ECMWF Workshop on: 'New developments in predictability'
13-15 November 1991

SHORT RANGE SKILL PREDICTION

Klaus Fraedrich
Institut für Meteorologie
1000 Berlin 41, Federal Republic of Germany

1. INTRODUCTION

Methods for predicting the skill (or accuracy) of weather forecasts may be categorized as either stochastic-dynamical or ensemble predictions: Stochastic-dynamical forecasts incorporate uncertainties of the initial conditions (and of the model) leading to a deterministic problem in stochastic terms. Thus the model predictions provide forecasts of means and variances (*Epstein,* 1969). For routine use with numerical weather prediction (NWP) models, however, this technique still seems to be beyond computer technology. Ensemble forecasts lead to an average prediction and, in addition, the forecast skill can be estimated from the dispersion of the ensemble. The required multiple forecasts may be obtained from a set of (randomly) perturbed initial conditions (*Leith,* 1974). Or, alternatively, ensembles of lagged forecasts based on previous initial analyses may be used, which are available on routine basis (*Hoffmann and Kalnay* 1983, *Kalnay and Dalcher* 1987). Applied to medium and extended range NWP forecasting, the prediction of the forecast skill is related mainly with the limits of predictability and the breakdown of the forecast. In short term NWP forecasting, however, one is generally not concerned about the predictability limit but about the growth rate of errors and their more precise predictions. These purely numerical methods describe basically a NWP model's sensitivity on initial conditions which, if the model performs well, leads to a measure of the predictability of the atmosphere.

On the other hand, one may define the predictability (or the forecast error) as an atmospheric variable (like the temperature or pressure at a weather station) to be predicted. Then, at least in principle, there are two basic methods to estimate the forecast of weather variables from NWP models, Model Output Statistics (MOS) and Perfect Prog: (i) The Model Output Statistics (*Glahn and Lowry*, 1972) relates the observed weather (predictand) to the variables forecast by the NWP model. In order to use this approach, a sample of model forecast data (and observations) must be available to develop the statistics. (ii) The Perfect Prog approach (*Klein, Lewis and Enger*, 1959) is based on the assumption that NWP forecasts are perfect. Thus an observed weather variable (predictand) can be related to other variables observed at the same time. Two examples of these approaches towards forecast-skill predictions will be briefly discussed (and some preliminary results presented).

The MOS-approach leads to a statistical-numerical regression scheme (*Leslie, Fraedrich and Glowacki,* 1989) which predicts the model forecast error (predictand) from the initial analysis and the model forecast at neighbouring points (predictors). Here it can be interpreted as an extension of a statistical correction scheme by which regression relationships between the NWP output and the desired variable (the forecast error) are established. This technique has been developed at the Australian Bureau of Meteorology Research Centre (BMRC). (ii) The perfect-prog approach to short term predictability forecasts is based on the estimates of the Liapunov exponents determined from ensembles (or clusters) of past weather analogues (*Fraedrich and ZiehmannSchlumbohm,* 1991). The members of an ensemble are 3 hourly single station (mesoscale) weather trajectories which are embedded in large scale regimes of similar structure defined by cluster analysis. The observed mesoscale dynamics evolves in a phase space spanned by time delay coordinates. In this sense it is guaranteed that the faster developing mesoscale dynamics is appropriately embedded in the slower synoptic scale regime defined by a subregion of the daily weather map.

2. MODEL FORECAST ERROR (MFE) PREDICTIONS: A STATISTICAL-NUMERICAL REGRESSION

METHOD: The statistical prediction of the model forecast errors is based on multiple linear regression methods. Predictands are the model forecast errors (MFE's) at grid points; predictors are initial analyses and the model forecasts (both detrended). The regression technique applied (*Glowacki,* 1988) is based only on a small number (about 10) of proximate grid points, because up to 36 hours ahead the forecast errors are found to be uncorrelated with distant locations. Thus, at a grid point, j, the MFE, e_j is estimated:

$$e_j = b_j + \sum_{s=1}^{k} e_{j,s} R_{j,s}^{t}$$

$$e_{j,s} = C_o + \sum_{i=1}^{n} C_i P_i^a + \sum_{i=n+1}^{2n} C_i P_i^m$$

where the bias is b_j; $e_{j,s}$ is the estimated model forecast error at j using the predictors of the subregion s, the multiple correlations are $R_{j,s,}$ and k is the number of subregions, which include the point j; t is a power law weight. Furthermore, the model forecast error $e_{j,s}$ depends on the number n of gridpoints within the subregions s, the analysis and model predictors are P^a and P^m associated with the regression coefficients C_i for each point j in all subregions s.

APPLICATION: The regression technique has been applied to NWP-model forecasts and analyses of the Australian Bureau of Meteorology operational regional model in the Australian region: The basic NWP grid has a resolution of 150 km and consists of 65x40 points. The 13x8 point subgrid was used for the MFE regressions. This subgrid was further subdivided into overlapping subregions (of $n=10$ points) to reduce the number of predictors as large distances do not effect the predictand. Finally, four 5x3 point areas were chosen for localized MFE-forecasts (FIGURE 1). Predictors are the analysis and the 24 hour prediction of the sea level pressure field; the predictand is the 24 hour rms forecast error. The regression coefficients are calculated monthly from 5 years (1980-84) archived operational numerical analyses and forecasts to obtain a 24 hour MFE-prediction scheme. It has been tested by an independent data set for the summer period, December-February 1984/85, and for winter, June-August 1985.

ACCURACY MEASURES OF MFE-PREDICTIONS: The independent sea level pressure forecast errors (FIGURE 2a) are evaluated by the following measures: significance of the correlations, categorical predictions, and local forecast error. The significance of the predicted and observed MFE must be tested before MFE-predictions can be used. The student's t-test, $t=r[(N-2)/(1-r^2)]^{1/2}$ with sample size N, was employed at each gridpoint using the null-hypothesis of zero correlation. Categorical predictions: Four forecast classes can be identified in the MFE-distribution, (1) very good, (2) good, (3) mediocre, and (4) poor to very poor. They are are associated with the rms MFE-intervals < 2.5, 2.5 to 2.9, 3.0 to 3.4 and ≥ 3.4 mb, respectively. The skill score (*Panofsky and Brier,* 1958), $S=(C-E)/(T-E)$, is based on the expected number of correct forecasts E, the total number of forecasts T and the number of correct chance forecasts $C=\Sigma P_i O_i/T$ for $i=1,\ldots4$, where P_i and O_i are the subtotals of predicted and observed values of each category. A chi-squared test needs to be applied to reject/accept that the MFE-predictions were obtained by chance. Local forecast errors: Four subregions have been chosen to provide the forecast quality in specific areas: two in the Australian tropics (east and west: TE, TW); two in the higher latitudes of the Australian continent (southeast and southwest: SE, SW).

RESULTS OF MFE-PREDICTIONS: The method was applied to a summer and a winter season (December-February 1984/85, June-August 1985, see FIGURE 2). The following results are noted: Correlation coefficients were calculated at all 104 gridpoints leading to averages of 0.54 for summer and 0.51 for winter. The null-hypothesis of zero correlations was rejected for $r>0.2$ at the 99% level (for $r>0.28$ at the 99.9% level) based on the Student's t-test. A scatter diagram (FIGURE 2) supported the high level of correlation between predicted and observed rms-MFE and

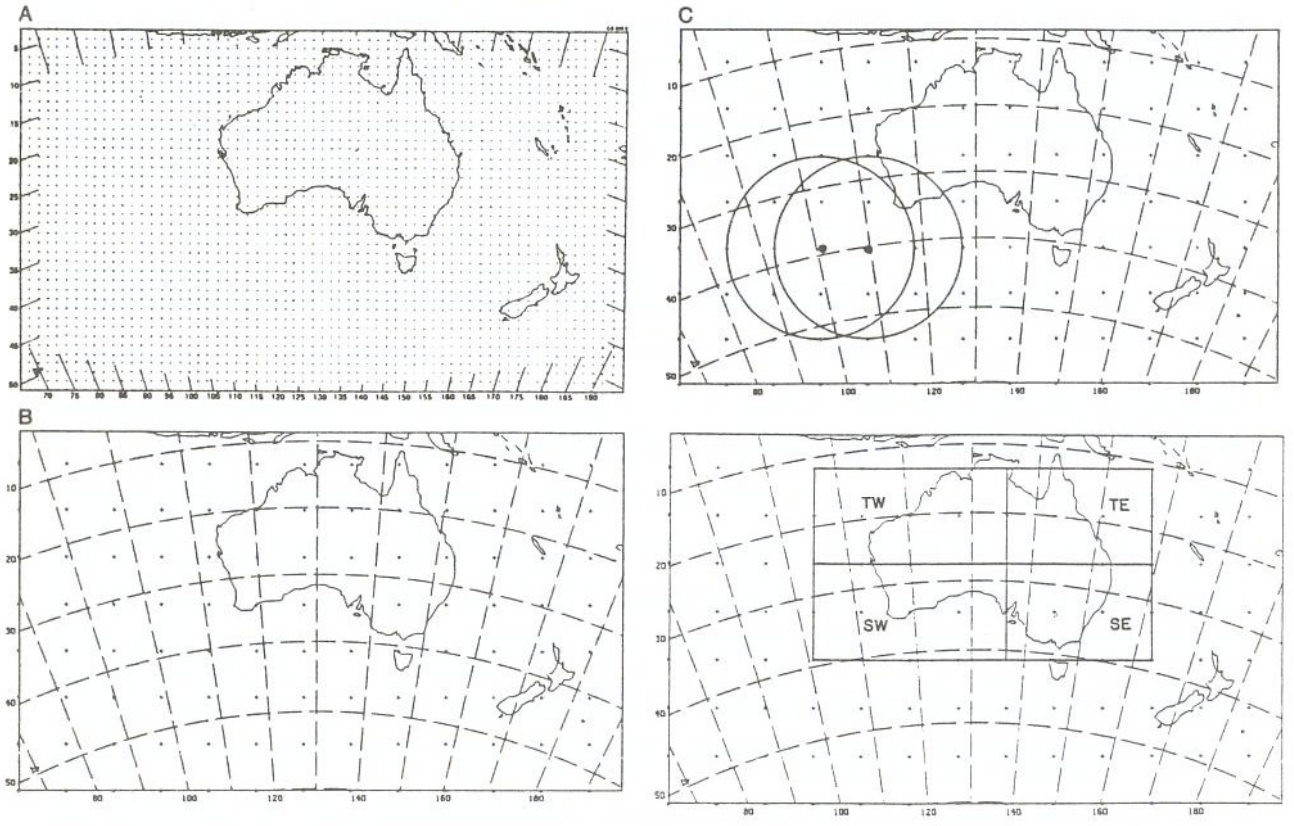

Fig. 1 The 65x40 point Australian region NWP grid of 150 km resolution; the 13x8 subgrid used for model forecast error (MFE) predictions; the overlapping subregions used for multiple linear regression; the local subregions for local MFE predictions.

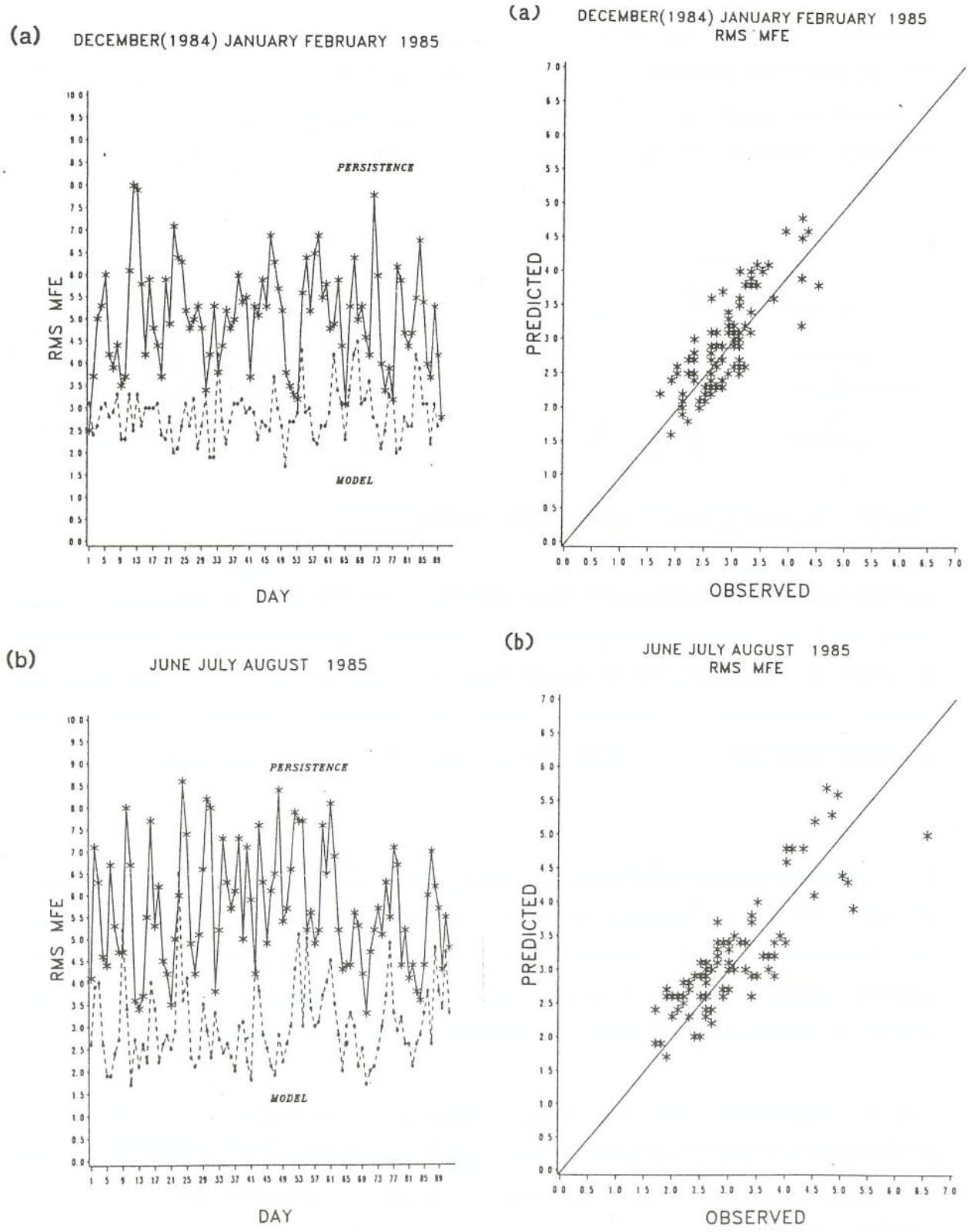

Fig. 2 Actual rms 24 hour model forecast errors (dashed line) for the Australian Bureau of Meteorology operational regional model for summer and winter (left); rms errors of persistence (full line) are also shown. Scatter diagram for predicted and actual rms 24 hour model forecast errors for summer and winter (right).

showed only very small bias (see also the distributions in FIGURE 3). From the 4x4 contingency table of categorical predictions the summer (winter) skill score was S = 0.27 (0.25). The subsequent chi-squared test lead to a rejection of the null-hypothesis that MFE-predictions were obtained by chance on a 99.9% level.

OBSERVED/PREDICTED:		1	2	3	4
1	very good	13; 10	9; 5	3; 1	0; 0
2	good	5; 9	11; 17	5; 8	1; 1
3	mediocre	1; 3	6; 6	10; 10	4; 3
4	poor	0; 0	2; 1	8; 8	12; 8

Table 1: Contingency table of forecast errors (summer; winter)

The four small (5x3 point) local areas within Australia allow a detailed MFE-prediction to analyse errors which may be systematic, orographically induced or related to a particular weather system at a given day. The averaged correlations in these small areas are above $r > .52$ (both winter and summer). In particular in the tropical eastern and southeastern Australian region, the high correlation coefficient $r > 0.55$ indicates the usefulness of the skill prediction scheme.

3. SHORT-TERM PREDICTABILITY ESTIMATES BY PHASE-SPACE ANALOGUES

METHOD: The analysis consists of two steps: Large scale flow patterns are defined as clusters of daily weather maps and the associated short term predictability is determined by the largest Liapunov exponent using 3-hourly single station observations characterising the mesoscale; this exponent is a measure of the exponential growth-rate of infinitesimally small 'errors'.

First, the synoptic scale patterns are defined by cluster analysis (proposed by *Ward*, 1963, see also *Wishart*, 1969) of daily observations. Given a base point (Berlin), its regional synoptic scale environment (a subregion of the weather map) is deduced from daily observed surface pressure anomalies: North-South gradient (NS) is defined by a pressure difference (Prague-Fanö); an East-West pressure gradient (EW) by another difference (Potsdam-Uccle), and the pressure mean of the four stations, $<P>$. These three variables span the synoptic scale state space characterizing the direction and strength of the large scale flow and the amplitude of the pressure system. The

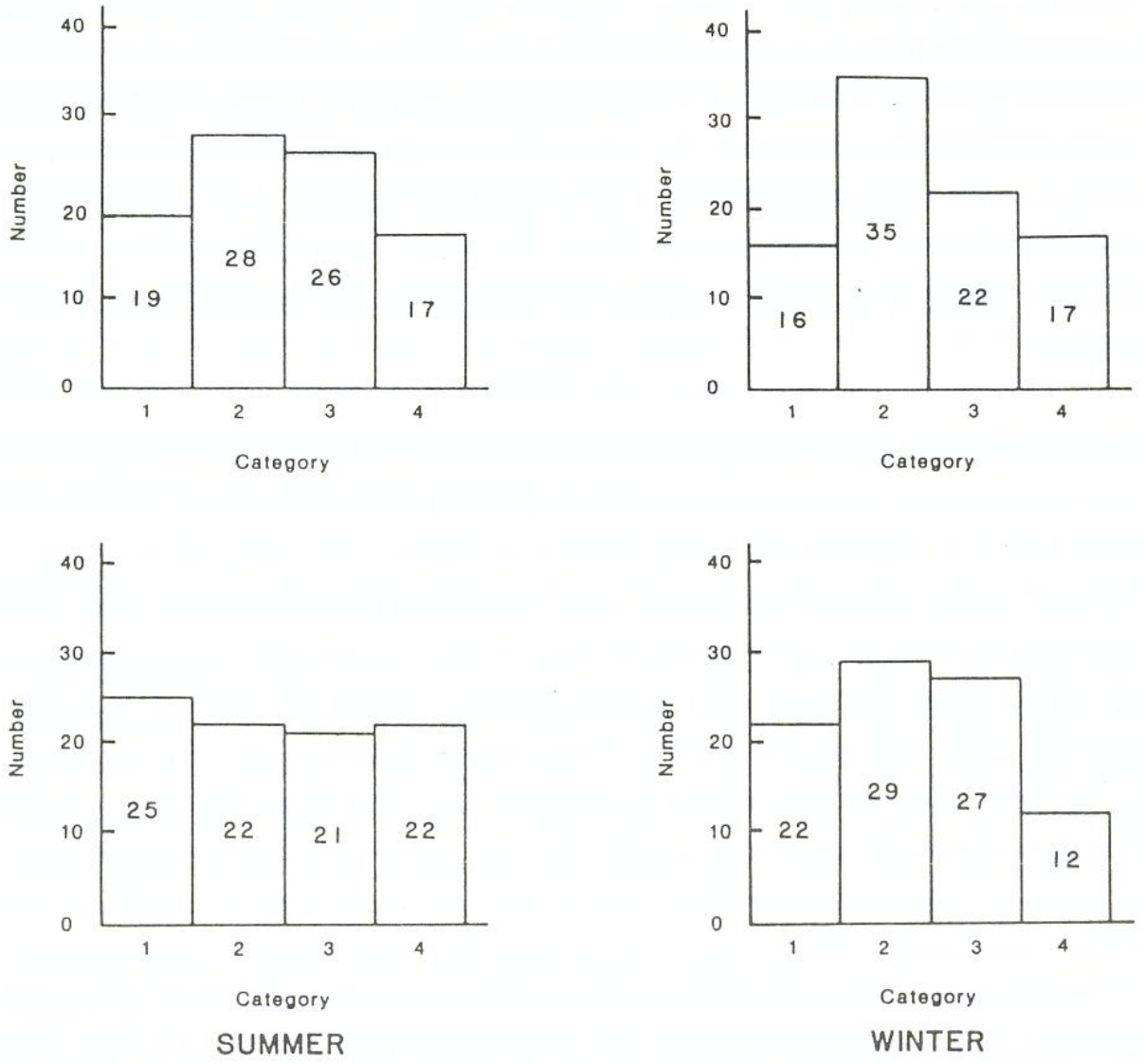

Fig. 3 Histograms showing numbers of predicted and observed rms 24-hour model forecast errors in four categories for summer and winter.

subsequent cluster analysis is based on the Euclidean distances in phase space between all possible pairs of realisations (leading to a symmetric distance matrix with zero main diagonal). The cluster analysis procedure commences from the two nearest neighbours, which are averaged to replace its two predecessors with the weight 2. Now the procedure is iterated with the revised distance matrix comparing the remaining pairs of weather states; new nearest neighbours are identified, averaged and the procedure is repeated until only one map (the climate mean) remains. That is, the climate mean is comprised by the average of two clusters, both of which are generated by two predecessors etc. (see also Wallace et al., 1991). The cut-off for clustering can be defined by the measure of homogeneity (or heterogeneity) in terms of (a sum of) cluster variances. Now large scale regimes can be identified by centroids (*Molteni, Tibaldi and Palmer,* 1990) defined by the mean and standard deviation of each cluster. Within these regimes the short term predictability can be deduced.

Secondly, given the large scale regime, the short term predictability is determined from 3-hourly single station weather observations (at Berlin) by the estimated Lyapunov exponent. It is determined by the Wolf et al. Algorithm (*Wolf, Swift, Swinney and Vastano*, 1985) using only those pieces of the single station mesoscale weather time series which belong to the same large scale cluster: 'Given the time series $\underline{x}(t)$, an m-dimensional phase portrait is reconstructed with delay coordinates, i.e. a point on the attractor is given by $\{\underline{x}(t), \underline{x}(t+T),\ldots,\underline{x}(t+(m-1)T\}$ where T is an almost arbitrarily chosen delay time. We denote the distance between these two points $L(t_0)$. At a later time t_1, the initial length will have evolved to the length $L'(t_1)$. The length element is propagated through the attractor for a time short enough so that only small scale attractor structure is likely to be examined. If the evolution time is too large we may see L' shrink as the two trajectories which define it pass through a folding region of the attractor. This would lead to an underestimation of λ_1. We now look for a new data point that satisfies two criteria reasonably well: its separation $L(t_1)$ from the evolved fiducial point is small, and the angular separation between the evolved and replacement elements is small. If an adequate replacement point cannot be found, we retain the points that were being used. This procedure is repeated until the fiducial trajectory has traversed the entire data file, at which point we estimate

$$\lambda_1 = \frac{1}{t_M - t_o} \sum_{k=1}^{M} \ln \frac{L'(t_k)}{L(t_{k-1})}$$

where M is the total number of replacement steps, t_{k+1}-t_k.'

ESTIMATING THE SHORT TERM PREDICTABILITY CLIMATE: A first application of the procedure is confined to a relatively small data set of 33 winter months (January 1953-84 or 992 maps). Weather regimes are defined by the centroids of the 5 major clusters (FIGURE 4) which account for about 90% of all weather maps. These regimes represent two high pressure patterns (1 and 5; anticyclonic), 2 low pressure structures (2 and 3, cyclonic) and one (4) with pressure near the climate mean. The transition matrix between these clusters shows a relatively high degree of regime persistence. Relatively frequent transience is observed from regime 4 to 1, which occurs after the passage of a disturbance when, starting with a cold air outbreak, a high pressure system follows from the west. A Markov chain approach leads to the eigenvalues, r_i for i=1,...,5, of the transition probability matrix: (r_i) = (1.0, 0.53+/-i0.01, 0.33, 0.24) which characterize the time evolution of a stochastic process. The first eigenvalue $r_1=1$ is attached to the climate mean state vector given by the relative occurence of the weather regimes: (0.35, 0.20, 0.16, 0.15, 0.14); the second and third give a strong decay associated with a small-amplitude oscillation (that is the passage of disturbances), and the last two also contribute to the decay towards the climate mean state. The largest decay rate is $\ln\mathrm{Re}(r_2)=-0.63$ per day.

FROM/TO	Σ	EW	NS	P	1	2	3	4	5	λ_{TE}	λ_P
1. South-East-High	307	4.4	-5.1	9.5	.71	.12	.02	.05	.10	**1.13**	**1.29**
2. South-West-Low	198	4.9	4.8	-7.6	.10	.63	.13	.11	.03	1.02	1.17
3. West-Low	153	-7.7	6.3	-8.2	.03	.16	.51	.19	.11	1.00	1.15
4. North-West-Mean	130	-3.9	-8.6	1.3	.45	.05	.08	.32	.10	1.01	**1.00**
5. West-High	119	-4.8	3.6	8.5	.15	.06	.18	.14	.47	**0.91**	1.10

Table 2: The five synoptic scale clusters which classify about 90% of all daily patterns: EW and NS are the zonal and meridional pressure differences, P is the mean pressure level, Σ the number of objects. Estimates of the transition probabilities and of the local Liapunov exponents for equivalent potential temperature, λ_{TE}, and pressure, λ_P, are also given (in 1/day).

Short term predictability can now be estimated for these dominating weather regimes from 3-hourly time series of surface pressure, P, and the equivalent temperature, TE=t+1.55p(v) at Berlin; p(v) is the water vapour pressure in mb. The estimates of the Liapunov exponents (λ_{TE} and λ_P) are made only from those pieces of trajectories which commence within the same cluster or large scale weather regime. The embedding dimension of the delay coordinate phase space is chosen to cover

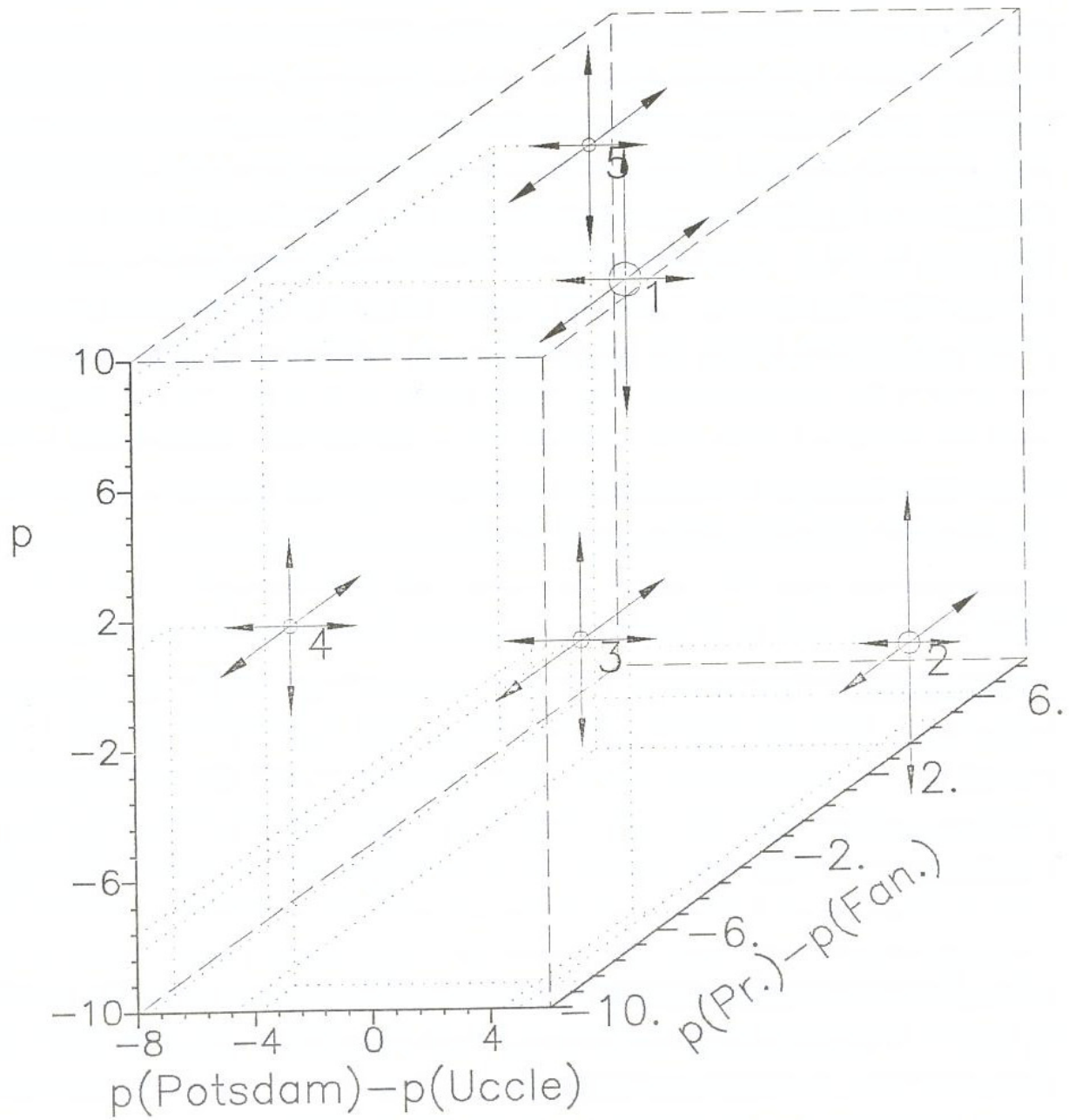

Fig. 4 Position of centroids of the five clusters representing regional weather regimes in terms of east-west and north-south pressure differences (Potsdam-Uccle, Prague-Fanö) and the four station means. Their locus is given by the respective means of the clusters, arrows indicating the standard deviation, the circle size is proportional to the number of objects in the cluster.

one day: m=5 for T=6h (or m=9 for T=3h). The calculations of the Liapunov exponents are based on a 27 hours evolution time to obtain the estimate of the spread of the nearby trajectories from independent (that is non-overlapping) data. These regime related Liapunov exponent estimates are shown in TABLE 2; the results do not change considerably when varying the embedding m and the sampling time T. Furthermore, the extreme values (bold) are significantly different from one another on a 99% level using the two-tailed t-test statistic $t=(\lambda_a - \lambda_b)/(s_a^2/n_a - s_b^2/n_b)^{1/2}$, with variances s_i^2 and sample size n_i. That is, the null-hypothesis can be rejected that the λ-estimates are obtained from the same sample.

The following results should be noted: (i) The estimates of the Liapunov exponent show differences between the thermodynamic and dynamic variables (λ_{TE} and λ_P) and the weather regimes $\lambda(i=1,...5)$. (ii) The West-Low and West-High clusters (3) and (5) reveal the best predictability estimates (smallest error growth rate or Liapunov exponent) for λ_{TE}. λ_{TE} (3,5) < λ_{TE} (1,2,4): Westerly flow advects relatively homogeneous maritime air masses which lead to an enhanced predictability of the thermodynamic properties. (iii) The South-East-High cluster (1) shows worst predictability for both dynamic and thermodynamic properties. It is known from other nonlinear system analyses that predictability is reduced in high pressure situations with weak gradients (*Kepenne and Nicolis*, 1989; see also summer versus winter forecast skill). There is a local downstream effect from the Alps (cluster 1) on the Berlin cloudiness and temperature depending sensitively on the direction of the meridional flow. (iv) The cluster sequence 2-3-4 represents the passage of a disturbance centered about the regime (3). $\lambda_P(2,3) > \lambda_P(4)$: Wave disturbances are a less common feature behind the trough.

5. DISCUSSION

A statistical correction scheme (*Glowacki,* 1988) has been applied to predict model rms-forecast errors (MFE) from neighbouring gridpoints of a regional NWP model initial analysis and forecast using multiple linear regression. This leads to a MOS-type forecast error prediction scheme which shows the following results:

* Independent predictions of the MFE and the actual error are highly correlated (with correlation coefficients of 0.54 and 0.51 in summer and winter).

* Categorical forecast error predictions (four categories) of the rms MFE lead to highly significant information compared to chance.

* Forecast error predictions of local subareas of the Australian region are more useful than for the entire domain.

Furthermore, a basis for a single station Perfect Prog approach towards short term predictability forecasts has been introduced. It is related to an estimate of the Liapunov exponent from observed time series. The estimates can be obtained from analogues of a (predicted) short term or mesoscale weather trajectory embedded in a larger scale regime which are obtained from local past weather observations. However, a real time trial for testing the procedure has not been made.

REFERENCES

Epstein, E.S., 1969: Stochastic dynamic prediction. Tellus, 21, 739-759.

Glahn, H.R., and D.A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. J.Appl.Meteor., 11, 1203-1211.

Glowacki, T., 1988: Statistical corrections to dynamical model predictions. Mon.Wea.Rev., 116, 2614-2627.

Hoffmann, R.N., and E. Kalnay, 1983: Lagged-average forecasting. Tellus, 35A, 100-118.

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. Mon.Wea.Rev., 115, 349-356.

Kepenne, C.L. and C. Nicolos, 1989: Global properties and local structure of the weather attractor over western Europe. J.Atmos.Sci. 46, 2356-2370.

Klein, W.H., B.M. Lewis, and I. Enger, 1959: Objective prediction of five day mean temperature during winter. J.Meteor., 16, 672-682.

Leith, C., 1974: Theoretical skill of Monte Carlo forecasts. Mon.Wea. Rev., 102, 409-418.

Leslie, L.M., K. Fraedrich, and T.J. Glowacki, 1989: Forecasting the skill of a regional numerical weather prediction model. Mon.Wea.Rev., 117, 550-557.

Molteni, F., S. Tibaldi, and T.N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. Q.J.R.Meteorol.Soc., 116, 31-67.

Panofsky, H.A., and G.W. Brier, 1958: Some Applications of Statistics to Meteorology. Pennsylvania State University, 224 pp.

Wallace, J.M., X. Cheng, and D. Sun, 1991: Does low-frequency atmospheric variability exhibit regime like behaviour? Tellus, 43AB, 16-26.

Ward, J., 1963: Hierarchical grouping to optimize an objective function. J.Amer.Statist. Assoc., 58, 236-244.

Wishart, D., 1969: An algorithm for hierarchical classifications. Biometrics, 25, 165-170.

Wolf, A., J.B. Swift, H.L. Swinney, and J.A. Vastano, 1985: Determining Lyapunov exponents from a time series. Physica, 16D, 285-317.