

KLAUS FRAEDRICH

**Forecasting: Predictability, Probability, and  
Persistence**

Sonderdruck aus:

MITTEILUNGEN  
DER MATHEMATISCHEN GESELLSCHAFT  
IN HAMBURG

BAND XXVI  
2007



IM AUFTRAG DER GESELLSCHAFT HERAUSGEGEBEN VON  
T. ANDREAE · A. BARLOTTI · W. BENZ · H. KARZEL  
A. KREUZER · G. OPFER · G. TISCHEL · H. WEFELSCHEID

SCHRIFTLEITUNG  
H. KIECHLE — G. TISCHEL

---



## Forecasting: Predictability, Probability, and Persistence

By KLAUS FRAEDRICH

**Summary:** Weather forecasting, though dominated by numerical weather prediction (NWP) models, is discussed in terms of time series analysis: (i) Nearest neighbour search and scaling is linked to predictability estimates and ensemble prediction; ensemble forecasts of tropical cyclone tracks provide a practical example. (ii) Short term memory in weather time series is tested and used to estimate the future probability of precipitation by discrete Markov chains. (iii) Finally, the weather's short term memory allows a first order autoregressive process to serve as a surrogate atmosphere, for which persistence forecasts are made and evaluated. This conceptual model demonstrates predictability experiments and their analyses as they are common in a practical weather forecast environment.

### Contents

1. Introduction: Forecasting and memory
2. Predictability and prediction: Analog ensemble forecasts
3. Probability: Markov chain forecasting
4. Persistence in red noise: Conceptual forecast experiments
5. Conclusion

### 1 Introduction: Forecasting and memory

Forecasting the evolution of a non-linear system is closely linked with its memory. Short-term memory lies behind an exponentially decreasing auto-correlation with a finite integral time-scale. It allows predictions of the *first kind* (Table 1.1, following Lorenz), which show sensitive dependence on initial conditions caused by internally occurring instabilities at fixed boundary conditions. Daily numerical weather predictions (NWP), which are verified by comparing predicted with observed fields, are prominent examples. Predictability, that is prediction of the forecast error, can be estimated by ensemble forecasts based on one model only, as the bias in the model's climatology is anticipated to be small. Long-term memory is described by an auto-correlation with a power-law decline of certain magnitude, which (i) extends over many time scales and compartments of the climate system, (ii) leads to an infinitely large integral time-scale, and (iii) characterises low frequency variability. In this sense, predictability of processes with long-term memory, as observed in N-Atlantic sea surface temperatures (but not in precipitation), may possibly be utilized to extend forecasts of the first kind into the domain of second kind predictions.

Predictions of the *second kind* describe the response of systems to changing boundary conditions, which can be linked with structural stability. Climate scenario computations are prominent examples, for which boundary conditions are prescribed like doubling of the anticipated future atmospheric CO-2 content. Climate states associated with prescribed boundary conditions are characterized by long term equilibrium statistics, that is, moments and extremes. Predictability is estimated by super-ensembles from a set of climate models, differing in their long-term statistics.

Table 1.1: Prediction of the first and second kind (schematic).

<i>Time scale (in days)</i>	$10^{-1}$	1	10	$10^2$	$10^3$	$10^4$
Atmosphere	pred. 1.kind => prediction of 2.kind					
Atmosphere and Ocean	prediction of 1.kind => prediction of 2.kind					

In the following, prediction of the first kind will be analysed. For practical weather forecasting, numerical weather prediction (NWP) models play the eminent role; their advancement has been remarkable since the first concepts (Richardson 1922) emerged and first applications (Charney, Fjörtoft, and von Neumann 1950) were made. However, alternative methods are presented here; these are empirical-stochastic forecasts using time series analysis techniques. A brief glossary on predictability and prediction is presented first.

*Predictability analysis* interprets the budget of a forecast error in terms of the time evolution of, for example, the distance between forecast and verification. These analyses can be interpreted in analogy to a diffusion process in state space: (a) *Single particle* diffusion corresponds to an analysis of the verification trajectory only; its distance from the origin represents the error growth of *persistence forecasts*, which is an important reference for forecast quality evaluation. (b) *Two-particle* diffusion provides the standard frame for an individual forecast evolving in relation to its verification. (c) An *ensemble of particles* ('prediction plume') with diverging trajectories represents a forecast ensemble; defining appropriate ensemble members and relating their statistics to the verification trajectory is subject of ongoing research.

*Predictability experiments* provide the data for diagnosing the error budget. *External (practical)* predictability experiments are, for example, practical weather or climate forecasts. The bias due to differences between model and real climate (systematic error) is one of the problems met in analysing NWP-products in an imperfect model environment. *Internal (theoretical)* predictability is related to error budgets due to small perturbations in initial or boundary conditions generated by identical model atmospheres; that is, 'identical twin' experiments in a perfect model environment.

The outline of these notes is as follows: First (section 2 on predictability), *non-linear* analog methods are applied to search for nearest neighbours or ensemble members in phase space in order to extrapolate their time evolutions, which are known from the past data sets. Thus, topological properties like dimension

and entropy (predictability) are estimated and, for example, hurricane tracks are predicted. Utilizing the commonly neglected forecast errors and combining independent predictions can improve forecast accuracy. The methods have been developed in the eighties, but have been used in meteorology even before the seventies (Lorenz 1969). Next (section 3 on probability), methods to predict single station probability of precipitation (*PoP*) are presented, applied, and evaluated. Since Cooke (1905) probabilistic forecasting of local weather states, which implies a forecast of forecast errors, has a long tradition. In the sixties, Markov chains have been tested and applied to quantify probability of precipitation forecasts (Gabriel and Neumann 1962). Finally (section 4 on persistence) prediction and predictability experiments are performed analytically in a conceptual model environment simulating the practical weather forecast process: Persistence represents the forecast scheme, which is used to predict a red noise surrogate atmosphere. Section 5 summarizes the results.

## 2 Predictability and prediction: Analog ensemble forecasts

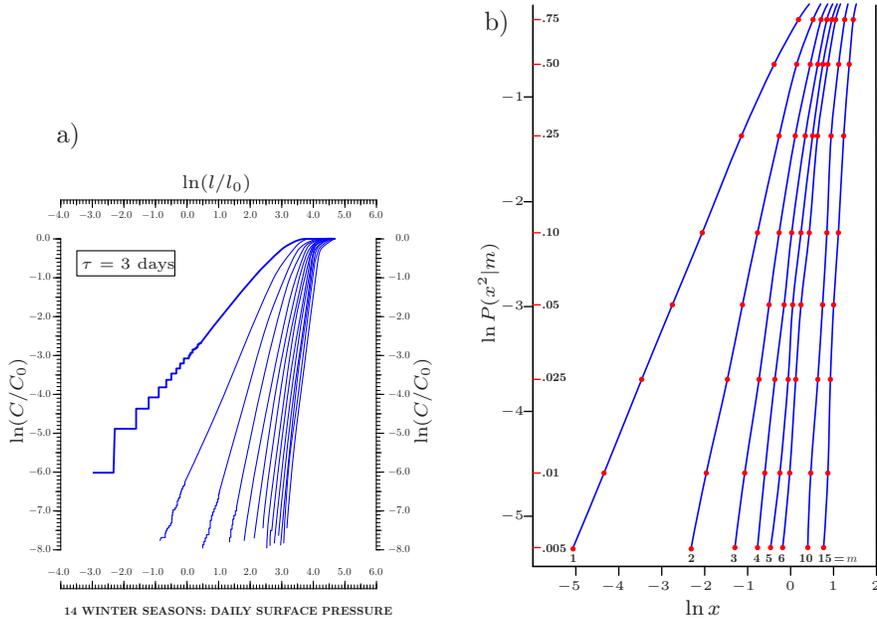
Predictability is quantified by the time evolution of an ensemble of nearest neighbour states (or analogs), which are embedded in a time-delay coordinate phase space, and is utilized to estimate future states from weather history. Employing the Grassberger-Proccacia (1984) algorithm the distance scaling of an ensemble of nearest neighbors yields the local *correlation dimension*  $D_2$ , and its spread in time yields the local *entropy*  $K_2$  (for reviews see Tsonis and Elsner 1989, Abarbanel et al. 1993, Götz 1995, for examples see Fraedrich 1986, 1988, and for some later developments see Fraedrich and Wang 1993, Shirer et al. 1995). Thus, atmospheric predictability is revealed by naturally occurring analogs (Lorenz 1969, revisited by Fraedrich and Leslie 1991) with analog ensemble forecasting being a practical application of non-linear systems analysis. Further development, for example, adapting a forecast error minimising metric in phase space for proper analog search has improved individual and ensemble mean forecasts of Hurricane tracks (Fraedrich and Rückert 1998, Fraedrich et al. 2003, Langmack et al. 2007). For general applications in daily weather forecasts, however, the method suffers from data scarcity, because the atmosphere's recurrence time is vast, even compared with its life span (van den Dool 1994).

### 2.1 Analog ensembles: Dimension and entropy

To make use of Takens' theorem on the reconstruction of the dynamics of a non-linear system evolving with time, consider an observed time series. The system's evolution can be displayed in a time-delay coordinate phase space of sufficiently large embedding dimension. Scaling the spread of initial ensemble members or nearest neighbours (perfect ensemble hypothesis) provides the  $D_2$ -dimension; here the spread characterizes the mutual distance between ensemble members and thus the initial error. The  $K_2$ -entropy estimates the time evolution

of the ensemble spread by increasing the embedding dimension of the time-delay coordinate space; it describes the main axes of an initially small sphere expanding into an ellipsoid as the trajectories diverge in time (perfect model hypothesis). Thus predictability can be defined and quantified; a first reliability test (Fraedrich 1988) with surrogate data may appear useful; a demonstration follows.

Take a single or vector time series  $\mathbf{y}(t)$  of observed local weather variables measured at time steps  $\tau$ . Pieces of this time series, which commence at  $t_i$  and last for  $(M - 1)\tau$  time steps, are used to define local weather states (or points) embedded in an  $M$ -dimensional phase-space spanned by time-delay coordinates,  $\mathbf{y}(t_i) = \{y(t_i), y(t_i - \tau), \dots, y(t_i - (M - 1)\tau)\}$ ; weather states at  $t_k$ , in the same phase space are called past weather 'analogs' or nearest neighbors  $\mathbf{y}(t_k)$  to the base point  $\mathbf{y}(t_i)$ , if both are independent  $|t_i - t_k| > \tau$  in this space and their Euclidean distance,  $d_{ik} = |\mathbf{y}(t_i) - \mathbf{y}(t_k)|$ , is small. Note that a delay time-lag  $\tau$  near de-correlation time-scale may guarantee linear independence of phase-space coordinates. Scaling the number statistics of analog-pairs (with distance in phase space of dimension  $m$ ) leads to estimates of the dimension of local weather dynamics and its entropy as a measure of predictability. Scaling behaviour is determined from the cumulative number distribution (correlation integral)  $C_M(l)$  of all  $K$  pairs of points in phase space.



**Figure 2.1** Correlation integrals (cumulative l-distance distribution),  $C_M(l)$ , changing with increasing embedding dimensions  $1, \dots, (M - 1)$ : (a) Local surface pressure and (b) surrogate data (AR(1)-process) determined analytically.

(i) *Analog ensembles*: Count the relative number of  $k = 1, \dots, K$  analogs  $\mathbf{y}(t_k)$ , whose distance from the local base point  $\mathbf{y}(t_i)$  is smaller than a prescribed threshold ( $d_{ik} < l$ ). This provides a cumulative number distribution, which increases to one with increasing threshold  $l$ . Here the Heaviside function is used with  $\Theta(a) = 0$  or  $1$ , if  $a > 0$  or  $a < 0$ . Repeating this procedure for all  $i = 1, \dots, K$  base points,  $\mathbf{y}(t_i)$ , and subsequent averaging gives the correlation integral (Grassberger and Procaccia 1983, 1984) in terms of the averaged cumulative number distribution (Figure 2.1):

$$C_M(l) = K^{-1} \sum_k^K \left[ \sum_i^{K-1} \Theta(l - d_{ik}) / (K - 1) \right].$$

(ii) *Perfect ensemble hypothesis* and distance scaling of the correlation integral (at fixed embedding) leads to the mean correlation dimension  $D$ . Consider the number of pairs of points, which are homogeneously distributed on a line (surface or volume) embedded in an  $M$ -dimensional space. This number grows according to a linear (quadratic or cubic) power-law:

$$C_M(l) \sim l^D$$

with  $D = 1, 2$ , or  $3$  at fixed  $M = 1, 2$ , or  $3$ . Likewise, local weather states, which evolve in a delay-coordinate phase space, describe an object with a dimension that characterises the dynamics of the system. The dimension is obtained by scaling of  $C_M(l)$  with  $l \geq 0$ , which occurs only at sufficiently high embedding  $M$ ; that is, if the pairs of points counted are perfect ensemble members in the  $M$ -dimensional  $l$ -sphere.

(iii) *Perfect model hypothesis* and time scaling of the correlation integral leads to the estimates of the local mean predictability (entropy or information loss): Extending the lengths of the weather trajectories from  $(M - 1)\tau$  to  $M\tau$ , that is, increasing the embedding dimension by one, reduces the number of analog pairs trapped in a sphere confined by the  $l$ -threshold. Their chance being trapped in the  $M$ -dimensional  $l$ -sphere decreases proportional to

$$C_M(l) \sim \exp(-M\tau L_k).$$

Thus predictability characterises the system's dynamics in delay coordinate phase space. It can be estimated by the sum of the positive Lyapunov (or characteristic) exponents, which contribute to the expansion of an infinitesimally small sphere of initially close trajectories into an ellipse with  $m$  expanding main axes:  $H = \sum_{m=1}^M L_m$  with  $L_m > 0$ , which is related to the Kolmogoroff-Sinai entropy.

(iv) *Interpretation*: Data analysis of weather variables reveals a relatively small dimension,  $D_2 \sim 7$ , which does not suggest that global weather has low dimensionality. Instead, this result represents merely the projection of the global weather dynamics onto a single station time series, which is associated with

an average local predictability (or entropy) corresponding to  $H \sim 2$  weeks. Estimating the dimensionality of the global system, however, requires a set of independent stations (and variables) separated by a distance of, say, the Rossby radius.

## 2.2 Tropical cyclone track forecasting

Theoretical predictability analysis provides the foundation for the methods of non-linear empirical forecasts based on historical weather analogs. One practical example is the prediction of tropical cyclone motion. Meteorological services in tropical regions regard that as their major weather forecasting problem, because Hurricanes are one of the most destructive natural hazards. Although hurricane position forecasting has been a major activity for many decades, it has defied rapid improvement, with an average of less than one per cent reduction per annum in the mean 48-hour position error over the past decade or so. It is worth noting that there are two distinct approaches to hurricane position forecasting: statistical regression techniques (CLIPER; see Leslie et al. 1990) blending climatology and persistence forecasts, and numerical weather prediction (NWP), whose skill has enhanced considerably due to the recent improvements in data assimilation. Both approaches may be combined to consensus or super-ensemble forecasts.

**Metric adaption:** Establishing empirical prediction tools requires the theory, building, and the validation of the forecast model. Here, a non-linear empirical prediction system is developed utilizing the analog-Ansatz described above to obtain ensemble forecasts where nearest neighbours of today’s weather are provided by their historical analogs in the time-delay coordinate phase space; that is, instead of all nearest neighbour pairs we take all neighbours to the reference point in phase space, given by the trajectory to be extended into the future by analog ensemble forecasting.

*Theory:* The future weather state (forecast) is estimated by analog evolutions of historical weather trajectories. These analog evolutions comprise an ensemble of size  $K$ , which are nearest neighbours of the reference (present) state and, from their ‘historical futures’, the real ensemble forecast is estimated. Such empirical forecast schemes require, in general, a finite set of parameters, which are optimized to obtain, in the average, a good forecast *plus* an estimate of the forecast error. The procedure presented here has been used by McNames (2000) extending Fraedrich and Rückert’s (1998) metric adaption procedure. Given the vector time series  $\mathbf{x}(t) = (y_1(t), \dots, y_N(t))$ , say the lat-long position of a hurricane ( $N=2$ ):

$$\mathbf{x}(t) = \begin{pmatrix} y_1(t), y_1(t - \tau), \dots, y_1(t - (M - 1)\tau) \\ \vdots \\ y_N(t), y_N(t - \tau), \dots, y_N(t - (M - 1)\tau) \end{pmatrix}$$

(i) Two parameters are sufficient for optimal forecasting: the time-delay embedding of the phase space and its metric, that is, the window length  $M$  and the weights  $\lambda_l$ , which characterize the system's memory duration and decay rate, respectively. The weights depend both on the vector component and its delay. The distance  $d_k$  between the basic state and its  $k$  neighbours (identified by  $t_k$ ) is needed for estimate the ranking weights  $r_k$  of the ensemble members

$$d_k = \sum_{n=1}^N \sum_{m=1}^M \lambda_l^2 (y_n(t-m) - y_n(t_k-m))^2 \text{ with } l = m + (k-1)M.$$

(ii) Two measures define the ensemble forecast, which are deduced from the 'historical futures' of the analog ensemble: the ensemble size,  $k = 1, \dots, K$ , and ranking,  $r_k$ , of the ensemble members. The weighted ensemble mean forecast,  $F_n(\mathbf{x})$ , is determined after the ranking has been estimated by minimizing the forecast error  $C$  (or cost function):

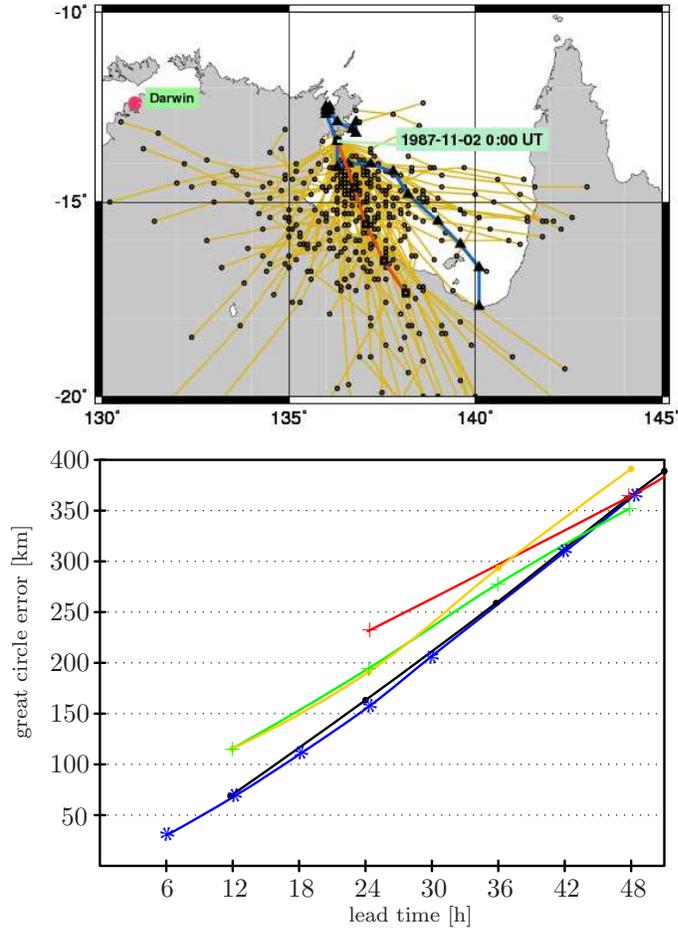
$$F_n(t_k) = \sum_{k=1}^K r_k^2 y_n(t_k + 1) / \left( \sum_{k=1}^K r_k^2 \right).$$

The ranking,  $r_k^2 = (1 - d_k^2/d_{K+1}^2)$ , is estimated to characterize the relevance of the  $k$ -th ensemble member in providing the ensemble mean prediction; the rankings are positive between zero and one and thus suitable to yield convergence of the optimization process (see also Fraedrich and Rückert 1998, McNames 2002, Langmack, H., F. Sielmann and K. Fraedrich, 2007). The cost-function  $C$  to be optimized measures the squared errors (Euclidean distance) comparing hindcasts (past forecasts) by the weighted ensemble mean  $F(t_k)$  with the respective verification,  $y(t_i + j + 1)$  commencing at  $t_i$ . Forecasts issued for lead time  $j + 1$ , depend on the historical analogs extending up to that lead time  $j = 0, \dots, J$ . The cost function includes averaging over the  $i = 1, \dots, I$  hindcast trials:

$$C = (1/IJ) \sum_{i=1}^I \sum_{j=1}^J |y_n(t_i + j + 1) - F_n(t_k + j)|^2.$$

*Model-building and validation:* The model building phase consists of two steps to estimate total number  $K$  and ranking  $r_k$  of the ensemble members by minimizing the cost-function  $C$ . The gradient algorithm is employed to minimize the cost-function. Applying the PARTAN algorithm, a recursive approach (McNames 2002) enables efficient computation. Now, ensemble mean and spread (variance) can be deduced. Altogether, 12 components appear to be optimal for the final prediction model: the  $N = 3$  state variables: position differences in latitude and longitude and its cosine of latitude, for a one day window with  $M = 4$  six-hourly time-lags. More variables result in larger errors. Note that

the dimension of Hurricane tracks in the Australian basin region is estimated to be about  $D \sim 8$  (Fraedrich and Leslie 1989); the upper bound for the embedding dimension,  $2D+1$  is thus satisfied, following Whitney's embedding theorem. A validation is presented for the Australian Hurricane basin (Figure 2.2) for the tropical cyclone Jason and the great circle forecast errors for a decade.

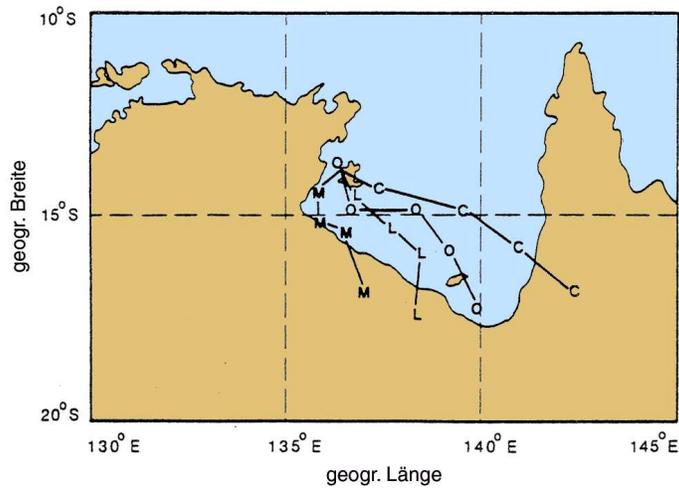


**Figure 2.2** Hurricane ensemble forecasts in the Australian basin for the tropical cyclone Jason (top, 10-12 February 1987, observed in blue, ensemble mean in red, ensemble members in yellow) and great circle forecast error of forecast track position versus lead time (bottom, 1991 to 2000): this method (blue), analog ensemble (Fraedrich et al. 2003, black), United Kingdom Meteorological Office (UKMO, red), Australian Bureau of Meteorology (BOM, green) and CLIPER (Climate-Persistence, yellow).

**Consensus forecasts:** Optimal weighting of super-ensemble forecasts (that is, forecasts made by models of differing bias) has received much attention in economics, management and statistics literature. In meteorology it is also known that consensus forecasts provide, in the average, more accurate results than the individual forecasts, which comprise the consensus. Although, according to Thompson (1977), this is the incontrovertible fact, it is recognized only recently. The following linear error minimizing multivariate combination forecast of tropical cyclone position ( $X^*$ ,  $Y^*$ ) shows considerable forecast improvement:

$$\begin{aligned} \text{combination} \quad X^* &= a_1 X_A + a_2 Y_A + a_3 X_B + a_4 Y_B + a_5 \\ Y^* &= a_6 X_A + a_7 Y_A + a_8 X_B + a_9 Y_B + a_{10} \end{aligned}$$

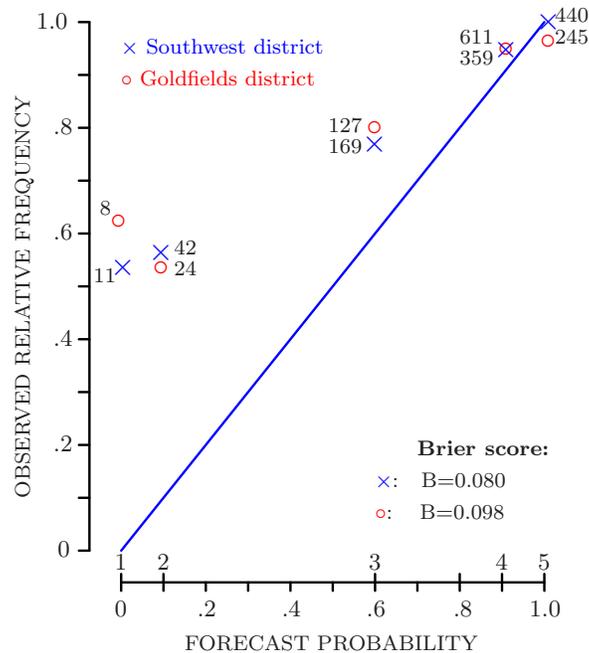
Here the numerical weather prediction (NWP) model (subscript  $A$ ) is combined in a simple manner with an empirical forecast system (subscript  $B$ ) based on non-linear analog and statistical regression techniques (CLIPER). Independent forecast trials show that the 24 and 48 hour position errors of the combination forecast can be reduced by 15 - 20% compared to the best individual scheme (Figure 2.3). Practical implementation has been reported (more in Leslie and Fraedrich 1990). Predictions by super-ensembles based on a set of different forecast models are being tested to supplement the classical single model ensemble forecasts modulated by initial conditions.



**Figure 2.3** Linear error minimising combination (multivariate) of short-term forecasts of tropical cyclone tracks: Comparison of forecast tracks of Australian tropical cyclone Jason (10-12 February 1987): observation (O), CLIPER (C), NWP model (M), linear combination (L).

### 3 Probability: Markov chain forecasting

**Who started it all?** A century ago the world's first probabilistic weather forecasts were issued in Western Australia and evaluated about eighty years later (Monthly Weather Review: Cooke 1906, Fraedrich and Leslie 1987). During the year 1905, daily weather forecasts for two districts in Western Australia were amended by quantitative weights of the forecasters' confidence in their predictions. These weights range from 1 to 5 or from 'not likely at all' to 'almost absolute certainty'. To evaluate skill and reliability (Figure 3.1), probabilities (0, 10, 60, 90 and 100%) are assigned to Cooke's weights. This trial shows that forecasts with high confidence weights were most frequently predicted, so that Brier scores (rms error of probability forecasts) attain high values. Since Cooke's time, automatic recording of long records of station data, sophisticated statistical and dynamical techniques and powerful computing devices have improved forecast methods to provide a distribution of possible future weather states and to issue probability forecasts.



**Figure 3.1** Reliability diagram (see also Figure 3.3) of the first probability weather forecasts for two districts near Perth in Western Australia during 1905. The first two of the five confidence weights have been quantified a posteriori in probabilistic terms (abscissa); the number of predictions is also indicated and the Brier scores (lower right).

### 3.1 Probability of precipitation (*PoP*)

Probability forecasts imply the prediction of forecast accuracy. Here, development, application, and analysis of a statistical-empirical model are briefly demonstrated. Although such forecast schemes serve mainly as a reference for the performance of more sophisticated numerical models, they have proven to be practically useful in very short-term forecasting of local weather, tropical cyclone motion and climate anomalies (long range or seasonal). Probability of precipitation, *PoP*, is a classical example of predicting both weather state *and* forecast accuracy, both of which are required from a forecaster. Three steps are discussed: model building, forecasting, forecast evaluation.

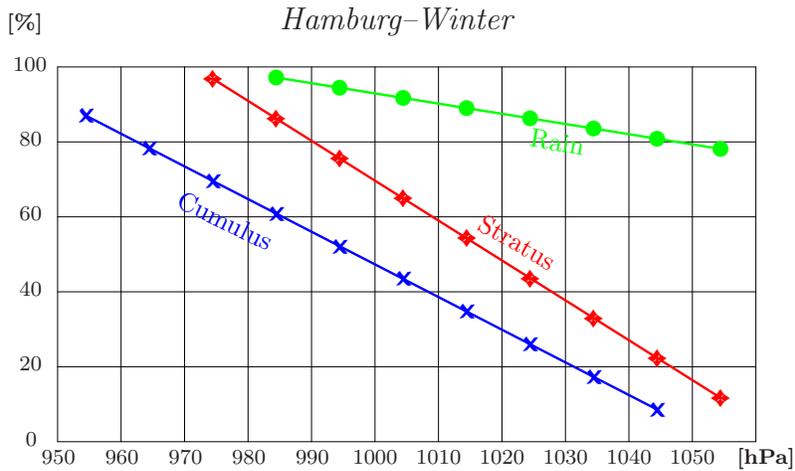
**Rainfall Markov chain:** Theory and model building consist of analysing internal time statistics (distributions of first passage times, period lengths etc.), testing the order of the process and the Markov properties of the underlying time series (for example, Fraedrich and Müller 1983, Kirk and Fraedrich 1998): Weather at a single station is best characterized by two extreme ‘sky-states’, sunshine and rainfall, so that three mutually exclusive weather states appear almost naturally: These two opposing extremes are ( $i = 1$ ) cloud-free including fair weather cumuli and ( $i = 3$ ) rainfall; the weather state  $i = 2$  lies between the two defined by stratus clouds. In practice, the rain state at a station is defined by measurable precipitation during the previous three hours or, according to the international weather code, by observing rain nearby. Now, a first-order Markov chain,  $p_i = P_{ij}p_j$ , consists of a state transition probability matrix,  $P_{ij}$  with  $\sum_j^n P_{ij} = 1$ , mapping an initial probability state vector,  $\mathbf{p} = p_j$  with  $\sum_j^n p_j = 1$  into its future state,  $\mathbf{p}^T = p_i$ , a component of which is the probability of precipitation,  $p_{i=3} = PoP$ . These transitions based on a single time-step memory show considerable improvement over

- (a) *chance* *PoP*-forecasts with equal transition probabilities,  $P_{ij} = 1/n$  for all  $i, j$ ;
- (b) *climate* *PoP*-forecasts (zero-order Markov chain) which are represented by the equilibrium transition matrix,  $P_{ij} = C_i$ , whose column entries are climate state probabilities emerging from the transition probability matrix, and
- (c) *persistence* *PoP*-forecasts whose transition probabilities are defined by the identity matrix,  $P_{ij} = I$ :

Markov chain	$p_i = P_{ij}p_j$ with $p_{i=3} = PoP$ ;
integral time-scale	$\tau_i = \sum_{r=0}^{r=N-1} P_{ii}^{r-1}(1 - P_{ii})$ of state $i$ ;
regression model	$PoP(i, r) = a(i, r) + \sum_{k=1}^3 b_k(i, r)X_k$ .

Finally, this basic three state first-order Markov chain can be extended by linearly incorporating covariates of other relevant weather variables observed at the single station and another nearby upstream location. The lead time

$r$  is the number of hours ahead; the covariates  $X_k$  for  $k = 1, 2, \dots$  are the cloud cover, pressure, temperature, low clouds and zonal wind, introduced through the empirical regression coefficients  $a, b_k$ . Including covariates from neighbouring upstream stations improves the model performance considerably. A simplified version of the *PoP*-model is conveniently drafted for Hamburg (and other stations) as a probability versus surface pressure diagram given three basic Markov weather states, cloud-free or cumulus, stratus and rain (Figure 3.2).



**Figure 3.2** Single station short-term forecast of probability of precipitation in Hamburg (12-hour *PoP*-prediction) for winter: Graphic display of the rainfall Markov chain (with surface pressure as the only covariate).

### 3.2 Forecast verification

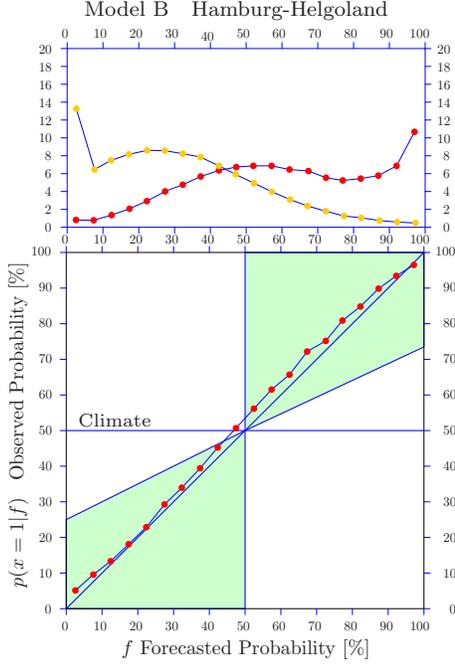
The relationship between the forecasts and observations can be conveniently analysed by two methods which rely on the joint forecast and verification distributions:

- (a) The performance measures of specific aspects of the relationship between  $F$  and  $X$ .
- (b) The basic joint, conditional, and marginal distributions themselves and their summary measures as, for example, means, variances etc.

*Performance* of probabilistic forecasts,  $F = PoP$ , is measured by the commonly used Brier score  $B = \langle (F - X)^2 \rangle$  with  $PoP$  lying between 0 and 1, and the verification  $X$  is either 0 or 1; smaller Brier scores correspond to better forecasts. The following values taken from independent Hamburg *PoP*-forecasts show how forecasts improve with better models: persistence ( $B \sim 0.29$ ), climate (0.25), simple Markov-chain without covariates (0.19), with pressure only

(0.17), with single station covariates (0.15) and with covariates of an upstream station (0.14).

*Joint forecast-verification distributions* provide more details on the forecast-verification couple  $(F, X)$  as described in the reliability and sharpness diagrams (Figure 3.3):



**Figure 3.3** Verification of Hamburg *PoP*-forecasts (with all co-variates including the upstream station Helgoland): Sharpness (top) and reliability (bottom) diagram.

First, the reliability diagram relates predicted rainfall probabilities, *PoP*, to the verifying observed relative frequencies of rainfall occurrence. A diagonal implies the ideal reliability of the model which can be achieved by *PoP*-forecasts (presented in 5% *PoP*-intervals); the vertical and horizontal lines corresponds to climate mean probability forecasts and the half-angle area including the ideal diagonal describes bounds for the real *PoP*-forecast reliability. Finally, as this result may also be achieved by forecasts mostly issued in the meaningless  $PoP \sim 50\%$  neighbourhood, the conditional forecast-verification distributions  $p(F|X)$  need to be analysed: The sharpness diagram associates *PoP*-forecasts (abscissa) with the verifying occurrence (ordinate) of rain and no-rain (yellow, red circles).

A model that blends Hamburg data with an upstream station, yields *PoP* forecasts, whose performance shows the required separation between large and small *PoP*-values. *Combining probability forecasts* in an error-minimising fashion is a particularly simple low-cost method for forecast improvement (Fraedrich and Leslie 1987), for example, combining single station *PoP* and categorical NWP rainfall forecasts:

$$\text{combination} \quad PoP^* = \alpha F + (1-\alpha)G.$$

The  $\alpha$ -weight describes the correlation between the two probabilistic forecast schemes,  $F$  and  $G$ , and the verification  $X$ :

$$\alpha = \{ \langle XF \rangle \langle XG \rangle + \langle G^2 \rangle \langle FG \rangle \} / \{ \langle F^2 \rangle + \langle G^2 \rangle - 2 \langle FG \rangle \}.$$

It is derived by minimizing the Brier-score of the combination, that is  $dB/d\alpha = 0$ . For example, the (unlikely) combination of two unbiased and uncorrelated binary forecasts of the same performance as obtained by climate mean predictions yields a 50% improvement of skill. In the practical forecast environment, a Markov-NWP combination is particularly useful; for Melbourne, percent correct NWP rainfall forecasts improved by 5% from 83 to 88%, at almost zero cost.

#### 4 Persistence in red noise: Conceptual forecast experiments

Forecast experiments are performed for a (stochastic) surrogate atmosphere representing the imperfect model environment; that is, individual and ensemble persistence forecasts in a red noise atmosphere are analysed analytically (see Fraedrich and Ziehmann 1994, 1995). These experiments show, in a qualitative sense, similarities with external (or practical) NWP predictions.

##### 4.1 Persistence and red noise

Persistence forecast experiments ('the weather remains in its present state') in a red noise environment are analysed analytically. The surrogate atmosphere is red noise generated by a first order auto-regressive Gaussian process, AR(1), which includes stochastic forcing  $z_i$  added with each time step  $\Delta$ . The process is discrete in time and continuous in the state variable  $X$ , represented by fluctuations,  $X(t) = \langle X \rangle + X'(t)$ , about zero mean  $\langle X \rangle = 0$ . Time (sample) averaging is denoted by the brackets,  $\langle \rangle$ , and the prime  $X' = X$  describes the anomalies. The stochastic AR(1)-process is the discrete-time analogue of the continuous-time Langevin equation with short-term memory. Thus, it is well suited as a surrogate atmospheric variable and a null-hypothesis in significance testing for atmospheric flow phenomena.

Time integration of the auto-correlation,  $C(r)/\sigma^2 = a^r$ , defines the integral time scale,  $\tau = 1/(1-a)$  or  $1/\tau = 1-a \sim \ln(1/a)$ , as a measure of the life time of a perturbation. *White noise* ( $a = 0$ ) forcing  $z_i$  with zero mean  $\langle z_i \rangle = 0$  is defined by the auto-covariance,  $\langle z_i z_j \rangle = q\delta(t_i - t_j)$  vanishing for  $i \neq j$ . The related white noise spectrum  $S_z = \sigma_z^2 \Delta$  can be defined by the unit time step,  $\Delta$ , so that  $q = \sigma_z^2$ . The variance of the response,  $\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle$ , is related to the noise or random forcing intensity  $\sigma_z^2 = \sigma^2(1-a^2)$ , which is  $1-a^2$  of the total variability. The deterministic part of the fluctuations contributes by  $a^2 \neq 0$ , with the signal to noise ratio of  $a^2/(1-a^2)$ .

*Red noise* ( $0 < a < 1$ ): This first order auto-regressive process is one of the simplest non-trivial processes simulating many observed aspects of the variability in the atmosphere. For  $\omega > 1/\tau$ , the spectrum  $S$  drops by a  $\omega^{-2}$  power-law. A large (small) red noise parameter,  $a$ , describes weather regimes with large (small) integral time scales,  $\tau$ , which are associated with small (large) intensities of the white stochastic forcing spectrum  $S_z^2$ . For sufficiently low frequencies  $\omega < 1/\tau$ , the response spectrum flattens to white noise. Note that the *random*

walk ( $a = 1$ ) commencing at  $X(t) = 0$ , reaches  $X(t + r) = \sum_i z_i$ ,  $i = 1, \dots, r$ , after  $r$  time steps. It is non-stationary as its variance grows linearly with time  $\langle X^2(t + r) \rangle = \langle \sum_i z_i^2 \rangle = r\sigma_z^2$ , because  $\langle X(t + r) \rangle = r \langle z_i \rangle = 0$ .

Table 4.1 Red noise atmosphere: AR(1)-process

autoregressive process	$X(t) = aX(t-1) + z_t = a^r X(t-r) + \sum a_i^{r-1} a^i z_{t-i}$
auto-covariance	$C(r) = \langle X(t)X(t+r) \rangle = \sigma^2 a^r$
spectrum	$S(\omega) = S_z^2 \{1 + a^2 - 2a \cos(\omega)\}$
integral time-scale	$\tau = \sum_{r=0}^{\infty} C(r)/\sigma^2 = 1/(1-a)$

**Chance, climate, and persistence:** These forecasts serve as reference predictions  $F(r)$  for the lead time  $r$ . They are commonly evaluated by the mean squared (ms) forecast error  $E = \langle (F - X)^2 \rangle$  sample averaged  $\langle \rangle$  over the forecast experiments; anomaly correlations are another measure of accuracy frequently used by numerical meteorological centres. *Chance* forecasts,  $F_R$ , select initial values at random; the *ms*-error corresponds to the squared distance between independent weather states, which is twice the variance of the system,  $2\sigma^2$ . *Climate* is predicted,  $F_C$ , by the climate mean; the *ms*-error corresponds to the system's variance and defines an accuracy threshold to obtain a predictability limit  $T$ . *Persistence* ( $F_P$ ) is a fundamental reference forecast, because only forecasts better than persistence have skill in the forecast of the time derivative, which is the goal of prediction. Persistence,  $F_P(r)$ , predicts future weather states,  $X(t)$ , by the initially observed state  $X(t_0)$  commencing with the observation at the time  $t_0 = t - r$ :  $F_P(r) = X(t_0) = X(t - r)$ . Its error,  $X(t) - X(t - r)$ , is evaluated after the lead time  $r$ . Forecast  $F(r)$  and verification  $X(t)$  are analysed as a pair of trajectories evolving in state space, whose squared Euclidean distance defines the squared error  $e^2 = (X - F)^2$ , averaged over all verification pairs  $\langle e^2 \rangle$ .

Table 4.2 Chance, climate, and persistence forecasts  $F(r)$  for lead time  $r$  and their mean squared (*ms*) forecast errors  $E = \langle (F - X)^2 \rangle$  in the red-noise atmosphere ( $\langle \rangle$  forecast climatology).

model	forecast	ms error	comment
chance	$F_R(r) = X_i$	$E_R(r) = 2\sigma^2$	independent states reference forecast
climate mean	$F_C(r) = \langle X \rangle$	$E_C(r) = \sigma^2$	
persistence	$F_P(r) = X(t - r)$	$E_P(r) = 2\sigma^2(1 - a^r)$	AR(1): $\alpha = a; \beta = 0$
combination	$F^*(r) = \alpha F_P + \beta F_C$	$E^*(r) = \sigma^2(1 - a^r)$	

**Predictability experiments** (random and systematic errors): Mean squared errors, which grow with increasing lead time  $r$ , show two domains that deserve particular analysis. The *short term memory* affects initial errors, error growth-rates and exit-times. Vanishing initial error,  $E(r = 0) = 0$ , characterises persistence forecasting. A limit of predictability is reached at a lead time when

the prediction error exceeds that of a reference forecast, which is conveniently defined by the error of a forecast by the climate mean,  $\sigma^2$ . Predictions at lead time  $r > T$  have passed this predictability limit,  $T = \ln(2)/\ln(a)$  (setting  $E_P(r = T) = \sigma^2$ ). As  $T$  is proportional to the integral time scale  $\tau$  of weather regimes,  $\ln(1/a) - 1/\tau$ , it follows that the effective time span for viable weather forecasts is limited by the life span of its most energetic phenomenon. Finally, saturation is reached at large lead times when forecast and verification become independent, their correlation vanishes and the mean squared error  $E(r \rightarrow \infty) = 2\sigma^2$  reaches the level of the mean squared distance between two randomly chosen weather states (that is, a chance forecast).

In an imperfect model environment like this, the mean squared error is separable into a systematic,  $SE$ , and non-systematic or random component,  $RE$ . Given the initial anomaly  $X_0$ , its persistence forecast,  $F(r) = X_0$ , up to lead time  $r$ , needs to be compared with the proper verification time series, which has commenced at this initial anomaly, that is the  $AR(1)$  time series  $X(t) = a^r X(t-r) + \sum_i i = 0^{r-1} a^i z_{r-i}$ , commencing from the  $r$ -th step backward. Now, averaging the *ms*-error over a sample of the forecast-verification pairs,  $i\bar{c}$ , conditional on a fixed initial anomaly,  $X_0$ , yields the conditional forecast error,  $E(r|X_0) = X_0^2(1 - a^r)^2 + \sigma^2(1 - a^{2r})$  (see Figure 4.1a, b). Averaging over all squared conditional anomalies  $X_0$  leads to the unconditional error budget and to the distinction between the forecast error's systematic and non-systematic or random components,  $SE = \sigma_0^2(1 - a^r)^2$  and  $RE = \sigma^2(1 - a^{2r})$ , which combine to  $E = SE + RE$  (Figure 4.1c). The systematic error is smaller than the random error and both approach unity for infinitely large lead times or  $\sigma^2$ . At the limit of predictability  $r = T$ , the systematic (non-systematic) error attains 1/4 (3/4) of the climate variance. The initial error growth rate vanishes for systematic errors but is finite for the random part.

Persistence forecasts averaged over the same initial anomalies  $X_0$  show the systematic error increasing with the distance of the initial condition from the climate mean, for which forecasts are expected to be better ( $SE = 0$ ).

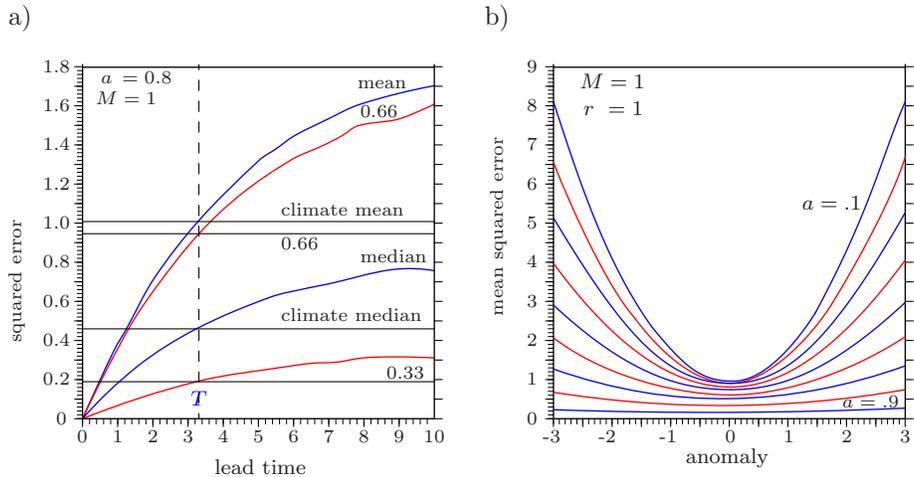
Table 4.3 Error statistics of imperfect model experiments: Forecasts  $F(r)$  and verification  $X$

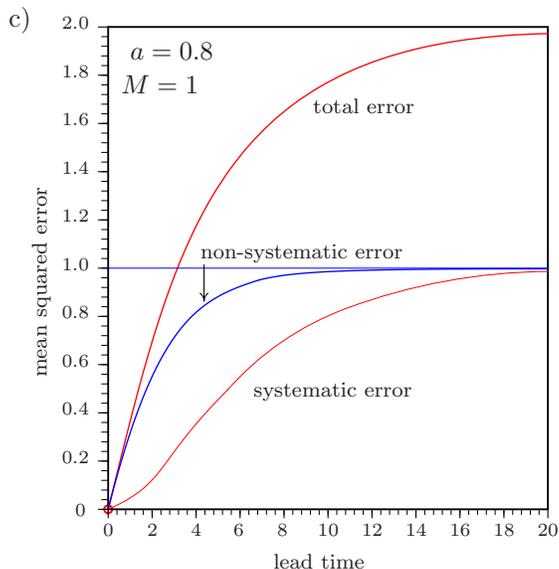
<i>error statistics</i>	<i>definitions</i>	<i>persistence in red noise</i>
error	$e = X - F$	$e(t, r) = X(t) - X(t - r)$
mean squared error	$E(r) = \langle e^2 \rangle$	$E(r) = 2\sigma^2(1 - a^r)$
conditional	$E(r X_0)$	$\langle X_0^2 \rangle (1 - a^r)^2 + \sigma^2(1 - a^{2r})$
systematic	$SE = \langle (\langle F \rangle - \langle X \rangle)^2 \rangle$	$\sigma^2(1 - a^r)^2$
non-systematic	$RE = E - SE$	$\sigma^2(1 - a^{2r})$
error growth law	$E_r = f(E, r)$	$E(\infty)\{1 - E/E(\infty)\}/\tau$
predictability limit	$E(r = T) = \sigma^2$	$T = \ln(2)/\ln(1/a)$
saturation error	$E(r = \infty)$	$2\sigma^2$

*Error growth:* Laws of error growth,  $E_r = f(E, r)$ , are similar to a Verhulst-type equation, where growth is confined by a quadratic saturation error feedback. Persistence in red noise confines the error growth rate by a linear term:  $E_r = 2\sigma^2 a^r \ln(1/a)$ . Substituting  $a = (1 - E/\sigma^2)^{1/r}$  gives a law of error-growth,  $E_r \sim E(\infty)\{1 - E/E(\infty)\}/\tau$ , which may replace the generally used Verhulst equation of confined growth. Note that, not unexpected, the initial growth,  $E_r(r = 0) \sim 1/\tau$ , is large in processes with very short-term memory and the growth decreases with increasing error size  $E(r)$ . That is, the often misinterpreted generalization that large errors grow slower, holds only for predictability experiments in the same forecast environment ( $a = \text{const.}$ ) where error growth decreases when approaching saturation. If regimes change ( $0 < a < 1$ ), the conditioning climatology has to be included in the error growth analysis.

*Lower and upper bounds of predictability:* Persistence plus half-trend forecasts of the state  $X(t)$  have been successfully used in empirical seasonal forecasting:  $F_T(r) = X(t - r) + \frac{1}{2}\{X(t - r) - X(t - 2r)\}$ . The performance of this scheme shows also interesting similarities with NWP models: Smaller scales (that is, processes with shorter memory or smaller  $a$ ) tend to have larger initial errors,  $E(r = 0) = 2\sigma^2(1 + a)$ ; smaller initial errors  $E(r = 0)$  are associated with smaller initial error growth rates,  $E_r(r = 0) = 2E(r = 0)(1 - a)(1 + a)^{-1} \ln(1/a)$ . Both results lead to useful predictability estimates which are similar to results obtained from practical weather forecasts:

- (i) A *lower bound* of the limit of predictability  $T$ , which depends on the initial error ( $r = 0$ ), is obtained by a linear extrapolation to  $E(r = 0) = 0$ . In this sense the lower bound characterizes the forecast potential once analysis schemes improve and initial errors tend to zero.
- (ii) An *upper bound* may be reached when, in addition, the systematic forecast error can be reduced to zero.





**Figure 4.1** Error of persistence forecasts of a red noise atmosphere: (a) Time evolution of the squared error distribution: Mean, median, upper and lower terciles (0.66, 0.33) of persistence and climate mean predictions, (b) error at lead time  $r = 1$  depending on initial anomaly-conditions, and (c) systematic, non-systematic and total error.

## 4.2 Ensemble forecasts

Ensemble forecasts are introduced to estimate future probability distribution of the atmospheric states to

- (i) improve the forecast using an ensemble mean and
- (ii) estimate the forecast error by the ensemble spread or variance. Ensemble forecasts and forecast error statistics is analysed.

*Perfect model/ensemble hypothesis:* The perfect model/ensemble hypothesis provides the background for introducing an ensemble of individual forecasts,  $F_i(r)$  for  $i = 1, \dots, M$ , and its ensemble average,  $[F_i] = M^{-1}(\sum_{i=1}^M F_i)$ , to predict a field variable  $X$ . A *perfect ensemble* ( $r = 0$ ) consists of ensemble members  $F_i$ , which are chosen such that the mean distances between all members  $[d_{ij}^2] = M^{-1} \sum_i (\sum_j (F_i - F_j)^2 / (M - 1))$ , represents the analysis error  $[e_i^2] = [d_{ij}^2]$  (see similarity to analog forecasting, section 2). In a *perfect model* ( $r > 0$ ) the same holds: The mean squared distance between all members  $[d_{ij}(r)^2]$  corresponds to the mean squared error of all individual forecasts,  $[e_i(r)^2] = [(i(r) - X)^2] = [d_{ij}(r)^2]$ . Thus the following results can be deduced (see Table 2.5):

- (i) The mean-squared error of the ensemble average forecast is about half of the mean of the squared errors of the individual forecasts:  $E_M = 2[e_i^2](M+1)/M$ .
- (ii) There is a linear relation between the ensemble spread  $S_M$  and the error of the ensemble mean,  $E_M = S_M(M+1)/(M-1)$ , because  $[d_{ij}^2] = 2S_M M/(M-1)$ .

Table 4.4 Perfect model/ensemble hypothesis: ensemble and error statistics

<i>ensemble statistics</i>	$[(F_i - X)^2] = ([F_i] - X)^2 + [(F_i - [F_i])^2]$
ensemble mean	$F_M = [F_i] = [F(r+i)] = M^{-1} \sum_{i=0}^{M-1} F(r+i)$
spread ( $X = 0$ )	$S_M = [(F_i - [F_i])^2] = \sigma^2 \{1 - (1+a)M^{-1}(1-a)^{-1} + 2a(1-a^M)M^2(1-a^{-2})\}$
distance ( $X = F_j$ )	$[d_{ij}^2] = [(F_i - F_j)^2]$
ensemble mean and	$E_M = [(F_M - X)^2] = [e_i^2] - S_M$ $= \sigma^2 \{1 + (1+a - 2a^r(1-a^M))M^{-1}(1-a)^{-1} - 2a(1-a^M)M^{-2}(1-a)^{-2}\}$
systematic error	$SE_M = \sigma^2 \{a^r - (1-a^M)M^{-1}(1-a)^{-1}\}$

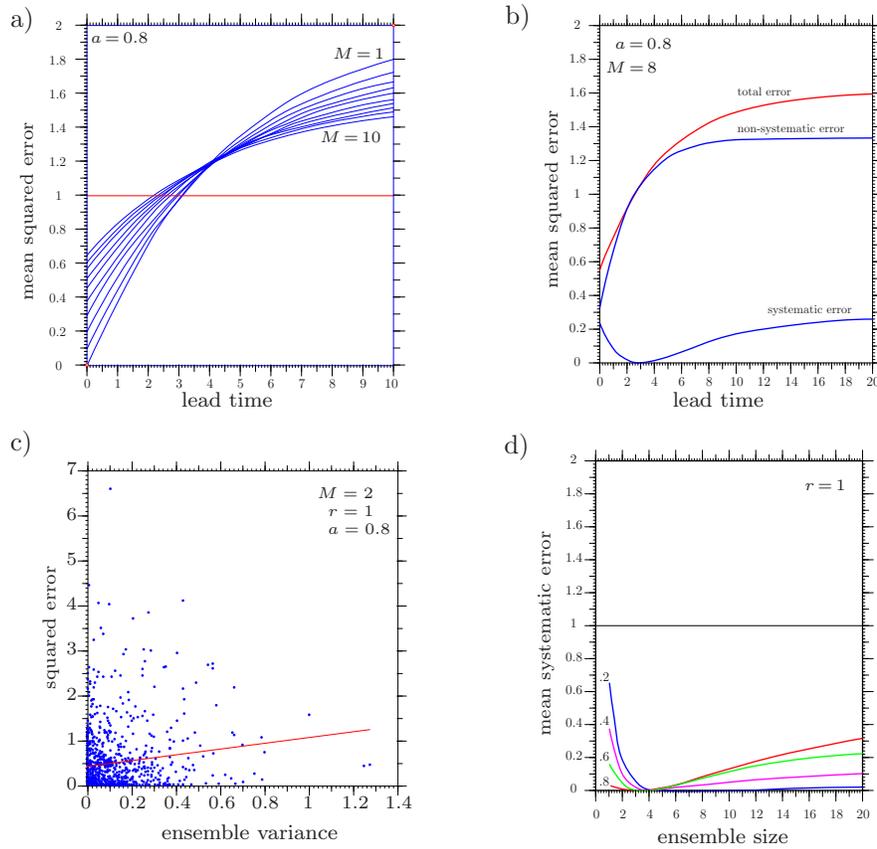
*Imperfect model/ensemble (lagged-average forecasts):* Practical forecasts are almost always made in an imperfect model/ensemble environment. Here it is simulated by an ensemble mean of  $M$  lagged persistence forecasts of the red noise atmosphere,

$$[F_i(r)] = [F(r+i)] = M^{-1} \sum_{i=0}^{M-1} F(r+i), \text{ where } F(r+i) = X(t - (r+i)).$$

The error and ensemble statistics can be deduced analytically: The predictability statistics include mean and spread (or distances) of the ensemble members and the conditional (systematic, non-systematic) errors made by ensemble mean forecasts. The results are summarised (Figure 4.2):

The ensemble mean and individual forecast errors,  $E_M(r)$  and  $E(r)$ , show that the lagged-average ensembles of persistence forecasts are, in general, worse than the latest individual forecast (before the predictability limit  $T$ , Table 4.3, is reached).

The ensemble spread  $S_M$  is independent of the lead time  $r$  but changes with ensemble size  $M$ . Furthermore, there is no direct error-spread correlation as suggested by the perfect model/ensemble case. Instead, the following is noted. Depending on lead time  $r$  and red noise memory  $\tau$  (or  $a$ ), there is an optimal ensemble size ( $M = 8$ ), whose members provide an optimal error-spread ( $E_M$ ,  $S_M$ )-correlation (0.31); the mean generates minimum systematic error ( $r = 1$ ,  $\tau = 5$ ,  $a = 0.8$ ).



**Figure 4.2** Lag-averaged persistence ensemble mean forecasts in a red noise atmosphere (autocorrelation  $a = 0.8$ ):

- (a) Error changing with lead time and ensemble size,
- (b) systematic, non-systematic and total error for  $M = 8$  ensemble members,
- (c) scatter diagram of error versus ensemble spread,
- (d) systematic error changing with ensemble size  $M$  (lead time  $r = 1$ ).

## 5 Some concluding remarks

Weather is non-linear and noisy and, therefore, provides a natural area for scientists studying dynamics and statistics alike. It is not surprising that the work of some explorers in these fields, for example, G. Walker, L.F. Richardson, and E.N. Lorenz, has stimulated both fundamental and applied research even outside the meteorological community. Here we illustrate forecasting research by some examples to characterize weather and climate related key words like predictability, probability and persistence.

**Predictability:** Predictions of the first kind link forecast research with the dynamical framework. Methods to estimate dimension and entropy from time series are used to demonstrate analog ensemble forecasts, to which metric adaptation and ensemble size are added to minimize the ensemble mean forecast error. Ensemble spread and forecast error prediction are briefly discussed.

**Probability:** Single station probability of precipitation is introduced to demonstrate a classical linear prediction model for both weather state and forecast skill, and its performance evaluation, which is an essential though frequently neglected subject. Tropical cyclone track forecasting is a multivariate application of both a linear and a non-linear approach combining numerical and empirical techniques; the linear part is an error minimising combination, the non-linear part is an analog-scheme demonstrating substantial forecast improvements (including error prediction) when past forecast errors are incorporated as variables.

**Persistence:** A toy model ‘persistence in red noise’ describes predictability analysis analytically in the practical forecast environment where systematic model errors need to be considered. Many aspects of predictability analysis are covered ranging from ensemble prediction to various forecasts limits. A discussion on techniques based on optimal growth and breeding of initial errors is postponed to later lectures, for which a suitable toy model needs to be developed.

In summarizing, with predictability as leitmotif the lecture presents closely related subjects: Short term memory, analog ensemble forecasting and nearest neighbour phase space statistics, dimension and entropy of dynamical systems; probabilistic, ensemble and forecast skill predictions; and an idealised conceptual model for practical weather forecasting techniques in terms of time series diagnostics.

**Acknowledgement:** These lecture notes update presentations at the European Geophysical Union (2006), Chaos and Geophysical Flows (2003), and the Enrico Fermi School (1998). Parts of chapter 2.2 about tropical cyclone track forecasting are taken from the diploma thesis of Hans Langmack. Gerhard Tischel’s professional and patient persistence in editing the manuscript and, in particular, the figures, together with Frank Sielmann’s support are gratefully acknowledged.

## References

- Abarbanel, H.D.I., R. Brown, J.H. Sidorowich and L.S. Tsimring, 1993: The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.* **65**, 1331-1392.
- Charney, J.G., R. Fjortoft and J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237-254.
- Cooke, W.E., 1906: Forecasts and verifications in Western Australia. *Mon. Wea. Rev.* **34**, 23-24.

Klaus Fraedrich

- Fraedrich, K., 1986: Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.* **43**, 419-432.
- Fraedrich, K. and L.M. Leslie, 1987: Evaluation of techniques for the operational, single station, short term forecasting of rainfall at a mid-latitude station (Melbourne). *Mon. Weather Rev.* **115**, 1645-1654.
- Fraedrich, K. and L. M. Leslie, 1987: Combining predictive schemes in short-term forecasting. *Mon. Wea. Rev.*, **115**, 1640-1644.
- Fraedrich, K. and L.M. Leslie, 1989: Estimates of cyclone track predictability, I: Tropical cyclones in the Australian region. *Q.J.R. Meteorol. Soc.*, **115**, 79-92.
- Fraedrich, K., 1988: El Nino/Southern Oscillation predictability. *Mon. Wea. Rev.*, **116**, 1001-1012.
- Fraedrich, K. and L.M. Leslie, 1991: Predictability studies of the Antarctic atmosphere using both analogue and chaos theory. *Austral. Meteorol. Mag.* **39**, 1-9.
- Fraedrich, K. and R. Wang, 1993: Estimating the correlation dimension of an attractor from noisy and small datasets based on re-embedding. *Physica* **D65**, 373-398.
- Fraedrich, K. and C. Ziehmann, 1994: Predictability experiments of persistence forecasts in a red noise atmosphere. *Q. J. R. Meteorol. Soc.* **120**, 387-428.
- Fraedrich, K. and C. Ziehmann, 1995: Praktische Vorhersagbarkeit: Persistenz in rotem Rauschen. *Meteorol. Zeitschrift* **NF4**, 139-149.
- Fraedrich, K. and B. Rueckert, 1998: Metric adaption for analog forecasting. *Physica* **A253**, 379393.
- Fraedrich, K., C.C. Raible and F. Sielmann, 2003: Analog ensemble forecasts of tropical cyclone tracks in the Australian region. *Weather and Forecasting*, **18**, 3-11.
- Gabriel, K.R. and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90-95
- Grassberger, P. and I. Procaccia, 1984: Dimensions and entropies of strange attractors from a fluctuating dynamics approach. *Physica* **D13**, 34-54.
- Götz, G., 1995: Predictability of non-linear dynamical systems. *Időjárás (Quart. J. Hungarian Meteor. Service)* **99**, 1-32.
- Kalnay, E. and R. Livezey, 1983: Weather predictability beyond a week: An introductory review. From: *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*. International School of Physics 'Enrico Fermi' Course **88**, 311-346.
- Kirk, E. and K. Fraedrich, 1998: Probability of precipitation: Short-term forecasting and verification. *Contrib. Atmos. Phys.*, **71**, 263-271.
- Langmack, H., F. Sielmann and K. Fraedrich, 2007: Analog-Vorhersagen von Hurrikan-Zugbahnen. *Promet*, **33**, 65-70.

- Leslie, L.M., G.F. Holland, M. Glover and F.J. Woodcock, 1990: A climatological-persistence (CLIPER) scheme for predicting Australian region tropical cyclone tracks. *Aust. Met. Mag.*, **38**, 87-92.
- Leslie, L.M. and K. Fraedrich, 1990: Reduction of tropical cyclone position errors using an optimal combination of independent forecasts. *Weather and Forecasting*, **5**, 158-161.
- Lorenz, E.N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130-141.
- Lorenz, E.N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.* **26**, 636-646.
- Lorenz, E.N., 1991: Dimension of weather and climate attractors. *Nature* **353**, 241-244.
- McNames, J., 2002: Local averaging optimization for chaotic time series prediction. *Neurocomputing* **48**, 279-29.
- Murphy, A.H. and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115** (1987), 1330-1338.
- Murphy, A.H. and R.L. Winkler, 1992: Diagnostic verification of probability forecasts. *International Journal of Forecasting*, **7**, 435-455.
- Popper, K.R., 1972: *Of Clocks and Clouds*. From: *Objective Knowledge*. Oxford, At the Clarendon Press, 206-255.
- Richardson, L.F., 1922: *Weather Prediction by Numerical Process*. Cambridge University Press, pp. 236.
- Richardson, L.F., 1926: Atmospheric diffusion on a distance neighbour graph. *Proc. Roy. Soc.*, London, **A110**, 709-737.
- Shirer, H.N., C.J. Fosome, R. Wells and L. Suci, 1995: Estimating the correlation dimension of atmospheric time series. *J. Atmos. Sci.* **54**, 211-229.
- Tennekes, H., 1991: Karl Popper and the accountability of numerical forecasting. In *New Developments in Predictability*, ECMWF Workshop proceedings, Reading, UK.
- Thompson, P.D., 1977: How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.*, **114**, 228-229.
- Tsonis, A.A. and J.B. Elsner, 1989: Chaos, strange attractors and weather. *Bull. Amer. Met. Soc.* **70**, 14-23.
- van den Dool, H.M., 1994: Searching for analogs: How long must we wait? *Tellus* **46A**, 314-324.
- Wolf, A., J. Swift, H.L. Swinney and J.A. Vassano, 1985: Determining Lyapunov exponents from time series. *Physica* **16D**, 285-317.

Eingegangen am 31.05.2007

Klaus Fraedrich  
Meteorologisches Institut  
Universität Hamburg

