1989-2000 ($R_{buildings}$=+35.6%, $R_{brigdes}$=+3.87%)

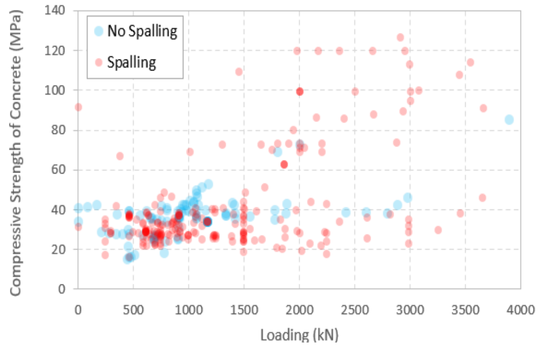
Fig. 2 Examples of correlations

the number of failed buildings and bridges in the US as reported in (Wardhana and Hadipriono 2003a, b) and compare that to the number of received PhDs in civil engineering (as per the National Science Foundation ("Surveys|NCSES|NSF" 2022)). Fig. 2(a) shows that there is a strong positive correlation (+84.9%) between the number of failed buildings in the US and the number of awarded PhDs in civil engineering between 1991-1996. Surprisingly, this plot may mistakenly imply that uptake in PhD graduates leads to a rise in structural failures. Within the same timeframe (see Fig. 2(b)), this correlation drops to a weak +20.6% when the number of awarded PhDs is contrasted against the number of failed bridges. This plot supports the notion that more PhD graduates in civil engineering do not strongly correlate with a rise in bridge failures. This interpretation is also not correct.

Now, plotting the number of failed buildings and bridges in the US against the number of received PhDs in civil



(a) Marginal dependence



(b) Conditional independence

Fig. 3 Marginal vs. conditional association between the compressive strength of concrete and loading in fire-exposed RC columns

engineering between 1989-2000 yields a weak correlation of +35.5% and +3.87%, respectively (see Fig. 2(c)).

The above is but one example of why correlations alone do not entail causation simply because statistical relations do not exclusively confine causal relations. In fact, the presented correlations herein, as well as those that could be derived from Fig. 2[6], represent *marginal* associations (i.e., two variables are independent while ignoring a third) as opposed to *conditional* associations (e.g., two variables are independent given a third) and hence are unlikely to have a definite causal effect. While this particular example demonstrates shocking claims (yet, supported by authentic data), let us look at a more practical example.

Fig. 3 reinforces the difference between marginal associations and conditional associations in a more related demonstration. This figure examines the relationship between the applied loads on fire-exposed reinforced concrete (RC) columns against the compressive strength of concrete in each corresponding column. Fig. 3(a) shows that the marginal dependence[7] between these two variables is moderate and positively linear, where columns subjected to larger loads are also made from high strength concrete (i.e., an increase in loading (X) co-occurs with an increase in

[6] For completion: substituting the outlier in 1993 in Fig. 2(b) by 40 failures, yields a strong 89.2% positive correlation when the number of awarded PhDs is contrasted against the number of failed bridges.

[7] Simply, marginal independence implies that knowledge of Y's value does not affect our belief in the value of X. Also, conditional independence implies that knowledge of Y's value does not affect our belief in the value of X, given a value of Z.

compressive strength (Y)). However, when these two variables are conditioned on a third variable (occurrence of fire-induced spalling, Z, see Fig. 3(b)), then one can clearly see that X is independent ($\perp\!\!\!\perp$) of Y given the occurrence of spalling (Z), as spalling seems to occur regardless of any marginal association between the applied loading and strength of concrete.

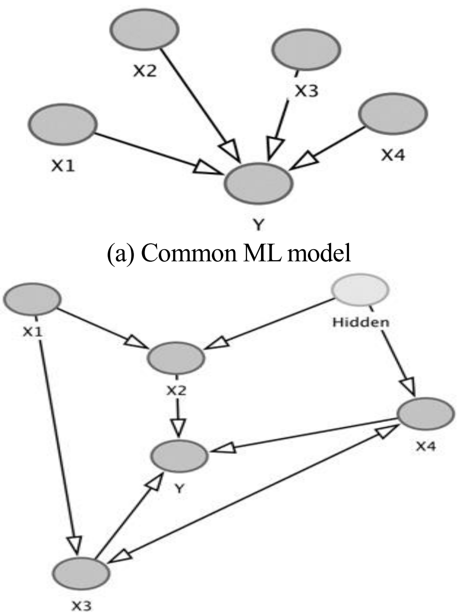
2.2 The causal ladder

This section outlines some of the key definitions of causality accepted in the open literature and then articulates key assumptions and algorithms often adopted to establish causation[8]. A good start to causality is to discuss how causality differs from traditional statistical and empirical methods, often favored by our domain. Our discussion will revolve around the causality rung system (or, more formally, the ladder of causation) pioneered by Pearl[9] (Pearl J 2018a). In this system, Pearl identifies three levels of causality.

The first rung builds upon the observations we *see* and hence stems from pure *statistical* associations. For example, an associational exercise would simply be the expectation that an uncoated metallic load bearing element (say, a plate girder) is bound to experience some level of corrosion after a few years of service. In this event, our observations of how metallic structures are likely to corrode come in handy to lay some form of association. Such an association can answer the following simple question, *what is the expected level of corrosion an uncoated steel plate girder could undergo after 10 years of service?*

Mathematically, this associational question resembles the conditional probability of the occurrence of corrosion (y), given that we observed an uncoated steel plate girder (x) - i.e., $P(y|x)$. An experienced engineer may be able to provide us with a good and possibly accurate estimation of the expected magnitude of such corrosion if we are to provide this engineer with information pertaining to the grade of steel, surrounding environmental conditions, etc. It is in this rung that *statistical* inference and empirical analysis reside, and we can build predictive models to predict such a phenomenon solely from the provided data[10].

The second rung of causality extends further than associations (the things we see) and includes *interventions* (the things we can *do*). Interventions are tasks that we *do* or *carry out* to explore questions such as what will happen to the system *if* we change one or more of its variables?



(a) Common ML model

(b) Acausal model

Fig. 4 An illustrative example of (a) A common ML model and (b) A causal model

Let us continue our plate girder example from an interventional perspective. To do so, we will devise an experiment where we intervene on this girder to see how a particular variable we identify (say, the addition of anti-corrosion coating) may or may not affect the corrosion of this girder. Mathematically, this interventional question resembles the use of the *do*-calculus $P(y|do(x))$, which can be translated as the probability of event y (occurrence of corrosion) given that we intervene and set the value of x (addition of coating) and subsequently observe the same event (10 years of installation)[11]. As one can see, an accurate and quantitative answer to this question cannot be answered from data alone - unless we have previous results from an earlier experiment or, more conveniently, an accompanying new experiment in which we test (i.e., intervene) on a new girder.

The same can also be said for machine learning (ML) enthusiasts. At this point in time, most supervised ML approaches are applied in a similar manner to that shown in Fig. 4(a) where a number of independent variables (X_1, X_2, X_3, …) are selected to predict an outcome (Y). In contrast, Y occurs via a DGP (an example is shown in Fig. 4(b)). Thus, in reality, we need to identify the DGP first to accurately predict Y. This crucial difference distinguishes between a predictive statistical model and an interventional model.

The third rung builds on the former two rungs to realize counterfactual reasoning of causality. This rung is dedicated to answering questions of counterfactuals (the things we can

[8] In all cases, we confine this discussion to aspects relating to structural engineering. For completion, a philosophical and historical discussion and a more in-depth review on causality can be found elsewhere (Klemme 2020, Pearl 2009a) and (Holland 1986), respectively. In addition, this paper tries to limit the amount of mathematical background behind causality to maintain a smooth flow, noting that dedicated works on the mathematical conditions and background required for establishing causality are readily available in notable textbooks (Pearl 2009a, Rubin 2005) and very recent review papers (Forney and Mueller n.d., Glymour *et al.* 2019, Spirtes and Zhang 2016). A discussion on such items (including *do*-calculus, conditional/interventional distributions, etc.), while worthy of our time and landscape of this study, is deemed too technical for an introductory work on causality in this domain.

[9] For a discussion on causality from the lens of the potential outcomes (PO), please refer to the following (Imbens 2020, Rubin 2005). For brevity, we will discuss the PO approach in a later section.

[10] Just as many machine learning model do.

[11] There is a subtle difference between $P(y \mid x)$ and $P(y|do(x))$ wherein the former describes which values Y would likely take on when X happened to be x (i.e., observational distribution), while the latter describes the values of y when X is set to x (i.e., interventional distribution).

imagine). For example, what would have happened to the expected level of corrosion of our plate girder after 15 years of exposure had this particular girder been fabricated from aluminum?

This question could be translated to $P(y_x|x', y')$. Namely, the probability that event $Y=y$ would be observed had X been x, given that we observed X to be x' and Y to be y'. This question is not empirically testable since we cannot test the *exact same girder* twice, especially if it is fabricated from two different materials and examined in two different exposures. Simply, counterfactuals aim to answer individual-unit questions (while interventions answer population-level questions (Dablander 2020)).

A logical solution to the above would be to run two experiments on identical girders (where one is fabricated from steel and the other from aluminum). Such an experiment will likely give us a realistic answer to our counterfactual question. Fundamentally and philosophically, this may solve our query even though we did not test the same girder under different conditions.

Yet, one can think of many examples were running an experiment may not be possible or ethical (e.g., what would have happened to the occupants of a collapsed building, X, had the earthquake been twice as harsh? etc.). This former example, among many others, emphasizes the need for a causal approach.

As one can see, while our domain incorporates all three rungs, we tend to heavily prioritize the first rung with a few ideas from the second (mechanically). Little has been paid to identifying the causal structure of DGP in our domain. While this author opts not to speculate on the above, two notions come to mind. It is possible that identifying DGPs: 1) is not only a complex endeavor but can be labeled as ambitious too, and 2) may not be warranted since methods spanning from the first two rungs could possibly provide us with practical resolutions for some of our engineering problems. This work hopes to present a case for our engineers on the value of seeking causality.

2.3 Regression and causality

Arriving at a given phenomenon's causal structure can be equivalent to a tool or an equation that represents how the governing factors interact to generate the phenomenon of interest. Unlike a traditional regression-based predictive tool/equation, a causal model goes beyond the realm of statistics and correlation to establish evidence for how causes lead to effects (vs. how correlation or association can be used to predict a phenomenon). According to Wasserman, "*prediction is about passive observation, and causation is about active intervention*" (Wasserman 2021).

In addition, regression is used to estimate the *influence* of dependent variable(s), X_1, X_2.., etc., on an independent variable, Y. By defining one variable to be an X and another as a Y, then the direction of the relationship is explicitly stated, and may imply a reversal. Fundamentally, these directions should then be defensible in theory[12].

In the event that such a relationship is linear, then a unit change in Y can be represented via a coefficient that is tied to a corresponding X (while keeping all other variables constant). Unfortunately, such an estimate is bound to be flawed if there is some unmeasured common cause between the X's themselves and/or with Y.

A solution to remedy the above aims to incorporate additional dependent variables. However, such a solution has been shown to cause instability (since more variables may also have other/common causes or may, in fact, be an effect as opposed to a cause of Y) (Glymour *et al.* 1994). This is one of the key differences between regression and causation.

Another key difference pertains to the exclusive nature of the time hierarchy. From this lens, some variables may naturally occur before, together, or after others. For example, the application of loading may cause beams to crack in the short run. Poorly maintained beams may also suffer from creep and settlement issues which also add to cracking in the long run. Incorporating this natural process is elemental to maintaining the true essence of the phenomenon on hand.

A reminder to the reader is that predictive regression hopes to project the values of the dependent variable given its independent variables, while causality starts by discovering/inferring the relationship between any two individual variables.

2.4 Definitions and big ideas

Hume (Klemme 2020) defines causality from two perspectives, a philosophical relation, and a natural relation. In the former, a cause is "*an object precedent and contiguous to another, and where all objects resembling the former are placed in like relations of precedency and contiguity to those objects that resemble the latter*". On the other hand, the latter states that a cause is *"... an object precedent and contiguous to another, and so united with it that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other*". A simpler definition that is also attributed to Hume is also provided herein, "*We may define a cause to be an object, followed by another, [. . .] where, if the first object had not been, the second had never existed*". The second part of this definition also describes a key companion concept to *causality*, which is *counterfactual*.

In terms of counterfactuals, Lewis (Lewis 1973) defines causality as, "*We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects - some of them, at least, and usually all -*

[12] For example, say that an engineer notices that heavily loaded columns (Y) are often cast with high strength materials (X). Then, this engineer regresses the loading on material strength. The resulting equation can probably estimate the expected magnitude of loading on a column of some material grade with good accuracy. However, this same equation does not imply that a reversal is true (i.e., a column cast from a high strength material may, or may not, be subjected to a high levels of loading). The above is especially true if such a relationship was arrived at from a statistical analysis on a sample size that contains a certain range for Y and X. Now, say that the sample size is large enough to cover all ranges. The same reversal may not still hold true unless we ensure that there are no confounders, no measurement errors, or any form of bias, etc.

would have been absent as well." For its similarity to Hume, we confine our discussion in this paper to Lewis' definition.
From a structural engineering perspective, one may say the following, a given Beam A would deflect upon loading - implying that the application of loading triggers (or is the cause of) the deflection of Beam A. Similarly, one may then infer the counterfactual that if it was not for the application of loading, Beam A would not have deflected. While these two examples may be perceived as logical, they as equally fundamental as they showcase the importance of causality in our domain.

How can we establish causality? To establish causality, we need to leverage causal models. Such models aim to describe (mathematically or otherwise) the data generating mechanism responsible for producing the observations we see. Before we dive into such models, let us start our discussion by demonstrating the broader concepts of causal discovery and causal inference necessary to realize causal models.

Causal discovery (also referred to as causal structure search) seeks to discover the underlying structure of causal relations between variables pertaining to a DGP by analyzing observational data. Let us revisit Beam A. Say that this beam is simply supported, made from one homogenous material, has a depth, *D*, a width, *W*, a span, *L*, and deflects by a magnitude of *X* upon the application of loading, *P*. Suppose we carry out experiments where we intervene on *D*, *W*, *L*, and *P*, to observe how the deflection changes in response to our interventions, then we will be able to collect a series of observations pertaining specifically to each of our interventions. These observations can help us pinpoint the DGP responsible for the deflection of Beam A - so that we can: 1) Predict the deflection of this beam without having to carry out costly/lengthy experiments, or 2) Identify a suitable combination of interventions to limit the deformation of Beam A to a predefined limit.

A *causal structure* can be represented mathematically by causal models. A causal model spells out the probabilistic (in)dependence of variables and the effects of interventions (actual or hypothetical changes on one/some/all variables) (Nogueira *et al.* 2022). Such a model can be further presented by a set of equations (namely, structural equation models (SEMs)) and/or accompanying graphs (i.e., Directed Acyclic Graphs[13] (DAGs)). An SEM represents assumed causal relationships between the involved parameters in an <u>equational</u> format that may entail a parametric or non-parametric form, while a DAG <u>visually</u> represents the causal structure of the DGP.

One should keep in mind that a DAG contains edged arrows that flow in one direction (e.g., $A \rightarrow B$, meaning *A* causes *B*, etc.) and does not allow for a circular flow of information. Fig. 5 shows an example of a DAG and a corresponding SEM. In this figure, *A* is the direct cause of *B*, *C* and *D*. *B* is the direct cause of *C*. $A \rightarrow B \rightarrow C$ is called a path. In addition, A also serves as an indirect cause of *C*,

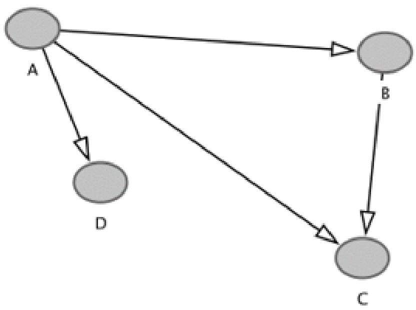


Fig. 5 Illustration of a DAG [corresponding SEMs are, D: $=f_1$(A), B: $=f_2$(A), C: $=f_3$(A, B). Note that error terms are assumed independent and are not explicitly shown in the figure. Also, :=implies a causal statement]

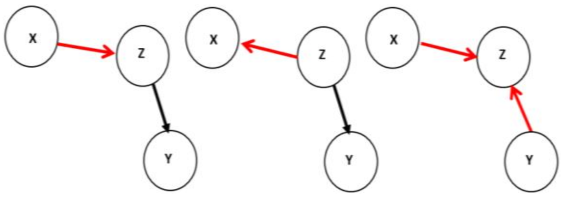


Fig. 6 Illustration of DAGs for mediators, confounders, and colliders (from left to right) [Note: the different arrowheads and colors are adopted for illustration purposes]

given its role on *B*. Further, a node, say *B*, is a descendent of *A* (and so on). DAGs are interpreted causally at the interventional and counterfactual levels, while associational levels are used to describe conditional independencies between variables. There is more to SEM and DAGs than the allowable page limit to this work. For additional sources, the reader is encouraged to review the following (Bollen and Pearl 2013, Forney and Mueller n.d.).

As one can see from Fig. 5, there is a number of ways two parameters (or nodes) can be tied together. Other than direct causes and indirect causes, three main relationships are of interest to causal analysis: mediators, moderators, and colliders (see Fig. 6). To parallel cited works, this study will remain true to the terminology used in such works, hence, *X* denotes a <u>cause</u>, *Y* is an <u>outcome</u> (target), and Z is a <u>third variable</u> that could be a mediator, moderator, or a collider.

A mediator (or a chain) is a pattern in the form of $X \rightarrow Z \rightarrow Y$. This format implies that *X* causally affects *Y* *through Z*[14]. For example, fire (*X*) generates elevated temperatures (*Z*), which in turn degrade the sectional capacity of a load bearing member (*Y*).

A companion to mediators is moderators, often defined as variables that change the *size* (or *direction*) of the relationship between variables. For instance, the magnitude of deformation (*Y*) in a given beam can be amplified during fire conditions. In this instance, temperature rise (*Z*) can lead a beam to undergo larger deformations under the same level of loading (*X*). Such larger deformations may not be observed under ambient conditions from *X*. Thus, temperature rise (*Z*) <u>amplifies</u> the deformation (*Y*).

Reversing the arrow that extends from *Z* to *X* changes

[13] A graph is a is a mathematical object that consists of nodes and edges. A DAG is a graph with directed edges. An example is shown in Fig. 5.

[14] For example, isolating $A \rightarrow B \rightarrow C$ from Fig. 5 into a new DAG can be considered a chain.

this pattern into a confounder (common cause or a fork). Here, $X \leftarrow Z \rightarrow Y$, Z causally affects X and Y, effectively labeling Z as a confounder and inducing a *non-causal association* between X and Y. This formation can be tricky to investigate since a confounder can be a variable we do not observe, do not often consider, do not know it exists, or have data on.

A hypothetical example would be the following: seismically active regions (Z) witness limited number of structural failures (X) despite having a large number of structures (Y) than other regions, i.e., low structural failures (X)←seismically active regions (Z)→high number of structures (Y). In this instance, the confounder Z makes X and Y statistically correlated despite the lack of a direct causal link between X and Y[15].

Finally, a collider turns up when the arrows of the moderator flip towards Z. Now, Z is no longer a cause but rather a common effect of X and Y. For example, harsh environmental conditions (X) and poor maintenance (Y) can lead to structural issues (Z) such as cracking and corrosion, thus, harsh environmental conditions (X)→structural issues (Z)←poor maintenance (Y). If we are to condition on Z[16], then the association between X and Y in the mediator and confounder blocks the flow of association while doing the same on the collider induces non-causal association.

2.5 Assumptions needed to establish causality

There are several core assumptions tied to causal discovery and causal inference. These assumptions form the basis for establishing causality - especially from observational data. As such, these are described herein, and rich examples and corresponding mathematical details can be found elsewhere (Huntington-Klein 2021, Nogueira *et al.* 2022, Pearl 2013, Scheines n.d.).

We start with the *causal Markov assumption*. This assumption states that a variable, X, is independent of each other variable (except X's effects) conditional on its direct causes. For example, X in the left DAG in Fig. 6 is independent of Y, given (or conditional on) Z. A companion to the Markovian assumption is the *d-separation* criterion (Pearl 2009a). This criterion establishes whether X is independent of Y, given Z (i.e., $X \perp\!\!\!\perp Y|Z$), by associating the notion of independence with the separation of variables in a causal graph. For example, a path is d-separated if it contains 1) a chain or a fork such that the middle node is not in Z, or 2) a collider such that the middle node is not in Z, nor its descendants are in Z.

Causal faithfulness implies that any population produced by a causal graph has the independence relations obtained by applying d-separation to it. Embracing this assumption eliminates all cases of unfaithfulness (i.e., independences that are not a consequence of the causal Markov condition or d-separation) from consideration. According to Heinze-Deml *et al.* (2018), the causal Markov and faithfulness assumptions suggest that d-separation relationships in a causal DAG have a one-to-one correspondence with conditional independencies in the distribution of that particular graph. An example of unfaithfulness is one where X is a cause of Y and Y is a cause of Z, but X and Z are independent of each other (Allen 2020).

Finally[17], we cover the c*ausal sufficiency* assumption. This assumption refers to the absence of hidden or latent parameters that we do not know nor are aware of, i.e., our data contains measurements on all of the common causes of the selected variables responsible for a DGP (Scheines n.d.)[18].

2.6 Graphical methods

Now that the fundamentals of causality are presented, it is time to move into the graphical methods that can be used to arrive at causal models and causal graphs[19].

Causal models have a causal structure and hence arriving at such models requires us to obtain the underlying structure of the problem on hand. While we may not know such a structure, we are likely to have data that we can exploit to arrive at a causal structure (which may turn out to be the ground truth or one that we are comfortable with labeling as a causal structure pending a series of assumptions arising from domain knowledge).

Fig. 7 shows illustrations of different granularities to causal search methods. For example, a *directed graph (DG)* is a graph that allows loop feedback, and a *partially directed graph (PDG)* may contain both directed and undirected edges. As mentioned earlier, a DAG assumes acyclicity. In some instances, a causal DAG may not be easily identifiable, but rather the search method returns a set of equivalent DAGs. Such graphs are referred to as *Markov*

[15] An explanation would be that a seismically active region might home a metropolitan. Such a metropolitan would house a large population which requires the development of many structures. In addition, structures of seismically active regions are often required to comply with codal provisions that enforce additional detailing and requirement to mitigate damage.

[16] Building on the above discussion and referring to Fig. 6 while doing so may raise the question of how to distinguish a confounder from a collider (from the data or when domain knowledge is limited). This task can be completed via the *back-door* and *front-door adjustments* (Pearl J 2018b). In the confounding formation, $X \leftarrow Z \rightarrow Y$, the back-door adjustment states that 1) no node in Z is a descendent of X, and 2) any path between X and Y that begins with an arrow into X (known as a back-door path) is blocked by Z, then controlling for Z blocks all non-causal paths between X and Y. In the event that the confounding factor is *unobservable* or *hypothetical*, then the front-door adjustment comes to the rescue. In this process, we add a new parameter that we may assume cannot be caused directly by the confounding factor. Now, we can use the aforementioned back-door adjustment to estimate the effect of the new parameter on the outcome. For an exhaustive description of the above two adjustments, please refer to the following (Correa and Bareinboim 2017, Glynn and Kashin 2018).

[17] Other assumptions also exist such as Gaussianity of the noise distribution, one or several experimental settings, and linearity/nonlinearity, acyclicity - see (Heinze-Deml *et al.* 2018).

[18] In hindsight, assigning a series of regressors, X_1, X_2, X_3, to predict Y does not imply that these regressors are causes of Y. In the event that a thorough domain knowledge exists on such regressors and that these regressors satisfy causal principles, then a regression analysis can indeed turn causal. The latter is, unfortunately, rarely checked in our domain.

[19] A thorough review on graphical model can be found in (Allen 2020). In addition, Reviewer no. 2 has pointed out the merit in describing other types of methods such as n-tuple relations in the First order Logic. The reader is encouraged to visit the following references for an in-depth discussion on such methods (Džeroski 2009, Kovalerchuk and Vityaev 2000, Mitchell 1997, Muggleton 1991).

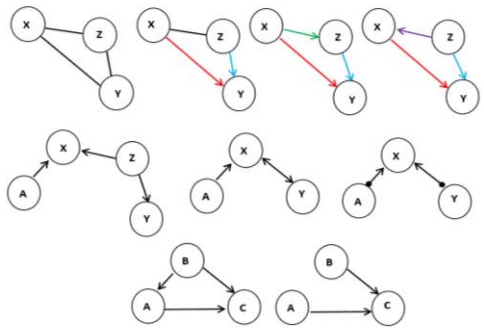


Fig. 7 Graphs [Note from the top row and going right: graph (skeleton), CPDAG, DAG Markov equivalent 1, DAG Markov equivalent 2. Middle row left and going right: Assumed true DAG, MAG, and PAG. Bottom row: DAG, DAG with an intervention on A (i.e., do(A=a)]

equivalence (i.e., encode the same set of d-separation relationships). For example, $X \rightarrow Z \rightarrow Y$, $X \leftarrow Z \rightarrow Y$, and $X \leftarrow Z \leftarrow Y$ contain the same conditional independence (i.e., $X \perp\!\!\!\perp Y | Z$).

Completed Partially Directed Acyclic Graphs (CPDAGs) can be used to represent the above set of equivalent graphs. In a CPDAG, an edge is only directed if there is only one graph in the Markov equivalence class with an edge in that direction, and if there is uncertainty about the direction, this particular edge is left undirected.

In the event that the search method identifies the presence of unobserved variables or confounders, then such a method may return an *Acyclic Directed Mixed Graph (ADMG)*. This graph presents the unobserved variable with bidirected edges $X \leftrightarrow Y$, meaning there exists a possible confounder between X and Y.

Maximal Ancestral Graph (MAG) have the capability to represent the same features of ADMGs as well as the presence of selection bias (preferential omission of data points from the samples (Bareinboim *et al.* 2014)). *Partial Ancestral Graphs (PAGs)* are similar to CDPAGs as they also represent the set of equivalent MAGs (see Fig. 7 for an illustration of the aforenoted graphs).

The above graphs can home a number of relationships. In principle, a causal search algorithm will attempt to tie the variables involved via such relationships, as discussed below.

$X \rightarrow Y$ where X is a cause of Y

X - Y where X is a cause of Y or vice versa

$X \leftrightarrow Y$ where there is an unmeasured confounder

$X \circ\!\!\rightarrow Y$ either X is a cause of Y, or there is a confounder

$X \circ\!\!-\!\!\circ Y$ either X is a cause of Y, and/or vice versa, and/or there exists a confounder

2.7 Causal search methods and causal machine learning packages

There are two primary search approaches to structure discovery[20]. These are *constraint-based* and *score-based*. This section covers key algorithms often used in each front, and other algorithms can be found elsewhere (Heinze-Deml *et al.* 2018, Nogueira *et al.* 2021, Vowels *et al.* 2021).

Constraint-based approaches arrive at causal structures by testing for constraints or conditional independencies as a means to construct a graph that reflects and satisfies such constraints. In other words, these approaches seek to only find the graphs that correctly represent the independence relationships through hypothesis testing (Nogueira *et al.* 2022). Additional rules are then applied to determine the relationship between each variable. As noted in the previous section, it is quite possible to find multiple graphs that fulfill a given set of conditional independencies. Therefore, constraint-based approaches are likely to output a graph representing Markov equivalence (CDPAGs or PAGs). Some of the commonly used constraint-based algorithms include *Peter-Clark (PC) algorithm* (Spirtes *et al.* 2000) and its variants, *Fast Causal Inference (FCI*) and *Inductive Causation (IC)* (Yu *et al.* 2016).

On the other hand, *score-based* approaches identify graphs by assigning a relevance score (say, Bayesian Information Criterion) to candidate graphs. Since these approaches are expected to score every candidate graph, then they turn computationally expensive, and hence such approaches are often accompanied by greedy heuristics. Examples may include *Greedy Equivalence Search (GES)*, Fast Greedy Equivalence Search (FGES), *Greedy interventional equivalence search (GIES)*, etc.

While constraint-based approaches have been noted to be more efficient as they output one graph with clear semantics, they lack an indication of relative confidence in the output graph. On the other hand, score-based counterparts provide a series of equivalent graphs with an added confidence metric(s). Hence, researchers moved to explore hybrid approaches to maximize the output of causal searches (Triantafillou and Tsamardinos 2016). Such a hybrid approach may include *structural agnostic modeling (SAM)* and *causal additive models (CAM)* (Spirtes and Zhang 2016). Table 1 summarizes the above discussion and refers to algorithmic and machine learning (ML) packages for various causal discovery tools.

In a similar manner to commonly used ML algorithms, the outcome of a causal analysis can be evaluated via metrics. A number of metrics are available and include those traditionally used in classification problems (M and M.N 2015, Naser *et al.* 2021) - especially if the ground truth mechanism (or one deduced from domain knowledge) is available. When the ground truth is not available, the performance of the causal discovery algorithm is evaluated on how well the causal structure predicts the phenomenon at hand.

2.8 Causal inference and causal machine learning packages

Causal inference seeks to study the possible effects of altering a given causal system. Simply put, causal inference

[20] Vowels *et al.* (Vowels *et al.* 2021) and Heinze-Deml (Heinze-Deml *et al.* 2018) identify additional approaches such as causal association rules, causal forest, and causal networks. These were not presented here for simplicity. In addition, approaches pertaining to time series data, mixed data, or Bayesian methods were also left out.

Table 1 ML packages for causal discovery (full description of each package and incorporate algorithms can be found in its original source)

	An R package for causal discovery and inference that includes:	Ref.
Pcalg	*Constraint-based learning algorithms*: • PC, FCI, Really fast causal inference (RFCI), Greedy interventional equivalence search (GIES) *Score-based learning algorithms*: • GES, Greedy interventional equivalence search (GIES) *Hybrid learning algorithms*: • Adaptively restricted GES (ARGES)	(pcalg 2022)
bnlearn	A Python package for causal discovery, estimating causal parameters, and performing inference. This package includes: *Constraint-based structure learning algorithms*: • PC (the *stable* version), Grow-Shrink (GS), Incremental Association Markov Blanket (IAMB), Fast Incremental Association (Fast-IAMB) *Score-based structure learning algorithms*: • Hill Climbing (HC), Tabu Search (Tabu), *Hybrid structure learning algorithms*: • Max-Min Hill Climbing (MMHC), Hybrid HPC (H2PC), General 2-Phase Restricted Maximization (RSMAX2) *Local discovery algorithms*: • Chow-Liu, ARACNE *Bayesian network classifiers*: • Naive Bayes, Tree-Augmented naive Bayes (TAN)	(Bnlearn 2020)
Tetrad *Py-causal* *R-causal*	Provides access to desktop Software (*Tetrad*), as well as Python (*Py-causal*) and R (*R-causal*) packages created by the Center for Causal Discovery in Pittsburgh University. Tetrad includes a series of causal algorithms, including PC, FCI, and GES.	(Center for Causal Discovery 2022)
Causal discovery toolbox	A Python and R packages for causal discovery that are mainly based on observational data. These packages include: • ANM (Additive noise model for pairwise causality). • GNN (Generative Neural networks). • IGCI (Information Geometric Causal Inference). • Jarfo (an ensemble method to build causally relevant features).	(Kalainathan and Goudet 2022)
TIGRAMITE	A Python package for causal discovery for time series datasets. This package includes: • *PCMCI* (a causal discovery framework for large-scale time series datasets). • *PCMCIplus* (a companion to PCMCI, and can identify the full, lagged, and contemporaneous, causal graph under the standard assumptions of Causal Sufficiency, Faithfulness, and the Markov condition). • *LPCMCI* (for causal discovery applicable for large-scale times series and allows for latent confounders/lag-specific causal relationships).	(TIGRAMITE 2022)

Comparison of packages

Package	*Pcalg*	*bnlearn*	*Tetrad* *Py-causal* *R-causal*	*Causal discovery toolbox*	*TIGRAMITE*
Programming language	R	Python	Java, Python, and R	Python and R	Python
Type of causal search	Constraint-based & Score-based	Constraint-based	Constraint-based & Score-based	Constraint-based & Score-based	Constraint-based
Type of data*	C/N/M	C/N/M	C/N/M	Depends on the selected algorithm	C/N & timeseries

*C: Categorical data. N: Continuous data. M: Mixed data.

infers how interventions, treatments, and manipulations systematically alter the observations. To realize such inference, a *causal structure* that describes the phenomenon is expected to be available. Going back to our example of a plate girder. Say that we were able to discover one possible causal structure for the DGP responsible for the corrosion of

Table 2 Metrics used for causal discovery analysis (Heinze-Deml *et al.* 2018; Nogueira *et al.* 2022)

Metric	Description
Missing or extra edges	Number of edges that are missing or present in the original model but not in the generated one.
Incorrect adjacencies	Number of undirected edges that are present in the generated model but not in the original one.
Incorrect and correct directed edges	Number directed edges that are missing present in the generated model that were correctly directed.
Structural hamming distance	Sum of missing edges, extra edges, and incorrectly directed edges.
Adjacency precision	Correctly predicted adjacencies/predicted adjacencies.
Adjacency recall	Correctly predicted adjacencies/true adjacencies.
Arrowhead precision	Correctly predicted arrowheads/predicted arrowheads.
Arrowhead recall	Correctly predicted arrowheads/true arrowheads.
Area-above-curve	AOC=1−AUC, where AUC (area-under-curve) measures the area below the graph of false positive rate and true positive rate.

Table 3 ML packages for causal inference (full description of each package and incorporate algorithms can be found in its original source)

	Packages for causal discovery	Ref.
DoWhy	A Python package for causal reasoning that provides a systematic interface for causal discovery and inference for modeling, estimating, and refuting causal assumptions. DoWhy combines graphical methods and potential outcome methods.	(Sharma and Kiciman 2019)
EconML	A Python package that can be used to estimate individualized causal responses from observational or experimental data. This package incorporates a series of interpretable methods as well.	(EconML 2022)
CausalGAM	An R package that implements estimators for average treatment effects and a standard regression estimator through generalized additive models.	(CausalGAM 2020)
FLAME	An R package for causal inference.	(FLAME 2022
CausalNex	A Python package for causal reasoning (discovery and inference).	(Beaumont 2021)
Causal ML	A Python package with a standard interface for causal inference and one of the few packages to estimate ITE.	(Uber Technologies 2020)

Comparison of packages

Package	*DoWhy*	*EconML*	*CausalGAM*	*FLAME*	*CausalNex*	*Causal ML*
Programming language	Python	Python	R	R	Python	Python
Type of estimands	ATE, ATT, CATE	ATE, ATT, CATE	ATE, ATT	ATE, CATE	Bayesian Networks	ATE, ATT, CATE, ITE
Type of data*	C/N	C/N	-	C	Primarily N	C/N

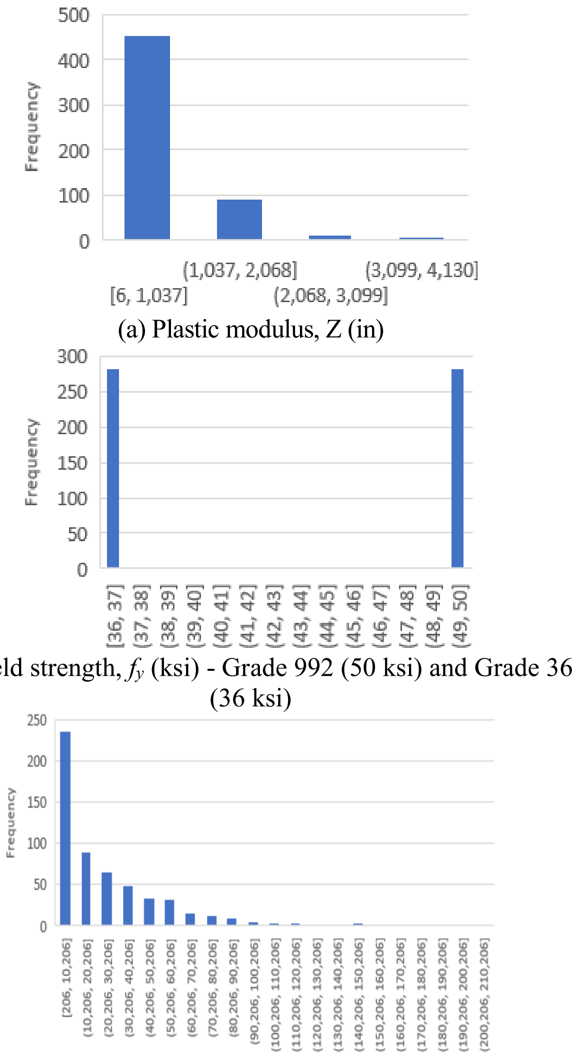
this plate girder. Then, using this discovered causal structure, we can answer questions about how altering the properties of this plate girder would influence its corrosion.

We can also causally infer about a particular phenomenon that we do not know its causal structure. In this instance, the causal inference procedure would require the availability of data. Such data can be examined via a series of causal assumptions to realize their corresponding causal effects. These assumptions and rules are discussed herein[21].

In the Potential Outcome (PO) approach, as pioneered by Rubin (Rubin 2005), causal effects can be estimated by comparing the *potential outcomes* of a given intervention/treatment. In reality, only one outcome can be observed since we cannot intervene on a unit and then un-intervene on the same unit at the same time. As such, causal inference is effectively a missing data problem, for which we can only observe one/some but not all of the outcomes. The potential outcomes are noted with Y. Y_0 notes the potential outcome for the untreated/benchmarked unit(s), and Y_1 notes the potential outcome for the treated/intervened upon unit(s). It is common to denote the intervention or treatment with a T.

Say, we are interested in inferring the benefits on flexural capacity obtained by casting beams using ultra-high performance concrete (UHPC) instead of traditional concrete, i.e., normal strength concrete (NSC). One way to carry out this investigation is by casting two identical groups of beams, Group A made from NSC (Y_0) and Group B made from UHPC (Y_1). Each group will home three

[21] For simplicity and to provide our engineers with a wide perspective on causality, I will be adopting the Potential Outcome (PO) approach, instead of the Pearlian approach favored in causal discovery, here. Please note that both approaches present equivalent concepts (Imbens 2020). A key concept in PO is the notion of randomization - often satisfied when setting upon experiments in structural engineering (i.e., we randomly assign intervention to load bearing specimens in the accompanying examples above).



(a) Plastic modulus, Z (in)

(b) Yield strength, f_y (ksi) - Grade 992 (50 ksi) and Grade 36 (36 ksi)

(c) Moment capacity, M (k.in)

Fig. 8 Frequency of identified variables in the compiled database

identical beams (with the exception of concrete type) such that Group A [$A_{0,1}$, $A_{0,2}$, $A_{0,3}$], and Group B [$B_{0,1}$, $B_{0,2}$, $B_{0,3}$]. Then, we test all beams under bending and report our findings. To realize an answer (or a series of answers) to our question, we will need to causally infer such answer(s).

The simplest approach is to average the results of the flexural capacities of Group A and Group B and compare these averages (or expectations ($\mathbb{E}$)). Mathematically, this results in the following

$$\text{The difference in average (expectation) between Group A and Group B}=\mathbb{E}[Y1-Y0] \tag{2}$$

Eq. (2) mirrors that which we could have arrived at using the Pearlian approach from intervention as noted by the do-operator (see Eq. (3))[22].

$$\mathbb{E}[Y|do(T=1)]-\mathbb{E}[Y|do(T=0)] \tag{3}$$

Both Eqs. (2) and (3) estimate the Average Treatment Effect (ATE).

Additional causal effects also exist, such as the conditional average treatment effect (CATE)[23] and the Average Treatment Effect on Treated (ATT) - What is the expected causal effect of the treatment for individuals/units in the treatment group?, as well as the Individual Treatment Effect of unit (ITE) - What is the expected causal effect of the treatment on the outcome of a specific unit? etc.

To continue our example in order to estimate ATT, we would want to visit Group B. Then, ATT would be the $\mathbb{E}[Y_1-Y_0|T=1]$. It is obvious that ITE may not be easily nor possibly estimated from this example since we can only test each exact beam under one consideration. On a more positive note, the difference in expectation between each identical beam can give a close estimate of ITE.

The PO approach extends further than the limited space of this work to cover a wide range of applications and conditions (i.e., propensity scoring, confoundness, missing data, matching, etc.), which can be reviewed herein (Imbens and Rubin 2015). There are a number of ML packages that can be used in causal inference investigations. These are listed in Table 3.

3. Case studies

This section describes three case studies to be highlighted in this work. The first two revolve around causal discovery (via constraint-based and score-based algorithms)[24] as carried out via the *Pycausal* package, and the third covers causal inference.

3.1 Causal discovery

3.1.1 Case no. 1: Obtaining the causal structure for a known DGP

The first case study hopes to verify our domain knowledge of a known and simple DGP. The DGP of interest pertains to the moment capacity, M, of W-shaped steel beams, which was discussed in Eq. (1). As mentioned earlier, the moment capacity of such beams can be estimated by multiplying the plastic modulus, Z, with the yield strength of steel, f_y, $Z \rightarrow M \leftarrow f_y$, thus effectively labeling M as a collider.

Let us explore two algorithms (namely, the PC algorithm, a *constraint-based,* and the FGES algorithm, a *score-based*). To execute this analysis, a database comprising all W-shaped structural steel sections adopted from the American Institute of Steel Construction (AISC) design manual (as provided in (AISC 2022)) was used. In this database, the plastic modulus of each steel section was multiplied by two steel Grades (A992 and A36). Therefore,

[22] Note that $p(Y=y|\mathrm{do}(X=x)) = p_m(Y=y|X=x) = \sum_z (Y=y|X=x)p(Z=z)$, where p_m is the manipulated distribution arising from our intervention. A review on do-calculus can be found in (Pearl 2009b), and a sample solved example is shown at a later section.

[23] Simply, an ATE conditioned on a subset of the population.

[24] The used algorithms are the PC and FGES in their *default* settings (given the illustrative nature of this study). For more information on these algorithms, please refer to their original sources (Heinze-Deml *et al.* 2018, Ramsey *et al.* 2017).

Table 4 Statistics on collected database

	Z (in)	f_y (ksi)	M (k.in)
Minimum	5.73	36.00	206.28
Maximum	4130.00	50.00	206500.00
Average	594.24	43.00	25552.46
Std. dev.	687.79	7.01	30252.23
Skewness	2.16	0.00	2.31

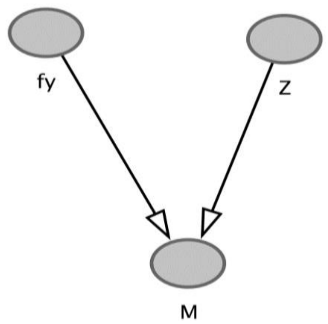


Fig. 9 Result from case study no. 1 (successfully resembling a collider)

Table 5 Statistics on collected database

	b (mm)	r (%)	f (MPa)	K	C (mm)	P (kN)
Minimum	152.0	0.3	15.0	0.5	13.0	0.0
Maximum	514.0	11.7	126.5	1.0	64.0	5373.0
Average	325.0	2.3	42.0	0.5	33.4	1335.2
Standard deviation	72.6	1.5	24.2	0.1	7.7	999.1
Skewness	0.7	2.5	1.9	3.6	-0.4	1.6

this database contains three variables, two independent (Z and f_y) and one dependent (M) - see Fig 8 and Table 4.

The compiled database and two above algorithms was used in this analysis, and the results of this analysis are shown in Fig. 9. As one can see, the generated DAG resembles that of our domain knowledge and is identical in both algorithms. This is both expected and comforting. As mentioned, only score-based algorithms are fitted with the capability to fit score-of-fitness to the developed DAG. This score was -19236.4 based on the default Bayesian Information Criterion (BIC)[25].

3.1.2 Case no. 2: Obtaining the causal structure for unknown DGP

In this case study, we are interested in uncovering the unknown DGP of how RC columns spall under fire conditions. In this pursuit, and to demonstrate some critical differences between causal discovery analysis and a typical ML analysis, a particular database of fire-exposed RC columns, which has been explored via a common and explainable ML algorithm in an earlier work of the author (Naser and Kodur 2022), was incorporated herein.

[25] BIC score is calculated as *chi-square statistics - degrees of freedom×log(sample size)*. Low values of BIC imply the goodness of the model.

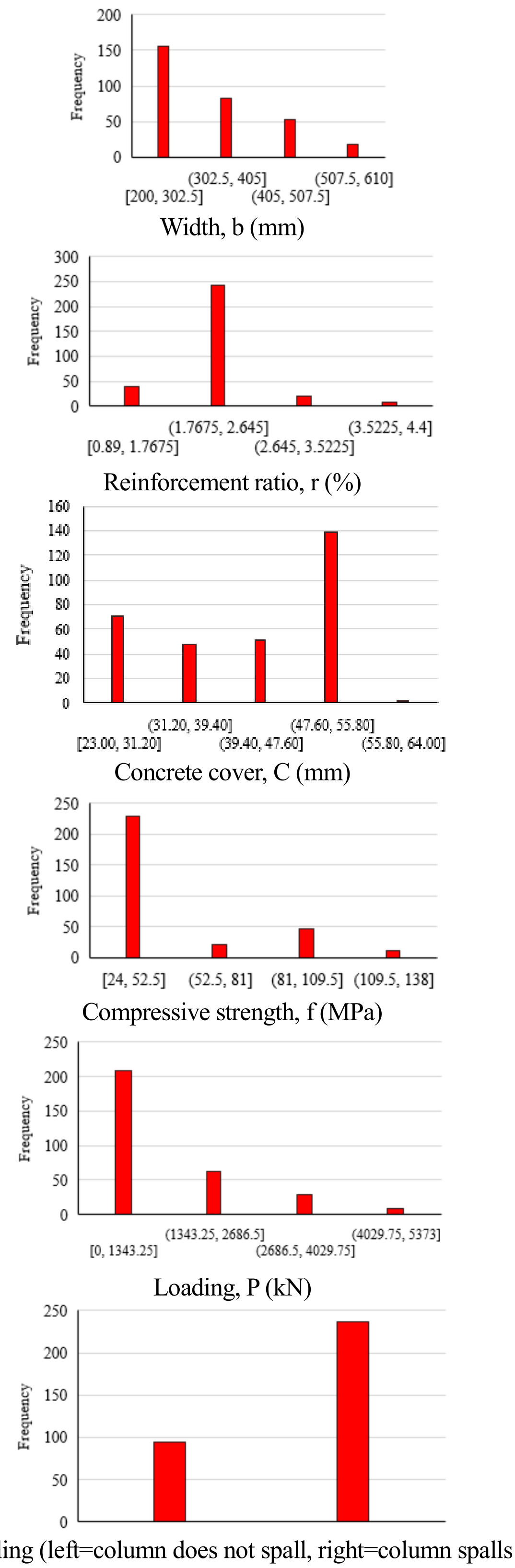
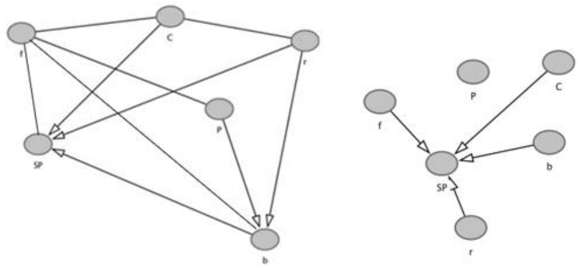
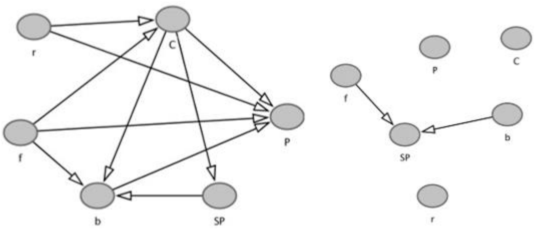


Fig. 10 Frequency of identified features of selected RC columns in the compiled database

The compiled database includes the following variables: column width and depth, b, steel reinforcement ratio, r,



(a) The outcome from the PC algorithm



(b)The outcome from the FGES algorithm (BIC = 0.0)

Fig. 11 Comparison between DGP [Note: left without domain knowledge, right: with domain knowledge]

concrete compressive strength, *f*, concrete cover to reinforcement, *C*, magnitude of applied loading, *P*, and Spalling, *SP* (Yes, or No). A visual illustration of the aforenoted variables via histograms, along with additional descriptive statistical details, are shown in Table 5 and Fig. 10. It is worth noting that the former analysis conducted via an explainable ensemble has identified the following five variables as the most important in terms of their influence on fire-induced spalling of RC columns:

C (100%), *f* (95%), *b* (88%), *r* (66%), and *P* (44%). It is equally worth noting that our domain knowledge dictates that none of the dependent variables can cause each other (Hertz 2003, Khoury 2000, Kodur 2000, Sanjayan and Stocks 1993).

Similar to the first case study, the constraint-based PC algorithm and the score-based FGES algorithm were applied to identify the causal structure of the DGP. The outcome of this analysis is shown in Fig. 11. This figure shows two possible structures per algorithm. The first (left) is obtained by using all the variables (i.e., each algorithm will have to determine the relationships between variables as well as which variable is the outcome of interest), and the second (right) is obtained by identifying the spalling (SP) as the outcome and that none of the dependent variables can cause each other.

Fig. 11 (left) shows that the PC algorithm was not able to create a DAG, while the FGES algorithm managed to do so. A look into these figures also shows that the PC algorithm identified *C*, *r*, *f*, and *b* to have possible relations. Surprisingly, this algorithm does not relate *P* to spalling. On the other hand, the FGES only notes one direct causal relation between *C* to spalling. It is clear that both of these figures do not provide us with sufficient information to uncover a likely DGP.

Moving to Fig. 11 (right), this figure shows that both PC and FGES algorithms manage to create a DAG each. These two DAGs do not agree on the DGP. Of all the DAGs, the DAG created by 11a(right) is the closest we can get to our

Table 6 Data used in the causal inference example (total beams=716)

Element	New system (T=1)	Old system (T=0)
Branch A	81 out of 87 improved (93%)	243 out of 280 improved (87%)
Branch B	190 out of 262 improved (73%)	60 out of 87 improved (69%)
Totals	271 out of 349 improved (78%)	303 out of 367 improved (83%)

domain knowledge as obtained from notable studies on spalling that qualitatively paint a picture of the possible DGP behind spalling (Hertz 2003, Khoury 2000, Kodur 2000, Sanjayan and Stocks 1993) - despite the fact that this DAG does not tie P to spalling. This causal analysis may shed some grey over the suitability of the former ML analysis in (Naser and Kodur 2022) (as that particular analysis cannot be assumed to uncover the causality of spalling but rather a prediction tool from the first rung of the causal ladder). Hence, a more in-depth analysis of such DGP is better reserved for a dedicated study that is under exploration at the moment[26].

3.2 Causal inference

3.2.1 Case no. 3: Inferring the benefits of alternative strengthening systems

The following is a case study about inferring the causal effects of whether the adoption of a new strengthening system (i.e., made from fiber-reinforced polymer composites, T=1) is more beneficial than an existing system (e.g., crack filler, T=0) on the capacity of RC beams (*C*). the data were collected from two branches of the same firm (Branch A and Branch B) to be analyzed to decide if this firm can adopt the new system across all of its branches. To do so, we need to estimate the causal effect of the treatment (*T*) on the sectional capacity, as listed in Table 6.

A look at Table 6 shows that the new system outperforms the old system in both branches individually (93% vs. 87% and 73% vs. 69%, respectively). A second glance at this table indicates that the same system underperforms when compared in terms of the totals (78% vs. 83% for Branch A and Branch B, respectively). In all cases, it is unclear how the collected data can be interpreted and analyzed. This intriguing data is an example of Simpson's Paradox (Blyth 1972, Wagner 1982)[27] where a proper causal inference investigation is warranted to estimate the true causal effects.

The chief engineer (*E*) is a proponent of the new system. Thus, being the chief engineer affects adopting the new system (T = 1), which in turn affects sectional capacity (*C*). This can be translated into the DAG to the left in Fig. 12.

[26] Another reason to not delve into the DGP in this study is that preliminary analysis conducted using other algorithms (i.e., RFCI and FAS) identified the presence of confounders. Tackling such confounders is hectic and is best left for a future work.

[27] Please note that this is a hypothetical example that stems from those showcased in the works of Pearl (Pearl J 2018a) and Fabian (Dablander 2020).

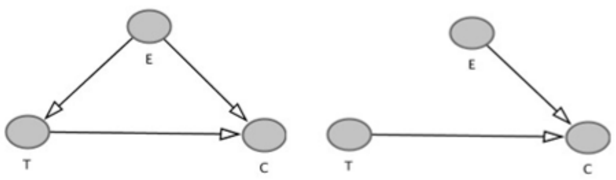


Fig. 12 DAG for the case study on hand

Fundamentally, this firm is interested in the probability of improvement if both branches are forced to adopt or not adopt the new system. In other words, the difference between these two probabilities is indeed the average causal effect in the population of RC beams listed in Table 6. This estimate resembles that which can be arrived at via Eqs. 2 and 3.

We will need to then calculate the $p(C=1|\text{do}(T=1))$ and $p(C=1|\text{do}(T=0))$, which can be arrived at from the DAG to the right of Fig. 12. For simplicity, these calculations are shown herein:

$$p(C=1|\text{do}(T=1)) = p(C=1|T=1, E=0)\, p(E=0)+p(C=1|T=1, E=1)\, p(E=1)$$
$$=81/87\times(87+280)/716+190/262\times(262+87)/716=0.83$$

And,

$$p(C=1|\text{do}(T=0)) = p(C=1|T=0, E=0)\, p(E=0)+p(C=1|T=0, E=1)\, p(E=1)$$
$$=243/280\times(87+280)/716+60/87\times(262+87)/716=0.78$$

Then, using Eq. (3) yields:

$\mathbb{E}[Y|do(T=1)]-\mathbb{E}[Y|do(T=0)]=0.83-0.78=0.05$.

Please note that this estimate (+5%) infers that, on average, 5% more RC beams would benefit if retrofitted using the new system. Surprisingly, this is opposite to the conclusion one might have drawn from the aggregated data in Table 5.

4. A note on causality and machine learning

This section presents additional insights and observations that arose during this work, which may spark future discussion on causality in our domain.

The bulk of ML works published in our area heavily rely on the notion of predictive models. A linear approach is adopted in which data is first collected and then fed into ML algorithm(s) for training and testing. The model is deemed *fit* once it satisfies a set of criteria and performance indicators. Revisiting Fig. 4 raises a series of questions. First, a common ML model aims to predict a value for an outcome (Erdal *et al.* 2018, Martí-Vargas *et al.* 2013, Yaswanth *et al.* 2021). This prediction may simply arise due to non-causal relations.

Second, the ML model treats all inputs from a data-driven perspective. As such, this model may alter such inputs (engineer features) in order to balance the required precision and/or simplify the model (among other items such as energy trade-offs, etc.). Most importantly, a typical ML model maximizes the variance between predictions and ground truth and assumes that the data points are *independent and identically distributed* (iid) [28]. The aforenoted is a common assumption in ML and is critical to be satisfied. Moreover, ML models rarely check for confoundness and selection bias. In some instances, ML users/engineers may not be aware of the presence of such items.

On the other hand, a causal model hopes to estimate a causal effect, and hence its fitness arises from the accuracy of its estimates. Such a model emphasizes the role of the input variables and seeks evidence (domain knowledge and judgment[29]) to identify such a role in establishing causality, minimizing/eliminating confoundness, and selection bias. An important note to remember is that a causal model is interpretable by nature as it conveys the DGP, however, a ML may not be[30].

While this study is a proponent of causality, it is worth noting that we may not need to truly establish causality and build causal models for all of our problems. In some problems, data-driven model continues to be of aid. In others, we do need to have interpretable ML models. For certain problems, we may not be in need of knowing the DGP as our domain knowledge can overcome such a need. It is also equally important to remember that establishing causality is challenging, and we may lack the data or tools to overcome such challenges. Perhaps future advancements in the coming years will help us find new ways to realize such a goal.

5. Conclusions

This work presents some of the big ideas behind causality, causal discovery, and causal inference from the lens of civil and structural engineering. The primary motivation of this study is to showcase the merit of adopting causal principles to identify and understand civil engineering-based data generating processes. In this pursuit, the Pearlian approach and the Potential Outcome approach are discussed and examined via case studies. The following list of inferences can also be drawn from the findings of this study:

- Identifying the data generating process is a natural pursuit for discovering new knowledge. This becomes an important goal for many of our complex problems.
- Causality in civil engineering is an exciting and uncharted area of research. Our domain is expected to thrive from future success in this area significantly. We advise to adopt causal models in lieu of traditional empirical, mathematical, and machine learning models.
- There is a dire need to integrate causal principles into

[28] Each data point was generated without reference to others, and the underlying distributions in the DGP are the same for all of the points.

[29] For transparency, some ML models also apply domain knowledge to justifying the rationale behind selecting or neglecting inputs.

[30] There could be a debate with regard to explainable AI (XAI). However, XAI primarily explains the reasoning behind a prediction from a data perspective, and not with regard to DGP. Thus, at a fundamental level, an XAI model may still remain a blackbox since our data may not be complete, may include confounders, etc. Additional discussion on XAI and recent advancements on this front can be found in (Dosilovic *et al.* 2018, Kovalerchuk *et al.* 2021).

commonly used ML methods as a means to maximize the output of such methods.

• A variety of causal discovery and causal inference algorithms and packages exist. Such packages may help accelerate the use of causality in our area. Yet, at the moment, none seem to be created by civil engineers and very little, if any, is utilized by them.

References

AISC (2022), AISC Shapes Database v15.0H, American Institute of Steel Construction Database, Chicago, IL, USA. https://www.aisc.org/search/?query=shapesdatabase&pageSize=10&page=1.

Allen, G.I. (2020), "Handbook of graphical models", *J. Am. Stat. Assoc.*, **115**(531), 1555-1557. https://doi.org/10.1080/01621459.2020.1801279.

Bareinboim, E., Tian, J. and Pearl, J. (2014), "Recovering from selection bias in causal and statistical inference", *Proceedings of the National Conference on Artificial Intelligence*, Québec City, Québec, Canada, July.

Quantumblacklabs/Causalnex (2021), Causalnex: A Python Library That Helps Data Scientists to Infer Causation Rather Than Observing Correlation, https://github.com/quantumblacklabs/causalnex.

Beyzatlar, M.A., Karacal, M. and Yetkiner, H. (2014), "Granger-causality between transportation and GDP: A panel data approach", *Transp. Res. A: Policy Pract.*, **63**, 43-55. https://doi.org/10.1016/j.tra.2014.03.001.

Blyth, C.R. (1972), "On Simpson's paradox and the sure-thing principle", *J. Am. Stat. Assoc.*, **67**(338), 364-366.

Bnlearn (2020), Bnlearn - Bayesian Network Structure Learning. https://www.bnlearn.com/

Bollen, K.A. and Pearl, J. (2013), "Eight myths about causality and structural equation models", *Handbook of Causal Analysis for Social Research*, Springer, Dordrecht, The Netherlands.

Bunge, M. (1979), *Causality and Modern Science*, Courier Corporation, North Chelmsford, Massachusetts, USA.

CausalGAM (2020), CRAN - Package CausalGAM, https://cran.r-project.org/web/packages/CausalGAM/index.html.

Center for Causal Discovery (2022), Data Science Research - Center for Causal Discovery, https://www.ccd.pitt.edu/people/data-science-research.

Chambliss, D.F. and Schutt, R.K. (2013), "Causation and experimental design", *Making Sense of the Social World: Methods of Investigation*, SAGE Publications, New York, NY, USA.

Correa, J.D. and Bareinboim, E. (2017), "Causal effect identification by adjustment under confounding and selection biases", *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, San Francisco, CA, USA, February.

Dablander, F. (2020), "An introduction to causal inference", *PsyArXiv*, **2020**, 1-15. https://doi.org/10.31234/osf.io/b3fkw.

Dimensions (2021), Dimensions.ai., https://www.dimensions.ai/.

Dosilovic, F.K., Brcic, M. and Hlupic, N. (2018), "Explainable artificial intelligence: A survey", *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, Opatija, Croatia, May.

Džeroski, S. (2009), "Relational data mining", *Data Mining and Knowledge Discovery Handbook*, Springer, Boston, MA, USA.

EconML. (2022), EconML - Microsoft Research, https://www.microsoft.com/en-us/research/project/econml/2022).

Erdal, H., Erdal, M., Şimşek, O. and Erdal, H.İ. (2018), "Prediction of concrete compressive strength using non-destructive test results", *Comput, Concrete*, **21**(4), 407-417. https://doi.org/10.12989/cac.2018.21.4.407.

Forney, A. and Mueller, S. (2022), "Causal inference in AI education: A primer", *J. Causal Inference*, **10**(1), 141-173. https://doi.org/10.1515/jci-2021-0048.

Gibb, A., Lingard, H., Behm, M. and Cooke, T. (2014), "Construction accident causality: Learning from different countries and differing consequences", *Constr. Manag. Econ.*, **32**(5), 446-459. https://doi.org/10.1080/01446193.2014.907498.

Glymour, C., Schemes, R., Spirtes, P. and Meek, C. (1994), "Regression and causation", Report CMU-PHIL-60; Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, USA.

Glymour, C., Zhang, K. and Spirtes, P. (2019), "Review of causal discovery methods based on graphical models", *Front. Gene.*, **10**, 524. https://doi.org/10.3389/fgene.2019.00524.

Glynn, A.N. and Kashin, K. (2018), "Front-Door versus back-door adjustment with unmeasured confounding: Bias Formulas for front-door and hybrid adjustments with application to a job training program", *J. Am. Stat. Assoc.*, **113**(523), 1040-1049. https://doi.org/10.1080/01621459.2017.1398657.

Heinze-Deml, C., Maathuis, M.H. and Meinshausen, N. (2018), "Causal structure learning", *Annual Review of Statistics and Its Application*, **5**, 371-391. https://doi.org/10.1146/annurev-statistics-031017-100630.

Hertz, K.D.D. (2003), "Limits of spalling of fire-exposed concrete", *Fire Saf. J.*, **38**(2), 103-116. https://doi.org/10.1016/S0379-7112(02)00051-6.

Holland, P.W. (1986), "Statistics and causal inference", *J. Am. Stat. Assoc.*, **81**(396), 945-960.

Huntington-Klein, N. (2021), *The Effect : An Introduction to Research Design and Causality*, Chapman and Hall/CRC, Boca Raton, FL, USA.

Imbens, G.W. (2020), "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics", *J. Econ. Liter.*, **58**(4), 1129-1179. https://doi.org/10.1257/jel.20191597.

Imbens, G.W. and Rubin, D.B. (2015), *Causal Inference: For Statistics, Social, and Biomedical Sciences*, Cambridge University Press, New York, NY, USA.

Kalainathan, D. and Goudet, O. (2022), Causal Discovery Toolbox Documentation - Causal Discovery Toolbox 0.5.23 documentation, https://fentechsolutions.github.io/CausalDiscoveryToolbox/html/index.html.

Khoury, G.A. (2000), "Effect of fire on concrete and concrete structures", *Prog. Struct. Eng. Mater.*, **2**(4), 429-447. https://doi.org/10.1002/pse.51.

Klemme, H.F. (2020), "Hume, David: A treatise of human nature", *Kindlers Literatur Lexikon*, Stuttgart Stuttgart, Germany.

Kodur, V.K.R. (2000), "Spalling in high strength concrete exposed to fire: Concerns, causes, critical parameters and cures", *Advanced Technology in Structural Engineering*, American Society of Civil Engineers, Reston, VA, USA.

Kovalerchuk, B., Ahmad, M.A. and Teredesai, A. (2021), "Survey of explainable machine learning with visual and granular methods beyond quasi-Explanations", *Studies in Computational Intelligence*, Springer, Cham, Switzerland.

Kovalerchuk, B. and Vityaev, E. (2000), *Data Mining in Finance: Advances in Relational and Hybrid Methods|Guide Books*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Lewis, D. (1973), "Causation", *J. Philos.*, **70**(17), 556-567. https://doi.org/10.2307/2025310.

Hossin, M. and Sulaiman, M.N. (2015), "A Review on evaluation metrics for data classification evaluations", *Int. J. Data Min. Knowl. Manag. Pr*, **5**(2), 1.

https://doi.org/10.5121/ijdkp.2015.5201.
Martí-Vargas, J.R., Ferri, F.J. and Yepes, V. (2013), "Prediction of the transfer length of prestressing strands with neural networks", *Comput. Concrete*, **12**(2), 187-209. https://doi.org/10.12989/cac.2013.12.2.187.
Michotte, A. (2017), *The Perception of Causality*, Routledge, New York, NY, USA.
Mitchell, T. (1997), *Machine Learning*, McGraw Hill, New York, NY, USA.
Muggleton, S. (1991), "Inductive logic programming", *New Gen. Comput.*, **8**, 295-318. https://doi.org/10.1007/BF03037089.
Naser, M.Z. and Alavi, A.H. (2021), "Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences", *Arch. Struct. Constr.*, **2021**, 1-19. https://doi.org/10.1007/s44150-021-00015-8.
Naser, M.Z. and Kodur, V.K. (2022), "Explainable machine learning using real, synthetic and augmented fire tests to predict fire resistance and spalling of RC columns", *Eng. Struct.*, **253**, 113824. https://doi.org/10.1016/j.engstruct.2021.113824.
Nogueira, A.R., Gama, J. and Ferreira, C.A. (2021), "Causal discovery in machine learning: Theories and applications", *J. Dyn. Games*, **8**(3), 203. https://doi.org/10.3934/jdg.2021008.
Nogueira, A.R., Pugnana, A., Ruggieri, S., Pedreschi, D. and Gama, J. (2022), "Methods and tools for causal discovery and causal inference", *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, **12**(2), e1449. https://doi.org/10.1002/widm.1449.
pcalg (2022), Methods for Graphical Models and Causal Inference [R package pcalg version 2.7-6], Comprehensive R Archive Network (CRAN).
Pearl, J. (2009a), "Causal inference in statistics: An overview", *Stat. Surv.*, **3**, 96-146. https://doi.org/10.1214/09-SS057.
Pearl, J. (2009b), *Causality*, Cambridge University Press, Cambridge, UK.
Pearl, J. (2013), "Causal diagrams and the identification of causal effects", *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge, UK.
Pearl, J. and Mackenzie, D. (2018a), *The Book of Why: The New Science of Cause and Effect-Basic Books*, Basic Books, New York, NY, USA.
Pearl, J. and Mackenzie, D. (2018b), *The Book of Why: The New Science of Cause and Effect, Notices of the American Mathematical Society*, Basic Books, New York, NY, USA.
Ramsey, J., Glymour, M., Sanchez-Romero, R. and Glymour, C. (2017), "A million variables and more: The fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images", *Int. J. Data Sci. Anal.*, **3**, 121-129. https://doi.org/10.1007/s41060-016-0032-z.
Rubin, D.B. (2005), "Causal inference using potential outcomes", *J. Am. Stat. Assoc.*, **100**(469), 322-331. https://doi.org/10.1198/016214504000001880.
Salmon, W.C. (2003), *Causality and Explanation*, Oxford University Press, Oxford, UK.
Sanjayan, G. and Stocks, L.J. (1993), "Spalling of high-strength silica fume concrete in fire", *ACI Mater. J.*, **90**(2), 170-173. https://doi.org/10.14359/4015.
Scheines, R. (1996), *An Introduction to Causal Inference*, Carnegie Mellon University, Pittsburgh, PA, USA.
Schölkopf, B. (2019), "Causality for machine learning", *arXiv preprint*, **1911**, 10500.
Sharma, A. and Kiciman, E. (2019), "DoWhy: A Python package for causal inference", https://github. com/microsoft/dowhy.
Spirtes, P., Glymour, C. and Scheines, R. (2000), "Causation, prediction, and search (Springer lecture notes in statistics)", *Lecture Notes in Statistics*, MIT Press, Cambridge, MA, USA.
Spirtes, P. and Zhang, K. (2016), "Causal discovery and inference: concepts and recent methodological advances", *Appl. Informat.*, **3**(1), 1-28. https://doi.org/10.1186/s40535-016-0018-x.
Surveys|NCSES|NSF (2022), https://www.nsf.gov/statistics/surveys.cfm.
Thelwall, M. (2018), "Dimensions: A competitor to Scopus and the Web of Science?", *J. Informetr.*, **12**(2), 430-435. https://doi.org/10.1016/j.joi.2018.03.006.
TIGRAMITE (2022), GitHub - Jakobrunge/Tigramite: Tigramite is a Python Package for Causal Inference with a Focus on Time Series Data, https://github.com/jakobrunge/tigramite.
Tong, T. and Yu, T.E. (2018), "Transportation and economic growth in China: A heterogeneous panel cointegration and causality analysis", *J. Transp. Geogr.*, **73**, 120-130. https://doi.org/10.1016/j.jtrangeo.2018.10.016.
Triantafillou, S. and Tsamardinos, I. (2016), "Score based vs constraint based causal learning in the presence of confounders", *CEUR Workshop Proceedings*.
Uber Technologies (2020), About Causal ML - Causalml Documentation, Uber Technologies Inc., San Francisco, USA. https://causalml.readthedocs.io/en/latest/about.html.
Vowels, M.J., Camgoz, N.C. and Bowden, R. (2021), "D'ya like DAGs? A survey on structure learning and causal discovery", *ACM Comput. Surv.*, **55**(4), 1-36. https://doi.org/10.1145/3527154.
Wagner, C.H. (1982), "Simpson's paradox in real life", *American Statistician*, **36**(1), 46-48.
Wardhana, K. and Hadipriono, F.C. (2003a), "Analysis of recent bridge failures in the United States", *J. Perfor. Constr. Facil.*, **17**(3), 144-150. https://doi.org/10.1061/(ASCE)0887-3828(2003)17:3(144).
Wasserman, L. (2021), "Causal inference", Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, USA.
Yaswanth, K.K., Revathy, J. and Gajalakshmi, P. (2021), "Artificial intelligence for the compressive strength prediction of novel ductile geopolymer composites", *Comput. Concrete*, **28**(1), 55-68. https://doi.org/10.12989/cac.2021.28.1.055.
Yu, K., Li, J. and Liu, L. (2016), "A review on algorithms for constraint-based causal discovery", *arXiv preprint*, **1611**, 03977.