Causal model evaluation

Katja Weigel¹,², Kevin Debeire²,³, Evgenia Galytska¹,², Soufiane Karmouche¹,², Jakob Runge³,⁴ and Veronika Eyring²,¹

¹University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany
²Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
³Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institute of Data Science, Jena, Germany
⁴Technische Universität Berlin, Institute of Software Engineering and Theoretical Computer Science, Berlin, Germany

AGCI Workshop: Exploring the frontiers in Earth system modeling with machine learning and big data
07.06.2022
Outline

- Introduction: Causal model evaluation
- Examples
  - A) Causal model evaluation for CMIP6
  - B) Arctic-Mid-latitude teleconnections
  - C) Coupled Modes of Climate Variability
Causal Model Evaluation (CME)

Example from Runge et al., Science Adv., 2019

- **Hypothesis:** tropical Pacific surface temperatures (Nino 3.4) causally affect temperatures over British Columbia (BCT), but not the other way round.
- **Ground truth:** Nino -> BCT

Runge et al., 2015

- Multi-variate
- Non-linear

**Goal:** derive causal networks from obs & models

**Data:**
- 3-day-mean sea level pressure (slp)
- CMIP5 simulations (preindustrial, historical)
- Reanalysis data (NCEP) for slp, CRU for precip

**Maximum time lag for causal links:** 30d; Nr of nodes: 50

**Time-lagged correlation**

**Causal discovery**

Nowack et al., Nat Commun., 2021

Causal fingerprinting estimates direct links among the nodes.
Causal Model Evaluation: Network comparison

Network similarity metric ($F_1$)

- Based on the existence or non-existence of links in a network relative to a given reference network (pair-wise)
- Fraction of links in B that also occur in A
- $F_1=1$: perfect model match
- $F_1=0$: worst model match

AGCI workshop concluded:
Owing to different model performances against observations and the lack of independence among models, there is now evidence that giving equal weight to each available model projection is suboptimal.

Nowack et al., Nat Commun., 2021
F1-score: network similarity metric
S-score: measure for precipitation modelling skill relative to observations (Taylor, 2001)

Potential for emergent relationships to constrain projections

Nowack et al., Nat Commun., 2021
A) Causal model evaluation for CMIP6: Multi-model climate projection from the ScenarioMIP of CMIP6

(Tebaldi, Debeire, Eyring et al., ESD, 2021)
A) Causal model evaluation for CMIP6: Pair-wise comparison of causal networks

- **Sea Level Pressure** data
- **Time period**: 1979-2014
- **20 CMIP6 models**: 3-10 ensemble members historical + 2 SSP scenarios (SSP5-8.5 and SSP2-4.5)
- **Reference datasets**: NCAR-NCEP and ERA5.

→ CMIP6 models which share the same atmospheric model have more similar causal networks

Debeire et al, in preparation
Causal model evaluation for CMIP6: F1-score vs projected precipitation change over land

- Global Land Precip.[2050-2099] - [1860-1910], Reference dataset: ERA5, Time resolution: 1 day
- A parabolic relationship between SLP causal networks F1-score and Precipitation change over land under SSP5-8.5 scenario is found (similar to Nowack et al. 2021 for the CMIP5 RCP8.5 scenario)

Debeire et al, in preparation
B) Arctic-Midlatitude teleconnections: Climate change in the Arctic

The warming is stronger in the Arctic. 
*IPCC, 2021, AR6.*
B) Arctic-Midlatitude teleconnections: ERA5 & HadISST 1979-2019

- Monthly mean data December-January-February (DJFs) but max time delay 5 months
- Data: ERA5 (Hersbach et al., 2020) and HadISST (Rayner et al., 2003), 1979-2019
- Many instantaneous links, one causal link between: Aleut-SLP and Ok-SIC
- Several links known from literature, e.g.
  - TAS is related to Ural and Sib-SLP
  - Connection between Sib-SLP and BK-SIC
  - PV is associated with NAO and U

Galytska et al, in preparation
B) Arctic-Midlatitude teleconnections: CMIP6 historical 1979-2019

- Monthly mean data, December-January-February (DJFs) but max time delay 5 months
- Data: ERA5 (Hersbach et al., 2020) and HadISST (Rayner et al., 2003), 1979-2019
- CMIP6: 19 models (Eyring et al., 2016, O’Neill et al., 2016):
  - Historical (1979-2019)

Majority of the links are detected in most of analysed CMIP6 historical simulations, except:

- NAO and vflux
- vflux and Aleut-SLP
- Aleut-SLP and Ok-SIC

Galytska et al, in preparation
B) Arctic-Midlatitude teleconnections: CMIP6 SSP5-8.5 2059-2099

- Monthly mean data, December-January-February (DJFs) but max time delay 5 months
- Data: ERA5 (Hersbach et al., 2020) and HadISST (Rayner et al., 2003), 1979-2019
- CMIP6 19 models (Eyring et al., 2016, O’Neill et al., 2016):
  - SSP5-8.5 (2059-2099)

Majority of the links are detected in most of analysed CMIP6 historical simulations, except:

- NAO and vflux
- vflux and Aleut-SLP
- Aleut-SLP and Ok-SIC

Galytska et al, in preparation
C) Coupled Modes of Climate Variability

Atlantic Multidecadal Variability (AMV):
Monthly SST anomalies averaged over the North Atlantic region [0-60N, 80W-0W] minus the global mean [60N-60S].
*Trenberth and Shea (2006)*.

ERSSTv5
1900-2014 (ANN)

Pacific Decadal Variability (PDV):
Standardized principal component (PC) timeseries associated with the leading EOF of area-weighted monthly SST anomalies over the North Pacific [20-70N, 110E-100W] minus the global mean [60N-60S].
*Mantua et al. (1997).*
C) Coupled Modes of Climate Variability: Mutual interactions between tropical Atlantic and tropical Pacific. *Meehl et al. (2021)*

Idealized schematic of mutual interactions between tropical Atlantic and tropical Pacific.

As a **response to AMV**+(-) SST anomalies: Tropical Pacific knows SST-(+) anomalies (**opposite sign**)

As a **response to PDV**+(-) SST anomalies: Tropical Atlantic knows SST+(+) anomalies (**same sign**)

Mainly through **Walker Circulation**

Contributions from: **tropical Walker Circulation**, extratropical teleconnections (PNA, PSA) associated with + precipitation and convective heating anomalies.
C) Coupled Modes of Climate Variability: Regime-oriented Causal Discovery

Reanalysis data:
ERSSTv5 (PDV, AMV) and ERA20C_ERA5 (PNA, PSA1)

- PCMCI+ to estimate causal effects from time series
- Detrended yearly time series for AMV, PNA, PDV, PSA1; 1900-2014 (115 time steps).
- Regime-oriented: Analysis on different time periods (depending on the phases of AMV and PDV)

Karmouche et al, in preparation
C) Coupled Modes of Climate Variability: Regime-oriented Causal Discovery

Reanalysis data:
ERSSTv5 (PDV, AMV) and ERA20C_ERA5 (PNA, PSA1)

CMIP6 example:
MIROC6 r20i1p1f1

Karmouche et al, in preparation
C) Coupled Modes of Climate Variability: 
F$_1$-score: Causal Network comparisons (Reanalysis vs CMIP6 LE)

Most models exhibit highest F$_1$-scores for **PDV/AMV Out of phase** (light-green box) 

Karmouche et al, in preparation
Summary

- Causal Model Evaluation is a useful supervised Machine Learning tool that goes beyond traditional methods: process oriented!
  - It can provide the potential for emergent relationships to constrain projections.
  - Similar climate models detected as ‘similar’:
  - Providing opportunities for weighting models.
- These results from Nowack et al., 2021 are confirmed for CMIP6 models
- Artic-Mid-latitude teleconnections:
  - Observed links are detected in most of CMIP6 historical as well as SSP5-8.5 future simulations.
- Coupled Modes of Climate Variability:
  - Links between AMV, PDV, PNA, and PSA: Agreement to observations comparable for most models.
  - Higher agreement, if regimes are separated: AMV and PDV out of phase/in phase. This demonstrates the importance of regime oriented analysis.
Data

- **Sea Level Pressure** data
- **Time period:** 1979-2014
- **20 CMIP6 models:** 3-10 ensemble members historical + 2 SSP scenarios (SSP5-8.5 and SSP2-4.5)
- **Reference datasets:** NCAR-NCEP and ERA5.

<table>
<thead>
<tr>
<th>CMIP6 model</th>
<th>Number of ensemble members</th>
</tr>
</thead>
<tbody>
<tr>
<td>CanESM5</td>
<td>14</td>
</tr>
<tr>
<td>IPSL-CM6A-LR</td>
<td>11</td>
</tr>
<tr>
<td>CNRM-CM6-1</td>
<td>10</td>
</tr>
<tr>
<td>MIROC6</td>
<td>10</td>
</tr>
<tr>
<td>UKESM1-0-LL</td>
<td>10</td>
</tr>
<tr>
<td>MPI-ESM1-2-LR</td>
<td>10</td>
</tr>
<tr>
<td>MIROC-ES2L</td>
<td>10</td>
</tr>
<tr>
<td>MPI-ESM1-2-HR</td>
<td>10</td>
</tr>
<tr>
<td>CNRM-ESM2-1</td>
<td>10</td>
</tr>
<tr>
<td>INM-CM5-0</td>
<td>10</td>
</tr>
<tr>
<td>MRI-ESM2-0</td>
<td>7</td>
</tr>
<tr>
<td>EC-Earth3-Veg</td>
<td>5</td>
</tr>
<tr>
<td>HadGEM3-GC31-LL</td>
<td>5</td>
</tr>
<tr>
<td>HadGEM3-GC31-MM</td>
<td>4</td>
</tr>
<tr>
<td>EC-Earth3</td>
<td>4</td>
</tr>
<tr>
<td>CESM2</td>
<td>4</td>
</tr>
<tr>
<td>BCC-ESM1</td>
<td>3</td>
</tr>
<tr>
<td>BCC-CSM2-MR</td>
<td>3</td>
</tr>
<tr>
<td>ACCESS-CM2</td>
<td>3</td>
</tr>
<tr>
<td>CESM2-WACCM</td>
<td>3</td>
</tr>
<tr>
<td>MPI-ESM1-2-HAM</td>
<td>3</td>
</tr>
<tr>
<td>KACE-1-0-G</td>
<td>3</td>
</tr>
<tr>
<td>NorESM2-LM</td>
<td>3</td>
</tr>
<tr>
<td>ACCESS-ESM1-5</td>
<td>3</td>
</tr>
</tbody>
</table>
Multi-model climate projection from the ScenarioMIP of CMIP6

(Tebaldi, Debeire, Eyring et al., ESD, 2021)

→ Larger uncertainties for the CMIP6 projected ensemble mean of precipitation over land
A) Causal model evaluation for CMIP6: Causal networks F1-score vs Precipitation over land bias (S-score)

→ Similar to Nowack et al. (2021), significant linear correlations are found on a global scale and regional scales (except for Africa and South-Asia). Better fingerprints are associated with smaller land precipitation biases.
### B) Arctic-Midlatitude teleconnections: Potential processes/actors

<table>
<thead>
<tr>
<th>Actor</th>
<th>Area</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Temperature</td>
<td>65-90°N, zm</td>
<td>TAS</td>
</tr>
<tr>
<td>Sea Ice Area Fraction</td>
<td>70-80°N, 30-105°E</td>
<td>BK-SIC</td>
</tr>
<tr>
<td></td>
<td>50-60°N, 140-160°E</td>
<td>Ok-SIC</td>
</tr>
<tr>
<td>Sea Level Pressure</td>
<td>45-70°N, 40-85°E</td>
<td>Ural-SLP</td>
</tr>
<tr>
<td></td>
<td>40-65°N, 85-120°E</td>
<td>Sib-SLP</td>
</tr>
<tr>
<td></td>
<td>45-80°N, 160-260°E</td>
<td>Aleut-SLP</td>
</tr>
<tr>
<td>Heat Flux</td>
<td>45-75°N, zm, 100hPa</td>
<td>vflux</td>
</tr>
<tr>
<td>Polar Vortex</td>
<td>65-90°N, zm, 100-10hPa</td>
<td>PV</td>
</tr>
<tr>
<td>North Atlantic Oscillation</td>
<td>20-90°N, 500hPa</td>
<td>NAO</td>
</tr>
<tr>
<td>Zonal wind</td>
<td>50-70°N, zm</td>
<td>U</td>
</tr>
</tbody>
</table>
B) Arctic-Midlatitude teleconnections: ERA5 & HadISST 1979-2019

- Monthly mean data December-January-February (DJFs) but max time delay 5 months
- Data: ERA5 (Hersbach et al., 2020) and HadISST (Rayner et al., 2003), 1979-2019

<table>
<thead>
<tr>
<th>Link</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAS and BK-SIC</td>
<td>Cohen et al., Nature Climate Change, 2020</td>
</tr>
<tr>
<td>TAS and Ural-, Sib-SLP</td>
<td>Luo et al., Journal of Climate, 2016</td>
</tr>
<tr>
<td>Ural-SLP and Sib-SLP</td>
<td>Tyrlis et al., Quarterly Journal of the Royal Met. Society, 2020</td>
</tr>
<tr>
<td>Sib-SLP and BK-SIC</td>
<td>Siew et al., Weather and Climate Dyn., 2020</td>
</tr>
<tr>
<td>vflux - PV</td>
<td>Kretschmer et al., Journal of Climate, 2016</td>
</tr>
<tr>
<td>PV and NAO, U</td>
<td>Smith et al., Nature communications, 2022</td>
</tr>
<tr>
<td>Aleut-SLP and Ok-SIC</td>
<td>Ogi et al., Journal of Climate, 2015</td>
</tr>
</tbody>
</table>
F1-score (Nowack et al., 2020)

Better agreement in DJFs
Extratropical teleconnections

Atmospheric Modes of Variability

Defined using monthly area-weighted sea level pressure (PSL) anomalies

PNA (DJF)
PSA1 (ANN)

ERA20C_ERA5 1900-2014

Pacific – North American Pattern (PNA): leading EOF in the region [20-85N, 120E-120W]

C) Coupled Modes of Climate Variability: Regime-oriented Causal Discovery

Reanalysis data:
ERSSTv5 (PDV, AMV) and ERA20C_ERA5 (PNA, PSA1)

CMIP6 example:
MIROC6 r20i1p1f1

CMIP6 example:
CanESM5 r11i1p2f1

CMIP6 example:
CanESM5 r12i1p2f1

CMIP6 example:
CanESM5 r17i1p2f1