On the use of statistics in complex weather and climate models

Andreas Hense

Meteorological Institute University Bonn

Together with..

- Heiko Paeth (Bonn)
- Seung-Ki Min (Seoul)
- Susanne Theis (Bonn)
- Steffen Weber (Bonn, WetterOnline)
- Monika Rauthe (Bonn, now Rostock)
- Rita Glowienka-Hense

Overview

- Some general remarks concerning complex models of the atmosphere / the climate system and statistics
- Use of statistics in numerical weather prediction
  - ensemble prediction
  - calibration
- Use of statistics in climate change simulations
  - Defining a signal and its uncertainty
  - Detecting a signal in observations

Climate Simulation and Numerical Weather Prediction

- Randomness in the climate system / atmosphere originates from highdimensionality and nonlinear scale interactions
- Randomness in climate models and NWP models arises additionally
  - from parametrizations
  - from model selection and construction

Climate Simulation and Numerical Weather Prediction

- Modelling a high dimensional system requires scale selection in space κ and time τ
- Simulation time $T < \tau$ a NWP / inital condition problem
- $T >> \tau$ climate problem
- Urban/Micro climatology $T \sim 1$ d, $\tau \sim$ min or h
- climate simulations embedded into NWP
- detailed precipitation with $T \sim 10$ d

Climate Simulation and Numerical Weather Prediction

- The deterministic view
  - – e.g. wrong NWP forecast due to model errors
  - – e.g. Any modeled climate change in a climate simulation with perturbed greenhouse gase forcing is due to this external forcing.
- More illustrative:
  - – „We predict in two days advance the sunny side of the street“
  - – „We predict in two days advance which tennis court in Wimbledon will have rain“

Climate Simulation and Numerical Weather Prediction

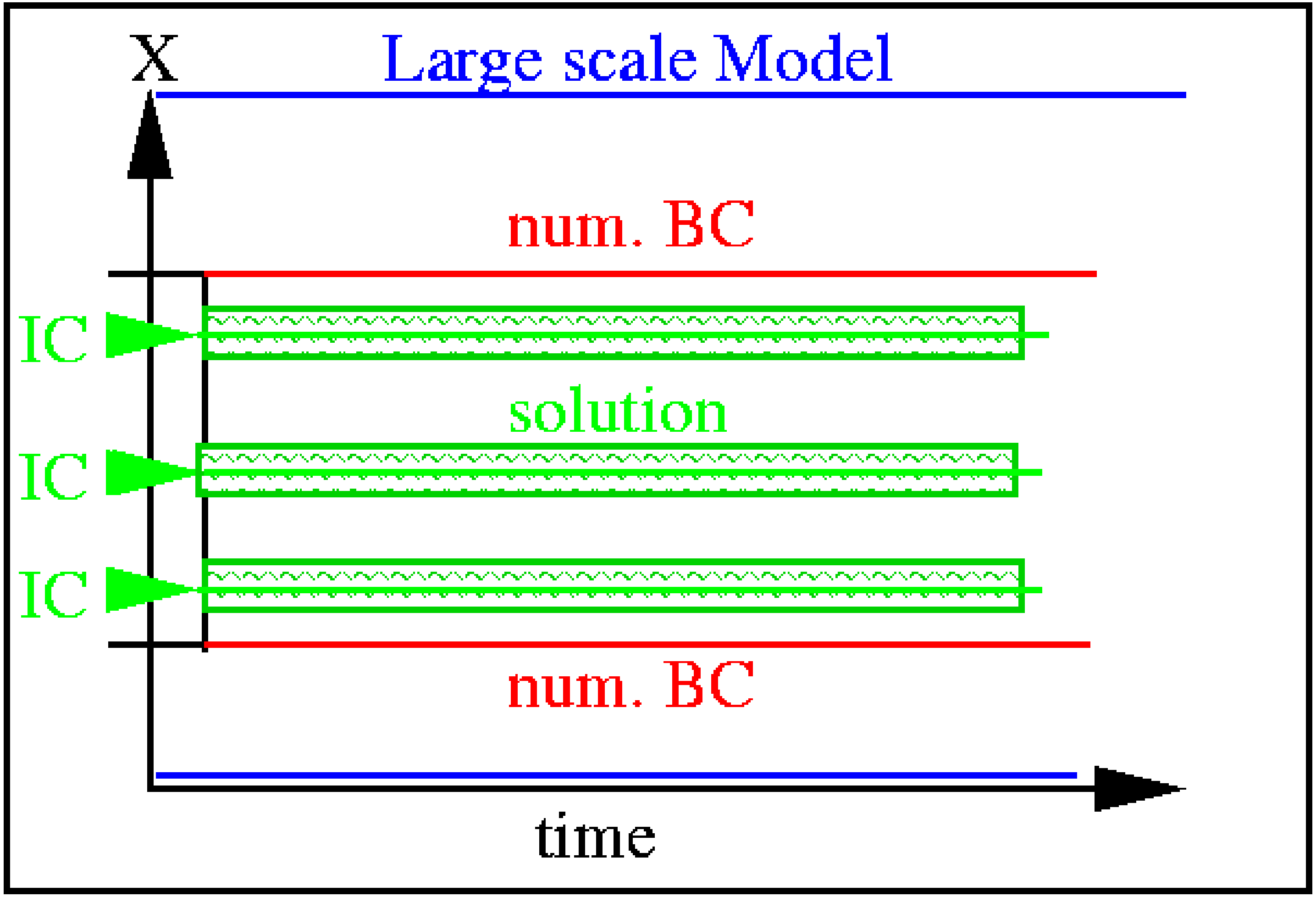
- General formulation of the problem
  - Analysis of the joint pdf of simulations ***m*** and observations ***o***
  - p(***m***|***o***) for model validation and selection
  - description of the observation process, mapping of ***o*** on ***m*** with some unknown parameterset χ
  - maximize p(m, χ | o): calibration, model output statistics MOS

NWP examples

- The generation of model ensemble
  - with precipitation as a (notoriously) difficult variable
  - generation of precipitation is at the end of a long chain of interactions
  - involves scales from the molecular scale up to relevant atmospheric scales 1000 km
  - highly non Gaussian
    - positive definite
    - most probably fat tailed

Generation of NWP ensembles

- Sampling uncertainty in initial conditions
- Sampling uncertainty in boundary conditions
  - physical bc at Earth‘s surface
  - numerical bc
- Sampling uncertainty in parameter constellations
- Using the limited area weather forecast model of the German Weather Service DWD (7km * 7km, 35 vertical layers, 177 * 177 gridpoints)



Numerical weather prediction is a scenario description of future states of the atmosphere

Sampling of parameter uncertainty: NWP models become stochastic models

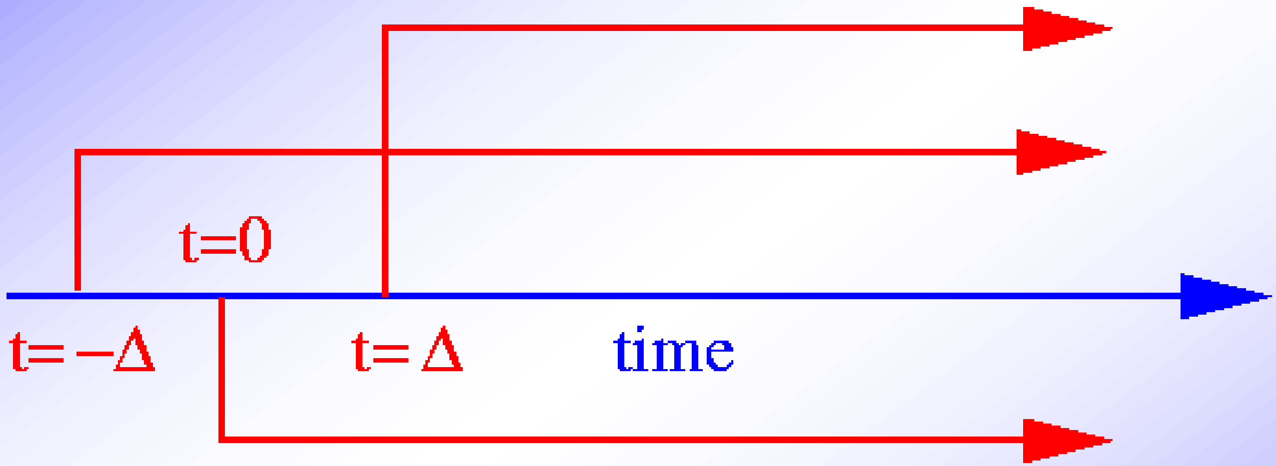
$$H = -D\vec{\nabla}T_{lc}$$

$$\frac{\partial T}{\partial t} \sim -\vec{\nabla}(D\vec{\nabla}T_{lc})$$

$$D = \bar{D} + D'$$

$$D' \in NV(0, \sigma_D)$$

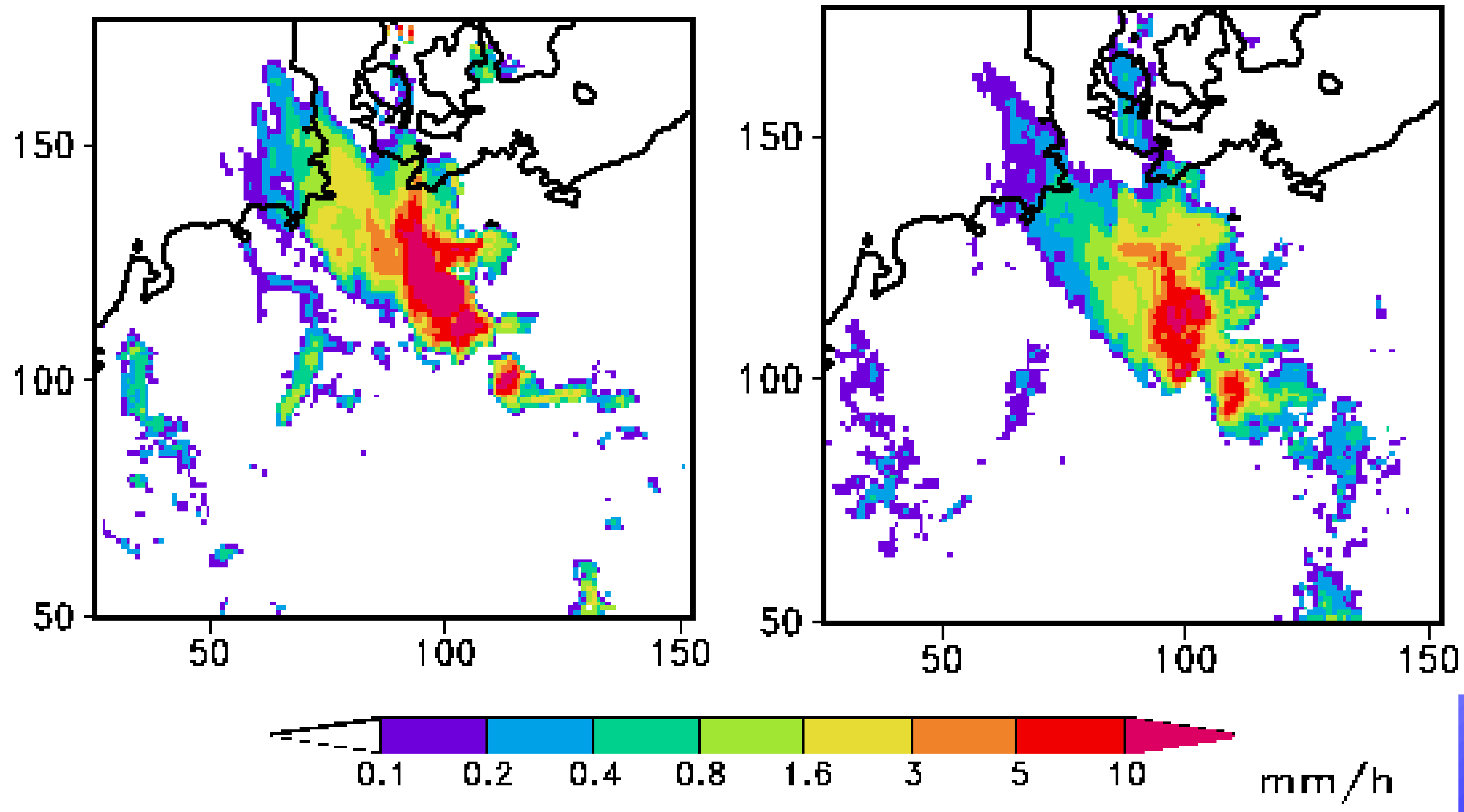
Sampling uncertainty in initial conditions



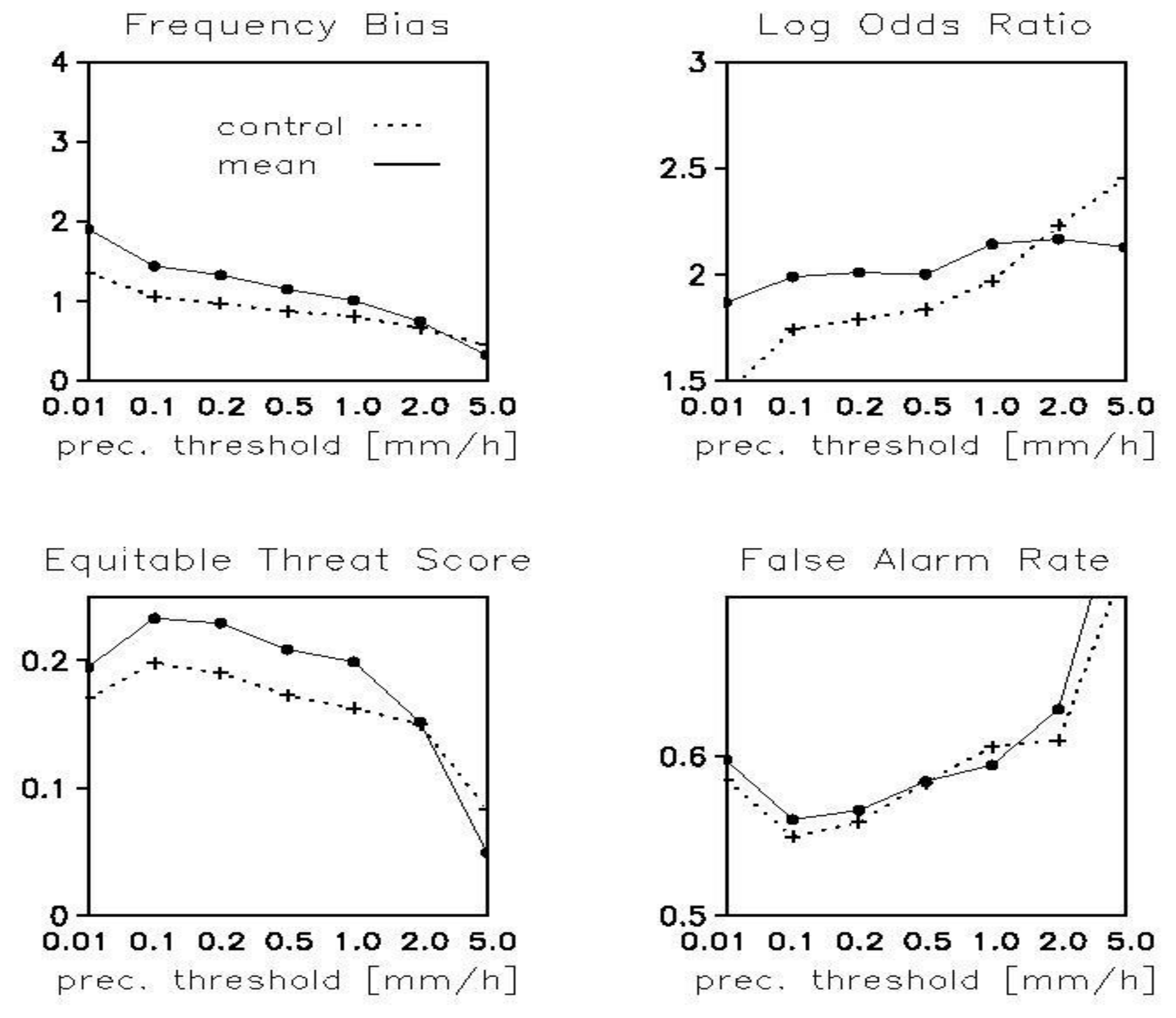
Most probably not a correct sampling !

Deterministic forecast

10 member ensemble std deviation



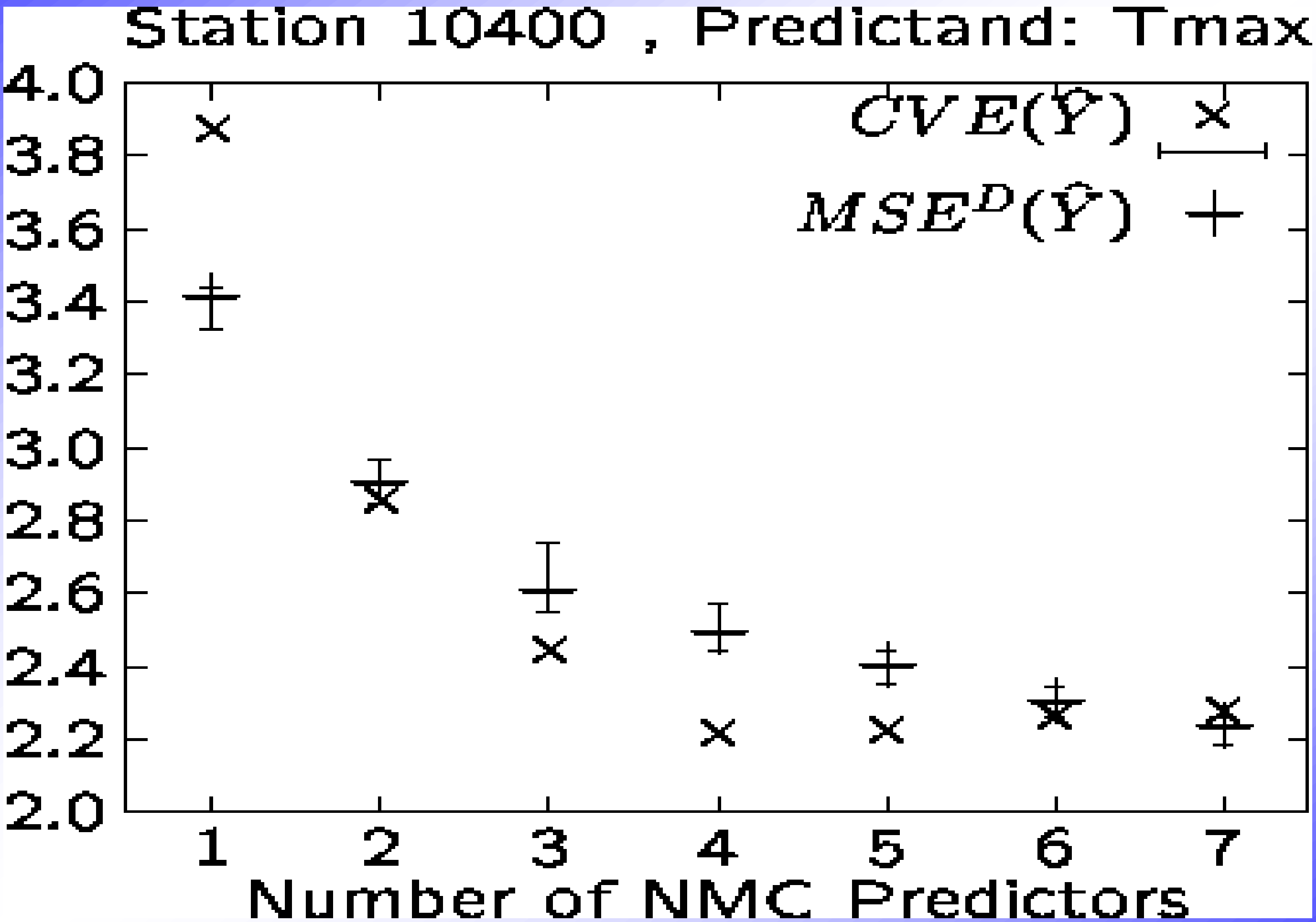
Experimental verification, mean



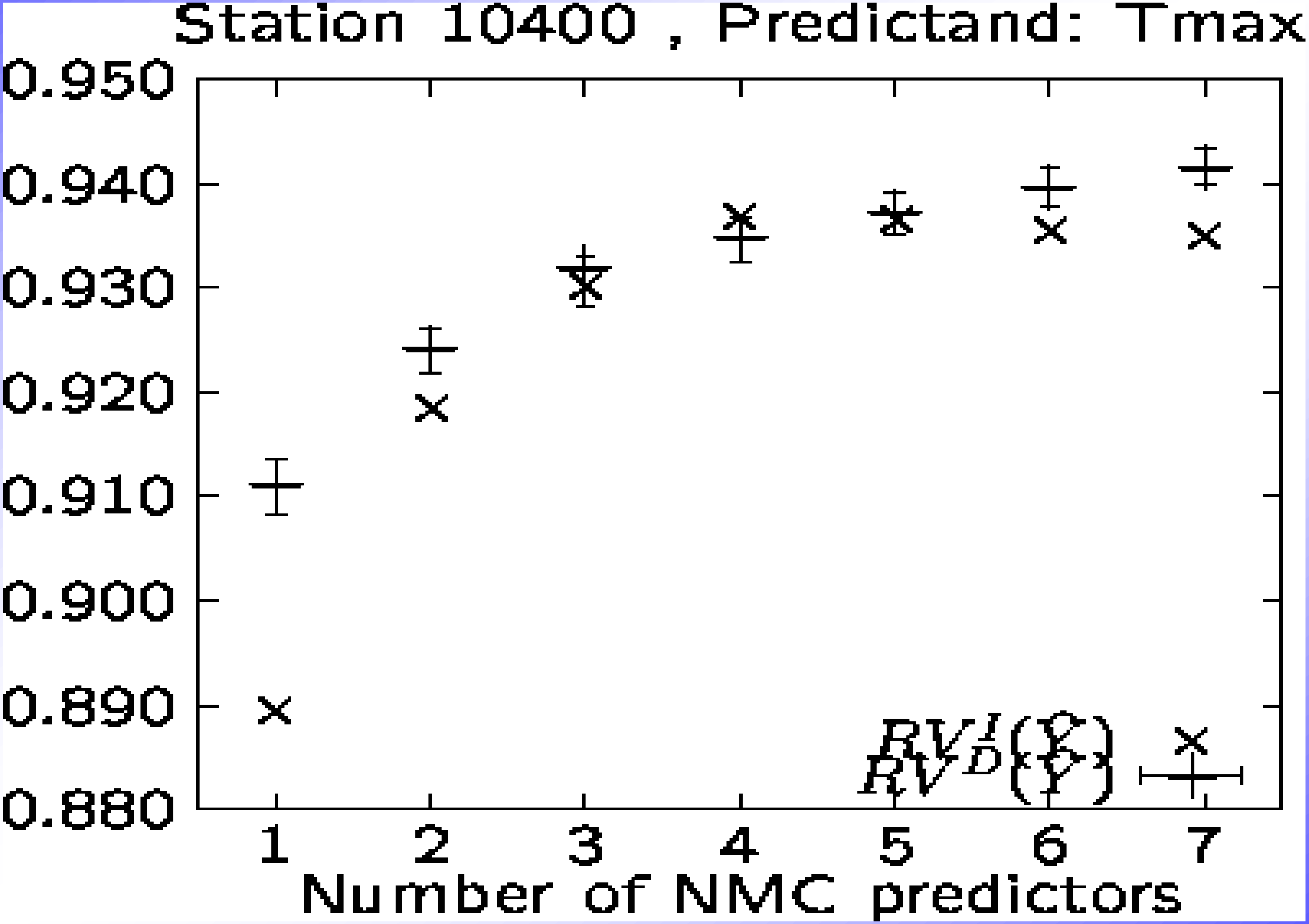
Calibration of weather forecasts MOS

- Weather forecasts NMC on a 1° * 1° grid
- single station observations every three hours
- not a fully developed Bayesian scheme yet
- but
  - multiple correlation with stepwise regression to select large scale predictands
  - and cross validation

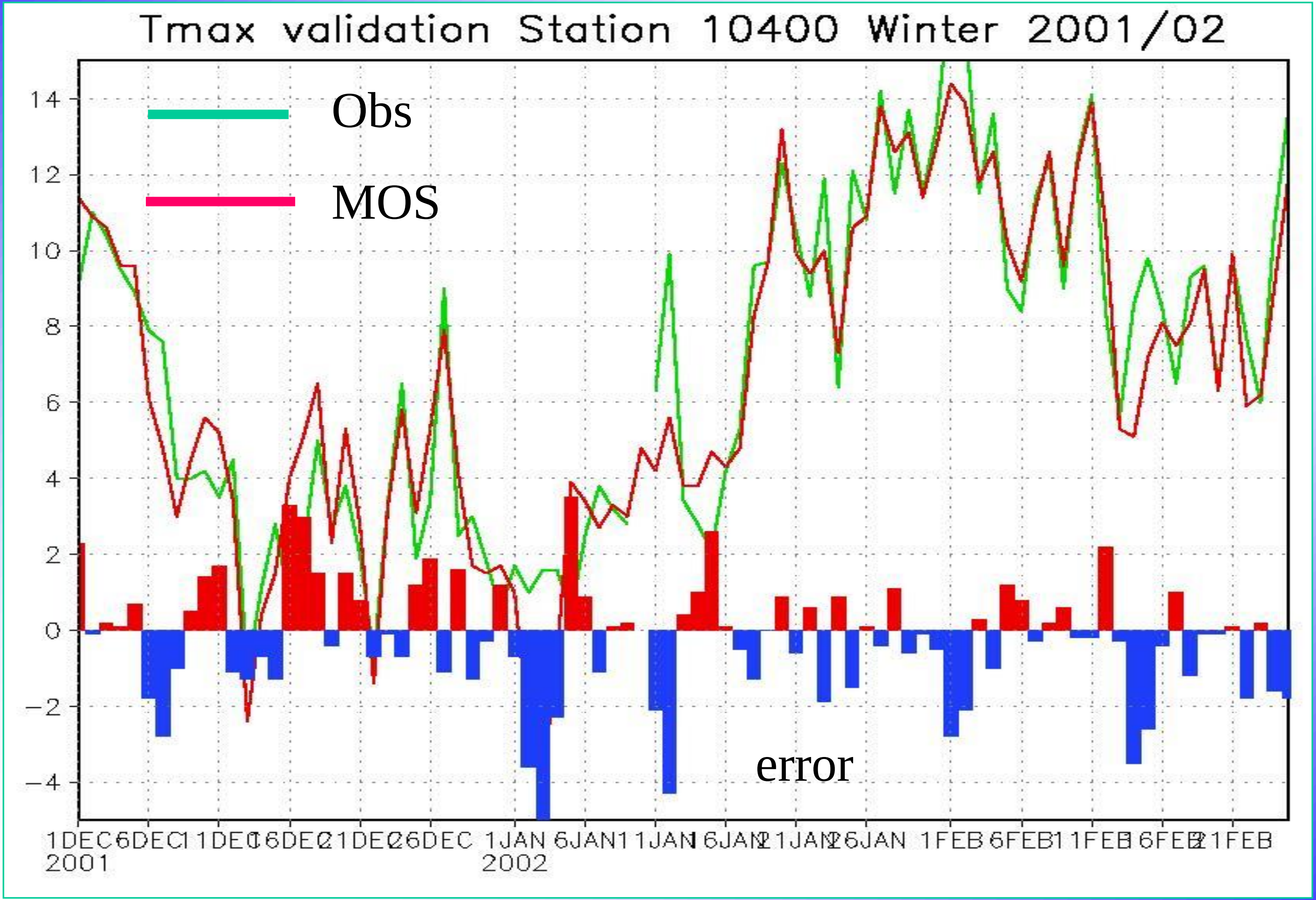
Calibration error statistics mean square error



Calibration error statistics, explained variance



Application: Daily T_{max} Winter 2001/02



Climate change model simulations

- Predicting changes of climate statistics ***p(m,t)*** due to changes in physical boundary conditions
  - changes in ***p(m,t)*** relative to $p(m,t_0)$ due to increasing greenhouse gase concentrations e.g. CO_2 (t) and other anthropogenic forcings
  - changes in ***p(m,t)*** relative to $p(m,t_0)$ due to solar variability, volcanic eruptions (natural forcings)
  - distinguish between anthropogenic and natural forcing effects

Climate change model simulation classical view

- Compare modeled anthropogenic changes with observed changes
  - if projection of observed changes onto modeled changes are larger than an unforced background noise level: reject Null hypothesis of unforced climate variability
  - requires the assumption of a „significant“ model change
  - which time/space scales and variables allow for these significant changes?

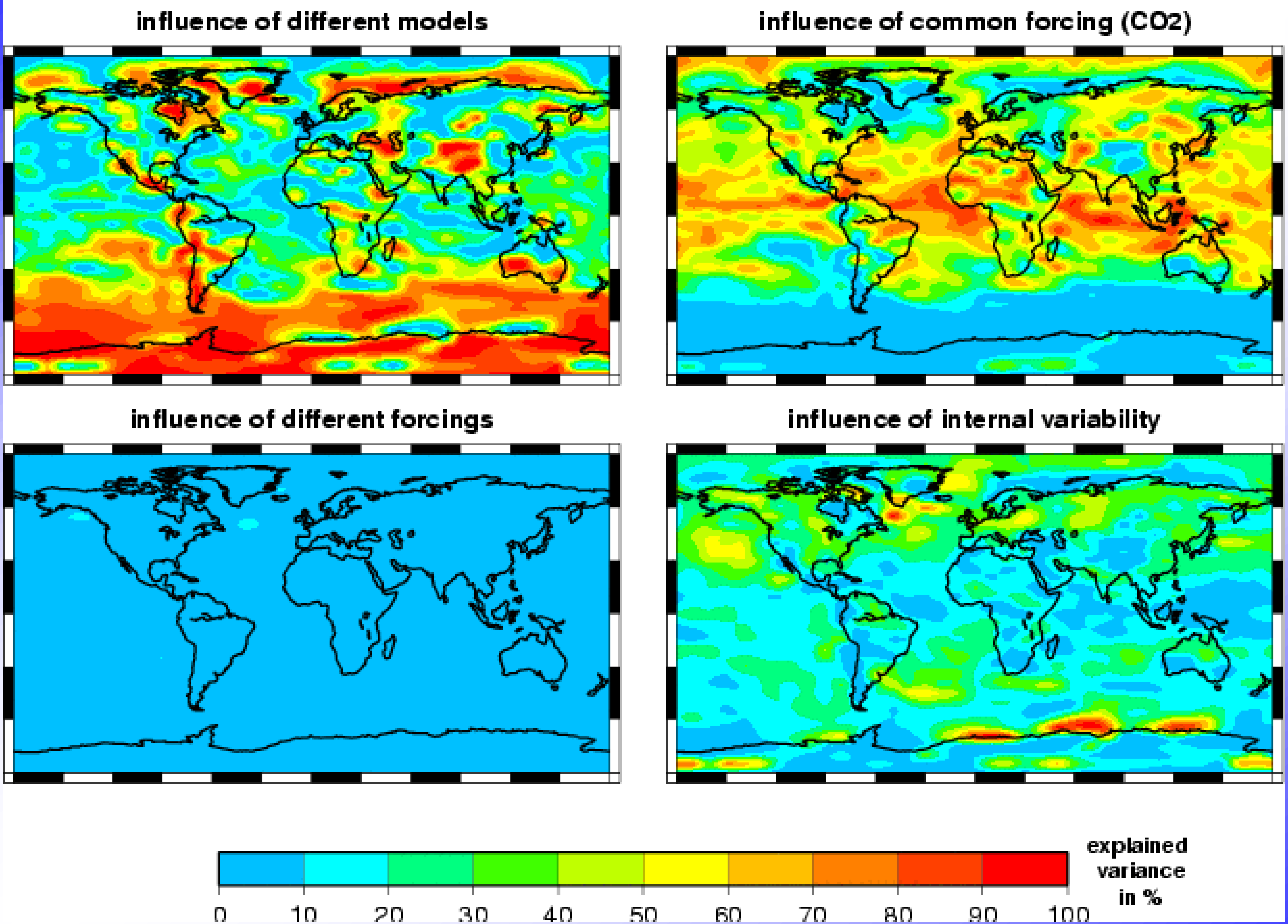
Climate change simulation with GHG forcing

- Sampling uncertainty in initial conditions
  - ensemble simulations (typically 5 or 6 members)
- Sampling inter-model uncertainty
  - two model example: ECHAM3/T21 and HADCM2
  - multimodel example: 15 different models from IPCC data server

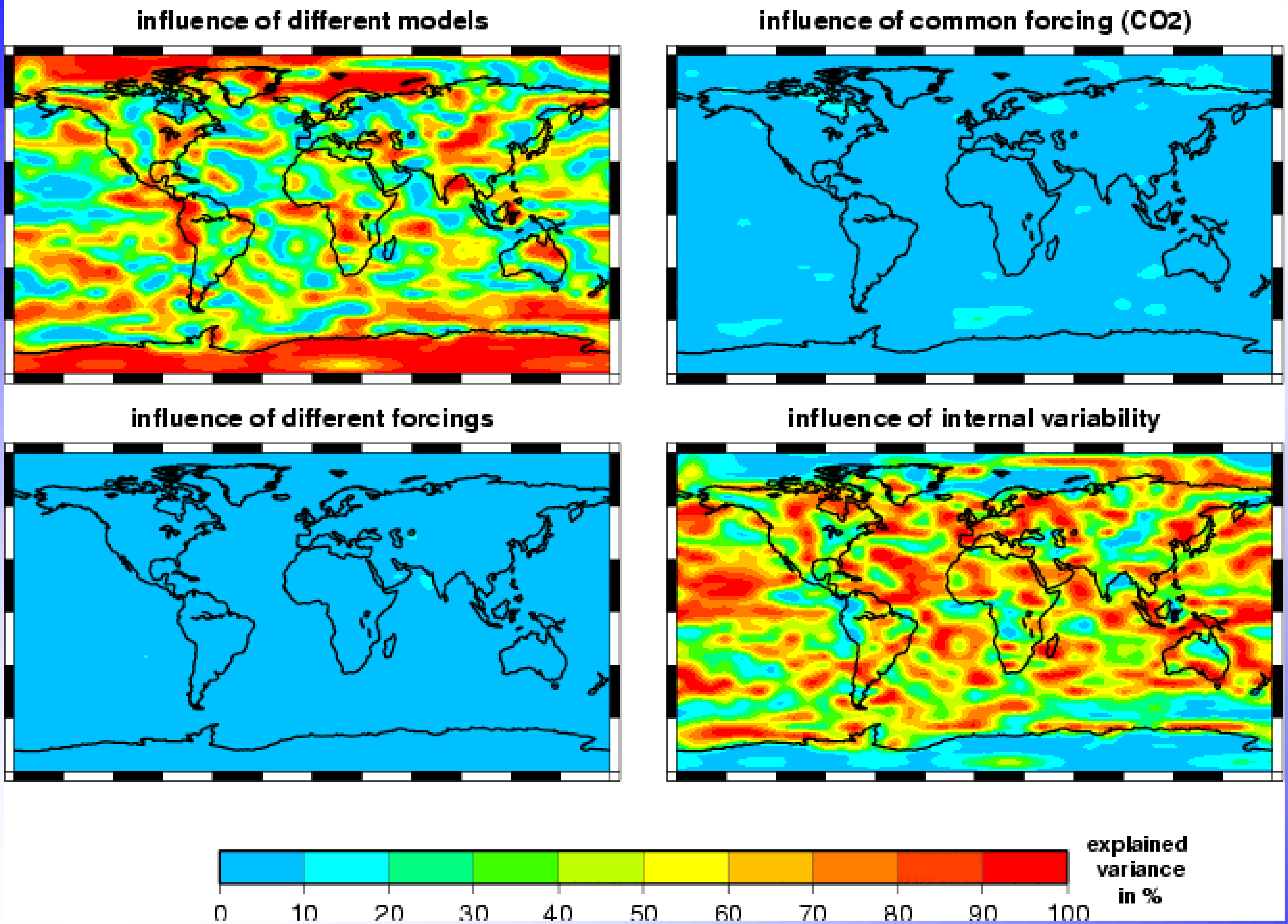
Climate change simulations with GHG forcing

- Two model case: precipitation and near surface temperature
- multi model case: Arctic oscillation/North Atlantic oscillation as a driving agent for regional climate variability in Europe
- classical 2-way analysis-of-variance
  - $\boldsymbol{x_{i,l,k} = a + b_j + c_l + d_{i,l} + e_{i,l,k}}$
  - $\boldsymbol{b_i}$: common GHG signal as function of time i
  - $\boldsymbol{c_l}$: bulk inter-model differences
  - $\boldsymbol{d_{i,l}}$: inter model-differences in GHG forcing

2W-ANOVA of CO2 scenario ensembles: annual means of T2M

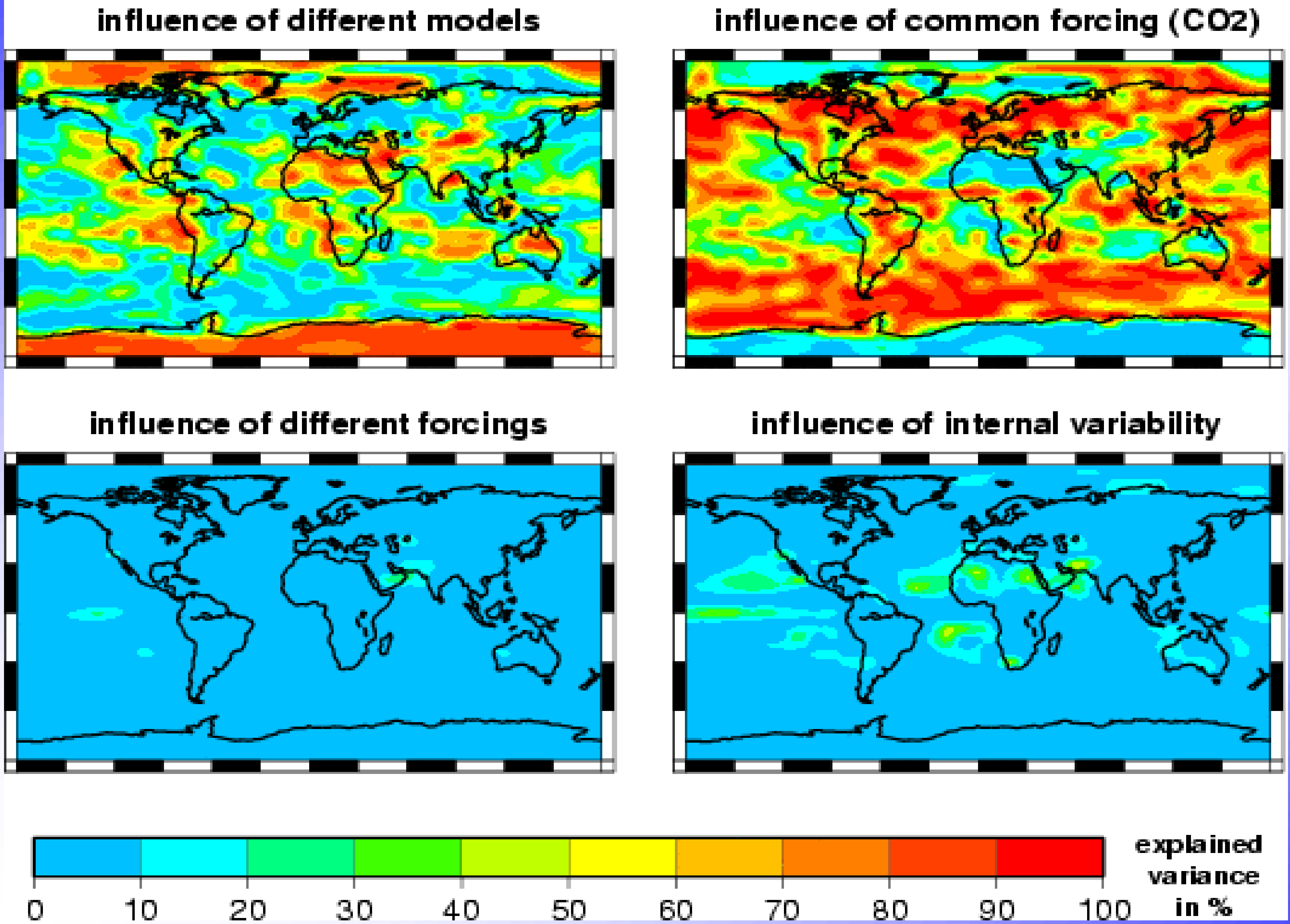


2W-ANOVA of CO2 scenario ensembles: annual sums of PRE

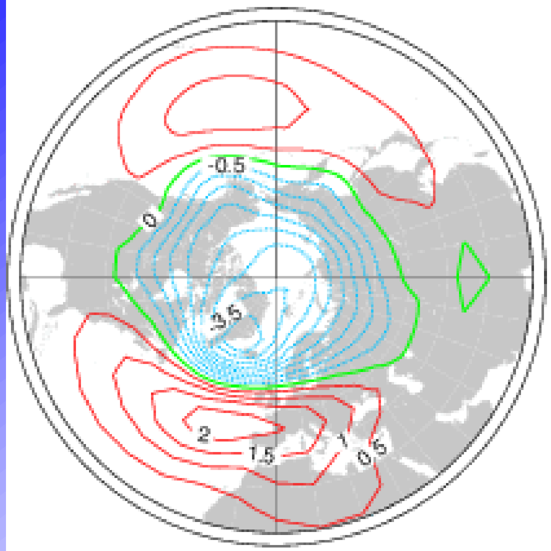


2W-ANOVA of CO2 scenario ensembles: 10-year filtered annual sums of PRE

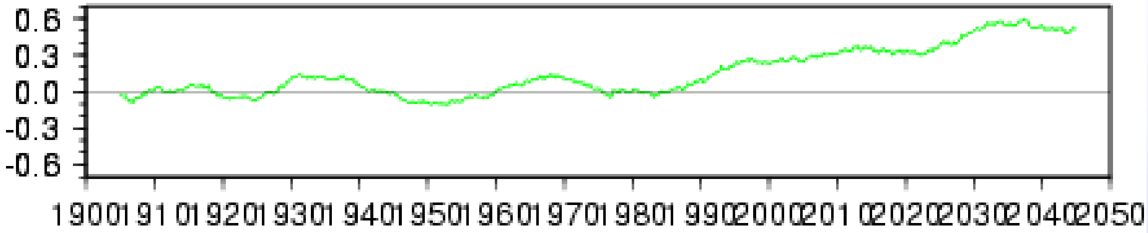
ECHAM3/LSG vs. HADCM2, 1880-2049, globe



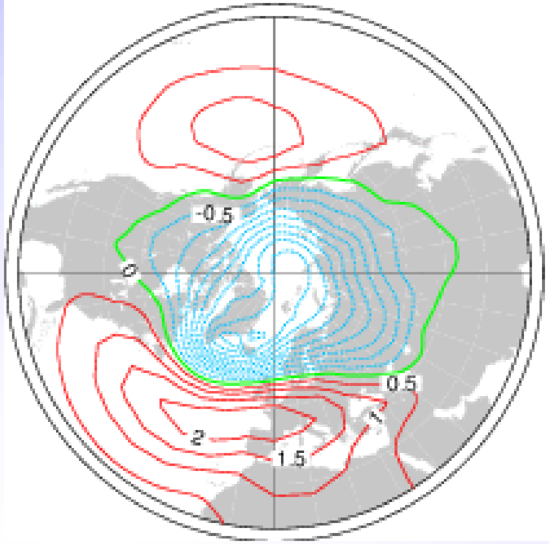
Superensemble EOF1 (20.3 %)



Superensemble 10-year lowpass filtered PC1

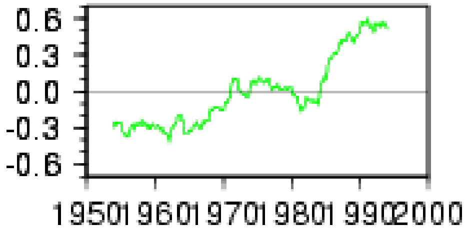


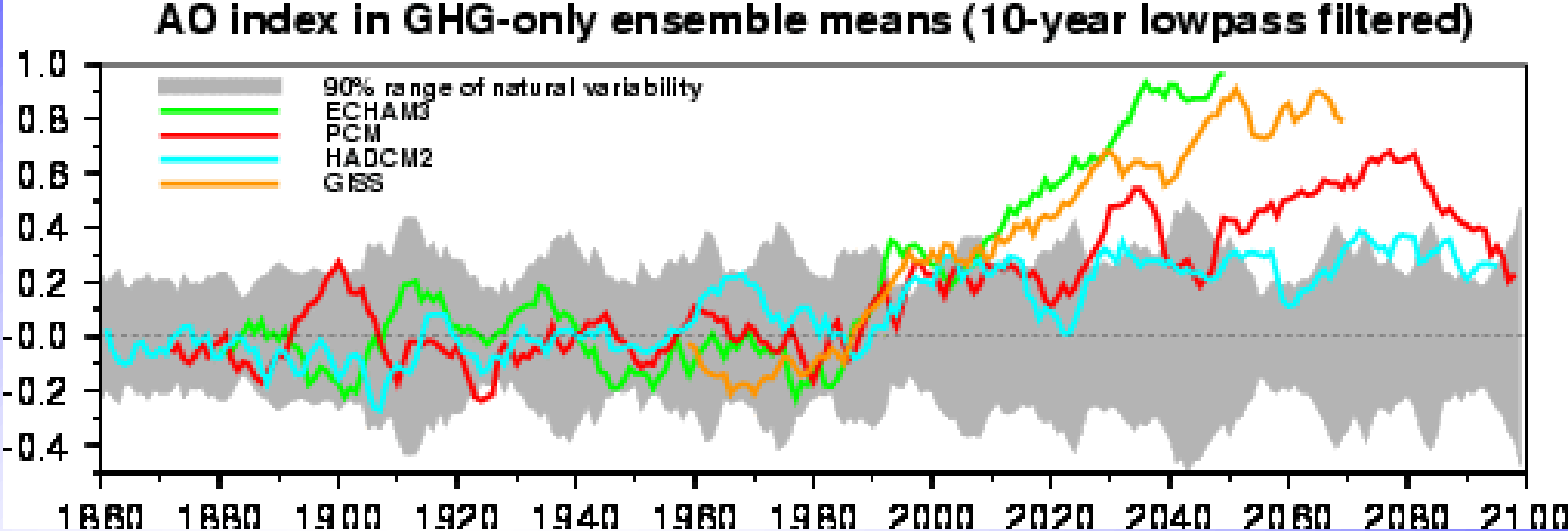
NCEP EOF1 (18.4 %)

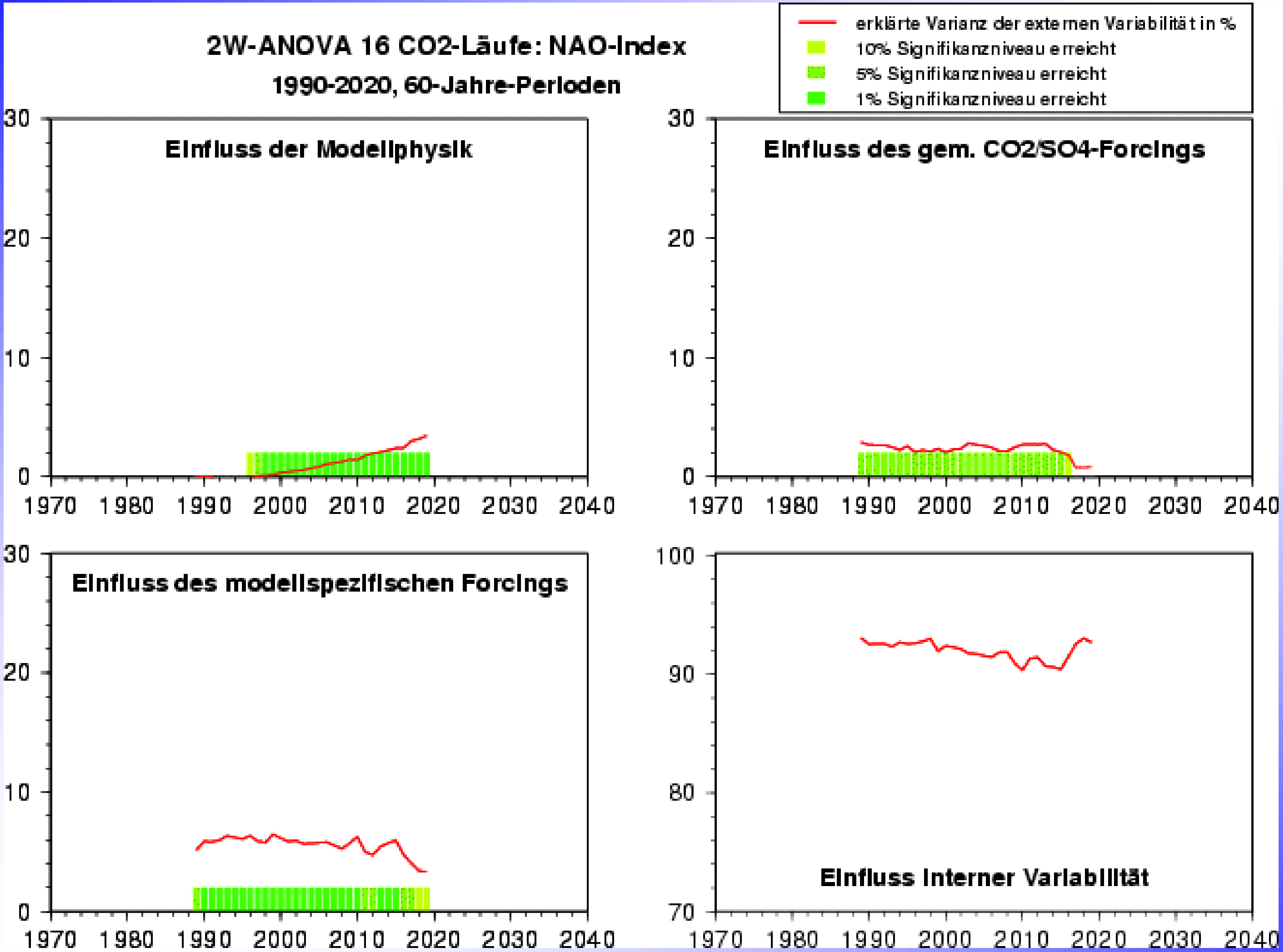


EOF1 correlation: 0.97
Super. trend (1974-2013): 0.86 hPa/100a
NCEP trend (1954-1993): 2.23 hPa/100a
PC1 correlation (trend periods): 0.88

NCEP 10-year lowpass filtered PC1

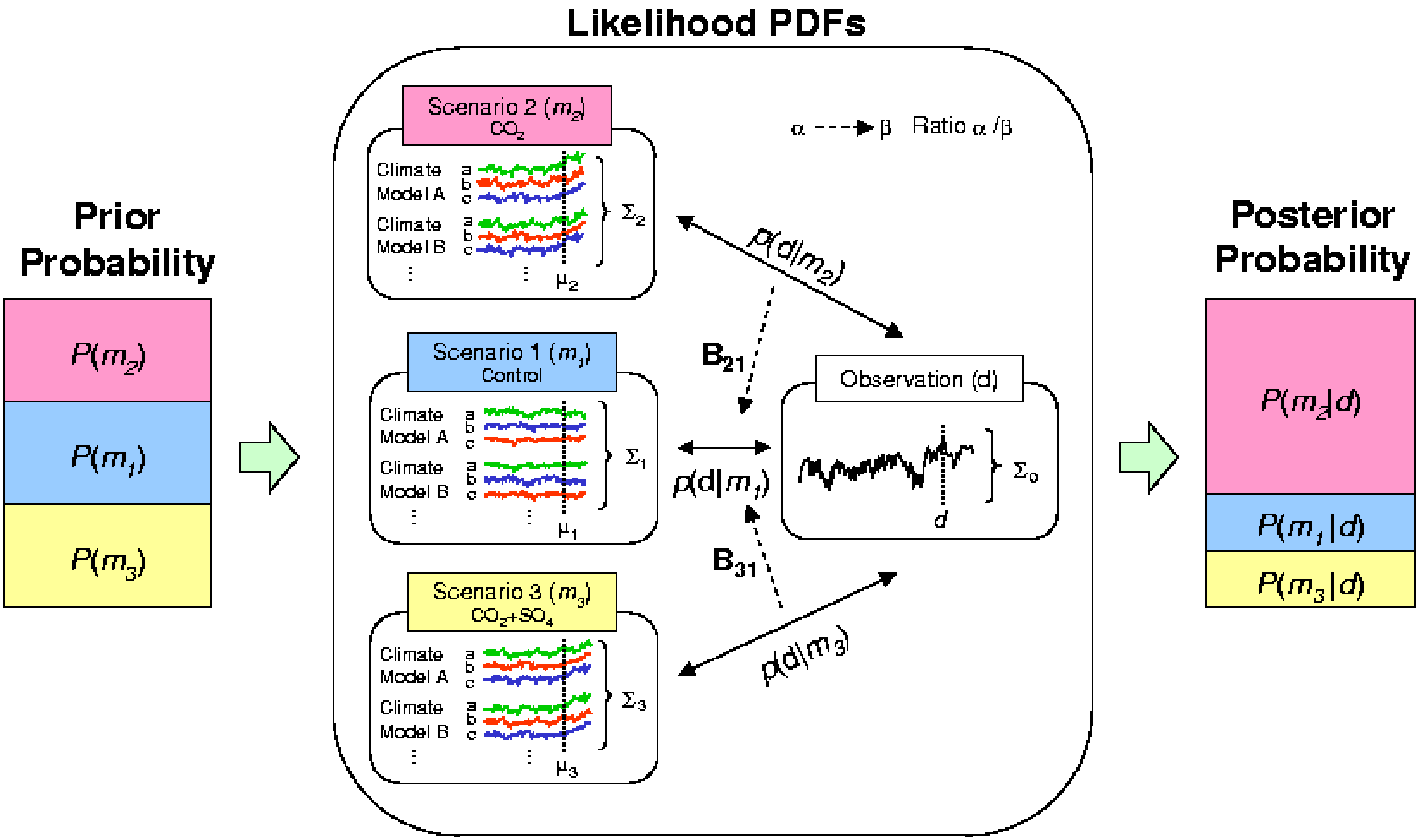



AO index in GHG-only ensemble means (10-year lowpass filtered)
90% range of natural variability
ECHAM3
PCM
HADCM2
GISS
1.0
0.8
0.6
0.4
0.2
-0.0
-0.2
-0.4
1860
1880
1900
1920
1940
1960
1980
2000
2020
2040
2060
2080
2100


2W-ANOVA 16 CO2-Läufe: NAO-Index
1990-2020, 60-Jahre-Perioden
erklärte Varianz der externen Variabilität in %
10% Signifikanzniveau erreicht
5% Signifikanzniveau erreicht
1% Signifikanzniveau erreicht
Einfluss der Modellphysik
Einfluss des gem. CO2/SO4-Forcings
Einfluss des modellspezifischen Forcings
Einfluss interner Variabilität
0
10
20
30
70
80
90
100
1970
1980
1990
2000
2010
2020
2030
2040

Climate change model simulations
Bayesian view

- Available a set of hypothesis /scenarios $\boldsymbol{h}_{\mathrm{i}}$
  - – unforced variability $\boldsymbol{i=1}$
  - – GHG forced
  - – GHG + sulphate aerosol forced
  - – solar/volcanic forced
- for each hypothesis / scenario we have a prior $\boldsymbol{o\,(h_i)}$
- Selection of $\boldsymbol{h_i}$ based on a given observation
  - – computation of Bayes factor from likelihood
  - – decision based on posterior $\boldsymbol{p(h_i|o)}$



Prior Probability

$P(m_2)$

$P(m_1)$

$P(m_3)$

Likelihood PDFs

Scenario 2 (m_2)
CO_2

Climate Model A a b c

Climate Model B a b c

⋮ ⋮ μ_2 Σ_2

Scenario 1 (m_1)
Control

Climate Model A a b c

Climate Model B a b c

⋮ ⋮ μ_1 Σ_1

Scenario 3 (m_3)
CO_2+SO_4

Climate Model A a b c

Climate Model B a b c

⋮ ⋮ μ_3 Σ_3

α - - -▶ β Ratio α/β

$p(d|m_2)$

$\mathbf{B}_{21}$

$p(d|m_1)$

$\mathbf{B}_{31}$

$p(d|m_3)$

Observation (d)

d Σ_0

Posterior Probability

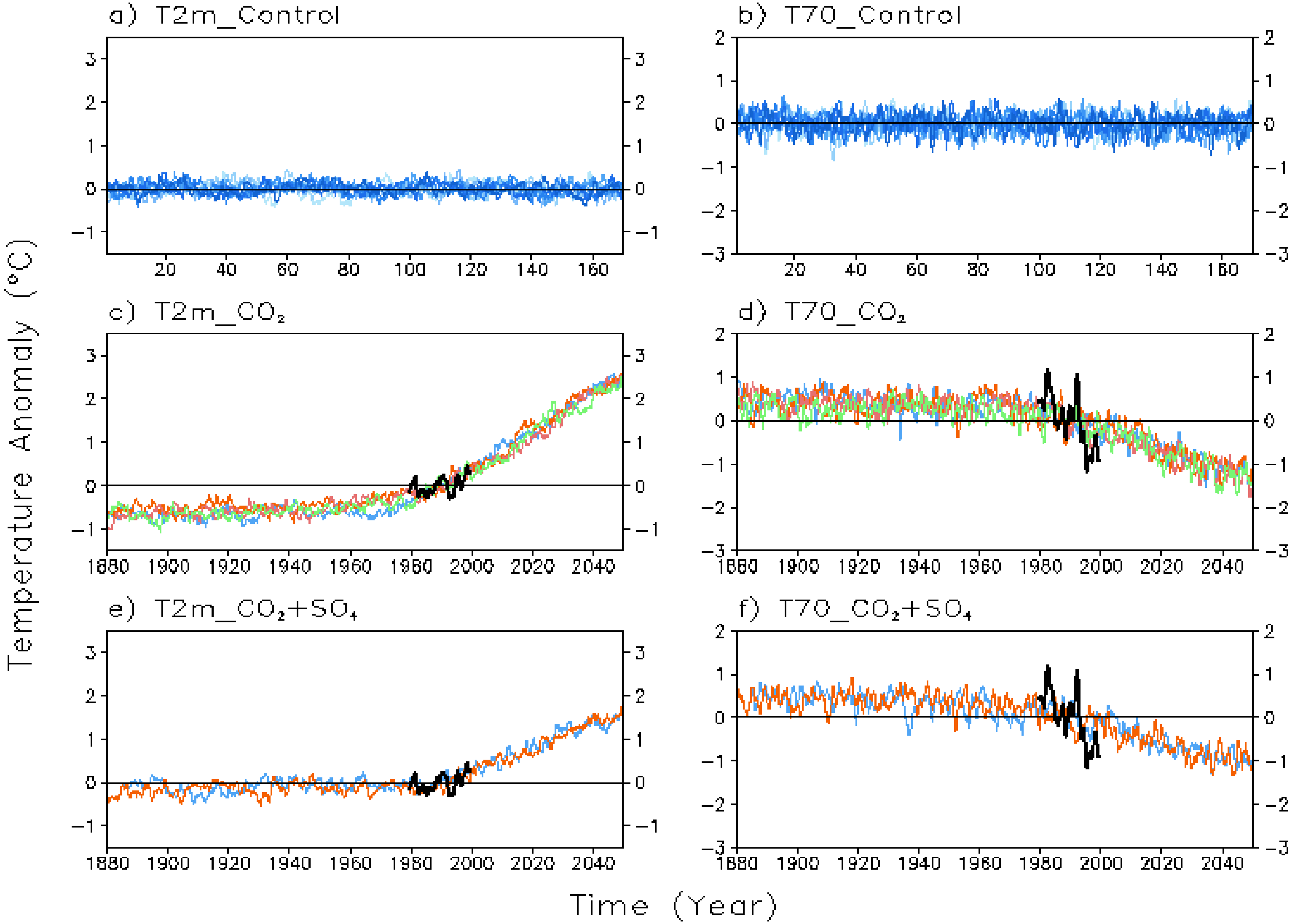
$P(m_2|d)$

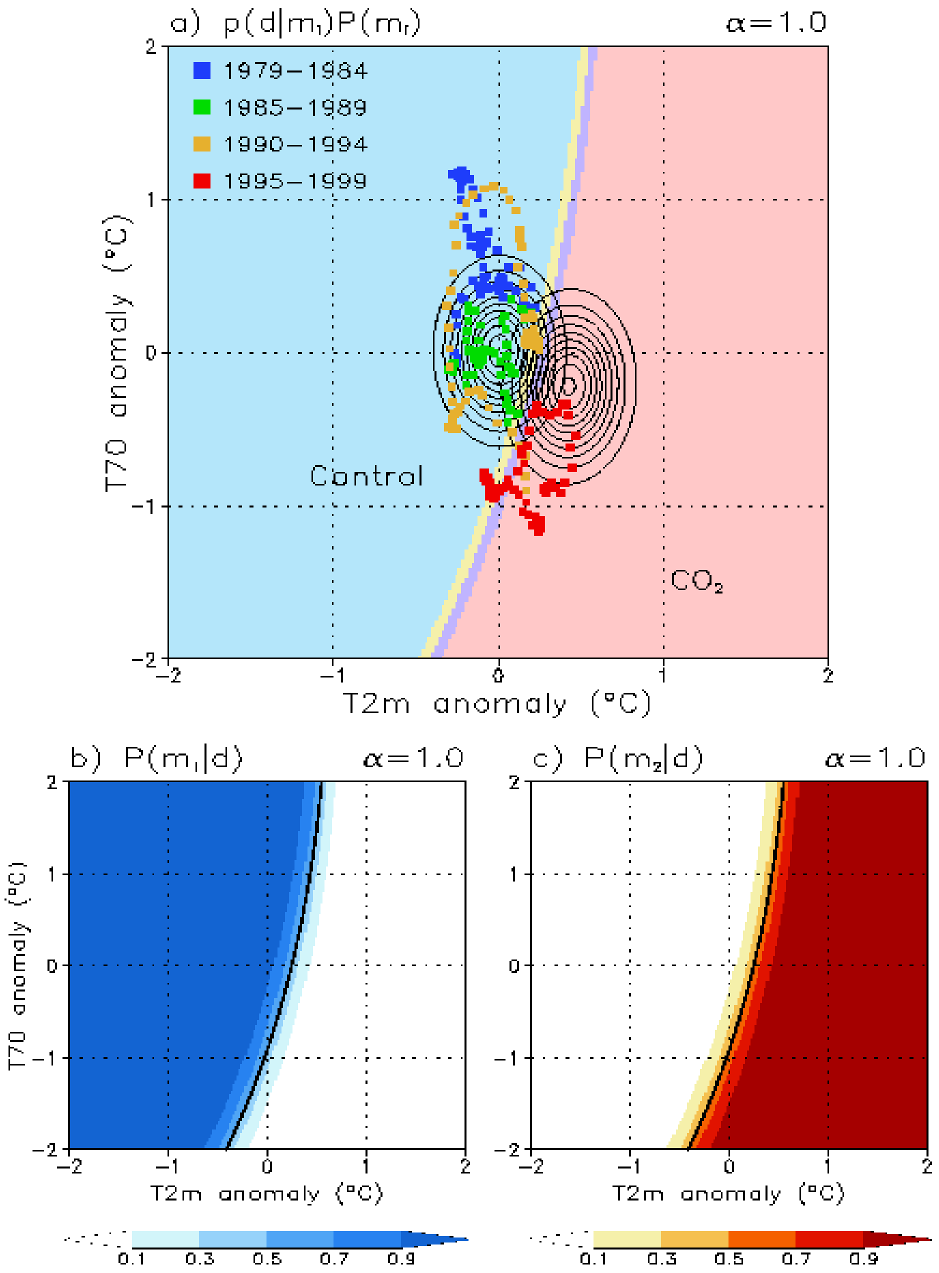
$P(m_1|d)$

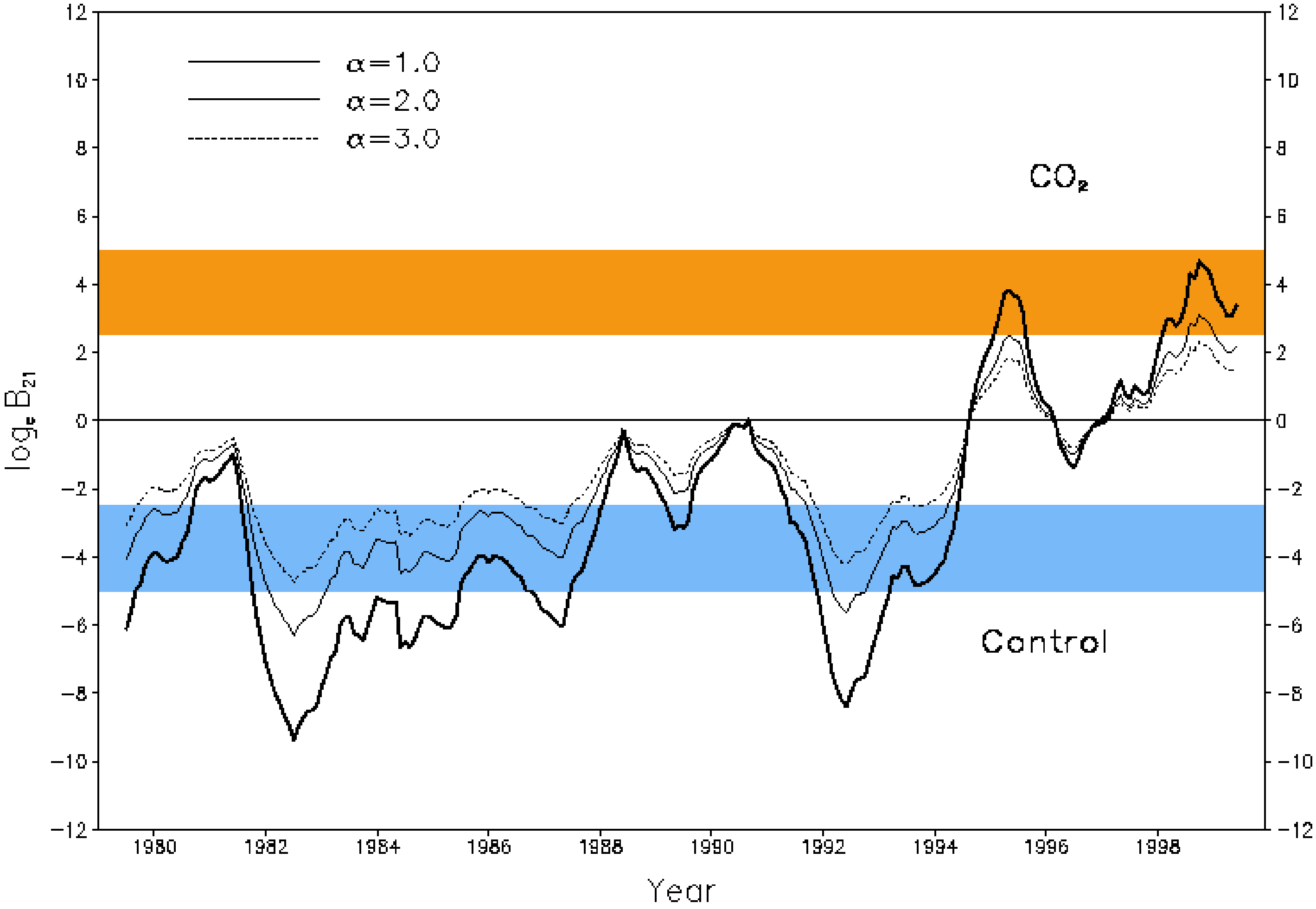
$P(m_3|d)$

Climate change model simulations
Bayesian view

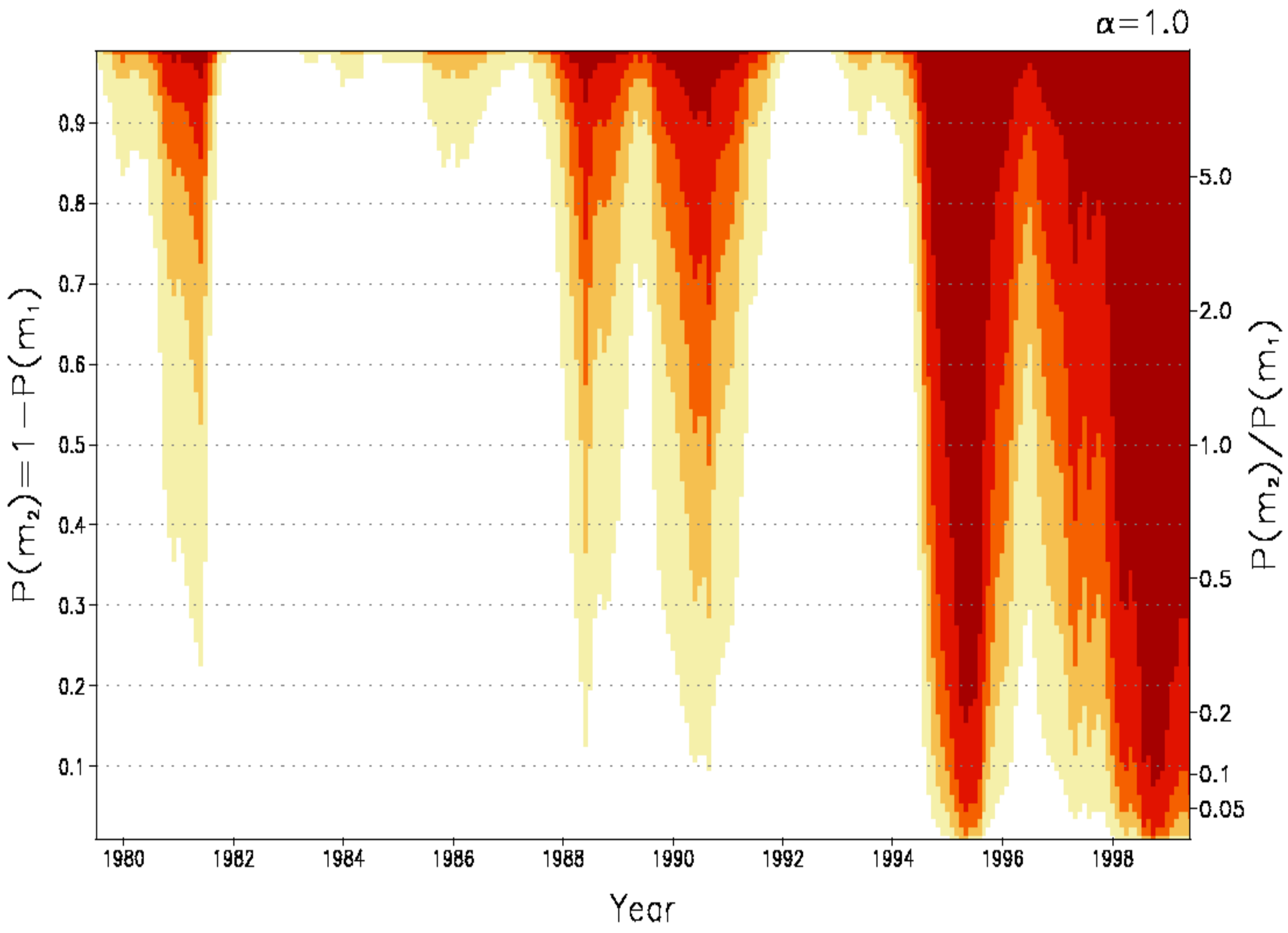
- 2-dimension example: using Northern hemisphere mean temperatures near surface and lower stratosphere
- observations 1979 - 1999 moving annual means
- model signal: linear change between 1990-2010 in model year 2000
- 5 member ensemble ECHAM3/T21 GHG only
- 3 member ensemble ECHAM3/T21 GHG+S-Ae


a) T2m_Control
b) T70_Control
c) T2m_CO₂
d) T70_CO₂
e) T2m_CO₂+SO₄
f) T70_CO₂+SO₄
Temperature Anomaly (°C)
Time (Year)
20 40 60 80 100 120 140 160
1880 1900 1920 1940 1960 1980 2000 2020 2040
3 2 1 0 −1
2 1 0 −1 −2 −3


a) p(d|m₁)P(m₁)
α=1.0
1979–1984
1985–1989
1990–1994
1995–1999
Control
CO₂
T70 anomaly (°C)
T2m anomaly (°C)
2
1
0
−1
−2
b) P(m₁|d)
α=1.0
c) P(m₂|d)
α=1.0
T70 anomaly (°C)
T2m anomaly (°C)
T2m anomaly (°C)
0.1
0.3
0.5
0.7
0.9
0.1
0.3
0.5
0.7
0.9

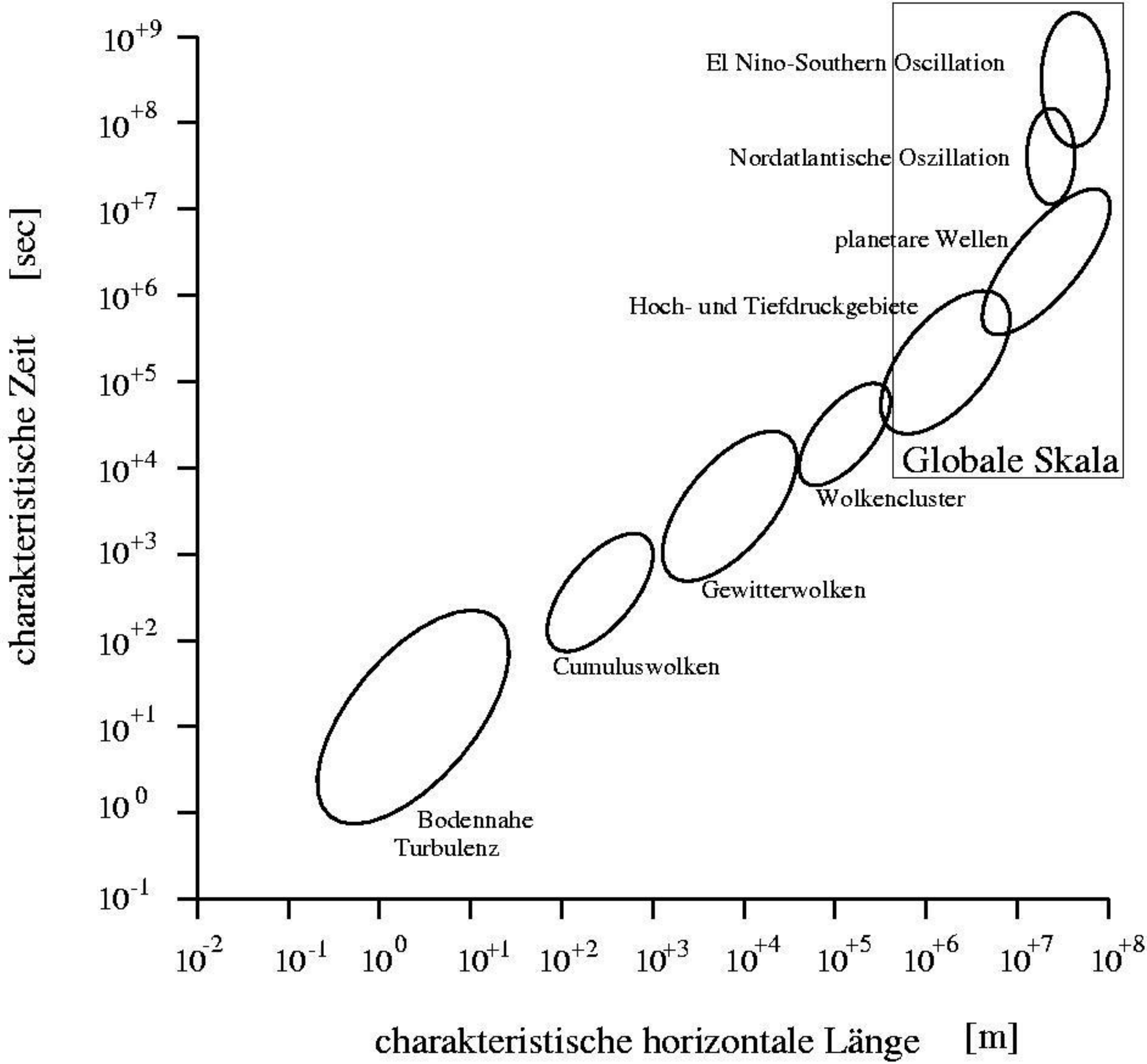


α=1.0
α=2.0
α=3.0
CO₂
Control
log_e B₂₁
Year
12
10
8
6
4
2
0
−2
−4
−6
−8
−10
−12
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998


α=1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
P(m₂)=1−P(m₁)
5.0
2.0
1.0
0.5
0.2
0.1
0.05
P(m₂)/P(m₁)
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
Year

Conclusion

- Weather prediction and climate system models simulate parts of the real Earth system
  - starting from these complex models: need to **introduce statistical aspects** at various levels
  - starting from observations: pure **data-based models need a guidance**: use physics / chemistry of complex models
- we need quantitative statements about **future changes and their uncertainties** of the real system either the next day, the next decade or century



charakteristische Zeit [sec]
10^{+9}
10^{+8}
10^{+7}
10^{+6}
10^{+5}
10^{+4}
10^{+3}
10^{+2}
10^{+1}
10^{0}
10^{-1}
El Nino-Southern Oscillation
Nordatlantische Oszillation
planetare Wellen
Hoch- und Tiefdruckgebiete
Globale Skala
Wolkencluster
Gewitterwolken
Cumuluswolken
Bodennahe Turbulenz
10^{-2}
10^{-1}
10^{0}
10^{+1}
10^{+2}
10^{+3}
10^{+4}
10^{+5}
10^{+6}
10^{+7}
10^{+8}
charakteristische horizontale Länge [m]