Nonlinear Processes in Geophysics (2003) 10: 373–383

Nonlinear Processes in Geophysics

Linear and nonlinear post-processing of numerically forecasted surface temperature

M. Casaioli[1], **R. Mantovani**[2], **F. Proietti Scorzoni**[3], **S. Puca**[3], **A. Speranza**[4], **and B. Tirozzi**[3]

[1]Institute of Atmospheric Physics, National Research Council, Rome, Italy
[2]Physics Department, University, Bologna, Italy
[3]Physics Department, University "La Sapienza", Rome, Italy
[4]Mathematics and Informatics Department, University, Camerino, Italy

Received: 25 February 2002 – Revised: 22 July 2002 – Accepted: 17 October 2002

Abstract. In this paper we test different approaches to the statistical post-processing of gridded numerical surface air temperatures (provided by the European Centre for Medium-Range Weather Forecasts) onto the temperature measured at surface weather stations located in the Italian region of Puglia. We consider simple post-processing techniques, like correction for altitude, linear regression from different input parameters and Kalman filtering, as well as a neural network training procedure, stabilised (i.e. driven into the absolute minimum of the error function over the learning set) by means of a Simulated Annealing method. A comparative analysis of the results shows that the performance with neural networks is the best. It is encouraging for systematic use in meteorological forecast-analysis service operations.

1 Introduction

The gridded fields produced by numerical models for forecast-analysis of meteorological fields are currently used in a variety of applications. It often happens that the information contained in the direct model output is insufficient for the proposed practical use, because the definition of some process is missing or incomplete. It is necessary then to include in deterministic or statistical form, information derived from other sources (geography, climatology, additional observations, etc.), in order to achieve the necessary detail concerning the process in question. This procedure is known as post-processing of model output. A very common type of post-processing is the so-called down-scaling, i.e. interpolating the gridded fields in some area to a higher resolution than that of the original grid. This paper deals with the problem of down-scaling two-meter temperatures (T2m),[1] produced by the operational European Centre for Medium Range Weather Forecasts (ECMWF) model onto the location of a number of meteorological stations spread over the Italian region of Puglia (Fig. 1). More specifically, our research consists of the comparative examination of different post-processing methods as applied to T2m post-processing over Puglia.

Several post-processing techniques are commonly employed in the operative forecasting centres and we refer to them as standard techniques. Since the results of their application are often not reported in the open literature, we refer here only to some papers containing information useful for our work.

An initial post-processing can be performed by applying altitude corrections. This technique has been applied to good effect, for example, by Deidda et al. (2000) in the Chilean Andes region, at an altitude of about 2600 m, where elevation errors were of obvious relevance.

Kalman Filters (KF) are widely used for post-processing. Previous work by Cacciamani (1993) on the use of such a technique in post-processing ECMWF numerical T2m onto stations located in Emilia-Romagna (another Italian region) was useful in our analysis. The use of the thickness between isobaric surfaces as an input parameter, rather than the two-meter temperature, is another common practice. An application that was instructive for us is discussed in a paper by Massie and Rose (1997).

Neural Networks (NN) do not address the problem of access to information remarked above with the standard techniques: an ever increasing number of papers have been published on the subject (see, for example, Hsieh and Tang, 1998; Marzban and Stumpf, 1998; Shao, 1998; Hall et al., 1999; Koizumi, 1999; Narasimhan, 2000; Tang et al., 2000) and new results are readily available in the open literature. However, obtaining a better performance by a NN than with standard techniques may be not simple. For example, a major shortcoming of back-propagation neural networks (as the one proposed in Tang et al., 2000) is the trapping in local

Correspondence to: A. Speranza
(antonio.speranza@unicam.it)

[1]This is the temperature adopted by the World Meteorological Organization as standard surface temperature: it is assumed to correspond to temperatures measured by thermometers at surface meteorological stations.

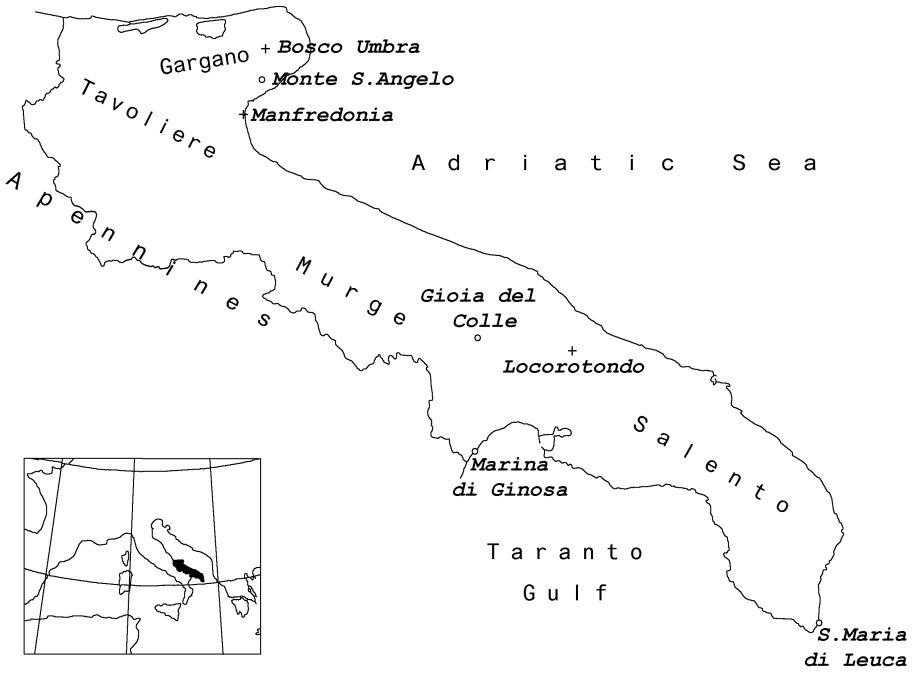

Fig. 1. Map of Puglia. The selected meteorological stations are marked: circles indicate SYNOP stations, crosses indicate SIMN stations.

minima of the error function. In this case, the ability of the learning algorithm in reaching the lowest values of the error strongly depends on the initial values. The approach that we propose in this paper for the global minima of the error function is through a procedure called Simulated Annealing (Geman and Geman, 1984): starting from a random initial state, the minima is reached after a suitable number of steps. In our case, we did not encounter any computational problem due to the convergence rate.

This work is just a preliminary step towards the setting up of post-processing routines for the high resolution (40 levels in the vertical, 10 km horizontal grid-size, coverage of the entire Mediterranean) Limited Area Model (LAM) of the Dipartimento per i Servizi Tecnici Nazionali (DSTN). Service operations with this model have just begun and, as a consequence, a sufficiently long-term series of output data is not yet available. In order to test different post-processing techniques, therefore, it seems natural to select the ECMWF forecast, which is currently used to provide the DSTN-LAM's initial and boundary conditions. The Italian stations of the ground-based WMO network (SYNOP) have a density comparable with the ECMWF model resolution (50 km). Also in our work, we used a higher-resolution surface network, of the Servizio Idrografico e Mareografico Nazionale (SIMN), which is more suitable for the set-up of a post-processing procedure for the DSTN-LAM service output, when a sufficient historical record will become available. The choice of Puglia is due to its smooth orography (which moderates the difficulty of the down-scaling problem) and to its climatic regime (the occurrence of extreme temperature events), together with the availability of reliable observations from stations in the SYNOP network (9) and in the SIMN network (70).

The paper is organized as follows: in Sect. 2 we describe the site and in Sect. 3 the data sets used in our work; Sect. 4 provides, by means of standard techniques, some statistical estimates against which to check the results from the application of the neural network described in Sect. 5; in Sect. 6 we draw our conclusions; technical details of the Monte-Carlo method are presented in Appendix A; Simulated Annealing is schematically described in Appendix B.

2 Site description

The geographical area considered in this paper, the Italian region of Puglia (Fig. 1), is an elongated strip of land located at the south eastern end of the Italian Peninsula. It consists of a plain (Tavoliere) and a steep headland (Gargano) in the north, a hilly center (Murge), a flat peninsula (Salento) in the south. The climate of this area is characterized by the occurrence of frosts. The surface observation coverage is good: 9 stations belonging to the SYNOP network and 70 stations to the SIMN network.

3 Data sets

Our work has been performed on a forecast data set and a control data set, consisting, respectively, of gridded ECMWF numerically forecast data and station observed surface temperatures.

3.1 Data sets for the application of the standard statistical post-processing techniques

In the application of standard post-processing techniques, we used a time-series of daily (in order to filter out the daily cycle) 12:00 UTC surface data, for the year 1996. The control set consists of T2m values measured at four stations in the SYNOP network, Monte Sant'Angelo (MSA), Gioia del Colle (GdC), Marina di Ginosa (MdG), S. Maria di Leuca (SMdL) and three stations in the SIMN network, Bosco Umbra (BoU), Manfredonia (Man), Locorotondo (Loc). These stations (their location is shown in Fig. 1) were selected to represent mountainous, hilly and maritime environments, respectively in the north, center, and south of the region. The basic geographic characteristics of the seven selected stations are summarized in Table 1. The forecast data set includes here two-meter temperatures, geopotential height (GPH) and temperature (T) over the three lowest pressure levels (1000 mb, 925 mb, and 850 mb) of the ECMWF global forecast model. For each selected station, the values of the above fields at the four grid-points nearest to the measuring location were considered. The ECMWF model resolution is 0.5° (about 50 km) for the year in question (1996). We used the 24-hour forecast (starting at 12:00 UTC) issued daily by ECMWF.

Two digital elevation models (see Figs. 2a, b) were used for altitude corrections: the ECMWF model orography, with a resolution of 0.5°, and the U.S. National Imagery and Mapping Agency Digital Elevation Model (NIMA DEM), with a resolution of 30 seconds of arc" (about 1 km).

3.2 Data sets for the neural network approach

In performing post-processing by means of a NN, a 10-year (1986–1995) record of control and forecast data was used. The control set consists of T2m daily (12:00 UTC) observations at the GdC SYNOP station, the forecast set of T2m daily (12:00 UTC) 24-hour ECMWF forecast at the four grid-points surrounding the selected station.

4 Standard post-processing techniques

The statistical relationship between forecast and observed temperatures is most easily visualized by means of scatter-diagrams as shown in Fig. 3. The linear regression line fitting the distribution of points in the diagram allows for the definition of the error components: systematic errors are measured by the intercept (when different from zero) and the slope (when different from one), while the dispersion of points around the line represents the random error component. Some basic statistical characteristics of the error distribution are measured by the following standard indexes:

$$\text{bias} = \frac{1}{N}\sum_{n=1}^{N}(To_n - Te_n), \tag{1}$$

$$\text{Mean Absolute Deviation (MAD)} = \frac{1}{N}\sum_{n=1}^{N}|To_n - Te_n|, \tag{2}$$

$$\text{correlation} = \frac{\sum_{n=1}^{N}(To_n - \overline{To}) \cdot \sum_{n=1}^{N}(Te_n - \overline{Te})}{\sqrt{\sum_{n=1}^{N}(To_n - \overline{To})^2} \cdot \sqrt{\sum_{n=1}^{N}(Te_n - \overline{Te})^2}}, \tag{3}$$

where Te_n and To_n are, respectively, the forecast and observed temperatures for the n-th day of the series, N is the total number of days, and the bar indicates the averaging operator.

For comparison between the observed and forecast data sets we tested two different definitions of forecast surface temperature: bilinear interpolation of T2m at the four grid-points nearest to the observing station in question and T2m at the grid-point nearest to the station itself. The difference between the results obtained with the two definitions is negligible. As a consequence, we discuss in this paper only the results obtained by using the bilinearly interpolated T2m. In the following Sects. 4.1–3 we describe the different post-processing techniques and in Sect. 4.4 we discuss the results of their application comparatively.

4.1 Vertical interpolation

Visual comparison of the ECMWF orography, in Fig. 2a, with the higher resolution NIMA DEM orography, in Fig. 2b, gives us an idea of how large the discrepancies arising in the numerical evaluation of the elevation at a single point can be. The quantitative estimate of the difference between the two digital elevation models at the seven selected stations is reported in Table 1. Stations characterized by a particularly large elevation difference (MSA and BoU) also exhibit a large systematic error: a bias of more than 4°C, as reported in Table 2 and an intercept considerably different from zero. The value of Mean Absolute Deviation (MAD) is, correspondingly, quite large. Scatter diagrams for MSA (Fig. 3a) and GdC (Fig. 3b) illustrate, respectively, the case of high and low elevation difference. The other stations manifest a similar behavior (basically in proportion to the magnitude of the elevation difference).

Under the above described conditions, an initial post-processing improvement can be obtained by simply correcting the elevation error with vertical interpolation. Starting from the geopotential height of the pressure levels at the four grid-points surrounding the station, we applied a post-processing algorithm which selected the upper and lower levels nearest to the station height at any given point and performed, at the station altitude, a vertical linear interpolation between the relative temperatures. This procedure can give quite satisfactory results when there are significant differences between the real altitude and model one (see, for example, Deidda et al., 2000 with application to the somewhat extreme case of the stations in the Andes).

Table 1. Geographical features of selected SYNOP and SIMN stations. ECMWF height is bilinearly interpolated from the four surrounding grid-points

Station	Network	Location	Longitude	Latitude	Actual-Z (m)	ECMWF-Z (m)
MSA	SYNOP	mountain, N	15°57' E	41°42' N	838	240
GdC	SYNOP	hill, C	16°56' E	40°46' N	345	246
MdG	SYNOP	seaside, C	16°53' E	40°26' N	2	194
SMdL	SYNOP	seaside, S	18°21' E	39°49' N	104	−16
BoU	SIMN	mountain, N	16°00' E	41°49' N	750	296
Man	SIMN	seaside, N	15°53' E	41°35' N	2	206
Loc	SIMN	hill, C	17°20' E	40°45' N	420	133

Z = station's height
N = north
C = center
S = south

Table 2. Statistical comparison between T2m observations at SYNOP and SIMN stations and ECMWF forecast, post-processed using various standard techniques

Station	T2mO (°C)	Post-processing method	T2mP (°C)	Bias (°C)	MAD (°C)	Correlation
MSA	11.1	None	16.0	−4.9	5.0	0.970
		Vertical interpolation	10.1	0.9	1.6	0.970
		Kalman filter	11.1	0.0	1.6	0.958
GdC	16.1	None	16.5	−0.3	1.9	0.969
		Vertical interpolation	14.8	1.3	2.1	0.970
		Kalman filter	16.2	0.0	1.7	0.959
MdG	17.4	None	16.6	0.8	2.0	0.936
		Vertical interpolation	17.9	−0.6	2.0	0.931
		Kalman filter	17.4	−0.1	2.0	0.913
SMdL	16.7	None	17.5	−0.9	1.4	0.956
		Vertical interpolation	16.5	0.1	1.3	0.957
		Kalman filter	16.7	0.0	1.5	0.933
BoU	13.2	None	17.4	−4.2	4.3	0.952
		Vertical interpolation	12.1	1.1	2.0	0.963
		Kalman filter	13.2	0.0	1.9	0.951
Man	20.0	None	17.2	2.8	3.0	0.956
		Vertical interpolation	18.2	1.8	2.3	0.949
		Kalman filter	20.0	0.0	1.9	0.941
Loc	17.2	None	18.8	−1.6	2.2	0.964
		Vertical interpolation	16.2	1.0	1.8	0.969
		Kalman filter	17.2	0.0	1.7	0.956

T2mO = annual average of observed temperature
T2mP = annual average of forecast temperature

4.2 Regression from thickness

An alternative approach to post-processing consists of using "predictors" variables that, although highly correlated with T2m, are not linked to the surface processes which are poorly represented in numerical models. In this type of approach, a typical variable used as a predictor in the post-processing of surface variables is the thickness of the atmospheric layer between isobaric surfaces at some height above the ground. Massie and Rose's (1997) analysis of a 1995–1996 numerical model thickness forecasts for Nashville (U.S.) provides a good example of correlation with observed T2m.

In this study, the chosen predictor for T2m is the mean temperature $\overline{T}$ of the layer between two isobaric surfaces, indicated as P_1 and P_2. $\overline{T}$ is connected to the thickness Δz of the layer by the relationship derived from the hydrostatic

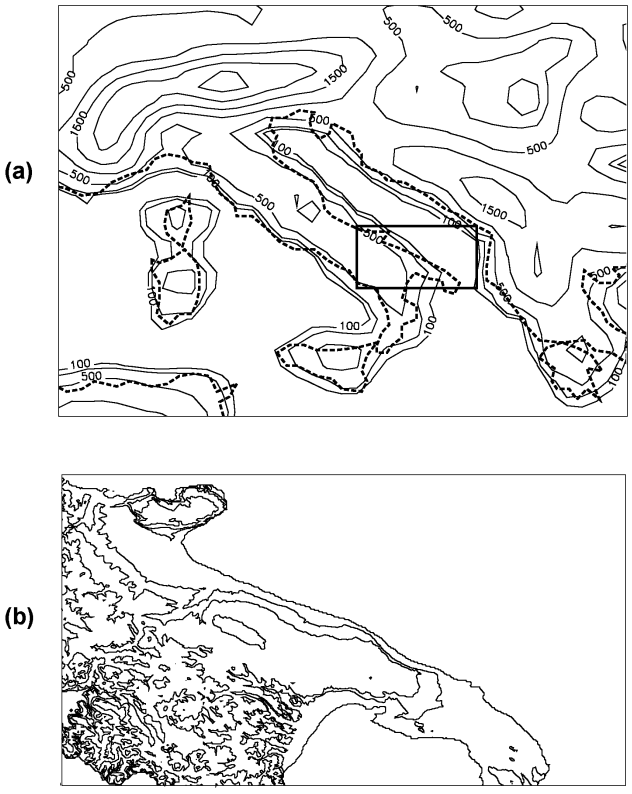

Fig. 2. **(a)** ECMWF model orography over Central Mediterranean Sea (resolution: 50 km); the rectangle delimitates the area in Fig. 2b; coastline (thick dashed line) is superimposed for clarity. **(b)** Detail over Puglia region of NIMA DEM orography (resolution: 1 km).

equation:

$$\overline{T} = \frac{g\Delta z}{R} \ln \frac{P_1}{P_2}, \tag{4}$$

where g is the gravitational acceleration and R is the gas constant. When a sufficiently long time series of $\overline{T}$ and T2m are known at a given location, a linear regression relationship can be established and values of T2m can be regressed from the forecast. To assess the effectiveness of this post-processing method, a model forecast at the three lower isobaric levels (1000, 925, and 850 mb) is used to calculate three different predictors.

4.3 Kalman filters

The parameters of the regression line derived from the scatter-diagram of the forecast-observed values at a station can be used in order to define a linear correction for the predicted value. The slope, A, and the intercept, B, of the regression line satisfy the relationship:

$$Te_n = A \cdot To_n + B + \eta_n, \tag{5}$$

where η_n is a random error term for the n-th day. A "corrected" forecast temperature, $T'e_n$, can then be constructed:

$$T'e_n = \frac{1}{A} Te_n - \frac{B}{A}, \tag{6}$$

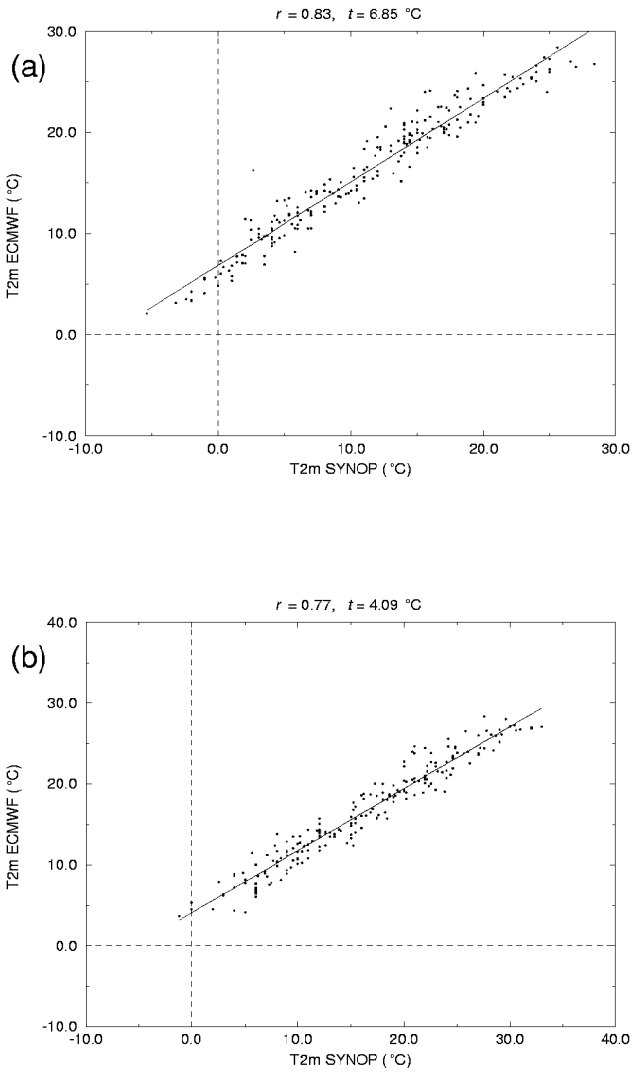

Fig. 3. Scatter-diagrams of ECMWF forecast T2m versus observed T2m: **(a)** MSA SYNOP station; **(b)** GdC SYNOP station. The slope and the intercept of the regression line are labeled with r and t, respectively.

differing from observations only for an error term, since $T'e_n - To_n = \eta_n / A$.

As new observations are inserted, in order to take into account time-variations of the systematic error component, KF varies the regression line parameters. The relation (5) then becomes:

$$Te_n = A_n \cdot To_n + B_n + \eta_n, \tag{7}$$

where A_n and B_n values, at each instant, are calculated by means of an algorithm which, in its simplest form, requires only a few previous values of the forecast error. As mentioned in the Introduction, an application of the above method to surface temperature in some Italian stations can be found in Cacciamani (1993). KF has two main practical advantages: it does not require long historical records and it minimises the bias over time scales, even when they are quite short.

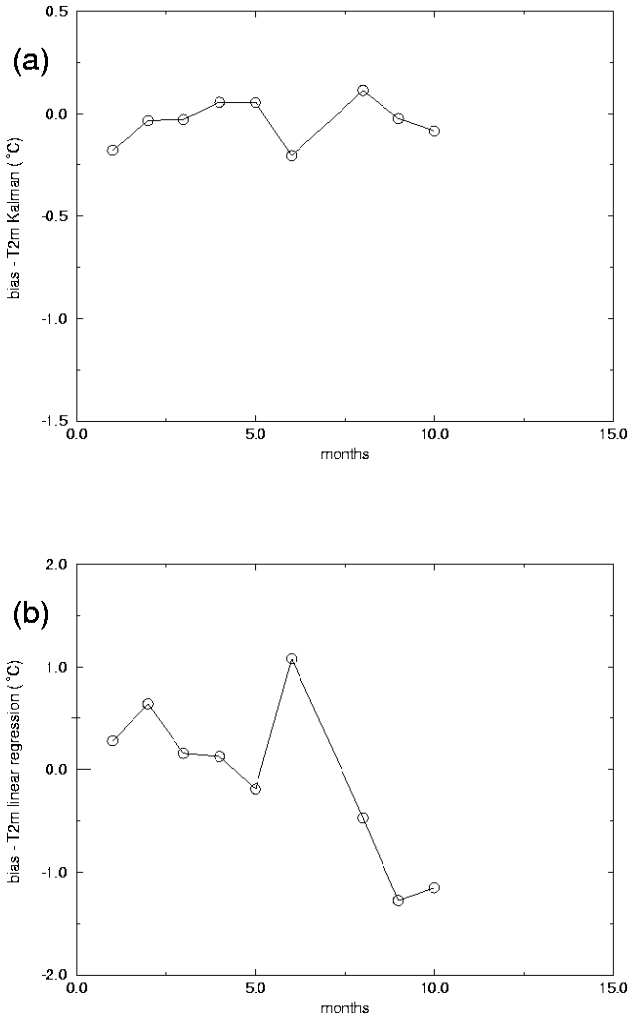

Fig. 4. Annual trend of monthly bias for MSA SYNOP station: **(a)** Kalman filter; **(b)** a posteriori linear regression of ECMWF T2m analysis against observed T2m (months numbered progressively starting from January).

4.4 Comparative application of standard methods

Table 2 reports the results obtained by applying all the above described post-processing procedures. As expected, vertical interpolation proves to be quite effective as a post-processing technique in the presence of strong elevation errors, since it eliminates almost all the systematic error component (Table 2). Vertical interpolation, however, may result in being less effective, or even disadvantageous (see, for example, the results for GdC) in other cases.

KF introduce a strong reduction in the systematic errors for all stations: the filtered series show very small bias. The filter performance is also good when monthly values of the bias are examined (Fig. 4). KF turns out to be, however, of little (if any) help in reducing the random error component of the MAD.

Consecutive application of different post-processing procedures exhibits the same statistical results as the application of the most powerful one. All procedures, moreover, have only a minimal effect on all (generally good) correlation values. We did not find any statistical inhomogeneity amongst the post-processed samples of observations at SYNOP and SIMN stations.

Table 3. Correlation between isobaric thickness forecast by ECMWF (for different layers) and observed T2m over the selected SYNOP stations. Correlation between forecast and observed T2m is also shown for comparison

Station	Forecast T2m	1000–925 mb layer	1000–850 mb layer	925–850 mb layer
MSA	0.970	0.962	0.958	0.952
GdC	0.970	0.963	0.958	0.951
MdG	0.936	0.926	0.920	0.936
SMdL	0.956	0.940	0.933	0.956

The results of the application of the thickness-regression method to temperatures measured at the SYNOP stations are also shown in Table 3. The correlation values are equal or slightly less than those obtained by directly using the T2m forecast. There is no evidence of different behavior, i.e. different correlation, in different layers.

5 Neural network post-processing technique

Neural networks are algorithms set up for solving the general problem of finding an unknown law linking two quantities. The data to be analysed may be chaotic or stochastic (Hertz et al., 1993). These algorithms can be applied to any nonlinear relationship, generating a sequence of patterns $\{\boldsymbol{x}^i, \boldsymbol{y}^i\}_{i=1}^P$ (where $\boldsymbol{x}^i$ and $\boldsymbol{y}^i$ are vectors with dimension N and L, respectively); it is, therefore, assumed that there exists a function $f : R^N \rightarrow R^L$ of the form:

$$\boldsymbol{y}^i = f(\boldsymbol{x}^i), \tag{8}$$

where the input vector $\boldsymbol{x}^i$ is mapped onto the output vector $\boldsymbol{y}^i$ by the action of the algorithm itself. The target of the search is the best approximation of the function f. Note that the function f is generic and can also contain random terms. Letting $\overline{\boldsymbol{y}}$ be the output of the neural network, we define the error function as:

$$E = \frac{1}{P} \sum_{i=1}^{P} |\boldsymbol{y}^i - \overline{\boldsymbol{y}}^i|, \tag{9}$$

where P is the number of input-output pairs (patterns) which we present to the network during the learning phase. The data set is divided into two subsets. One, the learning set, coincides with the collection of patterns on which the weights are estimated by means of the minimisation of Eq. (9). The generalisation skill of the NN is then evaluated by applying the error (9) on the second subset of data, called the testing set.

In the application of the neural network approach to the problem of post-processing the surface temperature, let $T_i^1, T_i^2, T_i^3, T_i^4$ be the ECMWF numerically forecast temperature for the i-th day at the four grid points surrounding the station and To_i the temperature measured in the same day at the output station. Then:

$$To_i = f(T_i^1, T_i^2, T_i^3, T_i^4). \quad (10)$$

In the simplest version of the post-processing procedure the input of the neural network would be $\boldsymbol{x}^i = (T_i^1, T_i^2, T_i^3, T_i^4)$, and the output Tr_i, forced to be statistically as close as possible to To_i. However, the structure of the input vector used in our work is more complex. In fact, experimentation with different structures shows that the simple NN defined above with the four temperatures forecast by the ECMWF for the i-th day is not sufficient to provide better results than the standard post-processing techniques. Satisfactory results are, instead, obtained by including in the input vector the values of temperature measured at the considered station over the previous six days $(To_{i-1}, ..., To_{i-6})$. Another change with respect to the basic approach expressed by Eq. (10), which turned out to be necessary, consists of forming the input vector with the first differences of the ECMWF temperatures: $(T_i^1 - T_{i-1}^1, ..., T_i^4 - T_{i-1}^4) = (dT_i^1, ..., dT_i^4)$, instead of the temperatures themselves. The input vector then becomes:

$$\boldsymbol{x}^i = (To_{i-1}, ..., To_{i-6}, dT_i^1, ..., dT_i^4). \quad (11)$$

This change is motivated by the fact that the slow decay shown by the autocorrelation function of the ECMWF forecast temperature (see Fig. 5a) would entail the use of the forecast temperatures of several previous days. On the other hand, such a long correlation time indicates the existence of a deterministic, low-frequency component in the forecast T2m. In our statistical application this component can be filtered out, and this is done by taking the first differences of the forecast temperature as in Eq. (11). The rapid decay of the first-difference series autocorrelation function (Fig. 5b) allows then the use of only one temperature difference at the four grid points near the observing station (i.e. four additional neurons) in the input vector.

The structure of the neural network adopted in processing the above determined data set is the following (the architecture is shown in Fig. 6):

- an input layer with $N = 10$ neurones;
- a hidden layer with $M = 14$ neurones;
- an output layer with only one neurone.

The optimal number of hidden neurons was found by running the learning procedure several times with different M-values; just one output neuron is needed for computing one single temperature as output. Since the range of temperatures over one year is rather large (from -2°C to 35°C, approximately), it is very difficult for a single neural network to reproduce the observations throughout the entire year. As a consequence,

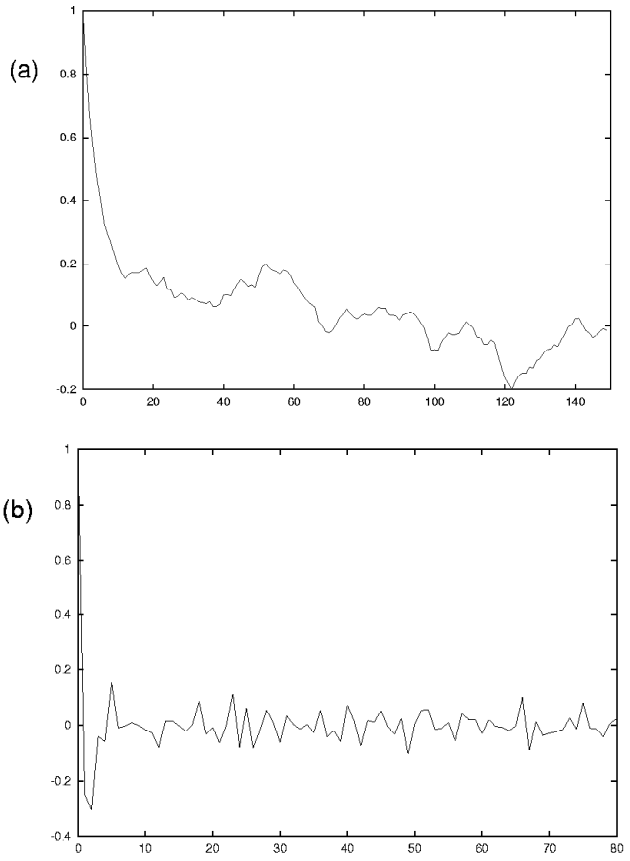

Fig. 5. Time evolution of the autocorrelation function for: **(a)** the original timeseries of the first season ECMWF forecast; **(b)** the timeseries of the first differences for the same case. The time is measured in days; the functions are calculated over one of the four grid-points surrounding GdC station.

Table 4. Computational seasons

Season	Beginning	End
First	21 May	20 September
Second	21 September	30 November
Third	1 December	28 February
Fourth	1 March	20 May

the year was divided into four "computational seasons", the temperatures of each one falling into a reduced range. For each one of these periods, defined in Table 4, we have a different network. Although the architecture is almost the same, there are differences among the four networks in:

- the number of patterns P;
- the constants:
  $I_s = \max\{To\}_s - \min\{To\}_s, m_s = \min\{To\}_s$
  $(s = 1, 2, 3, 4)$,

where $\{To\}_s$ indicates the set of the observation data To_i belonging to the season s. Each season has different maximum

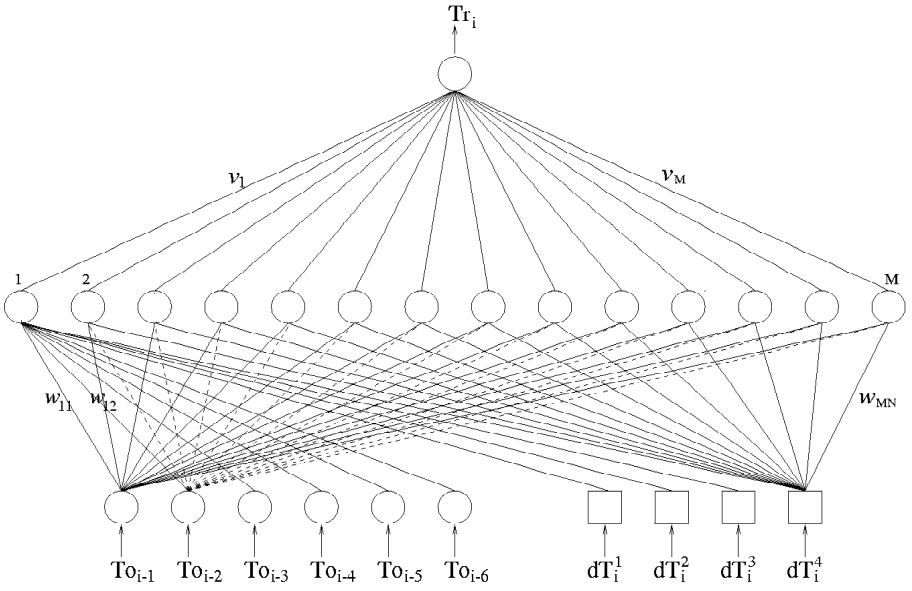

Fig. 6. Architecture of the multi-layer neural network for temperature forecast.

and minimum values. Since the output of the neural network has a value between 0 and 1, we must multiply by I_s and add m_s to the output.

Each input neurone is connected to the neurones of the hidden layer by "synaptic" interaction . The hidden layer's k-th neurone receives a signal equal to $\sum_{j=1}^{N} w_{kj} x_j^i$, and gives as output the signal:

$$h_k^i = \sigma_1\Big(\sum_{j=1}^{N} w_{kj} x_j^i - \theta_k\Big), \tag{12}$$

where θ_k is a threshold and $\sigma_1(t)$ is the nonlinear input-output function (logistic function) of the neurones:

$$\sigma_1(t) = \frac{1}{1+\exp(-\lambda_1 t)}. \tag{13}$$

The output of the hidden layer h_k^i is a nonlinear, increasing function of the signal $\sum_{j=1}^{N} w_{kj} x_j^i$, so that the k-th neurone becomes active if the input is larger than a certain threshold θ_k. The output of the single neurone of the output layer is obtained in an analogous manner. Synaptic potentials $v_1, ..., v_M$ connect the neurones of the hidden layer to the output neurone in such a way that the response of the neural network is:

$$\begin{aligned} Tr_{i,s} &= \sigma_2\Big(\sum_{k=1}^{M} v_k h_k^i - \theta_0\Big)\cdot I_s + m_s = \\ &= \sigma_2\Big(\sum_{k=1}^{M} v_k \sigma_1\Big(\sum_{k=1}^{M} w_{kj} x_j^i - \theta_k\Big) - \theta_0\Big)\cdot I_s + m_s, \\ s &= 1, 2, 3, 4. \end{aligned} \tag{14}$$

The best approximation of f (as defined in Eq. 8) is found through minimization of the error (9), performed by changing the values of the synaptic weights, the thresholds, and the parameters of the logistic functions.

In the learning phase, in order to find the global minimum, a Monte Carlo method was used to explore all the possible values of the free variables $\boldsymbol{W} = (w_{1,1}, ..., w_{M,N}, v_1, ..., v_M, \theta_1, ...\theta_M, \theta_o, \lambda_1, \lambda_2)$ belonging to a discrete bounded set. The Simulated Annealing algorithm (Carlson et al., 1996; Cohen et al., 1997) was adopted as the more effective Monte Carlo method under the conditions of our work. The Simulated Annealing skill in determining the global minimum was theoretically stated in the Geman and Geman theorem (Geman et al., 1984) and has been numerically checked many times (Feng and Qian, 1993). When applying the Simulated Annealing method, sometimes a practical problem is created by the modest algorithm convergence speed. In our work, however, the convergence speed was found to be adequate for operating efficiently.

One of the principal reasons for using of the Simulated Annealing method is the stability with respect to the choice of random initial conditions. This aspect was carefully tested during the application of the method (see Appendix A). In fact, the possibility of falling into a secondary minimum is the source of many difficulties encountered in applications.

The results of the application of the above described network to the temperature of the GdC station can be quantified by means of the learning and testing errors, defined by Eq. (9); note that this measure of the error is homogeneous with the MAD used for the standard methods. The expressions for the learning and testing errors for each season are:

$$E_L = \frac{1}{P}\sum_{i=1}^{P} |To_i - Tr_i|, \tag{15a}$$

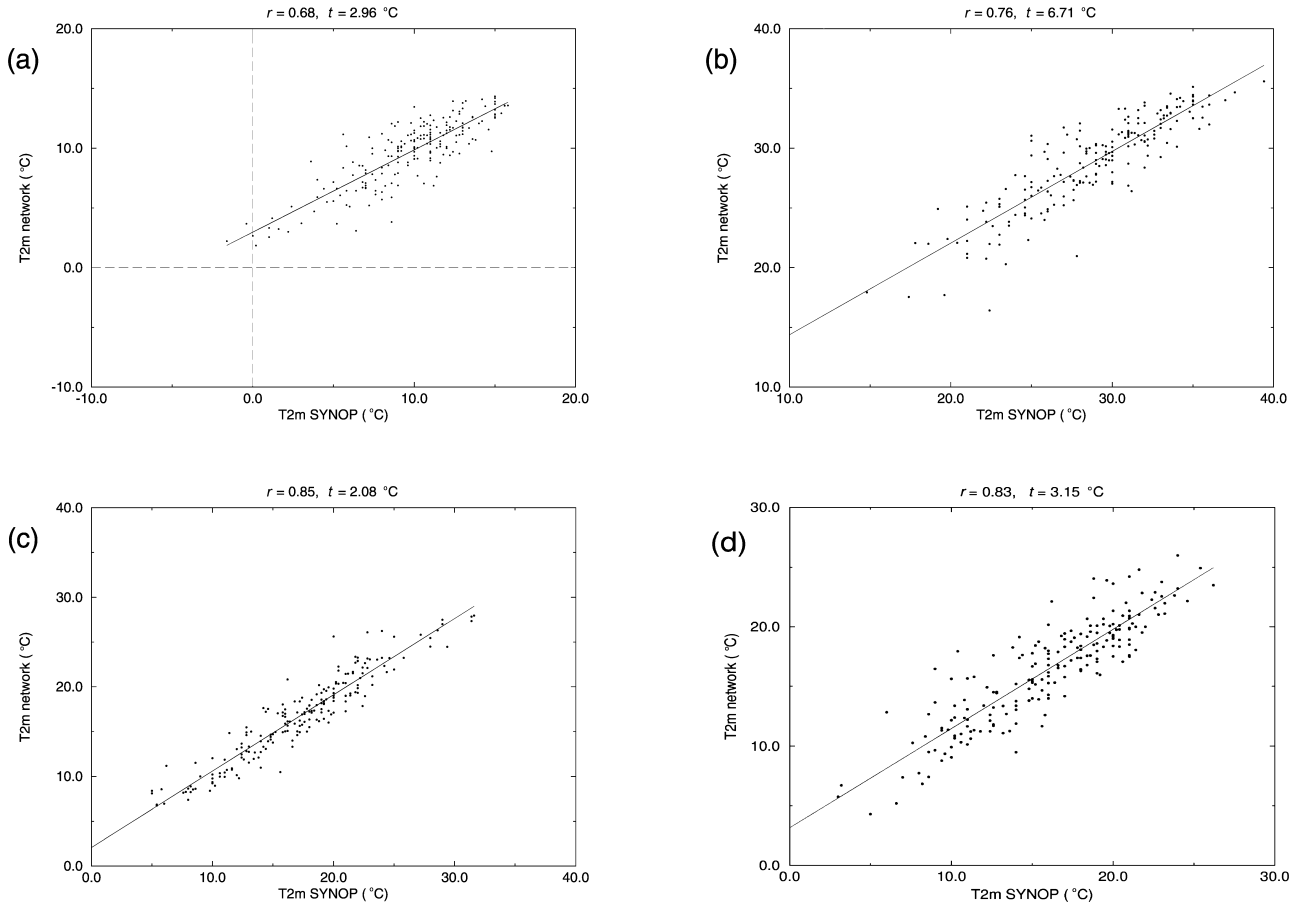

Fig. 7. Scatter-diagrams of the neural network T2m versus GdC T2m: **(a)** third season; **(b)** first season; **(c)** second season; **(d)** fourth season. The slope and the intercept of the regression line are labeled with r and t, respectively.

$$E_T = \frac{1}{P_T} \sum_{i=1}^{P_T} |To_i - Tr_i|, \tag{15b}$$

where P and P_T are the number of patterns used in the learning and testing processes, respectively. The algorithms produced the following results (Table 5):

- First network. In this case we obtained the smallest error of all the networks: $E_L = 1.3$°C with $P = 600$; $E_T = 1.4$°C with $P_T = 200$;
- Second network. Although this period is the longest one and contains more fluctuations than the others, we obtained similar results: $E_L = 1.6$°C with $P = 900$; $E_T = 1.6$°C with $P_T = 200$. We also verified that when increasing the number of elements for the learning set, the testing error tends toward the learning error;
- Third network. In this case $E_L = 1.4$°C with $P = 500$; $E_T = 1.4$°C with $P_T = 200$.
- Fourth network. This network is characterised by the longest convergence-time to the minimum in the learning process: $E_L = 1.6$°C with $P = 600$; $E_T = 1.7$°C with $P_T = 200$.

Figures 7a–d show, for visual comparison with Fig. 3, the scatter-diagrams for the testing set for each computational season. The MAD of these four neural networks oscillates between 1.4°C and 1.7°C. Further improvement can be obtained by changing the input and using moving averages over the series of temperatures at the four grid points and the GdC station. This will be the object of future work on the implementation of NN for service operations.

A linear neural network is also applied using the identity function, as an input-output function (Table 5). For all seasons the error is about one degree larger than in the nonlinear case.

A comparison of the neural network performance with that of standard methods (see the annual MAD values for GdC) shows that the nonlinear NN average testing error over a year (1.5°C) is smaller (Table 5). In particular, the average testing error is significantly less than the MAD given by the Kalman filter, which turns out to be the most powerful standard method (see Sect. 4). This is also true if the standard techniques are applied for the different computational seasons. In fact, except for the second (anomalous) computational season, the error of NN is consistently less than that of KF.

Table 5. Statistical comparison SYNOP T2m observations and ECMWF forecast, post-processed using various standard techniques and the linear and nonlinear NN for the different computational seasons. All the numbers in the table express MAD in °C

Post-processing method	First season	Second season	Third season	Fourth season	Average
None	2.0	1.1	1.6	1.5	1.5
Vertical interpolation	3.1	1.5	1.4	1.8	1.9
Kalman filter	1.8	1.3	1.5	1.9	1.6
Linear NN	2.5	2.1	1.9	2.4	2.2
Nonlinear NN	1.4	1.6	1.4	1.7	1.5

Table 5 clearly shows strong non-homogeneities. For example, the performance of the numerical model appears particularly good in the second computational season, while the performance of all the post-processing methods, including NN, remains comparable to that of the other seasons. We did not investigate further this point, since the ECMWF model is not the one that will be used in future service operations.

6 Conclusions

In this paper we present the results of a comparative study concerning the use of linear and nonlinear statistical techniques in the post-processing of gridded surface temperatures provided by numerical forecast models.

Among standard techniques the Kalman filter is the method yielding the least bias. This is easily minimised by the standard methods, to the contrary of the random error component of the MAD. As a consequence, we use the MAD to measure the neural network performance.

Linear and nonlinear NN are trained and tested with ECMWF T2m forecast and T2m observations over the GdC station, utilising an innovative learning algorithm, i.e. the minimisation of the MAD by means of a Monte Carlo method (based on Simulated Annealing). The major advantage of this procedure is its ability to reach the absolute minimum of the error function, by avoiding the "trapping" in some local minimum. On the contrary, NN based on more traditional back-propagation learning algorithms can converge only to the local minimum closest to the initial state: this can generate undesirable sensitivity to the learning initialisation, with consequent difficulties in achieving an optimal training. The main drawback of the Simulated Annealing is its sometimes slow convergence to the minimum of the error, making it necessary, in some cases, for the introduction of more complex methodologies, as, for example, in Carlson et al. (1996) and in Cohen et al. (1997), to provide a faster convergence. In our case, however, no problem of convergence was encountered.

Nonlinear NN post-processed temperature is characterized by an annual MAD error of 1.5°C, significantly smaller than the one associated with the best standard method, the Kalman filter (Table 5).

These results are encouraging for the development and application of our NN to the down-scaling problem in service operations. Moreover, we foresee further improvements. First, the network prediction error can be further reduced by means of the so-called compensation method. This consists of introducing into the hidden layer one or two additional neurones for any pattern of the training set that shows a non-decreasing contribution to the global error. Moreover, the data sets can be improved using a higher-resolution model output, more dense observation networks, and other sources of information. Further studies will be devoted to the above problems.

Appendix A

The Monte Carlo method used in our study works in the space of all the parameters $\boldsymbol{W} = (w_{1,1}, ..., w_{M,N}, v_1, ..., v_M, \theta_1, ..., \theta_M, \theta_0, \lambda_1, \lambda_2)$ by changing one component at each step, in agreement with a transition probability. This probability is given by $e^{-\beta \Delta E}$, if $\Delta E > 0$, and 1, if $\Delta E < 0$, where ΔE is the change in the error function when one component of the vector is changed. We obtain $\boldsymbol{W}'$ from $\boldsymbol{W}$ by taking randomly one component of $\boldsymbol{W}$, W_i, and adding or subtracting a minimal increment h with probability 1/2: $W_i' = W_i \pm h$. ΔE is then equal to $E(\boldsymbol{W}') - E(\boldsymbol{W})$. Thus, the phase space of all the possible weight vectors $\boldsymbol{W}$ is discretised and the Markov chain defined by this random process asymptotically visits all the states, since it is ergodic. The parameter β governs the probability of the amplitude of the jumps for the components of $\boldsymbol{W}$.

Appendix B

We offer here a schematic description of the Simulated Annealing method employed, in order to find the global minimum of the learning error. Simulated Annealing is a Monte Carlo method in which the transition probability changes in the process. In our case:

– If initially the weights, the thresholds, and the sigmoid parameters vector are inside the domain $\Gamma = [-10, 10]^m$, it will never go out during the evolution;

– The interval $I = [-10, 10]$ has been divided in $N_1 = 1000$ small sub-intervals and the elementary step h of the Monte Carlo method is: $h = |I|/N_1$;

– The configuration vector $\boldsymbol{W}$ contains the weights, the thresholds and the parameters λ_1, λ_2 of the sigmoid functions, σ_1, σ_2; so it can be represented as:

$$\boldsymbol{W} = (w_{1,1}, ..., w_{M,N}, v_1, ..., v_M, \theta_1, ..., \theta_M, \theta_0, \lambda_1, \lambda_2) \quad \text{(B1)}$$

– All of the initial values of the components of this vector are chosen randomly with uniform distribution in the interval [-10,10]. The optimal initial value of λ_1 and λ_2 is 1.4;

– A component of $\boldsymbol{W}$ is chosen with uniform probability;

– The new value of the j-th component W'_j is selected, with a probability of 1/2, between the possible W'_j values, obtained from W_j, either by an increase or decrease in h : $W'_j = W_j \pm h$;

– The fluctuation $\Delta E = E(\boldsymbol{W'}) - E(\boldsymbol{W})$ is computed and we examine the possibility of replacing W_j with W'_j. There are two cases:
a) If $\Delta E < 0$, then the new value of W_j is accepted: $W_j \rightarrow W'_j$;
b) If $\Delta E \geq 0$, $W_j \rightarrow W'_j$ with probability:

$$\frac{e^{-\beta \Delta E}}{1 + e^{-\beta \Delta E}}. \quad \text{(B2)}$$

This is done by extracting a random number with uniform probability in the interval [0,1]. If this number is less than (B2), then the transition is accepted; otherwise, W_j remains as it is. From (B2) it is evident that comparatively smaller values of β allow for bigger jumps in ΔE. In this sense, $T = 1/\beta$ can be regarded as a "temperature": a "hot" system jumps out from local minima, visiting all the parameter space, while a "cold" system rapidly converges to the nearest local minimum.

– The Simulated Annealing scheme is applied. According to this scheme, the value of b increases logarithmically: $\beta(n) = 1 + \log n$, where n is the number of times the process "visits" all $\boldsymbol{W}$ components. This "slow cooling" lets the system fall in the neighbourhood of the global minimum before converging to this one;

– This procedure is repeated until the learning error is less than a suitable value (2°C in our case). The logarithmic cooling may cause a very slow convergence rate, so that, generally, the control of the velocity of this process is a key practical issue in applying this method. Some authors deal with this problem in recent works (Carlson et al. 1996; Cohen et al. 1997). In our case, however, the network convergence is fast.

Acknowledgements. Part of this work has been made possible by the support of the Italian Space Agency (ASI) under grant N. ARS-99-60. We are grateful to DSTN (in particular SIMN and Ufficio Compartimentale della Puglia) for having provided us with numerical and observational data-sets, as well as technical support. We thank Prof. S. Albeverio for useful discussions and two anonymous reviewers for constructive criticism that helped improving the paper.

References

Cacciamani, C.: Extreme daily temperature forecasts in the Emilia Romagna region (North Italy) by Kalman filter technique of direct numerical weather prediction model outputs, LAM Newsletter, 22, 196–209, 1993.

Carlson, P., Danielsson, M., Dejardin, M., Jon-And, K., and Sjölin, J.: Simulated annealing and neural network as alternative methods for nonlinear constrained optimization, Nucl. Instrum. Methods Phys. Res., Sect. A, 381, 152–156, 1996.

Cohen, B., Saad, D., and Marom, E.: Efficient training of recurrent neural network with time delays, Neural Networks, 10, 51–59, 1997.

Deidda, R., Marrocu, M., and Speranza, A.: Meteorological forecasting for the European Southern Observatory in Chile, Nuovo Cimento C, 23, 1–23, 2000.

Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distribution and bayesian restoration of image, IEEE Trans. PAMI, 6, 721–741, 1984.

Hall, T., Brooks, H. E., and Doswell, C. A.: Precipitation forecasting using a neural network, Wea. Forecasting, 14, 338–345, III, 1999.

Hertz, J., Krogh, A., and Palmer, R. G.: Introduction to the Theory of Neural Computation, Addison-Wesley, 321 pp, 1993.

Hsieh, W. W. and Tang, B.: Applying neural network models to prediction and data analysis in meteorology and oceanography, Bull. Amer. Meteor. Soc., 79, 1855–1870, 1998.

Koizumi, K.: An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network, Wea. Forecasting, 14, 109–118, 1999.

Marzban, C. and Stumpf, G. J.: A neural network for damaging wind prediction, Wea. Forecasting, 13, 151–163, 1998.

Massie, D. R. and Rose, M. A.: Predicting daily maximum temperatures using linear regression and Eta geopotential thickness forecasts, Wea. Forecasting, 12, 799–807, 1997.

Mhaskar, H. N: Neural Networks for optimal approximation of smooth and analytic functions, Neural Computation, 8, 164–177, 1996.

Narasimhan, R., Keller, J., Subramaniam, G., Raasch, E., Croley, B., Duncan, K., and Potter, W. T.: Ozone modeling using neural networks, J. Appl. Meteor., 39, 291–296, 2000.

Shao, J.: Improving nowcasts of road surface temperature by a back-propagation neural network, Wea. Forecasting, 13, 164–171, 1998.

Tang, B., Hsieh, W. W., Monahan, A. H., and Tangang, F. T.: Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial pacific sea surface temperatures, J. Climate, 13, 287–293, 2000.