ON THE CONSTRUCTION OF TIME INVARIANT AND SEASONALLY DEPENDENT STATISTICAL CLIMATE MODELS

K. Hasselmann

Max-Planck-Institut für Meteorologie,
Hamburg, FRG

The large time scale separation between the natural variability of the atmosphere and the slower parts of the climatic system (oceans, cryosphere) introduces a natural statistical component into the structure of climate change. This statistical element is an essential input in stochastic climate models (Hasselmann, 1976) and plays an important role also in the construction of climate prediction models (Barnett and Hasselmann, 1979).

In statistical modelling it is useful to distinguish between statistical prediction models and the application of stochastic models to analyse the dynamical structure of the climatic system (Hasselmann, 1979). In system analysis applications it is generally more effective to work in the frequency rather than the time domain. Because of the causality side condition, however, the construction of optimal prediction models is more easily carried out in the time domain.

By fitting model auto-variance spectra to observed spectra of sea surface temperature (SST) variations and sea-ice fluctuations it has been shown that both SST and sea-ice variability can be described in the time scale range from a month to one or two years as the integral response of a slow (first order differential) system to white noise atmospheric forcing (Frankignoul and Hasselmann, 1977, Reynolds, 1978). Essentially the same model, extended to include a long time constant, deep diffusive ocean, was also able to reproduce qualitatively the observed climatic variability spectrum in the longer period range from a few years up to several thousand years (Lemke, 1977).

More sophisticated models of SST and sea-ice variability, including spatial coupling, can be constructed if the cross spectra of the observed anomalies at different locations are taken into account. However, as the models become more complex, the problem of the statistical significance of the model becomes important. In order to determine the degree of detail that can be inferred from a given data set it is necessary to test a hierarchy of models, starting from a simple model containing only a few adjustable physical parameters, and successively introducing additional parameters until a model is found which is consistent with the observed spectra at a given statistical significance level. The addition of further parameters then reduces the error level but at the same time also reduces the statistical significance of the model. The optimal model is the simplest physical model which is statistically consistent with the data. Application of these techniques to SST and sea-ice variability data yielded useful estimates of the fields of effective advection velocities and diffusion coefficients governing the evolution of SST and sea-ice anomalies, as well as the spatial structure of the white noise atmospheric forcing (Lemke, Trinkl and Hasselmann, 1980, Herterich, 1980).

The same problem of balancing the statistical significance of a model against good simulation of the data arises also in the construction of statistical prediction models. If the class of models to be fitted to the data is defined too widely at the outset, i.e. if the model contains too many adjustable parameters, no statistically significant model can be found. Thus the optimal model must again be determined from a model hierarchy, in which more complicated models are developed successively from simpler models. As the number of model parameters is increased, the skill of the model increases, but the statistical significance generally decreases. The point where the statistical significance falls below some prescribed level then defines the optimal model of the hierarchy, i.e. the maximal skill model which is still statistically significant. Application of these methods to multi-time lagged, multi-variable linear predictions of anomalies in the equatorial Pacific yielded significant predictions up to lead times of 6 to 10 months (Barnett and Hasselmann, 1979).

Clearly, the optimal significant prediction will depend rather critically on the *a priori* choice of model hierarchy. However, this is an unavoidable problem in all statistical modelling applications in which the statistical significance of the model is considered, since an unbiased statistical test always requires the *a priori* specification of the hypothesis to be tested.

Linear prediction models are normally constructed assuming statistically stationary processes. For short range climatic predictions, this assumption is rather questionable, since the statistical properties of the anomaly fields, even after subtraction of the mean annual cycle itself, clearly exhibit a strong seasonal dependence. It is a common view that the generalization of time-invariant prediction models to seasonally varying models is in many cases not feasible, because the many additional degrees of freedom of the model needed to represent the full annual cycle unavoidably degrades the statistical significance of the model. This problem can again be resolved, however, by introducing the additional degrees of freedom sequentially. A natural model hierarchy can be constructed, for example, by expanding both the model and the statistics (second moments) of the anomaly fields in a harmonic series with respect to the fundamental annual cycle. The importance of the additional degrees of freedom introduced through the annual modulation of the model must then be judged *a priori* in competition with the other degrees of freedom of the model (for example, the number of time lags, or the number of components retained in a principle expansion of the predictor fields). Based on this assessment, a one-dimensional hierarchy of models can then be defined and tested in essentially the same way as for the time invariant case. The unavoidable competition between the relative relevance of the time-invariant and seasonally dependent parameters of the model is partially offset by the fact that additional statistical information is available to test seasonally varying models, since not only the zero'th harmonic but also the higher annual harmonics of the second moment statistics are used in fitting the model. Thus if a strong seasonal modulation of the statistics is observed, seasonally dependent prediction models can generally be constructed with the same statistical significance as time invariant models containing a comparable number of predictors.

References

Barnett, T.P. and Hasselmann, K., 1979: Techniques of linear prediction, with application to oceanic and atmospheric fields in the tropical Pacific. Reviews of Geophysics and Space Physics 17, 949-968.

Frankignoul, C. and Hasselmann, K., 1977: Stochastic climate models, Part II, Application to sea-surface temperature anomalies and thermocline variability. Tellus 29, 289-305.

Hasselmann, K., 1976: Stochastic climate models, Part I, Theory. Tellus 28, 473-485.

Hasselmann, K., 1979: Linear statistical models. Dyn. Atmos. Oceans 3, 501-521.

Herterich, K., 1980: Inverse modelling of sea-surface temperatures in the North Pacific Ocean. (in preparation).

Lemke, P., 1977: Stochastic climate models, Part III, Application to zonally averaged energy models. Tellus 29, 385-392.

Lemke, P., Trinkl, E.W. and Hasselmann, K., 1980: Stochastic dynamic analysis of polar sea ice variability. (submitted to J.P.O.)

Reynolds, R.W., 1978: Sea surface temperature anomalies in the North Pacific Ocean. Tellus 30, 97-103.