ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR-
UND MEERESFORSCHUNG

Analysis of atmospheric circulation from climate model big data - Current approaches and future challenges

Dörthe Handorf, Klaus Dethloff, Ralf Jaiser, Annette Rinke

Section Physics of the Atmosphere

Research Department Potsdam, Germany

Alfred Wegener Institute

Helmholtz Center for Polar and Marine Research

AWI Data Science Symposium

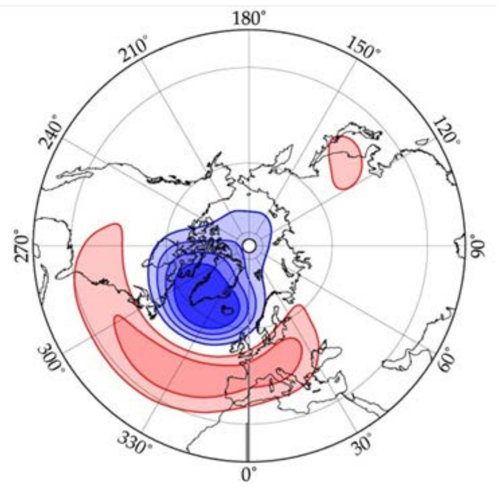
8./9.11.2017

HELMHOLTZ
GEMEINSCHAFT

Patterns of atmospheric variability – Atmospheric teleconnections

- preferred patterns of low-frequency variability (subseasonal to decadal timescales)
- refer to recurring and persistent, large-scale patterns of pressure and circulation anomalies that spans vast geographical areas
- are localised in definite regions (hemispheric-scale, basin-wide, continental)
- are a naturally occurring aspect of our chaotic atmospheric system
- can arise primarily due to the internal atmospheric dynamics, but can be impacted by external forcings
- reflect large-scale changes in the atmospheric wave and jet stream patterns
- influence temperature, rainfall, storm tracks over vast areas

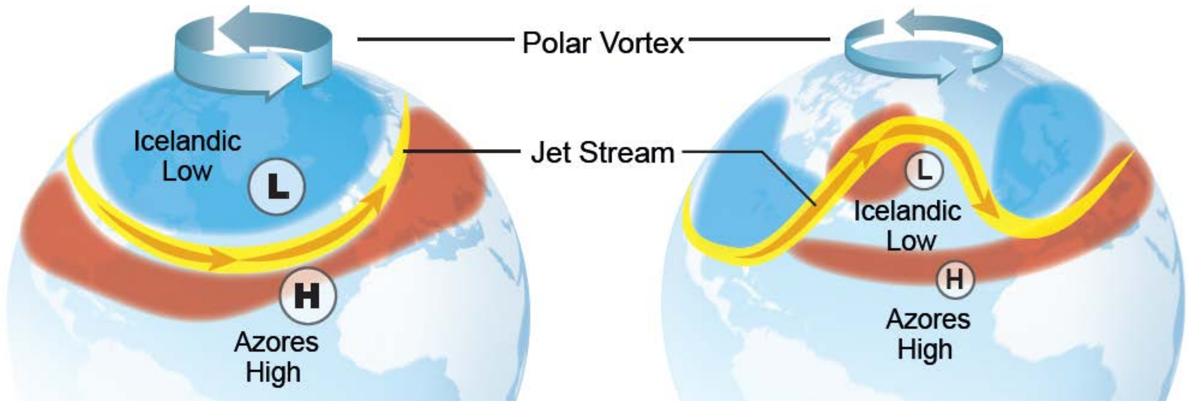
→ **Example: North Atlantic Oscillation (NAO)**
Most important pattern over North-Atlantic-European Region
Dipole structure over North Atlantic
associated with changes in location/intensity of North Atlantic jet stream and storm tracks



Patterns of atmospheric variability - N-Atl./Europe
The North Atlantic Oscillation (NAO)

Two states of atmospheric circulation over the North Atlantic-European sector

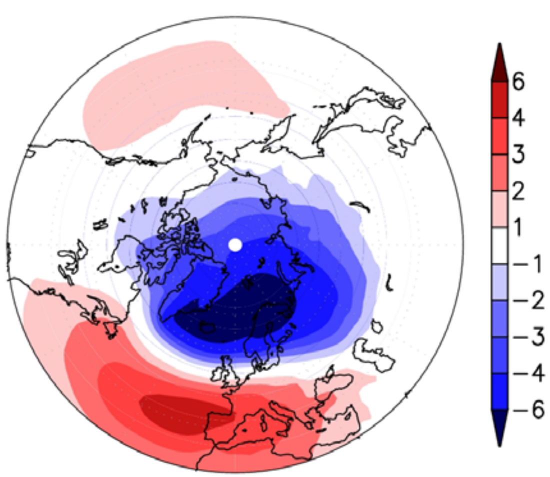
Zonal jet stream
Small-amplitude planetary waves

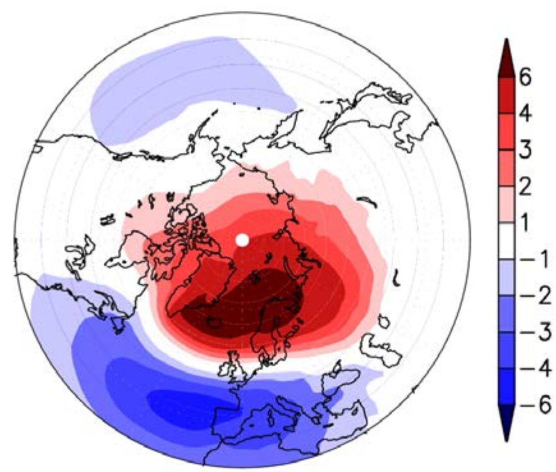


Meandering jet stream
Large-amplitude planetary waves

Corresponding **patterns of sea-level pressure anomalies**
(deviation from mean pressure distribution)

North-Atlantic Oscillation in positive phase (NAO+)





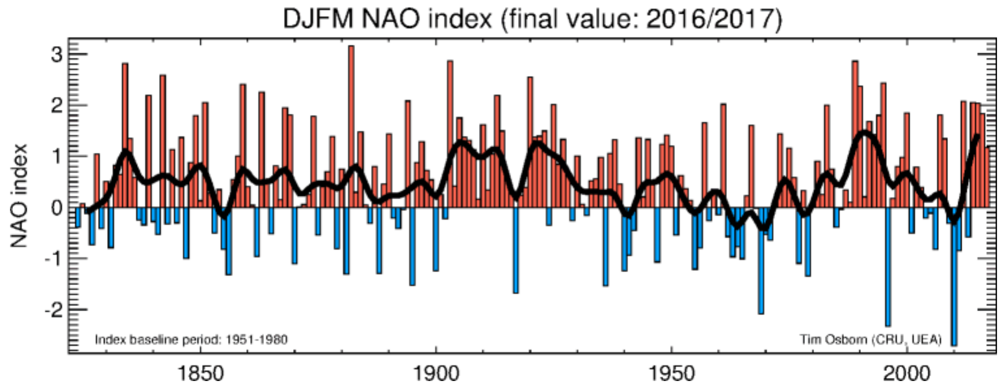
North-Atlantic Oscillation in negative phase (NAO-)

HELMHOLTZ
GEMEINSCHAFT

Patterns of atmospheric variability - N-Atl./Europe
The North Atlantic Oscillation (NAO)

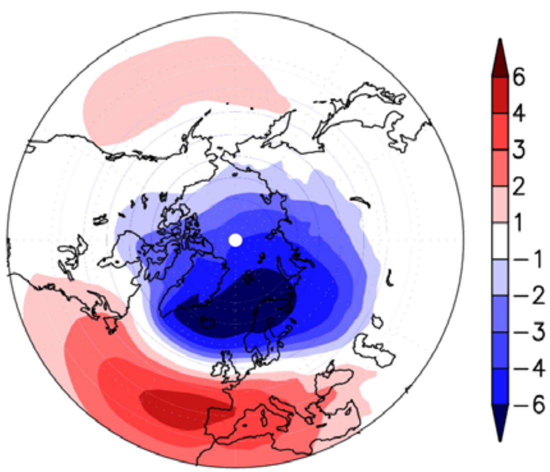
Two states of atmospheric circulation over the North Atlantic-European sector

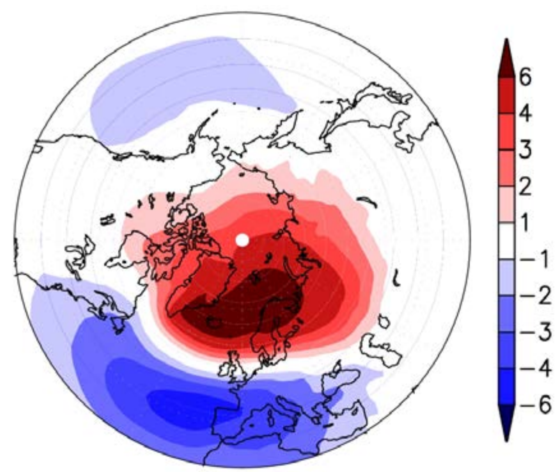
Interannual to decadal variability



Corresponding **patterns of sea-level pressure anomalies**
(deviation from mean pressure distribution)

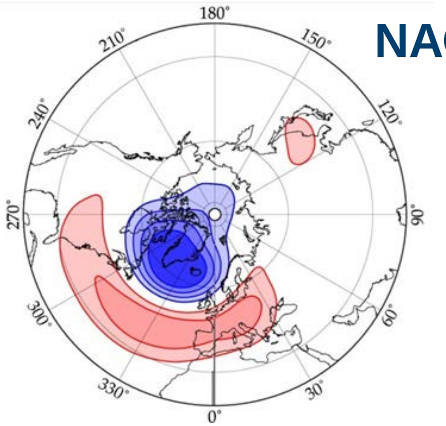
North-Atlantic Oscillation in positive phase (NAO+)





North-Atlantic Oscillation in negative phase (NAO-)

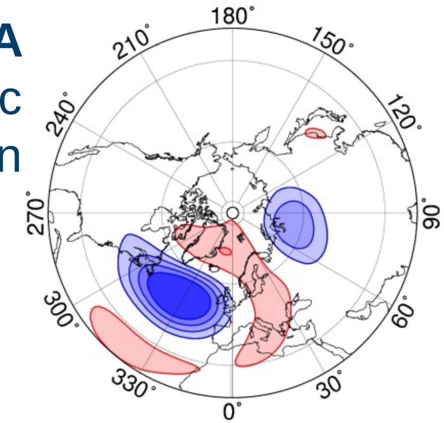
Patterns of atmospheric variability – Research Questions



NAO

➢ associated with changes in the North Atlantic jet stream and storm tracks

EA
East-Atlantic pattern



Research Questions

- ➢ Understand past, recent future changes in the spatial/temporal structure of teleconnection patterns
- ➢ Internal dynamics versus external forcing
- ➢ Potential predictability of teleconnections

Global, gridded data sets

- ➢ Reanalysis data sets
- ➢ Climate model simulations

Hypothesis-driven Approach

Step 1: Analysis of changes – Evaluation of climate models
Step 2: Development of hypothesis
Step 3: Provision of evidence
→ New model experiments
→ New analysis

Global gridded climate data sets: Reanalysis and Earth system models

Data amount of Reanalysis:

→ Example: ECMWF, global

ERA40	1957/09 to 2002/08			
	Sub-daily, Monthly	2.5°x2.5° / 1.125°x1.125°	60 levels 0.1 hPA top	
ERA-Interim	1979/01 to present			
	Sub-daily, Daily, Monthly	0.75°x0.75°	60 levels 0.1 hPA top	
ERA5	1979/01 to present			
	hourly, Sub-daily, Daily, Monthly	0.28°x0.28°	137 lev	0.01 hPA top

Expected: 5 Petabytes for ERA5

Global gridded climate data sets: Reanalysis and Earth system models

Data amount Reanalysis:
Example: ECMWF, global → ERA5 → 5 Petabytes

Data amount CMIPs Coupled model intercomparison projects

- **CMIP3** (for IPCC AR4 2007): 17 institutes (groups) and 25 models → 40 TB

– total years simulated: 70000
– individual models simulated on average 2800yrs

- **CMIP5** (for IPCC AR5 2013): 26 institutes (groups) and 60 models → 2 PB

– total years simulated: 330000
– individual models simulated on average 330000/60 = 5500 years

- Extrapolation for **CMIP6** data federation:

– CMIP6 has a more complex experiment structure than CMIP5
– 32 institutes (groups) and many model versions
– more models with higher resolution models
– 21 MIPs, many experiments, larger ensembles
– Expectations:
  → Volume: 150 PB
  → Number of files: 280 Mio Files

→ Climate big data?!

Analysis of teleconnection patterns
Empirical Orthogonal Function Analysis (EOF)

- **Empirical Orthogonal Function Analysis (EOF)**
  → reduce the dimensionality of the data
  → find the most important patterns explaining the variability
  → provide information about spatial structures and temporal scales

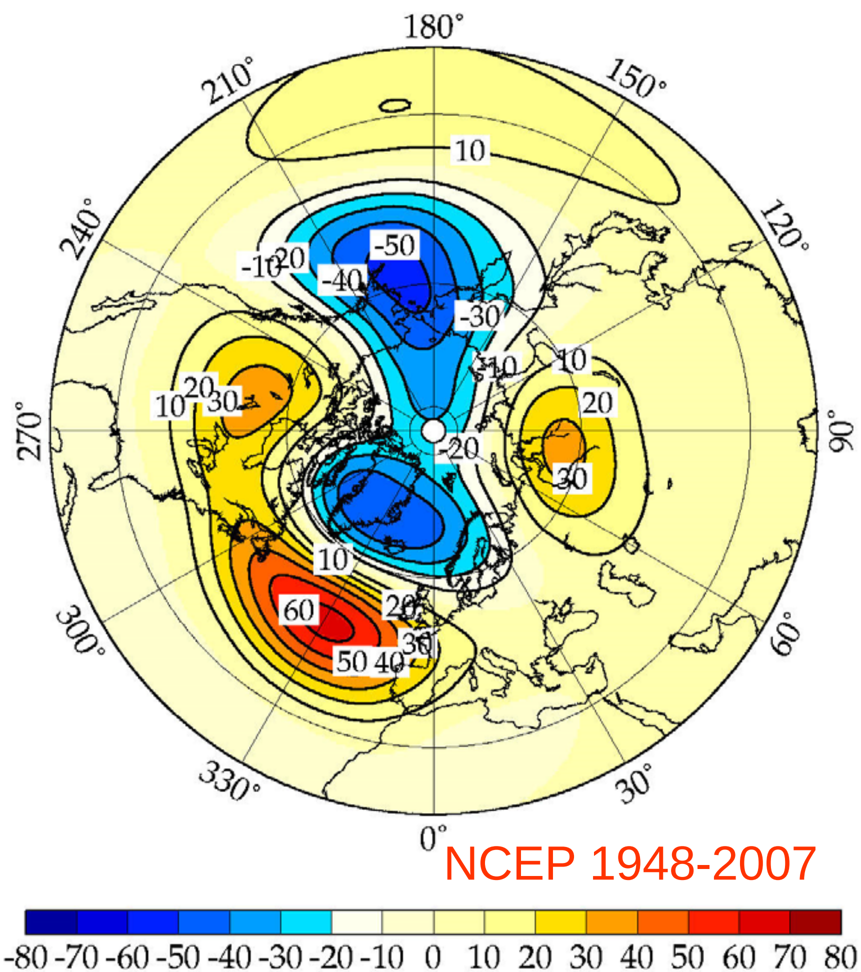
- Data field represented compactly in terms of EOFs: $Q'_i(t) = \sum_{j=1}^{N} \alpha_j(t) e_{ij}$

- **Principal components** $\alpha_j(t)$
  → represent projections of data onto the *j*-th EOF

- **Rotation of EOF** produce more localised patterns
  → Rotated EOFs = linear combination of first few EOFs,
  determined by minimisation of a functional (e.g. spatial variance)

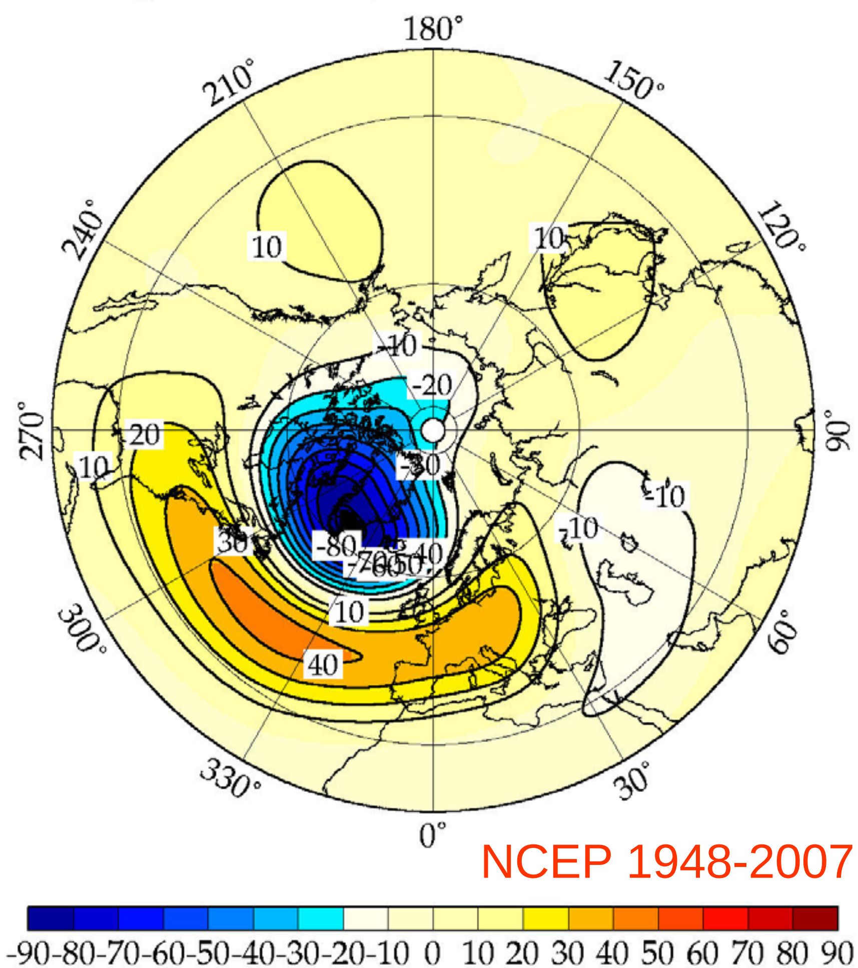
Analysis of teleconnection patterns
Empirical Orthogonal Function Analysis

EOF2 12.4% dominated by NAO



NCEP 1948-2007

ROT EOF1 13.4% NAO



NCEP 1948-2007

Analysis of teleconnection patterns in Reanalysis data and Evaluation of Climate models

- Analyses of monthly mean data of midtropospheric circulation
  → 500hPa geopotential height fields
- Analyses of dynamically active season of Northern Hemisphere (NH)
  → December, January, February data (DJF)
- Fields from 20°-90° N with removed seasonal cycle
- Evaluation with NCEP/NCAR and ERA40 Reanalyses

Analysed Experiments from CMIP3/CMIP5

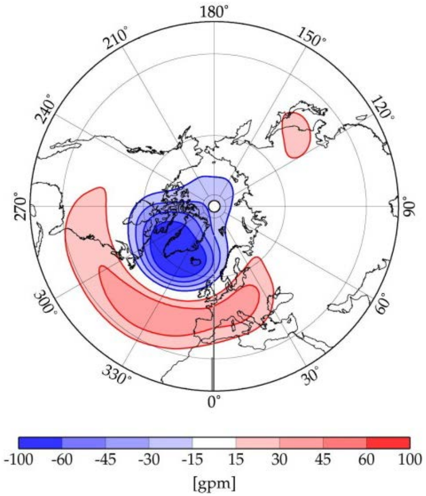
CMIP3 20th cen. simulation 1870-1999 Analyses of years 1958-1999	forced by observed atmospheric composition changes (anthropogenic & natural sources)	23 models
CMIP5 historical simulations 1850-2005 Analyses of years 1958-1999	forced by observed atmospheric composition changes (anthropogenic & natural sources) time-evolving land cover	46 models more comprehensive generally with higher spatial resolution

Technological approach

- Download of data from data centers
- Data analysis locally
- Software packages (MATLAB, R)
- Own software (FORTRAN, R) → use of libraries (NAG)

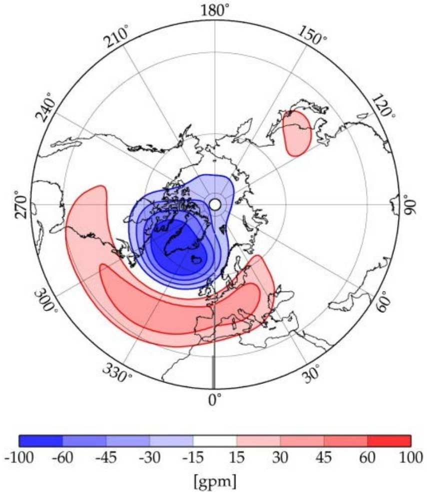
Teleconnections - Evaluation of spatial structure
CMIP3 ensemble - Period 1958-1999 - NAO

Reanalysis ERA40

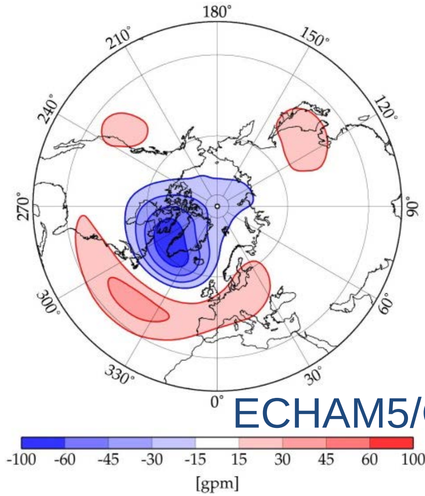


Teleconnections - Evaluation of spatial structure
CMIP3 ensemble - Period 1958-1999 - NAO

Reanalysis ERA40



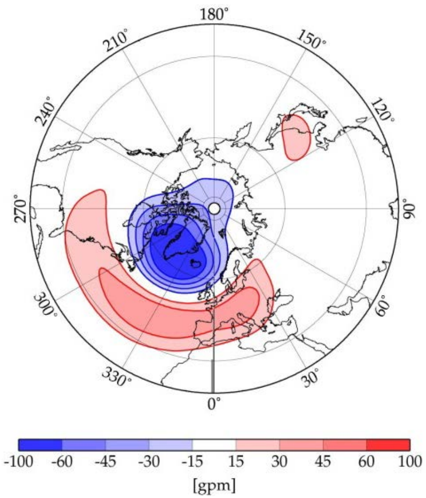
Single Model



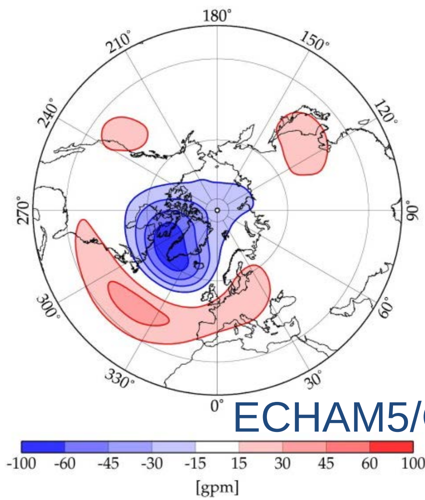
ECHAM5/OM1

Teleconnections - Evaluation of spatial structure
CMIP3 ensemble - Period 1958-1999 - NAO

Reanalysis ERA40

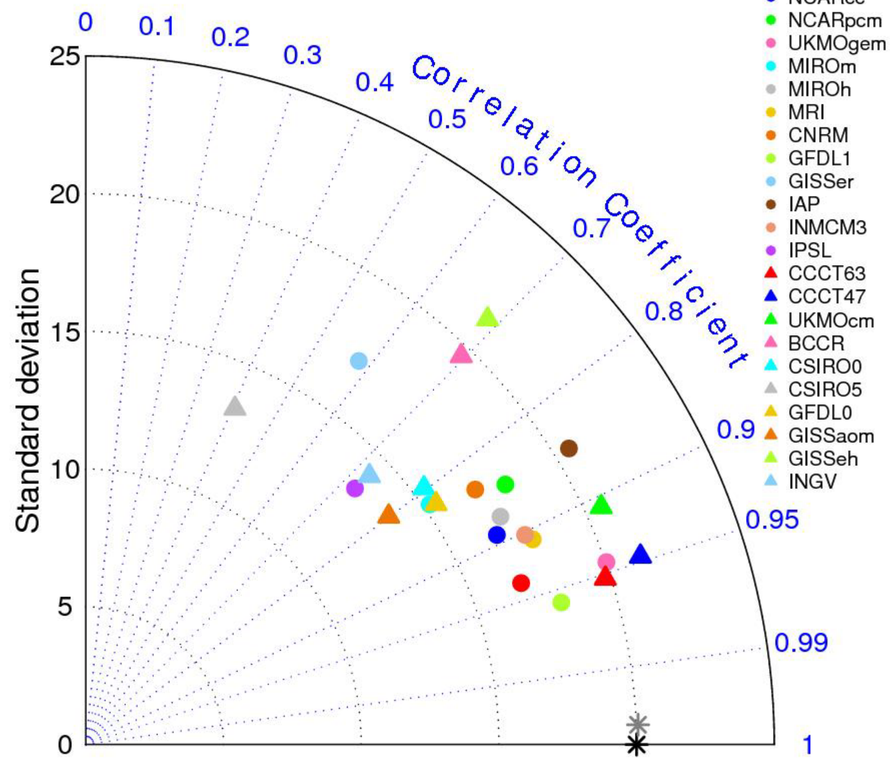


Single Model



ECHAM5/OM1

All Models

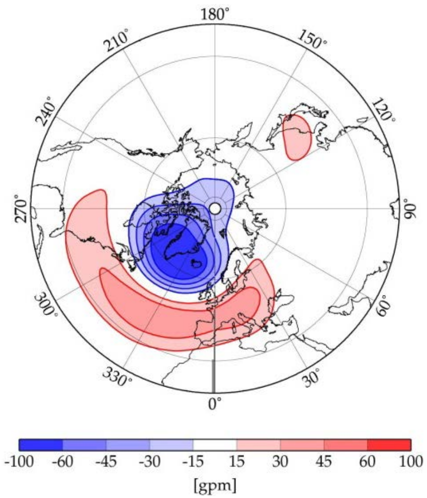


Taylor diagrams (Taylor, 2001)
→ Quantify similarity between different patterns
→ Compact summary of pattern statistics in terms of pattern correlation, root-mean-square difference and ratio of variances.

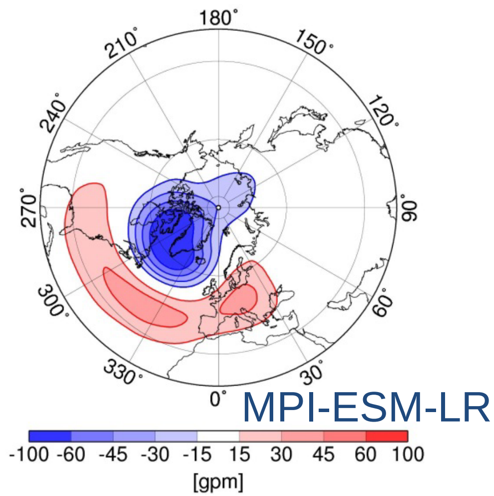
Handorf & Dethloff, Tellus 2012

Teleconnections - Evaluation of spatial structure
CMIP5 ensemble - Period 1958-1999 - NAO

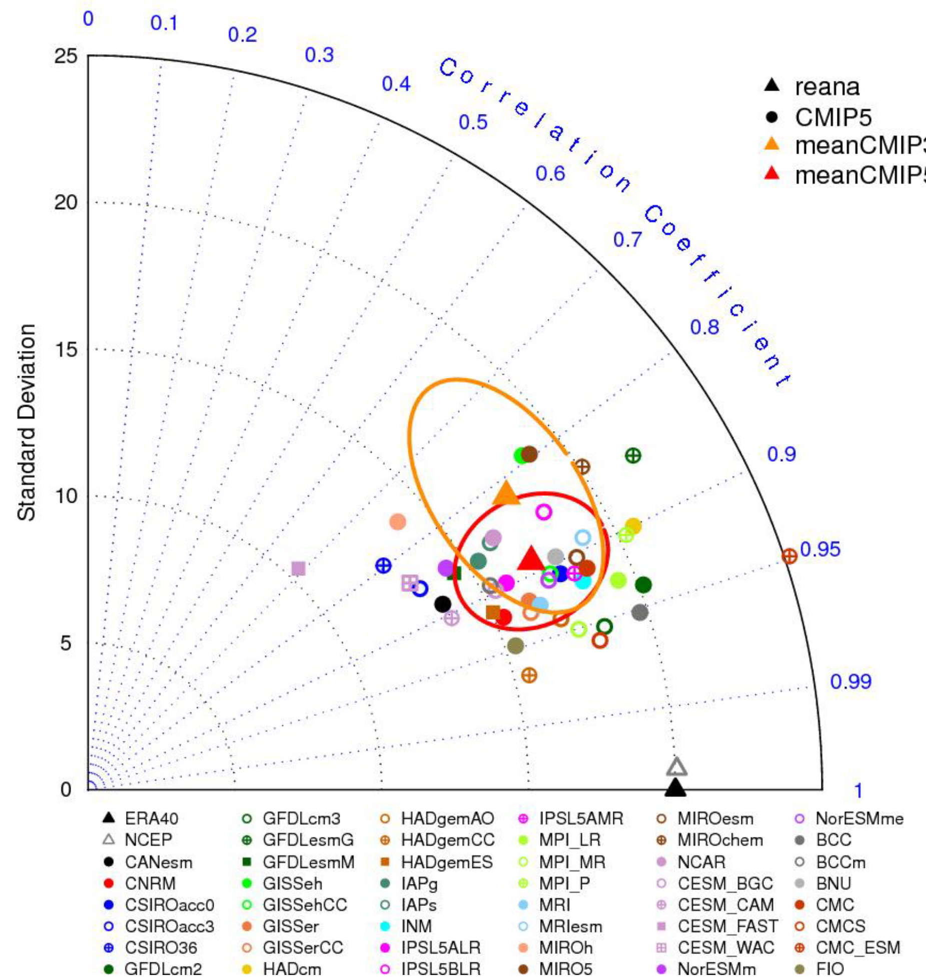
Reanalysis ERA40



Single Model



All Models



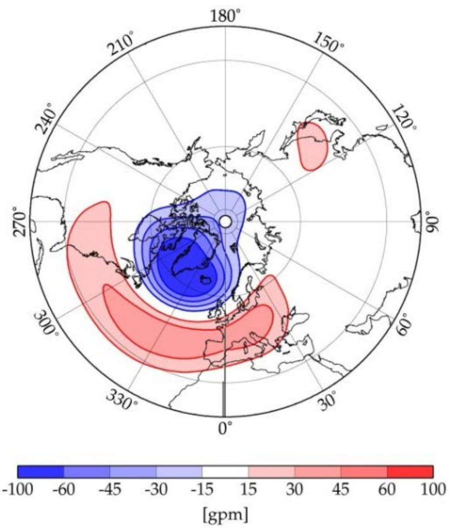
Spread in the skill of simulating spatial patterns over the CMIP5 ensemble

Improvement

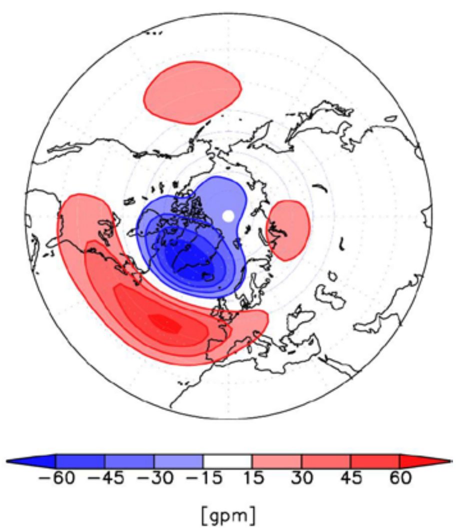
Dynamical reasons for limited skill of the CMIP3/5 ensemble in reproducing teleconnections

- Hypothesis 1: Deficiencies in atmospheric internal dynamics
- Teleconnections are related to variability of zonal wind **for the reanalyses**
  (gph and u fields are dynamically related, e.g., Athanasiadis et al., 2010; Li and Wettstein, 2012)

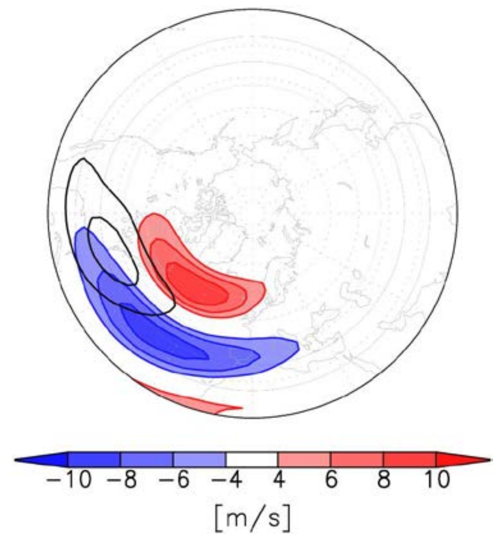
NAO pattern @500hPa



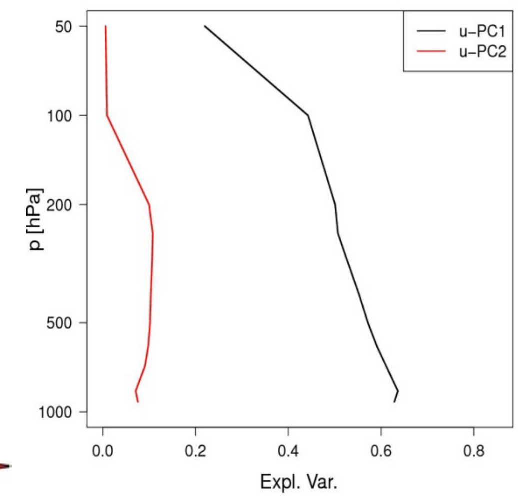
Regression onto Atlantic u-PC1@250hPa for GPH500



EOF1 of Atlantic zonal wind@250hPa



Profile of explained variance

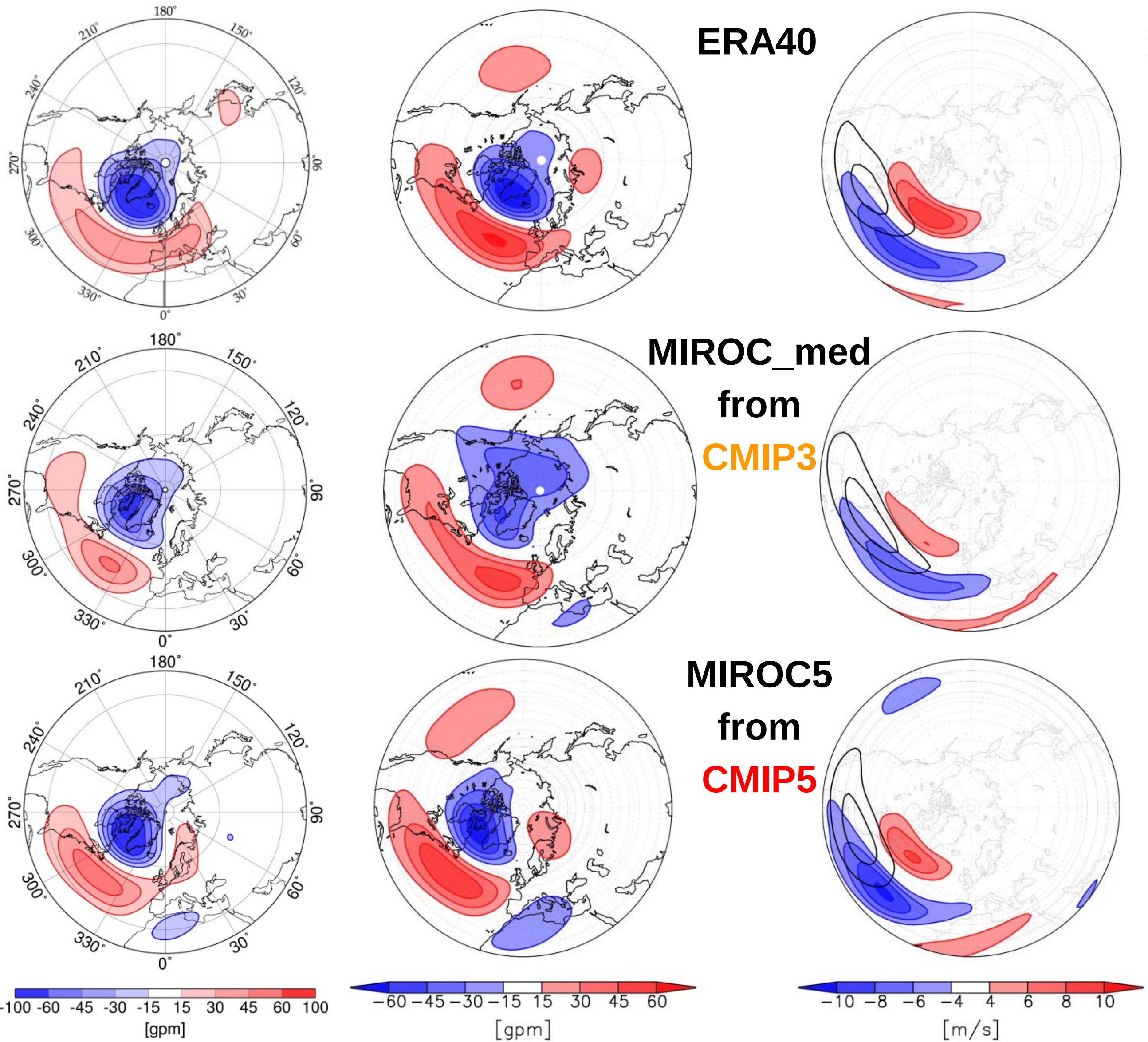


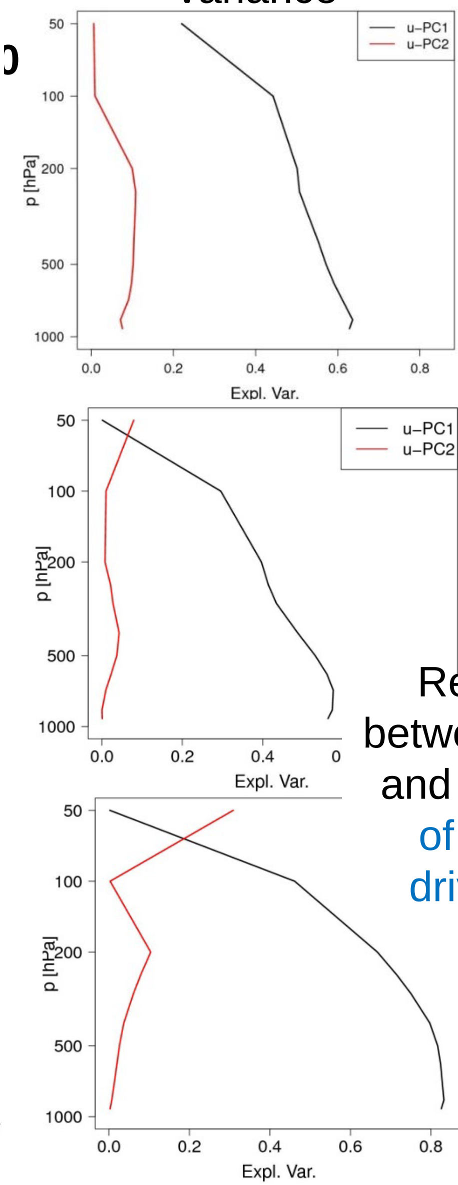
NAO closely related to EOF1 of Atlantic zonal wind (Position of eddy-driven jet)

EA closely related to EOF2 of Atlantic zonal wind (Intensity of eddy-driven jet)

Dynamical reasons for limited skill of the CMIP3/5 ensemble in reproducing teleconnections

NAO pattern @500hPa | **Regression** onto Atlantic u-PC1@250hPa for GPH500 | **EOF1** of Atlantic zonal wind@250hPa | Profile of explained variance

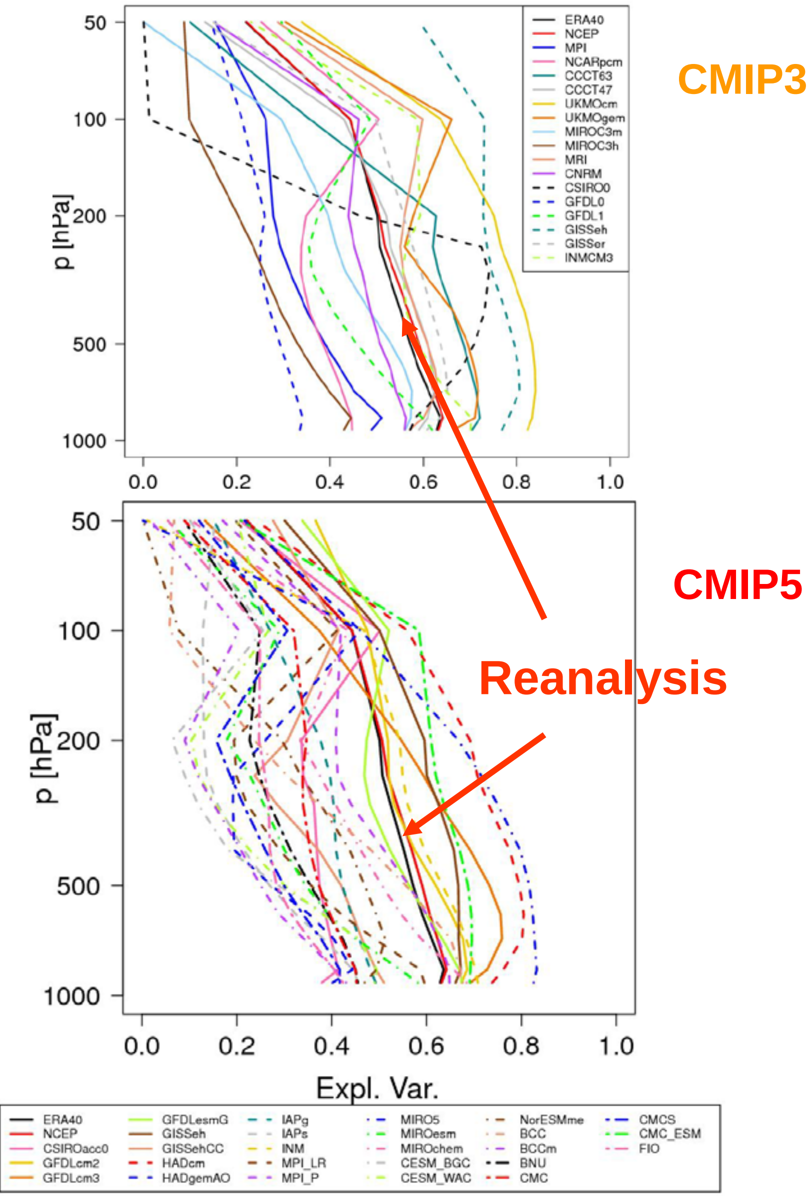




Relation between NAO and position of eddy-driven jet

Dynamical reasons for limited skill of the CMIP3/5 ensemble in reproducing teleconnections

Explained variance with NAO

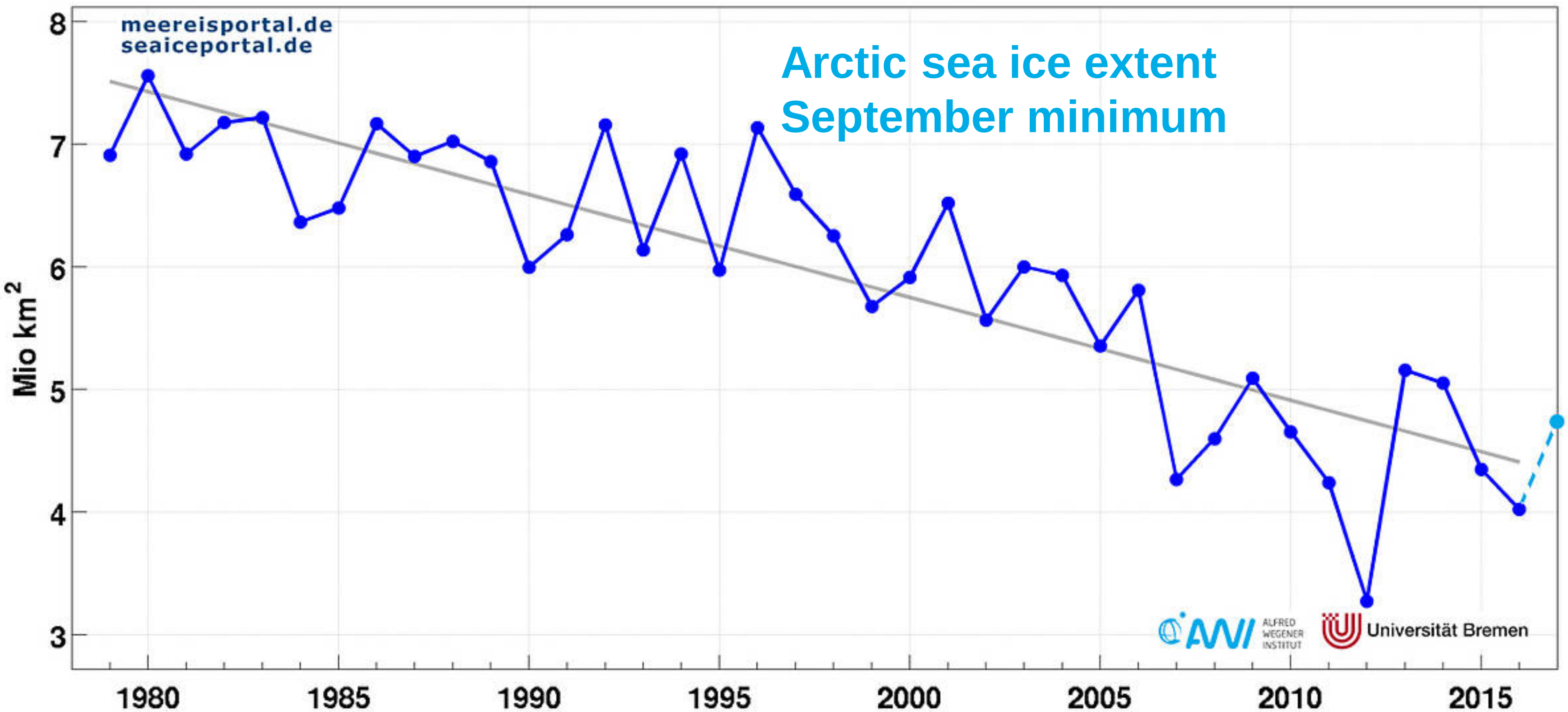


Structure of the relation between teleconnections and zonal wind variability captured by some (not all) models of the CMIP3 and CMIP5 ensemble (large spread)

The quality of the simulated teleconnection pattern is largely determined by the quality of the simulated zonal wind variability pattern

Impact of external forcings on teleconnections and their reproduction in climate models

- Hypothesis 2: Prefered state of atmospheric variability patterns is influenced by external forcing factors
- Extample: Forcing due to **changes in sea-ice**



Impact of sea-ice changes on atmospheric teleconnections

- Hypothesis 2: Prefered state of atmospheric variability patterns
  over the North-Atlantic-Eurasian region is influenced
  by Arctic climate changes (e.g. sea-ice changes)

Methods: Maximum Covariance Analysis (MCA):

- Statistical method detecting coupled patterns between pairs of climate fields
- Maximized covariance of time series associated to each pattern
- **Reanalysis data: ERA-Interim**

  September sea ice concentration 1979-2015

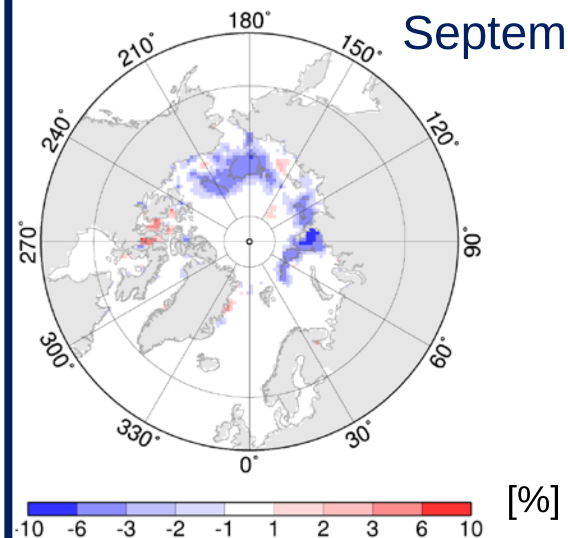
  Mean sea level pressure and geopotential height fields
  in Winter (February or DJF, 1979-2015)

- **Climate model data**:

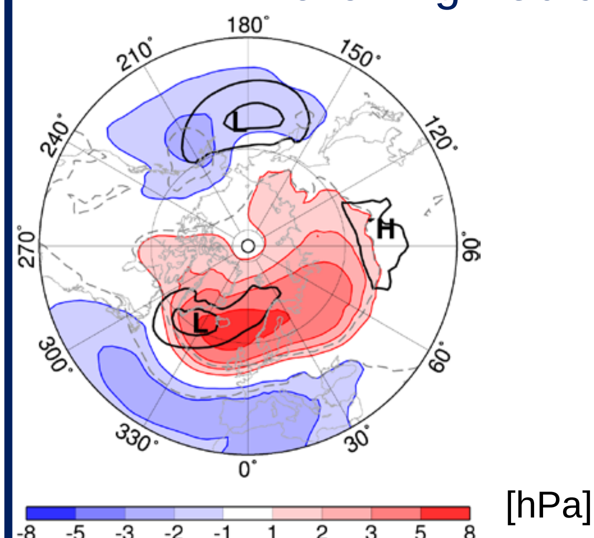
  AFES (Atmospheric general circulation model For Earth Simulator)
  Ensemble model simulations, 30 members
  AMIP-style, 1979-2014

Impact of sea-ice changes on atmospheric teleconnections – Reanalysis

Sea ice concentration
September (HadISST Data)



Sea level pressure
Following February (ERA-Interim)

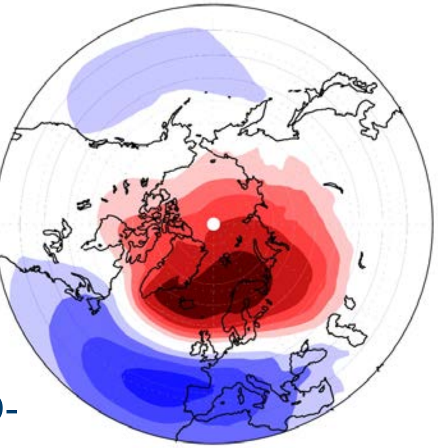


Planetary-scale response in February

Coupled Patterns 1979-2015

- Statistical relation between sea ice retreat and changes of atmospheric circulation patterns

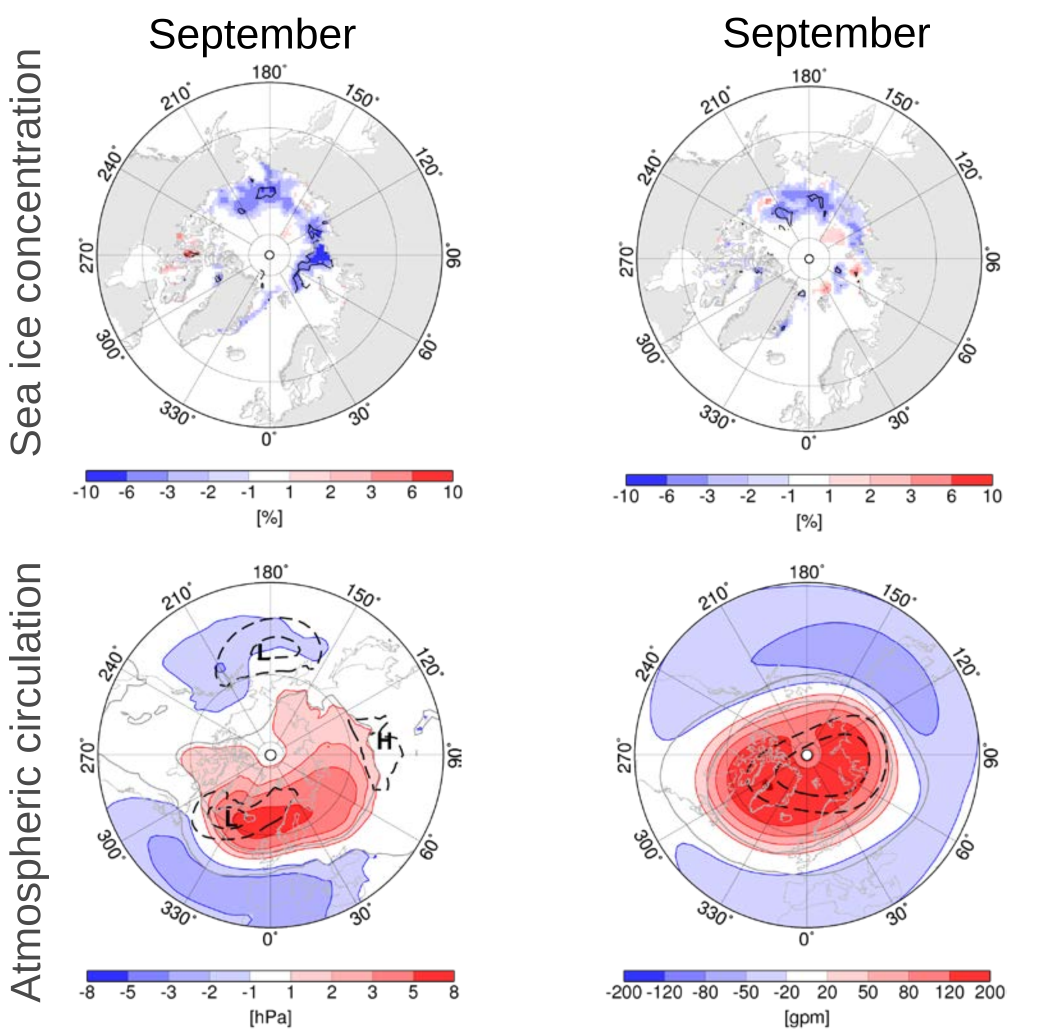
- Changes of centers of action, similarity with pattern of **NAO** in negative phase



Pattern of NAO-

HELMHOLTZ GEMEINSCHAFT

Impact of sea-ice changes on atmospheric teleconnections – Reanalysis



February
Sea level pressure
39% explained Covariance

February
GPH 50hPa
61% expl. Covariance

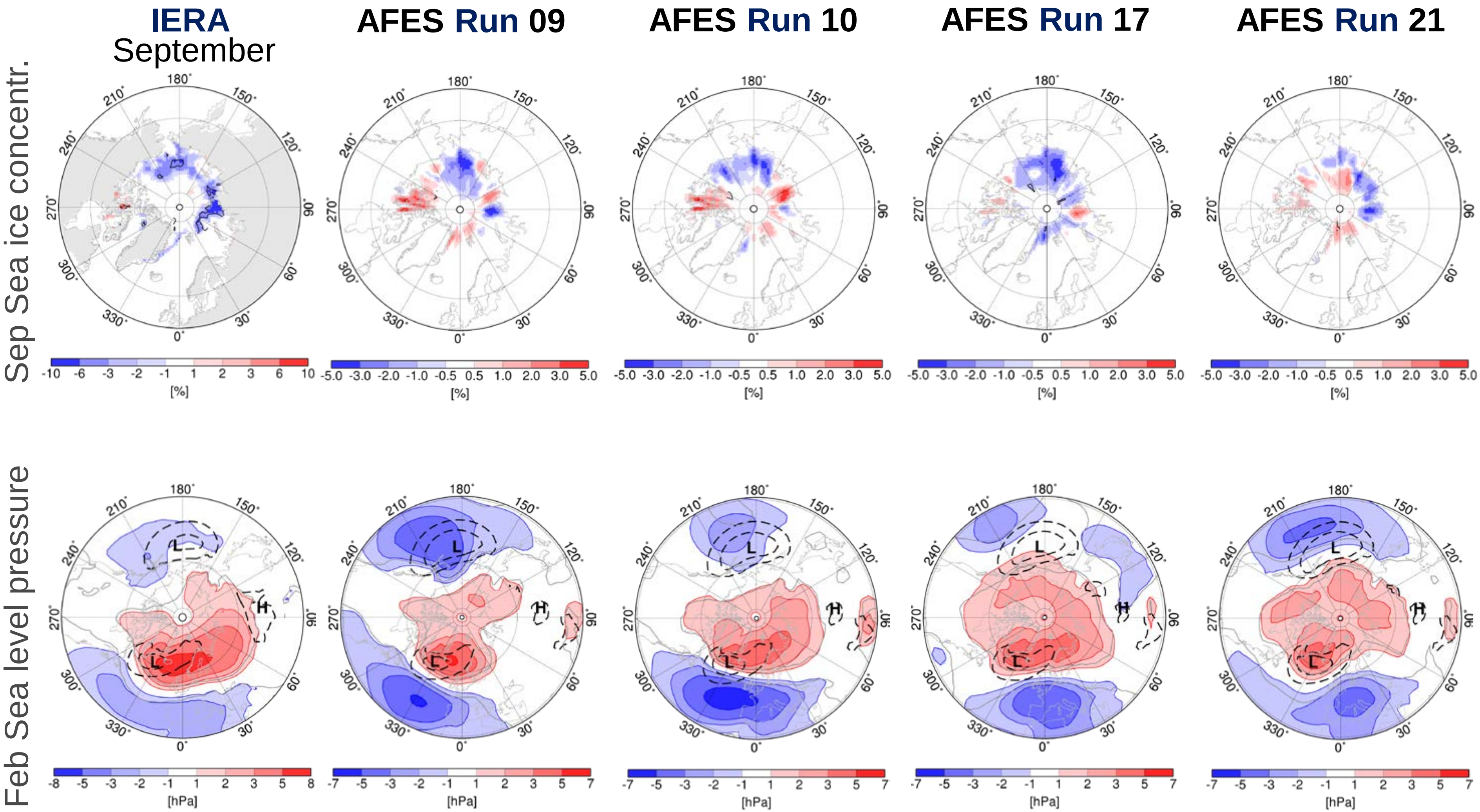
Planetary-scale response in Feb.

Coupled Patterns 1979-2015

- Statistical relation between sea ice retreat and changes of atmospheric circulation patterns
- Changes of centers of action, similarity with pattern of **NAO** in negative phase
- Associated changes in stratosphere→ Weaker stratospheric Polar Vortex

Jaiser et al. 2012, 2013, 2016
Handorf et al. 2015

Impact of sea-ice changes on atmospheric teleconnections – Ensemble model simulations

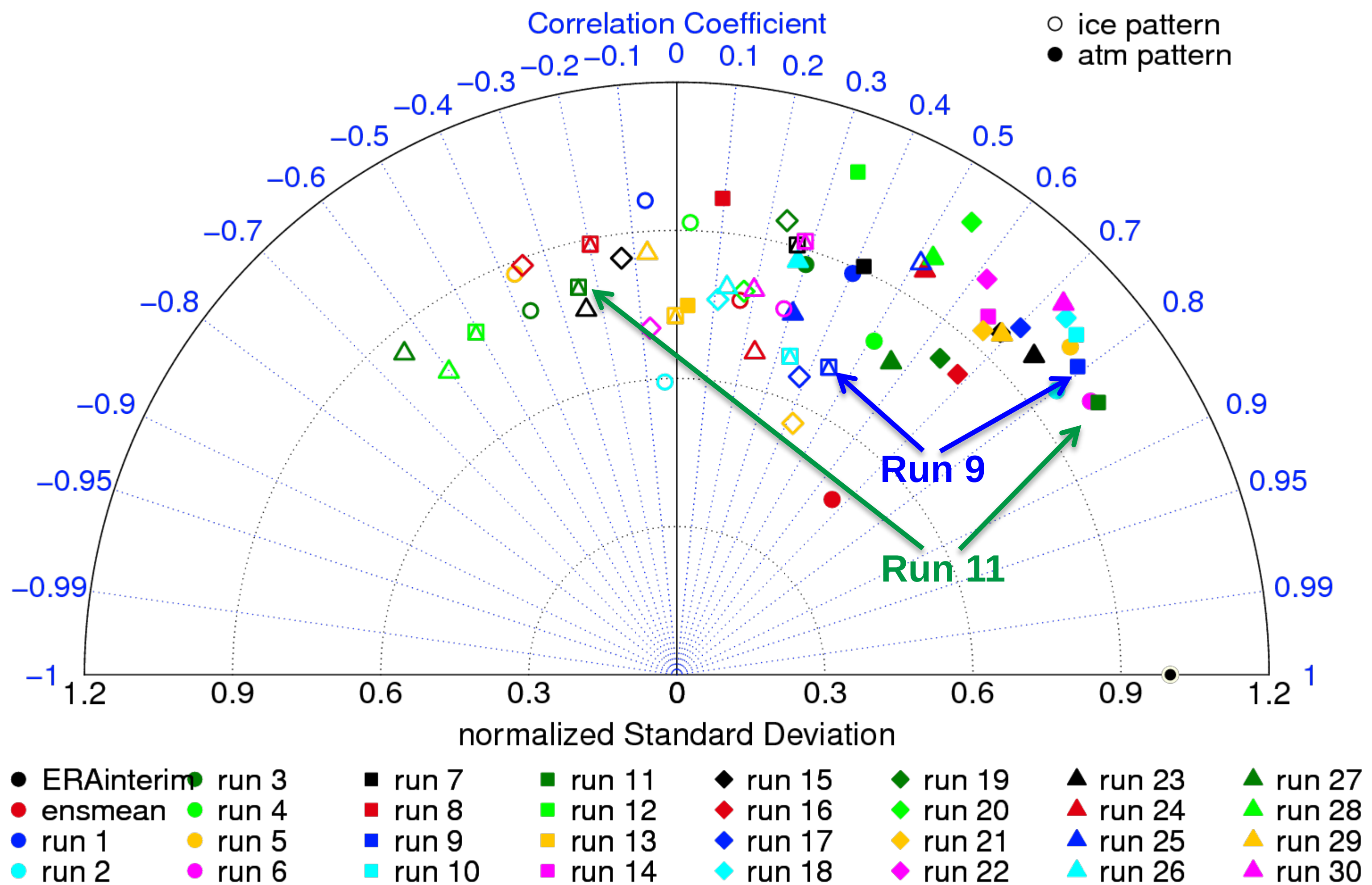


39% expl. Covar. 32% expl. Covar. 30% expl. Covar. 32% expl. Covar. 25% expl. Covar.

- Model: AFES (Atmospheric general circulation model For Earth Simulator)
- Ensemble model simulations, 30 members, AMIP-style

Impact of sea-ice changes on atmospheric teleconnections – Ensemble model simulations

- Coupled Patterns from ensemble model simulations → **Taylorplot**
- AFES Ensemble model simulations, 30 members



Impact of sea-ice changes on atmospheric teleconnections

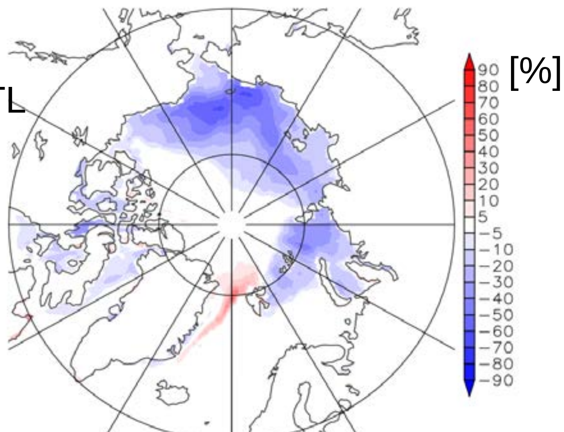
Hypothesis-driven Approach

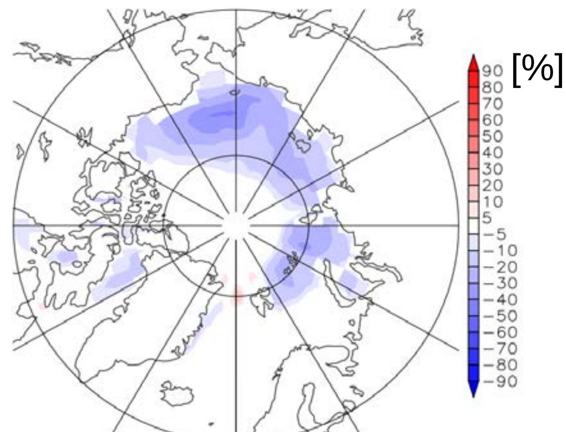
Step 1: Analysis of changes – Evaluation of climate models
Step 2: Development of hypothesis
Step 3: Provision of evidence
→ New specific model experiments
→ New analysis

Set-up of new specific model experiments

- Model: AFES → 2 model simulations, with 60 perpendicular years each
  - CNTL: High ice conditions as observed from 1979 to 1983
  - NICE: Low ice conditions as observed from 2005 to 2009
  - **Only sea ice is different between both runs**

Maps of sea ice concentration in fall (SON) for low minus high ice conditions





Nakamura et al. (2015, JGR); Jaiser et al. (2016)

Impact of sea-ice changes on atmospheric teleconnections

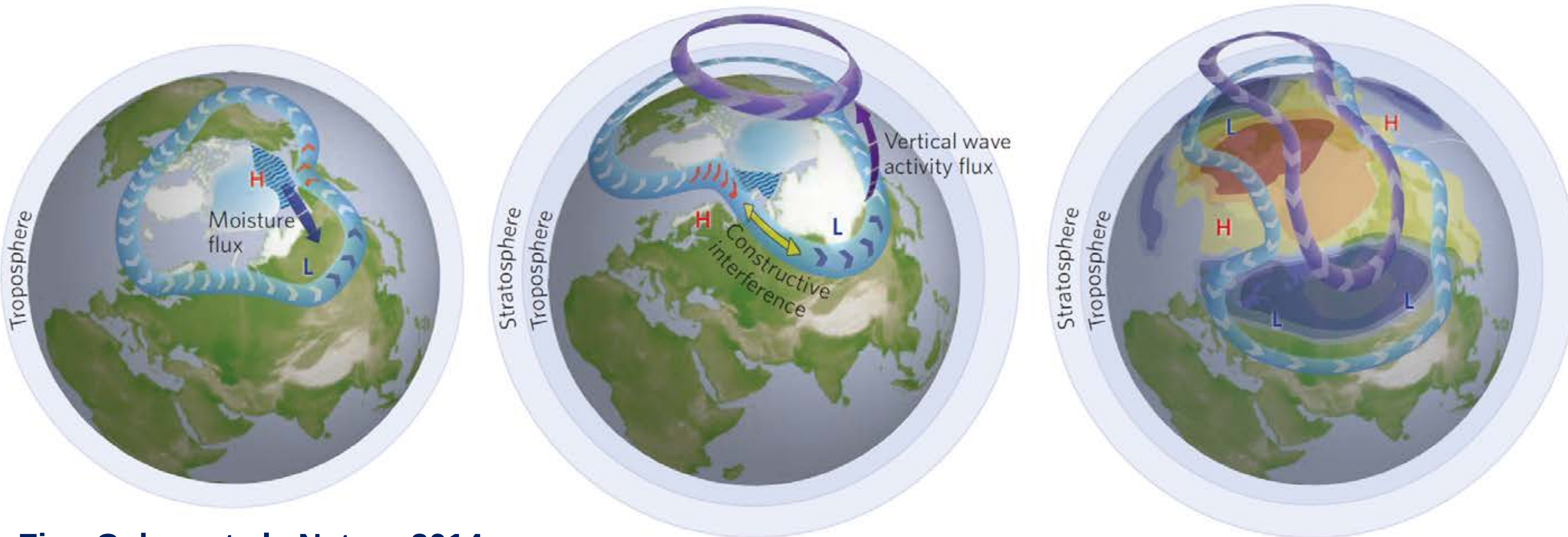


Fig.: Cohen et al., Nature 2014

Sea ice retreat

- Vertical heat- and moisture fluxes
- Increased baro-clinic instability (cyclones)
- Increase in snow cover over Sibiria

Forcing of planetary waves

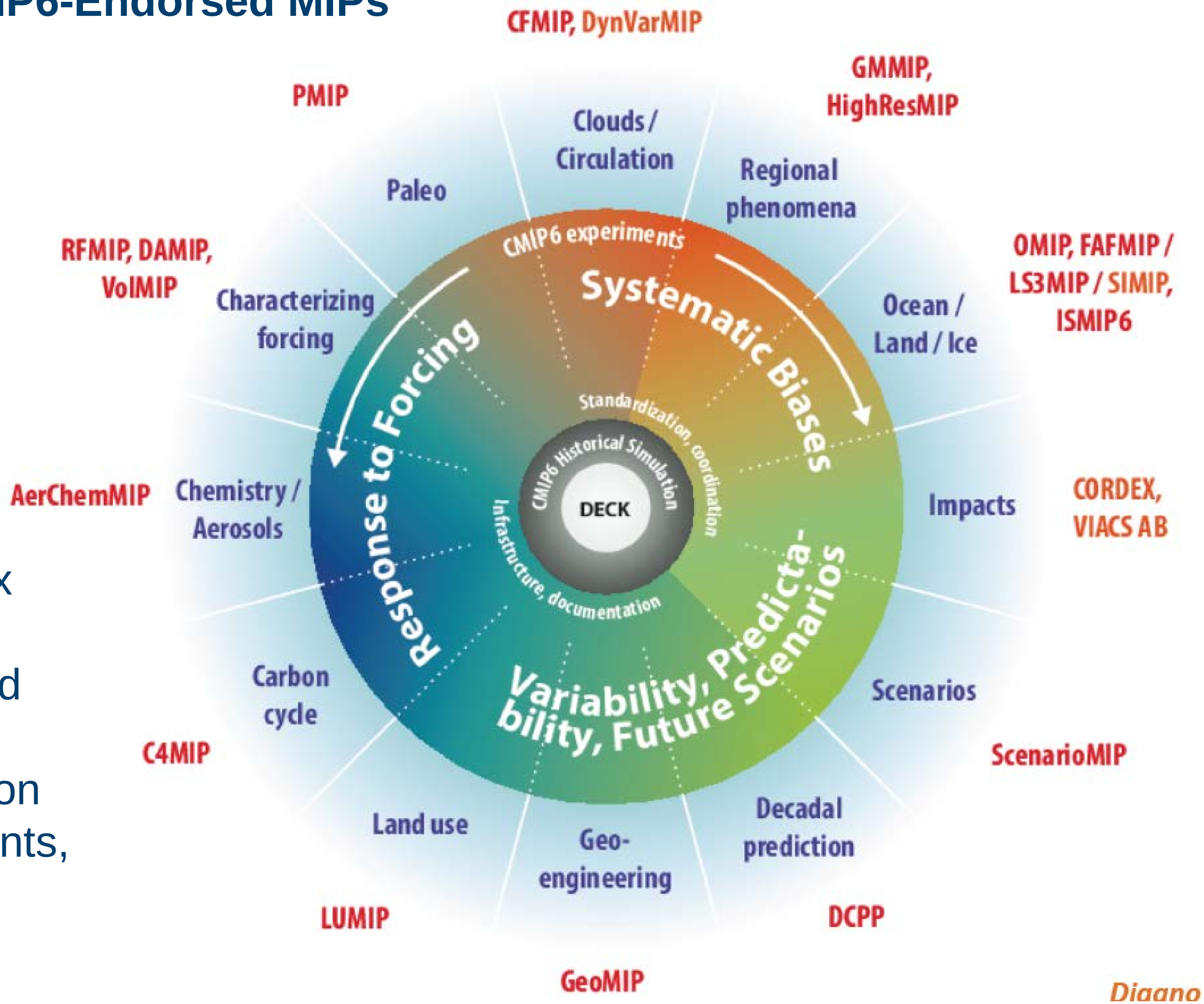
- Interactions between planetary and synoptic waves
- Diabatic forcing due to
  → changes in snow cover
  → ice anomalies in Nov.
- Decreased meridional temperature gradient

Enhanced planetary waves

- Enhanced vertical wave-propagation up to the stratosphere (EP-fluxes)
- Disturbance of stratospheric polar vortex
- Downward propagating signal
- negative NAO
  → colder European winter

Outlook: Big data and climate modelling – CMIP6

21 CMIP6-Endorsed MIPs



CMIP6 has a more complex experiment structure
→ 32 institutes (groups) and many model versions
→more with higher resolution
→21 MIPs, many experiments, larger ensembles
→ Expectations:
Volume: 150 PB
Number of files: 280 Mio

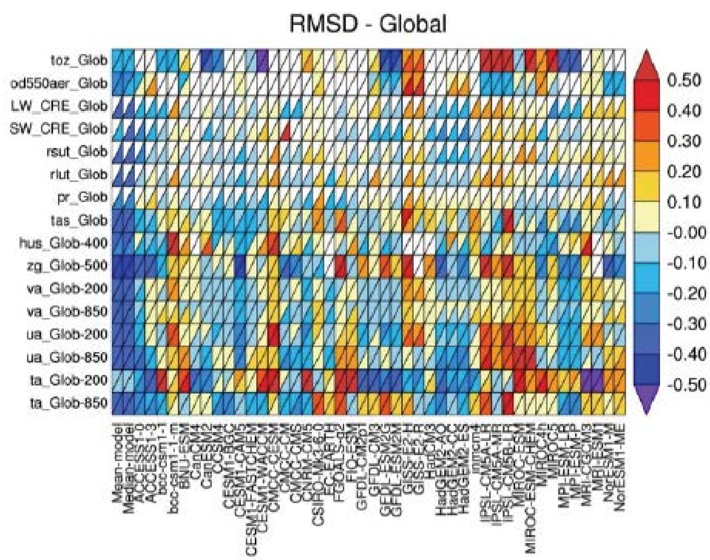
Eyring et al., Geosci. Mod. Develop., 2016

Outlook: Big data and climate modelling – CMIP6

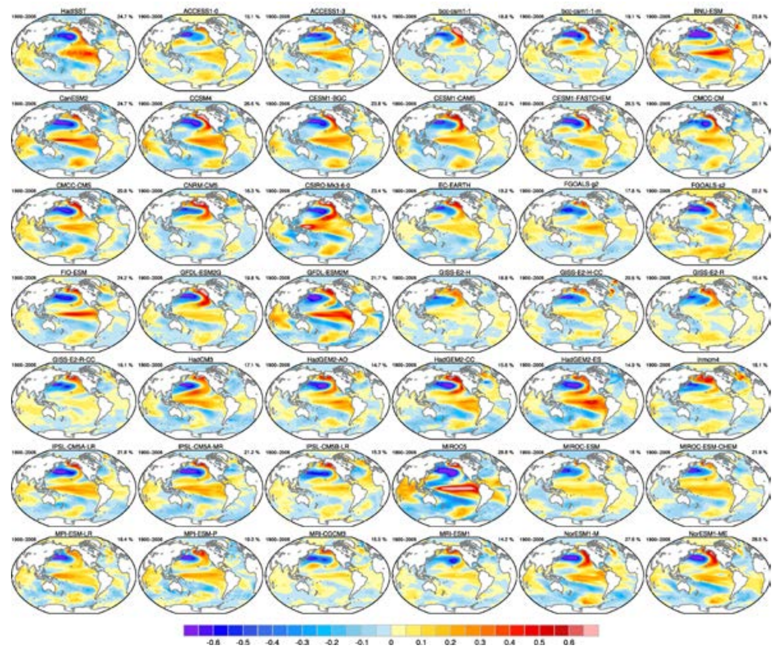
How to characterize the wide variety of models in CMIP6?
→ Routine Benchmarking and Evaluation Central Part of CMIP6 -

- **Evaluation tools** are provided such as
  the Earth System Model Evaluation Tool (ESMValTool, *Eyring et al., 2016)*
  the NCAR CVDP *(Phillips et al., 2014*)
  the PCMDI Metrics Package (PMP, *Gleckler et al., EOS, 2016)*
- will produce well-established analyses as soon as CMIP model output is submitted

Broad Characterization of Model Behavior



Rel. space–time root-mean square error calculated from the 1980–2005 climatological seasonal cycle of the CMIP5 historical simulations. Blue/red shading indicating performance being better/worse than the median of all model results.



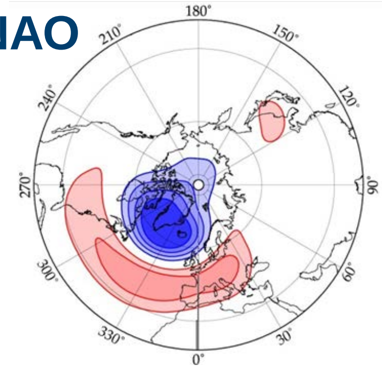
Pacific-Decadal Oscillation (PDO), 41 CMIP5 models and observations (upper left panel) for 1900–2005.

Open questions: Big data & data sciences in climate sciences

Hypothesis-driven Approach (with some data science)

Step 1: Analysis of changes – Evaluation of climate models
Step 2: Development of hypothesis
Step 3: Provision of evidence
→ New model experiments
→ New analysis

NAO



Current Technological Approach

- Download of data from data centers
- Data analysis locally
- Software packages (MATLAB, R)
- Own software (FORTRAN, R) → use of libraries (NAG)

- Our current technological approach for climate data analysis will be probably not applicable for CMIP6 and other future modelling activities
  → How can we benefit from routine benchmarking and evaluation within CMIP6?
  → How can we perform data analysis remotely given special software needs?
  → How to reduce the analytical bottleneck in scientific data analysis?
  → How to visualize results (large ensembles)?
- **There is a need for theory-guided/hypothesis-driven data science methods that blend the power of big data analytics with the caution of scientific theory and first principles.** (Faghmous & Kumar, 2014)