The Mathematics of Downscaling

Hans von Storch and Eduardo Zorita (Institute for Hydrophysics, GKSS Research Centre, 21502 Geesthacht, Germany)

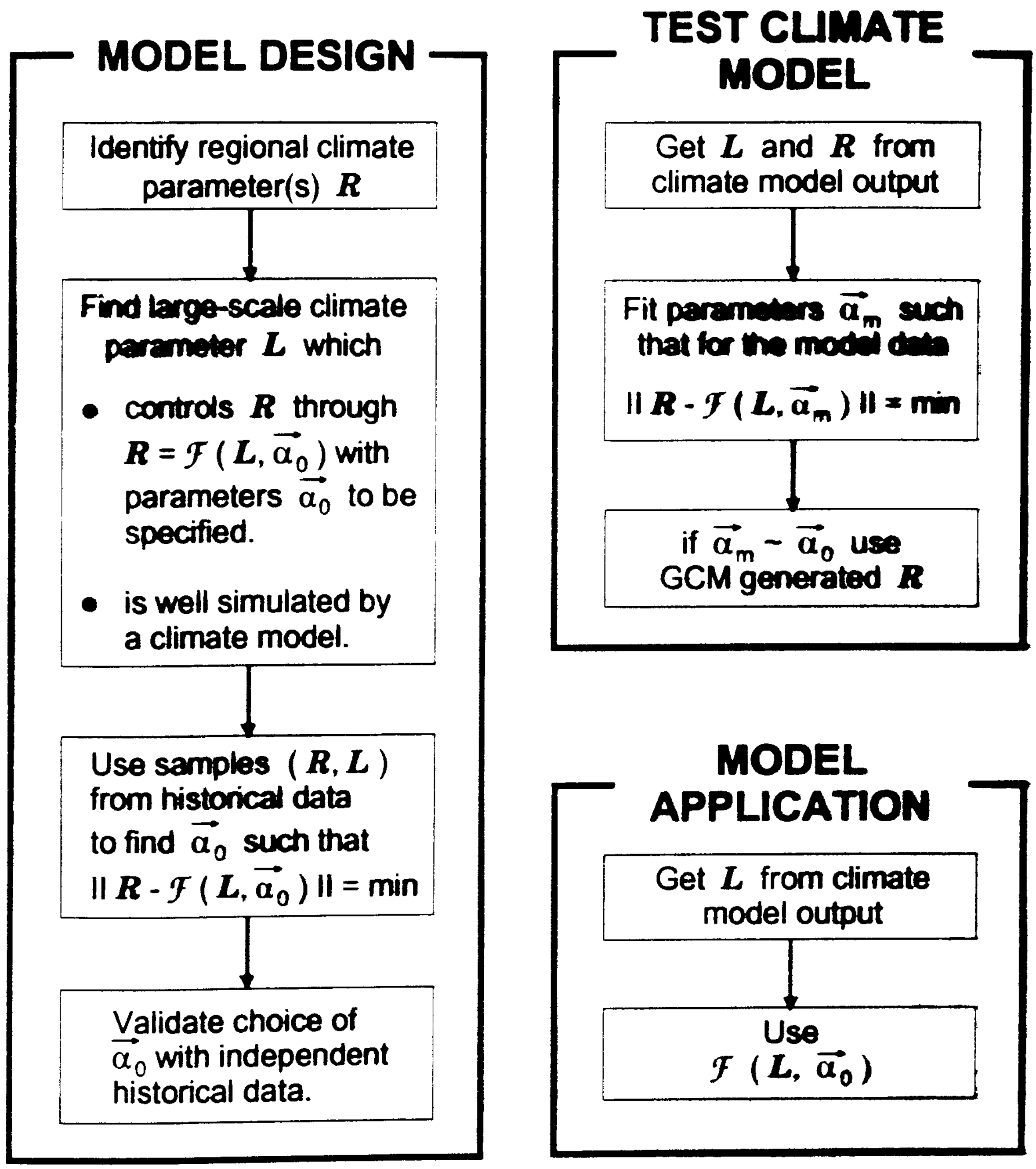
The basic idea of downscaling is to build dynamical, semi-empirical or purely empirical models which relate large-scale features of the state of the atmosphere or the ocean to local features of interest such as amounts of daily rainfall, storm-related sea level extremes or the blooming of snowdrops. The basic idea is illustrated, the large variety of different implementations is reviewed and the common mathematical structure and the underlying assumptions are discussed.

EGS Assembly 1996

Dear Hans

General Concept

von Storch, Zorita and Cubasch, 1993, J. of Climate
Hewitson and Crane, 1992, Geophys. Res. Lett.


MODEL DESIGN
Identify regional climate parameter(s) R
Find large-scale climate parameter L which
• controls R through $R = \mathcal{F}(L, \vec{\alpha}_0)$ with parameters $\vec{\alpha}_0$ to be specified.
• is well simulated by a climate model.
Use samples (R, L) from historical data to find $\vec{\alpha}_0$ such that $\| R - \mathcal{F}(L, \vec{\alpha}_0) \| = \min$
Validate choice of $\vec{\alpha}_0$ with independent historical data.
TEST CLIMATE MODEL
Get L and R from climate model output
Fit parameters $\vec{\alpha}_m$ such that for the model data $\| R - \mathcal{F}(L, \vec{\alpha}_m) \| = \min$
if $\vec{\alpha}_m \sim \vec{\alpha}_0$ use GCM generated R
MODEL APPLICATION
Get L from climate model output
Use $\mathcal{F}(L, \vec{\alpha}_0)$

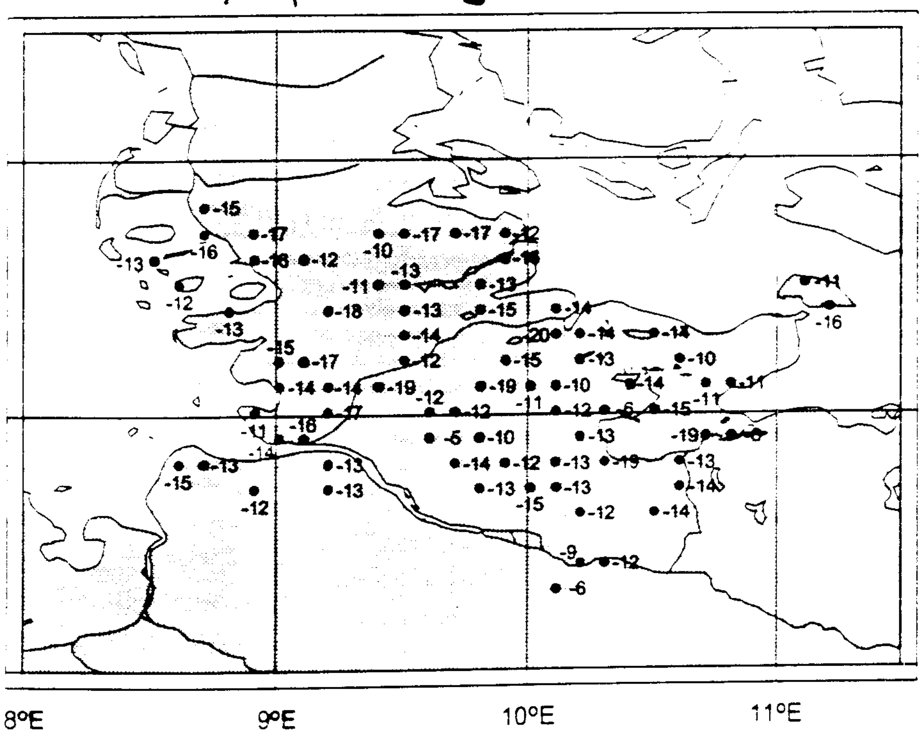
05HVSi.ds4

Example: Flowering of snowdrops in Northern Germany

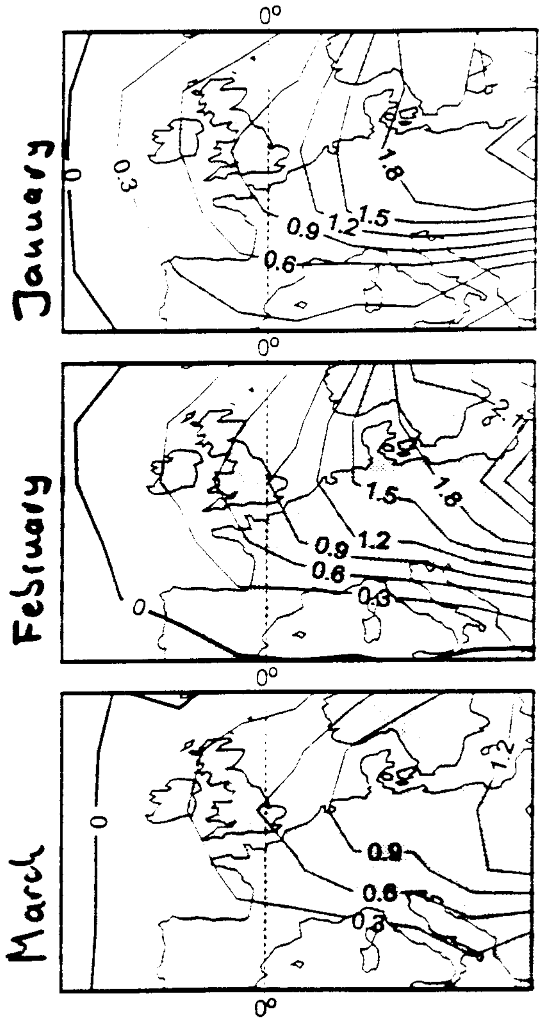
- see Maak and von Storch, 1996 (Int. J. Biometeor.)
- **R is the spatial distribution of the dates of the first flowering of snowdrops** in Schleswig Holstein
- **L is the European scale monthly mean temperature in January, February and March** before and while the snowdrops begin to flower.
- **Regression link established via a Canonical Correlation Analysis.**
- **Fit of regression model with 1971-90 observations. confirmation of model with independent data 1895-1900 and 1950-70.** Reconstruction of flowering dates from 1870 onwards.
- **Scenario for time of doubled carbon dioxide concentration.** GCM "2 CO_2" time slice experiment → snowdrops may flower two weeks earlier than presently.

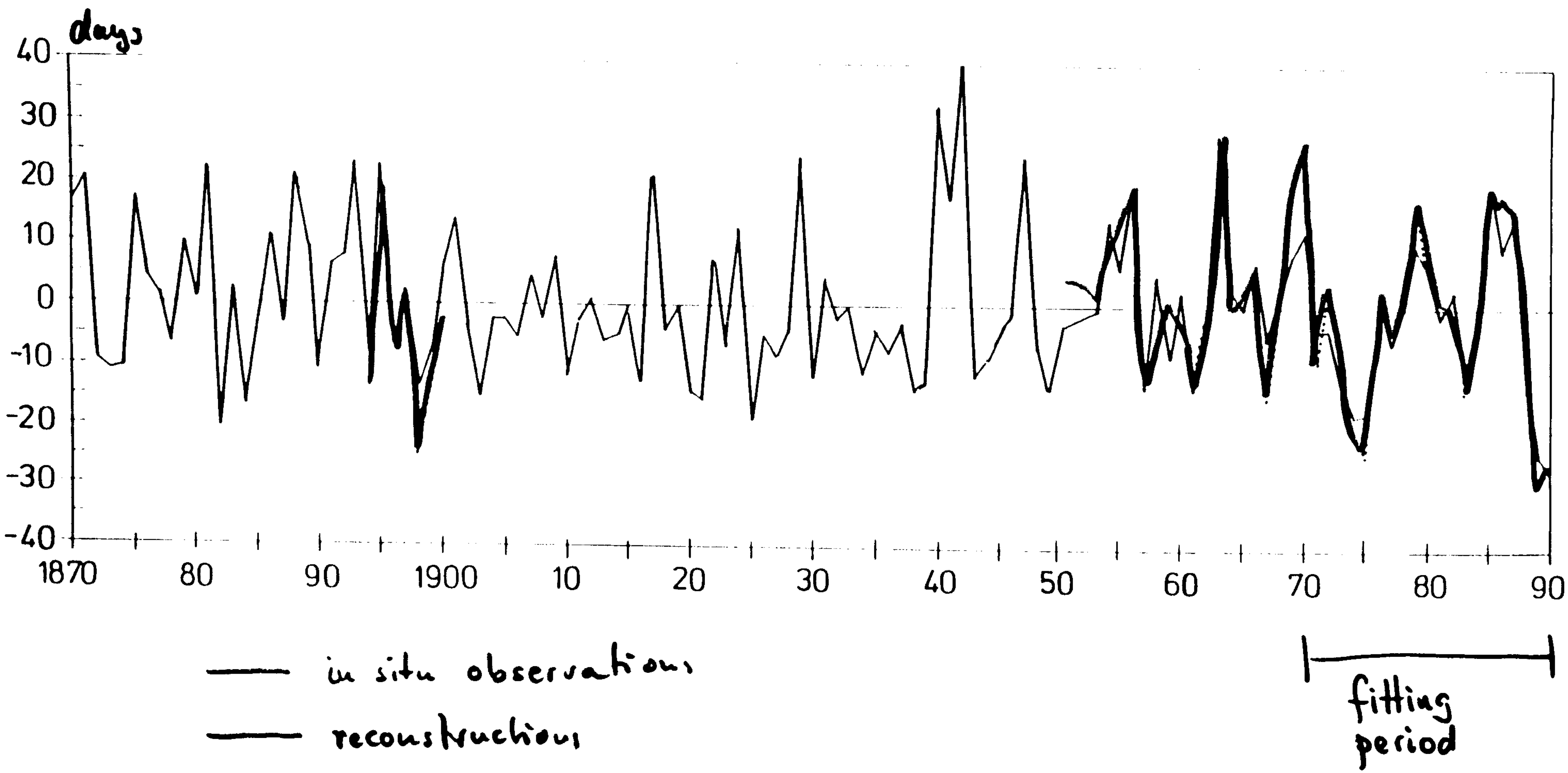
- diagram: patterns
- diagram: reconstuction and verification
- diagram: present and “predicted” distribution of dates

snowdrop flowering date



temperature





days
40
30
20
10
0
-10
-20
-30
-40
1870
80
90
1900
10
20
30
40
50
60
70
80
90
in situ observations
reconstructions
fitting period

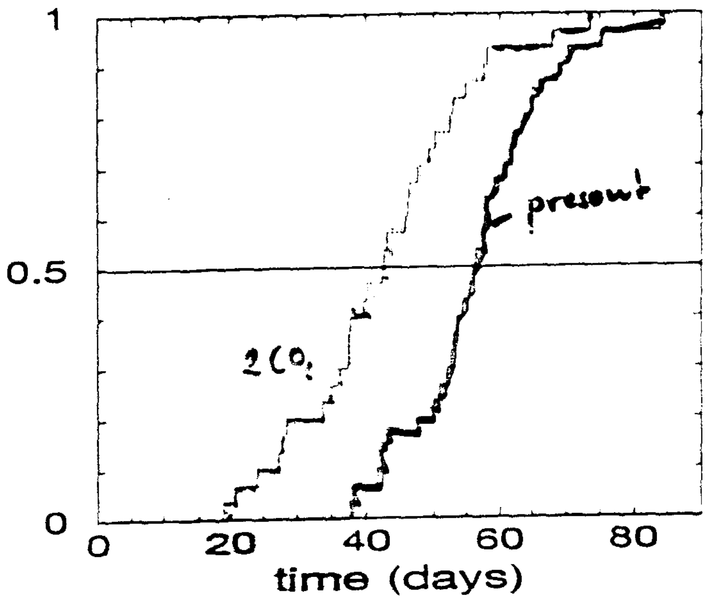


Figure 6: *Cumulative frequencies for the estimated bloom from the observed temperature of Jones and Briffa (solid line), the* $2 * CO_2$ *time-slice experiment (dotted line)* ~~*and the* $2 * CO_2$ *time-slice experiment (dashed line)*~~.

Downscaling . . .

- **. . . is related to**
  - synoptic climatology, and
  - the concept of parameterization of sub-grid scale processes
- **. . . can be done with dynamical models, empirical models and mixed models.** The level of technical complexity varies, but the basic concept is the same.
- **. . . may be seen as**
  - an interpolation problem, or
  - a conditional probabilistic problem

The Interpolation Ansatz

The local variable **R** is completely determined by the large scale **L**, i.e., the existence of a deterministic function $\mathcal{F}$ with

$$\mathbf{R} = \mathcal{F}(\mathbf{L}) \tag{1}$$

is postulated.

Then, the topography of $\mathcal{F}$ needs to be derived from a limited number of observations $(\mathbf{l}_k, \mathbf{r}_k)$. These observations may be exact or subject to uncertainties stemming from instrumental errors, analysis errors and other sources.

A climate change scenario $\mathbf{r}^*$ is derived by specifying an expected or hypothetical large-scale state $\mathbf{l}^*$ which is plugged into (1)

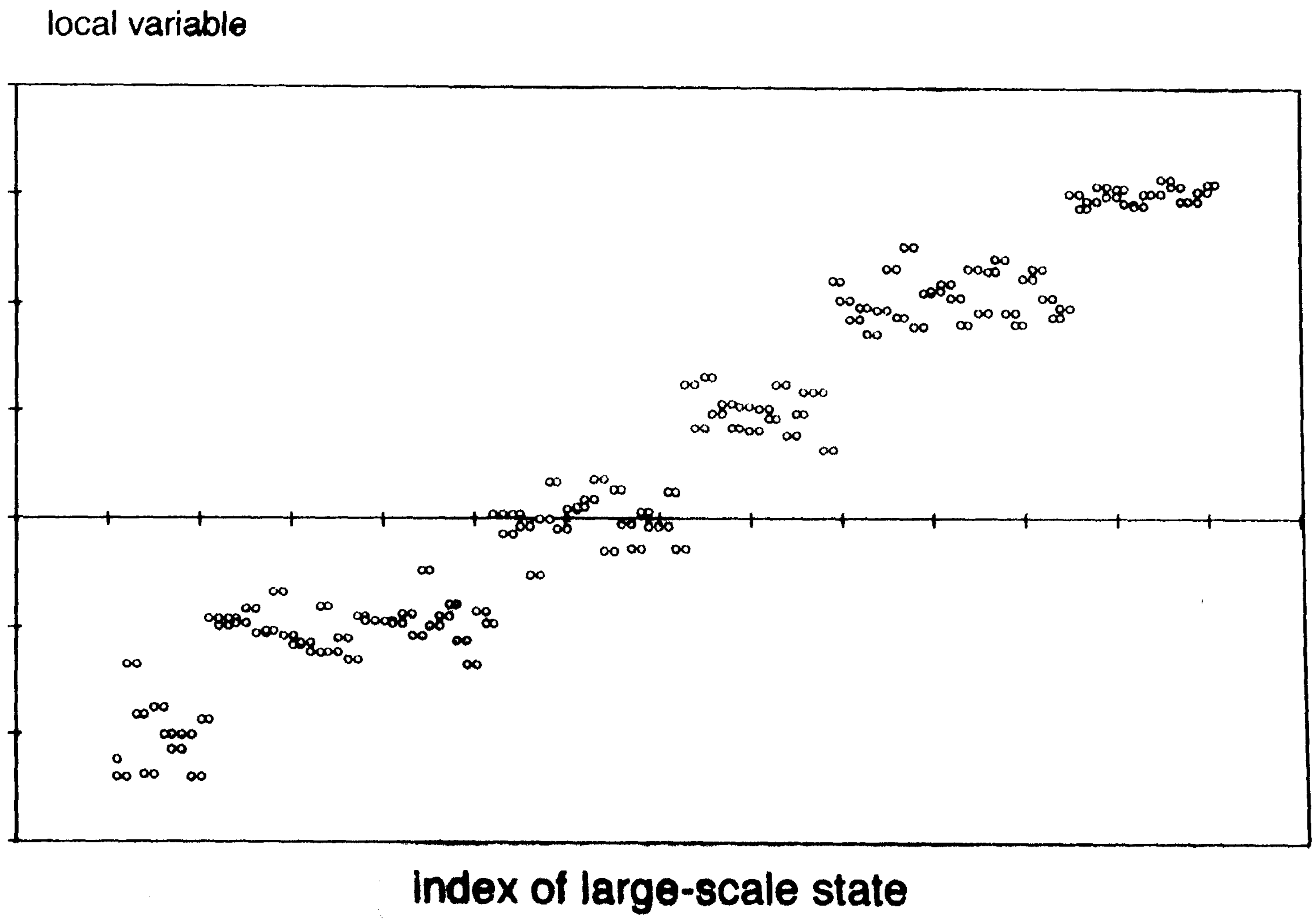
$$\mathbf{r}^* = \mathcal{F}(\mathbf{l}^*) \tag{2}$$

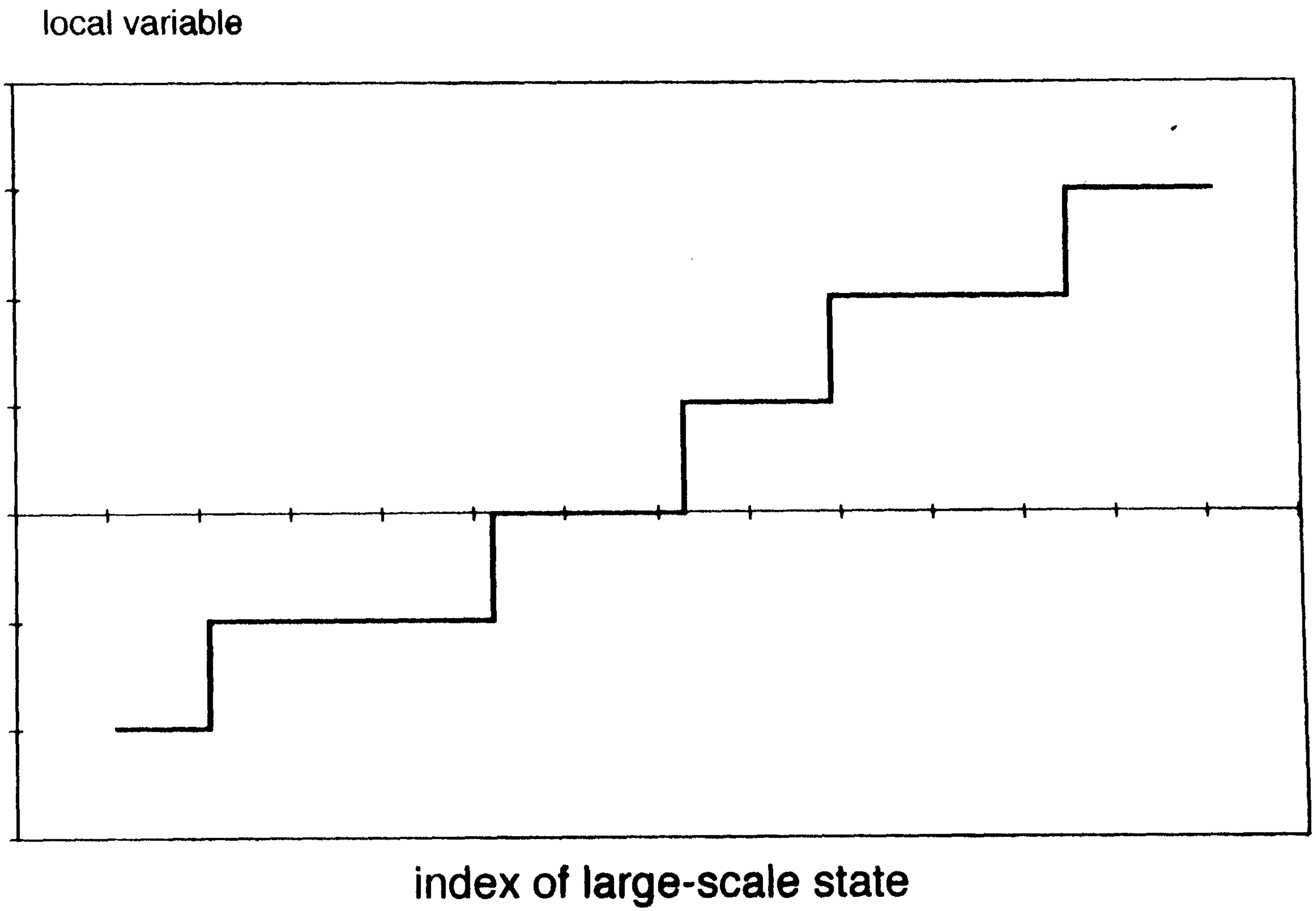
Working Assumption:

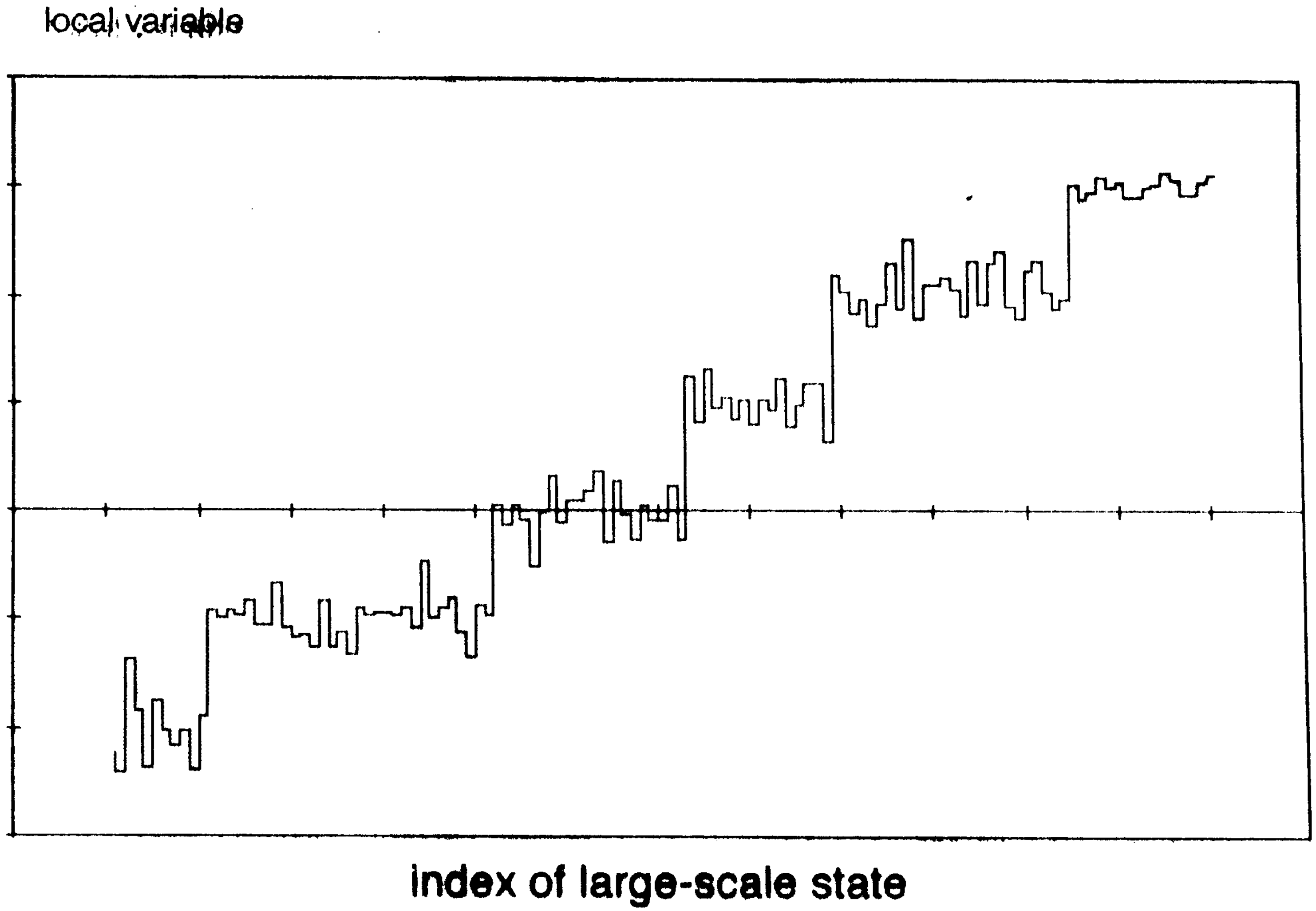
The topography $\mathcal{F}$ on the **L**-phase space is unchanged in the course of time (thus insensitive to different radiative regime). However, the system visits different parts of the **L**-phase space.

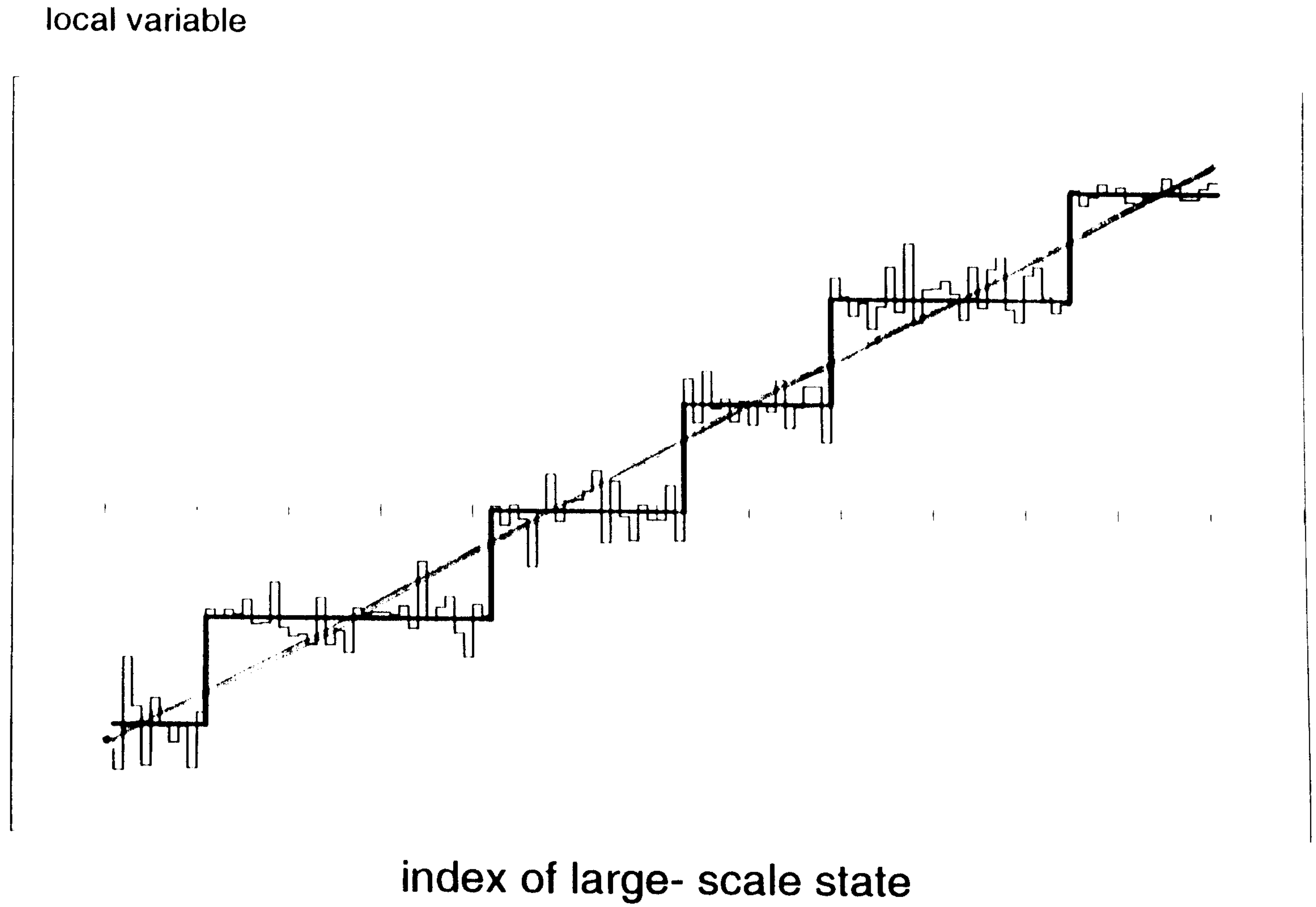
The interpolation of the multi-dimensional $\mathcal{F}$-topography may be pursued by various techniques, such as

- *kriging* and other geostatistical techniques,
- *neural networks* and other self-learning algorithms (e.g. McGinnis (1996), Hewitson (1996)),
- *splines* and piecewise constant functions determined by the nearest neighbor ("analogue") (e.g. Zorita et al. (1995), Cubasch et al. (1996)).


local variable
index of large-scale state


local variable
index of large-scale state


local variable
index of large-scale state


local variable
index of large- scale state

The Probabilistic Ansatz

The local variable $\vec{\mathbf{R}}$ is incompletely specified by the large scale $\vec{\mathbf{L}}$. The variability of $\vec{\mathbf{R}}$, as given by its density function $f_R(r)$, is caused by random processes, which are to some extent controlled by the large-scale state, which in turn is described by a density $f_L(l)$. Specifically, it is assumed that the probability density of $\vec{\mathbf{R}}$ may be factored such as (Katz and Parlange, 1996):

$$f_R(r) = \int f_{R|L=l}(r) f_L(l) dl \quad (3)$$

$$f_R(r) = \sum_l f_{R|L=l}(r) Prob(L = l) \quad (4)$$

if the large-scale state is described by the continuous or a discrete index $\mathbf{L}$.

Then, the *conditional probability densities* $f_{R|L=l}$ need to be determined with the help of a dynamical or statistical analysis.

A climate change scenario may then be constituted by determining a modified distribution $f_L(l)$ and by drawing one or several random samples from the random variable given by (3) or (4).

Working Assumption:

The conditional densities $f_{R|L=l}$ are the same under different climatic conditions.

Example: Sea Ice on the Western Baltic

The (estimated) probability distribution (in %) of (*strength of westerly flow, severity of Baltic Sea ice conditions*).
A: Distribution obtained from 104 years of data. From Koslowski and Loewe (1992).
B: Future distribution, if the marginal distribution of the westerly flow would change as indicated in the last row and if no other factors would control the ice conditions.

severity of the sea ice conditions	strength of the westerlies: strong	normal	weak	all
A: present distribution				
weak	21	11	2	34
moderate	20	14	7	41
severe	4	4	6	14
very severe	0	3	8	11
all	45	32	23	100
B: hypothetical future distribution				
weak	31	8	0	39
moderate	30	10	4	44
severe	6	3	3	12
very severe	0	2	4	6
all	67	23	11	101

The western Baltic Sea ice example is illustrative

but invalid

since it violates the working assumption that only the atmospheric circulation systematically affects the distribution of the severity of ice conditions. In this case, the systematic rise of temperature certainly is the major agent for systematical changes of sea ice statistics.

Dynamical Downscaling

In dynamical downscaling (e.g. Giorgi, 1990; Machenhauer et al., 1996), a regional climate model is integrated for an extended time with observed or simulated large-scale time-dependent boundary conditions.

This exercise may be understood as exposing the random variable "regional GCM" to a trajectory $\vec{\mathbf{L}}(t)$ in the $\mathbf{L}$-phase space. The system responds with a trajectory $\vec{\mathbf{R}}(t) = \mathcal{G}(\vec{\mathbf{L}}(t), t)$ in the $\mathbf{R}$-phase space. From this sample trajectory an estimate of the changed climate may be derived.

Großwetterlagen Approach: Deterministic Design

The **L**-phase space is subdivided into a finite number of areas "Großwetterlagen" $\mathcal{L}_k$, and for each of these a standard local weather R_k is specified:

$$R_k = \mathrm{E}(\mathbf{R}|\mathbf{L} \in \mathcal{L}_k) = \int r f_{R|L\in\mathcal{L}_k}(r)dr \qquad (5)$$

A scenario is then obtained as a weighted mean:

$$r^* = \sum_k R_k Prob(\mathcal{L}_k) \qquad (6)$$

The standard local weather may be determined by dynamical mesoscale models (Frey Buness et al., 1995; Fuentes and Heimann, 1996), or by empirical approaches (Bardossy and Plate, 1992; Conway et al., 1996; Enke and Spekat, 1996).

This approach is rooted in synpotic climatology. In the tradition of this approach, it usually relates instantaneous **R**- and **L**-states to each other.

Großwetterlagen Approach: Randomized Design

A conceptually different, but formally similar approach is to determine the set of all local weather states $\mathcal{R}_k$ observed during Großwetterlage $\mathcal{L}_k$:

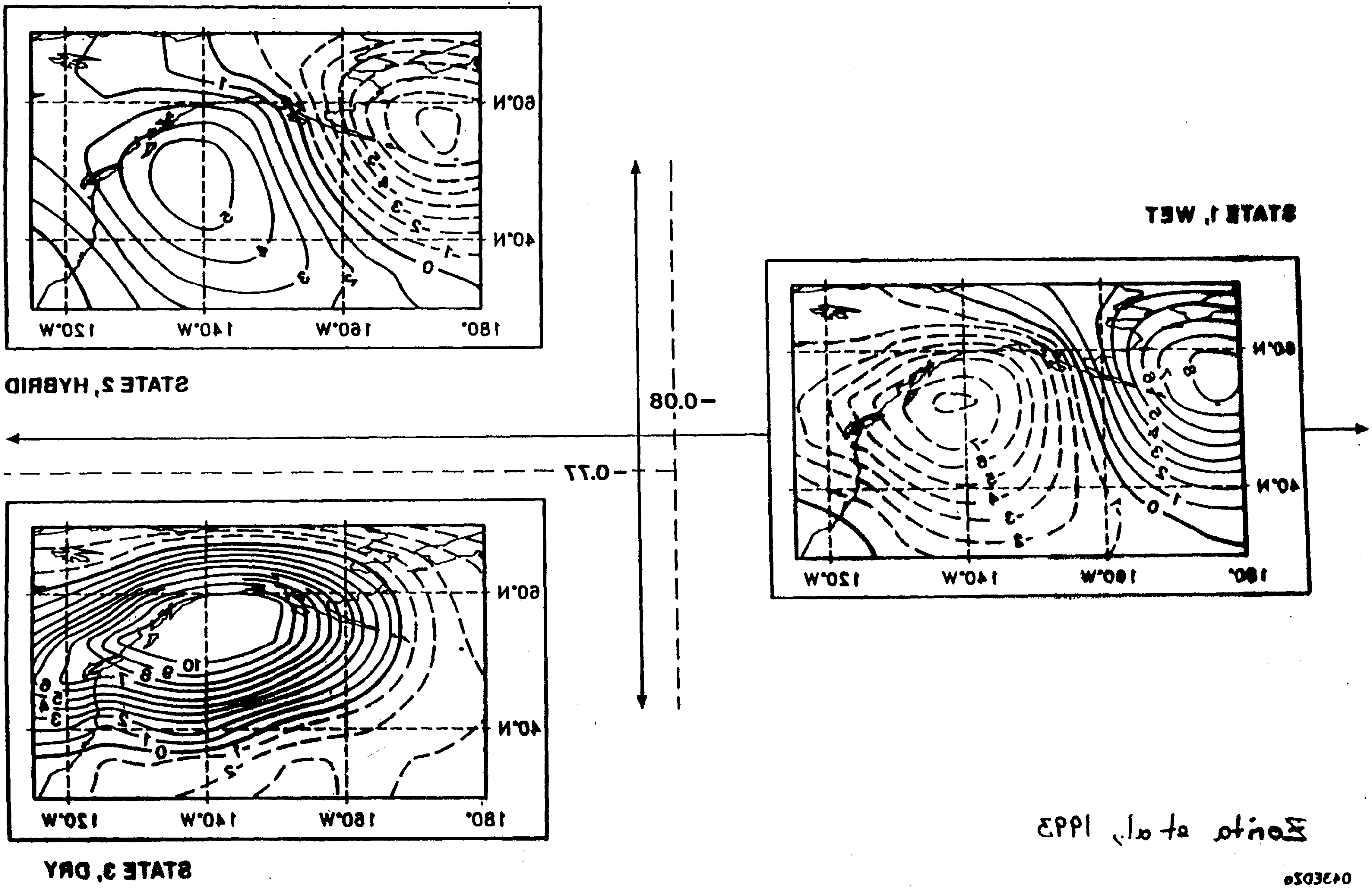
$$\mathbf{r}(t) \in \mathcal{R}_k \text{ if } \mathbf{l}(t) \in \mathcal{L}_k \tag{7}$$

A consistent local state $\mathbf{r}^*$ for a given large-scale state $\mathbf{l}(t)$ is obtained by a two-step approach: first, $\mathcal{L}_k$, with $\mathbf{l}(t) \in \mathcal{L}_k$, is identified; then $\mathbf{r}^*$ is randomly drawn from $\mathcal{R}_k$.

This approach was pioneered by Hughes et al. (1993), who optimized the division into Großwetterlagen by means of a CART algorithm.

A somewhat extreme approach was pursued by Zorita et al. (1995) who formed $\mathcal{L}_k$ sets which contained only one observation ("analogue approach").

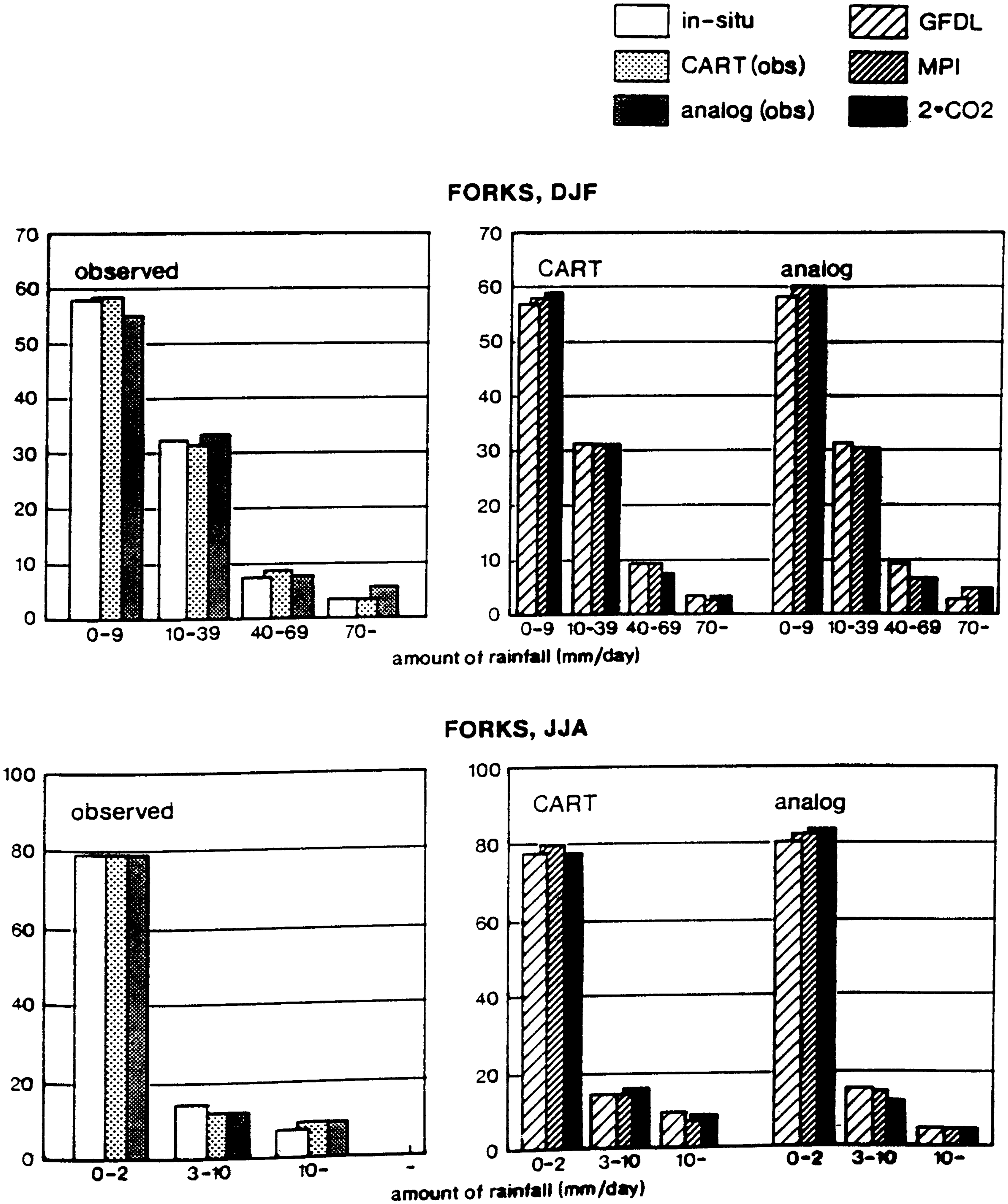
This approach is not limited to instantaneous large-scale local links in the tradition of synoptic climatology. Indeed, monthly mean spatial distributions have been used as $\mathbf{L}$ to specify a vector of *intramonthly* percentiles of precipitation (Cubasch et al., 1996).



Zorita et al., 1993

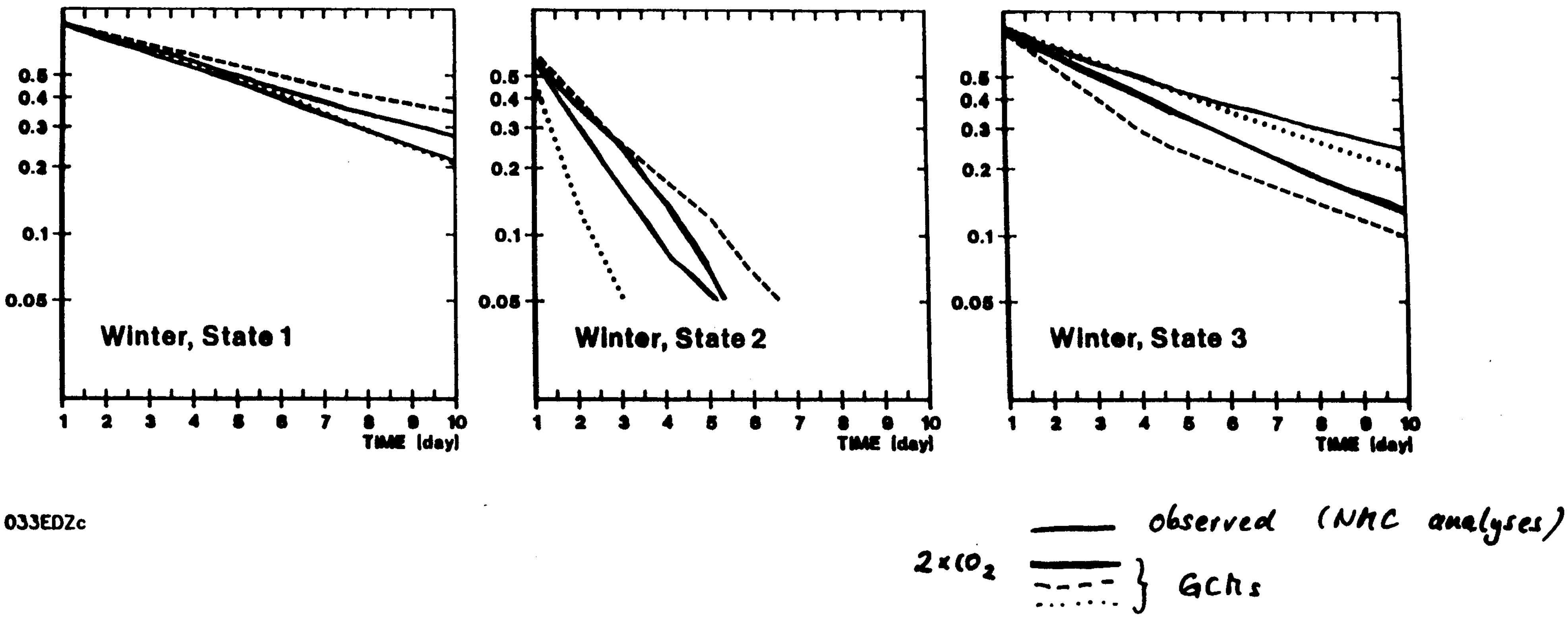
043ED20

Zorita et al. ....



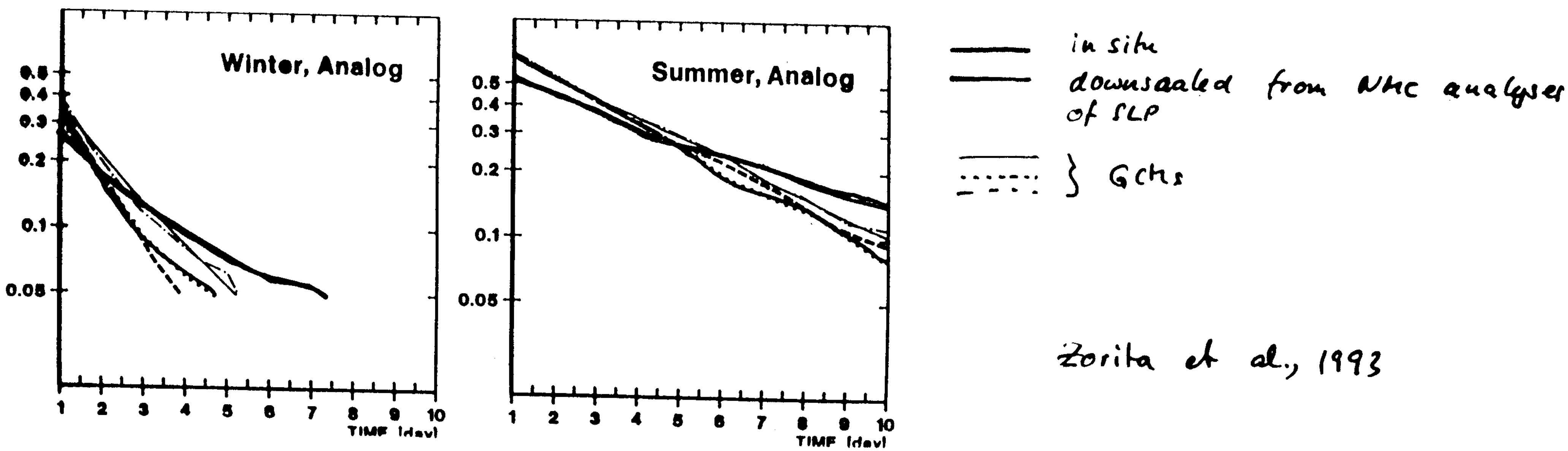
in-situ
CART (obs)
analog (obs)
GFDL
MPI
2•CO2
FORKS, DJF
observed
CART
analog
0-9
10-39
40-69
70-
amount of rainfall (mm/day)
FORKS, JJA
observed
CART
analog
0-2
3-10
10-
amount of rainfall (mm/day)

Log-survivor functions of CART weather states
(Zorita et al., 1993)



033EDZc

Log-survivor function
of storm interarrival time



Zorita et al., 1993

Regression Approach

$$\mathbf{R} = \mathcal{F}(\mathbf{L}) + noise \quad (8)$$

with some linear or nonlinear function $\mathcal{F}$ and a zero-mean noise term whose variance and time-memory may depend on $\mathbf{L}$ or not. When the noise term is disregarded, we are back to the interpolation problem; indeed the function $\mathcal{F}$ may be determined by any of the techniques mentioned there.
In most application (von Storch et al., 1993; Hewitson and Crane, 1992; Bürger, 1996; Kaas et al., 1996; Dehn et al., 1996), the function $\mathcal{F}$ was chosen to be linear. The system (8) is fitted to data in a low-dimensional subspace of $\mathbf{L}$, often accomplished by a combination of the EOF and CCA techniques.

When reproducing past data, the regression model is used *without* noise term; then the variance of the estimated local variable is too small since the unexplained variability is taken away.
When a scenario of future trajectories $\mathbf{R}(t)$ is needed then such an underestimation may be disadvantegous; then, the unexplained part should be modelled with an adequate stochastic process (say cyclostationary AR(1)) and the full model, *with* noise, should be run.

Often, the expression "downscaling" is limited to the empirically determined regression model (8). This limitation is inadequate.

Conclusion

Downscaling may be seen as an

- interpolation problem, or
- a conditional probability problem.

To some extent, downscaling is a late spin-off of synoptic climatology, but with major differences

- The goal is *not* to specify accurately individual numbers, as is required in weather predition. Instead, conditional statistics of variables of interest are generated.

- Downcaling is *not* limited to instantaneous dynamically direct links such as upper air flow and local response. Instead indirect links, for instance between monthly mean air pressure fields and a vector of intramonthly pecentiles of storm-related water levels at a tide gauge, may be usd as well.

- The „preditor“ *must* be well simulated by GCMs.

Downscaling may be used for

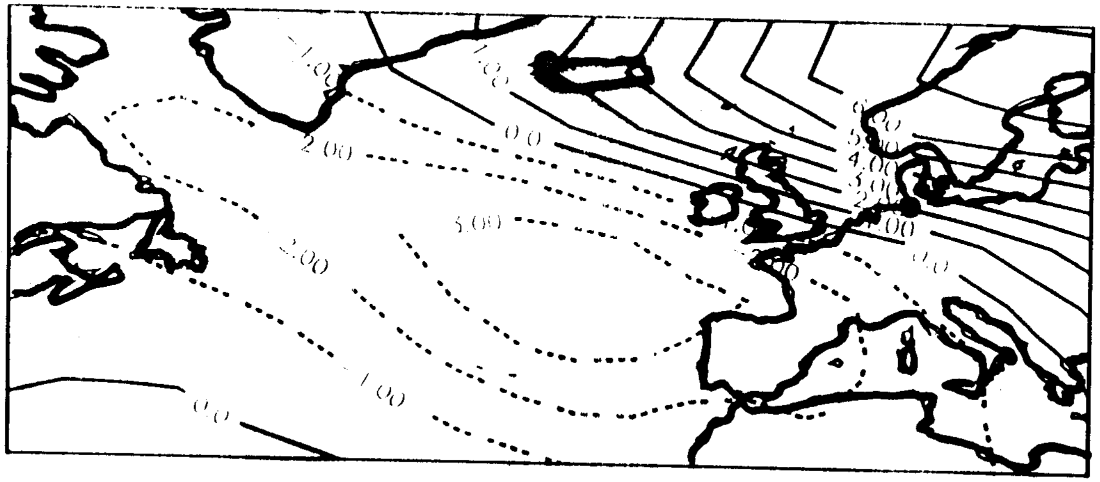
- the generation of lcoal climate change scenarios by postprocesing GCM output,

- the reconstruction of past local climate variations (as far as large-scale „specifying“ states are available), and

- for the verification of a GCM in simulating a local climate.

GCMs which simulate the local climate right, i.e., which have correctly incorporated the dynamical „large-scale - local link“ may be used to test the skill of statistical downscaling procedures in simulating *climate change*.

monthly mean SLP anomaly



intramonthly percentiles of high freq. water level anomalies

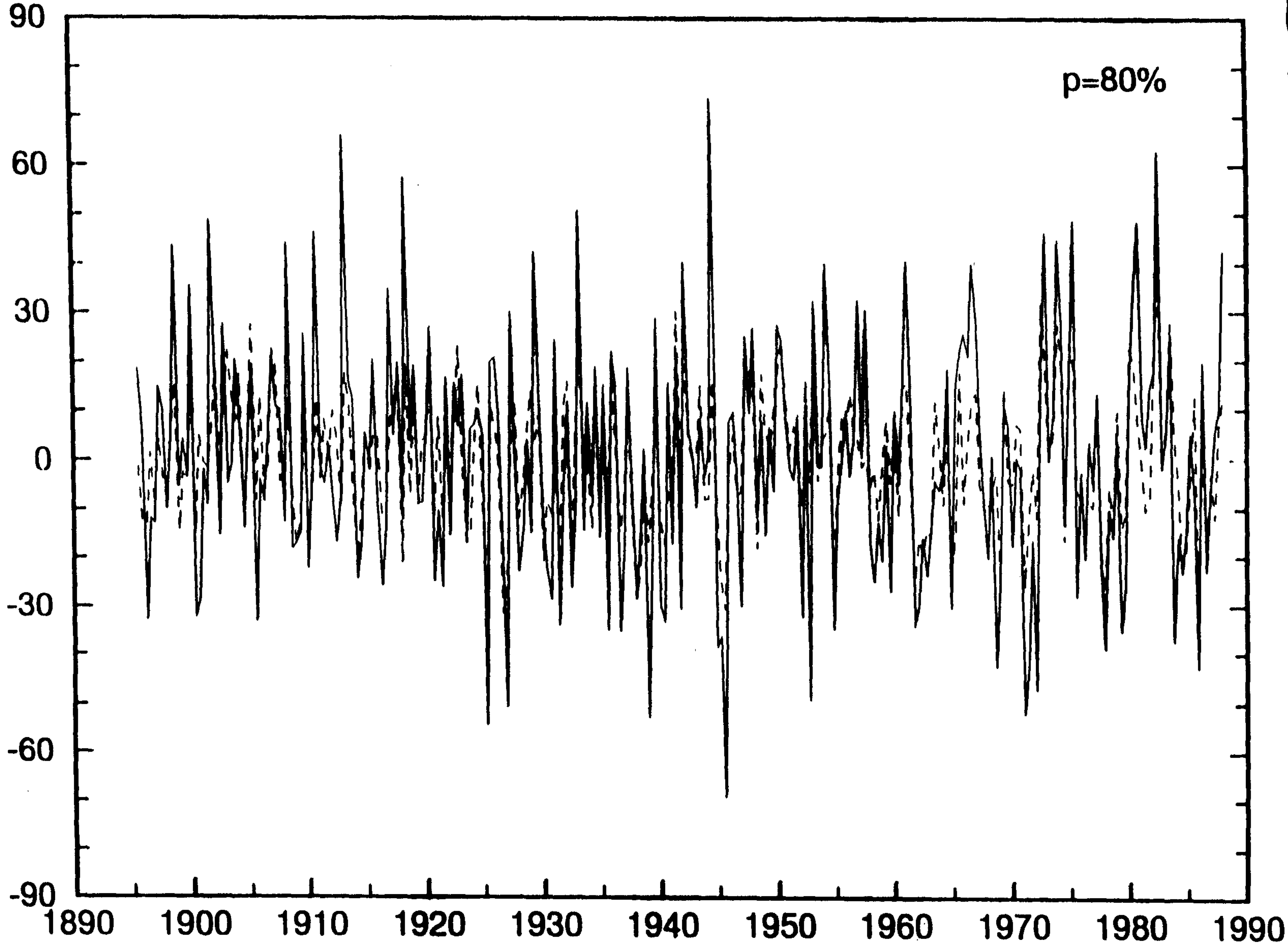
$q_{50\%}$ = -21 cm

$q_{80\%}$ = -16 cm

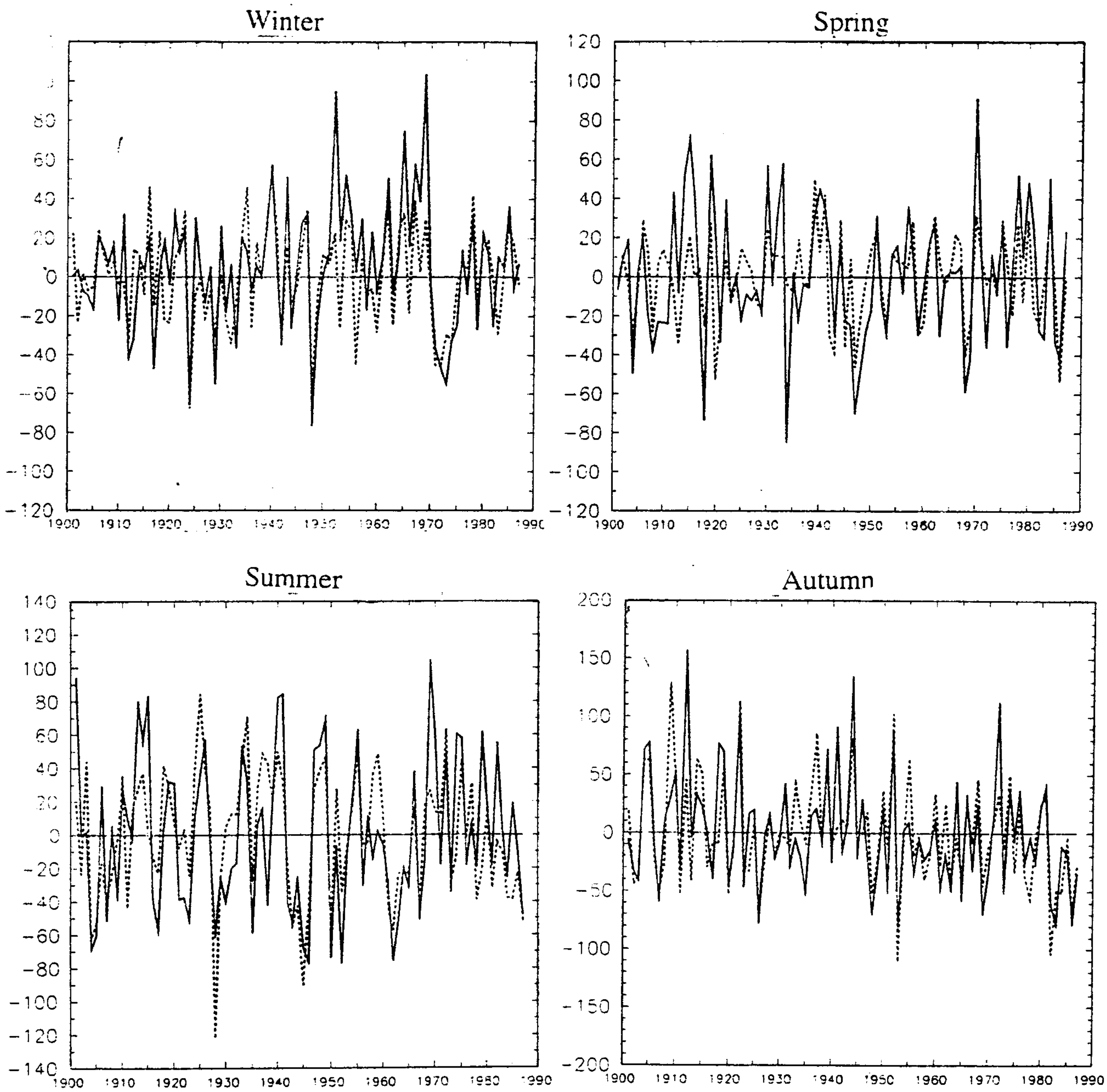
$q_{90\%}$ = -18 cm

at Cuxhaven

80% percentile of storm related intramonthly variability of water level in Cuxhaven.

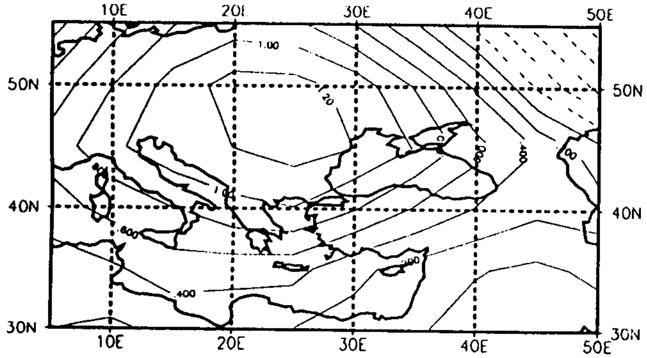


$\Delta_{2CO_2} \sim 7\,c$

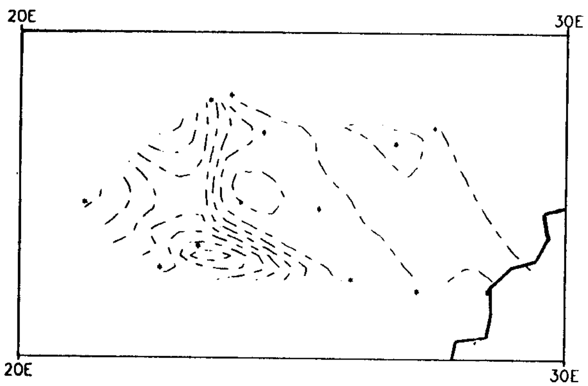


Winter
Spring
Summer
Autumn
120
100
80
60
40
20
0
-20
-40
-60
-80
-100
-120
140
-140
200
150
100
50
-50
-100
-150
-200
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990

CCA Autumn SLP 1901–1987



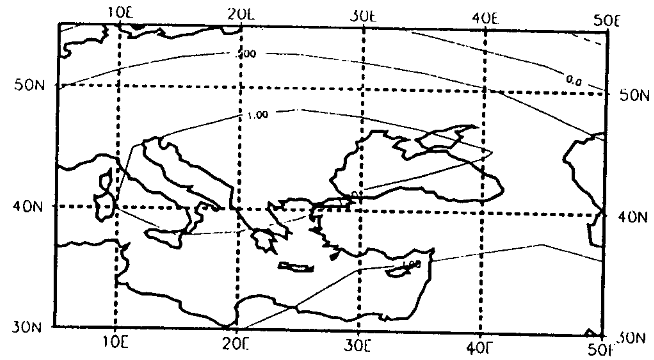
1 CCA cor: 0.81 , var: 0.2



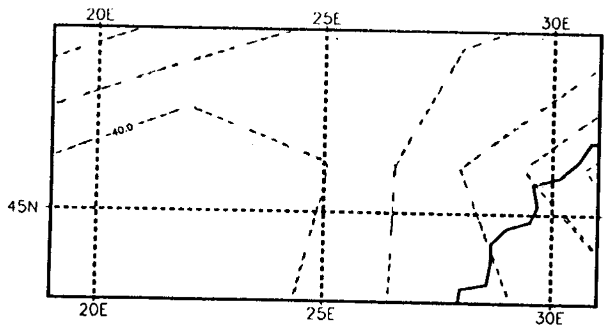
CONTOUR FROM -80.000 TO 60.000 CONTOUR INTERVAL OF 5.0000

1 CCA cor: 0.81 , var: 0.47

CCA SLP Autumn T42

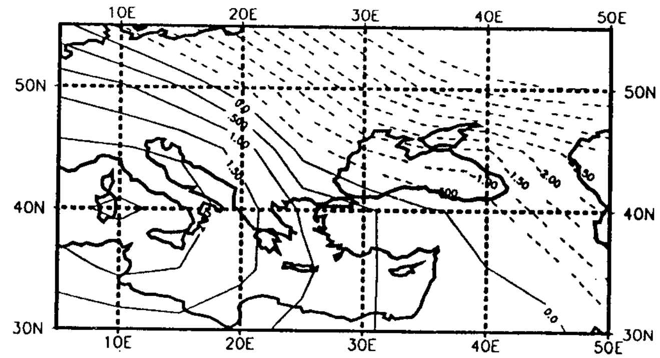


1 CCA cor: 0.82 , var: 0.1

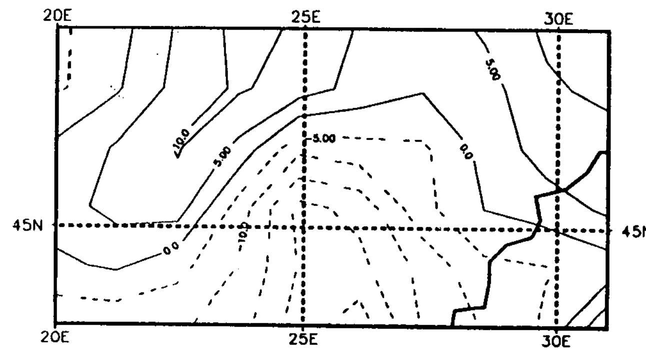


1 CCA cor: 0.82 , var: 0.6

CCA monthly autumn SLP T106

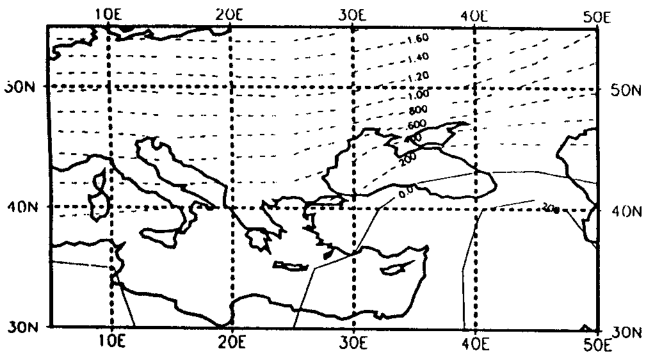


1 CCA cor: 0.84 , var: 0.7

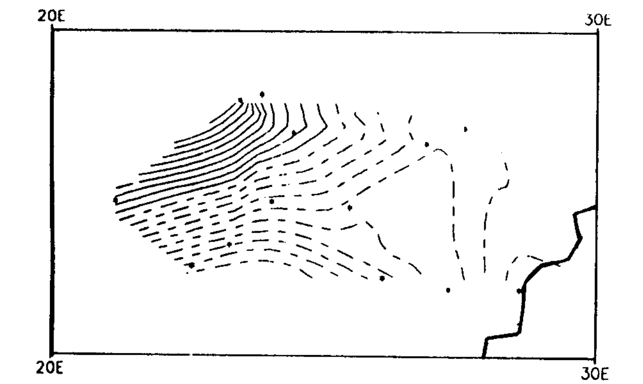


1 CCA cor: 0.84 , var: 0.3

CCA Autumn SLP 1901 – 1987



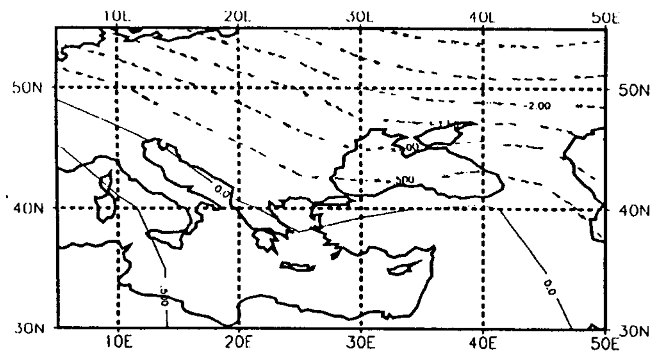
2 CCA cor: 0.63 , var: 0.3



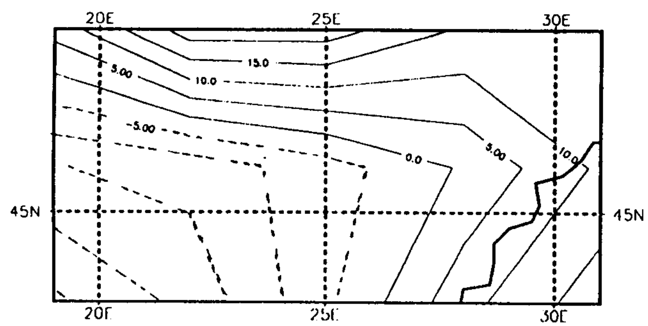
CONTOUR FROM -80.000 TO 60.000 CONTOUR INTERVAL OF 5.0000

2 CCA cor: 0.63 , var: 0.17

CCA SLP Autumn T42

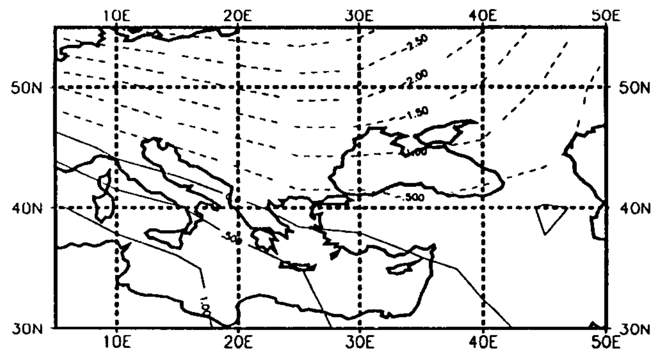


2 CCA cor: 0.62 , var: 0.5

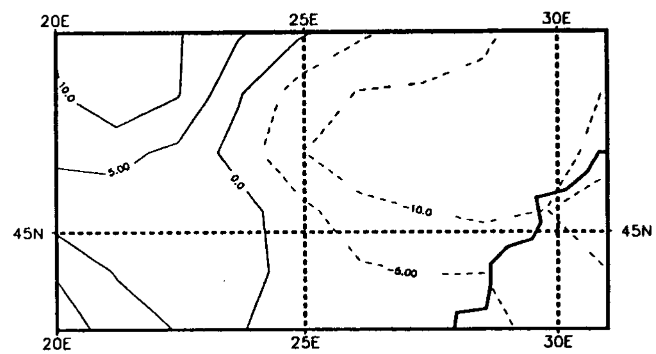


2 CCA cor: 0.62 , var: 0.1

CCA monthly autumn SLP T106

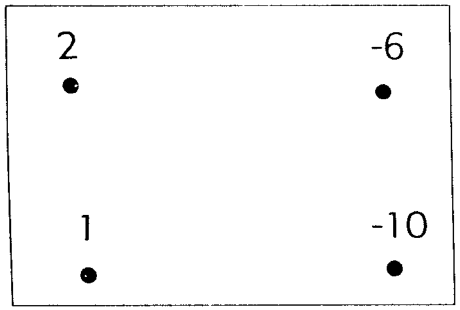


2 CCA cor: 0.54 , var: 0.2

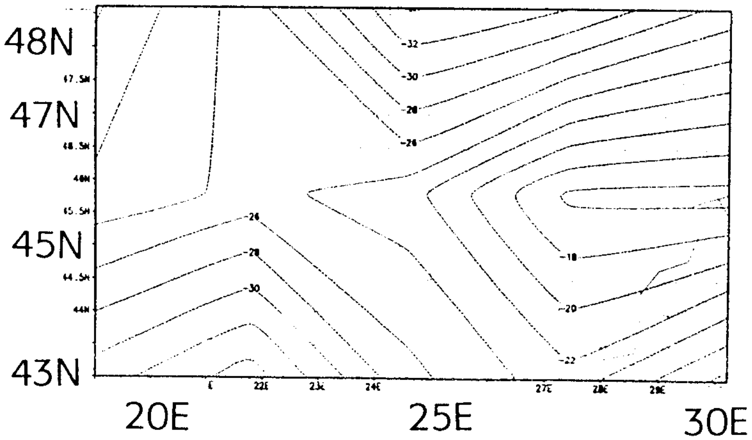


2 CCA cor: 0.54 , var: 0.2

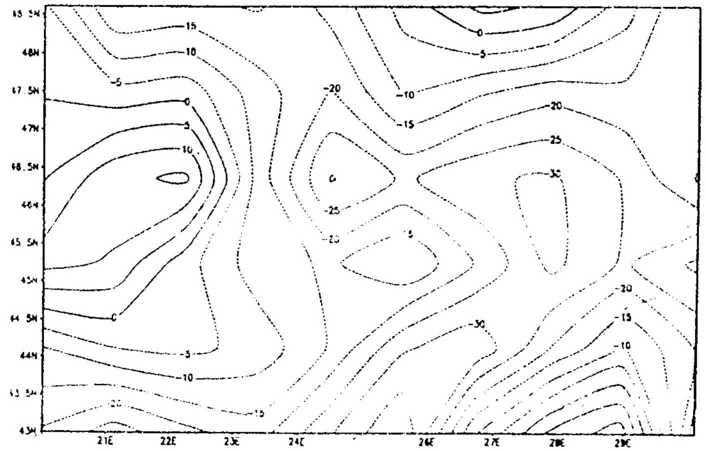
T 21



T 42 Autumn



T 106 Autumn



T21