CONSTRUCTION AND VERIFICATION OF STOCHASTIC CLIMATE MODELS

K. Hasselmann

Max-Planck-Institut fur Meteorologie
Hamburg, Germany

CAN THE BASIC FEATURES OF CLIMATE VARIABILITY BE SIMULATED WITH SIMPLE MODELS ?

Theories and speculations on the nature of climate and climate variability cover a wide range, but there exists general agreement that the climatic system represents a complex structure of coupled subsystems - generally divided broadly into the atmosphere, oceans, cryosphere and biosphere - of which each component represents a detailed discipline in its own right, and which interact across a wide spectrum of space and time scales in a complicated manner which ultimately determines the dynamics of the complete system [20].

If detailed models of the dynamics of each of the individual subsystems of the climatic system existed, and the necessary computing power were available, one could in principal try to simulate the dynamics of the complete climatic system by "brute force" by coupling the individual subsystems together in a comprehensive numerical model [21]. However, it is questionable whether this would be a useful exercise. One would have no guaranty that the unavoidable idealizations and simplifications introduced into the separate subsystems, even if verified for each of the subsystems individually, were still appropriate when the subsystems were subjected to different modes of response, involving different time and space scales, through their interactions with the rest of the system. It can be argued that the main challenge of climate modeling should be seen rather in the opposite approach, in the attempt to identify the structure of the governing interactions within the climate system and to express these in terms of rather simple climate

A. Berger (ed.), Climatic Variations and Variability: Facts and Theories, 481–497.

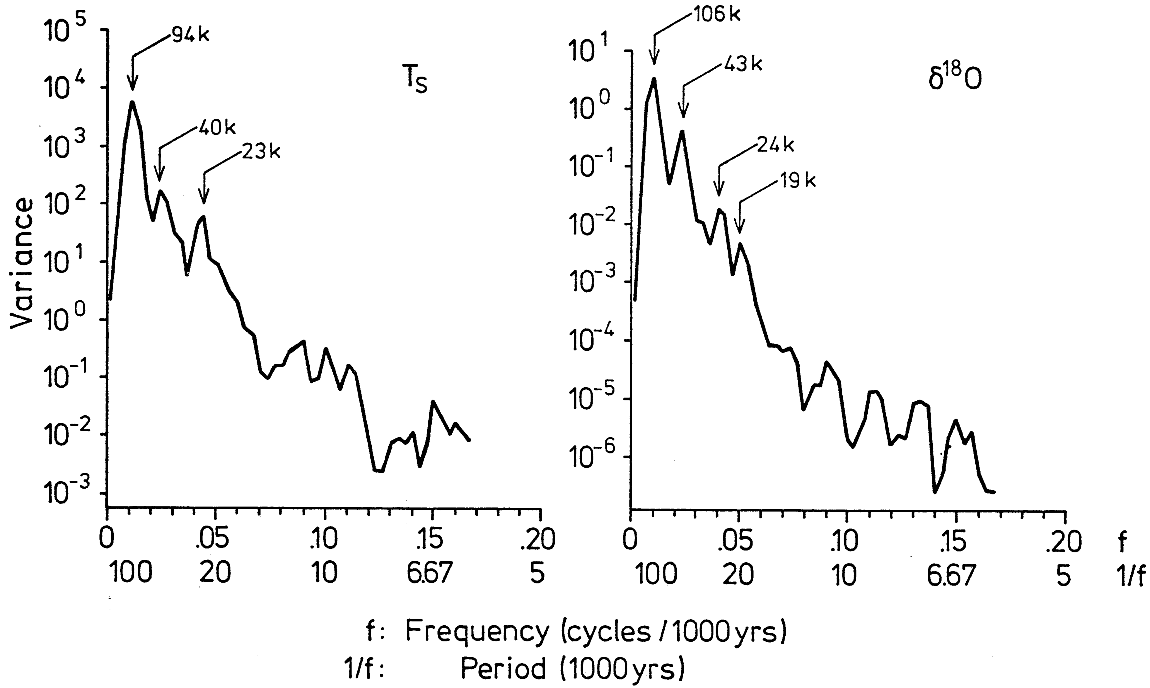

Figure 1 Variance spectra of paleoclimatic time series (inferred sea surface temperature T_s and $\delta^{18}O$ from Hays et al., 1976,[1].

models. Once such a first order picture has been established, the models can be iterated and gradually developed, through more detailed comparisons with data, into more sophisticated models.

It is of course not obvious that such an approach will succeed. Indeed, simple arguments will be presented in the following that the principal features of climate variability cannot be simulated by elementary climate models if the models are restricted to low order deterministic systems containing only a few degrees of freedom. However, it will be shown in the next section that more encouraging results are obtained if the class of models is extended to include stochastic forcing terms.

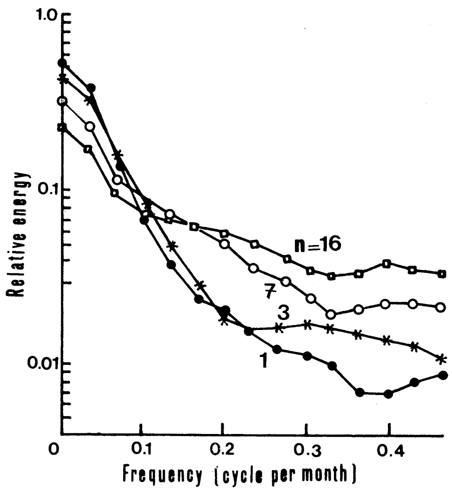

Figure 2 Variance spectra of selected amplitudes of empirical orthogonal functions of sea surface temperatures anomalies: n is the function mode number (after Barnett and Davis, 1975, [2]).

Some general guidelines on the type of climate model required to simulate climate variability may be derived from the statistical properties of observed climate fluctuations. Figures 1 - 3 show examples of the variance spectra of climatic fluctuations observed over two very different time scale ranges. The distributions may be regarded as representative of climatic variance spectra on all time scales. The characteristic feature of all spectra is a continuous red distribution, increasing towards low frequencies according to some power law

$$E(\omega) = \text{const.}\ \omega^{-2} \qquad (1)$$

where q typically lies in the range

$$1 < q < 2 \qquad (2)$$

The power law normally holds throughout the definition range of the spectrum, from the lowest resolution frequency (limited by the length of the record) up to the Nyquist cut-off frequency (limited by the sampling interval or the time resolution of the record).

An exception are the paleoclimatic spectra of Figure 1, which are terminated on the low-frequency side by a peak at a period near 10^5 years. This peak has been the subject of various speculations but has still not been conclusively explained. Also apparent in these spectra are weak peaks, barely statistically significant above the continuum, near periods of 2×10^4 years and 4×10^4 years, which have been attributed to variations of the earth's orbit in accordance with Milankovitch's theory [19]. However, apart from these features, the dominant characteristic of essentially all climatic time series, covering time scales from tens of thousands of years down to a few weeks, is the continuous red distribution of the variance spectrum without the occurrence of prominent peaks.

It is an interesting observation that phenomena with similar statistical characteristics (described by Press [3], as "flicker noise") are not limited to climatic fluctuations but are found widely in nature. Mandelbrot [4] mentions that Richardson discovered the same statistical properties for coastlines, and cites numerous other examples of one and higher dimensional configurations in nature which fall into the same statistical category.

The common feature of these processes is that they are basically non-analytical, i.e. the variances of the processes, or their derivations, if calculated formally by integrating over the spectrum from zero to infinite frequency, become infinite. Thus, the power law distribution (1), (2), if extended out to zero and infinite frequency, represents a process y(t) for which both the variance and the variance of the time derivation are infinite,

$$\langle y^2 \rangle = \lim_{\omega_L \to 0} \int_{\omega_L}^{\infty} F(\omega)\,d\omega = \infty \qquad (3)$$

$$\langle (\frac{dy}{dt})^2 \rangle = \lim_{\omega_u \to \infty} \int_0^{\omega_u} \omega^2 F(\omega)\,d\omega = \infty \qquad (4)$$

This elementary property immediately creates difficulties when one attempts to simulate climatic variability by simple deterministic models. Any set of coupled equations

$$\frac{dy_i}{dt} = F_i(y), \qquad i = 1, \dots n \tag{5}$$

describing the evolution of a system $y = (y_1, \dots y_n)$ with a finite number of degrees of freedom n will necessarily yield finite variances for the individual components of the system and their derivatives (provided the system contains no solutions which grow without bounds - a necessary condition anyway for the meaningful simulation of climate fluctuations). Although relatively low order nonlinear systems can be constructed which simulate fluctuations of a rather random appearance [5], the variance spectra of these fluctuations are necessarily concentrated in a finite band of frequencies and are therefore unable to reproduce the characteristic broad bandedess of climatic spectra which is responsible for the simultaneous divergence of the integrals (3) and (4) when the lower and upper frequency limits ω_L and ω_u approach zero and infinity, respectively.

It is possible to reproduce broad band variance spectra of the desired characteristics with numerical models in the limit as the number of degrees of freedom becomes very large. For example, general circulation models of the atmosphere containing 10^4-10^5 degrees of freedom generally reproduce reasonably realistic turbulence spectra exhibiting power law distributions, similar to the form (1), (2). These spectra generally extend from periods of several days to a short period cut-off of a few hours. The low frequency end of the spectrum can be extended further by coupling the atmospheric general circulation model to an ocean circulation and cryosphere model. However, this approach would defeat the main goal of climate modeling as defined here, namely to identify the basic processes which are responsible for the observed statistical properties of climate variability and to construct the simplest conceivable model which is able to reproduce the observed features.

A more fruitful approach under these circumstances may be to seek as starting point a simple model which is characterized by a broad band variance spectrum. This leads naturally away from low-order deterministic models to stochastic models. We shall find that this approach yields a simpler and more pertinent description of the role of the atmosphere in generating climatic fluctuations than the application of high resolution atmospheric general circulation models.

ELEMENTARY STOCHASTIC MODELS OF CLIMATE VARIABILITY

The conceptually simplest example of a broad band stochastic process is stationary white noise $x_i(t)$. This is characterized by a δ-function covariance function

$$\langle x_i(t+\tau)x_j(t)\rangle = C_{ij}\delta(\tau) \tag{6}$$

($\langle\ldots\rangle$ denotes ensemble expectation values)
and a frequency independent covariance spectrum

$$F_{ij}(\omega) = C_{ij}/2\pi \tag{7}$$

where C_{ij} = const.

If an elementary low-order deterministic climate model of the form (3) is extended to include white-noise forcing,

$$\frac{dy_i}{dt} = F_i(y) + x_i \tag{8}$$

one obtains in general a process which is characterized by a broad band, red-noise variance spectrum of a form qualitatively similar to (1), (2). The non-analytic properties of the climate variability implicit in a spectrum of this form are introduced directly through the non-analytical forcing function x_i. The transformation from a white noise input x_i to a red noise response y_i is a consequence of the time integration involved in solving (8). The form of the y-covariance spectrum depends in detail on the structure of the internal feedback function $F(\underset{\sim}{y})$. A number of rather simple feedback models have been investigated which were able to simulate observed climate fluctuations over various time scales reasonably well [6,7,8,9,10]. Some examples are given in Figures 3, 4 and 5.

Despite the reasonable agreement with observations, two basic questions arise in this modeling approach :

(i) What is the origin of the white noise forcing ?
(ii) The simulation of an observed variance spectrum by a model is clearly not a conclusive demonstration that the model is correct. How can the model then be verified reliably against data ?

The concept of white noise forcing is clearly an idealization. It is generally used in physics as an asymptotic approximation when the response of a system has a much larger characteristic time scale than the characteristic time scales of the forcing. In this case the forcing functions may be approximat-

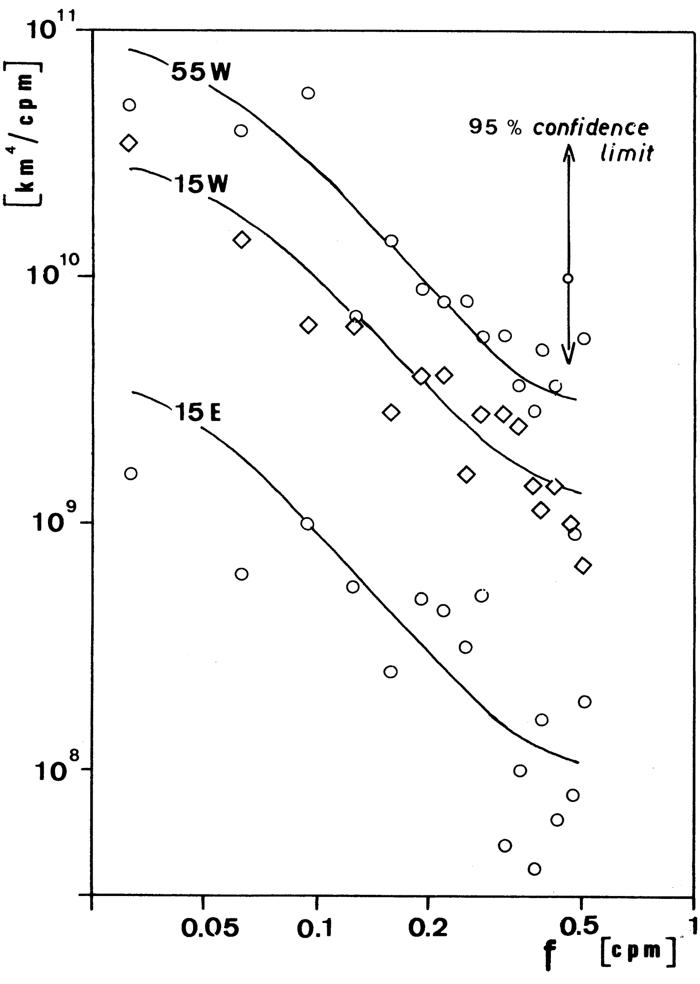

Figure 3 Variance spectra of Arctic sea ice anomalies at three typical longitudes (curves represent first-order Markov models, cf. second section).

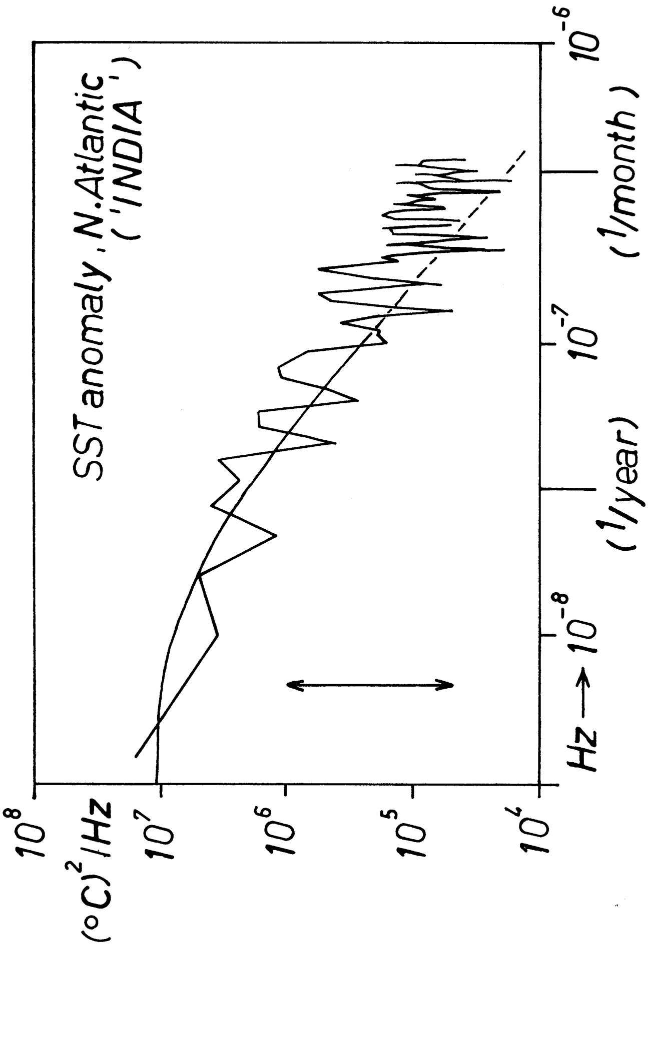

Figure 4 Comparison of observed sea surface temperature anomaly spectrum with prediction from simple mixed layer (copper-plate) model with stochastic atmospheric forcing and negative linear feedback. (from Frankignoul and Hasselmann, 1977, [7])

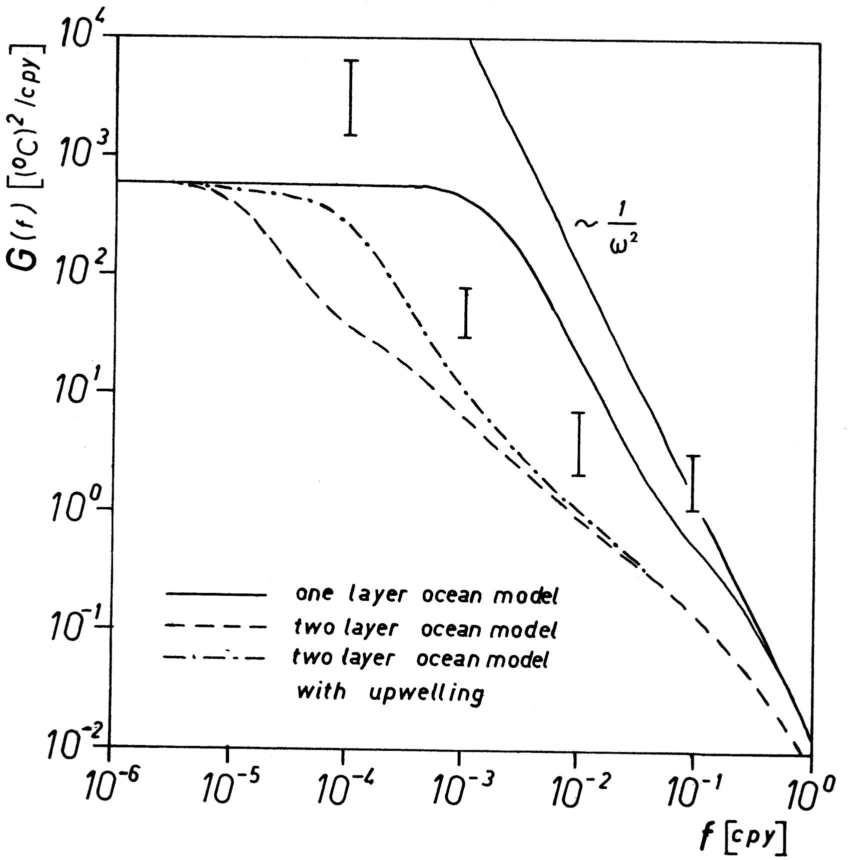

Figure 5 Influence of a second ocean layer on the variance spectrum of the surface temperature (at 55° latitude) for a stochastically forced Budyko-Sellers model. Vertical bars denote observations. (from Lemke, 1977, [8])

ed as a sequence of uncorrelated δ-functions. The response of an electronic system to the impulses of individual impacting electrons, or the Brownian motion of a large molecule bombarded by much smaller molecules, represent classical examples of this description.

In the case of climatic fluctuations, such a time scale separation exists between the turbulent motions of the atmosphere, with correlation times typically of the order of a few days, and the remaining components of the climatic system, the oceans, cryosphere and biosphere, which are characterized by response time scales ranging from months to several thousand years. In terms of the slow climatic variations in this time scale range, the statistical forcing by the much shorter period atmospheric weather disturbances (in the form of varying fluxes of heat, moisture and momentum at the earth's surface) may therefore be represented simply as (temporarily) uncorrelated white noise. The theory (and numerical simulation) of statistical climate fluctuations can be greatly simplified through this approximation.

The second problem, the verification that random forcing by atmospheric weather disturbances is indeed the principal cause of observed climate fluctuations (at least in some time scale ranges) is more difficult. We are faced here with the general problem of the construction and verification of statistical models from data, which we turn to in the following section.

VALIDITY AND STATISTICAL SIGNIFICANCE OF MODELS

The standard method of verifying a linear model is to compare the theoretical and observed auto-spectra and cross spectra (or covariance functions) of the input and output time series. The method provides a clear assessment of the skill of the model (i.e., the fraction of the output variance which can be expressed as a linear response to the input) and a comparison of the theoretical and observed transfer function relating the input and response. The method can be readily extended also to test nonlinear systems [11].

Although such techniques have been applied successfully to verify stochastic climate models in the short time scale range [7,12], they are often inapplicable in climate variability studies because the input data is not accessible. For example, in testing stochastic models of global sea surface temperature variations during the last few decades, or Northern Hemisphere temperature variations over still longer periods, it is virtually impossible to reconstruct reliable time series of the global short-term weather variability postulated as the white-noise forcing for these periods.

In these cases one is able to test the validity of the model only by comparing the theoretical and observed statistics of the "slow" components y_i of the climatic system. The "verification" therefore reduces to the demonstration that, for a given postulated model (8), the cross spectra of the output time series $y_i(t)$ can be adequately reproduced assuming a white noise input x_i characterized by a "reasonable" cross spectral coefficient matrix C_{ij} (cf. eq. (7)). Such a test is nevertheless nontrivial, since a complete set of functions $F_{ij}(\omega)$ depending on frequency must be simulated using only a set of constants C_{ij} and a few model coefficients for tuning.

Although the lack of input data makes it easier to fit a model to the remaining output data, it also makes it more difficult to satisfy the opposing requirement of statistical significance of the model. Clearly, a model containing a number of adjustable parameters can always be tuned to reproduce a given finite set of observed data, provided the number of free parameters is chosen sufficiently large. However, the significance of the model will then be small in the sense that some of the parameters or parameter combinations (and the physical processes which they represent) may be redundant, simpler models yielding equally acceptable simulations of the data within the errors of the data. An equivalent measure of the statistical significance of a model can be expressed in terms of the statistical errors of the model parameters. These arise in the present application through the (unavoidable) estimation errors of the cross spectra of the finite time series $y_i(t)$ to which the model is fitted (as well as through possible additional measurement errors). In general, the model parameter errors increase as the number of parameters increase. Thus the model in fact becomes less determined the more sophisticated the model and the better the apparent fit of the model to the data.

Various techniques have been developed to determine the optimal number of parameters of a model to yield the most suitable compromise between the conflicting requirements of a good fit to the data (as expressed in terms of "skill", "consistency" or model "validity") and a sufficiently high level of statistical significance of the resulting best fit model [13,14,15,16].

The application of such statistical model fitting techniques can be helpful not only in establishing the nature of certain kinds of climate variability, but also in determining the physical parameters describing interactions within the

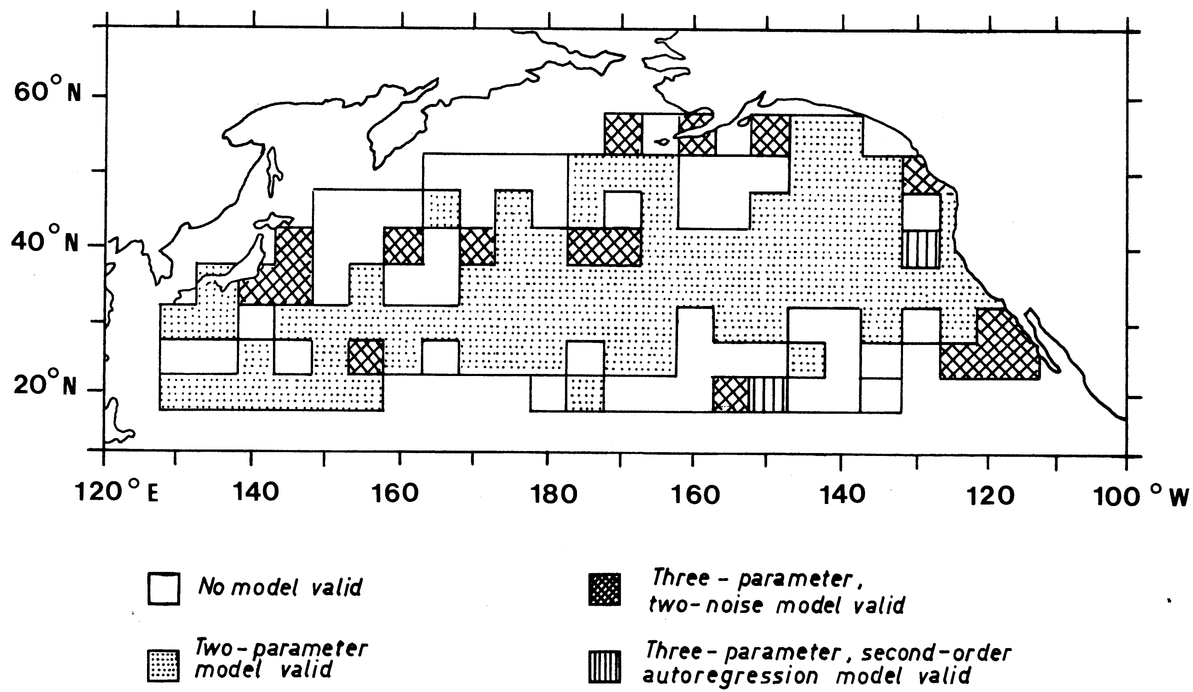

Figure 6 Region of validity of a local stochastic forcing mixed layer model of SST variability, based on auto-spectra alone. (from Reynolds, 1978, [9])

climatic system. As example, Figure 6 shows the region of validity in the North Pacific of a simple, locally decoupled, fixed depth (copper plate) model of the mixed layer of the ocean [9]. In this model, heat exchange between the mixed layer and the atmosphere occurs only through local white noise forcing and a negative feedback Newtonian cooling term proportional to the sea surface temperature anomaly. The model is seen to be valid in most of the interior of the North Pacific, but fails along the edges of the ocean and in the equatorial regions, presumably due to the presence of strong advecting currents. The auto-spectra of sea surface temperature fluctuations at individual locations were used as data base for the model test (as appropriate for a decoupled model). An analysis of the cross spectra of temperature fluctuations at different locations, however, shows that these cannot be adequately modeled by a decoupled model in most of the ocean and that the advection of sea surface temperature anomalies by currents should presumably be taken into consideration also in the interior ocean.

Figure 7 shows the region of validity of an extended model based on these results including advection by currents and horizontal diffusion [17]. In this case both auto spectra and cross spectra were used to validate the model. The empirically determined mixed layer currents obtained by a best fit to the data are shown in Figure 8. Reasonable agreement is found with standard current charts, determined largely from ship drift data (cf. Figure 9). Additional information which can be extracted from the model is the level and spatial correlation scale of the atmospheric white noise forcing and the effective diffusion coefficient characterizing the non-advective dispersion of sea surface temperature anomalies. Similar techniques have been applied to the study of sea ice anomalies [10].

In summary, it may be concluded that the construction and verification of stochastic climate models by statistical model fitting techniques is a useful tool not only for providing insight into possible mechanisms of climate fluctuations, but also for determining internal interaction coefficients of the climate system which are often difficult to measure directly.

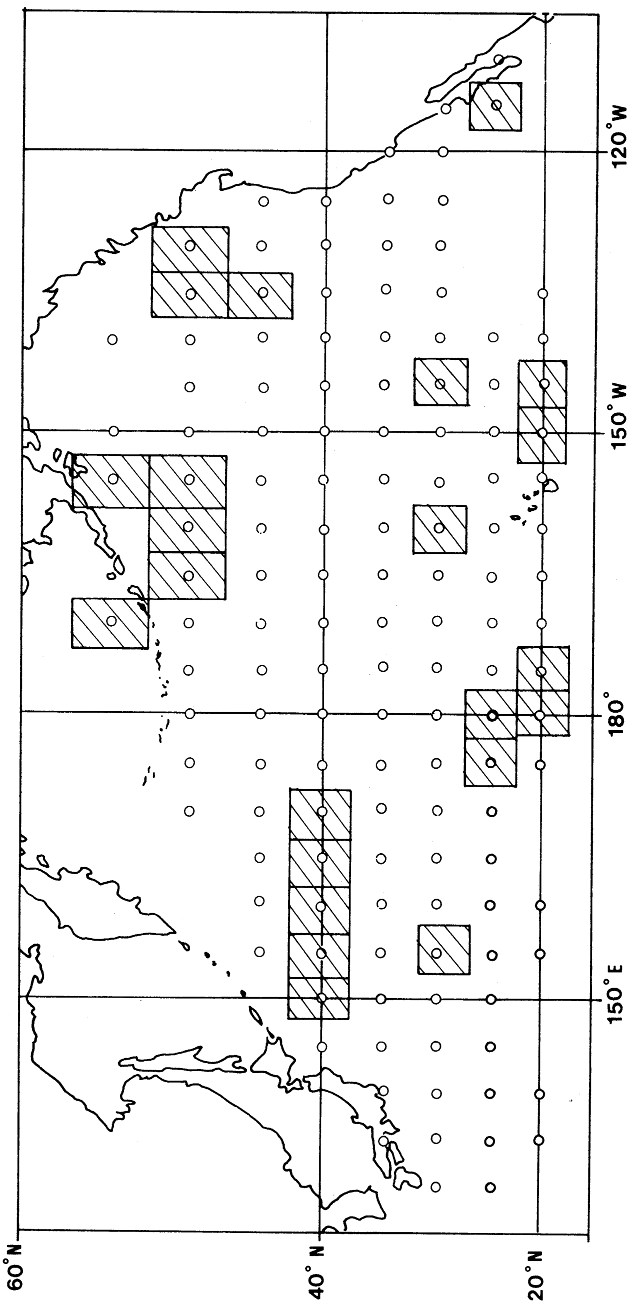

Figure 7 Region of validity of a stochastic mixed layer model including horizontal advection and diffusion, based on auto- and cross spectra of SST variability (shaded boxes represent invalid regions). (from Herterich and Hasselmann, 1981, [17])

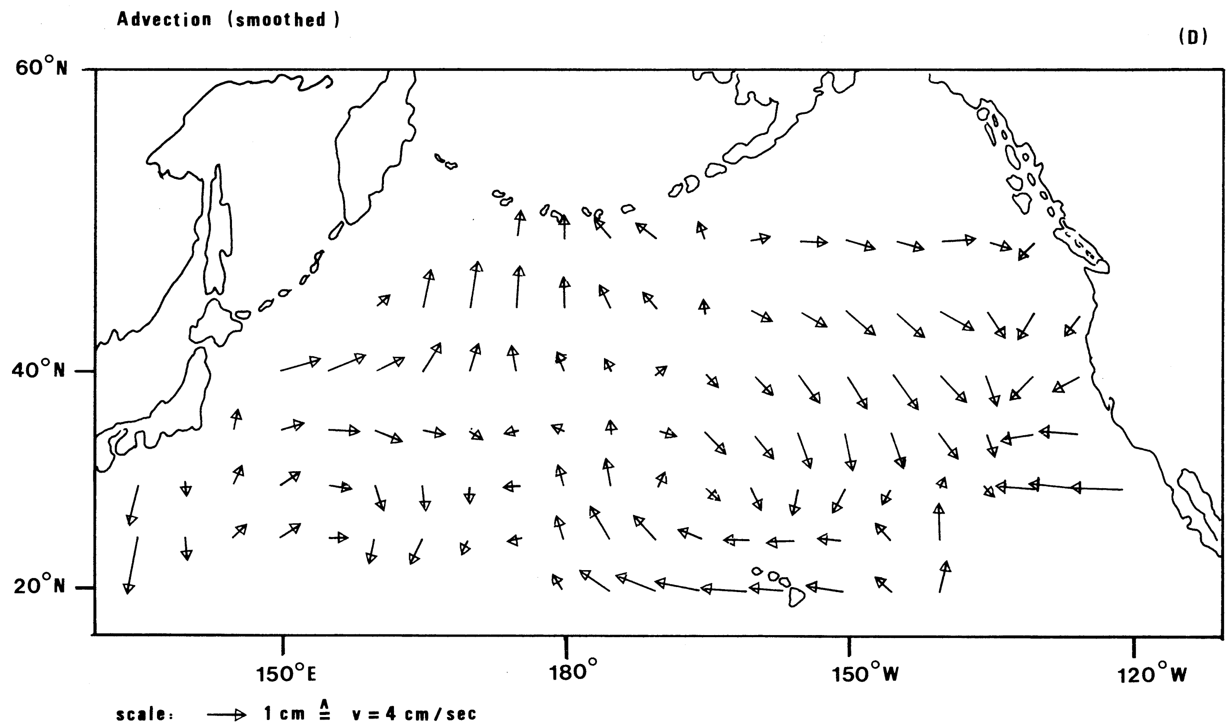

Figure 8 SST advection currents inferred by inverse model analysis of auto-spectra and cross spectra of SST variations. (from Herterich and Hasselmann, 1981, [17])

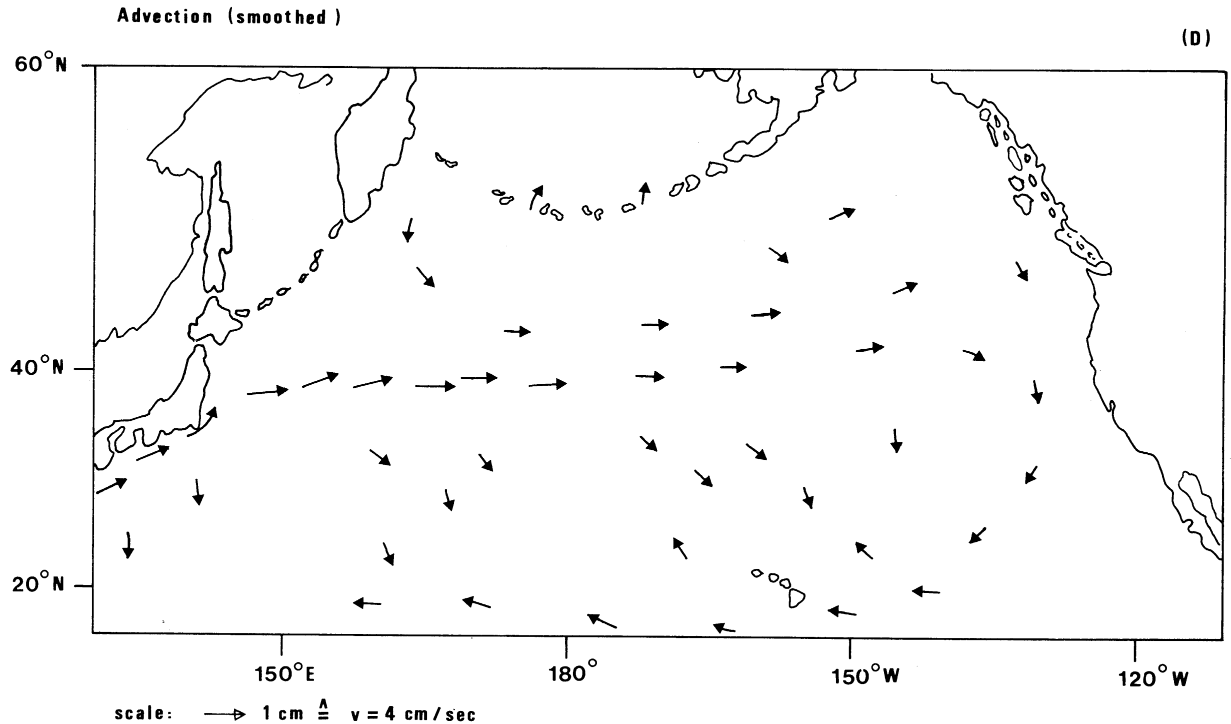

Figure 9 Surface currents derived from ship drift data (composite summer and winter current charts from Gorshkov, 1974, [12]).

REFERENCES

1. Hays, J.D., Imbrie, J. and Shackleton, N.J. : 1976, Science 194, pp. 1121-1132.
2. Barnett, T.P. and Davis, R.E. : 1975, in : "Proceedings of the Symposium on Long-Term Climatic Fluctuations", WMO Techn. Note no.421, pp. 439-450, WMO, Geneva.
3. Press, W.H. : 1978, Comments Astrophys. 7, pp. 103-119.
4. Mandelbrot, B.B. : 1977, "Fractals, Form, Chance and Dimension", pp. 365, W.H. Freeman and Co.
5. Lorenz, E.N. : 1965, Tellus 17, pp. 321-333.
6. Hasselmann, K. : 1976, Tellus 28, pp. 473-485.
7. Frankignoul, C. and Hasselmann, K. : 1977, Tellus 29, pp. 289-305.
8. Lemke, P. : 1977, Tellus 29, pp. 385-392.
9. Reynolds, R.P. : 1978, Tellus 30, pp. 97-103.
10. Lemke, P., Trinkl, E.W. and Hasselmann, K. : 1981, J. Phys. Ocean. (in press).
11. Hasselmann, K., Munk, W.H. and McDonald, G.J.F. : 1963, in : "Bispectra of Ocean Waves, Time Series Analysis, M. Rosenblatt (Ed.), pp. 125-139.
12. Frankignoul, C. : 1979, Dyn. Oceans Atm. 3, pp. 465-479.
13. Box, G.E.P. and Jenkins, G.M. : 1976, "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco.
14. Olbers, D.J., Muller, P., and Willebrand, J. : 1976, Phys. Earth Planet. Inter. 12, pp. 248-252.
15. Hasselmann, K. : 1979, Dyn. Oceans Atm. 3, pp. 501-521.
16. Barnett, T.P. and Hasselmann, K. : 1979, Rev. Geophys. and Space Phys. 17, pp. 949-968.
17. Herterich, K. and Hasselmann, K. : 1981, Generation, advection and diffusion of sea surface temperature anomalies in the North Pacific (in preparation).
18. Gorschkov, S.G. : 1974, "Atlas of the Oceans, vol. 1, Pacific Ocean", Ministry of Defense USSR Navy.
19. Berger, A. : 1981, Astronomical theory of paleoclimates. This volume.
20. Gates, L. : 1981, The climate system and its portrayal by climate models : a review of basic principles. I. Physical basis of climate. This volume.
21. Gates, L. : 1981, The climate system and its portrayal by climate models : a review of basic principles. II. Modeling of climatic changes. This volume.