Causal Discovery for Earth System Sciences

Section 3 – Functional Causal Models (FCMs)

Emiliano Díaz Salas Porras

Image and Signal Processing @ Universitat de València

Functional Causal Models (FCMs): Definition

Motivation: DAGs encode conditional independencies, but don't explain how variables influence one another. FCMs provide a **generative mechanism** behind observed data.

Definition: A **Functional Causal Model (FCM)** over variables $X_1, \ldots, X_n$ consists of:

- A collection of structural equations:

$$X_i = f_i(\mathrm{Pa}(X_i), N_i), \quad i = 1, \ldots, n$$

- Noise terms $N_1, \ldots, N_n$ assumed jointly independent

Interpretation:

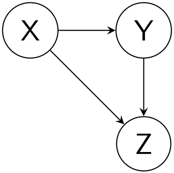
- Each variable is a function of its direct causes and a random noise term
- The FCM describes both the **mechanistic** and **probabilistic** elements of the system

Example: Functional Causal Model (FCM)

Variables: X, Y, Z
Structural Equations:

$$
\begin{aligned}
X &= N_X \\
Y &= 2X + N_Y \\
Z &= Y - X + N_Z
\end{aligned}
$$

Noise terms: $N_X, N_Y, N_Z \sim \mathcal{N}(0, 1)$, mutually independent



- Each variable is generated from its parents + noise
- The graph is induced by the structure of the equations

Properties of Functional Causal Models (FCMs)

Let an FCM be defined by:

$$X_i = f_i(\mathrm{Pa}(X_i), N_i), \quad \text{with } N_i \perp N_j \text{ for } i \neq j$$

Then the FCM has the following properties:

1. It induces a unique DAG $\mathcal{G}$
2. It induces a unique joint distribution $P(X_1, \ldots, X_n)$
3. $\mathcal{G}$ is a Markov I-map of the distribution P
4. The pair $(\mathcal{G}, P)$ forms a Bayesian Network
5. For any Bayesian Network $(\mathcal{G}, P)$, there exists an FCM that induces it
6. **However:** Some FCMs may induce **unfaithful** DAGs
7. **And:** Different FCMs (with different DAGs) may induce the same joint distribution

Takeaway: FCMs connect structure (graph) and mechanism (functions), but assumptions about function class and noise matter for identifiability

Exercise: FCM Can Induce Unfaithful DAGs

Consider the following Functional Causal Model:

$$\begin{aligned} X &= N_X \\ Y &= aX + N_Y \\ Z &= cX + bY + N_Z \end{aligned}$$

Assumptions: $N_X, N_Y, N_Z \sim \mathcal{N}(0, \sigma^2)$ i.i.d., mutually independent

Graph?:

Task: Find values of (a, b, c) for which $X \perp\!\!\!\perp Z$ (or $X \perp\!\!\!\perp Z \mid Y$), even though the DAG implies they are d-connected.

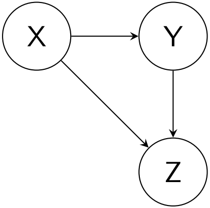
Exercise: FCM Can Induce Unfaithful DAGs

Exercise: Different FCMs Can Induce the Same Joint

Two Functional Causal Models (FCMs):

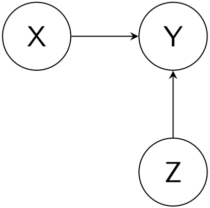
FCM 1:

$$X = N_X$$
$$Y = aX + N_Y$$
$$Z = cX + bY + N_Z$$



FCM 2:

$$X = N'_X$$
$$Z = N'_Z$$
$$Y = a'X + b'Z + N'_Y$$



Exercise: Different FCMs Can Induce the Same Joint

Task:

- Find parameters (a, b, c) and (a', b') such that both FCMs induce the same joint distribution
- Explain why the DAG is not identifiable from observational data alone

Causal Minimality for Functional Causal Models

Recall: An FCM defines a DAG $\mathcal{G}$ and induces a joint distribution P. We say that the FCM satisfies **causal minimality** if $\mathcal{G}$ is a minimal I-map of P.

Key Condition: Let $X_i = f_i(\mathrm{Pa}(X_i), N_i)$ be part of an FCM. Then the model is causally minimal if:

$$\forall X_j \in \mathrm{Pa}(X_i) : \quad X_i \not\perp\!\!\!\perp X_j \mid \mathrm{Pa}(X_i) \setminus \{X_j\}$$

Interpretation:

- Every parent has a **non-trivial influence** on its child
- No edge in the DAG is redundant — removing any would break the dependency

Why this matters:

- Allows testing for minimality from the FCM equations
- Important in identifiability, causal discovery, and functional modeling
- harder for Faithfulness as we would need to do global tests (possibly many)

Exercise: Hidden Conditioning Can Violate Reichenbach

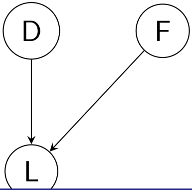
Variables:

- D = Distance to city (binary)
- F = Fire occurrence at time t (binary)
- L = Land cover at time $t+1$ (binary)

Functional Causal Model:

$$D = N_D \sim \text{Bernoulli}(0.5)$$
$$F = N_F \sim \text{Bernoulli}(0.5)$$
$$L = D \oplus F \oplus N_L$$
$$N_L \sim \text{Bernoulli}(0.01)$$

Notation: $\oplus$ denotes addition modulo 2 (XOR)



Exercise: Hidden Conditioning Can Violate Reichenbach

Task:

- Show that $D \perp\!\!\!\perp F$ marginally.
- Show that D and F become dependent when conditioning on $L = 0$.
- Explain why this looks like a violation of Reichenbach's principle, but is not.

Why Use Functional Causal Models (FCMs)?

1. **Generative Modeling - digital twins!**
   - FCMs specify how each variable is **generated** from its causes and independent noise
   - They model both **mechanism** and **probability**
2. **Automatic Markov Property**
   - Any FCM induces a DAG $\mathcal{G}$ that is Markov with respect to the joint
   - Markov factorization comes "for free" from the structure
3. **Convenient Language for describing change**
   - Describe stability/autonomy of the system, and therefore
   - Supports computing interventional and counterfactual distributions
4. **Extra Causal Information**
   - Interventional and counterfactual distributions which may distinguish between FCMs.
   - Asymmetries within markov equivalence class depending on the f's and N's can provide identifiability.

Interventions in Functional Causal Models

$$X_i = f_i(\text{Pa}(X_i), N_i)$$

An **intervention** modifies the FCM by changing structural equations.

Perfect Intervention: $\texttt{do}(X_i = x)$

- Replace $X_i = f_i(\cdot)$ with $X_i := x$
- Removes all incoming edges to X_i in the DAG

Imperfect Intervention:

- Replace $X_i = f_i(\text{Pa}(X_i), N_i)$ with another function $X_i := g(\text{Pa}(X_i), \tilde{N}_i)$
- May preserve some causal parents or alter noise distribution

Interventional vs Conditional Distributions:

$$P(Y \mid \texttt{do}(X = x)) \neq P(Y \mid X = x) \quad \text{(in general)}$$

Key Distinction:

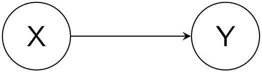
- **Conditioning** filters observations
- **Intervening** modifies the data-generating mechanism

Exercise: Conditionals vs Interventionals

$$X = N_X$$
$$Y = 4X + N_Y$$

Assume: $N_X, N_Y \sim \mathcal{N}(0,1)$, independent

DAG:



1. Compute the conditional distributions:

$$P(X \mid Y = y), \quad P(Y \mid X = x)$$

2. Compute the interventional distributions:

$$P(X \mid \mathtt{do}(Y = y)), \quad P(Y \mid \mathtt{do}(X = x))$$

3. Explain which ones differ and why

Exercise: Conditionals vs Interventionals

Solution: Conditionals vs Interventionals

1. Conditional Distributions

- $P(Y \mid X)$, $Y \mid X \sim \mathcal{N}(4X, 1)$
- $P(X \mid Y)$: Using linear Gaussian conditioning:

$$X \mid Y \sim \mathcal{N}\left(\frac{4}{17}Y,\ \frac{1}{17}\right)$$

2. Interventional Distributions

- $P(Y \mid \mathtt{do}(X))$: The structural equation for Y remains:

$$Y = 4X + N_Y \quad \Rightarrow \quad Y \mid \mathtt{do}(X) \sim \mathcal{N}(4X, 1)$$

- $P(X \mid \mathtt{do}(Y))$: The equation for Y is removed — Y is fixed by intervention, but X remains:

$$P(X \mid \mathtt{do}(Y)) = P(X) = \mathcal{N}(0, 1)$$

Conclusion:

- $P(Y \mid X) = P(Y \mid \mathtt{do}(X))$
- $P(X \mid Y) \neq P(X \mid \mathtt{do}(Y))$
- This reflects the asymmetry of cause and effect

When Does a Causal Effect Exist?

Motivation: Up to now, DAG gives us an intuitive notion of causes but no formal definition.

Definition: X has a causal effect on $Y \iff \exists N'_x : X \not\perp\!\!\!\perp Y$ in $P(Y \mid \mathtt{do}(X = N'_x))$

Equivalent Conditions:

- $X \not\perp\!\!\!\perp Y$ in the interventional distribution:
- $\exists x, x' : \quad P(Y \mid \mathtt{do}(X = x)) \neq P(Y \mid \mathtt{do}(X = x'))$
- $\exists x : \quad P(Y \mid \mathtt{do}(X = x)) \neq P(Y)$
- $\forall N'_x : N'_x$ has full support $\implies X \not\perp\!\!\!\perp Y$ in $P(Y \mid \mathtt{do}(X = N'_x))$ (Randomized trials)

Graphical criterion for causality

- If there is no directed path from X to Y, there is no causal effect
- If there is a directed path from X to Y there may not be a causal effect.
- If $\mathcal{G}$ is faithful to P then a directed path implies a causal effect.

Exercise: Myopia, Night Light, and Genetics

We examine whether **Night Light (NL)** exposure in infancy causes **Child Myopia (CM)**. We now include **Genetics (G)** as a common cause.

Two possible observed variable sets:

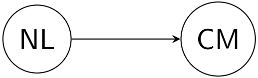
- (1) $\{\mathrm{NL}, \mathrm{CM}\}$
- (2) $\{\mathrm{NL}, \mathrm{PM}, \mathrm{CM}, \mathrm{G}\}$ where $\mathrm{PM} = \text{Parent Myopia}$

1. Suppose we observe that NL and CM are dependent in $P(\mathrm{NL}, \mathrm{CM})$.
   - What might we conclude in model (1)?
   - How does the conclusion change in model (2) if PM is included?
2. Analyze the interventional distribution:

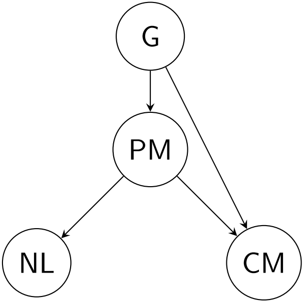
$$P(\mathrm{CM} \mid \mathtt{do}(\mathrm{NL}))$$

DAGs for the Myopia Example

Model 1: Observed = {NL, CM}



Model 2: Observed = {NL, PM, CM, G}



Counterfactuals in Functional Causal Models

What is a counterfactual?

- A statement about what **would have happened** under a different intervention, **given what actually happened**
- Example: "If the patient had not taken the drug, they would not have recovered"

Contrast: suppose $\mathcal{S}$ the original FCM. We then observe $X = x, Y = y$ and define a new FCM $\mathcal{S}'$ with noise variable N' with distribution $P(N|X = x, Y = y)$

- **Joint Distribution:** $P_{\mathcal{S}}(X, Y)$
- **Interventional:** $P(Y \mid \mathtt{do}(X = x))$
- **Counterfactual:** $P_{\mathcal{S}'}(Y \mid do(X = x'))$ — "What would Y have been, had X been x', given that we observed $X = x$ and $Y = y$?"

Counterfactuals in Functional Causal Models

Counterfactual evaluation steps:

1. **Abduction:** infer hidden noise values from observed data
2. **Action:** replace the structural equation for the intervened variable
3. **Prediction:** compute new outcome using same noise realization

Only possible using the full FCM — not the DAG or observational data alone.

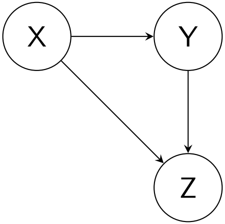
Exercise: Counterfactuals

FCM:

$$X = N_X$$
$$Y = X^2 + N_Y$$
$$Z = 2Y + X + N_Z$$

Assume: $N_X, N_Y, N_Z \sim \mathcal{N}(0,1)$, mutually independent

DAG:



Exercise: Counterfactuals

Suppose we observe: $(X, Y, Z) = (1, 2, 4)$

Tasks:

1. Compute $Y_{X=2}$: What would Y have been if X had been 2?
2. Compute $Z_{Y=5}$: What would Z have been if Y had been 5?
3. Compute $Z_{X=2}$: What is Z under a counterfactual change to $X = 2$?

Show: Counterfactuals are not transitive: Even though

$$\mathtt{do}(X = 2) \Rightarrow Y = 5, \quad \mathtt{do}(Y = 5) \Rightarrow Z = 10$$

it is not true that:

$$\mathtt{do}(X = 2) \Rightarrow Z = 10$$

Exercise: Counterfactuals

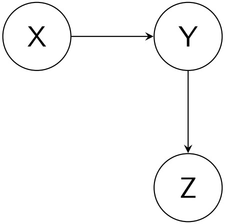
Exercise (Continued)

Modified FCM:

$$
\begin{aligned}
X &= N_X \\
Y &= X^2 + N_Y \\
Z &= 2Y + N_Z
\end{aligned}
$$

Assume: $N_X, N_Y, N_Z \sim \mathcal{N}(0, 1)$, mutually independent

Modified DAG:



Exercise (Continued)

Suppose we observe: $(X, Y, Z) = (1, 2, 4)$

Tasks:

1. Compute $Y_{X=2}$
2. Compute $Z_{Y=5}$
3. Compute $Z_{X=2}$

Show: Counterfactuals are now transitive:

$$X = 2 \Rightarrow Y = 5 \Rightarrow Z = 10 \quad \Rightarrow \quad X = 2 \Rightarrow Z = 10$$

Why? No direct path from X to Z — effect of X is fully mediated by Y

Exercise (continued)

Exercise: The Doctor's Dilemma

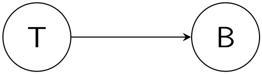
Variables:

- T = Treatment (0 = no treatment, 1 = treatment given)
- B = Blindness outcome (1 = patient goes blind, 0 = not blind)

FCM:

$$T = N_T \sim \text{Bernoulli}(0.5)$$
$$B = T \cdot N_B + (1 - T)(1 - N_B), \quad N_B \sim \text{Bernoulli}(0.01)$$

DAG:



Exercise: The Doctor's Dilemma

Scenario: We observe a patient with $(T, B) = (1, 1)$ — they were treated and went blind.

Task:

- What would have happened if the doctor had given $T = 0$?
- Compare the counterfactual $B_{T=0}$ to the interventional distribution $P(B \mid \mathtt{do}(T = 0))$
- What does this tell us about the limits of interventional analysis for individual decisions?

Hint: Use abduction to determine N_B from the observation.

Exercise: The Doctor's Dilemma

Solution: The Doctor's Dilemma

Step 1 – Abduction: Infer noise value

- Since $T = 1$, we have: $B = N_B$
- And $B = 1 \to$ so $N_B = 1$

Step 2 – Counterfactual: What if $T = 0$?

$$B_{T=0} = (1 - T)(1 - N_B) = (1 - 0)(1 - 1) = 0$$

Answer: The patient would NOT have gone blind without treatment!

Step 3 – Compare with Interventional Probability:

$$P(B = 1 \mid \mathtt{do}(T = 0)) = P(N_B = 0) = 0.01$$

Interpretation:

- Interventions describe population effects.
- Counterfactuals describe individual-level effects.
- For this patient, treatment caused blindness — even if it's rare!

Exercise: FCMs Can Differ in Counterfactuals

Two FCMs over variables X_1, X_2, X_3

FCM 1:

$$X_1 = N_1, \quad X_2 = N_2, \quad X_3 = \left\{ \begin{array}{ll} X_1 & \text{if } N_3 > 0 \\ X_2 & \text{if } N_3 = 0 \end{array} \right\} \cdot \mathbf{1}_{X_1 \neq X_2} + N_3 \cdot \mathbf{1}_{X_1 = X_2}$$

FCM 2:

$$X_1 = N_1, \quad X_2 = N_2, \quad X_3 = \left\{ \begin{array}{ll} X_1 & \text{if } N_3 > 0 \\ X_2 & \text{if } N_3 = 0 \end{array} \right\} \cdot \mathbf{1}_{X_1 \neq X_2} + (2 - N_3) \cdot \mathbf{1}_{X_1 = X_2}$$

Assume: $N_1, N_2 \sim \text{Bern}(0.5)$; $N_3 \sim \text{Uniform}\{0, 1, 2\}$

Task:

- Show that both FCMs induce:
    - The same DAG, ,joint and interventional distributions
- But they differ in counterfactuals — e.g., when $(X_1, X_2, X_3) = (1, 0, 0)$, what would X_3 have been if $X_1 = 0$?

Exercise: FCM provide additional info

Solution: FCM provide additional info

Observation: $(X_1, X_2, X_3) = (1, 0, 0)$

Step 1 – Abduction (infer N_3):

- Since $X_1 \neq X_2$ and $X_3 = 0$, then from both FCMs:

$$X_3 = X_2 \text{ if } N_3 = 0 \quad \Rightarrow \quad (N_1, N_2, N_3) = (1, 0, 0)$$

Step 2 – Counterfactual: What if $X_1 = 0$?

- $(X_1, X_2) = (0, N_2) = (0, 0)$
- We have $X_1 = X_2$ so we're in the second case of the equation
- And $N_3 = 0$
- **FCM 1:** $X_3 = N_3 = \mathbf{0}$
- **FCM 2:** $X_3 = 2 - N_3 = \mathbf{2}$

Solution: : FCM provide additional info

Conclusion:

- Both FCMs produce the same joint distribution and interventions
- But their counterfactual predictions differ because they use different equations for the same DAG
- **Counterfactuals encode extra information** — only available in the full FCM

Identifiability in Causal Discovery

Problem: We observe a joint distribution $P(X_1, \ldots, X_n)$ — can we identify the true causal DAG?

Under basic assumptions:

- FCM induces a DAG that is a Markov I-map of P
- Faithfulness: $I(P) = I(\mathcal{G})$
- Causal sufficiency: no hidden confounders
- Access to an oracle for conditional independence

Then: We can recover the **Markov Equivalence Class (MEC)** of the true DAG $\rightarrow$ But not the DAG itself

Limit of Identifiability:

$$\text{FCM} + \text{Faithfulness} + \text{Causal Sufficiency} + CI \text{ oracle} \quad \Rightarrow \quad \text{CPDAG}$$

MECs contain multiple DAGs: All DAGs in a MEC encode the same conditional independencies, but may differ in edge directions

Identifiability beyond Markov equivalence class

The Functional Causal model allows us to obtain identifiablity up to the actual DAG if we can make additional assumptions about the form of the equations.

It has been proven that the following types of FCMs are identifiable with certain small exceptions:

- $y = f(x) + N_x$ (Additive)
- $g(y) = f(x) + N_y$ (Post-Non linear)
- $y = f(x) + g(x) * N_y$ (Loction-Scale)

So if we apriori assume these cases we can identify them.

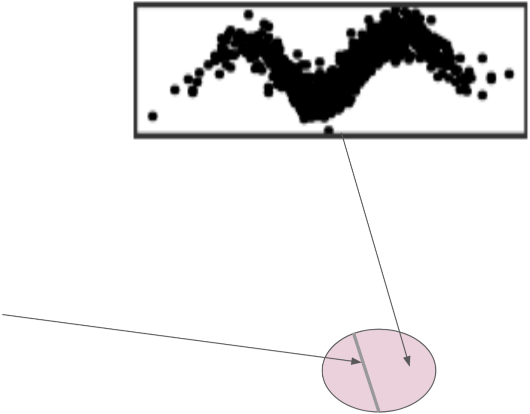
Identifiability beyond Markov equivalence class

FCM : Additive Noise model :

$$y = f(x) + z$$

Result: (Hoyer et al. 2009)

Exceptions: eg. linear gaussian

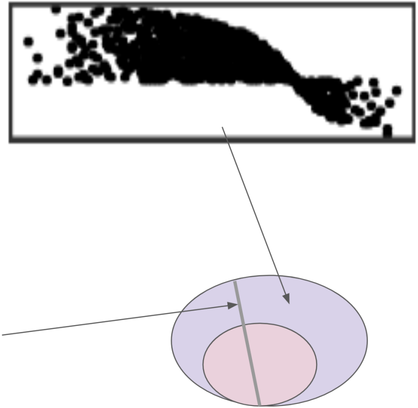


Hoyer, Patrick et al. "Nonlinear causal discovery with additive noise models.". *Neurips*. (2009).

Identifiability beyond Markov equivalence class

FCM : Post non-linear model :

$$y = g(f(x) + z)$$



Result: (Zhang and Hyvärinen, 2009)

Exceptions: eg. noise z generalized mixture of exponentials, f(x) two-sided, asymptotically exponential, f,g strictly monotonic

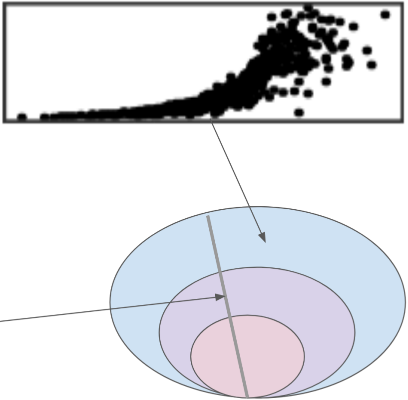
Zhang, Kun, and Aapo Hyvarinen. "On the identifiability of the post-nonlinear causal model.". UAI, (2009).

FCM : Location scale model :

$$y = f(x) + g(x) * z$$

Result: (Immer et al. 2023)

Exceptions: eg. noise z gaussian, x log-mix-rational log, f and g functions of polynomials of degree 2 or less.

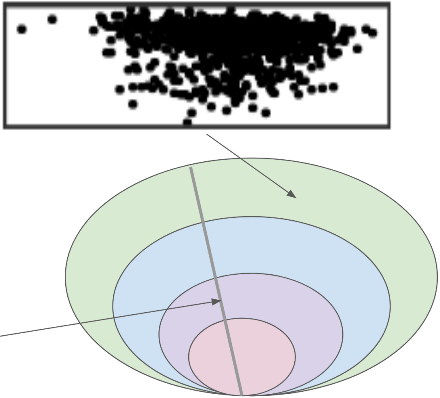


Immer, Alexander et al. "On the identifiability and estimation of causal location-scale noise models." *ICML*. (2023).

Identifiability beyond Markov equivalence class

FCM : general modular model :

$$y = f(x, z)$$

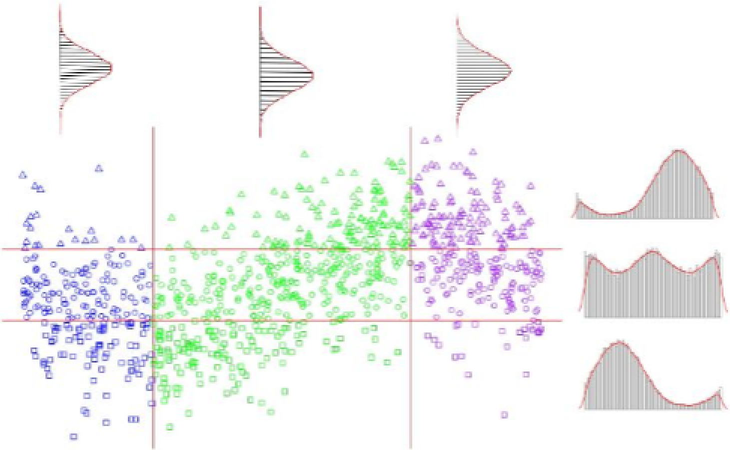


~~Result~~ Principle:
Independence of cause and mechanism (ICM) (Daniusis et al, 2010)
Equivalent to minimal complexity factorization

Exceptions: Both directions algorithmically independent. ie both factorizations have equal complexity

Daniussi, Povilas, et al. "Inferring deterministic causal relations." *UAI* (2010).

Identifiability beyond Markov equivalence class



- modularity/autonomy assumption
- $p(x)$ algorithmically independent from $p(y|x)$)
- ie no info about $p(y|x)$ in $p(x)$

python notebook