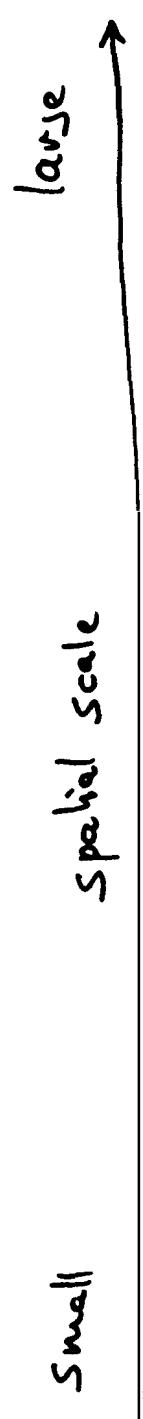


# Conditional Statistical Models:

A Discourse about the Local Scale  
in Climate Simulations.

Hans von Storch<sup>†</sup>  
Gloss Research Centre, Germany

† frequentist.



Few DGFs  
DETAILS matter  
described by numerical  
model

MANY DGFs  
statistics matter  
not completely repres.  
by numerical models

by uniflow.

LOCAL

SCALE

## Overview

- the local scale
- conditional statistical models
- Downscaling
  - Weather forecasting
  - "best guess" reconstruction
  - Simulating
- PARAMETRISATION
  - general
  - example: albedo in an Energy Balance Model (EBM)
- Conclusion.

# Conditional Statistical Models

local parameter  $X$

large-scale parameter  $G$

$$f_X(x) = \int f_{X|G=g}(x) f_G(g) dg$$

Then

$$E(X) = E_g(E_x(X|G))$$

$$\text{Var}(X) = \text{Var}_g(E_x(X|G)) + E_g(\text{Var}_x(X|G))$$

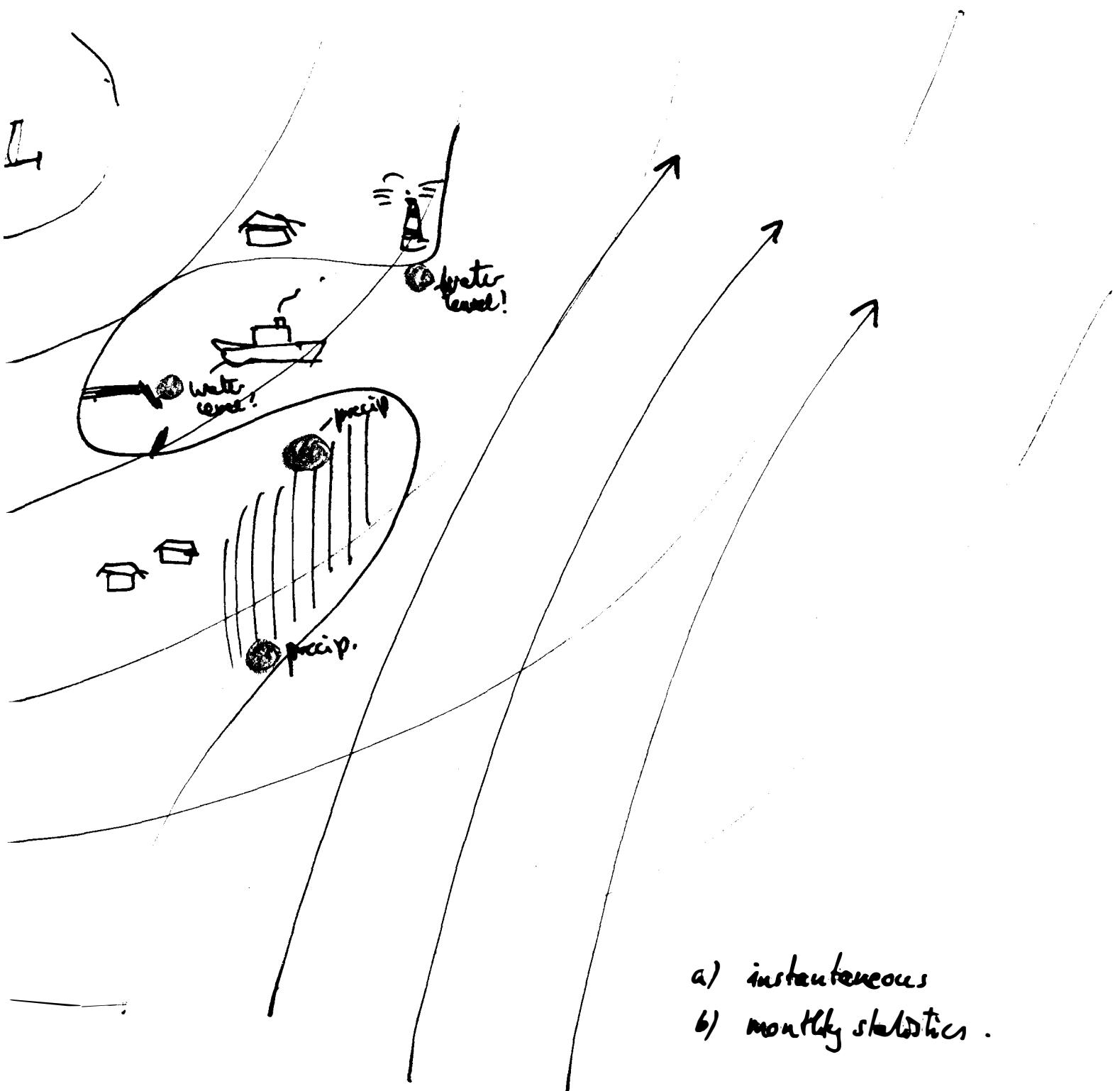
Example: Regression  $X_t = \mu_0 + \beta G_t + N_t$   
 $N_t$  and  $G_t$  independent  
 $N_t$  white (red) noise)

then  ~~$E_x(X|G)$~~   $E_x(X|G) = \mu_0 + \beta G$   
 $E(X) = \mu_0$  if  $E(G) = 0$

and  $\text{Var}_g(E_x(X|G)) = \beta \sigma_g^2$   
 ~~$E_g(\text{Var}(X|G)) = E_g(\mu_0 + \beta G_t + N_t - \mu_0 - \beta G_t)^2$~~   
 $= \sigma_u^2$

$$\Rightarrow \sigma_x^2 = \underbrace{\beta^2 \sigma_g^2}_{\text{externally induced}} + \underbrace{\sigma_u^2}_{\text{internal uncertainty / variability}}$$

# The downscaling problem



- a) instantaneous
- b) monthly statistics .

1. The General Nature of Weather Forecasting. The general problem of forecasting weather conditions may be subdivided conveniently into two parts. In the first place, it is necessary to predict the state of motion of the atmosphere in the future; and, secondly, it is necessary to interpret this expected state of motion in terms of the actual weather which it will produce at various localities. The first of these problems is essentially of a dynamic nature, inasmuch as it concerns itself with the mechanics of the motion of a fluid. The second problem involves a large number of details because, under exactly similar conditions of motion, different weather types may occur, depending upon the temperature of the air involved, the moisture content of the air, and a host of local influences.

In case of downscaling, the local  $X_t$  is assumed to be given by

$$X_t \sim \mathcal{P}(\vec{\alpha}; X_{t-1})$$

with some probabilistic model  $\mathcal{P}$  with parameters  $\vec{\alpha}$  (mean value, covariance matrix ...).  
For simplicity

$$X_t \sim \mathcal{P}(\vec{\alpha})$$

In downscaling applications

$$\vec{\alpha} = \mathcal{F}(G)$$

so that

$$X_t \sim \mathcal{P}(\mathcal{F}(G_t))$$

This set-up transforms a sequence of large scale states into a sequence of local random variables.

Regression case:  $X_t = \mu + N_t$

and  $\mu = \mu_0 + \beta G_t$

$$\Rightarrow X_t = \mu_0 + \beta G_t + N_t.$$

## BEST GUESS

To a given  $G$  belongs a local  $X$  which has not been observed. How to estimate  $X$ ?

$$\hat{X} = E[\mathcal{P}(f(G))] \quad \text{in general}$$
$$= \mu_0 + \beta G_L \quad \text{in the regression case}$$

GOOD:  $\hat{X}$  is in certain cases "best" in the sense of  $E(\|\hat{X} - X\|^2) = \min.$

BAD:  $\text{Var}(\hat{X}) < \text{Var}(X)$   
thus unsuited for further studies of processes influenced by  $X$ .

## Example

Snow drops (flowers!)  
in SCHLESWIG-HOLSTEIN  
(N. GERMANY)



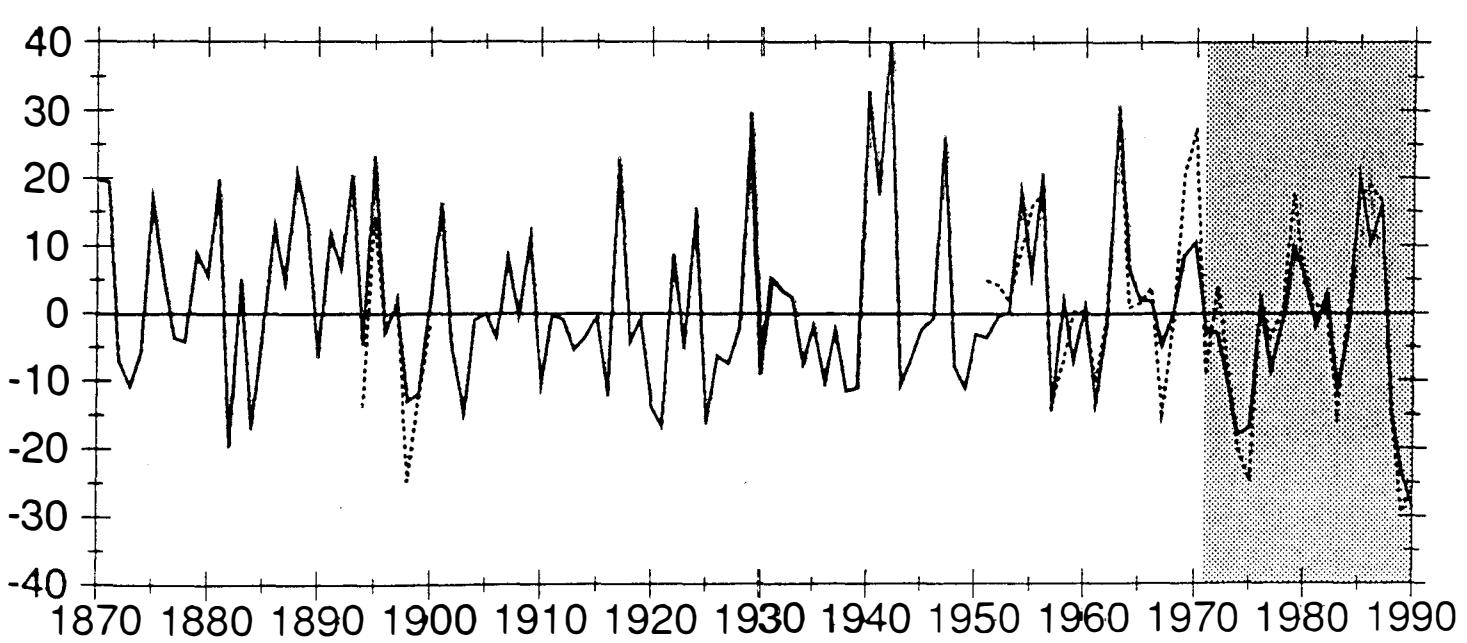
- $X_t$  = date of first flowering in year t.  
 $G_t$  = spatial distribution of JFM time average of temperature  
 $\hat{F}$  = regression

Fit of model: 1971-90

Verification of model: 1896-1900, 1954-70 ~~1960-70~~

Reconstruction "best guess": remaining years.

$$\begin{aligned} \text{Var}(X) &= (16.5 \text{ days})^2 \\ \text{Var}(\hat{X}) &= (14.9 \text{ days})^2 \\ \sigma_n &= 7 \text{ days.} \end{aligned}$$



## Simulation

- no "true  $X$ " existent! -

$$X^* \sim \mathcal{P}(\mathcal{F}(G)) \quad \text{in general}$$

$$= \mu_0 + \beta G_t + N_t \quad \text{in case of regression}$$

with  $N_t \sim \mathcal{N}(0, \sigma_n^2)$  independent of  $G_t$

THEN, in general:

BAD:  $E(\|X^* - X\|^2) > E(\hat{X} - X\|^2)$

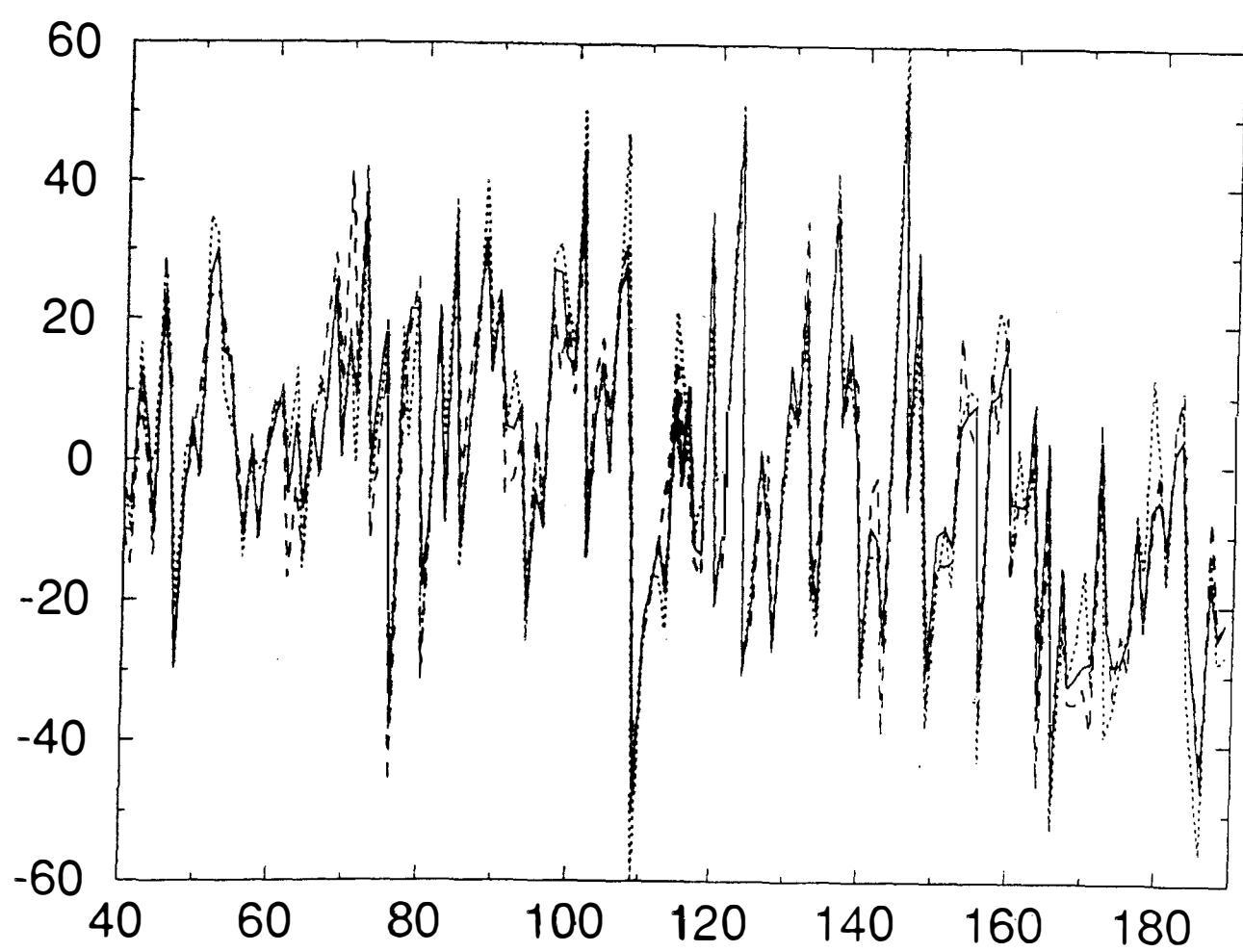
Good:  $\text{Var}(X^*) = \text{Var}(X)$

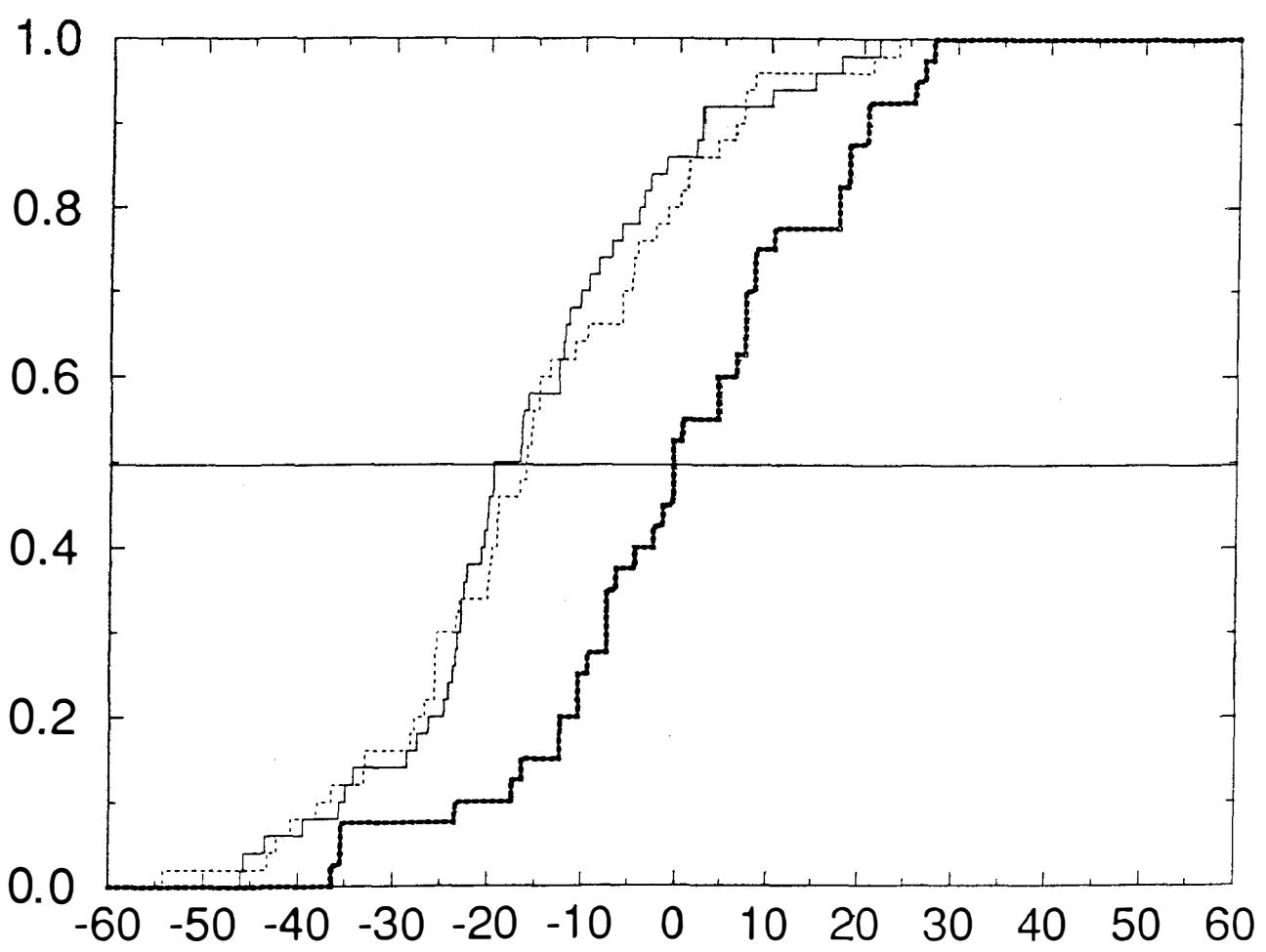
EXAMPLE: Snow drop flowering date

$$N_t \sim \mathcal{N}(0, (7\text{days})^2)$$

application to climate model date

- transient climate change experiments
- time slice ( $2 \times CO_2$ ) experiments.





## WEATHER GENERATORS

= statistical processes which

- generate realistically looking sequences of events
- reproduce certain statistics and relationships.

### Examples

- Snow drop case
- Markov process modelling of precip (Katz & Parlange, 96)
- CLUSTERING/CLASSIFICATION (Hughes et al., 93)

The G-phase space  $S$  is split into

$$S = \sum S_K.$$

so that the local feature  $X$  is "homogeneous" in  $S_K$ .

For any given  $G_t$ , first the subset  $S_K$  with  $L_t \in S_K$  is determined and then an observed state  $G_0$  with  $G_0 \in S_K$  is drawn (at random).

Then  $X^* = X_0$ .

- Analogue specification (Zorita et al., 95)

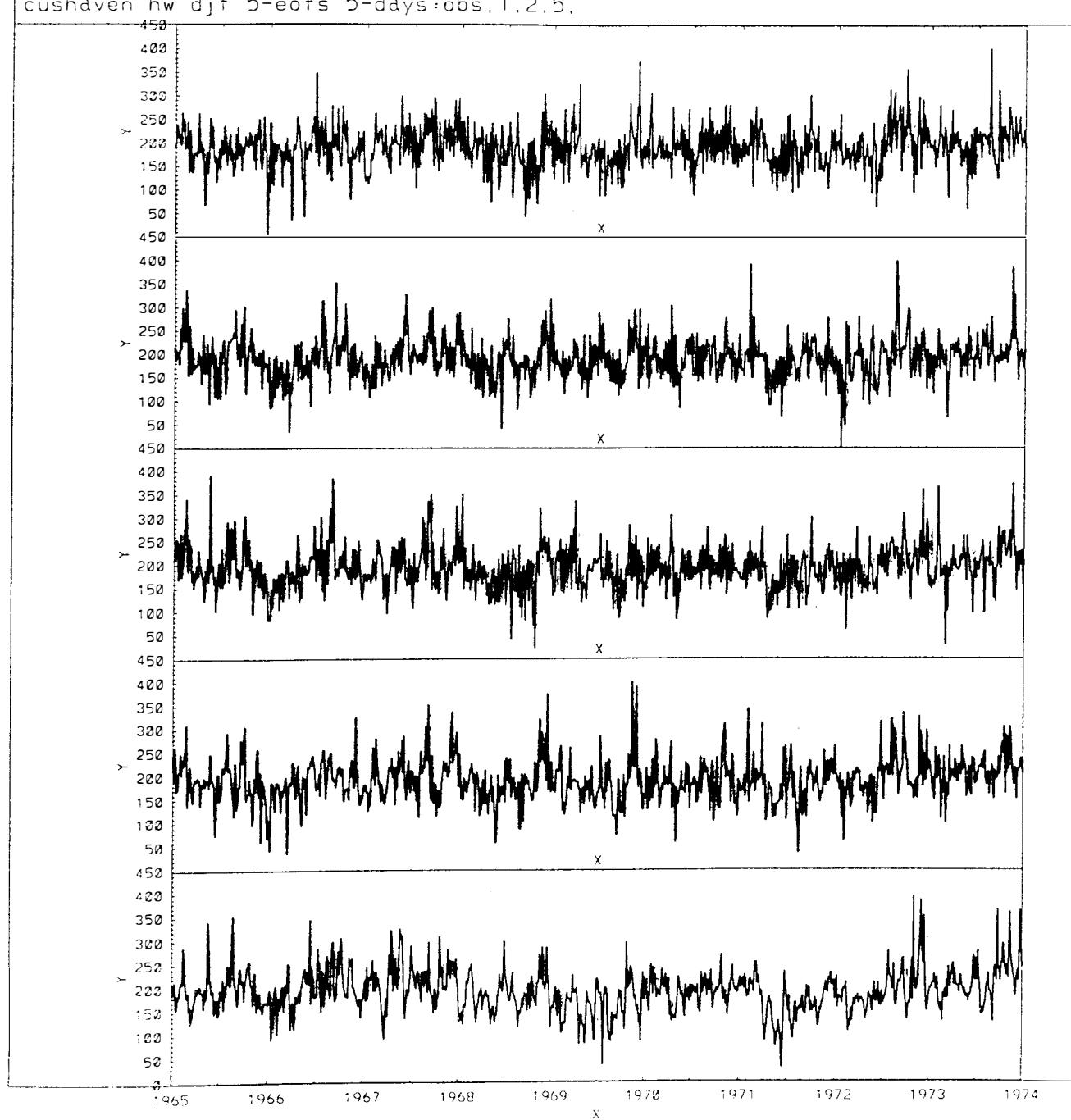
For any  $G_t$ , the nearest neighbor in the observational record is searched for.

Then

$$X^* = X_0$$

Example: Height of high tide in Cuxhaven  
(N. Germany).

cushaven hw djf 5-eofs 5-days:obs,1,2,5.



# The Local modifying the Global - PARAMETERIZATIONS of sub-grid scale processes

Let's consider a climate model

$$\frac{\partial \phi}{\partial t} = R(\phi)$$

and a spatial truncation  $\phi = \bar{\phi} + \phi'$

Then

$$(*) \quad \frac{\partial \bar{\phi}}{\partial t} = R_{\Delta x}(\phi) = R(\bar{\phi}) + R'(\phi')$$

Equation (\*) is not closed and cannot be integrated.

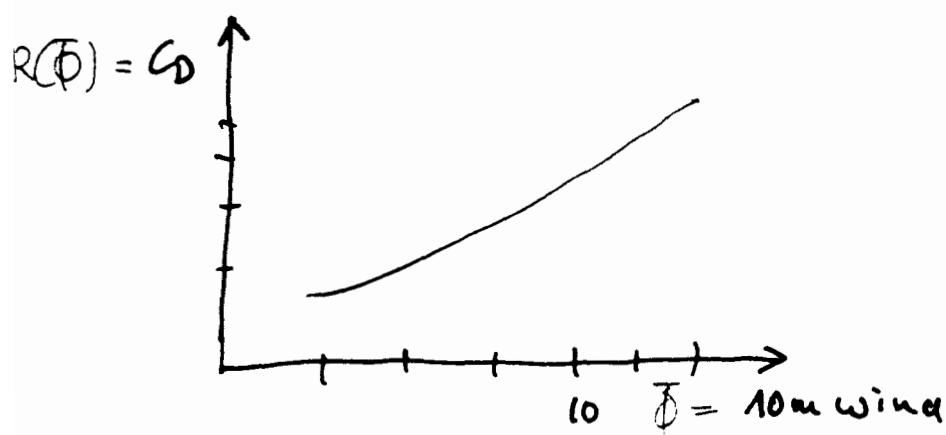
Options

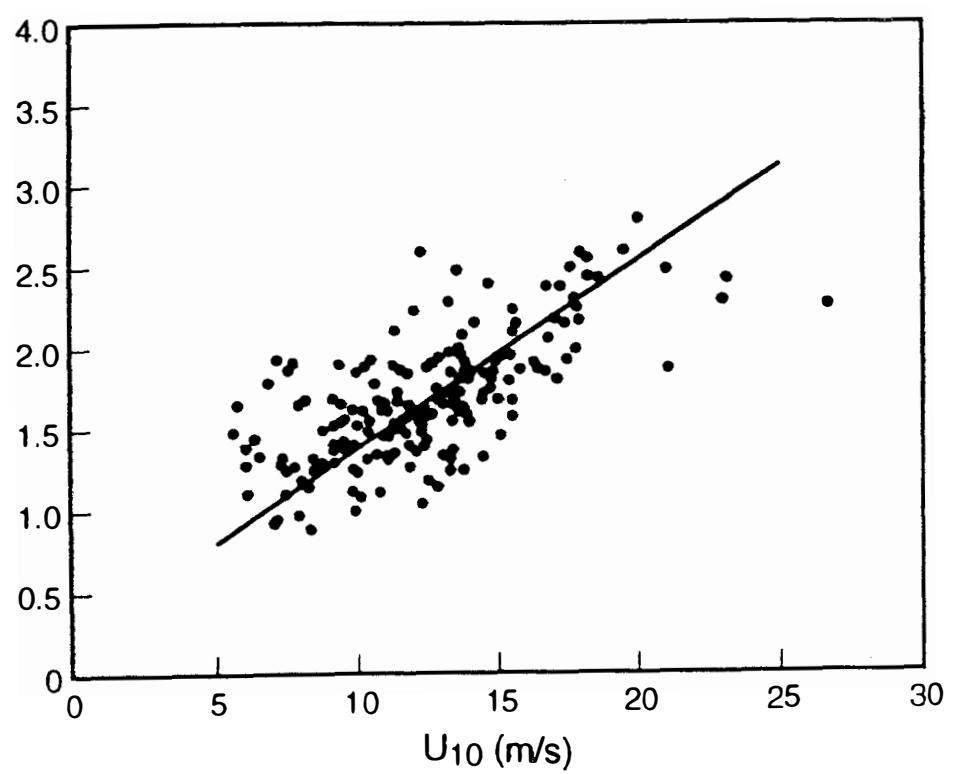
(a)  $R'(\phi') = 0$  .... in many cases done

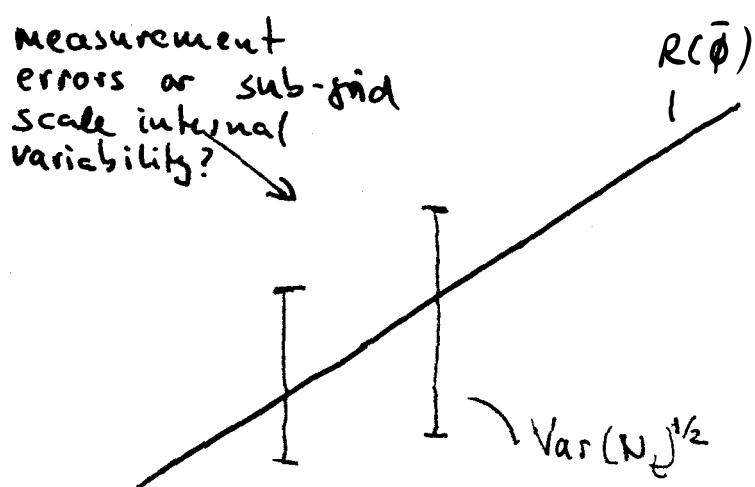
(b)  $R'(\phi') = Q(\bar{\phi})$  called "parameterization"

This ansatz implies that all variability on sub-grid scales unrelated to the resolved scale dynamics,  $\bar{\phi}$ , is irrelevant.

Example







c) randomised parameterization

$$R'(\phi') \sim \mathcal{P}(\vec{\alpha})$$
$$\vec{\alpha} = \mathcal{F}(\phi)$$

$$\Rightarrow R'(\phi') \sim \mathcal{P}[\mathcal{F}(\bar{\phi})]$$

(\*\*\*)

and often:  $\vec{\alpha} = Q(\bar{\phi}) + N_{\xi}$   
with  $N$  independent of  $\bar{\phi}$   
and adequately distributed.

(\*\*\*) is a randomized parameterization as it returns different values for the same resolved scale state.

It is formally identical with the randomized downscaling, with  $\bar{\phi} = G$ .

# DOES IT MATTER?

A demonstration with an Energy Balance Model

$$\frac{\partial \bar{T}}{\partial t} = c_w [S + L]$$

↑              ↑  
short    long wave radiation;  $\bar{T}$  = global mean temp

option b) conventional

$$S = \bar{S} = (1 - \alpha) S_0, \quad \alpha \text{ a function of } \bar{T}.$$

$$L = \bar{L} = b \bar{T}^4$$

$$\Rightarrow \frac{d\bar{T}}{dt} = \zeta(1 - \alpha(\bar{T})) S_0 + b c_w \bar{T}^4$$

two stable solutions, no "real" variability  
after initial adjustment phase.

option c) randomized

$$S = \bar{S} + S_0 N_t$$

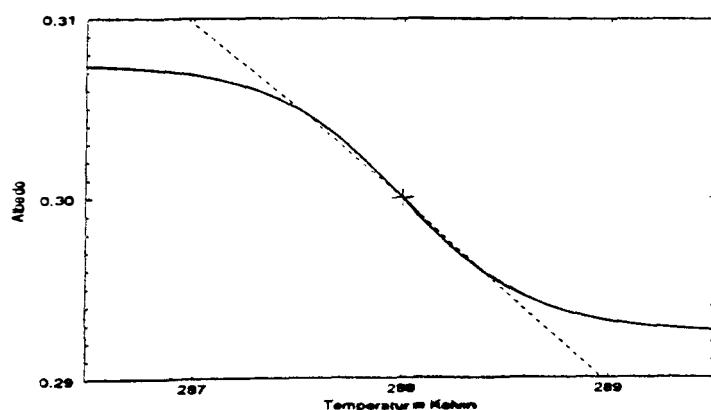
$$N_t \sim \mathcal{N}(0, 5\%)$$

two dynamical equilibria, full spectrum  
of variability, stationary dynamics  
Considerably richer than option b.

... this of course just Hasselmann's stochastic climate  
model ...

Yes, it matters, at least in certain situations.  
Nonlinearity is not a needed ingredient.

Abbildung 5.6: Temperaturabhängige Albedo  $\alpha(T_i)$  (durchgezogene Linie). Mögliche Gleichgewichtswerte des EBM's (gestrichelt) und ursprünglicher Gleichgewichtszustand (+).

$$\alpha(T_i) = 0,3 \cdot (1 - 0,025 \cdot \tanh(1,548 \cdot (T_i - 288K)))$$


Temperatur in Kelvin

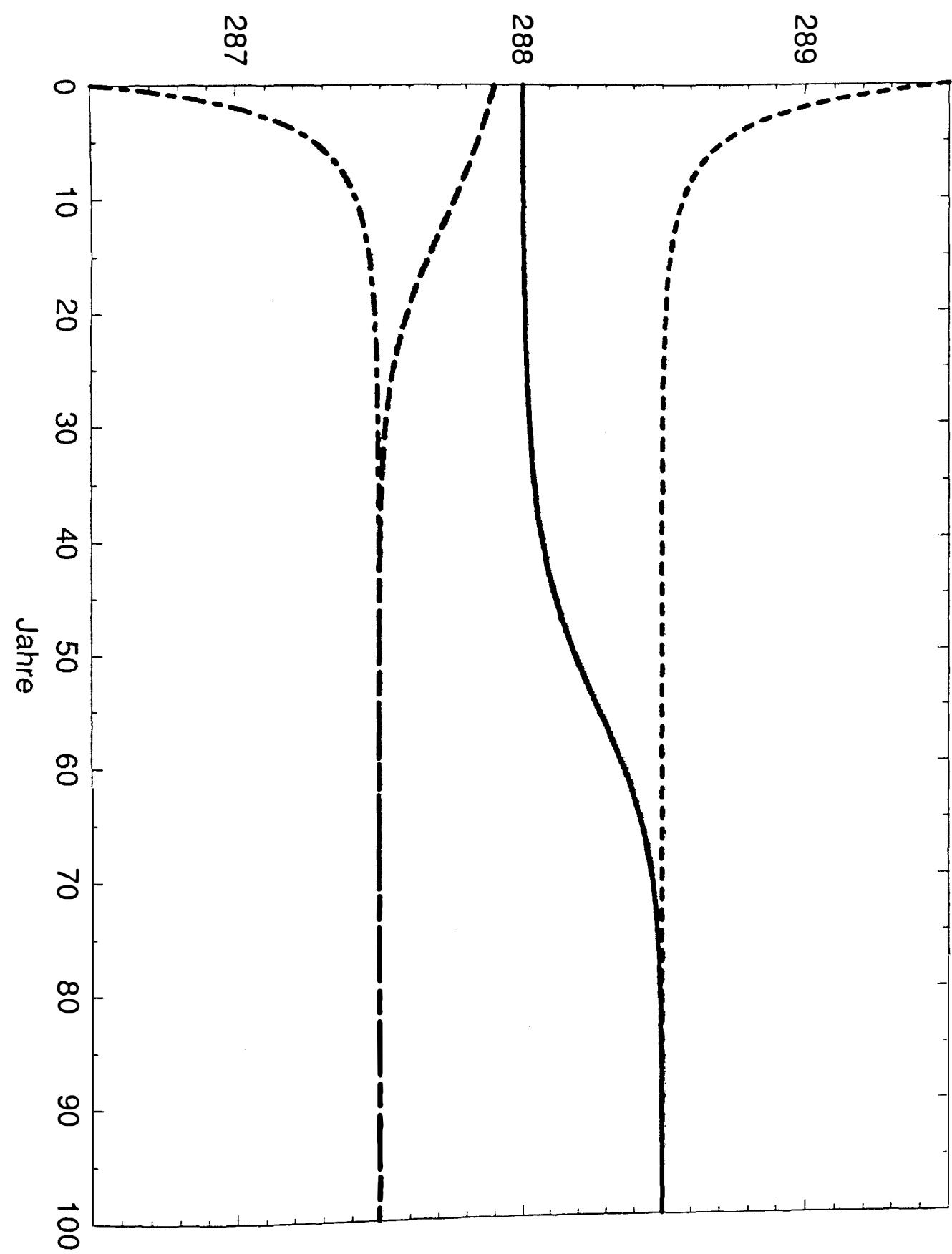
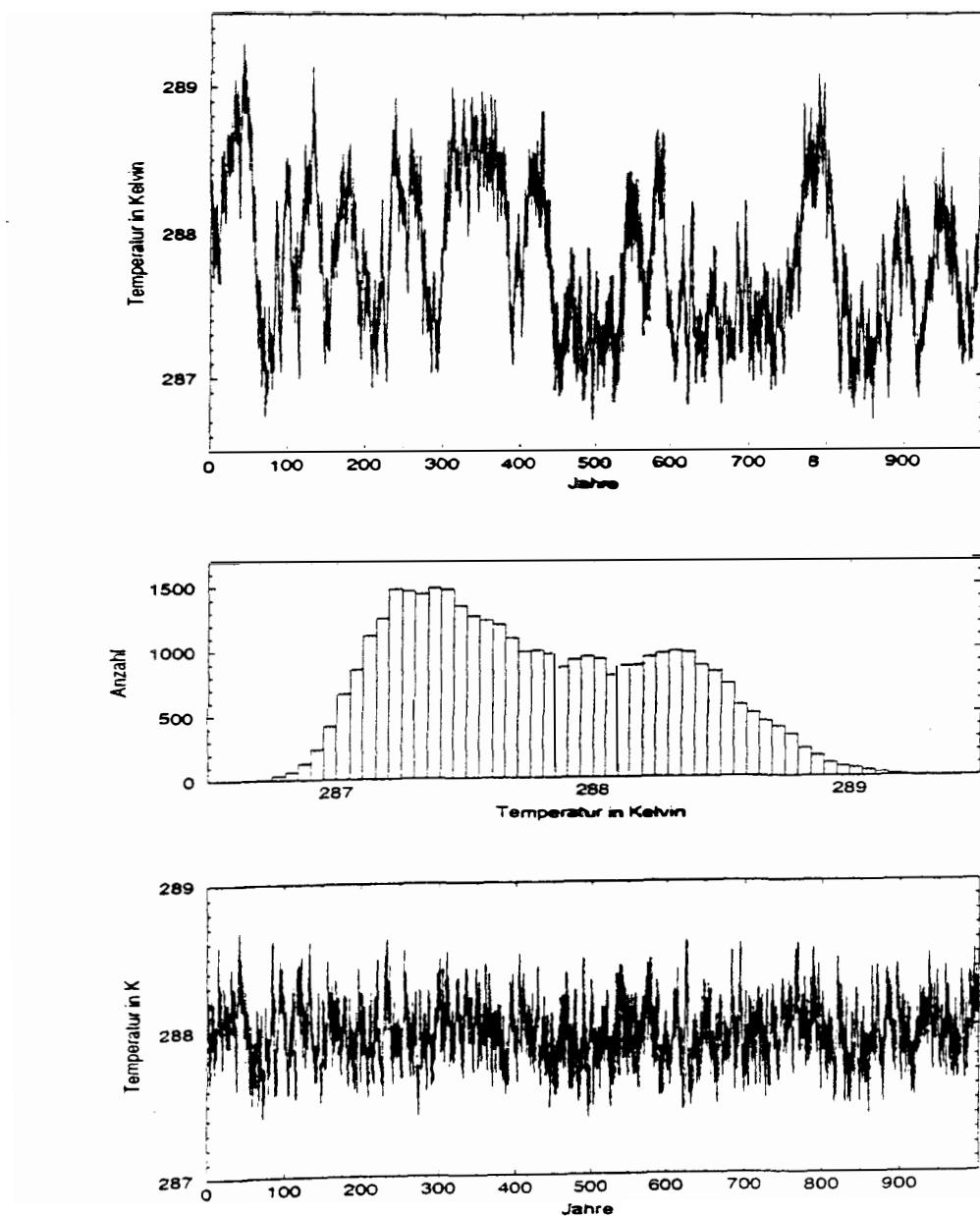


Abbildung 5.7: Oben: Langfristverhalten des EBMs bei temperaturabhängiger Albedo (aus Abbildung 5.6).

Mitte: Die Häufigkeitsverteilung des Ergebnisses.

Unten: Als Vergleich das Langzeitverhalten des EBMs bei konstanter Albedo.



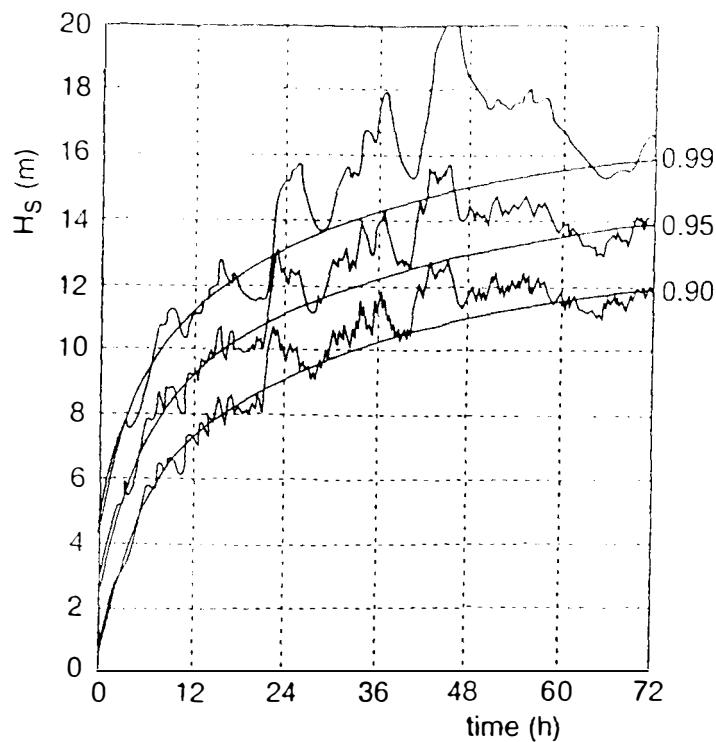


Fig. 4.33. Growth curves resulting from wind with a 0.25 turbulence level. Different correlation values  $\alpha$  between sequential wind values have been used, the smooth reference lines corresponding to  $\alpha = 0$ . The large oscillations result from those of the input turbulent wind (see figure 4.32). There is an up shift (of 2 m) between the diagrams for better visualization (after Cavaleri and Burgers, 1992).

## Conclusion

The Local in climate has two roles:

- it is the scale where climate acts upon climate-sensitive systems such as ecosystems, urban climate and coastal oceanography. It is the scale of applied climate research.
- it influences the development at the larger scales in a statistical manner.

For describing both roles in climate modelling, it should be acknowledged that only part of the local variability can be traced back to the larger resolved scales. Thus, a proper description should cast both the parameterization of local effects and the downscaling in a conditional statistical model.

Only when specific states are to be estimated, as for instance in forecasting, interpretation of proxy data etc, the specification by the non-random conditional expectation is meaningful.

It remains to be seen, if this approach is changing anything in a practical sense.