Causality

Bernhard Schölkopf and Jonas Peters
MPI for Intelligent Systems, Tübingen

MLSS, Tübingen

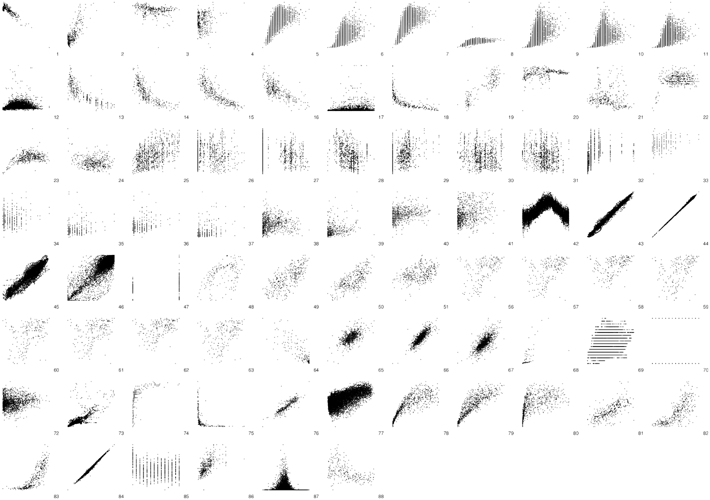
21st July 2015

	Overall
Treatment A: Open surgery	78% (273/350)
Treatment B: Percutaneous nephrolithotomy	**83%** (289/350)

Charig et al.: "Comparison of treatment of renal calculi by open surgery, (...) ", British Medical Journal, 1986

	Overall	Patients with small stones	Patients with large stones
Treatment A: Open surgery	78% (273/350)	**93%** (81/87)	**73%** (192/263)
Treatment B: Percutaneous nephrolithotomy	**83%** (289/350)	87% (234/270)	69% (55/80)

Charig et al.: "Comparison of treatment of renal calculi by open surgery, (...) ", British Medical Journal, 1986

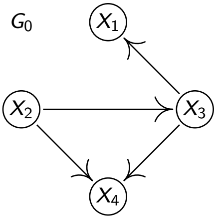


J. Mooij et al.: *Distinguishing cause from effect using observational data: methods and benchmarks*, submitted

Assume $P(X_1, \ldots, X_4)$ has been induced by

$$X_1 = f_1(X_3, N_1)$$
$$X_2 = N_2$$
$$X_3 = f_3(X_2, N_3)$$
$$X_4 = f_4(X_2, X_3, N_4)$$

- N_i jointly independent
- G_0 has no cycles

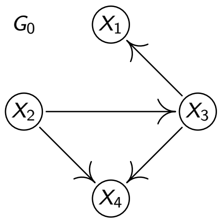


Functional causal model.
Can the DAG be recovered from $P(X_1, \ldots, X_4)$?

Assume $P(X_1, \dots, X_4)$ has been induced by

$$X_1 = f_1(X_3, N_1)$$
$$X_2 = N_2$$
$$X_3 = f_3(X_2, N_3)$$
$$X_4 = f_4(X_2, X_3, N_4)$$

- N_i jointly independent
- G_0 has no cycles

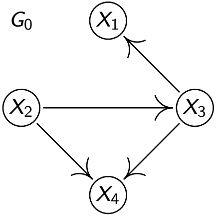


Functional causal model.
Can the DAG be recovered from $P(X_1, \dots, X_4)$? **No.**

Assume $P(X_1, \dots, X_4)$ has been induced by

$$X_1 = f_1(X_3) + N_1$$
$$X_2 = N_2$$
$$X_3 = f_3(X_2) + N_3$$
$$X_4 = f_4(X_2, X_3) + N_4$$

- $N_i \sim \mathcal{N}(0, \sigma_i^2)$ jointly independent
- G_0 has no cycles



Additive noise model with Gaussian noise.

Can the DAG be recovered from $P(X_1, \dots, X_4)$? **Yes iff f_i nonlinear.**

JP, J. Mooij, D. Janzing and B. Schölkopf: *Causal Discovery with Continuous Additive Noise Models*, JMLR 2014

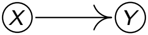
P. Bühlmann, JP, J. Ernest: *CAM: Causal add. models, high-dim. order search and penalized regr.*, Annals of Statistics 2014

S. Shimizu, P. Hoyer, A. Hyvärinen, A. Kerminen: *A linear non-Gaussian acyclic model for causal discovery*. JMLR, 2006

Consider a distribution generated by

$$Y = f(X) + N_Y$$

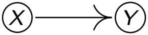
with $N_Y, X \overset{ind}{\sim} \mathcal{N}$



Consider a distribution generated by

$$Y = f(X) + N_Y$$

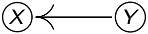
$$\text{with } N_Y, X \overset{ind}{\sim} \mathcal{N}$$



Then, if f is nonlinear, there is no

$$X = g(Y) + M_X$$

$$\text{with } M_X, Y \overset{ind}{\sim} \mathcal{N}$$

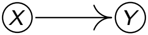


JP, J. Mooij, D. Janzing and B. Schölkopf: *Causal Discovery with Continuous Additive Noise Models, JMLR 2014*

Consider a distribution corresponding to

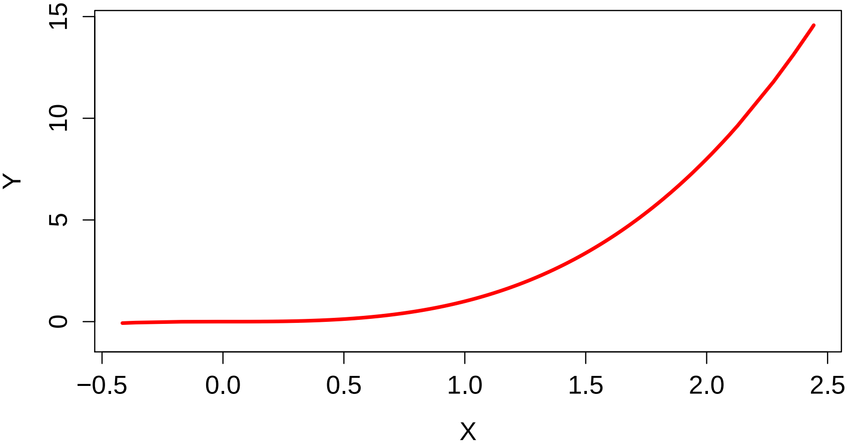
$$Y = X^3 + N_Y$$

with $N_Y, X \overset{ind}{\sim} \mathcal{N}$

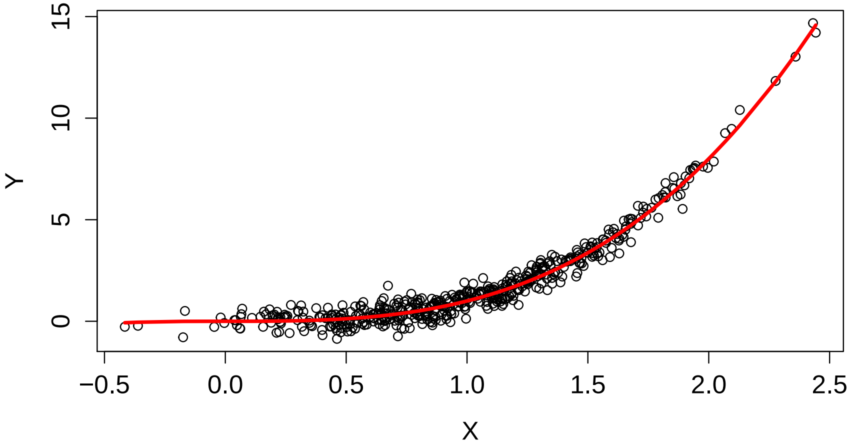


with

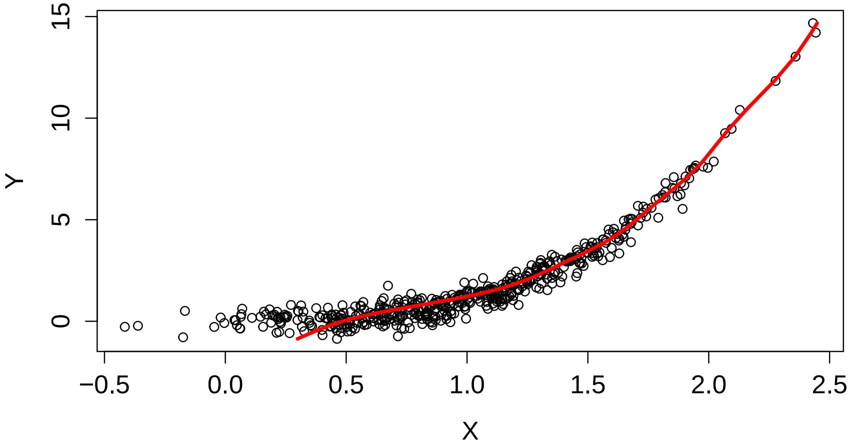
$$X \sim \mathcal{N}(1, 0.5^2)$$
$$N_Y \sim \mathcal{N}(0, 0.4^2)$$

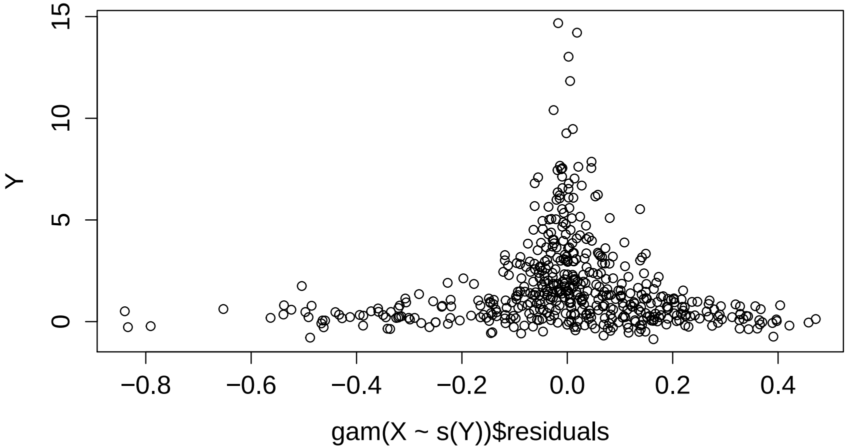


Y
X
0
5
10
15
−0.5
0.0
0.5
1.0
1.5
2.0
2.5



15
10
5
0
Y
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
X


15
10
5
0
Y
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
X


Y
0
5
10
15
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
gam(X ~ s(Y))$residuals

Surprise (under some assumptions):

$$2 \text{ variables} \Rightarrow p \text{ variables}$$

JP, J. Mooij, D. Janzing and B. Schölkopf: *Causal Discovery with Continuous Additive Noise Models, JMLR 2014*

Surprise (under some assumptions):

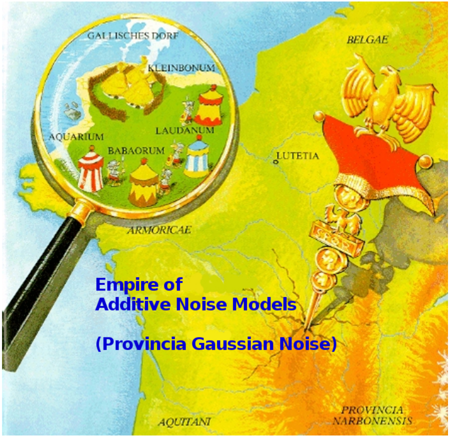
$$2 \text{ variables } \Rightarrow p \text{ variables}$$

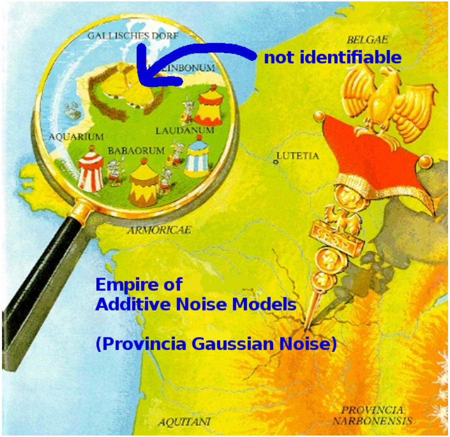
JP, J. Mooij, D. Janzing and B. Schölkopf: *Causal Discovery with Continuous Additive Noise Models, JMLR 2014*

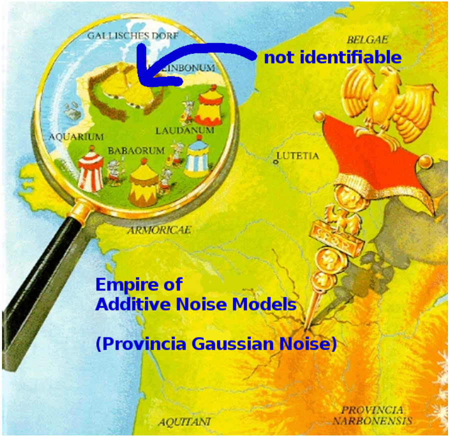
Let $P(X_1, \ldots, X_p)$ be induced by a ...

		conditions	identif.
structural equation model:	$X_i = f_i(X_{\mathbf{PA}_i}, N_i)$	-	✗
additive noise model:	$X_i = f_i(X_{\mathbf{PA}_i}) + N_i$	nonlin. fct.	✓
causal additive model:	$X_i = \sum_{k \in \mathbf{PA}_i} f_{ik}(X_k) + N_i$	nonlin. fct.	✓
linear Gaussian model:	$X_i = \sum_{k \in \mathbf{PA}_i} \beta_{ik} X_k + N_i$	linear fct.	✗

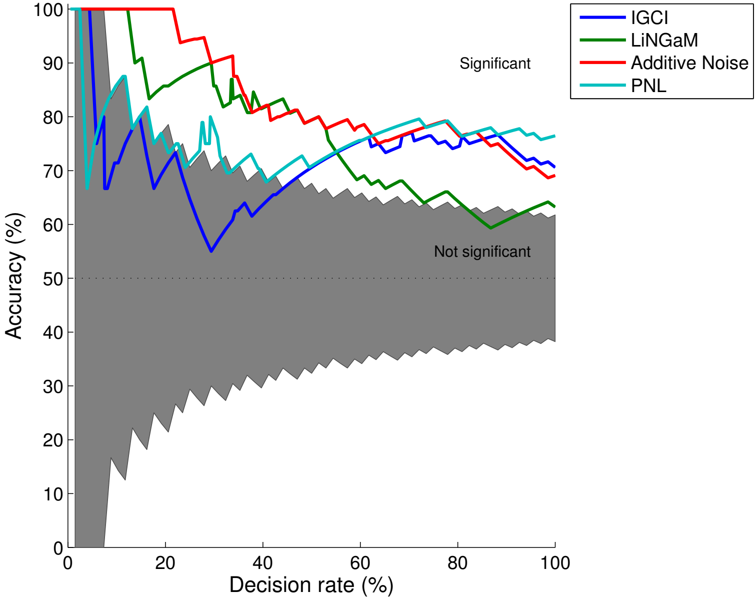
(results hold for Gaussian noise)


GALLISCHES DORF
KLEINBONUM
AQUARIUM
LAUDANUM
BABAORUM
BELGAE
LUTETIA
ARMORICAE
Empire of
Additive Noise Models
(Provincia Gaussian Noise)
AQUITANI
PROVINCIA
NARBONENSIS


GALLISCHES DORF
not identifiable
BELGAE
AQUARIUM
LAUDANUM
BABAORUM
LUTETIA
ARMORICAE
Empire of
Additive Noise Models
(Provincia Gaussian Noise)
AQUITANI
PROVINCIA
NARBONENSIS



~~GAUL~~ GAUSS
“the LINEAR”



see also

D. Lopez-Paz, K. Muandet, B. Schölkopf, I. Tolstikhin: *Towards a Learning Theory of Cause-Effect Inference*, ICML 2015

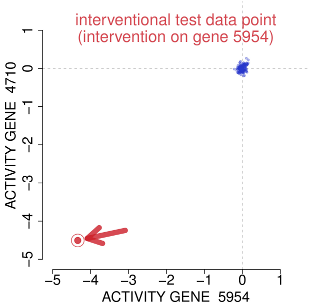
E. Sgouritsa, D. Janzing, P. Hennig, B. Schölkopf: Inf. of Cause and Effect with Unsupervised Inverse Regr., AISTATS 2015

Real data: genetic perturbation experiments for yeast (Kemmeren et al., 2014)

- $p = 6170$ genes
- $n_{obs} = 160$ wild-types
- $n_{int} = 1479$ gene deletions (targets known)

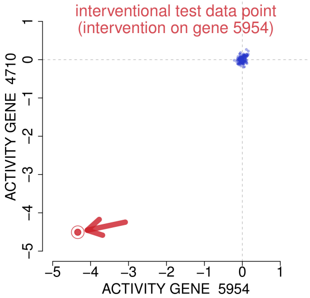
Real data: genetic perturbation experiments for yeast (Kemmeren et al., 2014)

- $p = 6170$ genes
- $n_{obs} = 160$ wild-types
- $n_{int} = 1479$ gene deletions (targets known)
- true hits: $\approx 0.1\%$ of pairs



Real data: genetic perturbation experiments for yeast (Kemmeren et al., 2014)

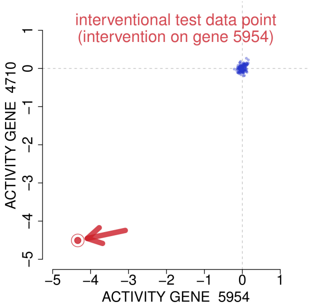
- $p = 6170$ genes
- $n_{obs} = 160$ wild-types
- $n_{int} = 1479$ gene deletions (targets known)
- true hits: $\approx 0.1\%$ of pairs



- "Invariant prediction" method: $\mathcal{E} = \{obs, int\}$

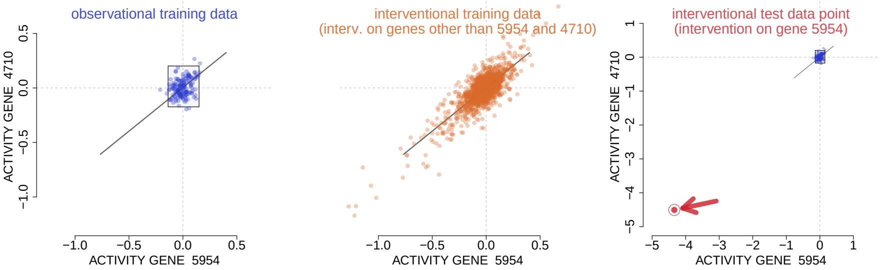
Real data: genetic perturbation experiments for yeast (Kemmeren et al., 2014)

- $p = 6170$ genes
- $n_{obs} = 160$ wild-types
- $n_{int} = 1479$ gene deletions (targets known)
- true hits: $\approx 0.1\%$ of pairs

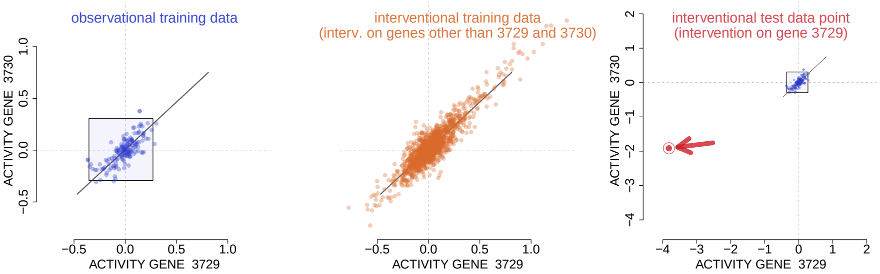


- "Invariant prediction" method: $\mathcal{E} = \{obs, int\}$

JP, P. Bühlmann, N. Meinshausen: *Causal inference using inv. pred.: identification and conf. intervals*, arXiv, 1501.01332
D. Rothenhaeusler, C. Heinze et al.: *backShift: Learning causal cyclic graphs from unknown shift interv.*, arXiv 1506.02494
M. Rojas-Carulla et al.: *A Causal Perspective on Domain Adaptation*, arXiv 1507.05333

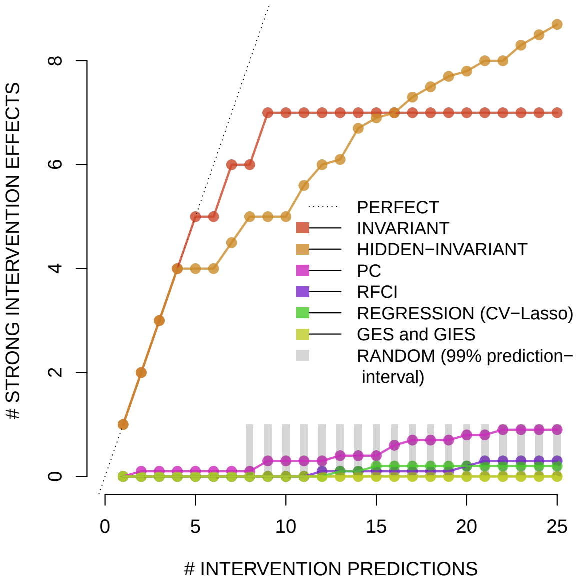


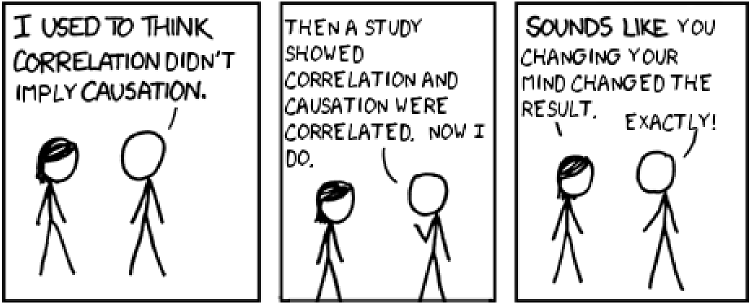
most significant pair



2nd most significant pair

3rd most significant pair


PERFECT
INVARIANT
HIDDEN−INVARIANT
PC
RFCI
REGRESSION (CV−Lasso)
GES and GIES
RANDOM (99% prediction−interval)
STRONG INTERVENTION EFFECTS
0
2
4
6
8
0
5
10
15
20
25
INTERVENTION PREDICTIONS



http://xkcdsw.com/3039

B. Watterson: *It's a magical world*, Andrews McMeel Publishing, 1996