Learning causal drivers of PyroCb

Emiliano Diaz Salas Porras, Gherardo Varando, Fernando Iglesias-Suarez, Gustau Camps-Valls, Kenza Tazi, Kara D. Lamb, Duncan Watson-Parris, Ashwin Braude, Daniel Okoh, Paula Harder, Nis Meinert

Talk prepared for Learning the Earth with AI and Physics (LEAP) group

http://isp.uv.es @isp_uv_es emiliano.diaz@uv.es

Motivation:

Causal discovery in Earth System science: **no experiments possible** on global scale, but different regimes act as “natural” interventions to create **experiment like data**.

Goal:

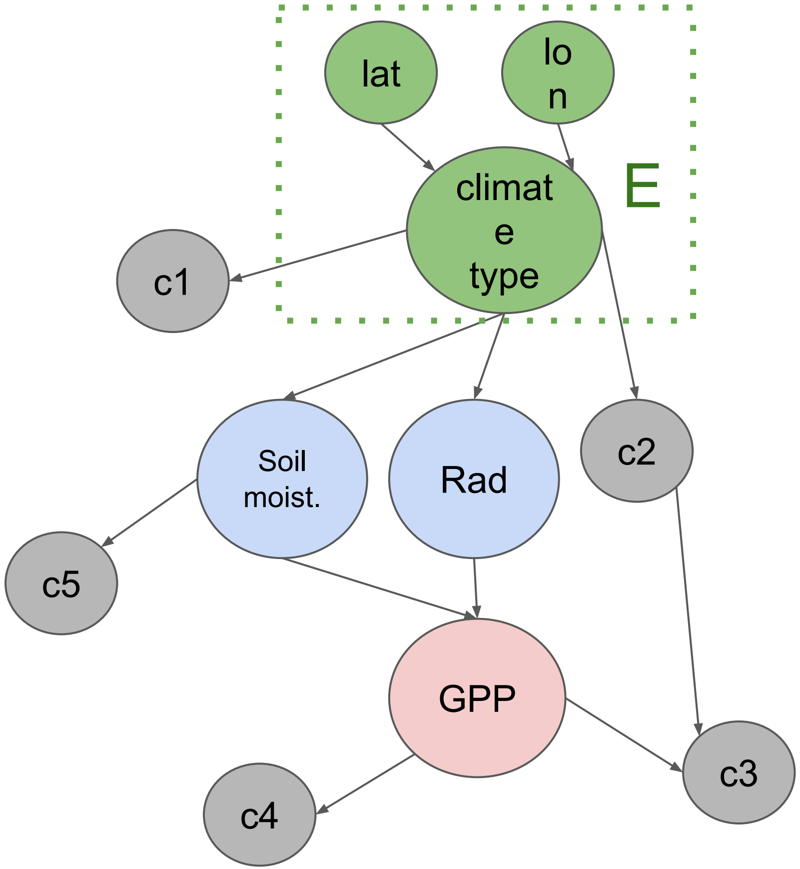
Can we use this heterogeneity to find causal drivers of phenomenon such as extreme wildfires (PyroCb) and Photosynthesis (GPP).

Use cases:

photosynthetic activity (toy model): can we separate direct causes of GPP from correlated variables (effects, shared common causes, indirect causes)?

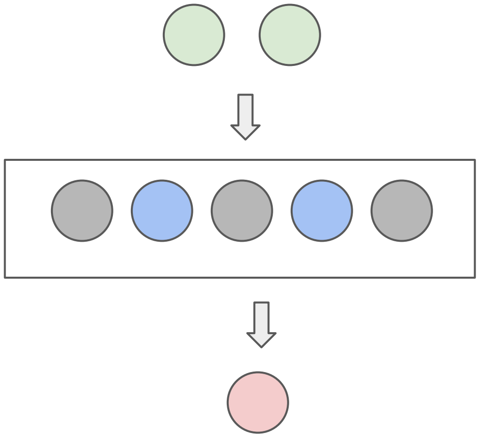
pyroCb ocurrence (real observations): why do some large fires generate pyroCb and others do not?

Toy GPP
casual model



Causal discovery in general: Learn the causal structure or DAG in th left.

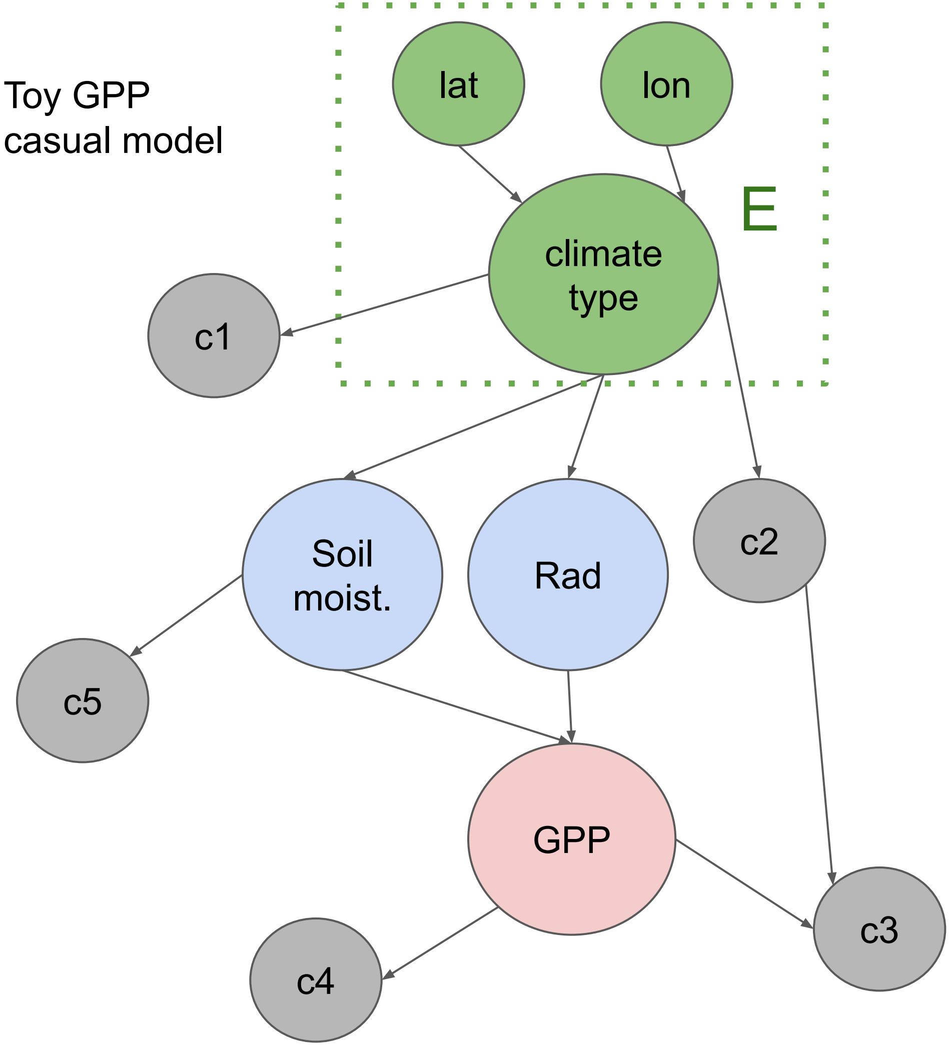
Additional context: we know the following



Two approaches.

- General Causal Discovery: find all conditional independencies
- Invariant Causal predictions: find the minimal set of variables S that satisfy $\boxed{Y \perp\!\!\!\perp E \mid X_{S^*}.}$

Toy GPP
casual model

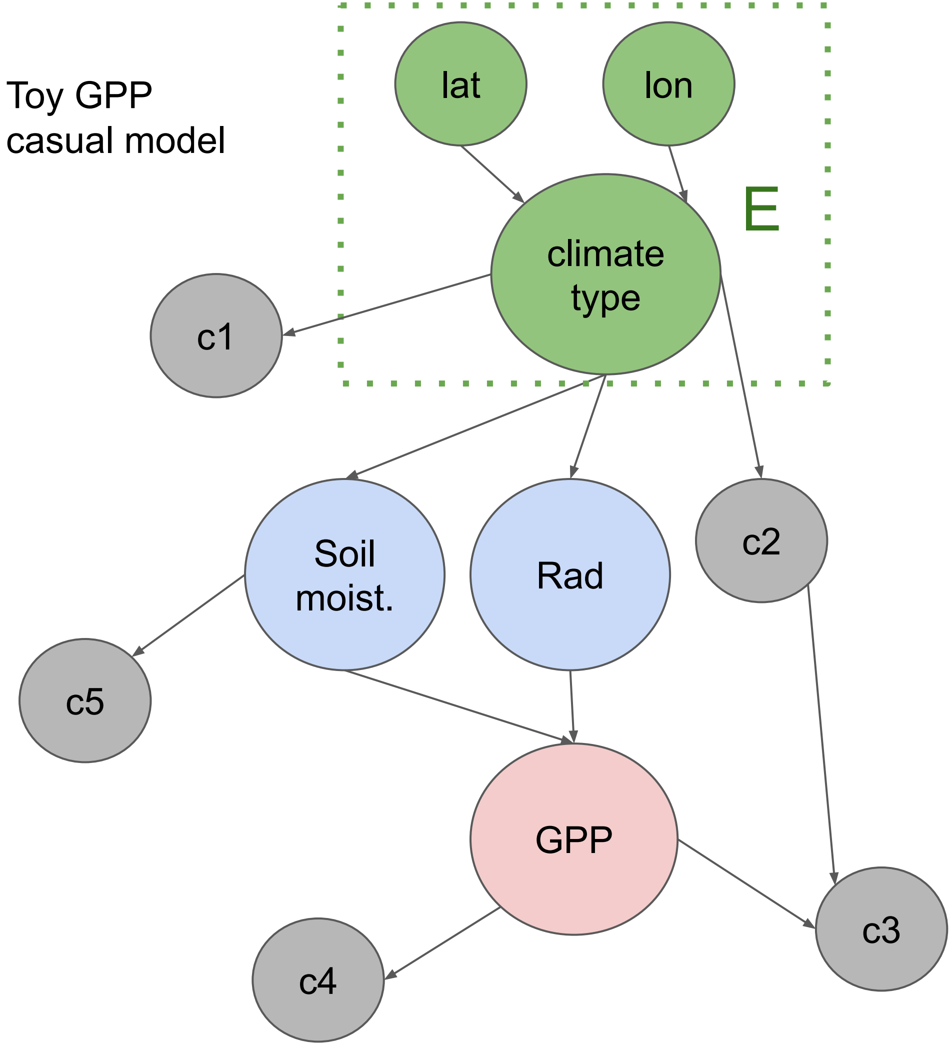


Invariant Causal Prediction (ICP) [Peters, J. et al 2016]:

Minimal conditional independence condition:

GPP independent of environment E given direct causes S*={soil moist., rad}

This is the minimal set S where this conditional independence holds



Invariant Causal Prediction (ICP) [Peters, J. et al 2016]:

Advantages of ICP:

- less ambitious than causal discovery: may concentrate on one target variable
- incorporate domain knowledge through choice of environment variable and potential causes
- Shown better performance than other conditional independence test based methods in some cases

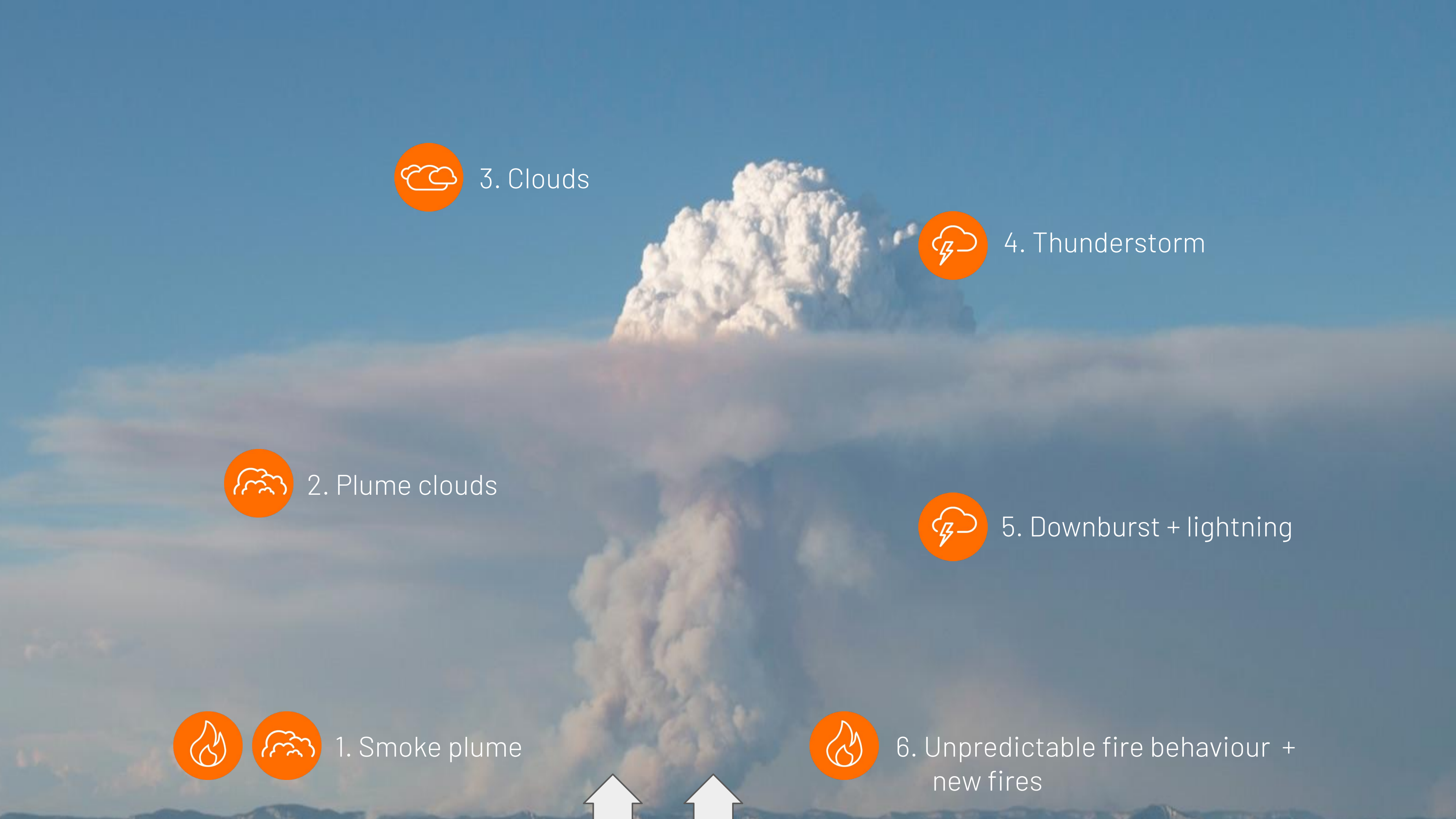
ICP algorithm

To find the causes of Y:

1. For each subset S_i of candidate predictors perform conditional independence test H_i:

$$\boxed{Y \perp\!\!\!\perp E \mid X_{S^*}.}$$

2. Take **intersection** of S_i where H_i is not rejected as causal predictors.


3. Clouds
4. Thunderstorm
2. Plume clouds
5. Downburst + lightning
1. Smoke plume
6. Unpredictable fire behaviour + new fires

28 variables total

atmospheric

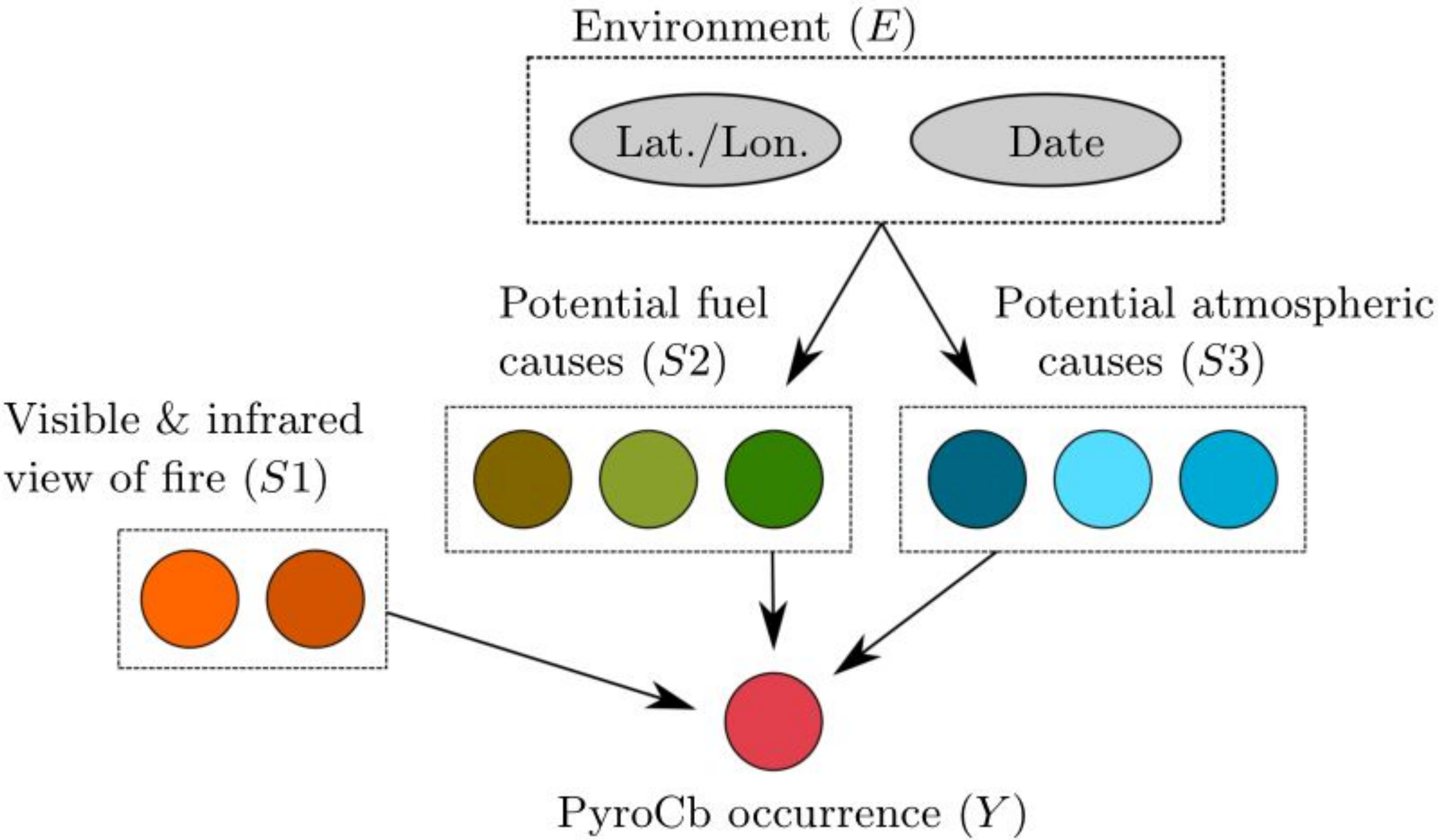
fuel

thermal

~ 100 pyroCb events comprising ~6k hourly observations in North America and Australia

Variable	Description	Sensitive to
$ch1$	0.47 μm	smoke, haze
$ch2$	0.64 μm	terrain type
$ch3$	0.86 μm	vegetation
$ch4$	3.9 μm	thermal emissions & cloud ice crystal size
$ch\{5,6\}$	$\{11.2, 13.3\}$ μm	thermal emissions & cloud opacity
$\{u,v\}$	$\{u,v\}$ comp. of wind at 250 hPa	upper-level dynamics which influence rising motion
$\{u,v\}10$	10 m $\{u,v\}$ component of wind	change in fire intensity and spread
$fg10$	10 m gusts since prev. post-processing	(same as above)
blh	boundary layer height	height of turbulent air at the surface
$cape$	convective available potential energy	energy for air to ascend into atmosphere
cin	convective inhibition	energy that will prevent air from rising
z	geopotential	energy needed for air to ascend into atmosphere as a function of altitude
$\{slhf, sshf\}$	surface {latent, sensible} heat flux	heat released or absorbed {from, neglecting} phase changes
w	surface vertical velocity	ascent speed of the plume from the wildfire
$cv\{h,l\}$	fraction of {high, low} vegetation	available fuel for the wildfire
$type\{H,L\}$	type of {high, low} vegetation	(same as above)
$r\{650,750,850\}$	rel. humidity at {650,750,850} hPa	condensation of vapour into clouds

From Tazi, K., et al 2022


Environment (E)
Lat./Lon.
Date
Potential fuel causes ($S2$)
Potential atmospheric causes ($S3$)
Visible & infrared view of fire ($S1$)
PyroCb occurrence (Y)

ICP algorithm

To find the causes of Y:

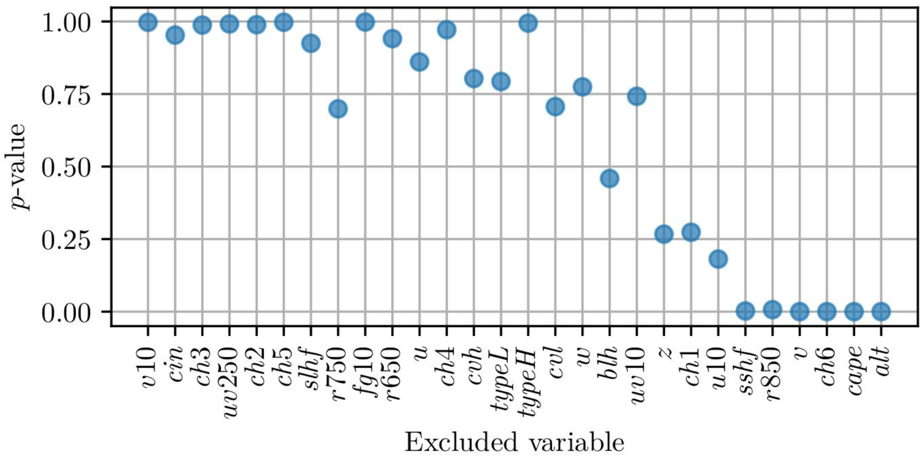
1. For each subset S_i of candidate predictors perform conditional independence test H_i:

$$\boxed{Y \perp\!\!\!\perp E \mid X_{S^*}.}$$

2. Take **intersection** of S_i where H_i is not rejected as causal predictors.

ICP: 28 variables in pyroCb dataset -> 250 million tests!

Greedy ICP: start with all candidate predictors and exclude one at a time -> 406 tests



Conditional independence test based on difference between **reduced** (Random Forest) model (excluding E) and **full** model (including E).

Use DeLong, E.R, et al (1988) test for comparing AUCs

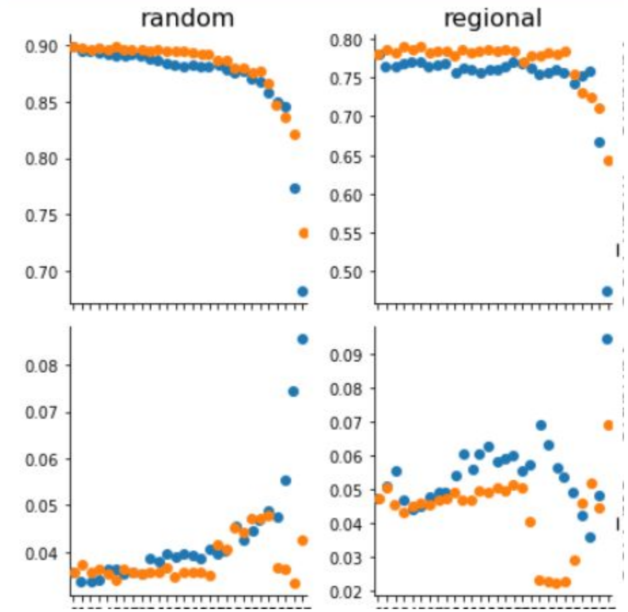
Plot shows p-value of H_i: $\boxed{Y \perp\!\!\!\perp E \mid X_{S^*}.}$ as we exclude variables with Greedy ICP

	variable	proxy for…
alt	altitude	energy needed to breach atmosphere
sshf	surface sensible heat flux	unstable boundary layer
ch6	13.3 μm reflectance	Very large and intense fire
r850	relative humidity at 850 hPa	Mid-tropospheric moisture source
v	component of wind at 250 hPa	Unstable atmosphere, conditions favorable for thunderstorms
cape	convective available potential energy	

CV type

Mean AUC

STD AUC



# of excluded variable

Greedy exclusion algorithm criterion

RF importance

greedy ICP p-value

Limitations of the ICP approach

ICP :

- number of hypothesis tests needed very large
- Dependence among predictors results in empty set inference
- Conditional independence tests hard in non-linear case

Greedy ICP

- order dependent- variables chosen for exclusion in beginning affect inference.

Invariant Causal Features

Can we use Neural Networks to:

1. learn a causal representation (get around ICP and Greedy ICP problems)

2. Learn latent environment -> identify our “quasi-experiments” (climatic type in GPP toy model)

$$\mathrm{L}(\mathrm{y},\,\mathrm{x};\,\mathrm{w}_E, w_x, \alpha, \beta) = L_1(y, x; w_E, w_x, \alpha, \beta) + \lambda ||\nabla_{w_E} L_1(y, x; w_E, w_x, \alpha, \beta)||_2^2$$

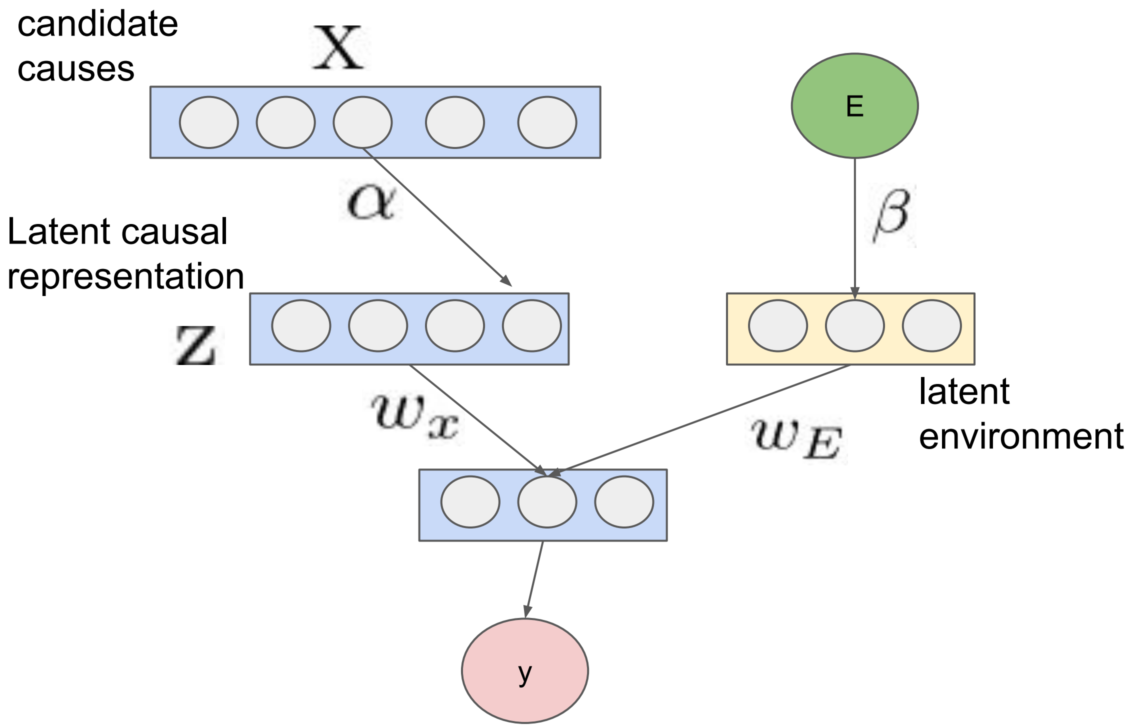
Prediction Loss:

First term the usual MSE or Cross Entropy loss

Second term in loss conditional independence (CI) proxy

- Favors representations that are invariant across E’s but not necessarily minimal/sparse!
- Use pruning of weights wrt the CI proxy in order to obtain sparsity -> intersection operation of ICP

$$L(y, x; w_E, w_x, \alpha, \beta) = L_1(y, x; w_E, w_x, \alpha, \beta) + \lambda ||\nabla_{w_E} L_1(y, x; w_E, w_x, \alpha, \beta)||_2^2$$



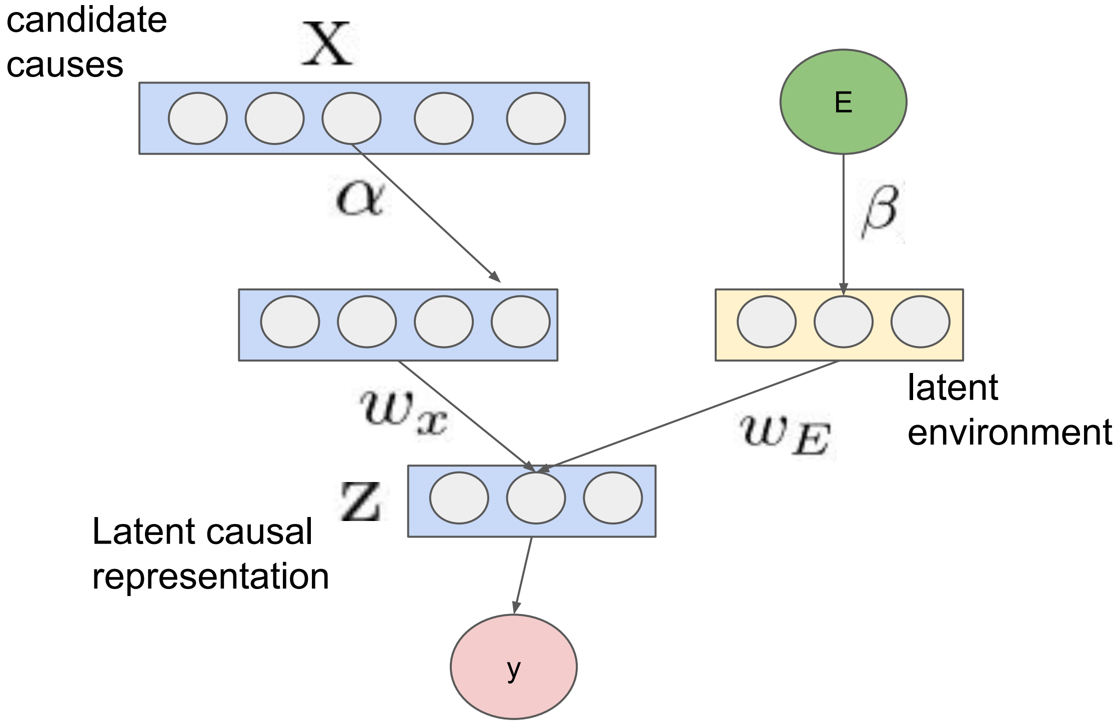
Each rectangle represents a fully connected (possibly deep) NN

We know have ONE architecture with reduced and full model (recall our conditional independence test)

In blue - “reduced model”

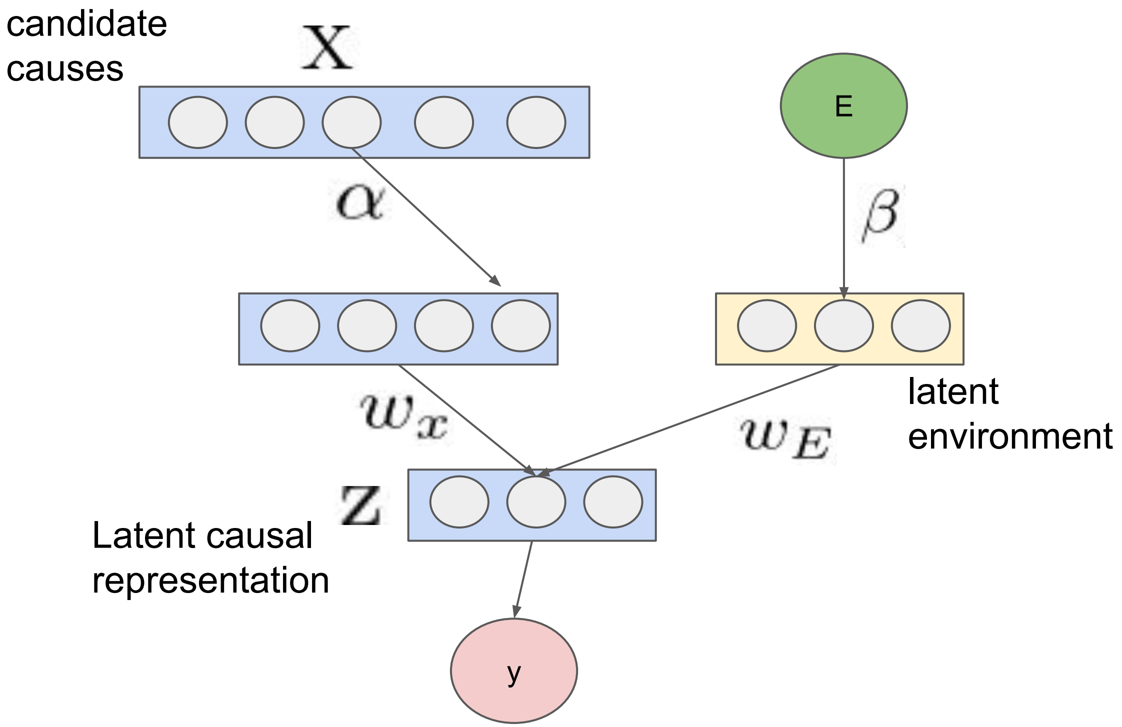
Blue and yellow - “full” model

$$L(y, x; w_E, w_x, \alpha, \beta) = L_1(y, x; w_E, w_x, \alpha, \beta) + \lambda ||\nabla_{w_E} L_1(y, x; w_E, w_x, \alpha, \beta)||_2^2$$



- Learn causal representation
- Learn latent environment

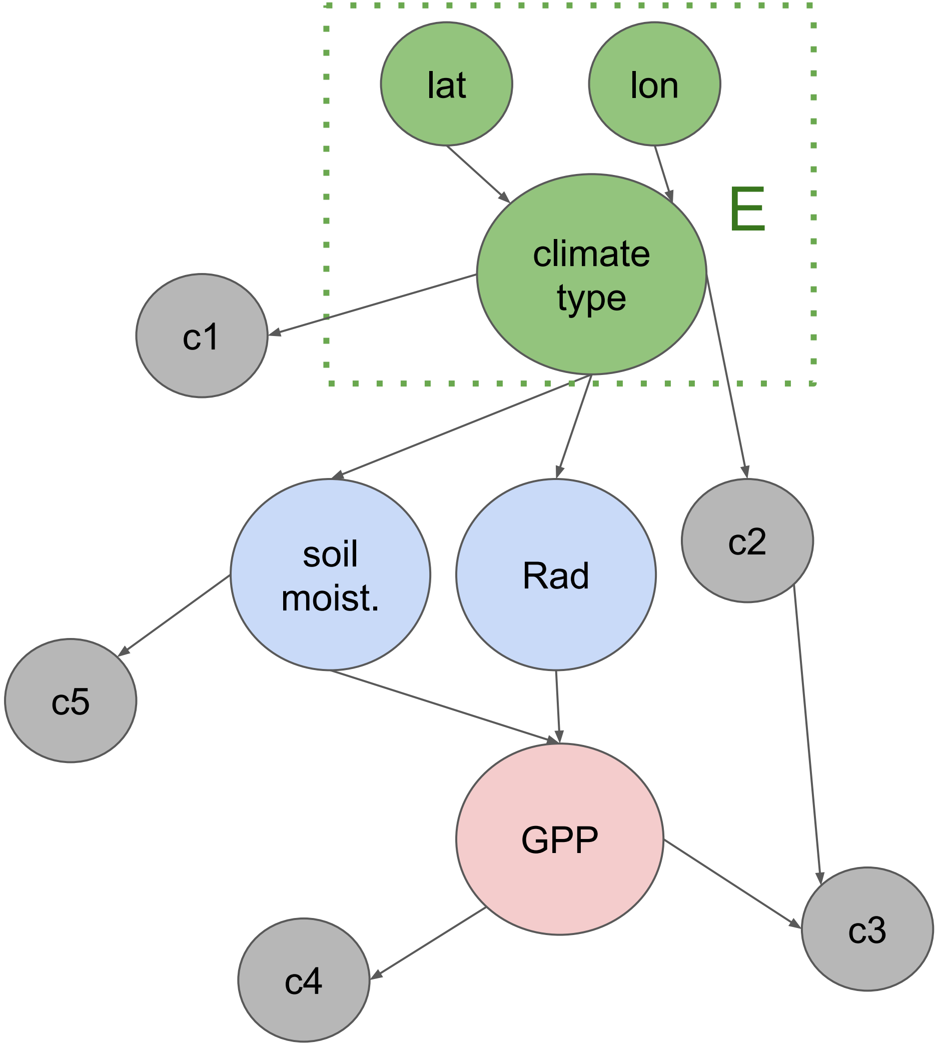
$$\mathrm{L}(\mathrm{y}, \mathrm{x}; \mathrm{w}_E, w_x, \alpha, \beta) = L_1(y, x; w_E, w_x, \alpha, \beta) + \lambda||\nabla_{w_E} L_1(y, x; w_E, w_x, \alpha, \beta)||_2^2$$



Adversarial training of reduced model and latent environment:

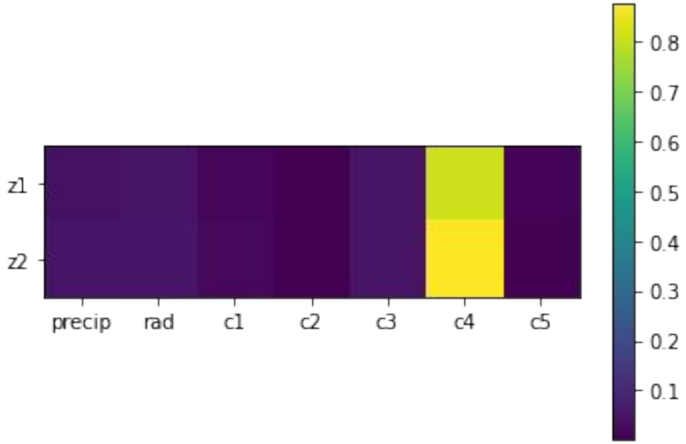
We don't want to use environment info for prediction. Use it to:

- enforce conditional independence proxy
- estimate latent environment



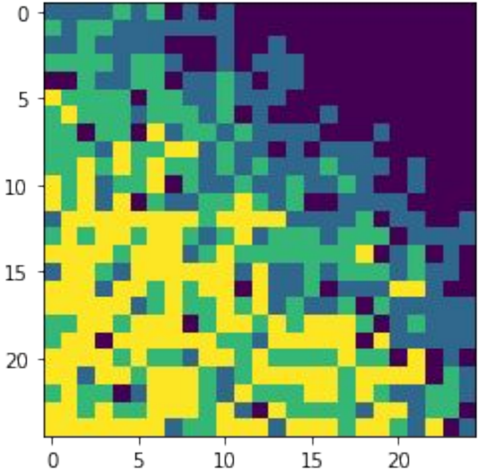
With toy GPP causal model, with **known ground truth** we test if we can learn:

1. causal representation
2. climatic type (latent environment)

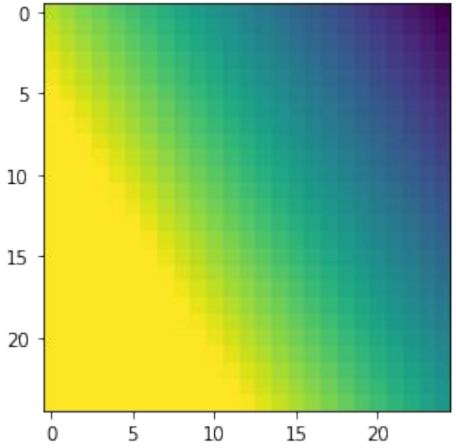


The representation is using c4 as a proxy for GPP

Ground truth climatic region



Estimated climatic region



This might be a way of investigating when environments create different conditions that can be exploited in causal discovery.

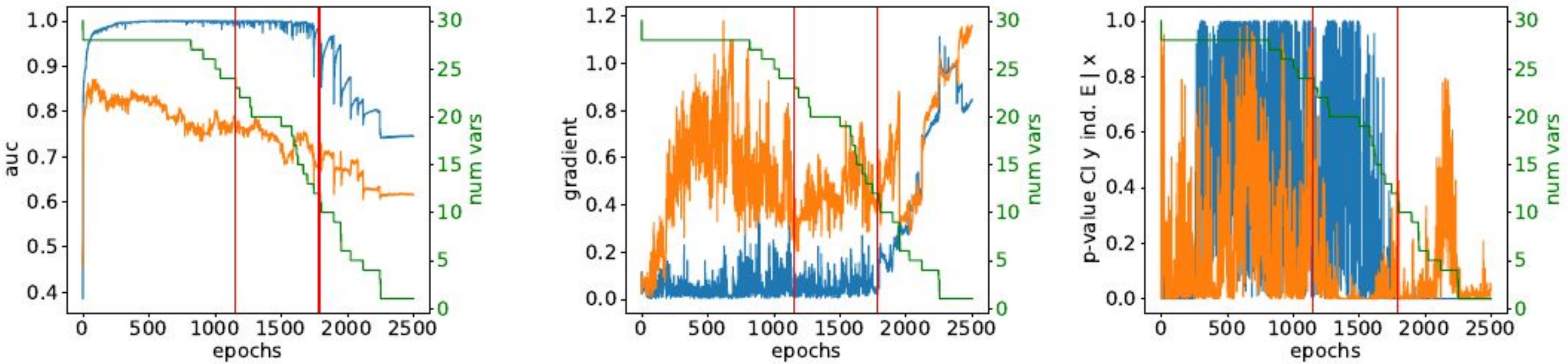


Figure 5.12: Training evolution of AIG algorithm: AUC, $||\nabla_{w_E} L_1(y, x; W)||_2^2$ and the p-value corresponding to the conditional independence test $y \perp\!\!\!\perp E|z$ for both train (blue) and test (orange) sets. In green the number of variables that have not been pruned is shown. Vertical red lines indicate two interesting causal representations to be further analyzed.

After 1800 epochs (marked in red) we maintain a test AUC of 68% with 12 variables used and test p-value of 0.59: we have a sparse model which satisfies causal CI condition.

Large test gradient may indicate we have overfit the environment variables.

Table 5.2: Variable exclusion sequence for greedy ICP and AIG.

Method	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ICP	v10	cin	ch3	uv250	ch2	ch5	slhf	r750	fg10	r650	u	ch4	cvh	typeL
AIG	ch5	alt	u10	typeH	ch4	cvh	sshf	u	ch1	v10	ch2	w	z	cvl
Method	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
ICP	typeH	cvl	w	blh	uv10	z	ch1	u10	sshf	r850	v	ch6	cape	alt
AIG	slhf	r850	typeL	ch6	fg10	r750	r650	uv10	uv250	v	blh	ch3	cin	cape

The AIG algorithm results put more emphasis variables such as cape, cin and blh which describe the instability of the atmosphere.

Altitude not considered as important but vegetation (ch3) is.

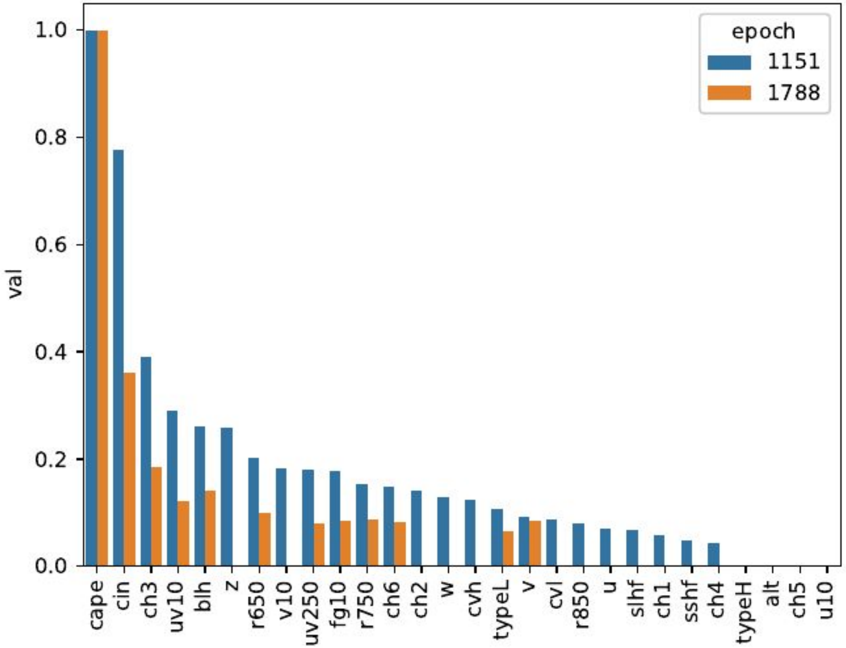


Figure 5.13: Causal representation found at two stages of training indicated in Figure 5.12. The magnitude of the bars represents the aggregated and normalized partial derivative of the loss function with respect to each input.

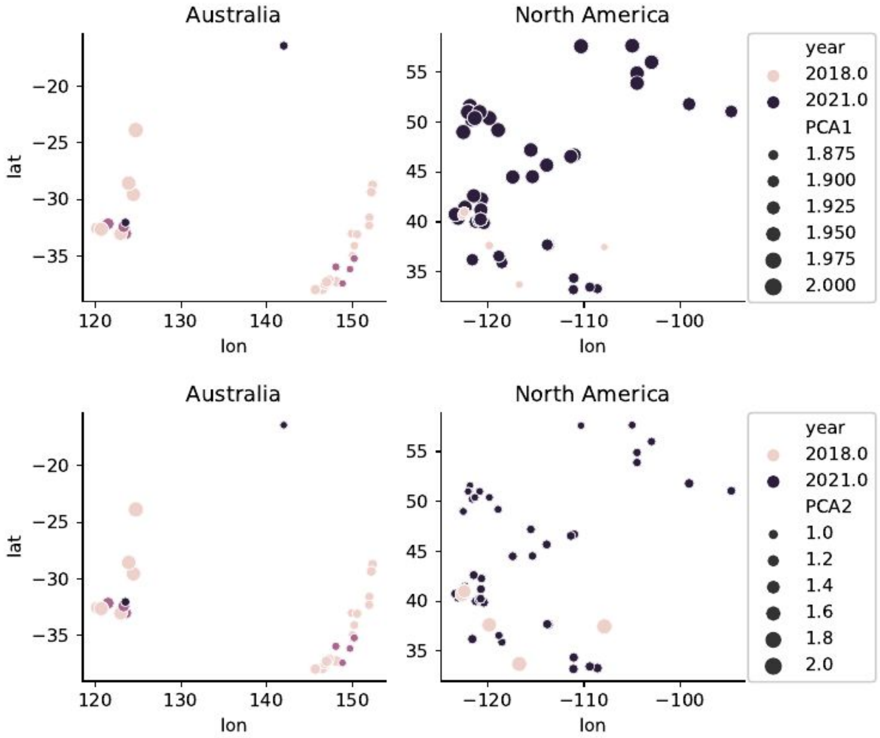


Figure 5.14: Representation of Environment. The first two PCA components of the environment representation z_E are shown using the size of the points. These represent more than 99% of the variance in z_E. The plot also illustrates the dependence of these components on the raw environment variables, longitude (x-axis), latitude (y-axis) and julian date (color of points).

Australia-North America distinction as an important distinction in terms of learning a stable causal representation

There may be some overfitting in the environment representation since, for example, in the south west of North America there appear to be a few pyroCb events that are assigned a much different representation than their surrounding neighbours (3 small dots in lighter color in top plot).

Take aways:

1. ICP unfeasible when large number of candidate predictors.

2. Greedy ICP finds a plausible set of causes for pyroCb but inference is unstable

3. Unclear if NN are effective in finding causal representation but may help to identify natural interventions which could help in causal discovery .

Next Steps:

1. Can we get NN to learn correct causal representation.
2. Can we use learnt environment in causal discovery with mixed data

Acknowledgements

This project was first developed as part of the Frontier Development Lab (FDL) Aerosols-Europe 2022 challenge

European Research Council (ERC) Synergy Grant “Understanding and Modelling the Earth System with Machine Learning (USMILE)” has provided funding for research subsequently.

References

Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society Series B, 78(5):947–1012, 2016. URL https://EconPapers.repec.org/RePEc:bla:jorssb:v:78:y:2016:i:5:p:947-1012.

Kenza Tazi, Emiliano Díaz Salas-Porras, Ashwin Braude, Daniel Okoh, Kara Lamb, Duncan Watson-Parris, Paula Harder, and Nis Meinert. Pyrocast: a machine learning pipeline to forecast pyrocumulonimbus (pyrocb) clouds. In NeurIPS 2022 Workshop-Tackling Climate Change with Machine Learning, 2022.