Causal discovery and deep learning to improve convection in climate models

Fernando Iglesias-Suarez¹, Veronika Eyring¹,², Pierre Gentine³,⁴, Tom Beucler⁵, Michael Pritchard⁶, Jakob Runge⁷,⁸, and Breixo Solino-Fernandez¹

¹Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institute of Atmospheric Physics, Oberpfaffenhofen, Germany
²University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany
³Department of Earth and Environmental Engineering, Columbia University, New York, USA
⁴Earth Institute and Data Science Institute, Columbia University, New York, USA
⁵University of Lausanne, Institute of Earth Surface Dynamics, Lausanne, Switzerland
⁶University of California, Department of Earth System Science, Irvine, USA
⁷German Aerospace Center, Institute of Data Science, Jena, Germany
⁸TU Berlin, Berlin, Germany
1. Motivation

2. Methods

3. Results:
   a) Causal input-output links
   b) Causally-informed neural nets (NNs) performance
   c) Optimization

4. Summary & key messages
1. Motivation
Climate models exhibit large uncertainties in their projections.

Global warming projections

- Historical 44
- SSP1-2.6
- SSP2-4.5
- SSP5-8.5

Large Uncertainties ➤ Clouds

CMIP6

Tebaldi et al., ESD (2021)

Super-parameterized Model

CAM: state fields
SP: parameterizations (physics)
Deep learning to represent subgrid processes in SPCAM

SPCAM ➔ Artificial Neural Network (NN) ➔ NNCAM

Idealized aquaplanet simulations with zonally symmetric SSTs

Artificial Neural Network (NN) learning subgrid processes as represented by the SP component

DL can be used to capture many advantages of CRMs.

Instabilities in the coupled system (NNCAM) found under a number of setups.

Solved via “brute force”, i.e., deepening the neural network.

➢ Lack of understanding of the fields driving subgrid scale processes

Zonal-mean temperature averaged over 5-year period.
From Fig. S1 in Rasp et al. (2018)

Gentine et al. (2018); Rasp et al. (2018)
Causal discovery: causally-linked fields in complex systems

The aim is to remove spurious links and optimise both, the inputs and the NN architecture.

Specific goals:
- Simpler NN’s without losing performance
- Stable prognostic runs less dependent on the NN’s architecture
- Generalization capabilities
2. Methods
The causal discovery framework (PCMCI) for SPCAM

Causally-informed

SPCAM  ➔  Artificial Neural Network (NN) 

emulating subgrid processes as represented by the SP component

System

Inputs \((I_t) = (I_t^1, ..., I_t^{N=94})\)

\([T(z), Q(z), V(z), PS, Sin, H, E]\)

Outputs \((O_t) = (O_t^1, ..., O_t^{N=65})\)

\([dT/dt(z), dQ/dt (z), F_{rad}, P]\)

PCMCI (e.g., Runge et al., 2019a,b, 2020)

- Algorithm: PC1
- Conditional test: Partial correlation
- Lag-time \((\tau): 1\)
- PC-alpha: 0.01

\[ I_t^i \rightarrow O_t^j \in pa(O_t^j) \]

\[ pa'(O_t^j) = \{ I_t^i : P(I_t^i \in pa_g(O_t^j)) > \text{quantile} \} \]
Causally-informed Neural Networks (proof-of-concept)

Causally-informed

SPCAM → Artificial Neural Network (NN)
emulating subgrid processes as represented by the SP component

NN (Rasp et al., 2018)

Single-NNs

Causal-NNs

I_{N=94} → O_{N=65}

Hidden layers
Number = 9
...

Hidden layers
Number = 9
...

Hidden layers
Number = 9
...

I_{N=94} → O_{(N=65)}^{i}

I_{N=\text{causal}} → O_{(N=65)}^{i} \in \text{pa}'(O_{t}^{i})
3. Results

a) Causally-linked inputs-outputs
b) Causally-informed NNs performance
c) Optimization

Iglesias-Suarez et al. (In Prep.)
Results: a) Causally-linked Inputs-Outputs

pc-alpha=0.01; threshold=0.2
Results: a) Causally-linked Inputs

pc-alpha=0.01; threshold=0.2
Results: a) Causally-linked Inputs

Solar insolation
Causal discovery: Causally-linked Inputs to Outputs in SPCAM

Remove spurious links?

Causally-linked Inputs-to-Outputs
(avg. ~15 % inputs per output)
3. Results

a) Causally-linked inputs-outputs

b) Causally-informed NNs performance

c) Optimization

Iglesias-Suarez et al. (In Prep.)
Offline: Competitive performance with only ~15% of the inputs

SingleNN/CausalSingleNN: Arch. As in Rasp et al. (2018), but ReLu activation & 15 Epochs
Online: Competitive performance with only ~15% of the inputs

* SingleNN/CausalSingleNN: Arch. As in Rasp et al. (2018), but ReLu activation & 15 Epochs
3. Results

a) Causally-linked inputs

b) Causally-informed NNs performance

c) Optimization

Iglesias-Suarez et al. (In Prep.)
Results: c) Optimization -> Threshold (offline)

- **Threshold definition**: spatially- & quantile-based approach
- **Optimization based on the 992 hPa level**

**Condition-1**: \( R^2_{\text{CAUSALNN-thr}} \geq R^2_{\text{SINGLENN}} \rightarrow R^2_{\text{SINGLENN}} > 0 \)

**Condition-2**: \( \max(\text{thr}) \)

[Diagrams showing \( R^2 \) vs \( \frac{dT}{dt} \) for different levels]
Online: within best performing cases

From Pritchard, as per Figure 4 in Ott et al. (2020)
1. Causality goes beyond correlation & removes spurious links;

2. causally-informed NNs retain generalization capabilities without loosing performance;

3. helps optimize both, inputs and (potentially) NN architecture

Work in progress…
Detection power depends on:

- **Dimensionality** of the Conditional independence (CI) test. Conditioning on the past of other adjacent links, increases the dimensionality of the CI test (i.e., partial correlation: $\sim$ number of iterations)
  
*Note the large number of inputs ($I_t = 94$) in a highly correlated system.

- **Multiple sequential testing.** Each link is tested with multiple conditioning sets.

- **Effect size.** As the dimensionality and sequential testing increase, the effect size of the CI can get very small, such that the probability it falls below the significance threshold increases.

Imagine the PC1 algorithm misses a “true” link due to high dimensionality and a very small effect size, then this could lead to “false” positives.

**false negatives (removal of true links) -> false positives (due to missing true links)**

Causal Markov Condition:
 dependence $\Rightarrow$ connectedness

Faithfulness assumption:
 independence $\Rightarrow$ no causal link

Sufficiency
 All common causes are “observed”
Optimization -> NN architecture

Causally-informed \( \frac{dQ}{dt} [820\text{hPa}] \) (35% of the original inputs)

More simple  More complex

\[ dQ/dt \quad [820\text{hPa}] \] (35% of the original inputs)

First result: CausalNN-CAM (arch-opt) unstable!
Work in progress…
pa'(dT/dt_{92hPa}): spatially-based vs pdf-based (II)

\[
\text{pa'}(O_t)\text{-ratio} = \{ \; l_t^i : \frac{\#(l_t^i \in pa_g(O_t^i))}{N_g} > \text{threshold} \; \}
\]

Total number of causal-inputs: 13 (13.8 %)

\[
\text{pa'}\text{-ratio [0.1]} = \{
T_{\text{hPa}}[820, 859, 957, 976, 992],
Q_{\text{hPa}}[912, 992],
V_{\text{hPa}}[936, 957, 992],
\text{Sin, H, E}
\}
\]

\[
\text{pa'}(O_t)\text{-pdf} = \{ \; l_t^i : P(l_t^i \in pa_g(O_t^i)) > \text{quantile} \; \}
\]

Total number of causal-inputs: 40 (42.6 %)

\[
\text{pa'}\text{-pdf [0.55]} = \{
T_{\text{hPa}}[3, 14, 232, 691, 763, 820, 859, 887, 912, 936, 957, 976, 992],
Q_{\text{hPa}}[445, 524, 609, 691, 763, 820, 859, 887, 912, 936, 957, 976, 992],
V_{\text{hPa}}[3, 7, 14, 859, 887, 912, 936, 957, 976, 992],
PS, \text{Sin, H, E}
\}
\]
Thresholds: spatially-based vs quantile-based

pc-alpha = 0.01; threshold (spatially) = 0.1

pc-alpha = 0.01; threshold (PDF) = 0.55
Causal Inputs-Outputs links are consistent across climates

pc-$\alpha=0.01$; quantile-based[thr]=0.59
Hyperparameter tuning (SHERPA): hyperparameters & ranges

Because our reference NN is Rasp et al. (2018) and we aim to “maintain (even increase) CausalNN’s performance while attaining sparser models”, we could focus on:

**Fixed parameters** (as in Rasp et al., 2018):
- **LeakyReLU coefficient**: 0.3
- **Learning rate**: 0.001
- **Learning rate decay**: 0.58

**Hyperparameter optimization** (parameters & ranges):
- **Algorithm**: GridSearch explore all possible combinations
- **Nodes per layer**: [32, 64, 128, 256, 512, 1024, 2048]
- **Number of layers**: [1-10]

**SPCAM data set** (1-month): training-shuffled & validation
**Epochs**: 9 with early-stopping (3 epochs patience)
Online: Annual zonal-mean tendencies cross-sections

SPCAM

CausalNN-CAM (t0.59)

dT/dt [10^{-5} K/s]  

dQ/dt [10^{-8} kg/kg/s]
Online: Time-series for global tropospheric T & Q

T (90N–90S; p >= 197.9 hPa)

SPCAM (two-ensembles) CausalNN-CAM (thr-opt)

Q (90N–90S; p >= 197.9 hPa; 10–3)

Time (daily-mean)