Statistical Single-Station Short-Term Forecasting of Temperature and Probability of Precipitation: Area Interpolation and NWP Combination

CHRISTOPH C. RAIBLE, GEORG BISCHOF, KLAUS FRAEDRICH, AND EDILBERT KIRK

Meteorologisches Institut, Universität Hamburg, Hamburg, Germany

ABSTRACT

Two statistical single-station short-term forecast schemes are introduced and applied to real-time weather prediction. A multiple regression model (R model) predicting the temperature anomaly and a multiple regression Markov model (M model) forecasting the probability of precipitation are shown. The following forecast experiments conducted for central European weather stations are analyzed: (a) The single-station performance of the statistical models, (b) a linear error minimizing combination of independent forecasts of numerical weather prediction and statistical models, and (c) the forecast representation for a region deduced by applying a suitable interpolation technique. This leads to an operational weather forecasting system for the temperature anomaly and the probability of precipitation; the statistical techniques demonstrated provide a potential for future applications in operational weather forecasts.

1. Introduction

Although, in general, numerical weather prediction (NWP) models are hard to beat in very short-term forecasting (up to 24 h), they do require a substantial amount of computation time and the model forecasts are not always stable at this timescale. In contrast, statistical schemes require little computation time to make a forecast and are adapted to the station's climate but, in general, do not include nonlinear behavior. Another important advantage of statistical methods over NWP models is that the latter often produces biased forecasts, whereas the former are usually unbiased.

Due to some of these obvious advantages, statistical methods have a long application history (Murphy 1998). This paper builds on a series of studies conducted in the past, starting with a Markov-chain model predicting daily rainfall for a single station (Gabriel and Neumann 1962). Fraedrich and Müller (1983) extended this model type to predict sunshine periods and probability of precipitation (PoP). This and related studies led to the introduction of these techniques in weather services (Miller 1984; Miller and Leslie 1984; Hess et al. 1989). Besides probability of precipitation and daily sunshine, the temperature plays obviously an important role in practical weather forecasting. Therefore, many have applied statistical methods to 6–24-h temperature forecasting

(Wilson 1985; Dallavalle 1996; Knüpffer 1996) as well as the long-term prediction (more than a month ahead) of monthly mean temperature (Nicholls 1980; Navato 1981; Madden 1981; Norton 1985). Thus, the first purpose of this paper is to set up a statistical model of short-term forecasting of the temperature anomaly and the PoP, and to verify these forecasts. Motivated by the forecast improvements achieved by combining independent short-term prediction schemes (Fraedrich and Leslie 1987a), a second purpose of this paper is to show the skill gained in linearly combining these statistical forecasts with the NWP model forecasts, the latter of which is the direct model output (DMO) of the Europe model (EM) of the German Weather Service (DWD). This offers an alternative to the widely used model output statistics (Glahn and Lowry 1972). Finally, the third aim is to provide and illustrate the performance of an operational statistical weather forecasting system that, associated with an interpolation method (Smith and Wessel 1990), leads to areal short-term predictions.

The outline of the paper is as follows: After a brief description of the database (section 2) the statistical models are introduced (section 3). The focus lies on the multiple regression technique for forecasting temperature anomalies and a brief review is given for the multiple regression Markov model to predict the probability of precipitation. These two statistical models are applied to single weather stations and results are presented for Hamburg Fuhlsbüttel as an example. In section 4, these forecasts are combined with the NWP model forecasts of the EM to improve the forecast accuracy. In section 5 the two statistical methods are applied to several

Corresponding author adress: Klaus Fraedrich, Meteorologisches Institut, Universität Hamburg, Bundesstr. 55, D-20146 Hamburg, Germany.
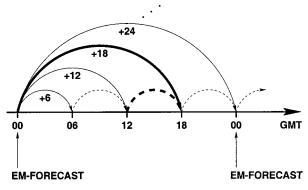E-mail: fraedrich@dkrz.de

FIG. 1. Construction of the NWP time series; the bold arrows refer to the example described in the text.

FIG. 2. Mean annual temperature cycle for the main observation times at Hamburg Fuhlsbüttel.

weather stations of central Europe to obtain an areal forecast for this region. Finally, section 6 summarizes the results and highlights the important features.

2. Data

The observed data (time series) of 68 DWD stations are used for model building and forecast experiments. These observed data contain the standard synoptic parameters at 6-hourly intervals with sample lengths varying between 8 and 46 yr. The longest time series data is available for Hamburg Fuhlsbüttel and is used to present and verify the different statistical models for temperature anomalies and probability of precipitation, and for the model combination. The combination is based on numerical weather predictions utilizing forecasts of the EM (Majewski 1991). This time series of model data starts in 1991 and ends in 1995. It contains, among other parameters, the 2-m temperature to the nearest $1/10°C$ and four different types of precipitation, convective rain, convective snow, large-scale rain, and large-scale snow, to the nearest $1/10$-mm. Therefore, the NWP probability of precipitation is defined as either 0 or 1, if any one of these four parameters has a value greater than 0 mm. The EM runs daily at 0000 UTC. Each run generates forecasts at 6-h intervals for projections of 6–48 h. To obtain forecasts for the missing generating times 0600, 1200, and 1800 UTC we use the forecasts of the corresponding 0000 UTC EM run. For example, the 6-h prediction at the 1200 UTC start time uses the 18-h prediction of the 0000 UTC run (see Fig. 1).

Strictly, the modified NWP time series cannot be compared with time series of forecasts of other models. This obvious disadvantage is outweighted by the advantage of updating the forecasts with new observations.

3. Single-station models

Statistical single-station forecast models are developed for both the temperature and the probability of precipitation. A multiple regression technique is used
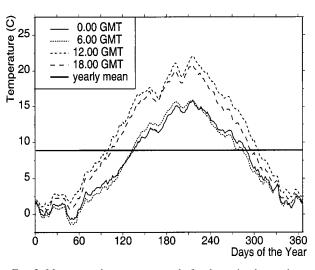
for the prediction of temperature anomalies (R model). This technique, in combination with a Markov process, is used to predict PoP (M model). After a description of the two models, the performance is presented using the weather station Hamburg Fuhlsbüttel as an example. A particular emphasis is taken on the joint (probability) distributions of forecasts and observations (Murphy and Winkler 1987). The denoted marginal and conditional distributions are used "to identify the strengths and weaknesses in the forecasts to provide the modeler and forecaster with feedback and a basis for improving the quality of such predictions" (Murphy et al. 1989). Furthermore, the skill of both models is analyzed with reference to simple standard forecast models, like the climate mean and persistence.

a. Temperature forecast: Multiple regression model and its verification

A multiple regression method (backward elimination) is used to build the forecast model for temperature anomalies. The anomalies are defined as deviations from a mean averaged over all years ($j = 1, 2, \ldots, J$) for a given calendar day ($i = 1, 2, \ldots, 366$) and each observation time ($k = 0000, 0600, 1200, 1800$ UTC) separately. Thus, for a certain day i of a year j at the main observation time k one obtains the temperature anomaly T_{Aijk}:

$$T_{Aijk} = T_{ijk} - \overline{T}_{ijk} = T_{ijk} - \frac{1}{J} \sum_{j=1}^{J} T_{ijk}, \quad (1)$$

where T_{ijk} is the observed temperature. To smooth these four annual cycles (Fig. 2) associated with the four main observation time k, a 7-day running mean is applied.

1) MULTIPLE REGRESSION MODEL

A classic multiple regression scheme (Draper and Smith 1981; Perrone and Miller 1985; Wilks 1995; von Storch and Zwiers 1998) is applied to each of the four temperature anomalies and to the whole year (annual model). In addition, it is possible to fit the model for each individual season (winter is defined as December–February; spring as March–May; summer as June–August; and fall as September–November); this will only be used for the application of the areal forecast (section 5). Furthermore, the time series data is divided into a model building part (1949–79) and an independent verification part (say 1980 to 1995 for Hamburg Fuhlsbüttel). The standard routine synoptic observation provides 20 parameters for applying the regression scheme. For practical use the number of covariates (or predictors) is reduced as follows: First, the comprehensive version of the model is built with all 20 covariates. Second, the covariates are sorted with respect to their importance [see Eq. (2)] using only the 0000 UTC start time with the 12-h projection. Third, models are reconstructed by successively reducing the number of the less important covariates. Fourth, the final model version is obtained when its rms errors (for all start and lead time combinations in the independent dataset) begin to show changes in the second digit (compared with the comprehensive model performance). Here it should be noted, that the ranking of the predictors [with respect to their importance, Eq. (2)] is different for each combination of start and lead times. Thus, a set of 12 regression coefficients are estimated for each start time (0000, 0600, 1200, and 1800 UTC) and for the 6-, 12-, 18-, and 24-h forecast projections. There are two types of covariates: (a) direct use of the synoptic data (e.g., cloud cover from the 0–9 synoptic code number as the predictor value) and (b) derived covariates. We use the dewpoint anomaly (treated exactly like the temperature) and the pressure anomaly from long-term annual mean value. The natural logarithm is applied to the covariate visibility due to the exponential characteristics (which is also shown in the use of the exponential code table). The meridional and zonal wind is derived from the wind speed and the wind direction.

Table 1 shows the 12 input covariates taken at the observation time 0000 UTC for the 12-h prediction of the annual Hamburg Fuhlsbüttel model. The covariates ($i = 1, \ldots, 12$) are sorted with respect to their importance I_i, which is defined as

$$I_i = \sum_{t=1}^{T} (a_i X_{it})^2, \qquad (2)$$

where the a_i are the corresponding regression coefficients and the X_{it} are the values of the covariates and all cases $t = 1, \ldots, T$ in the independent forecast sample. Additionally, Table 1 shows the corresponding regression coefficients for the 0000 UTC observation time and the 12-h forecast. The temperature is the most important covariate representing the persistence of the system. If a positive anomaly is observed, the positive regression coefficient causes a positive contribution for the forecast. The other covariates represent the climatological setting of the station. For example, the meridional wind component contributes negatively to the forecast. Two artificial covariates—the sine and cosine of the date—are used to test whether there is a remaining small contribution to the annual cycle, which appears not to be the case so that the main part of the annual cycle is excluded when utilizing temperature anomalies.

TABLE 1. The 12 covariates sorted with respect to their importance for the single-station temperature prediction; the regression coefficient refers to the 12-h forecast of the annual Hamburg Fuhlsbüttel model.

No.	Type of covariate	Regression coefficient
1	Temperature anomaly (1/10°C)	0.612
2	Cloud cover (0–9 synop code)	1.856
3	Ceiling (0–9 synop code)	−1.063
4	Zonal wind (Kt)	−0.188
5	Type of middle clouds (0–9 synop code)	−1.306
6	Type of low clouds (0–9 synop code)	−0.921
7	Dewpoint anomaly (1/10°C)	−0.352
8	Pressure (DPa)	0.579
9	Meridional wind (Kt)	0.282
10	Cosine of the date	2.426
11	LOG [Visibility (m) + 1]	−0.605
12	Sine of the date	0.464

2) FORECAST VERIFICATION

The basis of forecast verification for the independent sample 1980–95 is the joint distribution of the forecasts and the observations, from which the traditional summary performance measures can be deduced.

In Fig. 3 the bivariate histogram for the 12-h predictions versus observations is presented for the annual R model together with the corresponding scatter diagram using 1°C class intervals. The scatter about the diagonal gives a qualitative impression of the model reliability. Note that even for extreme anomaly conditions the scatter does not increase essentially and the joint distribution appears to be symmetric about and along the diagonal. We recognize that the dispersion of the errors is nearly Gaussian around the climate mean state of zero temperature anomaly indicating minimal or no forecast errors. The largest joint relative frequencies occur in the domain where the error between observation and forecast is below 1°C. Thus, all off-diagonal frequencies of the joint distribution correspond to forecasts with errors larger than 1°C. Adding all relative frequencies of the error interval ±2°C, we obtain 80% of the 6-h forecasts (not shown) and 70% of the 12-h forecasts. The 18- and 24-h predictions (not displayed) show similar behavior with the difference being, that the joint distributions become wider and plainer due to deteriorating forecast accuracy.
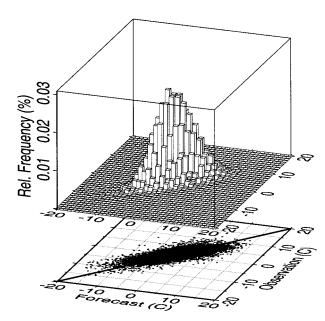
FIG. 3. The bivariate histogram of the joint distribution of the 12-h forecasts vs the observations and the corresponding scatter diagram for an annual model of Hamburg Fuhlsbüttel for the independent sample (1980–95).

As indicated above, traditional performance measures also play an important role in verification schemes. Moreover, they are summary measures of the joint distributions and can facilitate comparisons among prediction lead times and between different models. Table 2 shows the root-mean-square error (rms error) and the correlation coefficient between observations and forecasts. In this context the rms error is more important, because the correlation coefficient is more phase and less amplitude dependent. The R-model disposes of the highest accuracy for all lead times. The 6-h persistence and the first-order autoregressive or AR(1) process, which is a combination of an optimally weighted persistence and climate model plus white noise, show the same behavior. Only the climate mean forecast has an initial error due to the concept of this model. The performance of the 6-h persistence model indicates the system's tendency to remain in its initial state. The improvement of the AR(1) process being significantly lower compared to the R-model justifies the use of the more complicated R model.

TABLE 3. Definition of the classes for the weather states using codes of the World Meteorological Organization.

		First-order Markov chain with 3 classes		
Class		WW	C_L	W1
Cumulus	0	00–16, 18, 19–24, 28, and 30–49	0–4	0–4
Stratus	1	00–16, 18, 19–24, 28, and 30–49	5–9	0–4
Rain	2	17, 25–27, 29, and 50–99		5–9

b. PoP forecasts: Multiple regression Markov model and its verification

The PoP forecast technique is based on Markov-chain models (Gabriel and Neumann 1962; Fraedrich and Müller 1983) utilizing time series $X_t = i$, which can be defined for discrete states i (h or j) $= 1, \ldots, s$ occurring at discrete time steps $t = 1, \ldots, n$ (or m). In the theory of Markov chains, see, for example, Kemeny et al. (1976), the observed event $X_t = i$, which can be regarded as the outcome of a trial at time t, depends only on the outcome of the directly preceding trial $t - 1$. Thus, the outcome is not associated with a fixed probability, but with a pair corresponding to a conditional (or transition) probability (prob):

$$\text{prob}(X_t = j | X_1 = h, \ldots, X_{t-1} = i)$$
$$= \text{prob}(X_t = j | X_{t-1} = i) = p_{ij}. \quad (3)$$

The transition probabilities p_{ij} can be estimated by the maximum likelihood method (Anderson and Goodman 1958; Kemeny et al. 1976). Thus, the Markov-chain model consists of a transition probability matrix mapping an initial state probability vector linearly into a predicted state probability vector that defines a probability distribution: the vector components correspond to discrete classes or states $i = 1, \ldots, s$ of the probability distribution; they are nonnegative and sum up to unity.

In this study, the Markov-chain model utilizes three classes of discrete weather states (cumulus, stratus, and rain; see Table 3), which are mutually exclusive and collectively exhaustive.

Combining the two variables W1 (weather condition of the past 6 h) and WW (weather condition of the observation time) has the advantage that they cover the whole 6-h time interval compared to the amount of rain (RR) frequently measured over a 12-h interval (World Meteorological Organization 1988). The division of the

TABLE 2. Rms errors and correlation coefficient of all models and prediction lead times.

Forecast model	Rms error (°C) (h)					Correlation coefficient (%) (h)				
	0	6	12	18	24	0	6	12	18	24
Climate	3.9	3.9	3.9	3.9	3.9	—	—	—	—	—
6-h persistence	0.0	2.2	2.9	3.0	3.1	100.0	84.4	73.1	70.5	69.0
AR (1) process	0.0	2.1	2.7	2.8	2.9	100.0	84.6	73.1	70.7	69.0
R model	0.0	1.7	2.2	2.4	2.7	100.0	89.0	81.6	76.3	71.2

TABLE 4. Ranked list of the covariates.

No.	Type of covariate	Regression coefficient
1	Intercept	0.5622
2	Cloud cover (0–9 synop code)	0.0382
3	Pressure (DPa)	−0.0026
4	Log [Visibility (m) + 1]	−0.0030
5	Ceiling (0–9 synop code)	0.0071
6	Zonal wind (Kt)	−0.0019
7	Low clouds (0–9 synop code)	0.0051
8	Low cover (0–9 synop code)	0.0030
9	Pressure tendency (DPa)	−0.0069
10	Meridional wind (Kt)	0.0027
11	Temperature (1/10°C)	−0.0046
12	Dewpoint (1/10°C)	−0.0008

"no rain" class into two classes is supported by the fact that the covariate, low clouds (C_L), has a strong correspondence with the precipitation. The initial condition is defined with respect to the classes, WW and C_L, whereas the final condition uses WW, C_L, and W1 to determine the transition probabilities at the corresponding lead time of a first-order Markov-chain model forecast. For each class (0, 1, and 2) the Markov-chain model provides a transition probability for each forecast projection. For example, the 6-h forecast of an annual model yields for Hamburg Fulhsbüttel

$$(\hat{P}) = \begin{pmatrix} \widehat{p_{00}} & \widehat{p_{01}} & \widehat{p_{02}} \\ \widehat{p_{10}} & \widehat{p_{11}} & \widehat{p_{12}} \\ \widehat{p_{20}} & \widehat{p_{21}} & \widehat{p_{22}} \end{pmatrix} = \begin{pmatrix} 0.61 & 0.24 & 0.15 \\ 0.17 & 0.50 & 0.33 \\ 0.10 & 0.17 & 0.73 \end{pmatrix}.$$

For example, the probability of class 2 (rain) given an observation of class 1 (stratus) is 33% for the next 6 h.

1) MULTIPLE REGRESSION MARKOV MODEL

This Markov-chain model is now extended by combination with a standard multiple regression technique (Miller 1984; Miller and Leslie 1985; Kirk and Fraedrich 1998). The PoP (Kirk and Fraedrich 1998) for a given state $j = 1, 2, 3$ at current time t and for h hours ahead is

$$\text{PoP}(j, t, h) = \sum_{i=1}^{12} a_i(j, t, h) X_i(j, t, h), \qquad (4)$$

where $X_i(j, t, h)$ are the covariates and $a_i(j, t, h)$ the corresponding regression coefficients. These regression coefficients are estimated for each class (cumulus, stratus, and rain) and the 12 single-station standard synoptic data. Like the R model, it is possible to establish seasonal models (only used in section 5). This leads to 12 regression coefficients for each class and for each forecast projection. In contrast to the temperature forecast predicting for a time point (say the temperature in 6 h) the PoP forecast predicts for a time interval (say the PoP during the next 6-h time interval). Table 4 lists the 12 covariates of the 12-h prediction of an annual model

ranked in order of their importance. Here, the same modifications apply to the synoptic parameters as for the R-model, except for the temperature and the dewpoint anomalies being replaced by their actual values.

Note, that the covariate intercept, which is the most important covariate, indicates the transition probability that a pure Markov model would have without regression. All other covariates improve the performance of the M model significantly and reflect the climatological experience.

2) FORECAST VERIFICATION

To evaluate the PoP forecasts for the independent sample 1980–95 the joint distribution $p(\delta, \phi)$ of observations δ and forecasts ϕ is decomposed into the two conditional distributions $p(\delta | \phi)$, $p(\phi | \delta)$ and the two marginal distributions $p(\phi)$, $p(\delta)$. The conditional distribution $p(\delta | \phi)$ of the observations given the forecasts is called reliability; the conditional distribution $p(\phi | \delta)$ represents the forecasts given the observations and identifies as the discrimination ability or the conditional sharpness (Murphy and Winkler 1987). The diagonal line represents the perfect reliability, that is, $p(\delta = 1 | \phi) = \phi$ for all ϕ. The lower panel of Fig. 4 displays the reliability for 12-h prediction with small overforecasting $[\phi > p(\delta = 1 | \phi)]$ at lower probabilities and small underforecasting $[\phi < p(\delta = 1 | \phi)]$ for the rest.

The discrimination diagram illustrates that the model is able to differentiate between dry and wet days but that the general sharpness is relatively poor because the climatological chance of rain for the next 12 h is almost 50% in Hamburg. The traditional performance measure of probability forecasts, the half-Brier score (see Table 5), shows a significant improvement of the M model over standard reference forecasts. Note that the PoP forecasts are made for increasing time intervals (6, 12, and 24 h).

4. Forecast combination

The numerical weather predictions and the single-station stochastic model forecasts are combined in an error minimizing fashion. This procedure consists of two steps: downscaling and combination. The NWP model forecasts are areal forecasts whereas the forecasts of the statistical models are single-station forecasts. Before combination, it is necessary that these forecasts possess the same spatial representation. That is, the models' forecasts needs to be compatible in their space and time structure. Here, we apply the downscaling method reducing an areal forecast to a point forecast. This procedure is illustrated for the temperature anomaly at the station Hamburg Fuhlsbüttel. Next, the linear combination scheme is applied (appendix) to the temperature predictions (Thompson 1977; Fraedrich and Smith 1989) and to probability forecasts following Fraedrich and Leslie (1987a).
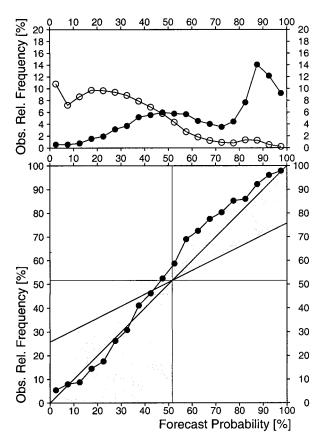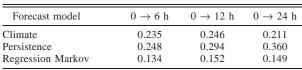
FIG. 4. Sharpness (upper panel) and reliability (lower panel) diagrams for the 12-h forecast; black dots represent rain events and white dots represent no-rain events; the shaded area is the region of positive skill in relation to climate.

a. Downscaling

The observations at the station Hamburg Fuhlsbüttel are compared with the forecasts of (a) nearest grid point 1, (b) with the mean of the nearest four grid points 1, 2, 3, 4, and (c) with the mean of the nearest nine grid points 1, 2, . . . , 9 (see Fig. 5). The rms error for temperature anomalies and the half-Brier score (Brier 1950) for probability of precipitation are used as performance measures.

Table 6 shows the rms errors of the NWP model forecasts for all lead times. The highest accuracy of 2.8°C is obtained by the prediction of the nearest grid point only. It is almost lead time independent due to the particular construction of the NWP time series (see Fig.
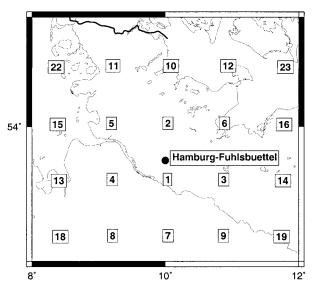

FIG. 5. Hamburg weather station (lower-left corner of the label) and the surrounding nearest grid points of the NWP (Europe Model).

1). The high rms error values of the time series associated with the four nearest grid points can be explained by their position. Most of them are preferably situated downstream of Hamburg and, therefore, underestimate the effect of the prevailing westerly flow. The choice of the nearest nine grid points compensates this effect, but this time series does not achieve the accuracy of the one with the nearest grid point. The values of this time series represent an average over too large an area underestimating local effects. Consequently, the time series built with the nearest grid point is the basis for the combination scheme for the temperature anomaly and also the probability of precipitation (not shown).

b. Combination of temperature forecasts

The important gains by the error minimizing combination of two independent predictions are shown for the forecasts ϕ_1 based on the multiple regression technique and the forecasts ϕ_2 of the nearest grid point of the NWP model. The linear combination ϕ_* of these two independent forecasts can be expressed as

$$\phi_* = a\phi_1 + b\phi_2, \qquad (5)$$

where *a* and *b* are the corresponding weights (for more details see the appendix). Again, the time series of the

TABLE 5. Half-Brier scores for PoP forecasts made by the M model and standard reference forecasts for the weather station Hamburg Fuhlsbüttel.

Forecast model	$0 \rightarrow 6$ h	$0 \rightarrow 12$ h	$0 \rightarrow 24$ h
Climate	0.235	0.246	0.211
Persistence	0.248	0.294	0.360
Regression Markov	0.134	0.152	0.149

TABLE 6. The rms error in °C of the NWP model forecasts for all lead times of prediction.

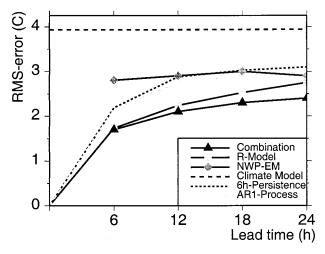| Lead time (h)	One grid point	Four grid points	Nine grid points
6	2.8	5.0	3.3
12	2.9	5.1	3.4
18	3.0	5.1	3.4
24	2.9	5.1	3.4

FIG. 6. Rms errors of independent temperature forecasts for all models with increasing prediction lead time.

TABLE 7. The regression and NWP model weights, *a* and *b,* for temperature forecast combination changing with increasing prediction lead time.

Lead time (h)	Weights	
	a	*b*
6	0.95	0.07
12	0.74	0.28
18	0.64	0.36
24	0.49	0.46

weather station at Hamburg serves as an example to demonstrate this scheme. Now, an increased model building part (1949–91) for fitting the R model is used because, for the combination, only the period 1991–95 is available. This latter period is divided into the fitting part (1991–93) for the coefficients for the linear combination and the independent verification part (1994–95). To obtain stable results, the partitioning into different seasons is not applied.

Figure 6 presents the rms errors of the combination forecasts and of the individual models used for the combination scheme. In addition, some simple standard models [climate, 6-h persistence, and the AR(1) process] are also presented. The rms error of the NWP model forecast is nearly constant. This results from the special construction of the time series supplying the time resolution of 6 h. Note, that the initial error of the NWP model forecasts cannot be estimated. For all lead times the R model accuracy is superior to the one of the NWP model that, however, decreases with increasing lead time. The best accuracy is achieved by the combination forecast. For the 6-h prediction the improvement can be neglected compared to the R-forecast quality. For the 24-h prediction the improvement by combination reaches 14% compared to the R model alone and 17% compared to the NWP model. Moreover, we see in Fig. 6 that the combination scheme gains nearly 12 h of lead time over the R model alone.

The lead time dependence of the combination weights *a* and *b* reflects the influence of the deterioration of the R model compared with the constant skill of the NWP model. In Table 7, the weight *a* corresponds with the forecast of the R model and *b* with the NWP model. The table shows that, for the 6-h forecast, the NWP model forecast can almost be ignored but, with increasing lead time, the importance of the NWP model increases at the expense of the R forecasts. For the 24-h prediction the weights attain similar magnitude, and the

contribution by each model to the combination is nearly the same. The sum of the weights totals more than unity. This implies that the NWP model forecasts have a cold bias.

c. Combination of PoP forecasts

Statistical PoP forecasts and NWP model forecasts are linearly combined following Fraedrich and Leslie (1987a). The downscaling for PoP (not shown) suggests also the EM time series built with the nearest grid point of the four types of precipitation, convective rain, convective snow, large-scale rain, and large-scale snow; if any of these parameters is greater than 0 mm at the grid point, PoP = 1; PoP = 0 otherwise.

The result for the combination of statistical PoP forecasts and NWP model forecasts is illustrated in Fig. 7. The half-Brier score is displayed changing with increasing integral lead time. Initial errors are not shown because PoP predicts a time interval. As recognized for temperature anomaly forecasts, the highest improvement using the combined forecast is achieved for the 24-h prediction with a gain of 18% compared to the M-model alone and 33% compared to the NWP model. Almost 18 h of lead time is gained via the combination method over the M-model. We also see that the 24-h
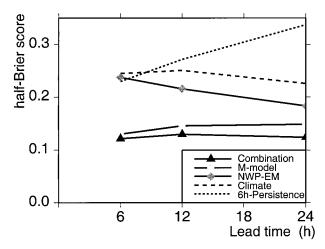

FIG. 7. Line diagram of the half-Brier scores of all models with increasing prediction lead time; initial errors are not included because PoP forecasts are integral measures of the predicting lead time interval.

TABLE 8. The combination weight a of the PoP forecast changing with increasing integral lead times.

Lead time (h)	Weight a
6	0.81
12	0.72
24	0.60

TABLE 9. Range of skill scores for all weather stations changing with increasing lead time.

Lead time (h)	Range of skill scores	
	Temp anomaly	Probability of precipitation
6	0.47–0.62	0.34–0.58
12	0.36–0.50	0.31–0.49
18	0.31–0.41	not available
24	0.25–0.35	0.25–0.39

predictions are more accurate than the 12-h forecasts. This behavior can be explained by the fact that the relative frequency of rain has increased from the unfavorable 50% level so that better predictions are possible. This is also illustrated by the half-Brier score of the climate mean forecast model.

For the linear combination forecast ϕ_* to be a probabilistic variable (as the two individual PoP forecasts) requires that the weights satisfy the relation $a + b = 1$ (Fraedrich and Leslie 1987a). Therefore, the combination scheme utilizes the weights a for the M-model forecasts and $b = 1 - a$ for the NWP model predictions; their magnitudes correspond with the accuracy of the individual forecast model contributing to the combination. The change of a (see Table 8) with increasing lead time also indicates the corresponding amount of accuracy that each single forecast contributes to the combination. The weight a of the M model decreases
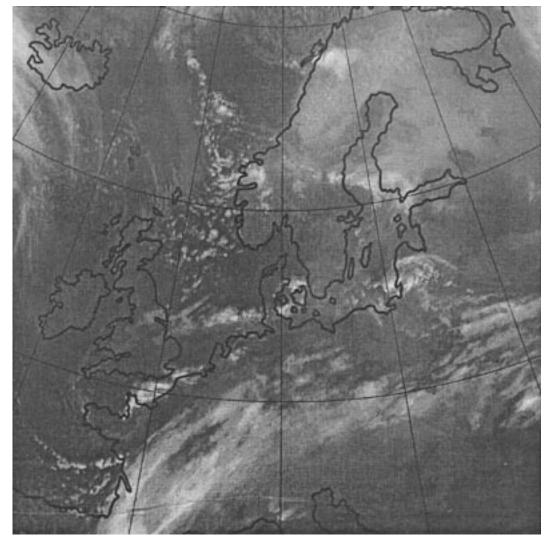

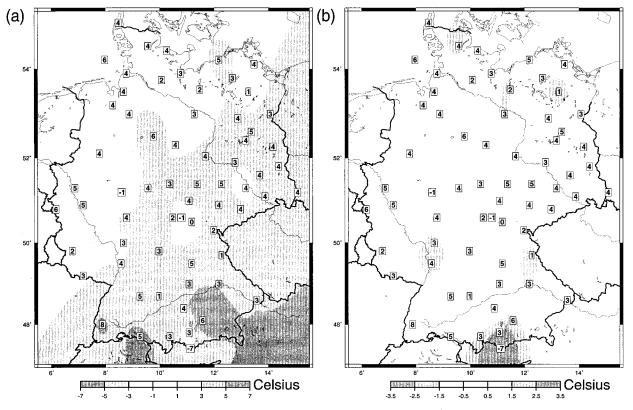FIG. 8. Infrared satelite image of central Europe for 12 Jan 1991.

FIG. 9. (a) The 12-h temperature anomaly forecast for 12 Jan 1991 initialized at 0600 UTC in 2°C contour intervals and (b) the difference between the observation forecast of the temperature anomaly in 1°C contour intervals; the corresponding temperature forecast is attached to the weather stations.

with increasing lead time. The highest improvement can be achieved, when both individual models have similar skill.

5. Area forecasts: Method and case study

The R model and the M model are fitted for 68 weather stations to deduce an areal forecast of temperature anomaly and of probability of precipitation for central Europe. Because of the different sample lengths, we choose two-thirds of the time series data as the model building part and one-third as the verification part.

a. Method

The interpolation technique of Smith and Wessel (1990) based on continuous curvature splines in tension is applied to both temperature and PoP forecasts of the irregularly distributed stations of central Europe to obtain areal forecasts. The performance range of the statistical models at these stations is presented in Table 9. It contains the range of the skill scores of the temperature anomaly and the PoP forecast based on rms errors and half-Brier scores (Brier 1950); the climate model is used as the reference model, because it represents the variability of the single-station time series.

For the temperature anomaly the geographical distribution of the skill scores is irregular; for the PoP forecasts we note the lower values in the northeast and the higher values in the southwest part of central Europe (the more continental part). Consequently, in the southwest the PoP predictions appear more accurate than in the northeast, which is influenced by the near coast. An illustrated example of the areal forecast is demonstrated by the following case study.

b. Case study (12 January 1991)

The synoptic situation is characterized by a strong upper trough over the British Islands and the Biscay with the zonal flow changing to a meridional one over central Europe for 11 January 1991. The next day (Fig. 8), the trough moves eastward with its axis and the associated cloud field extending from the Baltic Sea over the east of the Pyrenees to North Africa so that, south of the Danube, continuous rainfall and, in the north western part of central Europe, showers are observed. The area coverage, based on 68 weather stations, is presented for the 12-h forecast initialized on the main observation time 0600 UTC of 12 January. Figure 9a shows the predicted temperature anomaly (contoured in 2°C intervals) and at each station the predicted tem-
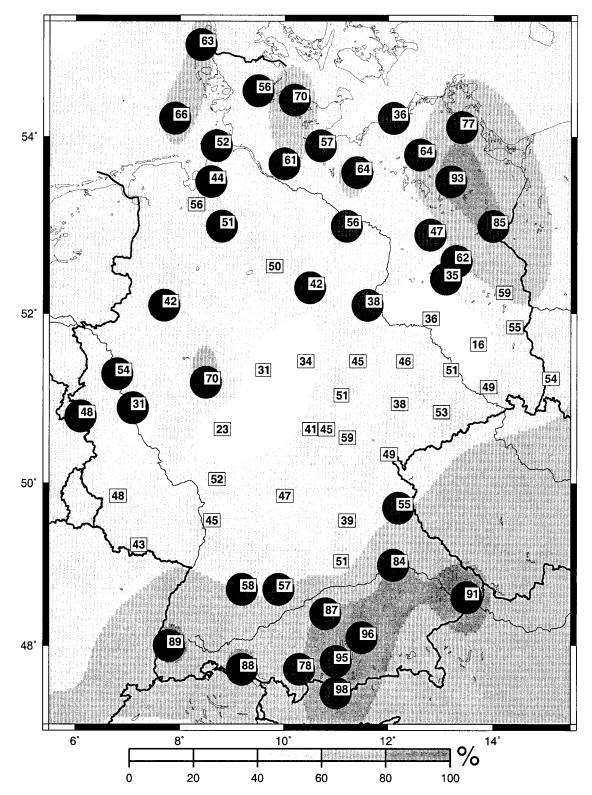
FIG. 10. Same as Fig. 9a but for 12-h forecast of PoP; circles around the weather station mark rainfall events.

perature (i.e., temperature anomaly plus long-term mean for the given day and the given time of day). Figure 9b shows the difference between the observation and the forecast (in 1°C intervals). Positive temperature anomalies are predicted for nearly the whole area increasing from the northwestern to the southeastern part except for central Europe. Comparison with the dominant weather situation shows that this decline of the positive temperature anomaly in the north of central Europe corresponds to the shift of the cloud field. The difference between forecasts and observations shows errors smaller than ±2.5°C for the 12-h prediction (±1.5°C for the 6-h forecast, not shown) in large areas of central Europe. The highest differences between forecasts and observations appear in the Alpine region, which is due to the short lengths of the time series (<10 yr) and the topography. But 89% of the weather stations show a difference that lies inside with their individual standard deviation bounds. In Fig. 10, the 12-h forecasts of PoP are presented in 20% increments. The observation "rain," as defined in Table 3, is marked by a shaded circle at the corresponding station. The precipitation south of the Danube, which is connected to the cloud field, and the rainfall in the northeast of central Europe, is well predicted (Fig. 10). The showers as subscale effects, however, are not so well predicted in the western part of central Europe in the second half of the day.

6. Summary and conclusions

Two statistical schemes are introduced for short-term (up to 24 h) single-station forecasts: a multiple regression model predicting temperature anomalies and a probability of precipitation forecast system. The verification of independent forecasts is shown in some detail using the most accurate scheme for the weather station Hamburg Fuhlsbüttel as an example. The verification methods applied are outlined by Murphy and Winkler (1987). The joint distribution allows an interpretation of the forecast performance itself. Even if extreme temperature anomalies are observed, the scatter of the corresponding forecasts increases only slightly. For example, 80% of the 6-h forecasts lie in a range of ±2°C. The rms error of the temperature forecasts achieves a magnitude of 1.7°C for the 6-h forecast (2.2°C for the 12-h forecast) and the half-Brier score of the PoP forecasts yields 0.13 for the 6-h forecast (0.15 for the 12-h forecast). The comparison with simple standard statistical models like the climate mean shows a significant improvement of accuracy obtained by the more advanced statistical models.

As the performance of the statistical models shows good results, the linear combination of these models with numerical weather predictions achieves further improvements in forecast accuracy. The combination temperature forecasts yield a 14% gain for the 24-h prediction with respect to the R model alone, and 17% gain with respect to NWP model. The combined PoP forecast achieves 18% with respect to the R model and 33% with respect to NWP model. While statistical models based only on observations are independent of NWP models, the MOS technique, to provide the best results, requires a recomputation whenever the NWP model is changing. Therefore, the linear combination scheme offers an alternative way to improve the direct model output (DMO) instead of the widely used model output statistics or Kalman filtering methods because only the weights need to be recomputed whenever the NWP model version changes.

The results of the statistical weather forecasting system illustrate the possibility of deriving an areal forecast for Europe using a standard interpolation technique that is economical in computing time. Due to the small technical expense, such schemes may be suitably introduced to regional weather forecast centers. Extensions are possible if the observations of neighboring weather stations are included. An operational statistical system running every 6 h is shown in the Internet (http://www.dkrz.de/wetter/prognosen/).

APPENDIX

Combination of Two Independent Predictions

The appendix describes the mathematical optimal linear method for combining two independent forecasts of continuous variables. The derivation of the combination scheme follows the method of Fraedrich and Leslie (1987a) for probability forecasts and results by Thompson (1977) and Fraedrich and Smith (1989).

Let ϕ_1 and ϕ_2 denote two independent forecasts of the same event in the future (e.g., the temperature anomaly in 6 h, and δ the corresponding observation of this event). The aim of a linear combining scheme is to reach a higher accuracy in predicting the observation δ than using each single forecast alone. For continuous variables, one of the classic measures of performance is the ensemble mean square error (mse).

The linear combination ϕ_* of the two independent forecasts can be expressed as

$$\phi_* = a\phi_1 + b\phi_2, \qquad (A1)$$

with the weights a and b. The corresponding mse for the combination forecast ϕ_* is

$$\text{mse} = \langle(\phi_* - \delta)^2\rangle$$

$$= \langle a^2\phi_1^2 + b^2\phi_2^2 + \delta^2 + 2ab\phi_1\phi_2 - 2a\phi_1\delta - 2b\phi_2\delta\rangle$$

$$= a^2\langle\phi_1^2\rangle + b^2\langle\phi_2^2\rangle + \langle\delta^2\rangle + 2ab\langle\phi_1\phi_2\rangle$$

$$- 2a\langle\phi_1\delta\rangle - 2b\langle\phi_2\delta\rangle. \tag{A2}$$

Here, $\langle\phi_1\phi_2\rangle$ is the covariance between the forecasts ϕ_1 and ϕ_2; $\langle\phi_1\delta\rangle$ and $\langle\phi_2\delta\rangle$ are the covariances between either ϕ_1 and the observation δ or ϕ_2 and the observation δ. In addition, $\langle\phi_i^2\rangle$ with $i = 1, 2$ is the variance of the corresponding forecast and $\langle\delta^2\rangle$ is the variance of the observation characterizing the time series. For a time series of temperature anomalies $\langle\delta^2\rangle$ is the climatological variability.

The objective of this combination scheme is to find the optimal weights a and b. Therefore a and b are chosen in such a manner that the mse is minimized by the combination forecast. Regarding mse as a function of a and b, the minimum can be found by deriving this function:

$$\frac{\partial\text{mse}}{\partial a} = 2a\langle\phi_1^2\rangle + 2b\langle\phi_1\phi_2\rangle - 2\langle\phi_1\delta\rangle = 0 \tag{A3}$$

$$\frac{\partial\text{mse}}{\partial b} = 2b\langle\phi_2^2\rangle + 2a\langle\phi_1\phi_2\rangle - 2\langle\phi_2\delta\rangle = 0. \tag{A4}$$

Solving this system of equations yields to

$$a = \frac{\langle\phi_2^2\rangle\langle\phi_1\delta\rangle - \langle\phi_1\phi_2\rangle\langle\phi_2\delta\rangle}{\langle\phi_1^2\rangle\langle\phi_2^2\rangle - (\langle\phi_1\phi_2\rangle)^2} \tag{A5}$$

$$b = \frac{\langle\phi_1^2\rangle\langle\phi_2\delta\rangle - \langle\phi_1\phi_2\rangle\langle\phi_1\delta\rangle}{\langle\phi_1^2\rangle\langle\phi_2^2\rangle - (\langle\phi_1\phi_2\rangle)^2}. \tag{A6}$$

Note that the weights do not fulfill the condition $a + b = 1$ like the optimal weights linearly combining probability forecasts (Fraedrich and Leslie 1987a). This condition of probability leads to only one unknown weight a.

REFERENCES

Anderson, T. W., and L. A. Goodman, 1958: Statistical inference about Markov chains. *Ann. Math. Stat.,* **28,** 89–110.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.,* **78,** 1–3.

Dallavalle, J. P., 1996: A perspective on the use of Model Output Statistics in objective weather forecasting. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., 479–482.

Draper, N. R., and H. Smith, 1981: *Applied Regression Analysis.* 2d ed. John Wiley and Sons, 709 pp.

Fraedrich, K., and K. Müller, 1983: On single station forecasting: Sunshine and rainfall Markov chains. *Contrib. Atmos. Phys.,* **56,** 108–134.

——, and L. M. Leslie, 1987a: Combining predictive schemes in short-term forecasting. *Mon. Wea. Rev.,* **115,** 1640–1644.

——, and ——, 1987b: Evaluation of techniques for the operational, single station, short-term forecasting of rainfall at a midlatitude station (Melbourne). *Mon. Wea. Rev.,* **115,** 1645–1654.

——, and N. R. Smith, 1989: Combining predictive schemes in long-range forecasting. *J. Climate,* **2,** 291–294.

Gabriel, K. R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence in Tel Aviv. *Quart. J. Roy. Meteor. Soc.,* **88,** 90–95.

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.

Hess, G. D., L. M. Leslie, A. E. Guymer, and K. Fraedrich, 1989: Application of a Markov technique to the operational, short-term forecasting of rainfall. *Aust. Meteor. Mag.,* **37,** 83–91.

Kemeny, J. G., J. L. Snell, and A. W. Knapp, 1976: *Denumerable Markov Chains.* Springer-Verlag, 210 pp.

Kirk, E., and K. Fraedrich, 1998: Probability of precipitation: Short-term forecasting and verification. *Contrib. Atmos. Phys.,* **72,** 263–271.

Knüpffer, K., 1996: Methodical and predictability aspects of MOS systems. Preprints, *13th Conf. on Probability and Statistics in Atmospheric Sciences,* San Francisco, CA, Amer. Meteor. Soc., 190–197.

Madden, R. A., 1981: A quantitative approach to long-range predictions. *J. Geophys. Res.,* **86,** 9817–9825.

Majewski, D., 1991: Numerical methods in atmospheric models. *Seminar Proc.,* Vol. 2, Reading, United Kingdom, ECMWF, 147–191.

Miller, A. J., and L. M. Leslie, 1984: Short-term single-station forecasting of precipitation. *Mon. Wea. Rev.,* **112,** 1198–1205.

——, and ——, 1985: Short-term single-station probability of precipitation forecasting using linear and logistic models. *Contrib. Atmos. Phys.,* **58,** 517–527.

Miller, R. G., 1984: GEM: A statistical weather forecasting procedure, short- and medium-range weather prediction research program (PSMP). *WMO Publi. Series,* **10,** 102 pp.

Murphy, A. H., 1998: The early history of probability forecasts: Some extensions and clarifications. *Wea. Forecasting,* **13,** 5–15.

——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.

——, B. G. Brown, and Y.-S. Chen, 1989: Diagnostic verification of temperature forecasts. *Wea. Forecasting,* **4,** 485–501.

Navato, A. R., 1981: An experimental short-term climate forecasting procedure for U.S. winter temperature and its verification. *Proc. Fifth Annual Climate Diagnostics Workshop,* Seattle, WA, National Oceanic and Atmospheric Administration, 360–363.

Nicholls, C., 1980: Long-range weather forecasting: Value, status, and prospects. *Rev. Geophys. Space Phys.,* **18,** 771–788.

Norton, D. A., 1985: A multivariate technique for estimating New Zealand temperature normals. *Wea. Climate,* **5,** 64–74.

Perrone, T. J., and R. G. Miller, 1985: Generalized exponential markov and model output statistics: A comparative verification. *Mon. Wea. Rev.,* **113,** 1524–1541.

Smith, W. H. F., and P. Wessel, 1990: Gridding with continuous curvature splines in tension. *Geophysics,* **55,** 293–305.

Thompson, P. D., 1977: How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.,* **105,** 228–229.

von Storch, H., and F. W. Zwiers, 1998: *Statistical Analysis in Climate Research.* Cambridge University Press, 510 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Wilson, L. J., 1985: Application of statistical methods to short range operational weather forecasting. Preprints, *Ninth Conf. on Probability and Statistics in Atmospheric Sciences,* Virginia Beach, VA, Amer. Meteor. Soc., 1–10.

World Meteorological Organization, 1988: International codes. Manual of Codes, 1, No. 306.