Advanced Methods of Time Series Analysis and their Application to Climate Research

Katja Matthes (1, 2), Christian Blume (2), and Illia Horenko (3)

(1) Helmholtz-Zentrum Potsdam
Deutsches GeoForschungsZentrum (GFZ), Potsdam, Germany

(2) Freie Universität Berlin, Institute für Meteorologie, Berlin, Germany

(3) Institute of Computational Science, Universita della Swizzera Italiana, Lugano
Switzerland

GFZ Helmholtz-Zentrum Potsdam

Freie Universität Berlin

UCLA/IPAM Workshop LA, 26.05.2010

HELMHOLTZ GEMEINSCHAFT

Outline

– Collaboration project description
– Introduction to stratospheric research questions
– Results from standard time series analysis
– Results from new methods (FEM-VARX, Neural Network)
– Conclusions and Outlook

GFZ Helmholtz-Zentrum Potsdam
Freie Universität Berlin

HELMHOLTZ GEMEINSCHAFT

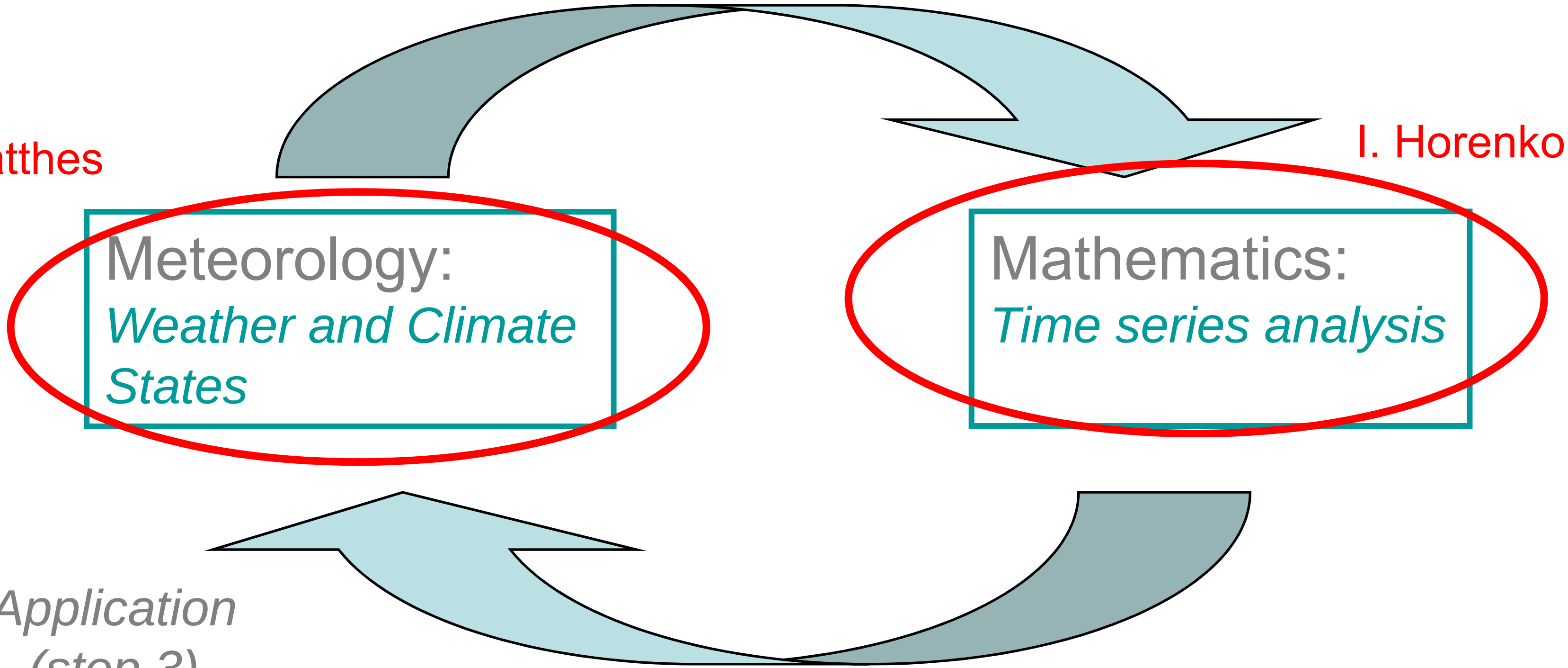
Work Program

CSS Center for Scientific Simulation

Meteorological questions and methodological requirements, data (step 1)

K. Matthes

I. Horenko



Application (step 3)

Mathematical methods, new approaches and tools (step 2)

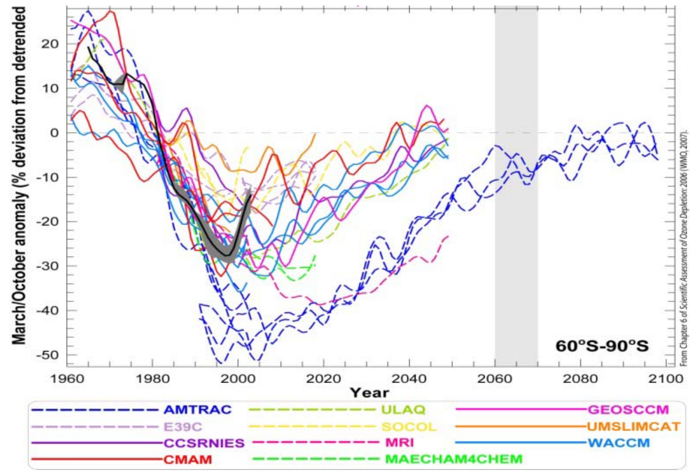
GFZ Helmholtz-Zentrum Potsdam

HELMHOLTZ GEMEINSCHAFT

Mathematical methods and tools

(I. Horenko)

Total column ozone anomalies



WMO (2007)

Meteorological data are multidimensional, nonlinear, nonstationary

Methodological Challenges (mathematics):

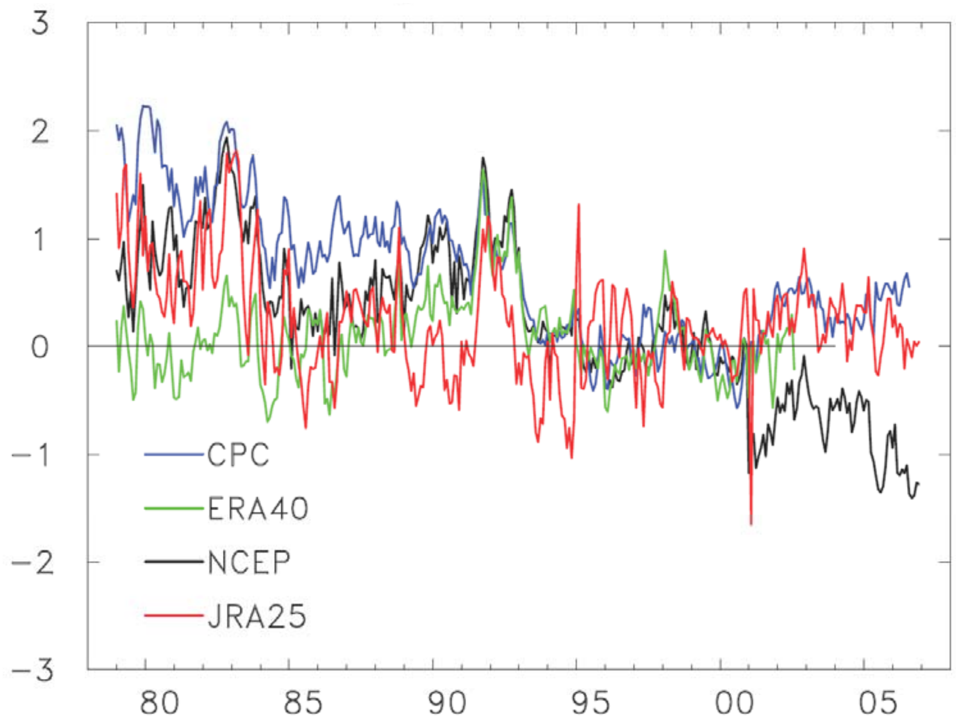
1) New methods of time series analysis (*nonstationary*)

2) *Apply theory of large deviations to climate data*
(Wenzel/Freidlin, O. Bühler)

Climate Variability and Change

Meteorology (Matthes)

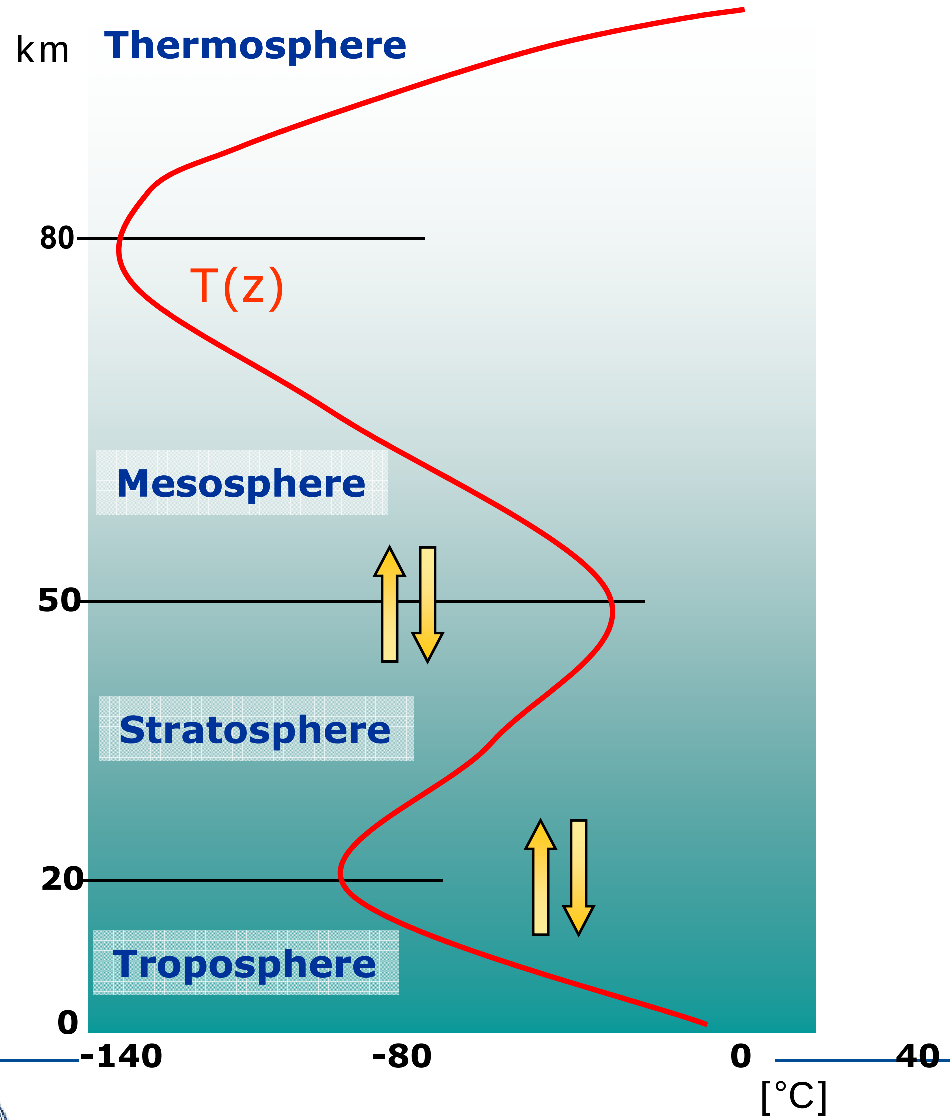
Time series of global averaged temperature anomalies at 100 hPa
from different observational data sets, ref. to the period 1995-2000



Randel et al. (2008)

Problems: • Identifying significant signals (trends, 11-year solar cycle, QBO)
• Attributing causes to observed changes

Introduction: The Stratosphere



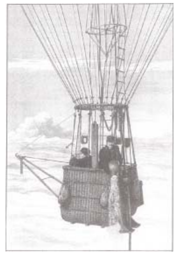
Why is the stratosphere interesting?

- Early warning system for climate change

- Decrease of ozone layer („ozone hole“) in the stratosphere

- Stratospheric dynamics influences tropospheric circulation

Observational data: Stratosphere

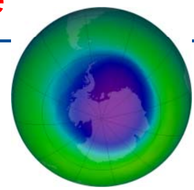



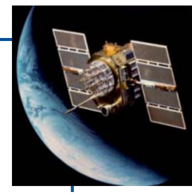
- Since 1881 — Scientific ballons (Mongolfièren)
- Since 1892 — ballons without passengers > 10 km
- 1901 — Ballon trip of Berson & Süring
- from 1901 — Use of free ballons from rubber

- 1948 — First hemispheric monthly mean stratospheric maps
- Since 1951 — American radiosondes from Berlin-Tempelhof


- 1957/58 — International Geophysical Year (IGY) :
  first international coordinated measurement problem
  first daily stratospheric analysis at FU Berlin
- 1960er — Extension of radiosonde and rocket sonde network
- 1969 — Start of first operational remote sensing satellite (Nimbus 3)

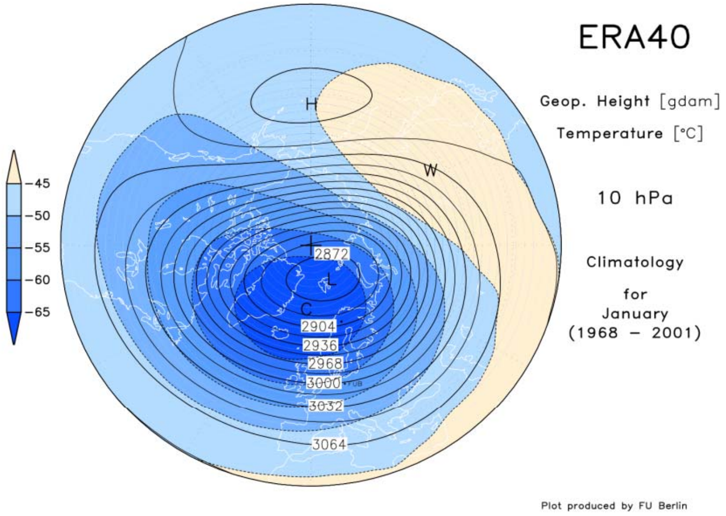
- From 1979 — continuos operational satellite measurements (Nimbus, NOAA)
- 1985 — Discovery of Antarctic ozone hole



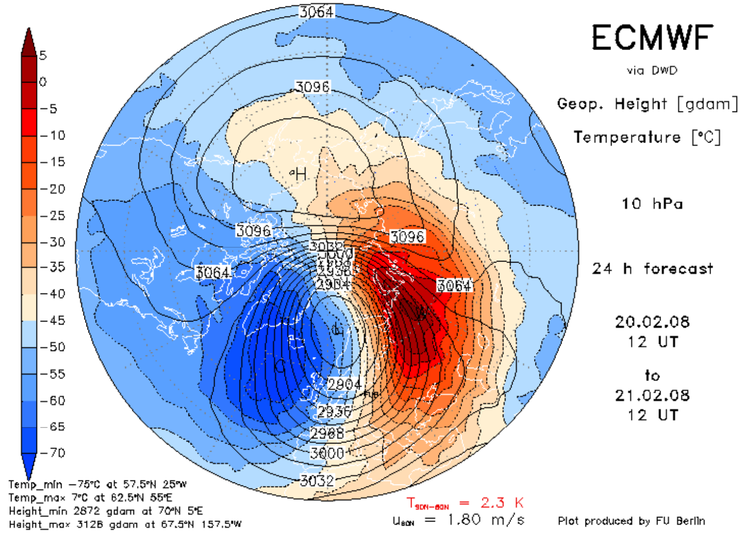


Stratosphere – Variability and Coupling

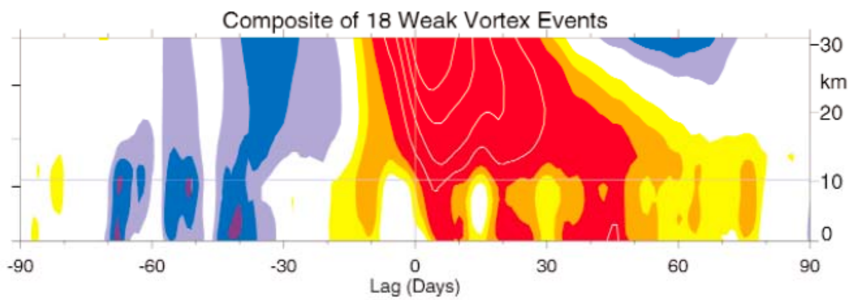
January – Mean State



Major Stratospheric Warming



Stratosphere - Troposphere coupling



Weather from above. A weakening stratospheric vortex (red) can alter circulation down to the surface, bringing storms and cold weather farther south than usual.

Baldwin and Dunkerton (2001)

GFZ Helmholtz-Zentrum Potsdam

HELMHOLTZ GEMEINSCHAFT

Why do we use middle atmosphere (MA) models?

- MA part of the whole atmosphere that interacts through dynamical, chemical and radiation processes with the troposphere

- GCMs are „experimental tools“ to understand physical processes (radiative processes, dynamical and chemical interactions) in the atmosphere(-ocean) system with the following questions:

1. How does the MA influences the troposphere? How does this change in a changing climate?

2. How does the troposphere impacts the stratospheric circulation?

3. How will the ozone layer change due to anthropogenic changes?

What are current MA-GCMs able to do?

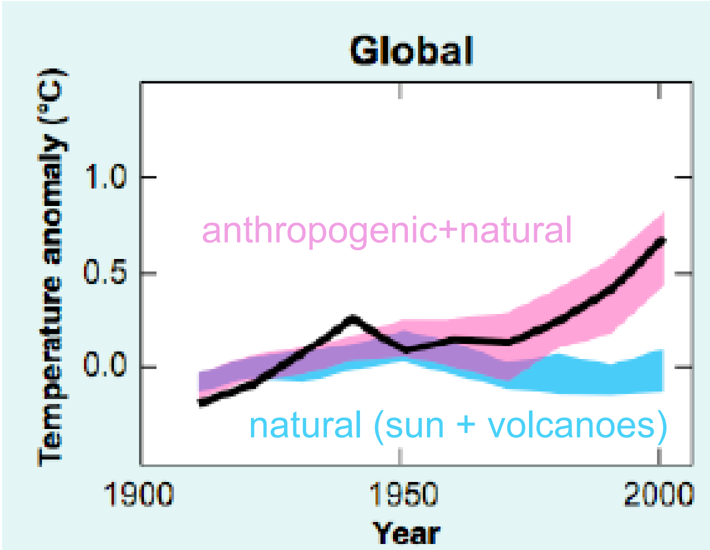
- MA-GCMs simulate the time evolution of large scale atmospheric structures through integration of the primitve equations under suitable initial and boundary conditions

- Ideal would be: only prescription of external forcings (solar radiation, sea surface temperatures and ice coverage) => Modell simulates temperature, Wind, clouds, and trace gase distributions as 3-D functions in space and time

- Reality: compromises required (scientific and computational restrictions)

small scale processes, e.g., radiative transfer, gravity waves and convection are parameterized

Recent climate debate

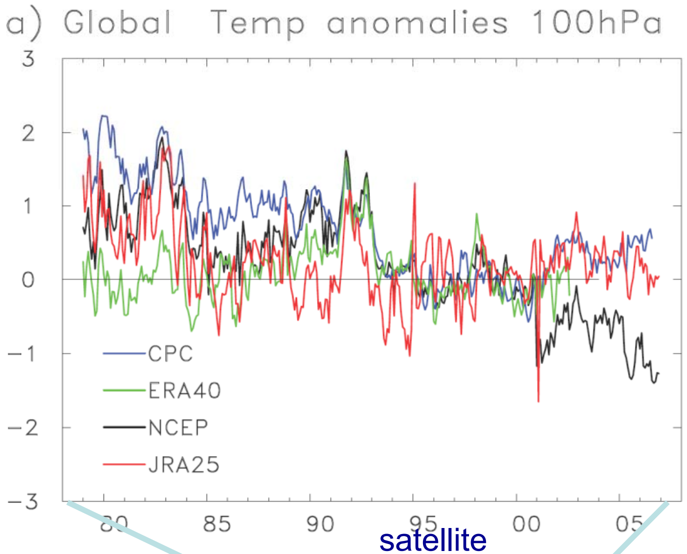


IPCC (2007)

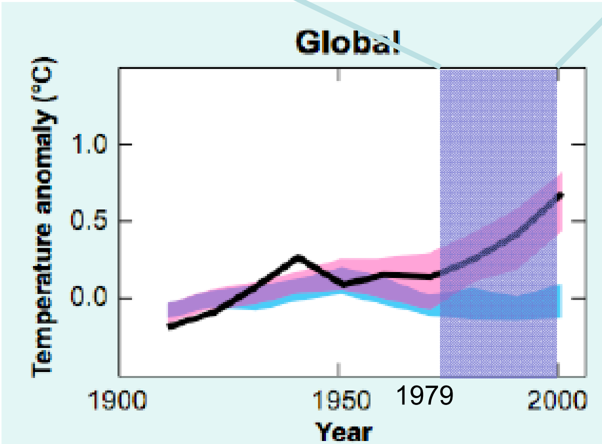
Problems

- in time and space limited observational data
- Modell uncertainties (stratospheric processes not included)
- Impact of natural and anthropogenic effects in observational data?

Role of the Stratosphere for future climate change?



satellite measurements



Cooling of the stratosphere

Randel et al. (2008)

Warming of the troposphere

IPCC (2007)

Motivation

- commonly used method: multiple linear regression (MLR) (Bodeker, 2001)

=> problem: atmospheric signals highly non-linear

- novel methods of multivariate time series analyses (Horenko, 2008, 2010): allow to analyze non-linear, non-stationary, non-markovian and multidimensional problems

=> Test of new methods with stratospheric datasets (observations as well as chemistry climate model output)

Example Data

Time series of total ozone averaged from 60S to 60N from 1963-2004:

2) ground-based and satellite data (TOMS+gb)
3) ouput from a chemistry climate model (CCM), i.e. NCAR WACCM from the recent past (1960-2004) from the WCRP/SPARC-CCMVal initiative

SPARC-CCMVal Report

STRATOSPHERIC PROCESSES
AND THEIR ROLE IN CLIMATE

SPARC

A project of the WMO/ICSU/IOC World Climate Research Programme

---

SPARC Report on the Evaluation of
Chemistry Climate Models

June 2010

Prepared by the SPARC CCMVal Group under the auspices of the
SPARC Scientific Steering Group.

Edited by V. Eyring, T. Shepherd and D. Waugh

WCRP - XXX
WMO/TD - No. XXXX
SPARC Report No. 5

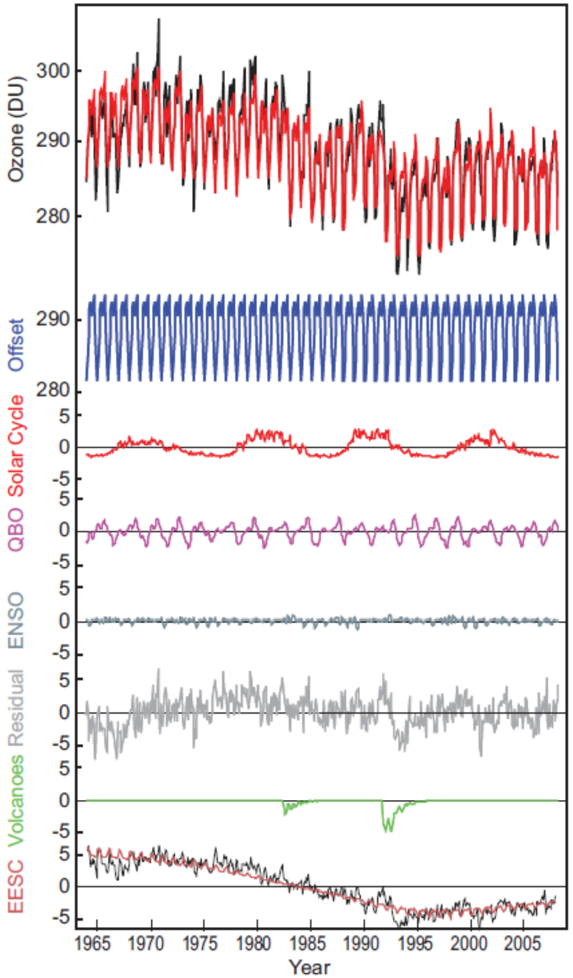
Chapter 8: Natural Variability of Stratospheric Ozone 305

Manzini and Matthes (2010)

V. Eyring, T. G. Shepherd, D. W. Waugh (Eds.), SPARC Report No. 5, WCRP-X, WMO/TD-No. X, http://www.atmosp.physics.utoronto.ca/SPARC, 2010..

Results from a Standard Methode: Multiple Linear Regression Analysis

Total Column Ozone from 60S to 60N



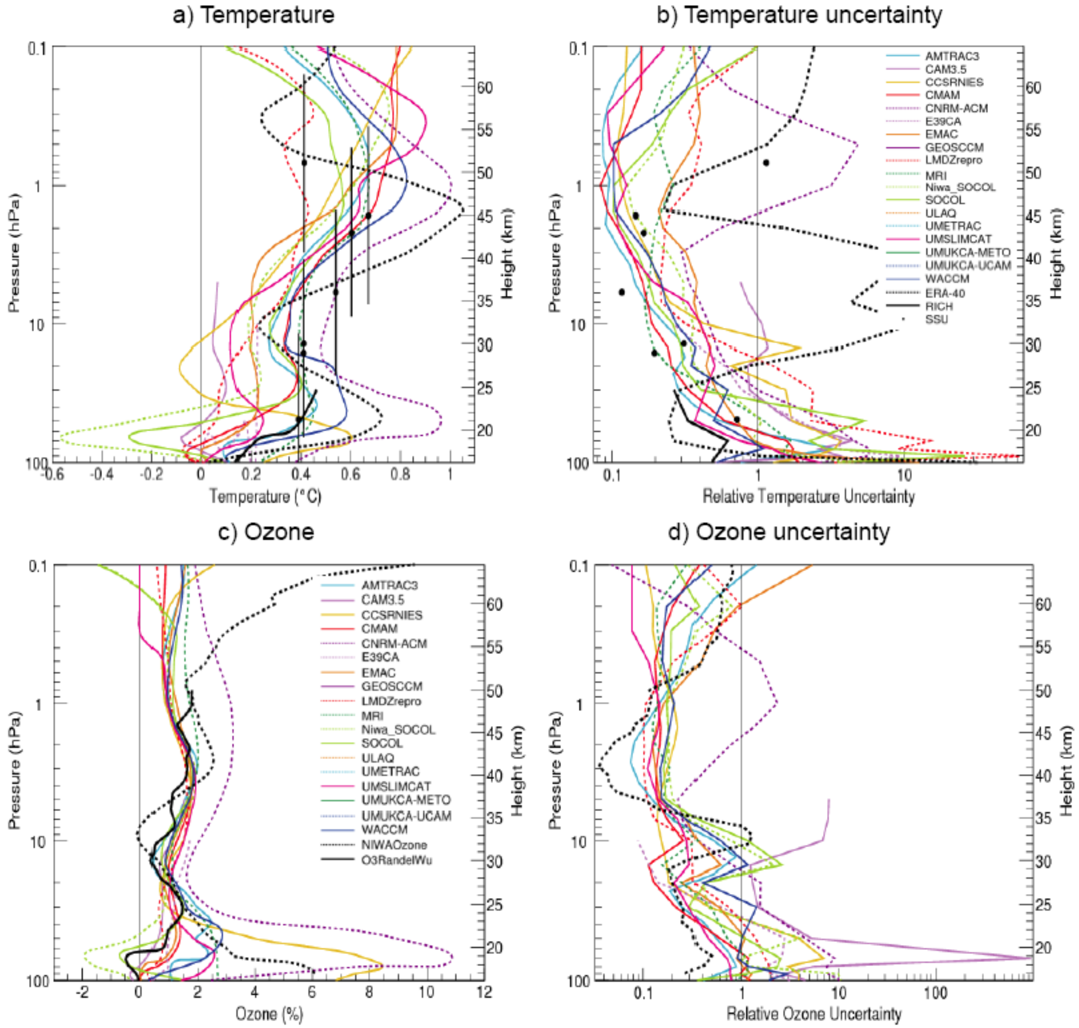
$$y(t) = \beta_{offs(N=4)} \times offset + \beta_{EESC(N=2)} \times EESC(t) + \beta_{QBO(N=2)} \times QBO(t) + \beta_{QBO\_or(N=2)} \times QBO\_orthog(t) + \beta_{sol(N=0)} \times solar(t) + \beta_{ENSO(N=2)} \times ENSO(t) + \beta_{Ag(N=2)} \times Agung(t) + \beta_{Elc(N=2)} \times ElChichon(t) + \beta_{Pin(N=2)} \times Pinatubo(t) + R(t)_{t=1,n}$$

Bodeker et al. (1998), (2001)

MLR assumes that external factors only have a stationary and linear influence which is certainly not true for highly non-linear processes in the atmosphere

GFZ Helmholtz-Zentrum Potsdam | Freie Universität Berlin | HELMHOLTZ GEMEINSCHAFT

Solar Signal in Tropical Stratospheric Ozone



Manzini and Matthes et al., SPARC CCMVal report, 2010

- non-linear interactions not considered (solar, QBO, ENSO, vulcanoes)

- very different reaction of observ. and modeling data to changes in basis functions

The FEM-VARX Method I (Horenko, JAS, 2010)

- possibility to separate external contributions such as the QBO, solar and seasonal cycles in observational data

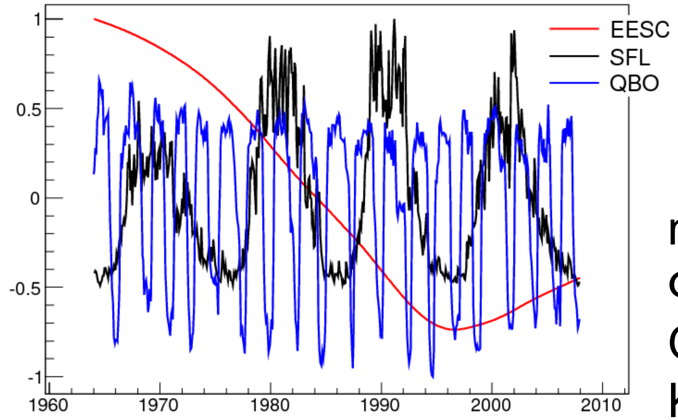
- goal: understand and separate e.g., anthropogenic and natural contributions in the atmosphere and make more reliable future predictions

- Purely data-driven approach for parameterization by means of nonstationary multivariate autoregressive factor (VARX) model combined with finite element method (FEM) clustering procedure

Procedure

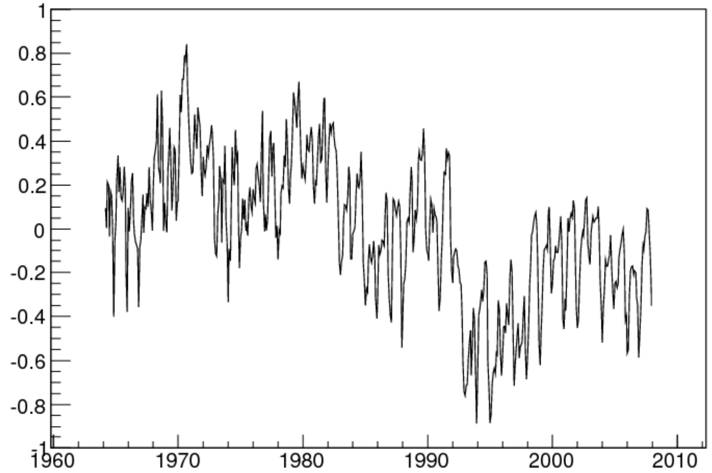
Deseasonalized and normalized

External factors u(t)



m (memory depth)
Φ (factor function)
Optimal C
K (number of cluster)

Input time series x(t)



FEM-VARX
(Horenko, JAS, 2010)

Output: model parameters

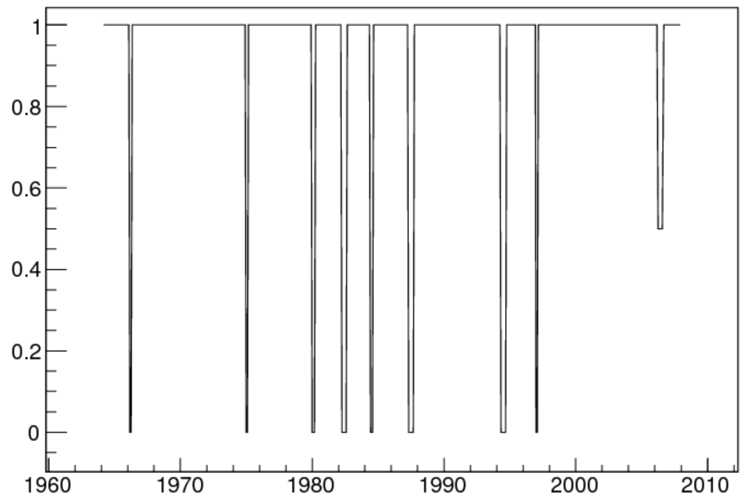
$\boldsymbol{\theta}(t) = [\mathbf{A}(t), \mathbf{B}(t), \mathbf{C}(t), \boldsymbol{\mu}(t)]$

$$\mathbf{x}_t = \boldsymbol{\mu}(t) + \sum_{q=1}^{m} \mathbf{A}_q(t)\mathbf{x}_{t-q\tau} + \mathbf{B}(t)\boldsymbol{\phi}\,[u(t)] + \mathbf{C}(t)\boldsymbol{\epsilon}_t.$$

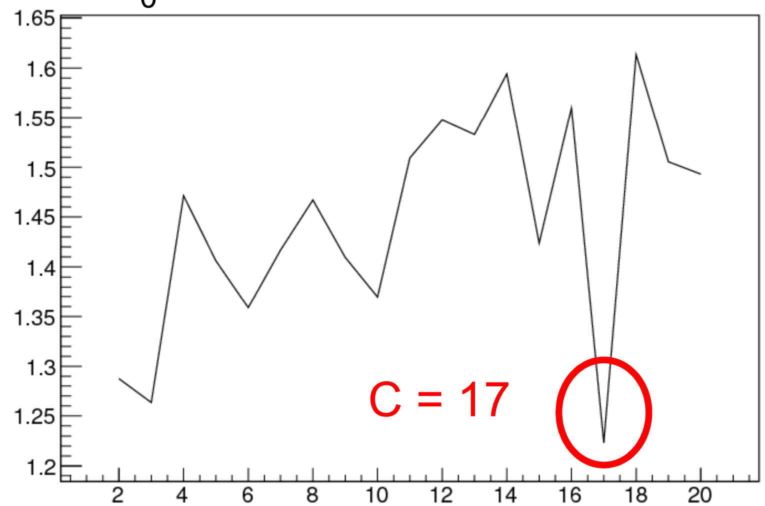
Choosing Optimal Settings

- bootstrap method => two statistically distinguishable cluster (K=2)
- Determination of optimal C (persistency parameter that defines maximal number of transition between cluster i and all other states) with Log-Likelihood method: optimal C when clustering functional L_0 minimal

Cluster affiliation function γ(t)



L_0 for different C



$$\mathbf{L}^{\epsilon}(\Theta,\Gamma,\epsilon^2) = \mathbf{L}(\Theta,\Gamma) + \epsilon^2 \sum_{i=1}^{K} \int_0^T [\partial_t \gamma_i(t)]^2 \delta t \rightarrow \min_{\Gamma \in W_{1,2}(0,T),\Theta}$$

Information Criterion

To select proper model order m and optimal functional from $\Phi(u(t))$ for external factors, standard tools of information theory applied, such as Akaike information Criterion (AIC): *opt. ratio between numb. of parameters and "model quality"*

model quality

$$\mathbf{AIC}\left(\Gamma(t),\Theta|\alpha_{opt},K,C,u_t\right)=\boxed{\mathbf{L}\left(\Gamma(t),\Theta|K,C,u_t\right)}+\alpha_{\mathrm{opt}}\boxed{\mathbf{N}\left(K,C,u_t\right)}\to\min$$

model compression factor

Horenko (2010)

Goal: quantification of relative influence of external factors to explain time series

The AIC Criterion

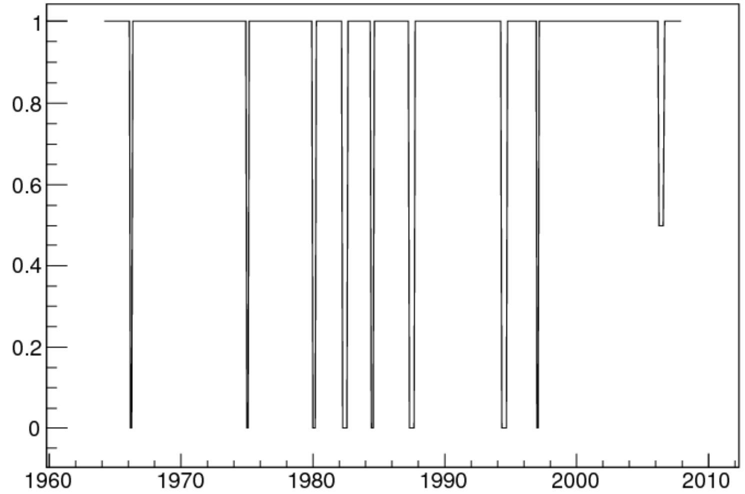
	m=1	m=2	m=3
With seasonal cycle	42.1	62.2	63.4
Without seasonal cycle	**31.7**	49.8	62.3

Φ	m=1	m=2	m=3
Exponential	32.1		
X^2	34.9		
X^3	31.9		
x^4	36.4		

⇒ Optimal AIC for m=1 (Markov process), 3 external factors and Φ linear

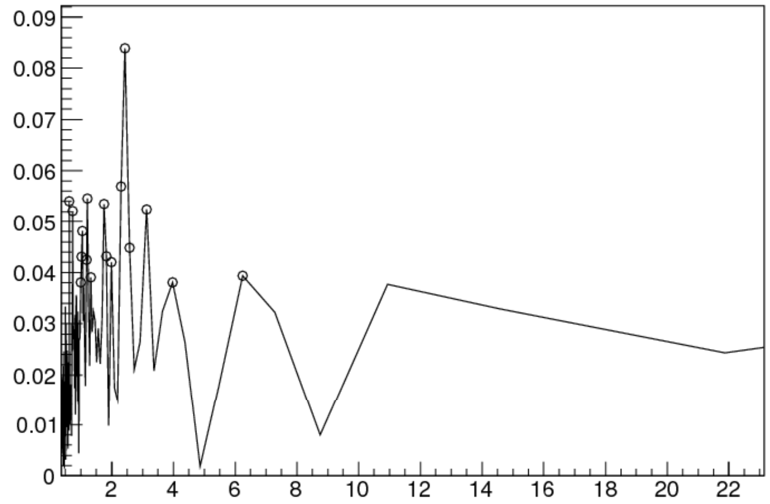
Example from Observations

Cluster affiliation function γ(t)



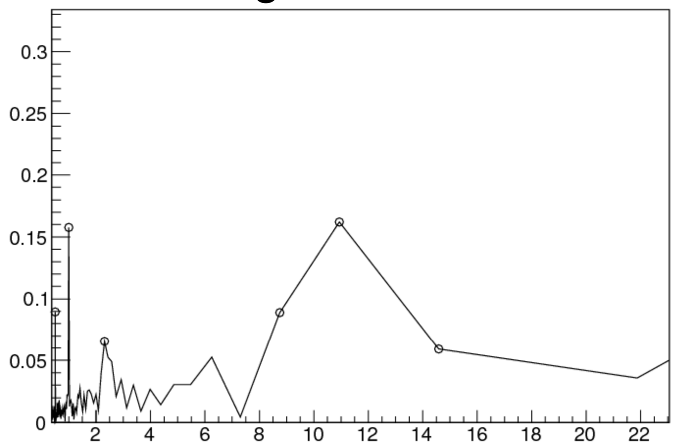
years

DFT of cluster affiliation



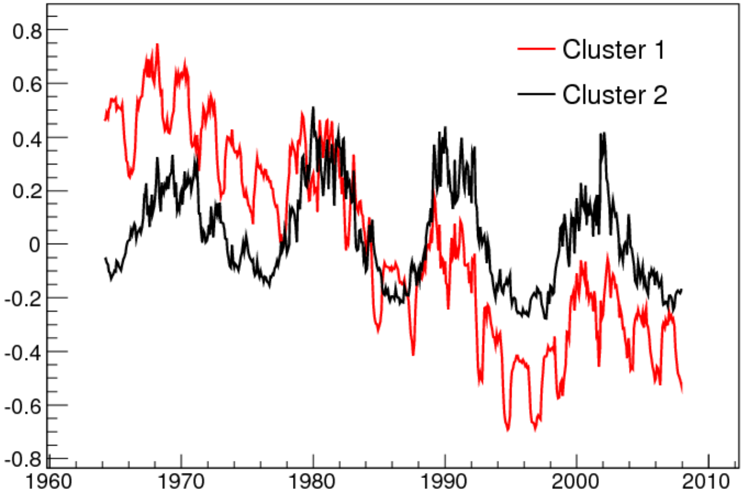
years/cycle

DFT Original Time series



Example from Observations

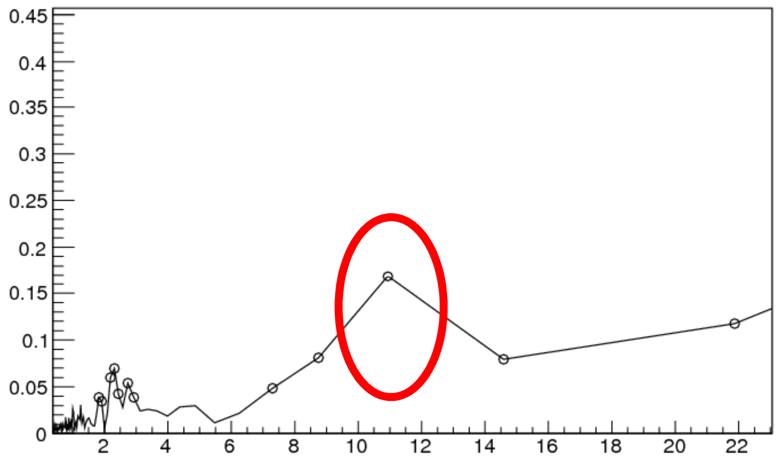
Expectation values of mean equilibrium positions E1(u) (Cluster 1) and E2(u) (Cluster 2)
u = vector of external factors



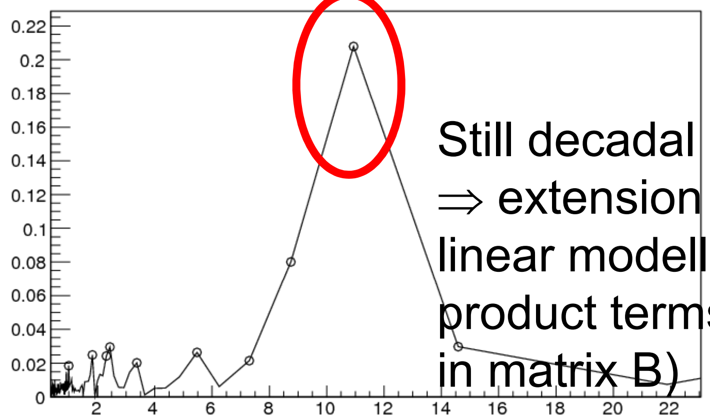
$$\mathcal{E}^{(i)}[u(t)] = \boldsymbol{\mu}^i + \sum_{q=1}^{m} \mathbf{A}_q^i \mathcal{E}^{(i)}[u(t)] + \mathbf{B}^i \boldsymbol{\phi}\ [u(t)]$$

(without noise)

DFT Cluster 1



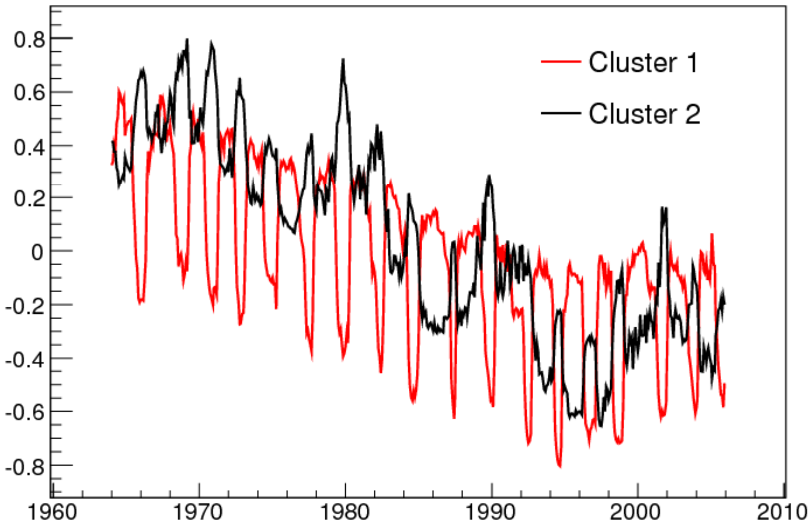
DFT Cluster 2



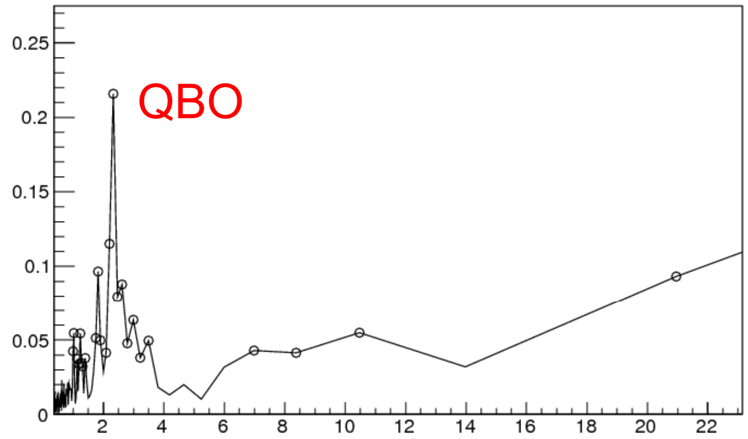
Still decadal signal $\Rightarrow$ extension of non-linear modell with product terms (SFL in matrix B)

Example from Chemistry-Climate Model

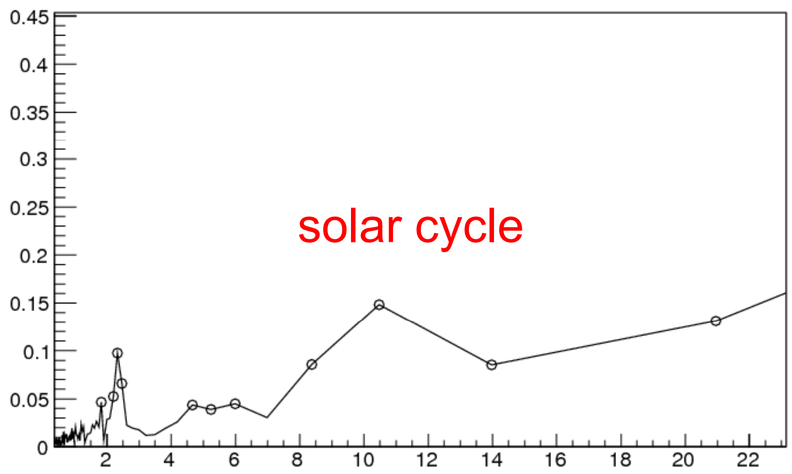
Expectation values of mean equilibrium positions E1(u) (Cluster 1) and E2(u) (Cluster 2)
u = vector of external factors



DFT Cluster 1



DFT Cluster 2



Neural Network Example

- Understanding the frequency, dynamics and causes of sudden stratospheric warmings (SSW) – „extreme events“

- Climate models (Charlton et al., 2007) are able to simulate SSWs in principle, but evaluating contributions of atmospheric variability factors remains a difficult task.

- Estimating the non-linear impact of factors like the QBO, ENSO, the Sun, the NAO, or stratospheric chlorine on wind and temperature data in the stratosphere (Labitzke et al., 2006; Camp & Tung, 2007; Calvo et al., 2009).

- Evaluating non-linear contributions of these variability factors to the probability of occurrence of SSWs.

Three Sets of Daily Data

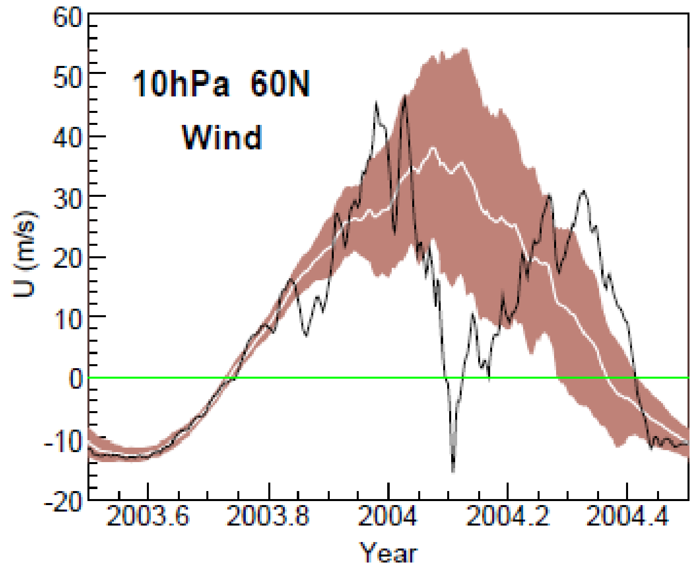
1. **ERA40** Reanalysis Extended from 1958 to 2008 (ERA40 + Operational ECMWF Analyses)
2. **NCEP**/NCAR Reanalysis from 1958 to 2008
3. **EMAC** CCMVal REF-B1 model output from 1960 to 2000 (Jöckel *et al.*, 2006; Morgenstern *et al.*, 2010)

The analyses were repeated with every time series starting at 1979 when the quality of data improved.

SSW Criteria at 10hPa (FUB)

$$U_{60\mathrm{N}} < 0$$
and
$$\Delta T := T_{90\mathrm{N}} - T_{60\mathrm{N}} > 0$$
and
Winter Period

→ Consider here: Nov. to Feb.



Idea

1. **Non-linear regression** on the time series $U_{60}(t)$, $T_{90}(t)$ and $T_{60}(t)$ with respect to the atmospheric factors
   - EESC (Equivalent Effective Stratospheric Chlorine, 3 year age)
   - MEI (Multivariate ENSO Index, NOAA)
   - QBO (Equatorial zonal wind at 50hPa, FU Berlin)
   - SFL (Solar Cycle, 10.7cm Radio Flux, NOAA)
   - NAO (North-Atlantic Oscillation, NOAA)

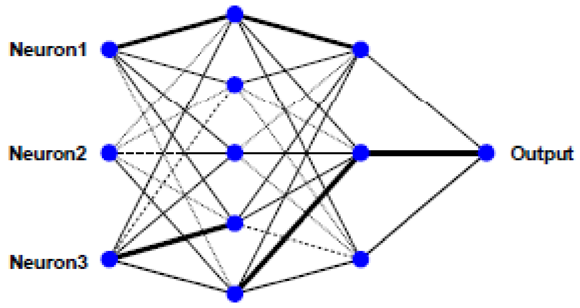
2. After regression, change the regressors and evaluate their **impacts** on the different time series

3. Look at the **change** in number of values that hold $U < 0$ ($\Delta T > 0$) to estimate the influence of certain regressors on the occurrence of SSWs

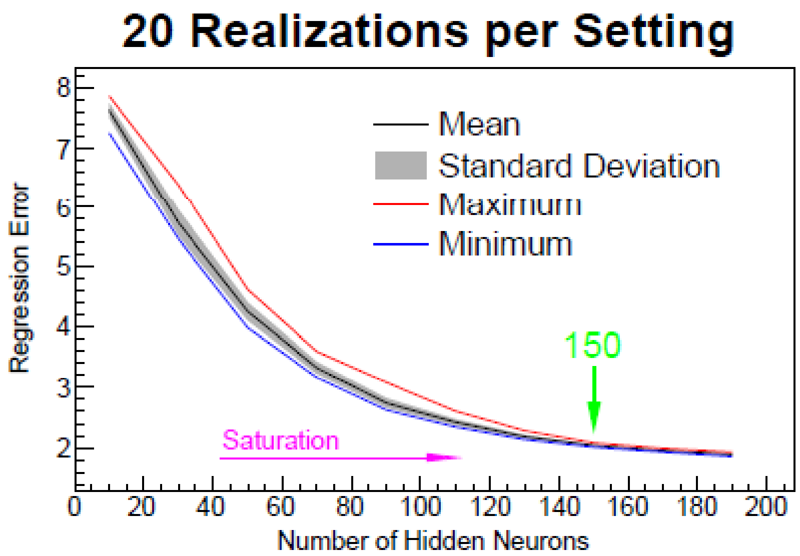
Statistical Model

A feed-forward **Artificial Neural Network** (ANN) performs the non-linear regression. (Bishop, 1995; Lu *et al.*, 2009)

Example Network



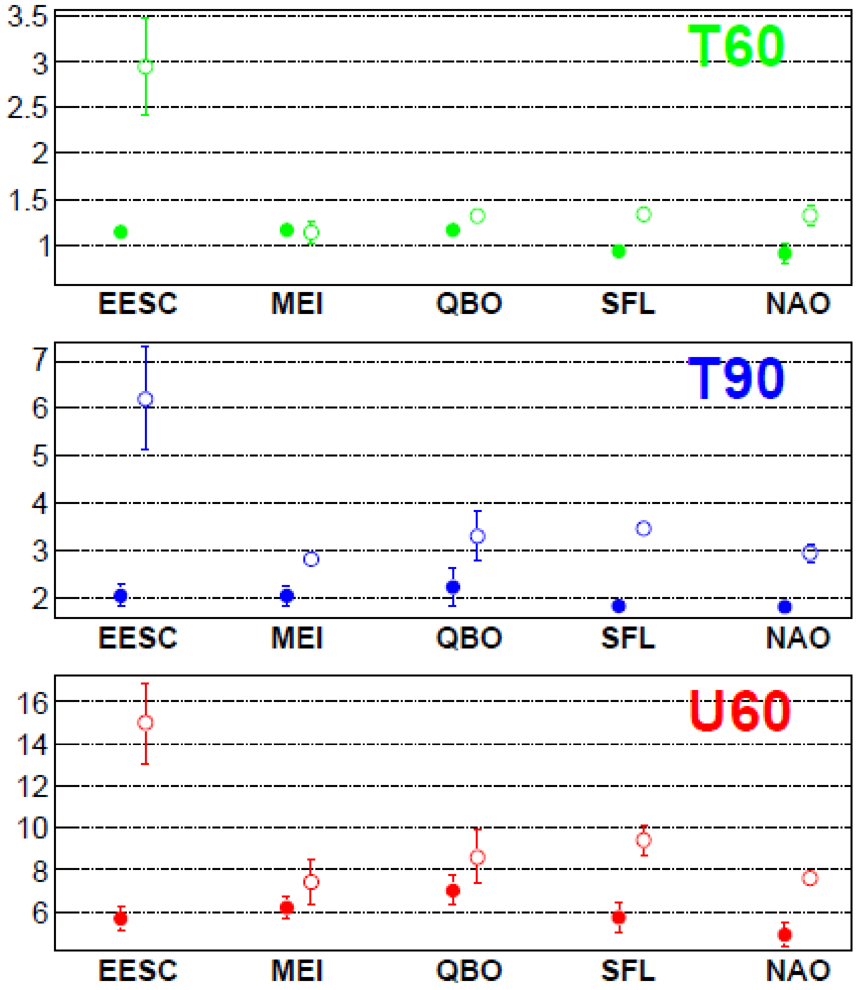
C++ Library at
`http://root.cern.ch`
Class: TMultiLayerPerceptron



In this analysis, the ANN has **5** normalized input neurons (regressors), **150** hidden neurons in one single hidden layer and **1** linear output neuron (response): **5,150,1**

⇒ The regression was successfully performed for both periods.

Impact (%) of Input Neurons for $\gamma = 0.1$



What is shown?

- Mean of the three data sets
- Shaded points: full time period (A)
- Hollow points: start at 1979 (B)

What does it mean?

- QBO and MEI make largest impact on ANN during (A)
- EESC, SFL and NAO make small impact during (A)
- During (B), SFL impact became large, QBO still important
- the largest impact makes EESC during (B) $\rightarrow$ Ozone loss

Conclusions and Outlook

- FEM-VARX successfully applied to stratospheric data
- Testing of extension of non-linear model to supress decadal signals
- Consideration of spatial patterns (via EOF)
- Repeat stratospheric warming example with FEM method and compare to results from neural network