Assimilation of PAGES2k temperature reconstructions using GCM ensemble member selection

UNIVERSITY OF BIRMINGHAM

Martin Widmann, Anastasios Matsikaris
University of Birmingham

Johann Jungclaus
MPI for Meteorology, Hamburg

PAGES2k-PMIP3 workshop

Lamont Doherty Earth Observatory, 3. June 2016

Content

1) Short overview on data assimilation (DA) approaches

2) Assimilation of PAGES2k temperature reconstruction using GCM ensemble member selection. Lessons learned on

   - information propagation on decadal timescales (online vs offline DA)

   - information propagation to sub-continental spatial scales

3) Comments on further development of DA (including for hydrological variables)

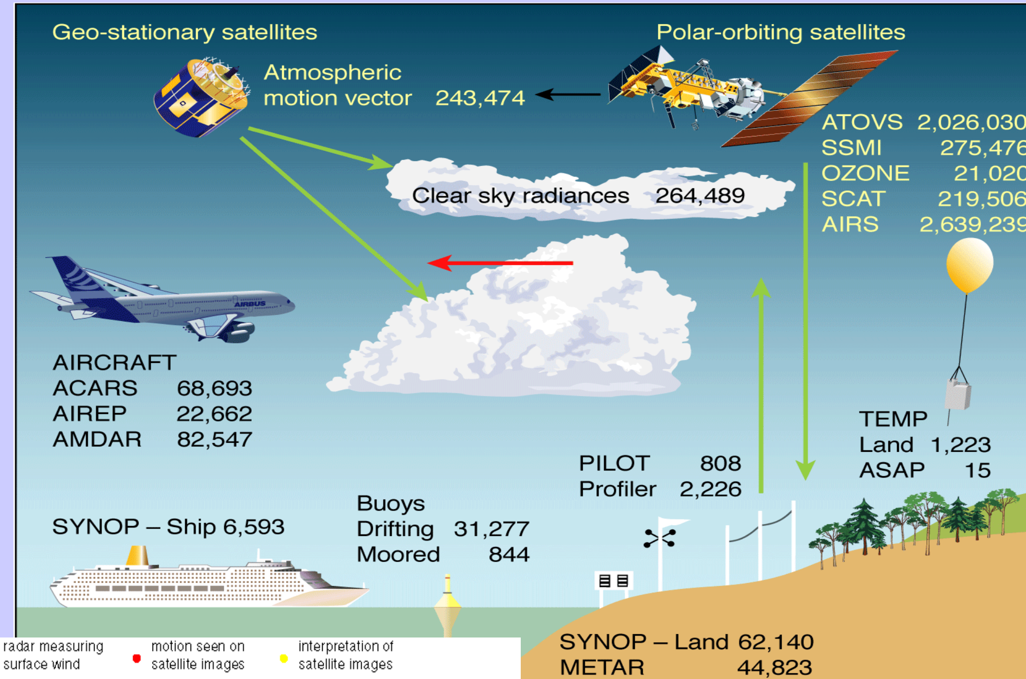
DA in weather forecasting and for atmospheric reanalyses

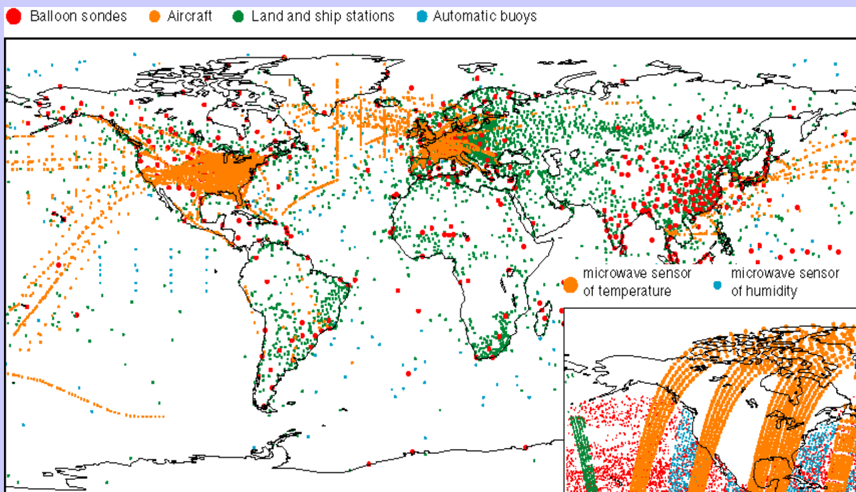
DA used to

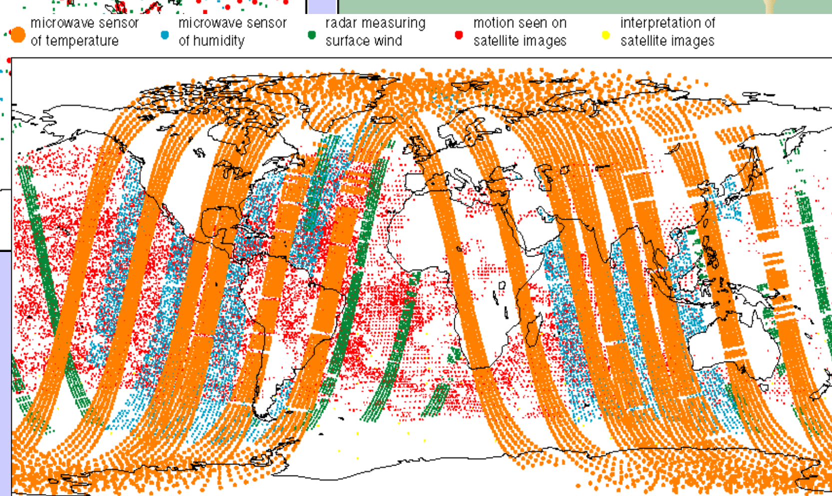
- define the initial conditions for weather forecast

- reconstruct atmospheric states for (second half of) 20th century (NCEP, ERA reanalyses)

Observations assimilated at ECMWF over 24 hours on 13 Feb. 2006





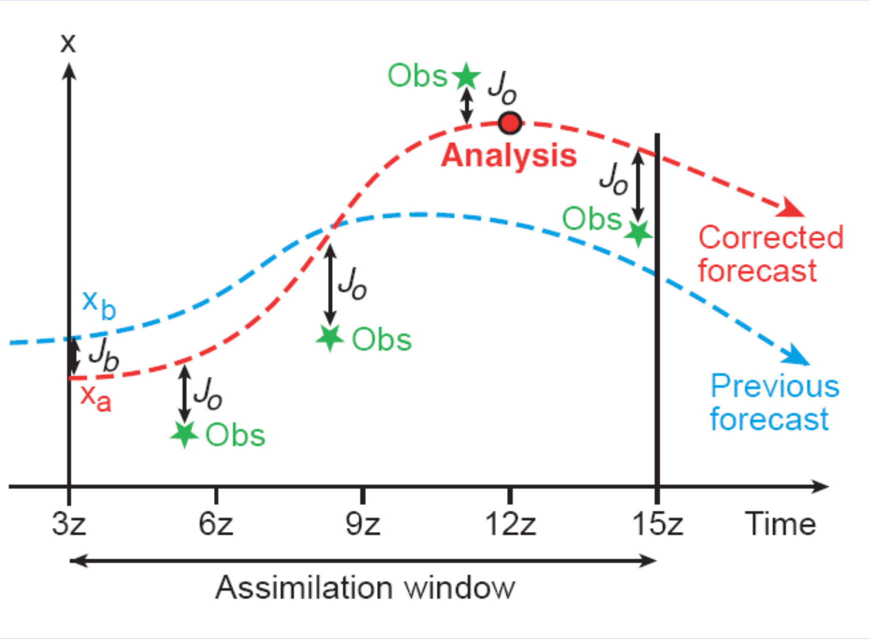


(courtesy ECMWF)

Variational DA in meteorology

The optimal analysis x^a is defined by the nonlinear least squares problem

$$\min \; J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)$$
$$+\frac{1}{2}\sum_{i=0}^{n}(H_i[\mathbf{x}_i] - \mathbf{y}_i^o)^T \mathbf{R}_i^{-1}(H_i[\mathbf{x}_i] - \mathbf{y}_i^o)$$



subject to

$$\mathbf{x}_i = S(t_i, t_0, \mathbf{x}^a)$$

(true states follow model equations S)

$\mathbf{x}_b$ - Background state (simulation)

$\mathbf{y}_i$ - Observations

H_i - Observation operator (or forward model, can be non-linear)

$\mathbf{B}$ - Background error covariance matrix

$\mathbf{R}_i$ - Observation error covariance matrix

Solved using adjoints, needs good linear approximation of dynamical system

Sequential data assimilation and Kalman Filter

If we base the analysis x_k^a at time k only on observations at time k and on background fields x_k^b at this time the analysis solves

$$\min \quad J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_k^b)^T \mathbf{B}_k^{-1}(\mathbf{x}-\mathbf{x}_k^b)$$
$$+\frac{1}{2}(H_k[\mathbf{x}]-\mathbf{y}_k^o)^T \mathbf{R}_k^{-1}(H_k[\mathbf{x}]-\mathbf{y}_k^o)$$

Approximate solution (exact for linear system) is given by

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{B}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{B}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1} (\mathbf{y}_k^o - H_k \mathbf{x}_k^b)$$

(analysis = forecast + weight * (observation – observation estimated from forecast)

where $$\mathbf{H}_k = \frac{\partial H_k}{\partial x}$$ is the linearised observation operator

(following Swinbank et al. 2002)

Data assimilation for the climate of the last millennium

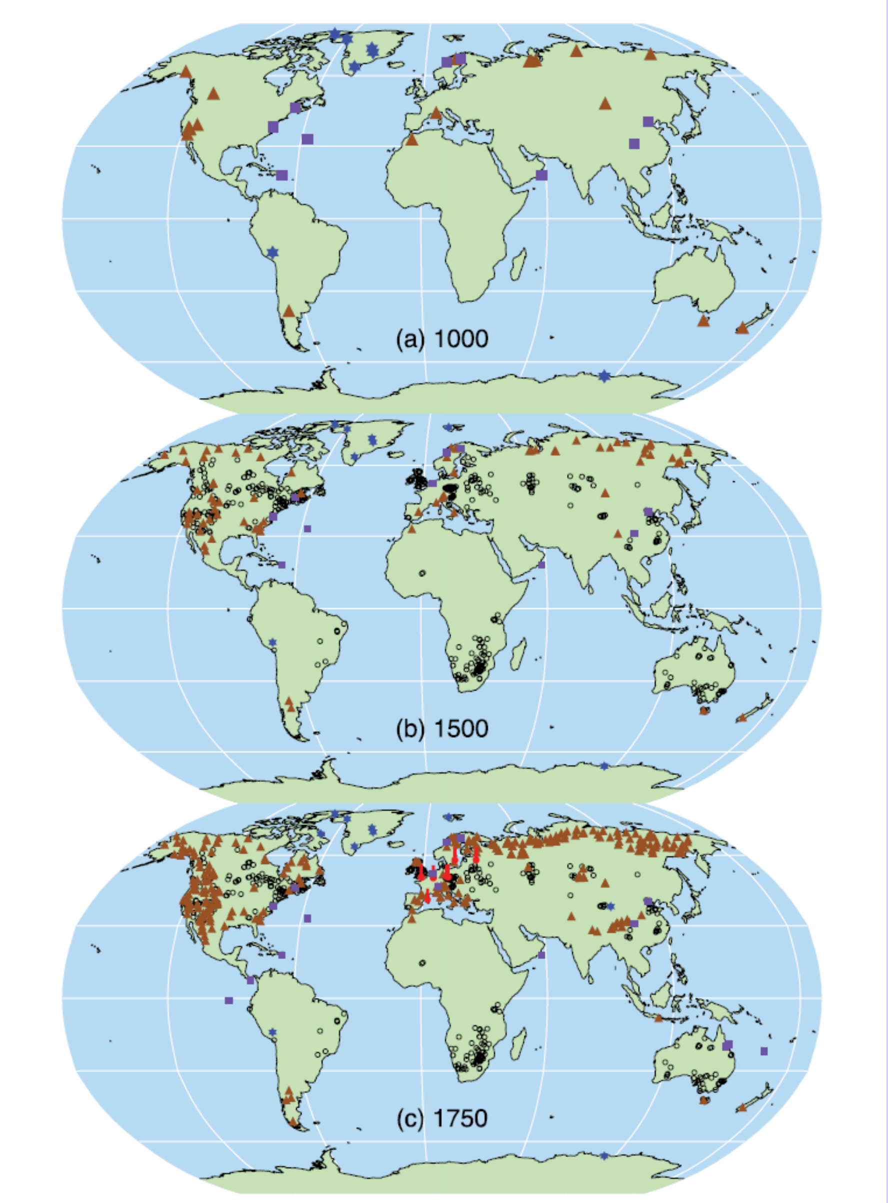
Challenge because empirical estimates constrain only

- a few locations or large-scale patterns (i.e. a low-dimensional subspace)
- seasonal and longer variability

Using standard assimilation methods is not straightforward, because

- methods need to be efficient enough for long simulations
- (model and proxy errors unknown)
- technical/mathematical problems with observations integrated over long periods (e.g. linearisations and adjoints)

Proxy sites back to 1000/1500/1750 AD



Data assimilation for the climate of the last millennium

Approach 1

Use EMIC or GCM ensemble simulations and chose ensemble members consistent with proxy evidence for temperature.

(Goosse et al. 2006, 2010, 2012.; Crespin et al. 2015, Bhend et al. 2012, Matsikaris et al. 2015, 2016a/b)

Approach 2

Prescribe atmospheric circulation with target states based on proxy evidence or idealized states

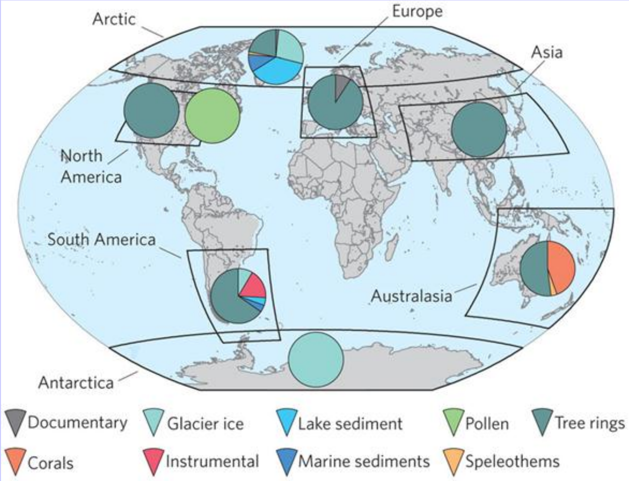
- **forcing singular vectors with EMIC (van der Schrier et al. 2005, 2007)**

- **pattern nudging with GCM (Widmann et al. 2010)**

Approach 3

Ensemble Kalman filter based on stationary ensemble (Steiger et al. 2015, 2016)

Data assimilation with ECHAM6/MPI-OM



Model: ECHAM6/MPI-OM (T31L31, MPI-ESM-CR))

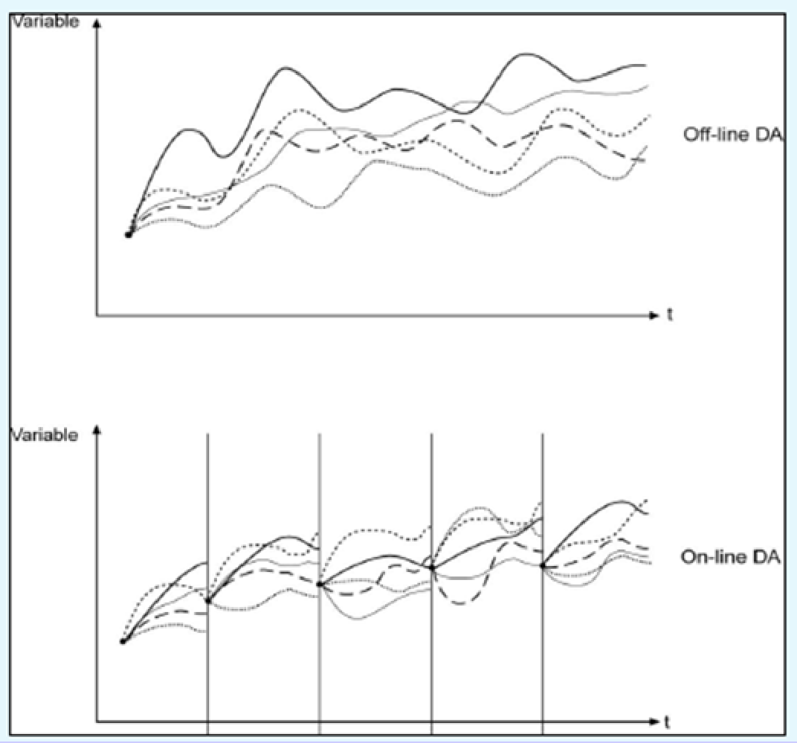
Data: PAGES2k NH continental reconstructions

Method:

- **ensemble member selection, on-line and transient off-line**

- **cost function evaluated for decadal means**

$$CF(t) = \sqrt{\sum_{i=1}^{k} \left(T^{i}_{mod}(t) - T^{i}_{prx}(t)\right)^{2}}$$

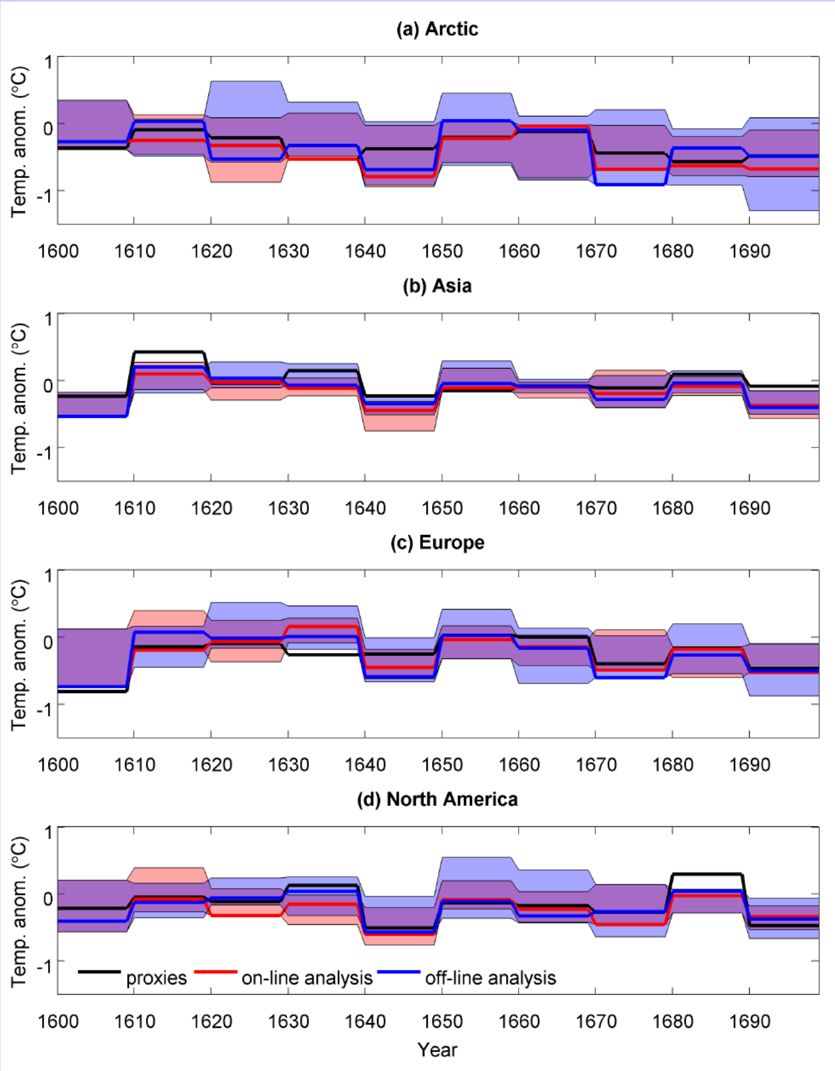


Experiments:

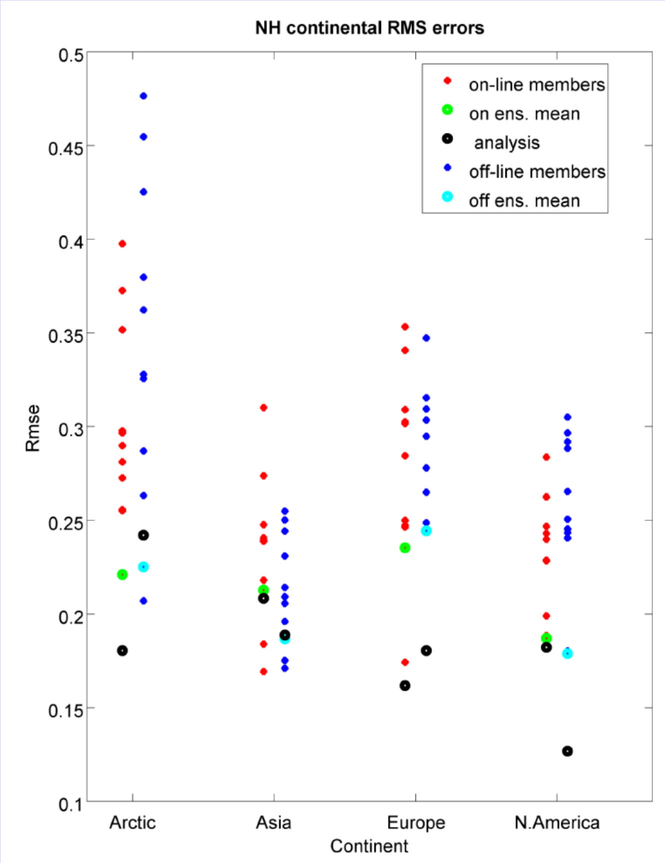
- **1600 - 1700 AD (10 ensemble members)**
- **1750 - 1950 AD (20 ensemble members)**

(Matsikaris, Widmann and Jungclaus, Climate of the Past, 2015, 2016; Clim Dyn. 2016)

Online vs offline ensemble member selection with ECHAM6/MPI-OM



(Matsikaris, Widmann and Jungclaus, Climate of the Past, 2015)



similar skill of on-line and off-line DA

no useful information propagation (IP) on decadal timescales; potential reasons:

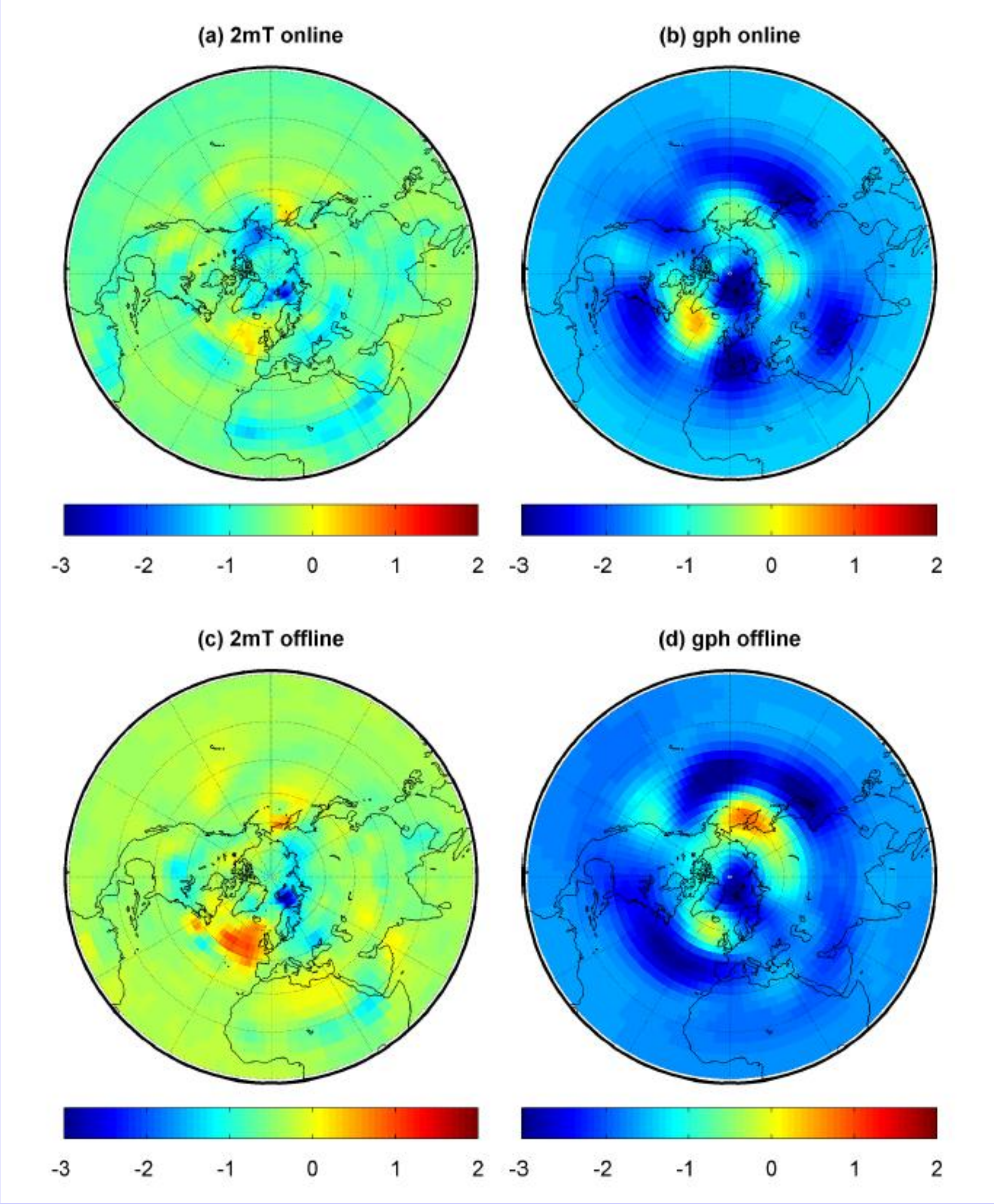
- there is no IP in the model (and reality)
- there is IP, but it is wrong
- initial conditions are wrong, ocean state is not confined

Data assimilation for palaeoclimate with ECHAM6/MPI-OM

1640 – 1649 AD
2m temperature and 500 hPa gph anomalies wrt 1961-1990 AD

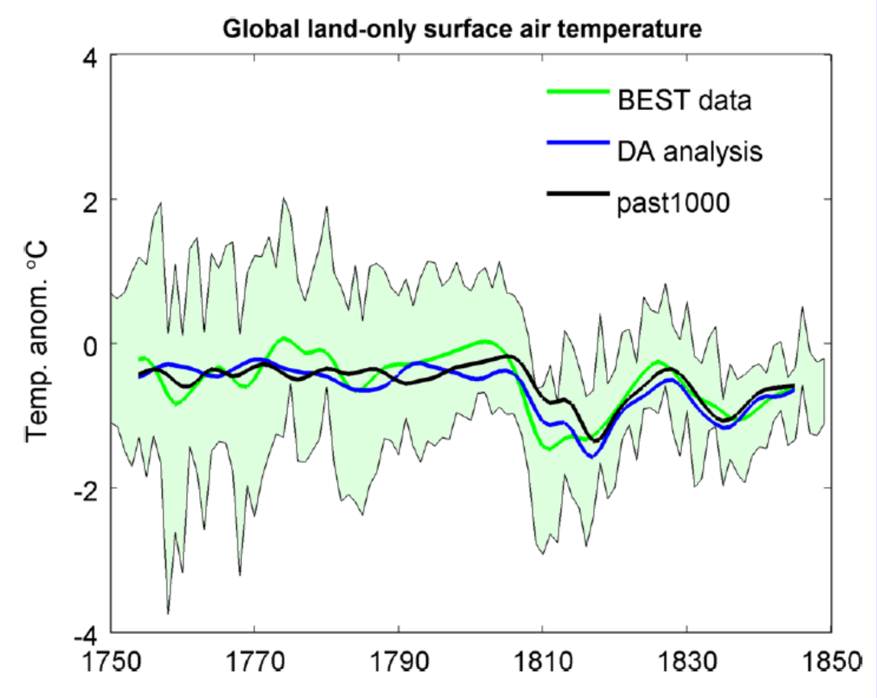
Assimilation of PAGES 2k continental temperature anomalies

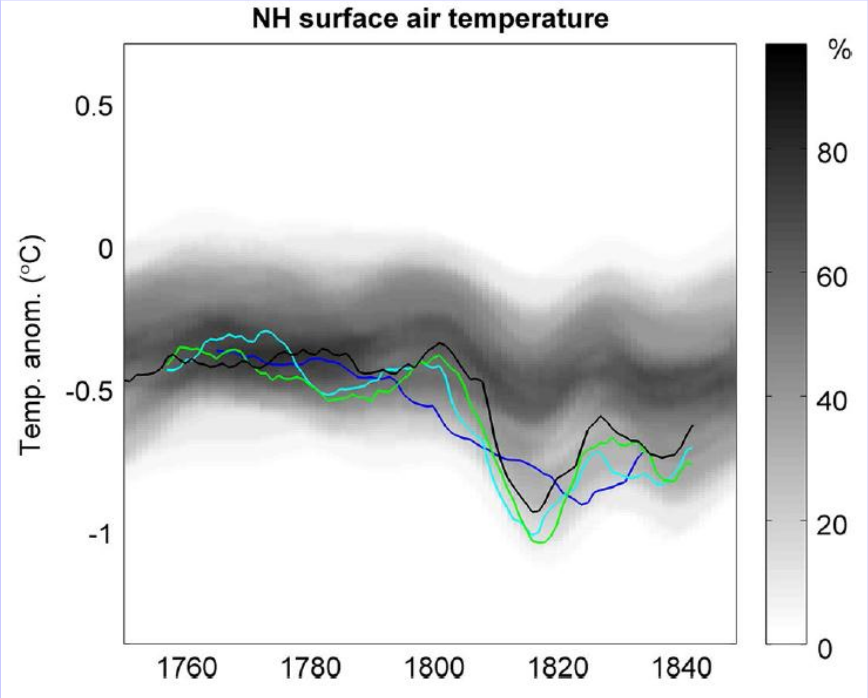
Empirical reconstruction for NAO index (Luterbacher): - 0.28



(Matsikaris, Widmann and Jungclaus, Climate of the Past, 2015)

Global and northern hemispheric temperatures 1750 - 1850 in DA with ECHAM6/MPI-OM





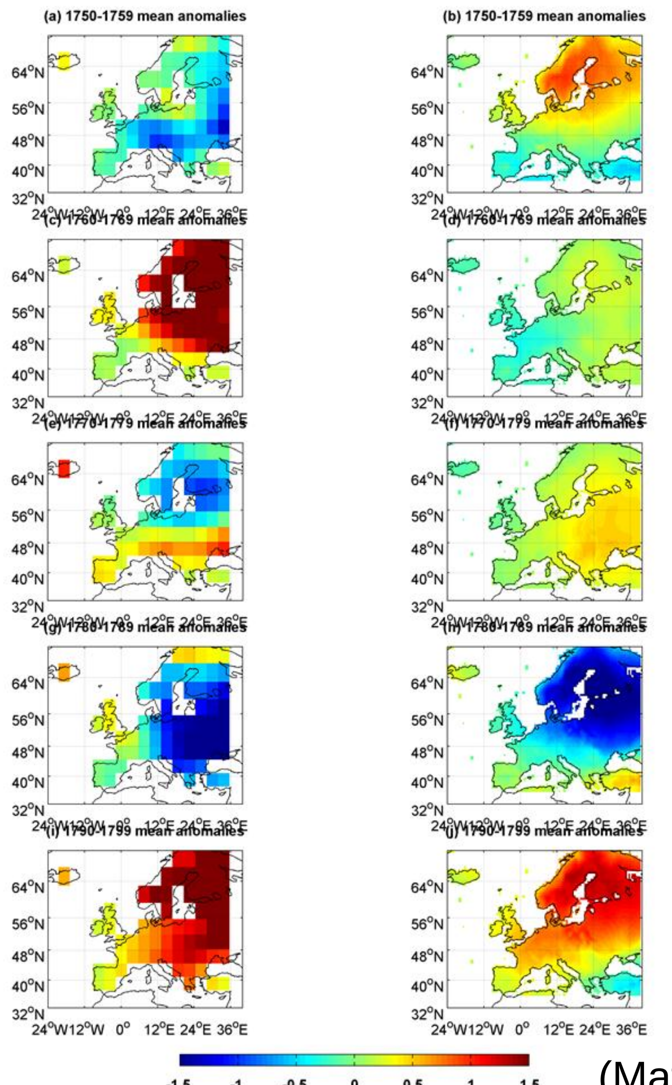
Standard forced and DA simulation are similar, forced variability dominates

Consistent with reconstructions

(Matsikaris, Widmann and Jungclaus, Climate Dynamics, 2016)

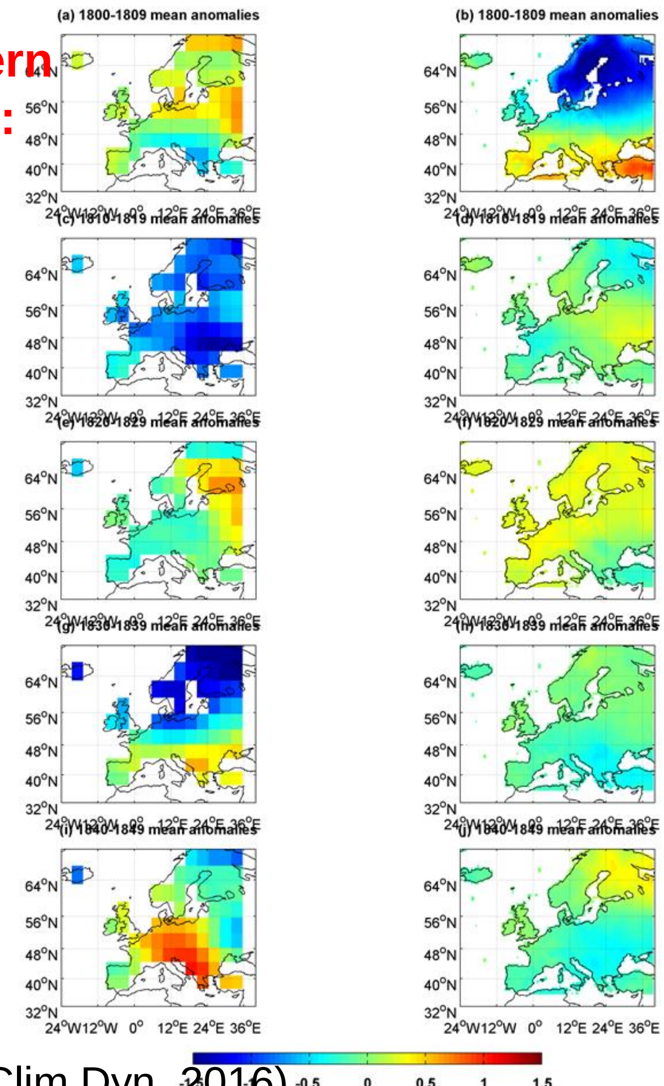
Decadal mean winter temperatures 1750 - 1850 simulated (DA) and empirical reconstruction (Luterbacher)

Simulated **Luterbacher**



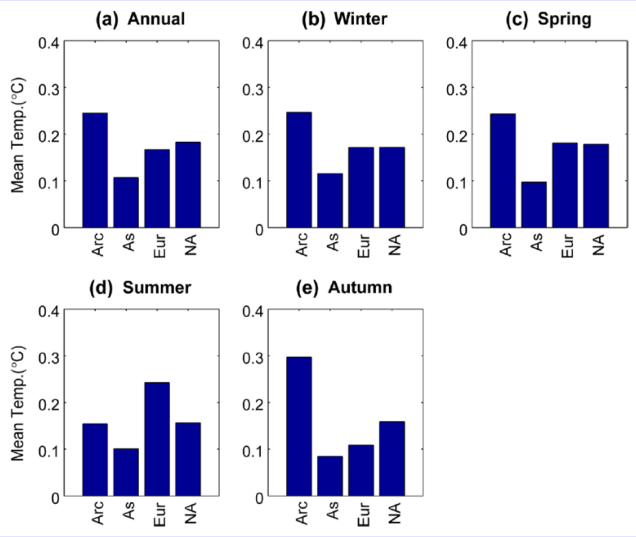
mean pattern correlation: - 0.03

Simulated **Luterbacher**



(Matsikaris et al., Clim Dyn. 2016)

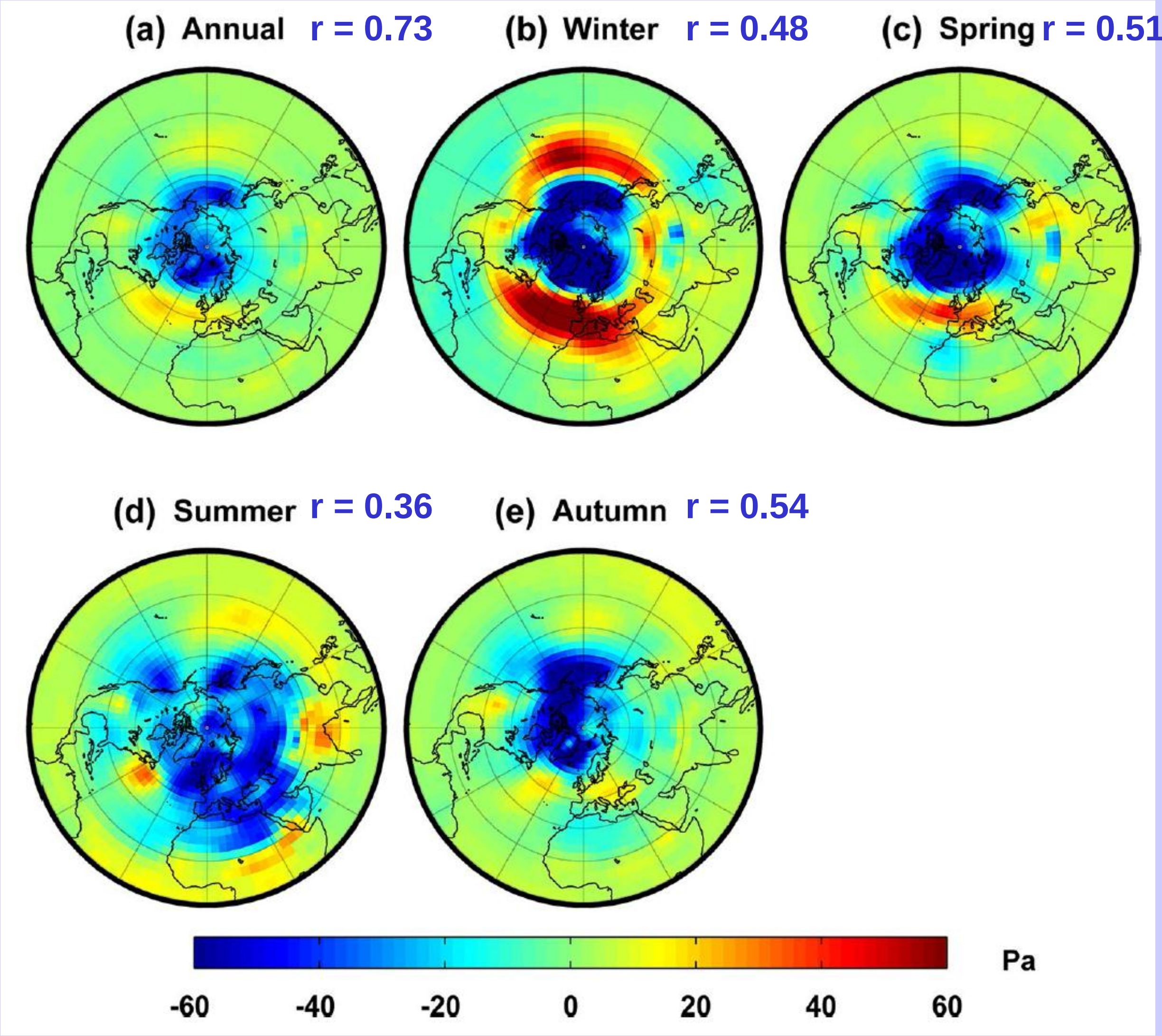
Coupling (MCA) between temperature for NH PAGES2k regions/seasons and SLP in 1000 year GCM control run



Temp: Same sign in
all continents

SLP: NAM structure for
annual, winter, spring

Links are reproduced with DA for annual, winter and spring (r = 0.81, 0.82, 0.82), but not for summer and autumn (0.17, -0.01)



(Matsikaris, Widmann and Jungclaus, Climate Dynamics, 2016)

EOF 1 of NH SLP in ECHAM6/MPI-OM 1000 year control run

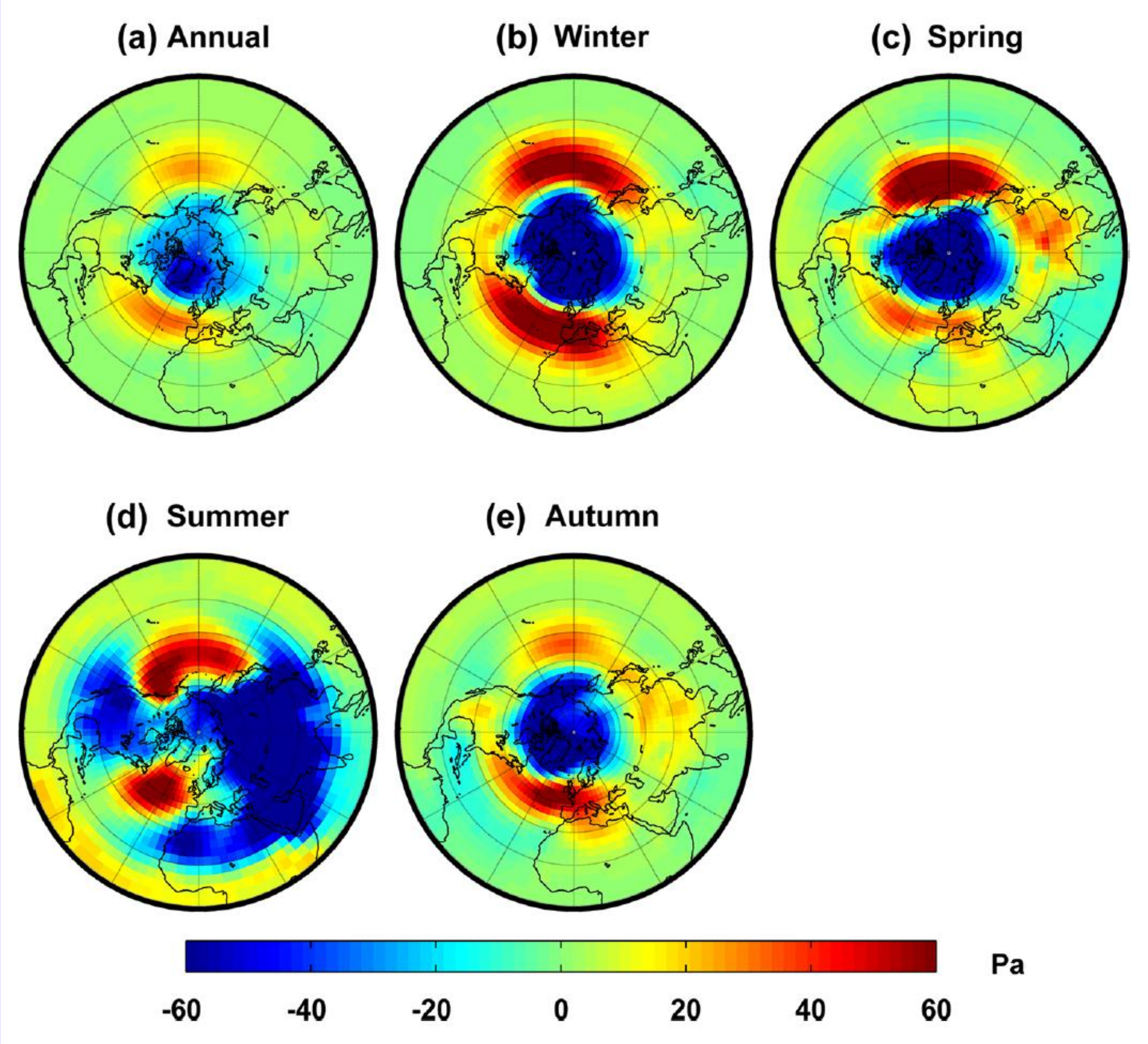
similar to NAM for annual and all seasons but summer

land – sea contrast in summer

similar to MCA patterns for annual, winter, and spring

DA reproduces the link when the MCA pattern is similar to EOF 1

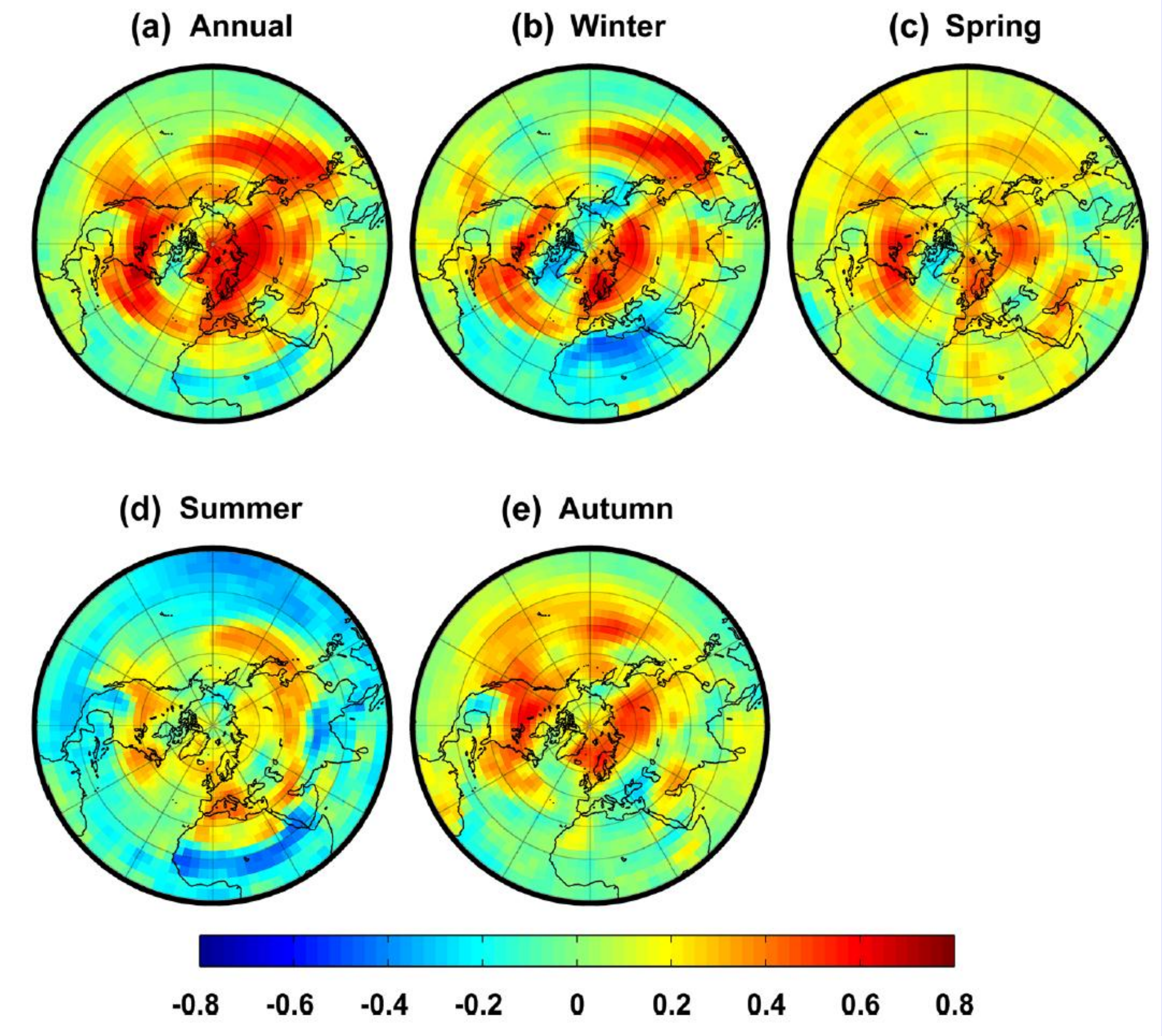
Even then there is still substantial unexplained variability in SLP TEC and in local temperature (given correct SLP TEC)



(Matsikaris, Widmann and Jungclaus, Climate Dynamcs, 2016)

Correlations SLP TEC1 with local temperatures in ECHAM6/MPI-OM 1000 year control simulation

Correlations in Europe are typically between 0 and 0.6

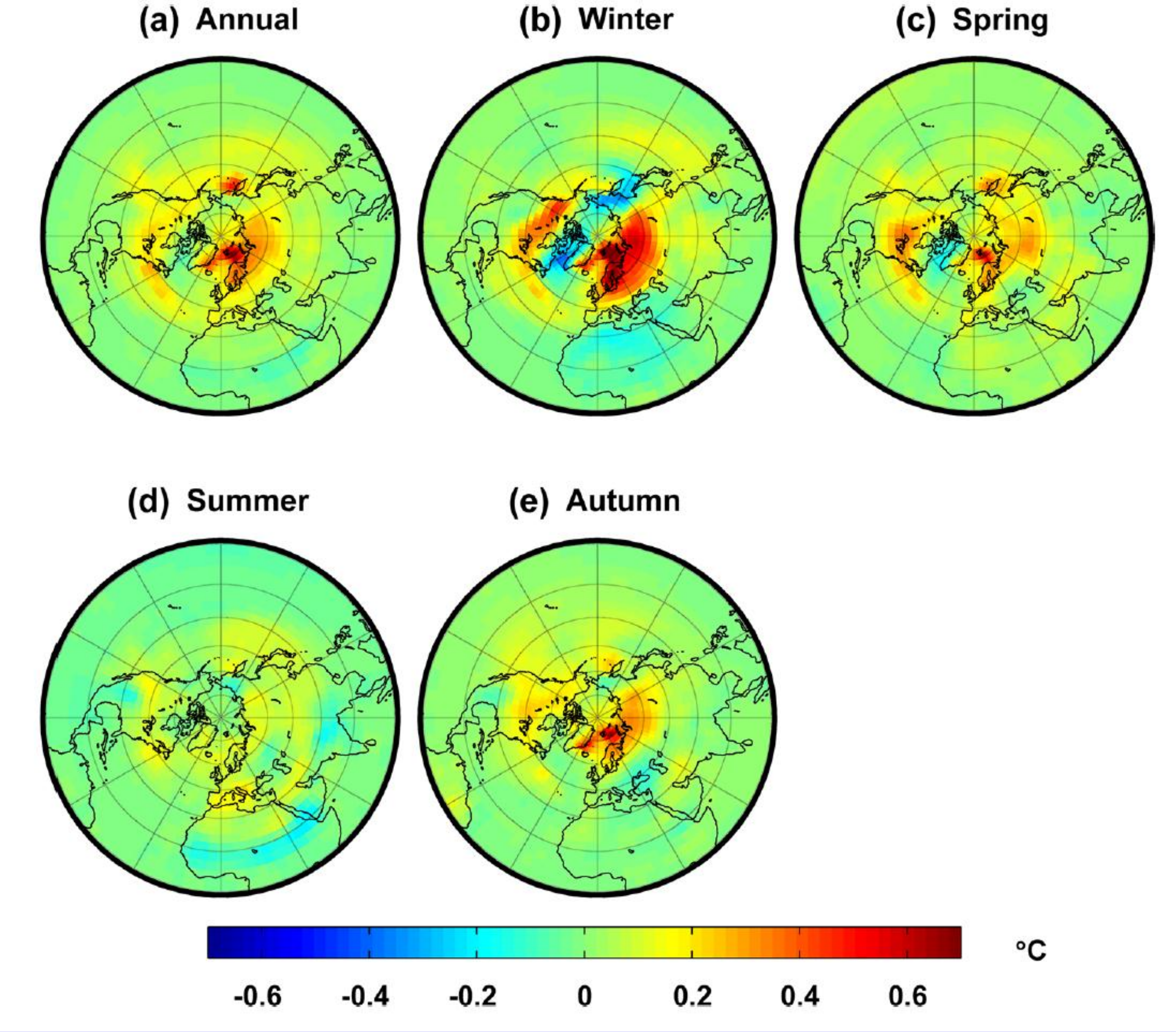


(Matsikaris, Widmann and Jungclaus, Climate Dynamics, 2016)

Regression coefficients local temperatures against SLP TEC1 in ECHAM6/MPI-OM 1000 year control simulation

Not uniform within PAGES2k regions

Better estimates for the SLP MCA patterns (e.g. annular modes) could be explained by using more regional temperatures.



(Matsikaris, Widmann and Jungclaus, Climate Dynamics, 2016)

Comments and (lots of) questions

Spatial and temporal scales

- State estimation and process undertanding require constraining leading circulation patterns

- Spatial variability in hydrological variables is higher than for temperature

- Assimilation on continental-scales is too coarse, but local scale might be too fine (errors and dimensionality of state space).

  Optimal scale is not known and may depend on variable and DA method. Avoiding upscaling based on teleconnection might be a good idea, but regional averages can be expected to be OK.

- Which temporal resolution should be used for DA?

Comments and (lots of) questions

Variables

- Which are suitable for DA? Isotopes, precipitation, PDSI, others?

- What do we know about errors of reconstructions for the different variables?

- How can we achieve good coverage in all seasons?

DA methods

- Ensemble member selection and/or EnKF? (or others?)

- Is there information propagation in time on the DA timescale?

- What ensemble size is needed? Might depend on method and variable (dimension of state space)

- Stationary online, transient online or offline? Offline can use existing simulations, onlince cannot.