Temperature forecast and dome seeing minimization

I. A case study using a neural network model

F. Buffa and I. Porceddu

Stazione Astronomica di Cagliari, Str. 54, Loc. Poggio dei Pini, I-09012 Capoterra (CA), Italy

Abstract. Dome seeing may strongly deteriorate the final sharpness of a point source astronomical image, reducing its Image Quality. Both the telescope enclosure and the mirrors may contribute to the dome seeing, if air convection is induced by differences of temperature between them and external air. The prediction of the external air temperature with respect to a given time interval allows one to preset in advance the air conditioning temperature value in the telescope enclosure. With the aim to study the neural networks capabilities and limits to make short term temperature prediction, a few case studies have been carried out by using an autoregressive neural network model. The actual goal is to understand if and with which constraints a neural network model can actually be used in a NTT-like dome (i.e. telescope is in open air when observing and heat sources are highly controlled when close or inside the telescope's dome) for steering the daytime air conditioning system. We do not present any interface with an actual telescope: this paper presents a feasibility study about the forecasting methodological approach rather than its operational application to a specific telescope.

The results show that on site output prediction of a neural network are competitive with respect to a linear prediction approach.

Key words: turbulence — methods: data analysis — methods: statistical — site testing

1. Introduction

Astronomical sites are chosen with reference to the local climate and seeing. The latter can be monitored and it can provide a key for understanding the general quality of the site: but the intrinsic excellence of a site is very often deteriorated by the astronomical facilities themselves, which

Send offprint requests to: I. Porceddu

introduce an additive level of seeing (Coulman 1985). The natural seeing, which is an intrisic site characteristic, has been normally considered the major limiting factor for reaching the telescope virtual top performance. There is now a general awareness about the crucial rôle which is played by the man induced seeing, and the need for minimizing it in order to get a quasi diffraction limited condition.

Several image quality series of data demonstrate that mirror seeing is a primary source of the point spread function broadening. The seeing contributed by a 3.5 m primary mirror warmer than the air outside the dome can range from 0.4 to 0.6 arcsec per Celsius degree (Zago 1986; Racine et al. 1991), a value which is comparable or higher than the natural seeing in good astronomical sites. Several laboratory studies of image deterioration by convection (Lowne 1979; Iye et al. 1991) confirm that microthermal activity grows significantly as soon as the temperature gradient between dome air and primary mirror is positive. The dome itself is a major source of seeing degradation (Bely 1987), and the seeing contribution is highly variable, depending on the telescope enclosure design. The Canadian-France Hawaii telescope and the Nordic Optical Telescope enclosures (Racine et al. 1991; Vernin & Muñoz-Tuñon 1992, 1994) show a low impact on the overall man induced seeing, while the ESO 3.6 m telescope is more critical (Faucherre 1995). Newest telescopes are designed and built taking care of this specific problem, and thermal analysis is modeled as specific item of the global dome design. Therefore, the telescope thermal environment needs a specific approach both in the dome design phase and in the telescope operational methodology. Heat sources inside and close to the observing floor are thermally controlled and air conditioning in the telescope enclosure is a normal procedure for telescopes committed to top performances.

This paper presents a feasibility study of temperature forecast in support to active air conditioning of a telescope's dome as part of a research activity which is done within the meteorological support to the italian national

telescope Galileo (TNG), located at the Observatorio Roque de Los Muchachos (ORM), in the Canary Islands. A neural network modelling is presented as compared to classic linear filter algorithm.

2. The need for predicting meteorological parameters

The monitoring and forecast of meteorological parameters has been rarely used by astronomers: e.g., cloud cover was checked visually and water vapour content was merely used. There is now an ever growing interest and need for an extensive application of atmosphere soundings to astronomical activities; newest telescopes are demanding on this side, being devoted to do queued observations. The Gemini and VLT projects, e.g., will adopt an operational model that will include at least 50% of the allotted time as queued: they will take advantage of an active scheduling of the observing facilities. For all of these high tech telescopes and instruments, weather and site monitoring will be very important in the selection of proposals for the coming nights. Therefore prediction algorithms are under contract for ESO/VLT for clouds and water vapour, aiming at developing custom made daily weather forecasts for Paranal: although that site must be seen as a special situation, this is the expected trend for the future ground based observatories.

2.1. Medium Range Forecasts and Local Area Models

Astronomers can in principle get support from the local weather forecast services. The European Center for Medium-Range Weather Forecast (ECMWF), e.g., predicts the behaviour of the atmosphere in the medium-range up to ten days ahead, providing numerical weather prediction maps on a synoptic scale every 12 hours. National Meteorological institutions receive these products through a dedicated telecommunications network and can further process the ECMWF output, providing an higher horizontal resolution in a short range prediction. The running atmospheric models are extremely demanding from the computing resources side: the recent availability of more powerful computers led the ECMWF to definitely increase the actual grid resolution to about 60 km. The latest numeric model also improves the vertical discretization, with 31 baric levels available instead of 19: the forecast standard error for temperature parameter in the medium range ECMWF model is well higher than 2 degrees. In case of the ECMWF model, the horizontal discretization grid as well as the error on the forecast of meteorological parameters is too high for astronomical purposes.

In case of short range prediction, a local area model (LAM) may be used, which takes into account more detailed information, mostly local sampled data and an higher resolution orographic model. The output of a LAM is "limited", both in time and space, being valid in the range from several hours to about two days and within a limited spatial area, but can offer a forecast map having grid points separated only 30 to 15 km and a prediction error on temperature as low as 2 Celsius degrees RMS on a 24-hours prediction map. This is true for an operational LAM running on a rough orography in a Mediterranean area (Picchedda and Chessa, private communication); a much better result is expected for a more stable region as the La Palma island is. The laminar flux due to the presence of the Atlantic Ocean may actually smooth the complex behaviour which is found to be the dominant characteristic on an european region. Therefore a LAM could virtually perform with a resolution approaching the Celsius degree, but the limit is still on the computing side.

As european development, the project High Resolution Limited Area Model (HIRLAM) is the most challenging operational example for a LAM. It is a project cooperation among several european countries, and the reference version is maintained at the ECMWF.

3. Temperature forecast and neural networks

As we have seen, the Medium-Range and Local Area prediction models may actually provide a set of predicted weather related parameters which could support an attempt to active scheduling of the observations. And the atmospheric mesoscale numerical prediction may be a key approach to seeing nowcast too, when service observing scheduling is needed (Bougeault et al. 1995). Nevertheless, the absolute errors which are shown by operational high resolution prediction models are much higher than the astronomical constraints. Moreover, the computing and human resources which are required for developing, servicing and operating Local Area Models gives them a very low flexibility.

Medium-Range and Local Area prediction models require a closed set of appropriate physical laws expressed in mathematical form, suitable initial and boundary conditions and an accurate numerical method of integrating the system of equations forward in time. Within the framework of a very short time range prediction, several different methodological approaches can be investigated, which do not require the knowledge of underlying physical laws. A standard *black box* configuration, where an input set of data is being processed and an output result is produced, can be a correct layout for our purposes.

Neural networks mime a *black box* model: this is accomplished using time series of meteorological parameters as an approach to forecast dynamical processes. Murtagh & Sarazin (1993) approached the temperature and seeing prediction by using a neural network model. A similar approach has been used here for temperature prediction as a feasibility study in support of a generic dome thermal environment control system. The neural network is a non linear approach to data series treatment which is highly

flexible, presents a low-cost from the computing point of view and may provide excellent results.

3.1. Multilayer Feed Forward Neural Networks

A Neural Network (NN) (e.g. Hecht-Nielsen 1991; Hertz et al. 1991) is a flexible mathematical structure which is capable of identifying complex non-linear relationships between input and output data sets. For these reasons NN models have been found useful and efficient, particularly in problems for which the characteristics of the process are difficult to describe using physical equations. NN are powerful objects having inference and generalisation capabilities; in fact, a NN which has been trained with a representative number of examples of a given process is able to extrapolate states not present in example data set.

The network topology we chose is the usual feed forward (FF) (Fig. 1a), which has been found to have high performances in input-output function approximation (Elsner 1992). In a typical three-layer FF NN the first layer connects the input variables and is called the input layer. The last layer connects the output variables and is called the output layer. Layers in-between the input and output layers are called hidden layers; there can be more than one hidden layer. The processing unit elements are called nodes (Fig. 1b): each of them is connected to the nodes of neighbouring layers. The parameters associated with each of these connections are called weights. All connections are "feed forward"; that is, they allow information transfer only from an earlier layer to the next consecutive layers. Nodes within a layer are not interconnected, and nodes in non adjacent layers are not connected. Each node j receives incoming signals from every node i in the previous layer. Associated with each incoming signal x_i is a weight w_{ji}. The effective incoming signal s_j to node j is the weighted sum of all incoming signals:

$$s_j = \sum_{i=0}^{n} w_{ji} x_i \qquad (1)$$

where $x_0 = 1$ and w_{j_0} are called the bias and the bias weights, respectively. The effective incoming signal, s_j, is passed through a non- linear activation function (called also transfer function or threshold function) to produce the outgoing signal (h_j) of the node. The most commonly used activation function is the sigmoid function. The characteristic of a sigmoid function is that it is bounded above and below, it is monotonically increasing, and it is continuous and differentiable everywhere. The sigmoid function we used is:

$$h_j = \tanh s_j = 1 - \frac{2}{\exp 2s_j + 1} \qquad (2)$$

in which s_j ranges from $-\infty$ to $+\infty$, but h_j is bounded between -1 and 1. In our scheme only signals processed in hidden units are passed through the activation function.
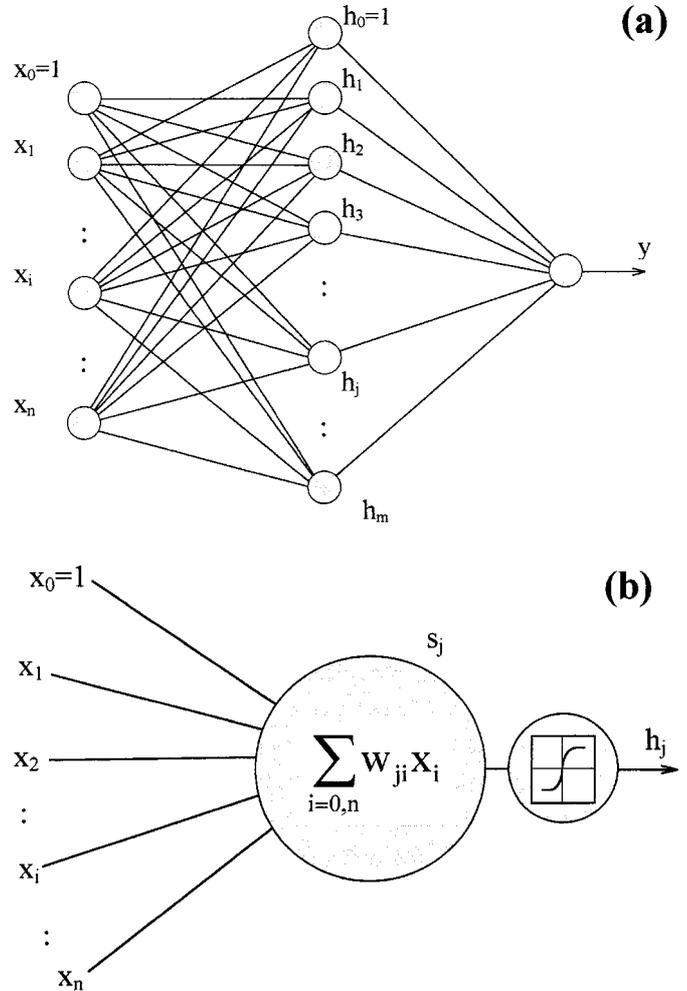
Fig. 1. The **a)** draw shows a three-layers feed forward neural net. A processing unit element is drawn in **b)**

To achieve weights optimisation a large number of "training" algorithms exists, each of which is characterised by a learning law that will drive the weight matrix to a location that yields the desired network performance. Due to its rapid convergence properties and robustness, we chose a Levenberg-Marquardt algorithm (Nørgaard 1995) as engine in the minimization procedure.

In order to avoid overfitting, the network's performances are usually measured using two different data set: the training set and the validation set. While the training set is used directly to train the network, the validation set is used only for the evaluation process. Another way to increase network's performances consists in removing of idle connections (pruning): one of most popular method is the so called "brain damage", which needs a retraining after each trial unit damage.

3.2. Autoregressive approach to temperature time series treatment

The temperature data series we used come from the Carlsberg Automated Meridian Circle (CAMC) automatic weather station, which provides several meteorological parameters with a 5 minutes time interval. The meteorological transducer for temperature monitoring is an AD590K, which can operate from −55 to +150 °C; it is positioned on a mast head at 10.5 metres above ground. In this paper the CAMC site is supposed to be representative of the temperature variations which can be found at the ORM, where both CAMC and TNG are operated.

In this paper we present a preliminary study of temperature forecast at ORM site using linear and non-linear autoregressive models.

These statistical models based on the original idea due to Box and Jenkins (BJ) represent the fundamental approach in system identification and time series studies since the early '70s (Box & Jenkins 1970). The basic idea of BJ approach is that if a system is (partially) governed by deterministic rules, the future behaviour may be in some extent modelled from the behaviour of the past states.

The *classic* linear autoregressive moving average with exogenous inputs (ARMAX) approach consists in modelling the (deterministic part of a) generic variable of time $T(t)$ at time $t = t_i + \Delta t$ (in our case T is the temperature) by the function $T'(t)$ defined as:

$$T'(t_i + \Delta t) =$$
$$\mathcal{F}[T(t_i), T(t_i - \Delta t), ..., T(t_i - k_T \Delta t);$$
$$P(t_i), P(t_i - \Delta t), ..., P(t_i - k_P \Delta t);$$
$$E(t_i), E(t_i - \Delta t), ..., E(t_i - k_E \Delta t); W)]$$

where $\mathcal{F}$ is a linear function (linear combination), T, P and E are the vectors containing the fitting regressors, W is a vector containing the weights of the linear combination, Δt is the time lag and k_T, k_P and k_E are the number of past regressors used for each estimation. $E(t_i) = T(t_i) - T'(t_i)$ is a recurrent (dynamic) term containing error (noise) propagation estimation. In our case we choose pressure as exogenous variable P. The choice of pressure has been suggested by an analysis of the cross-correlation structure between temperature and all the others meteorological variables collected by CAMC. The exogenous regressor P is omitted in moving average autoregressive (ARMA) scheme, while in the autoregressive scheme (AR) both P and E are omitted. In AR models, fixing $k_T = 1$, one obtains that the unique weight W_1 in linear combination is directly related to the autocorrelation factor of variable T at time step Δt.

Many authors assert that non-linear approach allows modelling of complex dynamics in climatic variables. For such a reason the classic BJ model may be reinterpreted

from a neural point of view (Fig. 2), the most important difference is that $\mathcal{F}$ changes in a non-linear function realised by the neural network giving the NLAR, NLARMA and NLARMAX models.
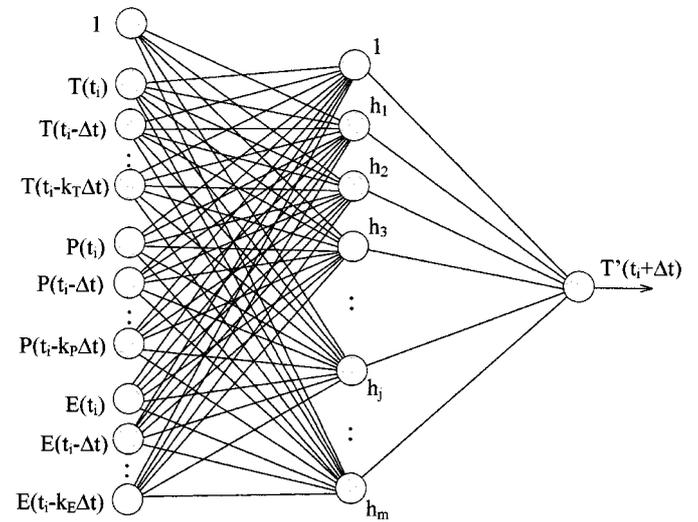
Fig. 2. Implementation of an ARMAX model thru a NN scheme. Input nodes are temperature (T), pressure (P) and error propagation estimation (E) values for different time lags. k_T, k_P and k_E are the number of past regressors used for each estimation

4. Results

In this section we analyze the output results of two distinct series of runs performed with two different approaches; a standard linear BJ filter (labelled L, which stands for *linear*) and an autoregressive NN (N, for *non- linear*). The data set we used for both series has been previously standardized (i.e. $\mu = 0$ and $\sigma = 1$), the series corresponds to the period April-May 1993 for the training data set and to the period May-June 1993 for the validation data set.

We suggest that both training and validation sets must be, in this particular kind of analysis, short and adjacent series; in fact, trends and variability features, having strong seasonal behaviour may be reproduced more easily (Aussem 1995). This kind of seasonal features may result confused in too long series. On the contrary this approach lacks in forecasting of "out of trend" events, such as strong pressure variations related to the approaching of climatic fronts, which are difficult to predict with a merely statistic approach.

In Table 1 results corresponding to L-runs are reported. A standard least squares algorithm has been used to refine fitting parameters on training set. Note that reported standard errors are evaluated on validation data set. Results of run L1 seem to be quite promising, confirming the strong autocorrelation component contained

in temperature time series. Runs L2 (Fig. 3) and L3 represent an attempt to forecast temperature with a time lag equal to 6 hours: note that the purely autoregressive model give better performances than the ARMAX one. In order to justify this fact, two explanations are possible:
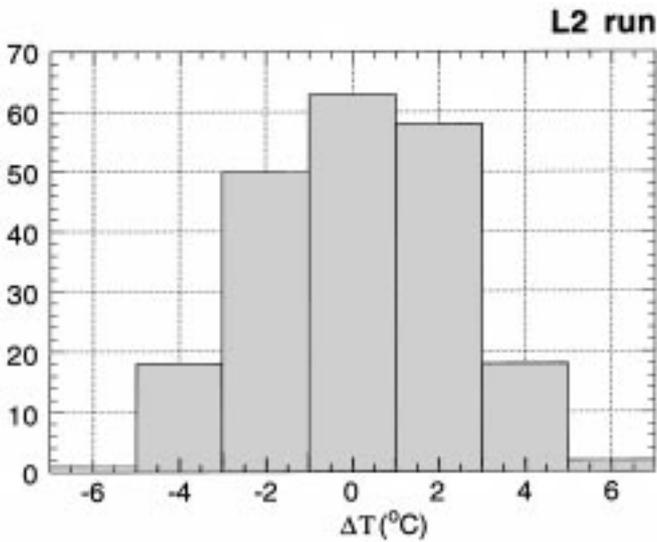

Fig. 3. Binned distribution of differences between observed and predicted temperatures for L2 run

- the contribution of pressure or pressure gradient in terms of additional information is poor or even useless;
- pressure and temperature are related in a non-linear way.

The first run of N type represents a NLAR attempt to perform a forecast with a time lag equal to 1 hour. Note that we found better performances using few units (only one in the shown case) in the hidden layer.

In runs from N2 to N5 we checked the NN capabilities with a lag of 6 hours. Runs N2, N3 and N4 use NLARN, LARMA and NLARMAX schemes respectively, while run N5 differs slightly from the previous ones. In fact we used 8 units in hidden layer and then we refined the network architecture with a pruning strategy called "Optimal Brain Surgeon" (Nørgaard 1995). After this procedure about 70% of unit connections have been pruned, confirming that in the present case a low number of units in hidden layer increases performances and network stability. Only results from run N2 are shown (Fig. 4) being representative of the sub-set with time lag equal to 6 hours.

In Table 2 the characteristics of six runs and estimation errors for the validation data set are summarised. BJ and NN predictions performances may be compared with predictions obtained with *Carbon Copy* technique. This
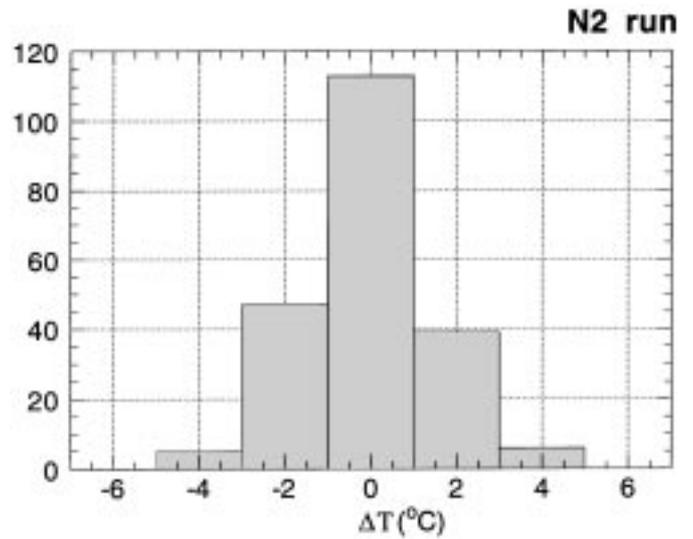

Fig. 4. Binned distribution of differences between observed and predicted temperatures for N2 run

method assumes that the value of T at time t is equal to T at time $t - 24$ (hours).

The estimated confidence level of the Carbon Copy analysis we carried out is about 2.1°C: this value limits the actual prediction capability to a time interval of 12 hours (Table 2). Figure 5 gives a different criterion to evaluate predictions in AR and NLAR approaches showing the correlation coefficients ρ between validation set and corresponding models versus forecast time step. Figure 5 also shows that both models are highly correlated with observed series, confirming the correctness in the choice of an autoregressive model. Moreover, NN shows greater correlation values than BJ.
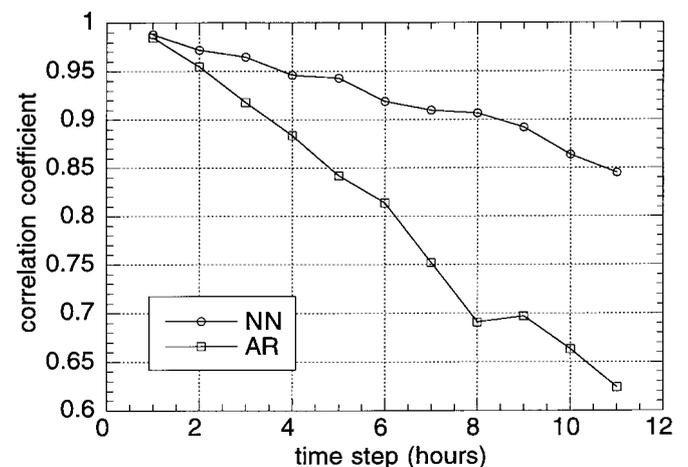

Fig. 5. Correlation coefficient between actual and predicted values as a function of prediction time in AR and NLAR cases

Table 1. Parameters and errors in validation sets of L-runs; regrs stands for regressor type, pts is the number of samples in training and validation set, while RMS is the evaluated standard error

Run	Type	k_T/k_P	lag (hours)	regrs	pts	RMS (°C)
L1	AR	1	1	T	1320	0.7
L2	AR	1	6	T	220	2.2
L3	ARMAX	2/2	6	T/P	220	3.1

Table 2. Parameters and errors in validation sets of N-series; regrs stands for regressor type, pts is the number of samples in training and validation set, while RMS is the evaluated standard error, nhu is the number of hidden units. The label (*p*) means that a pruning algorithms has been used (see text). Symbol (*) refers to NN topology before pruning

Run	nhu	Type	k_T/k_P	lag(hours)	regrs	pts	RMS (°C)
N1	1	NLAR	5	1	T	1320	0.6
N2	2	NLAR	10	6	T	220	1.3
N3	1	NLARMA	5	6	T	220	1.4
N4	1	NLARMAX	5/5	6	T/P	220	1.4
N5	8(*)	NLARMAX(*p*)	5/5	6	T/P	220	1.5
N6	1	NLAR	4	12	T	110	2.1

5. Discussion

Time series of climatological and meteorological parameters show a dynamic trend where the prediction accuracy decreases as the prediction time increases. Both NN and BJ models confirm this behaviour, although the neural network approach is slightly more efficient when predicting temperatures; moreover, the gap grows up as the time lag increases (Fig. 5).

Table 2 shows that N2, N3, N4 and N5 approaches are quite similar from the standard error point of view. On one hand the stochastic component represented by MA could be negligible; on the other hand the exogenus X component associated to the variable pressure seems not to include information which are able to improve the temperature prediction capability. And those case studies carried out using the whole set of meteorological parameters collected by CAMC show that prediction output is worse when one of them is added to the input set; in this case we may suppose that the studied case is purely autoregressive.

A further facet of the NN behaviour is that best performances are obtained for a topology with a minimum number of hidden units ($nhu = 1, 2$); note that the validation set is used both for evaluating performances as well as for optimising the nhu value. During calibration phase, whenever the number of hidden units has been increased, we found a better performance of the model with respect to the training set and a worse feedback when using the validation set. This could be explained in terms of a physical behaviour characterized by a certain degree of linearity. In fact, a NN topology with a single or 2 hu mimes a quasi-linear approach, possibly related to an effective linear behaviour of the selected time series.

6. Final remarks

Medium and short range numerical weather prediction models are unable to satisfy the astronomical constraints in terms of both temperature forecast accuracy and spatial resolution, although a local area model specifically designed for astronomical purposes could improve the performances of a standard model. From a theoretical point of view, an *astronomical* LAM could provide an actual RMS error below the 2 Celsius degrees: nowadays this is hard or impossible to accomplish. For the neural network model we have presented here, the extimated prediction capability shows an application threshold at 12 hours, being a limit to the application of such NN topology. Among the various attempts which can be carried out in order to improve the reported results in terms of both time prediction and accuracy, a new NN topology will be tested, where the medium-range forecasts provided by the ECMWF is used as an input node together with the local monitored time series. The new topology could in principle act as a post-processing engine for the ECMWF medium scale forecast, triggered by the local weather.

The Carlsberg Automated Meridian Circle telescope, which is operated jointly by Copenaghen University Observatory, the Royal Greenwich Observatory and the Real Instituto y Observatorio de la Armada en San Fernando, kindly provided us with the time series of meteorological parameters which have been used in this paper. The friendly support we got from Bob Argyle (RGO) is highly appreciated.

Dr. Marino Marroccu from CRS4, kindly discussed with us the Local Area Model topic.

References

Aussem A., 1995, Ph.D. Thesis

Bely P.-Y., 1987, PASP 99, 560

Bougeault P., De Hui C., Fleury B., Laurent J., 1995, Applied Opt. 34, 3481

Box G.E.P., Jenkins G.M., 1970, Time Series Analysis Forecasting and Control. Holden Day - San Francisco

Picchedda G., Chessa P. (private communication)

Coulman C.E., 1985, ARA&A 23, 19

Elsner J.B., 1992, Predicting time series using a neural network as a method of distinguishing chaos from noise, J. Phys. A: Math. Gen. 25, 843

Faucherre M., 1995, The Messenger 80, 5

Iye M., Noguchi T., Torii Y., Mikami Y., Ando H., 1991, PAPS 103, 712

Hecht-Nielsen R., 1991, Neurocomputing. Addison-Wesley

Hertz J., Krogh A., Palmer R.G., 1991, Introduction to the Theory of Neural Computation. Addison-Wesley

Lowne C.W., 1979, MNRAS 188, 249

Murtagh F., Sarazin M., 1993, PASP 105, 932

Nørgaard M., 1995, Neural Network based system identification toolbox, Tech. Report 95-E-773, Technical University of Denmark

Racine R., Salmon D., Cowley D., Sovka J., 1991, PASP 103, 1020

Vernin J., Muñoz-Tuñon C., 1992, A&A 257, 811

Vernin J., Muñoz-Tuñon C., 1994, A&A 284, 311

Zago L., 1986, SPIE Proc. "Advanced Technology Optical Telescope III", Tucson, 628, 350