Comment on "Causal inference using invariant prediction"

Qingyuan Zhao*†, Charles Zheng†, Trevor Hastie, and Robert Tibshirani

Department of Statistics, Stanford University

We congratulate Peters and his colleagues on this thought-provoking paper. Statistical inference of causality has been thoroughly studied in randomized experiments or observational studies but is seldom considered when data from both *observational* and *interventional* settings are available. Peters and his colleagues have made an important contribution by tackling this problem with their notion of invariant causal prediction (ICP).

At first look, ICP is a corollary of structural equation models, but we think its value might be much more substantial. Dawid [2000] noticed that causal researchers are predominately Laplacian determinists, for whom "nothing short of a functional model relating outputs to inputs will do as a description of nature". Peters and his colleagues provide an alternative approach that defines causality by *predictability* instead of *determinism*: two different concepts that are not logically connected [Hoefer, 2016]. In light of Breiman [2001]'s two cultures of statistics, determinism roughly corresponds to the data modeling culture and predictability is the spirit of Breiman's algorithmic modeling culture.

Bearing this difference in mind, Peters and his colleagues do not take a downright predictability approach in this paper. Rather, they consider two types of assumptions: invariant prediction in order to define causality and deterministic modeling assumptions such as linearity. This hybrid perspective becomes clear when comparing the assumptions in Equation (4) with those of expressions (24), (28) or (31). As a consequence, ICP can make causal discoveries only when the modeling assumptions are correct. The authors take this as a robustness property, but in our view it also limits the applicability in practice. We did not find in the paper a summary of the robustness of ICP, so we tried to outline in Table 1 the behavior of linear ICP when some of its assumptions are not met. We would welcome the authors' comments on this summary.

To test the empirical performance of ICP, we use the authors' software on a protein signaling network dataset. Sachs et al. [2005] collected a combination of observational and 9 interventional datasets to infer the causal structure of 11

*Correspondence address: qyzhao@stanford.edu

†Contributed equally.

	Issues	ICP's behavior
a)	Intervene on Y (or a missing cause)	$\bigcap_{\emptyset}$
b)	Non-linear, non-additive, and/or heteroscedastic	$\bigcap_{\emptyset}$
c)	Not enough interventions	False causal positives
d)	Small sample size	$\emptyset$
e)	Left out a confounder	$\bigcap_{\emptyset}$
f)	Left out an unconfounding predictor	okay
g)	Misspecified model or noise distribution	False positives

Table 1: Robustness properties of the ICP procedure. Under certain types of model misspecification, ICP will return a "model reject", denoted by $\cap_\emptyset$ (i.e. all subsets including the empty set are not invariant), rather than produce false positives. (a) when interventions are performed on Y, no predictor set can be invariant; (b) when the homoscedastic linear model is misspecified, the prediction rule will vary depending on the range of the predictors; (c) without enough interventions, the set of causal parents is unidentifiable, and non-causal invariant sets exist; (d) when the sample size is small, the hypothesis test for invariance has insufficient power to reject the invariance null, hence many sets are accepted as invariant; (e) if a confounder is left out, this is equivalent to intervening on Y; (f) when an uncounfounding predictor is left out, its effect is equivalent to i.i.d. noise; (g) under a misspecified noise model, the hypothesis test may not be sensitive to differences in the noise distribution, leading to low power.

proteins. Using their own method, Sachs et al. [2005] reportedly recovered 15 of the known directed arcs and discovered two new putative links (which are not shown), and missed three of the interactions which were known in the literature. In contrast, ICP only makes three causal discoveries. Among them, only one belongs to the known arcs, as we see in Figure 1. The poor performance of ICP on this dataset could be explained by the overly-restrictive linear model.

References

Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231, 2001.

A Philip Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, 2000.

Carl Hoefer. Causal determinism — The Stanford Encyclopedia of Philosophy, 2016. URL `http://plato.stanford.edu/archives/spr2016/entries/determinism-causal/`.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

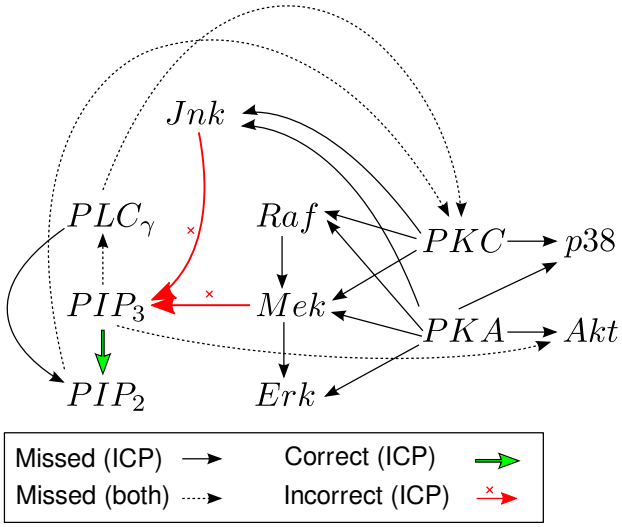

Figure 1: Application of ICP procedure to recover the protein signaling network, taking in turn each of the 11 variables as the response of interest and selecting the subset of environments in which the response was not perturbed. The invariant set for each variable can be identified as the parents of that variable in the graph. For nine of the 11 proteins, ICP rejected the model and reported no discoveries; for protein PIP2, ICP correctly identified one parent, PIP3; for protein PIP3, ICP reported Mek and Jnk as part of the invariant set, but these do not match any interactions known in the literature.