Using Causal Discovery Algorithms to analyze and predict the Stratospheric polar vortex

Marlene Kretschmer, 21.9.2017

Dim Coumou, Jakob Runge, Jonathan Donges

THE STRATOSPHERIC POLAR VORTEX

Strong Polar Vortex:

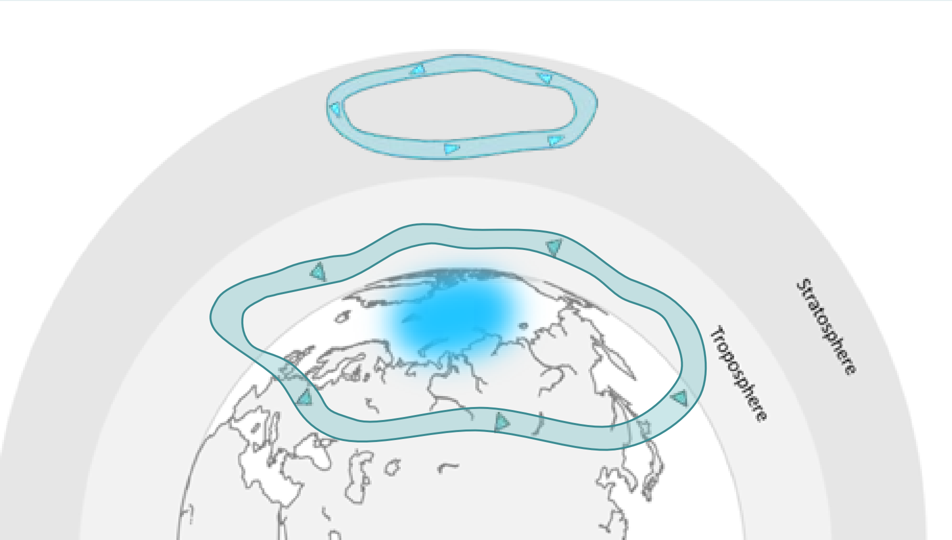
- Fast winds
- Strong, circumpolar flow

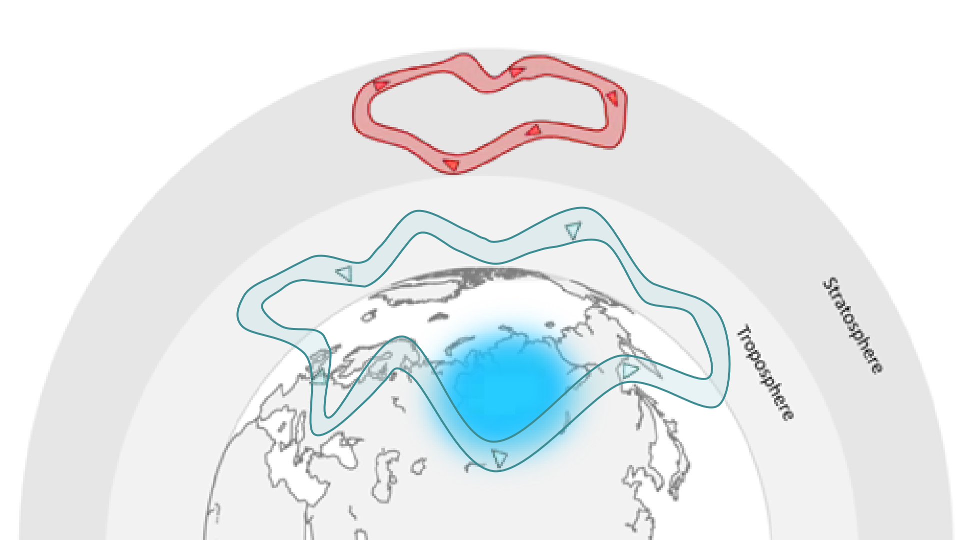
à Mild winters/+AO

Weak Polar Vortex:

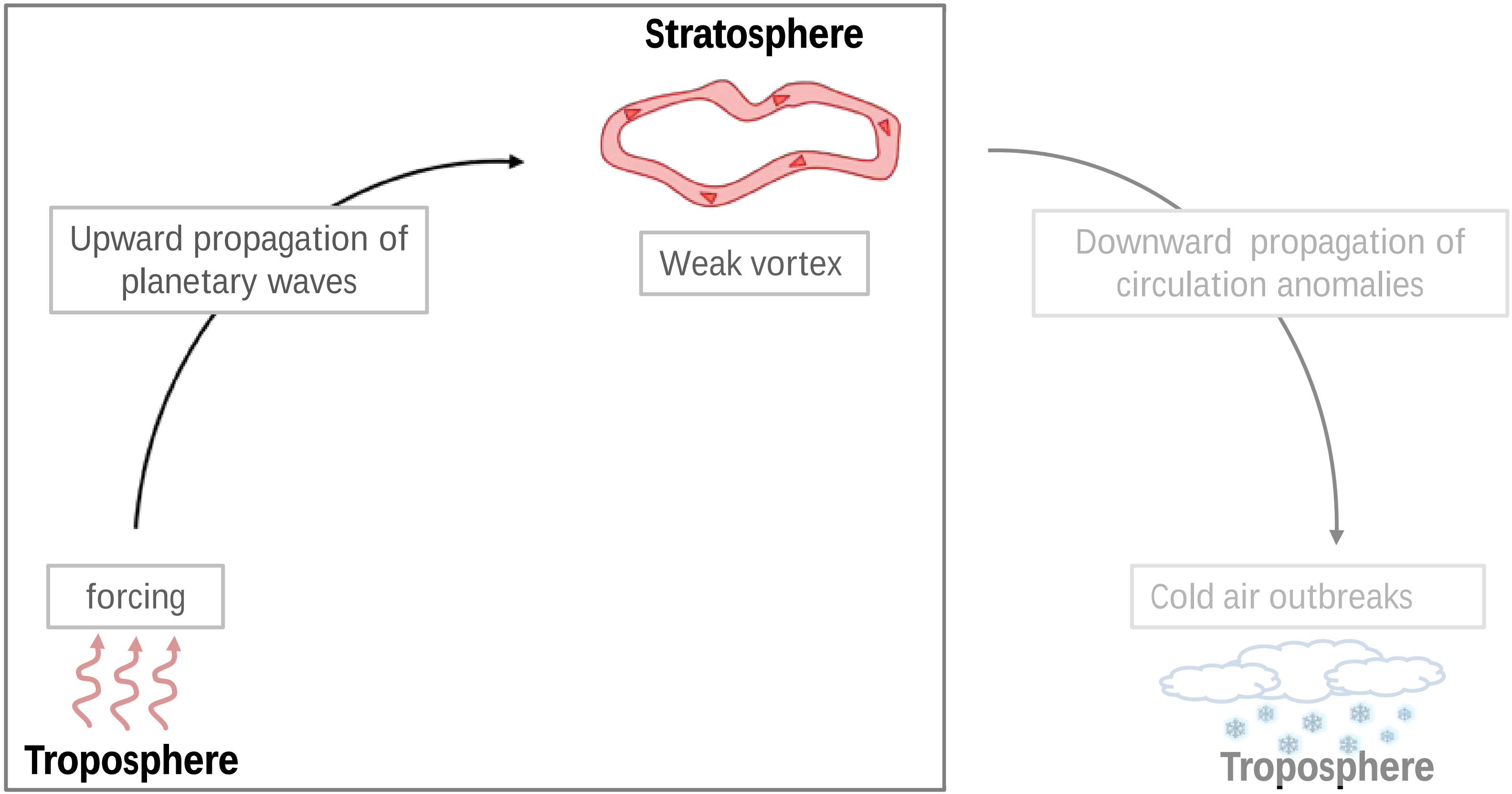
- Slow winds
- Weak, wavy flow

à Cold winters/-AO





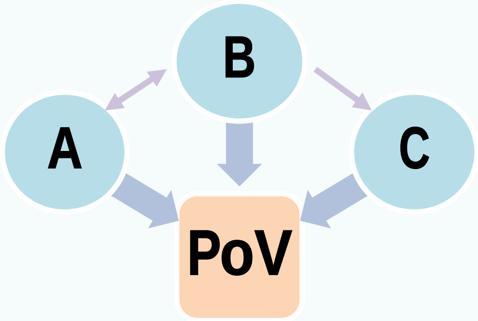
TROPOSPHERE - STRATOSPHERE - TROPOSPHERE COUPLING



OUTLOOK

PART I

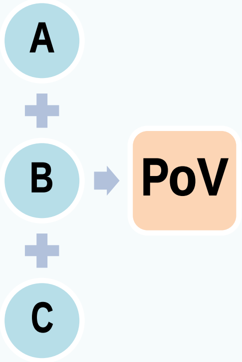
Causal effect networks for hypothesis testing



(Arctic) drivers of the polar vortex

PART II

Response-guided causal precursor detection for predictions

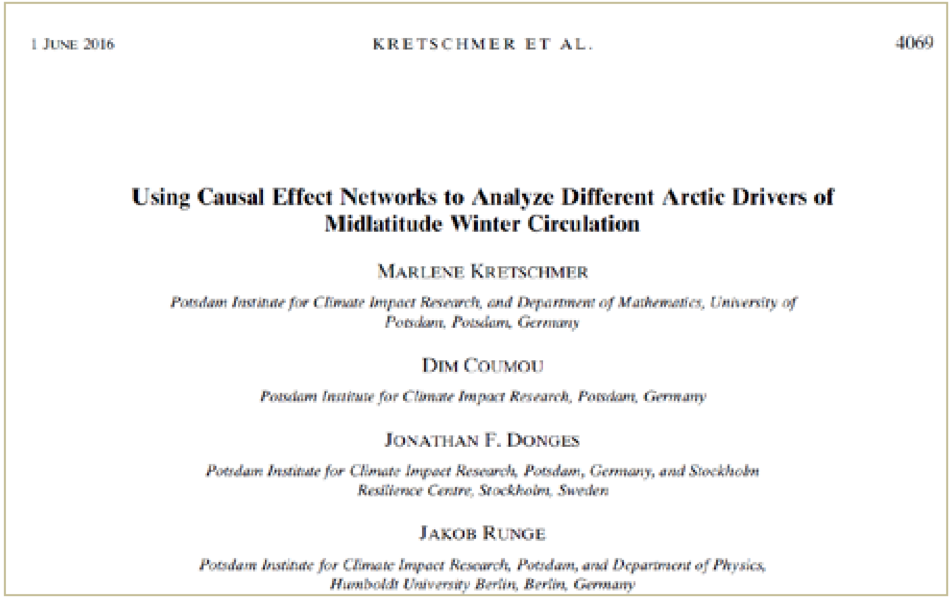


Predict polar vortex variability

Common tasks when studying teleconnections

PART I

Hypothesis testing with Causal Effect Networks



Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation

MARLENE KRETSCHMER

Potsdam Institute for Climate Impact Research, and Department of Mathematics, University of Potsdam, Potsdam, Germany

DIM COUMOU

Potsdam Institute for Climate Impact Research, Potsdam, Germany

JONATHAN F. DONGES

Potsdam Institute for Climate Impact Research, Potsdam, Germany, and Stockholm Resilience Centre, Stockholm, Sweden

JAKOB RUNGE

Potsdam Institute for Climate Impact Research, Potsdam, and Department of Physics, Humboldt University Berlin, Berlin, Germany

Hypothesis: Arctic drivers of Polar vortex variability

Low sea-ice concentrations in fall in the Barents-Kara Seas
(e.g. Kim et al, *Nat Comm*, 2014)

→

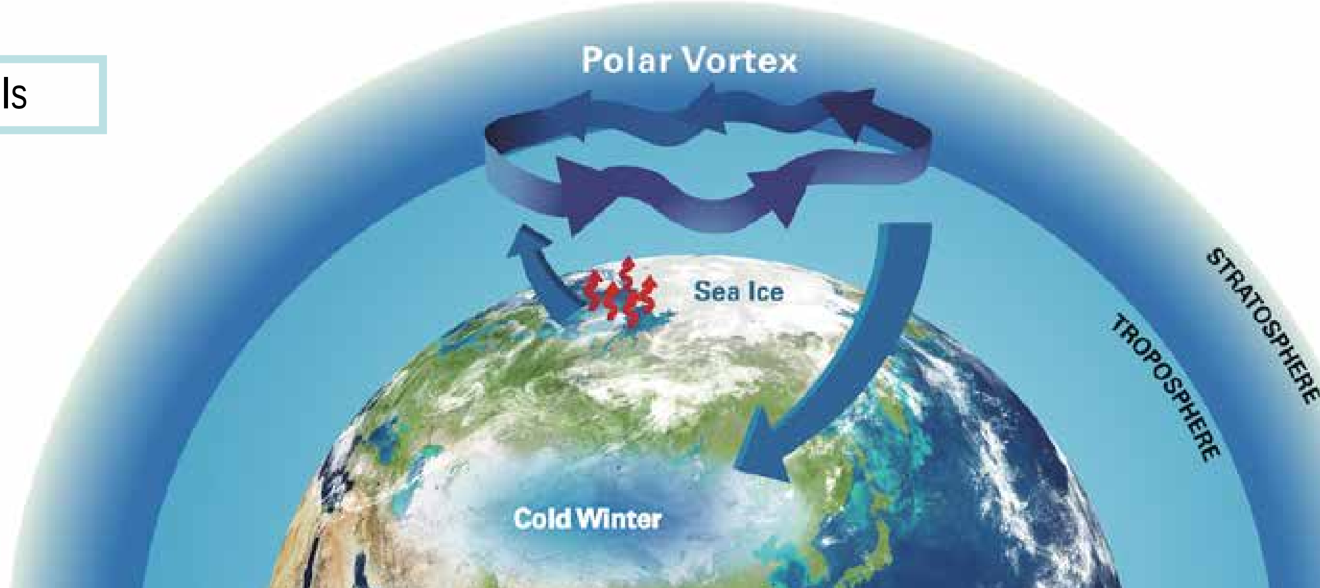
Enhanced snow cover in fall
(e.g. Cohen et al, *J Clim*, 2014)

↓

lead to higher surface pressures, upward wave propagation, weakening of polar vortex

↓

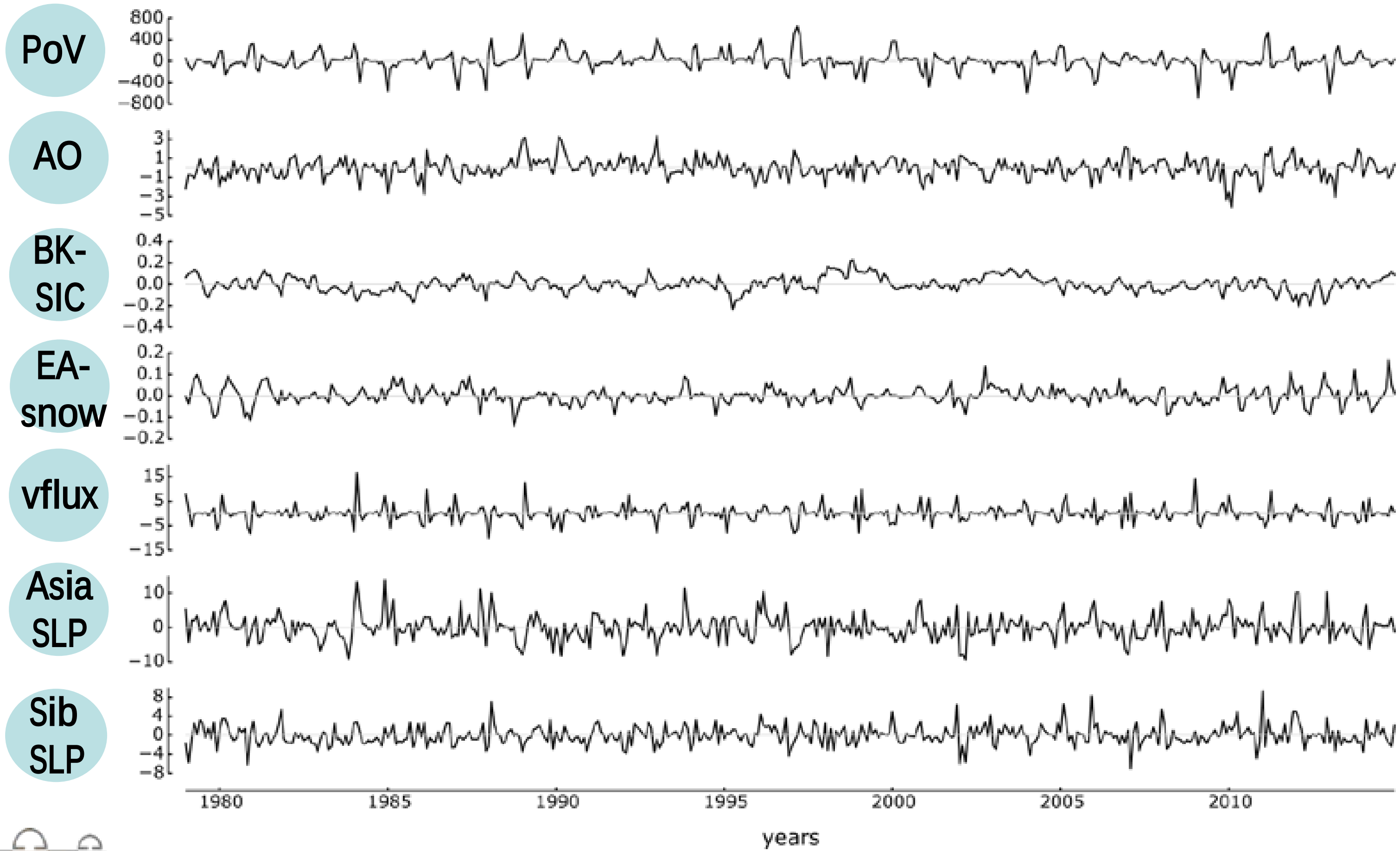
favor -AO and mid-lat cold spells



kretschmer@pik-potsdam.de

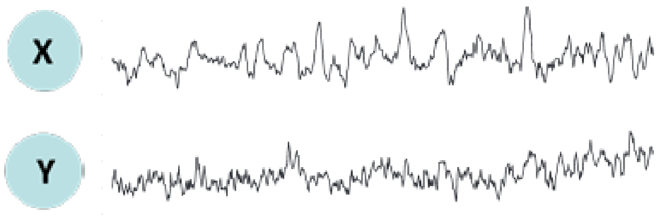
Data Selection: choose variables for each "*Actor*"

Abbreviation	Actor	Variable/Unit	Region (Level)
BK-SIC	Barents Kara sea ice	Sea ice area fraction	70 °- 80°N, 30°- 105°E
EA-snow	Eurasia snow cover	snow covered area fraction	40° - 80°N, 30°-180°E
Sib-SLP	Siberian High	Sea level pressure	40° - 65°N, 85° - 120°E
Ural-SLP	Ural Mountains sea level pressure	Sea level pressure	45° - 70°N, 40° - 85°E
v-flux	Vertical wave propagation	Pole-ward eddy heat flux v*T*	45° - 75°N (100 mb)
PoV	Polar Vortex	Geopotential height in m	65° - 90°N (10 - 100 mb)
AO	Arctic Oscillation Index	Geopotential height	20° - 90°N (1000 mb)



Monthly anomalies 1979-2014

Problems with Cross-Correlation

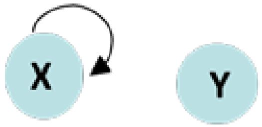


Assume X_{t-1} and Y_t correlate strongly:

e.g. $\rho(X_{t-1}, Y_t) = 0.7$

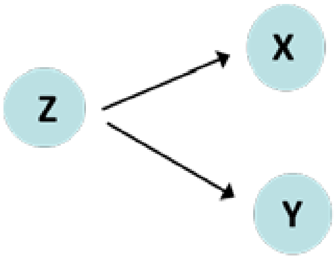
Does this mean X causes Y?

Auto-correlation



Indirect link

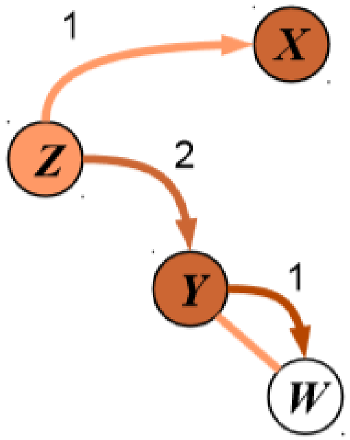
Common driver



Multi-variate approach is required

Causal Effect Network (CEN)

Step 1 Find causal links	Estimate the parent processes for each actor: Exclude spurious correlations due to auto-correlation, common drivers, indirect links
Step 2 Estimate link strength	Calculate the link strength via multiple linear regression



STEP 1: CALCULATE THE PARENT PROCESSES FOR POV

- **Which actors are significantly correlated with the PoV index?**

$P_0 = \{\text{v-flux}_{t-1}, \text{PoV}_{t-1}, \text{Ural-SLP}_{t-1}, \text{Ural-SLP}_{t-2}, \text{AO}_{t-1}, \boxed{\text{EA-snow}_{t-1}}\}$

<u>Test Hypothesis</u>

Does Eurasian snow cover influence the polar vortex with a lag of one month?

- $\rho(\text{EA-snow}_{t-1}, \text{PoV}_t) = -0.3$ $(a < 0.01)$
- $\rho(\text{EA-snow}_{t-1}, \text{PoV}_t \mid \text{v-flux}_{t-1}) = -0.1$ $(a > 0.01)$

$\rho(X, Y \mid Z)$ = partial correlation of X and Y given Z

- EA-snow and PoV are **conditionally independent**

Step 1: Calculate the parent processes for PoV

$P_0 = \{$v-flux$_{t-1}$, PoV$_{t-1}$, Ural-SLP$_{t-1}$, Ural-SLP$_{t-2}$, AO$_{t-1}$, EA-snow$_{t-1}\}$

$P_1 = \{\boxed{\text{v-flux}_{t-1}}$, PoV$_{t-1}$, Ural-SLP$_{t-1}\}$

<u>Test Hypothesis</u>

Does poleward heat-flux influence the polar vortex with a lag of one month?

- ρ(v-flux$_{t-1}$, PoV$_t$) = - 0.7 (a < 0.01)
- ρ(v-flux$_{t-1}$, PoV$_t$ | Ural-SLP$_{t-1}$) = - 0.6 (a < 0.01)
- ρ(v-flux$_{t-1}$, PoV$_t$ | PoV$_{t-1}$) = - 0.6 (a < 0.01)
- ρ(v-flux$_{t-1}$, PoV$_t$ | PoV$_{t-1}$, Ural-SLP$_{t-1}$) = - 0.6 (a < 0.01)
- v-flux and PoV are **conditionally dependent**

CEN Algorithm

Step 1
Find causal links

$P_0 = \{\text{v-flux}_{t-1}, \text{PoV}_{t-1}, \text{Ural-SLP}_{t-1}, \text{Ural-SLP}_{t-2}, \text{AO}_{t-1}, \text{EA-snow}_{t-1}\}$

$P_1 = \{\text{v-flux}_{t-1}, \text{PoV}_{t-1}, \text{Ural-SLP}_{t-1}\}$

$\mathcal{P}_{\text{PoV}} = \{\text{v-flux}_{t-1}, \text{PoV}_{t-1}, \text{Ural-SLP}_{t-1}\}$

Step 2
Estimate link strength

Linear regression:

$\text{PoV}^*_t = \beta_0 + \beta_1 \text{v-flux}^*_{t-1} + \beta_2 \text{Ural-SLP}^*_{t-1} + \beta_3 \text{PoV}^*_{t-1} + \varepsilon$

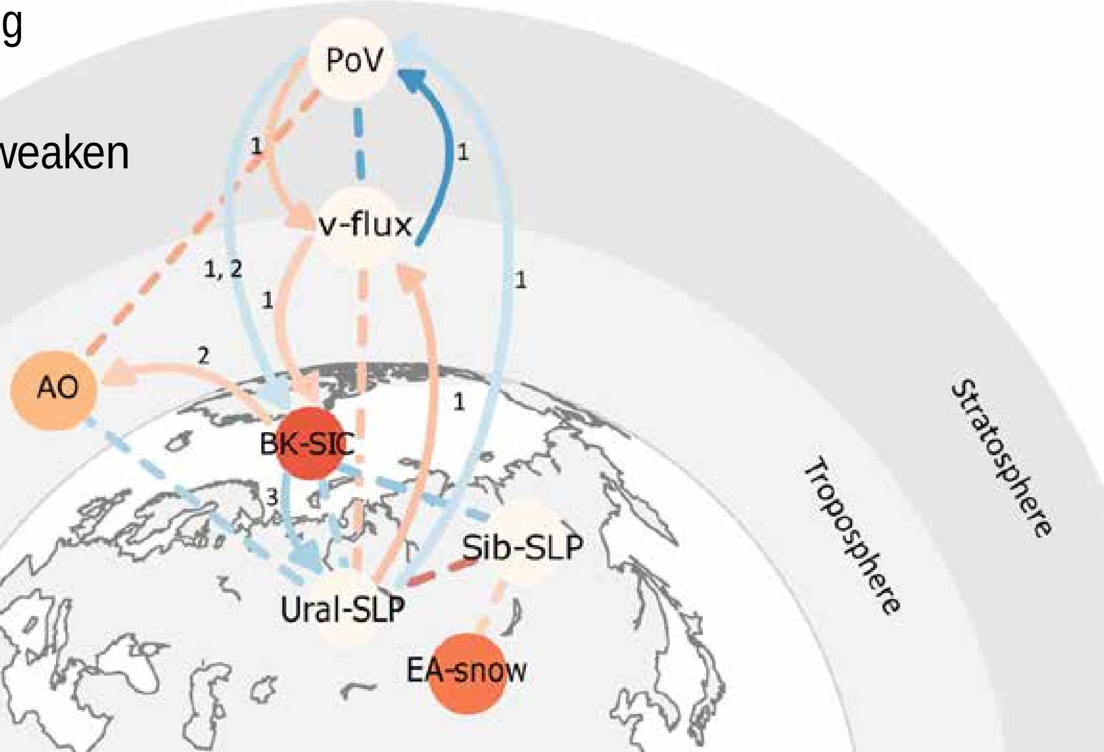
(We account for multiple-testing)

Repeat Step 1+2 for each actor

$\mathcal{P} = \{\mathcal{P}_{\text{AO}}, \mathcal{P}_{\text{BK-SIC}}, \mathcal{P}_{\text{EA-snow}}, \mathcal{P}_{\text{v-flux}}, \mathcal{P}_{\text{PoV}}, \mathcal{P}_{\text{Sib-SLP}}, \mathcal{P}_{\text{Ural-SLP}}\}$

RESULTS: CAUSAL EFFECT NETWORK

- CEN constructed for winter (DJF)
- Troposphere-Stratosphere coupling robustly found
- Low sea-ice conditions in fall can weaken polar vortex in winter
- Role of snow cover less robust



PART II

Response-guided causal precursor detection for predictions

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER

10.1002/2017GL074696

Key Points:
- A new scheme to detect causal precursors useful for empirical predictions is introduced
- Subseasonal causal precursors of the stratospheric polar vortex are identified
- Enables skillful predictions of extremely weak and strong vortex states at long lead times

Early prediction of extreme stratospheric polar vortex states based on causal precursors

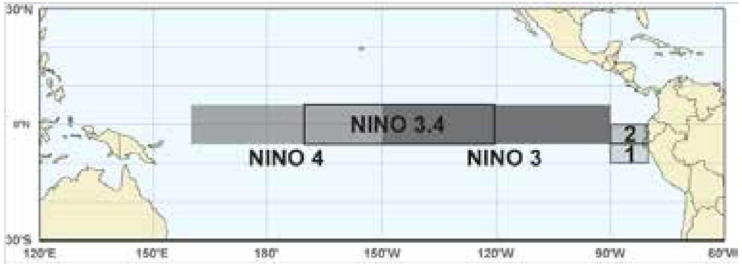
Marlene Kretschmer[1,2], Jakob Runge[3], and Dim Coumou[1,4]

[1]Potsdam Institute for Climate Impact Research, Earth System Analysis, Potsdam, Germany, [2]Department of Physics, University of Potsdam, Potsdam, Germany, [3]Imperial College, Grantham Institute, London, UK, [4]Department of Water and Climate Risk, Institute for Environmental Studies (IVM), VU University Amsterdam, Amsterdam, Netherlands

Abstract Variability in the stratospheric polar vortex (SPV) can influence the tropospheric circulation and thereby winter weather. Early predictions of extreme SPV states are thus important to improve forecasts of

Motivation: Limitations of CEN

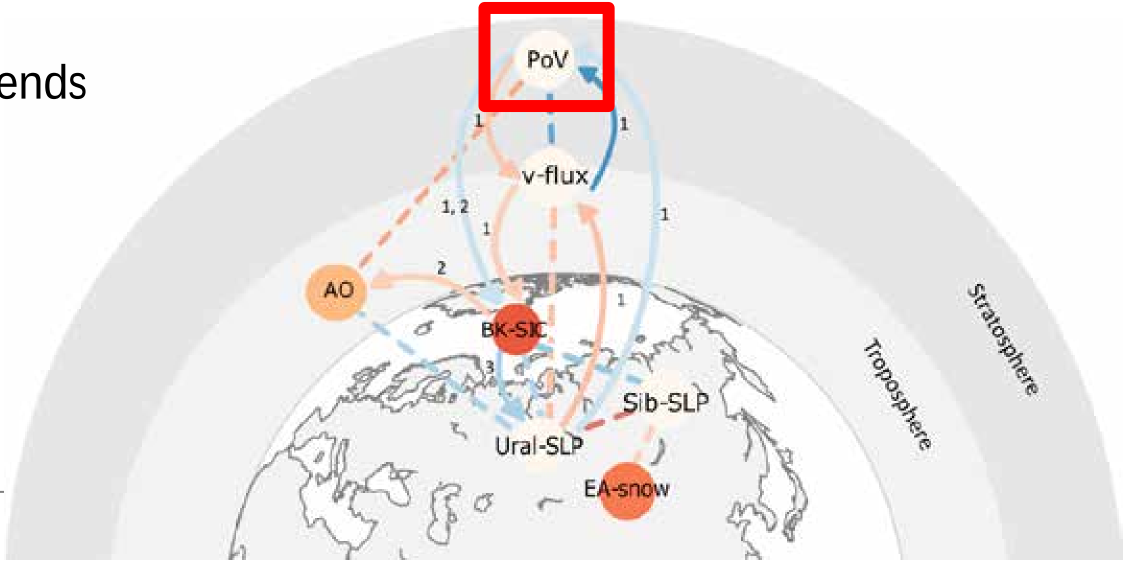
- CEN outcome depends on included actors
- Which processes and hypothesis should be considered?
- Over which regions should the indices be calculated?
  e.g. which ENSO index?



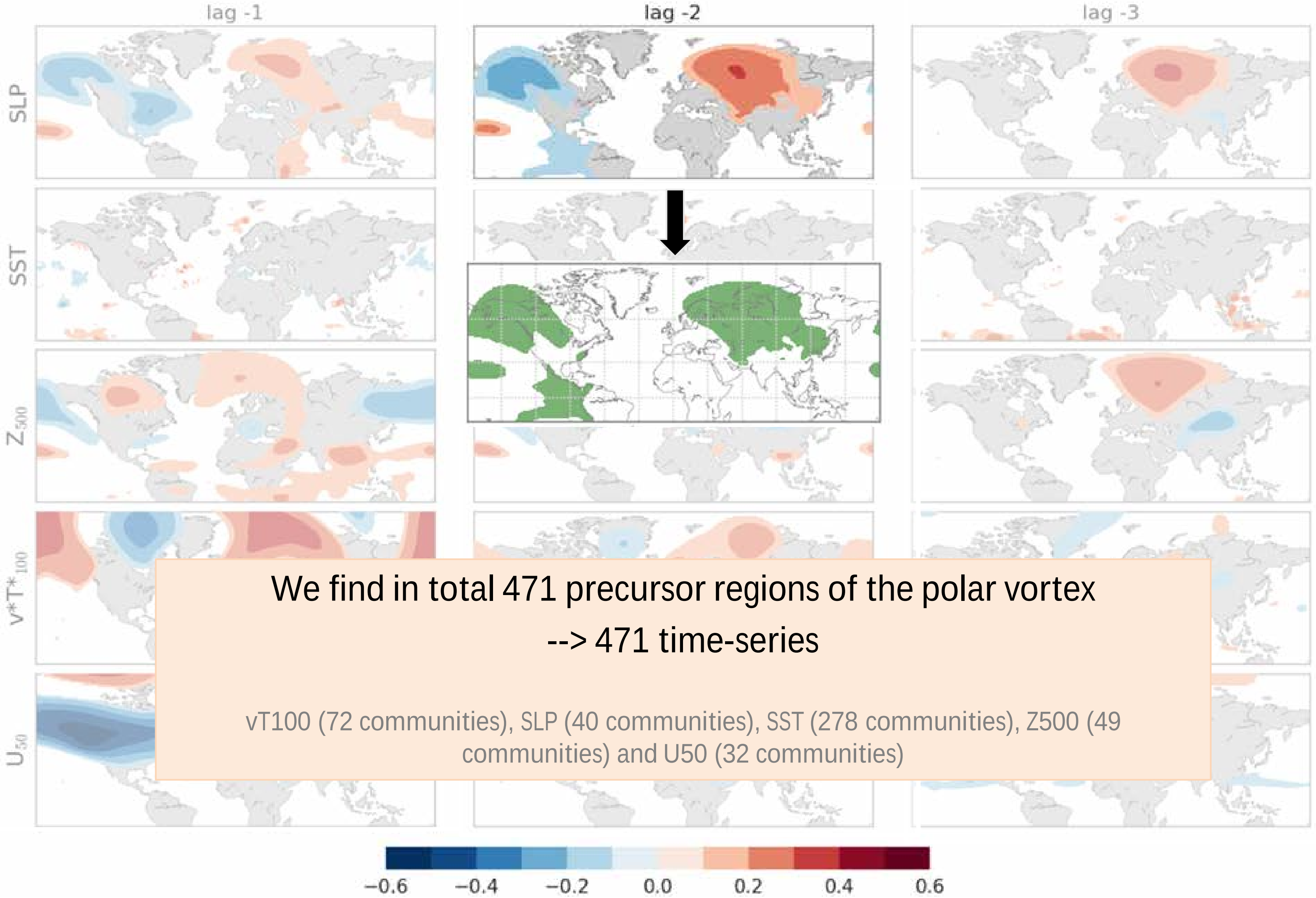
AIM
Algorithm that objectively identifies potential drivers of an index of interest

DATA AND PARAMETER SELECTION

- <u>Response variable</u>:
  Polar Vortex index in winter (NDJFM)
- <u>Potential drivers:</u>
  SST, SLP, GPH 500mb, v*T* 100mb, uwind 50mb
  from 20°S – 90°N
- Calculate anomalies, remove trends
- Half-monthly time-series
- Maximum lag = 4 months
  (8 time-steps)



kretschmer@pik-potsdam.de



We find in total 471 precursor regions of the polar vortex

--> 471 time-series

vT100 (72 communities), SLP (40 communities), SST (278 communities), Z500 (49 communities) and U50 (32 communities)

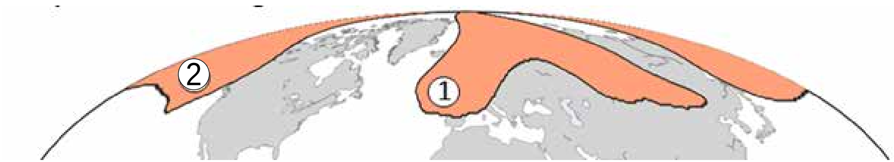
Detection of Causal Drivers

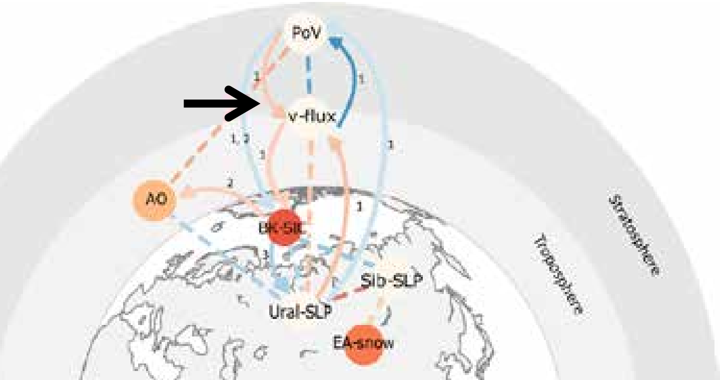
After we apply CEN-algorithm, from the 471 potential drivers only <u>3 causal precursors</u> remain

Region 1: PoV (lag-1)

Region 2: v*T* at 100hPa (lag-1)

Region 3: v*T* at 100hPa (lag-1)

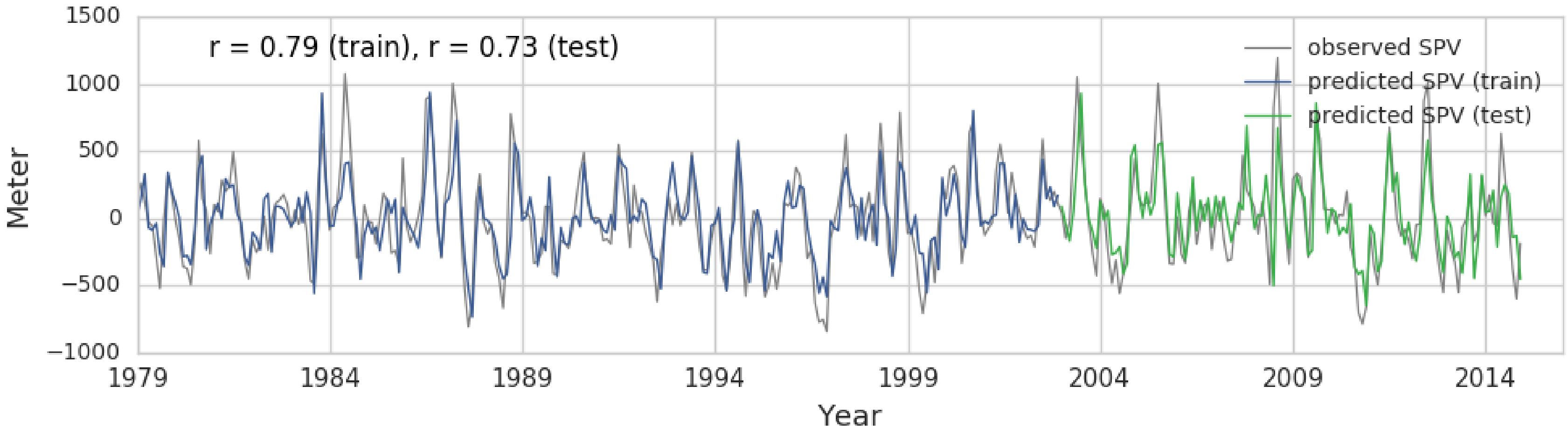




Linear Regression model:

$$\mathbf{PoV_t} = \beta_0 + \beta_1 \mathbf{PoV_{t-1}} + \beta_2 \mathbf{Region2_{t-1}} + \beta_3 \mathbf{Region3_{t-1}} + \varepsilon$$

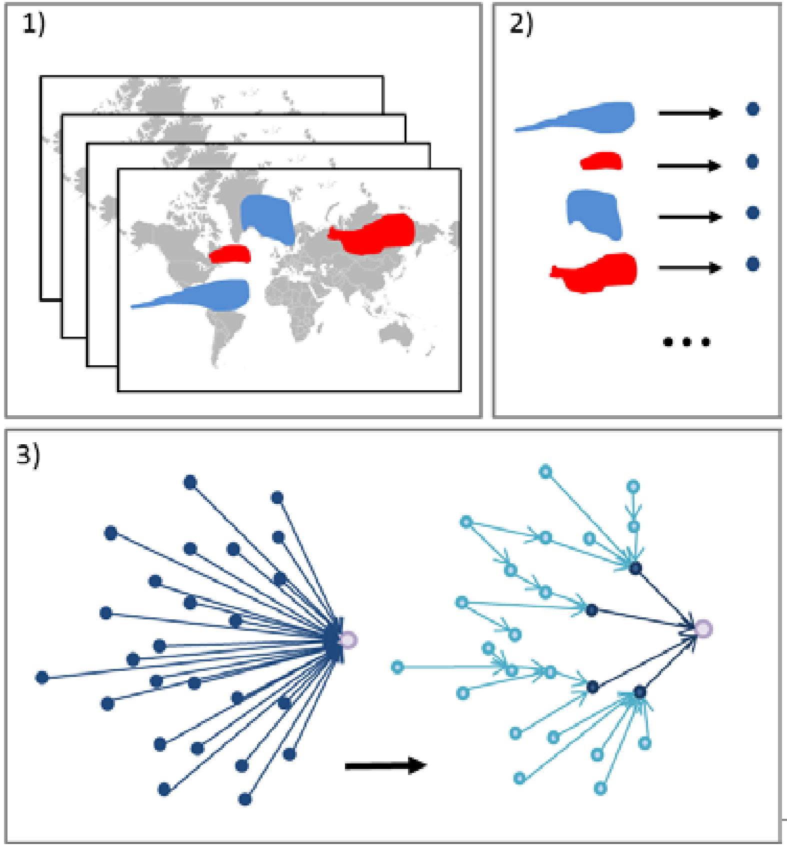
EVALUATION OF DETECTION SCHEME AND REGRESSION MODEL



Do RG-CPD for training data and apply to independent test data:

- Robust precursors
- No overfitting of model

RESPONSE-GUIDED CAUSAL PRECURSOR DETECTION (RG-CPD)



1) Detect regions in multi-variate data which correlate positively (red) or negatively (blue) with the response variable at different lags

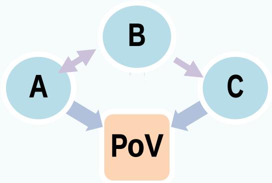
2) Take area-weighted averages of all regions creating time-series of precursors.

3) A causality test removes all non-causal links due to common drivers, auto-correlation or indirect links.

Kretschmer et al., (GRL 2017)

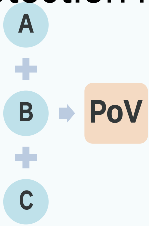
CONCLUSION

PART I
Causal effect networks for hypothesis testing



- Correlation analysis is limited in interpretability
- CEN-algorithm is multi-variate approach to overcome some limitations
- identifies and quantifies causal relationships
- Useful for hypothesis testing

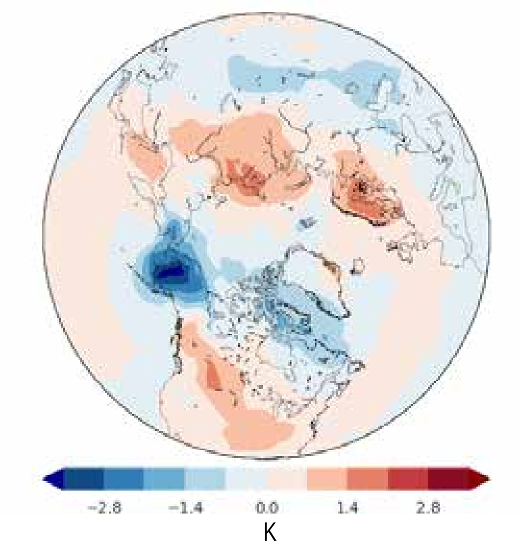
PART II
Response-guided causal precursor detection for predictions



- RG-CPD algorithm objectively detects causal precursors of a response variable
- Avoids overfitting of linear models and is therefore also suitable for predictions
- Restriction to longer lead-times can be used for early-warning systems

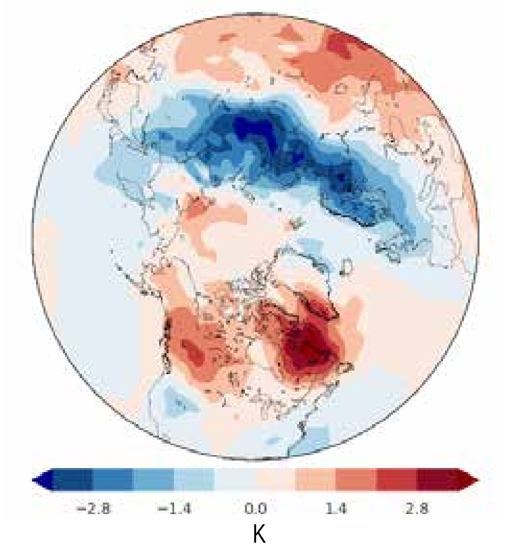
SURFACE TEMPERATURE RESPONSE

Strong Polar Vortex

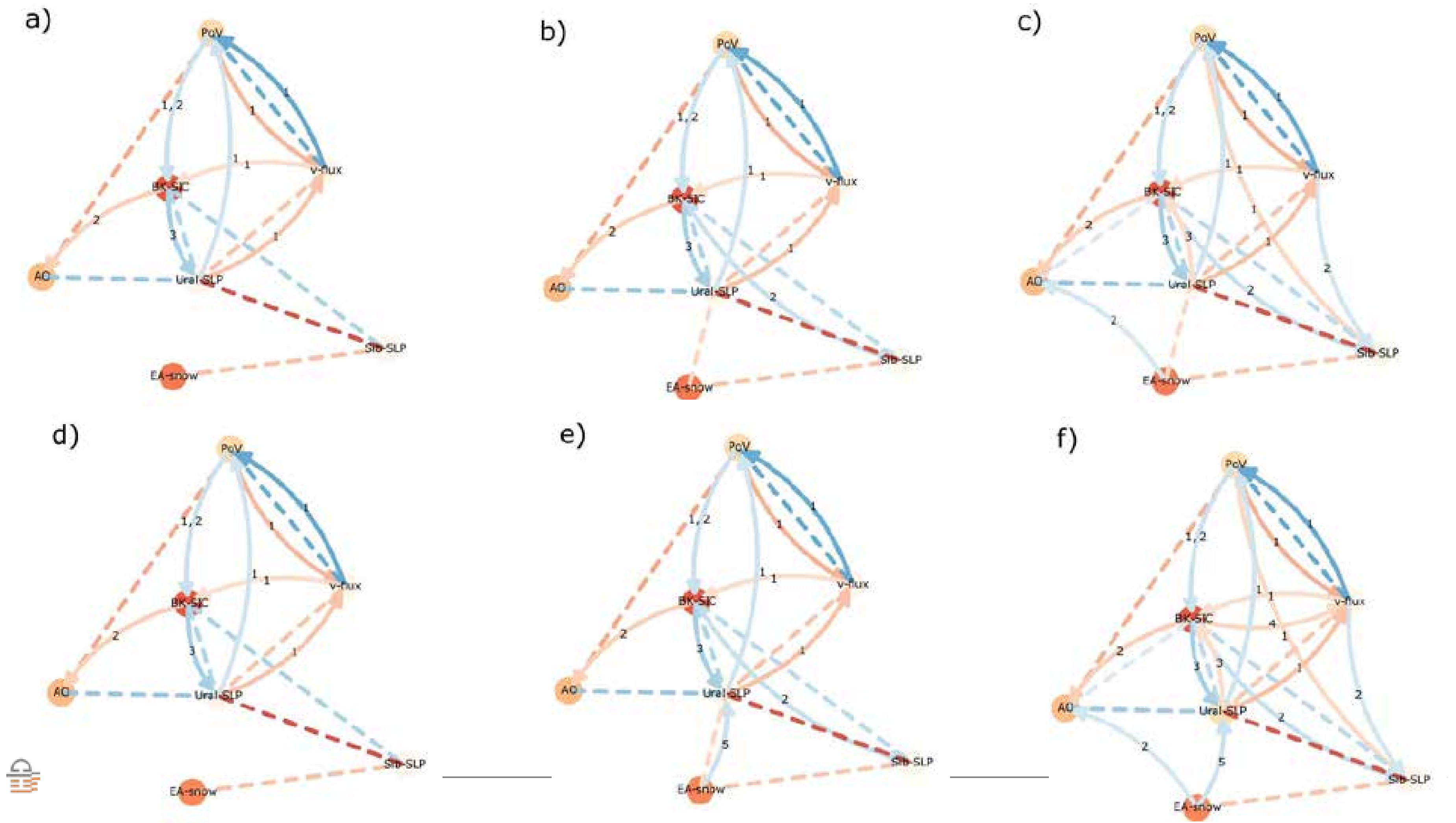


Mild winters, +AO

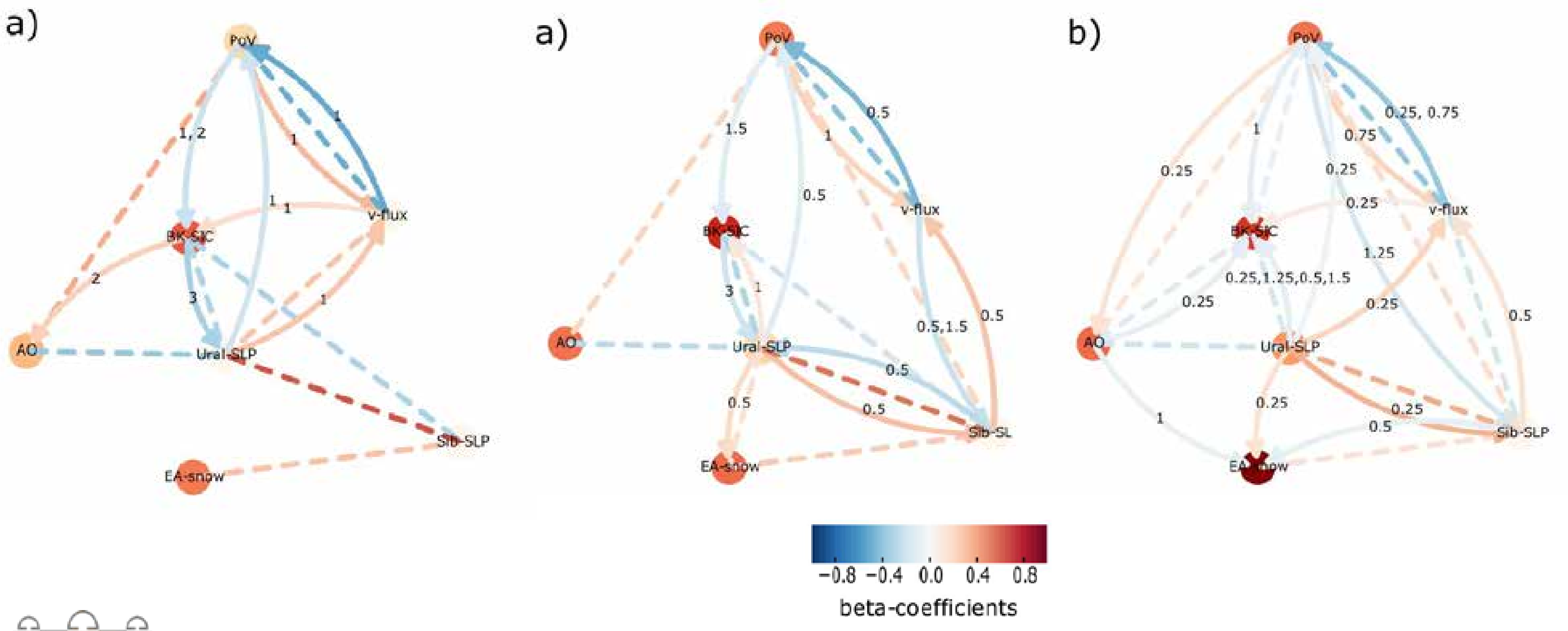
Weak Polar Vortex



Cold winters, -AO


a)
b)
c)
d)
e)
f)
PoV
v-flux
BK-SIC
AO
Ural-SLP
Sib-SLP
EA-snow

monthly 2-weekly weekly



Long lead-lag Prediction Skill

Receiver-operating-characteristic (ROC) false-positive-rate vs. true-positive-rate for different percentiles

- 64% of weak SPV states are predicted by our model with a false-alarm-rate of ~4% (odds-ratio = 42.3)
- For longer lead-times, models still correctly predict 42% (16-30 days ahead), 22% (31-45 days ahead) and 14% (46-60 days ahead) with associated odds-ratios of 10.3, 3 and 1.5.