Specific and generic attribution of climate extreme events

Alexis Hannart
CNRS

26 April 2016
References


Is this event attributable to human influence?
Context

Is this event attributable to human influence?

- general public and media
- policymakers and planners
- risk managers
- courts, insurers
- climate scientists

Need for causal answers from scientists.
Context

- **Conventional approach:**
  - define the event
  - derive the probability of the event in two model ensembles:
    - factual probability: $p_1$
    - counterfactual probability: $p_0$
  - derive a metric comparing $p_1$ to $p_0$ (FAR = 1-$p_0/p_1$ and/or RR = $p_1/p_0$)
  - make corresponding claim:
    - «80% of the event is causally attributable to human emissions»
    - «human emissions caused the risk of the event to increase by a factor 5»

- **Emerging approach:**
  - use «a physically based approach that is strongly linked to the occurrence»
  - in order to find out «why such an extreme occurred the way it did.»

Are these methods and causal claims justified?
Quotes:

- "Causality is the relationship between something that happens or exists and an effect, result, or condition for which it is responsible." (NAS report, 2016)
- "We can’t say anything for individual event, we can only say something about classes of event"
Context

- Quotes:
  - “Causality is the relationship between something that happens or exists and an effect, result, or condition for which it is responsible.” (NAS report, 2016)
  - “We can’t say anything for individual event, we can only say something about classes of event”

What is a cause?

How does one evidence a causal link from data?

How does one quantify uncertainty on a causal link?
Outline

- Context

- Causal theory

- Specific event attribution
Brief history of causality

- A much debated matter.
- A long philosophical tradition.
- A research topic in cognitive science / logic / computing science / statistics.

 emergence of a consensual framework (1980s-1990s).
Structural causal theory

  - A general theoretical corpus on causality.
  - Clear definitions and tools for causal evidencing.
  - Turing Award 2004.

- Diverse fields of applications:
  - Epidemiology.
  - Economics.
  - Social science.

- Diverse questions:
  - Legal attribution/compensation.
  - Policy elaboration/evaluation.
  - Certification/regulatory aspects.
Overview of the theory – the core

\[
\begin{align*}
PN &= P(Y_0 = 0 \mid Y = 1, X = 1) \\
PS &= P(Y_1 = 1 \mid Y = 0, X = 0) \\
PNS &= P(Y_0 = 0, Y_1 = 1)
\end{align*}
\]
Overview of the theory – probabilistic framework

- A probabilistic framework
  - handle all factors, observed and not observed.
  - allow for uncertainty on causal assessments.
  - notion of dependence between random variables.
  - notion of conditional independence.
Overview of the theory - probabilistic framework

- A probabilistic framework
  - handle all factors, observed and not observed.
  - allow for uncertainty on causal assessments.
  - notion of dependence between random variables.
  - notion of conditional independence.

- Let $X, Y, Z$ be random variables.
  - $X$ is independent of $Z$ conditionally on $Y$ iif

\[ P(Z \mid X, Y) = P(Z \mid Y) \]
Overview of the theory - conditional independence

- Conditional independence factorization:
  \[ P(X_1, X_2, \ldots, X_n) = \Pi_{i=1}^{n} P(X_i \mid P_i) \]

- Let X, Y, Z be random variables (e.g. binary).
  - X: barometer
  - Y: rain
  - Z: road wet
  - W: low pressure system

  \[ P(X, Y, Z, W) = P(W) \cdot P(X \mid W) \cdot P(Y \mid W) \cdot P(Z \mid Y) \]
Overview of the theory - oriented graphs

- Oriented graphs:
  - visual representation of the conditional independence structure of a joint distribution

\[
P(X, Y, Z, W) = P(W) \cdot P(X \mid W) \cdot P(Y \mid W) \cdot P(Z \mid Y)
\]
Overview of the theory - oriented graphs

- Oriented graphs:
  - visual representation of the conditional independence structure of a joint distribution
  - well-suited to encode causal dependence relationships.

- The joint PDF bears the signatures of causality:

  independent causes become dependent conditional on their common effect.

  dependent effects become independent conditional on their common cause.

- Causal relationships may be found by estimating the joint PDF.
Overview of the theory - oriented graphs

- Example:
  - X and Z are Gaussian iid
  - effect event: X+Z>2
Overview of the theory - oriented graphs

- Is it all it takes to model and evidence causality?

- Not quite: oriented graphs are ambiguous.

- The same causal graph may be compatible with two very different causal situations.

- And we still do not have a definition for « cause ». 
Overview of the theory - illustration

- **Situation A**
  
  - Hypothesis:
    - $X$ and $Z$ are Bernoulli independent ($p = 0.5$).
    - $Y = X \times Z$.
  
  - Graph of the joint PDF of $(X,Y,Z)$:

- **Situation B**
  
  - Hypothesis:
    - $X$ and $Z$ are Bernoulli independent ($p = 0.5$).
    - $Y = 1 - (1-X) \times (1-Z)$.
  
  - Graph of the joint PDF of $(X,Y,Z)$:
Overview of the theory – illustration

Situation A

Situation B
Overview of the theory - oriented graphs

- Is it all it takes to model and evidence causality?
- Not quite: oriented graphs are ambiguous.
- The same causal graph may be compatible with two very different causal situations.

Sufficiency versus Necessity
Overview of the theory - interventional probability

- probability of light **forcing** the switch to on/off.
- experimental context in which the switch is manipulated.

- probability of light **knowing** the switch is on/off.
- non-experimental context in which the switch is unconstrained.

\[ P(Y \mid do(X = x)) \neq P(Y \mid X = x) \]
Overview of the theory – core definitions

- PN = probability that Y does not occur when X is turned off, given that Y and X were occurring.
- PS = probability that Y occurs when X is turned on, given that Y and X were not occurring.
- PNS = probability that Y occurs when X is turned on and Y does not occur when X is turned off.

\[
\begin{align*}
\text{PN} &= \text{def } P(Y_0 = 0 \mid Y = 1, X = 1) \\
\text{PS} &= \text{def } P(Y_1 = 1 \mid Y = 0, X = 0) \\
\text{PNS} &= \text{def } P(Y_0 = 0, Y_1 = 1)
\end{align*}
\]
Illustration

**Situation A**

PN = 1, PS = \(\frac{1}{2}\), PNS = \(\frac{1}{2}\)

**Situation B**

PN = \(\frac{1}{2}\), PS = 1, PNS = \(\frac{1}{2}\)
Possible application to any causal situation

Switch X

Bulb Y

PN = .., PS = .., PNS = ..
Overview of the theory - necessary and sufficient causation

- How to calculate PN, PS and PNS in practice?
- A few convenient theorems are available:

\[
\begin{align*}
PN &= 1 - \frac{p_0}{p_1} + \frac{p_0 - P(Y_0=1)}{P(X=1,Y=1)} \\
PS &= 1 - \frac{1-p_1}{1-p_0} - \frac{p_1 - P(Y_1=1)}{P(X=0,Y=0)} \\
PNS &= P(Y_1 = 1) - P(Y_0 = 1)
\end{align*}
\]

under the assumption of monotonicity and where:
- \( p_1 = P(Y=1 \mid X=1) \): factual probability of the event
- \( p_0 = P(Y=1 \mid X=0) \): counterfactual probability of the event
Overview of the theory - necessary and sufficient causation

- How to calculate PN, PS and PNS in practice?

- A few convenient theorems are available:

$$PN = 1 - \frac{p_0}{p_1}, \quad PS = 1 - \frac{1 - p_1}{1 - p_0}, \quad PNS = p_1 - p_0$$

under the assumption of monotonicity and exogeneity and where:

- $p_1 = P( Y=1 \mid X = 1 )$: factual probability of the event
- $p_0 = P( Y=1 \mid X = 0 )$: counterfactual probability of the event
Critical requirement: experimentation

- In many fields, experimentation is possible.
- E.g. in life sciences, causal evidence can be collected through in vivo experiments (counterfactual = placebo)

![Diagram of group A receiving new drug and group B receiving placebo, with some people healed and others sick.](image)
Critical requirement: experimentation

- In climate science, *in situ* experimentation at large scale is not an option.
- The only option that remains is so called *in silico* experimentation on numerical avatars of Earth, i.e. no models => no causal evidence.

**group of Earths A: Greenhouse Gases**

**group of Earths B: No Greenhouse Gases**
Overview of the theory - necessary and sufficient causation

\[
\begin{align*}
\text{PN} &= 1 - \frac{p_0}{p_1}, \\
\text{PS} &= 1 - \frac{1 - p_1}{1 - p_0}, \\
\text{PNS} &= p_1 - p_0
\end{align*}
\]

- PN matches with the FAR under exogenity and monotonicity.
- PN provides a new interpretation of the FAR.
- PS provides a new information w.r.t. causality.
Overview of the theory - necessary and sufficient causation

![Graphs](image)
Overview of the theory - necessary and sufficient causation
Overview of the theory - necessary and sufficient causation
Overview of the theory - necessary and sufficient causation
Rare events are attributable in a necessary causation sense.

\[ PN = \max\{1 - \frac{p_0}{p_1}, 0\} \]

is the right quantity to look at under exogeneity.
Does FAR represent a fraction of attributable risk?

- anthropogenic emissions
- intense SACZ
- anticyclonic conditions
- heatwave

FAR = 0.8
Does FAR represent a fraction of attributable risk?

- Anthropogenic emissions
- Intense SACZ
- Anticyclonic conditions
- Heatwave

FAR = 0.8

FAR2 = 0.2?
Does FAR represent a fraction of attributable risk?

- **anthropogenic emissions**
  - **intense SACZ**
  - **anticyclonic conditions**
    - **heatwave**

PN1 = FAR1 = 0.8
PN2 = FAR2 = 0.9*
PN3 = FAR3 = 0.9*

FAR does not represent a share of causality. FARs do not sum up to one.
PN could be used instead.

*stylized
Does FAR represent a fraction of attributable risk?

- anthropogenic emissions
  - intense SACZ
    - anticyclonic conditions
      - heatwave

PN1 = FAR1 = 0.8
PN2 = FAR2 = 0.9*
PN3 = FAR3 = 0.9*

"Human emissions are very likely a necessary cause of the heatwave".
Conventional approach:
- define the event
- derive the probability of the event in two model ensembles:
  - factual probability: $p_1$
  - counterfactual probability: $p_0$
- derive metrics comparing $p_1$ to $p_0$: FAR = $1 - p_0/p_1$, RR = $p_1/p_0$
- make a statement:
  - « 80% of the event is causally attributable to human emissions »
  - « human emissions increased the risk of the event by a factor 5 »

Emerging approach:
- use « a physically based approach that is strongly linked to the occurrence »
- in order to find out « why such an extreme occurred the way it did. »

Are these methods and causal claims justified?
Context

- Conventional approach:
  - define the event
  - derive the probability of the event in two model ensembles:
    - factual probability: $p_1$
    - counterfactual probability: $p_0$
  - derive metrics comparing $p_1$ to $p_0$:
    \[ \text{FAR} = 1 - \frac{p_0}{p_1} \]
    \[ \text{RR} = \frac{p_1}{p_0} \]
  - make a statement:
    - « 80% of the event is causally attributable to human emissions »
    - « human emissions increased the risk of the event by a factor 5 »

- Emerging approach:
  - use « a physically based approach that is strongly linked to the occurrence »
  - in order to find out « why such an extreme occurred the way it did. »

Are these methods and causal claims justified?
Illustration: Socrates

S1: “Ingesting hemlock causes death”.

S2: “Socrates ingesting hemlock was the cause of Socrates death”.

- S1 is an epidemiological claim:
  - general tendency of the cause to bring about the effect at the population level.
  - requires a population level risk comparison: $p_0$, $p_1$.

- S2 is a medical claim:
  - diagnosis of an individual with an explanation of the causal chain.
  - requires an autopsy.
Q1: “Does ingesting hemlock cause death?”
Q2: “What caused Socrates death?”

- **Q1** is a generic question:
  - Policymaker on health.
  - Risk manager.

- **Q2** is a specific question:
  - Courts (litigation).
  - Media, public.
  - Scientists.
Illustration: Jonas

- Jonas is a snowstorm that occurred 22-24 January 2016 over Northeastern US.

S1: “Human CO2 emissions cause extreme snowstorms in the Northeast of the United States.”

S2: “Human CO2 emissions caused the extreme snowstorm Jonas.”
Definitions for “event”, “generic”, and “specific”

- Realizations $\omega$
- Specific realization $\omega^*$
- Sample space $\Omega$
- Subset $C$: cause event
- Subset $E$: effect event
Definitions for “event”, “generic”, and “specific”

\[ Y(\omega) = y : \]

- realizations \( \omega \)
- specific realization \( \omega^* \)
- sample space \( \Omega \)
- subset \( C \): cause event
- subset \( E \): effect event
Definitions for “event”, “generic”, and “specific”

The focus is on the entire sample space.

The focus is on a single realization.
Definitions for “event”, “generic”, and “specific”

The focus is on the entire sample space.
- Probabilistic
- Necessary causation, Sufficient causation

The focus is on a single realization.
- Deterministic
- Actual causation
Causal theory (Pearl): actual causation

(Actual cause) $\bar{X} = \bar{x}$ is an actual cause of $\varphi$ in $(M, \bar{u})$ if the following three conditions hold:

AC1. $(M, \bar{u}) \models (\bar{X} = \bar{x}) \land \varphi$. (That is, both $\bar{X} = \bar{x}$ and $\varphi$ are true in the actual world.)

AC2. There exists a partition $(\bar{Z}, \bar{W})$ of $\mathcal{V}$ with $\bar{X} \subseteq \bar{Z}$ and some setting $(\bar{x}', \bar{w}')$ of the variables in $(\bar{X}, \bar{W})$ such that if $(M, \bar{u}) \models Z = z^*$ for $Z \in \bar{Z}$, then

(a) $(M, \bar{u}) \models [\bar{X} \leftarrow \bar{x}', \bar{W} \leftarrow \bar{w}']^{-} \varphi$. In words, changing $(\bar{X}, \bar{W})$ from $(\bar{x}, \bar{w})$ to $(\bar{x}', \bar{w}')$ changes $\varphi$ from true to false,

(b) $(M, \bar{u}) \models [\bar{X} \leftarrow \bar{x}, \bar{W} \leftarrow \bar{w}', \bar{Z}' \leftarrow \bar{z}'^*] \varphi$ for all subsets $\bar{Z}'$ of $\bar{Z}$. In words, setting $\bar{W}$ to $\bar{w}'$ should have no effect on $\varphi$ as long as $\bar{X}$ is kept at its current value $\bar{x}$, even if all the variables in an arbitrary subset of $\bar{Z}$ are set to their original values in the context $\bar{u}$.

AC3. $\bar{X}$ is minimal; no subset of $\bar{X}$ satisfies conditions AC1 and AC2. Minimality ensures that only those elements of the conjunction $\bar{X} = \bar{x}$ that are essential for changing $\varphi$ in AC2(a) are considered part of a cause; inessential elements are pruned.
Causal theory (Pearl): actual causation

- How does one establish actual causation?
  - sensitivity studies in a model by intervening on some variables while prescribing others to their observed values under $\omega$

- Example:
  - $\omega$: switch X is turned on, then switch Z is turned on.

```
Switch Z is an actual cause
Switch X
Switch Z
Bulb Y
```

```
Switch X is an actual cause
Switch Z
Switch Z
Bulb Y
```
Causal theory (Pearl): actual causation

- How does one establish actual causation?
  - sensitivity studies in a model by intervening on some variables while prescribing others to their observed values under $\omega$

- Example:
  - sensitivity of precipitation in Boulder to the anthropogenic forcing when prescribing the circulation (Trenberth et al. 2015)

TPW, 13 Sept. 2013 (NCEP)
Causal theory (Pearl): actual causation

- How does one establish actual causation?
  - sensitivity studies in a model by intervening on some variables while prescribing others to their observed values under $\omega$

- Example:
  - sensitivity of precipitation in Boulder to the anthropogenic forcing when prescribing the circulation (Trenberth et al. 2015)

  Heuristic reasoning only, no quantification.

  How can this be done in practice?

  Data Assimilation is an option.
Illustration: Boudor extreme summer precipitation occurrence
Illustration: Boucher extreme summer precipitation occurrence
Illustration: Bouder extreme summer precipitation occurrence
Summary

- Causal theory can help:
  - clarifying semantics.
  - justifying the methods, possibly improving them.
  - clarifying the distinction generic/specific.

- An individual event can be defined as a realization.

- Causal diagnostic of singular events may be implemented using DA.

- Synergy with existing infrastructure at NWP centers.

- Research under way:
  - Articulating with actual causation.
  - Experiments using larger models (ICTP AGCM, WRF)
  - Implementation on real case studies.
Thank you

Founders of Dadaïsm, Zürich, 1915
Exploring the causal chain: likelihood ratio approach

**Objectives:**
- comparing the PDF (or likelihood) of $y$ in both worlds.
- identifying the features creating a gap.

**Difficulties:**
- $y$ is a very high-dimensional vector ($N \sim 10^8$).
- the same event will never be randomly simulated.

**Proposed solution:** Data Assimilation.
Outlook of Data Assimilation: origins

Observations: multiple sensors

State vector: atmospheric model

Numerical Weather Prediction requires to **initialize** the model every six hours with **new observations**.
Outlook of Data Assimilation: algorithms

Observations: multiple sensors

State vector: atmospheric model

**Goal**: deriving the PDF of $X$ conditional on $Y = y$

high dimensional Bayesian update.
The “primitive equations” of data assimilation

Assumptions: Hidden Markov model

- Dynamic equation:
  \[ X_{t+1} = M(X_t, F_t) + v_t \]
- Observational equation:
  \[ Y_t = H(X_t) + w_t \]
- \( v_t \) and \( w_t \) Gaussian error terms with covariance \( Q \) and \( R \);
- \( M \) is the model with \( F_t \) external forcing;
- \( H \) is the observation operator.
The “primitive equations” of data assimilation

Assumptions:
Hidden Markov model

- Dynamic equation:
  \[ X_{t+1} = M(X_t, F_t) + v_t \]
- Observational equation:
  \[ Y_t = H(X_t) + w_t \]
- \( v_t \) and \( w_t \) Gaussian error terms with covariance \( Q \) and \( R \);
- \( M \) is the model with \( F_t \) external forcing;
- \( H \) is the observation operator.

Solution:
Gaussian linear approximation

- Propagation equation:
  \[ x_{t+1}^f = M x_t^a \]
  \[ P_{t+1}^f = MP_t^a M' + Q \]
- Update equation:
  \[ x_t^a = x_t^f + K(y_t - Hx_t^f) \]
  \[ P_t^a = (I - KH)P_t^f \]
  \[ K = P_t^f H'(HP_t^f H' + R)^{-1} \]
The likelihood $f(y)$ is a by-product of data assimilation.

Solution: Gaussian linear approximation

- Propagation equation:
  \[
  x_{t+1}^f = M x_t^a \\
  P_{t+1}^f = M P_t^a M' + Q
  \]

- Update equation:
  \[
  x_t^a = x_t^f + K (y_t - H x_t^f) \\
  P_t^a = (I - KH) P_t^f \\
  K = P_t^f H' (H P_t^f H' + R)^{-1}
  \]

By-product: PDF of observation $y$

Likelihood equation:

\[
- \log f(y) = \sum_{t=0}^{T} \frac{1}{2} \log |\Sigma_t| + \frac{1}{2} d_t' \Sigma_t^{-1} d_t
\]

with:
\[
d_t = y_t - H x_t^f \\
\Sigma_t = H P_t^f H' + R
\]
Experiments in the ICTP AGCM model

geopotential height

surface temperature

truth

factual analysis

counterfactual analysis
Experiments in the ICTP AGCM model

- **surface temperature**
- **wind**
- **geopotential height**
Defining the probabilities of causation – how it works

Assumptions:
- switch X is on with proba p.
- other switch is on with proba q.
- switches are independent.
Defining the probabilities of causation – how it works

All possible states of the world

X

Y

proba = pq

X

Y

proba = p(1-q)

X

Y

proba = (1-p)q

X

Y

proba = (1-p)(1-q)

PN = P(Y_0 = 0 \mid Y = 1, X = 1)

PS = P(Y_1 = 1 \mid Y = 0, X = 0)
Defining the probabilities of causation – how it works

1. States satisfying to \( Y = 1, X = 1 \)

\[ \text{proba} = pq \]

\[ \text{PN} = P(Y_0 = 0 \mid Y = 1, X = 1) \]
Defining the probabilities of causation – how it works

States satisfying to $Y = 1, X = 1$

$\text{proba} = pq$

$PN = P(Y_0 = 0 \mid Y = 1, X = 1)$
Defining the probabilities of causation – how it works

States satisfying to $Y = 1, X = 1$

$\text{proba} = pq$

$PN = P(Y_0 = 0 \mid Y = 1, X = 1)$

$PN = 1$
Defining the probabilities of causation – how it works

All possible states of the world

\[
\begin{align*}
\text{proba} &= pq \\
\text{proba} &= p(1-q) \\
\text{proba} &= (1-p)q \\
\text{proba} &= (1-p)(1-q)
\end{align*}
\]

\[
\begin{align*}
\text{PN} &= P(Y_0 = 0 \mid Y = 1, X = 1) \\
\text{PS} &= P(Y_1 = 1 \mid Y = 0, X = 0)
\end{align*}
\]
Defining the probabilities of causation – how it works

1. States satisfying to $Y = 0, X = 0$

\[
\begin{align*}
\text{proba} &= (1-p)q \\
\text{proba} &= (1-p)(1-q)
\end{align*}
\]

\[
PS = P(Y_1 = 1 \mid Y = 0, X = 0)
\]
Defining the probabilities of causation – how it works

States satisfying to $Y = 0, X = 0$

\[ \text{proba} = pq \]
\[ \text{proba} = p(1-q) \]
\[ \text{proba} = (1-p)q \]
\[ \text{proba} = (1-p)(1-q) \]

$\text{PS} = P(Y_1 = 1 \mid Y = 0, X = 0)$
Defining the probabilities of causation – how it works

States satisfying to $Y = 0, X = 0$

\[
\text{proba} = (1-p)q \quad \text{proba} = (1-p)(1-q)
\]

\[PS = \frac{(1-p)q}{(1-p)q + (1-p)(1-q)} = q\]
Defining the probabilities of causation

1. Switch X
   Bulb Y

\[ PN = 1, \ PS = q, \ PNS = q. \]

2. Switch X
   Bulb Y

\[ PN = 1-q, \ PS = 1, \ PNS = 1-q \]

3. Switch X
   Bulb Y

\[ PN = 1, \ PS = 1, \ PNS = 1 \]
How good is model M1 in representing the observed trajectory y in comparison to model M2?
How good is model M1 in representing the observed trajectory $y$ in comparison to model M2?
How good is model M1 in representing the observed trajectory $y$ in comparison to model M2?
Context

How good is model M1 in representing the observed trajectory y in comparison to model M2?

Proposal: compute the likelihood of observations in each model.

Formulation: statistical model = numerical model + measurement error.

Computation technique: data assimilation.