Fuzzy Astronomical Seeing Nowcasts with a Dynamical and Recurrent Connectionist Network

Alex Aussem †‡, Fionn Murtagh[1] ∓, Marc Sarazin †

†Very Large Telescope Division,

European Southern Observatory, Karl-Schwarzschild-Str. 2,

D-85748 Garching bei München - Germany

‡Université René Descartes, UFR de Mathématiques et Informatique,

45, rue des Saints-Pères, 75006 Paris - France

∓ Space Telescope - European Coordinating Facility,

(Affiliated to Astrophysics Div., Space Science Dept., European Space Agency)

European Southern Observatory, Karl-Schwarzschild-Str. 2,

D-85748 Garching bei München - Germany

Telephone +49 89 32006-298, Fax +49 89 32006-480, Email fmurtagh@eso.org

Abstract

We assess a neural-based method for fuzzy astronomical seeing prediction, based on known meteorological variables at the same time-point. This multiple regression, termed *nowcasting* [10, 11], will allow the modern telescopes to be preset, a few hours in advance, in the most suited instrumental mode. The data used are extensive meteorological and seeing measurements partly made at Cerro Paranal in Chile, site of the Very Large Telecope (VLT). A fuzzy correspondence analysis is carried out to explore the internal relationships in the data. Then, a *time*- and *space* -recurrent network is used in combination with a fuzzy coding approach to capture the temporal regularities of the seeing series. Such a connectionist network is endowed with an internal dynamic by means of arbitrary recurrent time-delayed connections. The performance of the model is appraised and the results are compared with the fuzzy k-nearest neighbors method.

Keywords - neural networks, nearest neighbor method, time series prediction, fuzzy correspondence analysis.

[1] ***Corresponding author.***

1 Introduction

The present study is in the context of an envisaged Astronomical Weather Station, a future interface between the observer and the terrestrial environment of the VLT observatory. Its function is to improve the quality of observations and to guarantee efficient use of the telescope time. If a reliable astronomical seeing estimate is proposed, some time in advance, then the ground-based telescope can be set in the most suited mode and allow an efficient use of the observing time. The practical experience and theoretical knowledge of the meteorologists points to the inherent difficulty of deriving rules for local seeing forecasting. Therefore, prediction is sought without reference to any underlying physical model.

Several experiments have been carried out so far [3, 9], seeking to forecast quantitative-valued seeing measurements 24 hours in advance. However, non-linear multiple regression, including the k-nearest neighbor method and connectionist networks, were found to perform poorly on the seeing time series. In fact, seeing is not easy to predict; Jorgenson et al. [6] give concerns about its *somewhat* chaotic behavior.

Therefore, a somewhat different tack is taken in this study with a view to improving our seeing prediction abilities. We split the problem into two parts: (i) the prediction of meteorological conditions one day in advance, and (ii) the *nowcasting* of seeing according to these predicted parameters. This paper focuses on the latter, namely seeing nowcasting. From the statistics viewpoint, nowcasting is multiple regression, i.e. we attempt to regress a set a meteorological variables on a seeing variable. We assume that good estimates can be provided by an *ad hoc* meteorological model.

Furthermore, an intuitively plausible and interpretation-friendly form of coding was adopted, consisting of "fuzzifying" the seeing measurements. Such coding is particularly appropriate since seeing categories ("very good", "very poor", etc.) are especially helpful for interpretation. A fuzzy correspondence analysis is used to explore the initial relationships in the data. We find that specific meteorological variables can be related to good seeing conditions *at the same time point.*

The framework pursued in this article not only relates fuzzily coded seeing measurements with current meteorological measurements, but also attempts to capture the time-local determinants of the seeing dynamic. With this objective in mind, we discuss the use of a dynamical and recurrent neural network (DRNN) to make context-sensitive nowcasts. Recurrent neural networks have emerged in the past few years as a successful tool in time series prediction [1, 2, 5, 12], process modeling and adaptive control of dynamical systems. Mutual delayed interactions between neighboring units are particularly well suited to hold relevant context information pertaining to the past input sequence. Such a powerful architecture enables the network to identify dynamical systems without having to be explicitly fed with past inputs and outputs. In this respect, we concentrate on a fully recurrent structure in which the synapses are expressed as finite order autoregressive filters. The resulting architecture encompasses most of the earlier networks introduced for temporal processing tasks and is able to deal with systems whose order is unknown and discover contingencies spanning unknown temporal intervals. This paper examines the ability of such a *space* and *time* recurrent neural network to infer the rules that govern the fuzzy seeing evolution from the regularities in the present and past meteorological measurements. With a view towards demonstrating the effectiveness of our approach, the neuro-fuzzy method is compared with a k-nearest neighbor method.

2 A Fuzzy Framework

2.1 Meteorological Variables Used

The meteorological variables used were: wind direction and velocity, at 10m and 2.5m above ground; seeing at 5m; temperature, humidity and pressure at 2m; and ground temperature at -0.1m. Very high correlation among some of these variables led to us retaining the following: wind velocity in ms^{-1} at 10m above ground; standard deviation of the wind speed over 20 min. using 2-second samples; relative humidity, in percent; air temperature, in degrees at 2m above ground; pressure in mB; and seeing, measured at 5m above ground. Seeing is defined as the full width at half-maximum of a stellar image with a perfect large telescope at 0.5 μm wavelength and at zenith.

Unlike in [11], nightly median values are considered for these variables instead of hourly averages. This was done with the aim of giving an indication of the quality of the observing night. The data between spring 1989 and autumn 1991 were from Cerro Paranal. Construction then started at this site and the data for the four final seasons, summer 1992 to spring 1992 were collected from the nearby Cerro NTT. These data comprised a 605-tuples dataset.

2.2 Seasonally-Referenced Fuzzy Coding

A major lesson from time series analysis is that the *representation problem* plays an important role. A relevant representation can make information explicit and strip away obscuring clutter. Murtagh and Sarazin [11] adopted an intuitively plausible and interpretation-friendly form of coding. Each variable was categorized into three categories: low-valued (= 0), a linearly varying moderate-value, and high-valued (= 1). The boundaries between these categories were set at the 33rd and 67th percentiles. In a practical scenario, we believe it would be relatively easy to obtain high quality estimates of such variables, all the more so as we require only minimal information, i.e. high, low, or moderate. The coding of seeing was referenced to the season. We consider a season as a set of about 90 nights which are batched together in time. In fact, around 50 nights remained when missing values were removed. The seasonal referencing which required a relevant set of 33rd and 67th percentiles was provided by the immediately preceding "season" of 49 cases. This is a robust approach: extreme values are not allowed to play a disproportionate role.

When we recode seeing in this way, we can refer to low values as good seeing, and high values as bad seeing, again referenced to the season or other subset of the data on which we base our definitions of the 33rd and 67th percentiles. The Euclidean distance may be more easily used between vectors which have been recoded in this way since normalization of the very differently-valued variables has been taken care of. We can go further, however, making use of the notions of "high" and "low" which are used in this recoding approach. In order to be consistent with these notions, we have taken them as fuzzy values. The usual definitions of relatedness and dependence generalized for use in such a fuzzy framework are described in what follows.

2.3 Definitions of Possibility

"High" and "low" and the linearly interpolated intermediate values are fuzzy quantities. When dealing with such data, we therefore use fuzzy generalizations of traditional notions such as conditional probability. The fuzzy version of this is termed conditional possibility.

Consider three variables, A_i, B_i, and C_i, where index i corresponds to a case or record. Therefore A_i could be the temperature measured at the time-point i; B_i could be the seeing measured at the same time-point; etc. As discussed in [10] we have:

$$\text{Possibility of A and B} = \bigcap_i (A_i, B_i) = \sum_i \min(A_i, B_i)$$

$$\text{Conditional possibility of A, given B and C} = \bigcap_i (A_i, B_i, C_i) / \bigcap_i (B_i, C_i)$$

$$= \sum_i (\min_i(A_i, B_i, C_i)) / \sum_i (\min_i(B_i, C_i))$$

See [11] for further discussion and references.

2.4 Fuzzy Correspondence Analysis

Correspondence analysis is a dimensionality-reduction method for the analysis and display of data. It is approximate insofar as we generally consider only two-dimensional projections of a multivariate space. In presenting projections in such a planar display, we may look for proximities (indicating high correlatedness, or perhaps even causal relationships), oppositions (for example, between good and bad seeing), and various other interrelationships.

Input to correspondence analysis was, in all cases, based on the fuzzily coded data, as described in the previous section. In addition, the input data was "doubled" in the following way: each variable was not only fuzzily coded, but its complement ($1-$ the fuzzy value) was also used. Thus, values for both low humidity and high humidity were entered for each case. Given the complementary between each such pair of values, there is considerably redundancy in the input data.

The reasons for using this form of input coding were two-fold. Firstly, we make reference to a technical aspect of correspondence analysis, discussed in [11]. Correspondence analysis uses the chi-squared metric, as opposed to the use of the Euclidean metric in principal components analysis. The Euclidean metric is often more convenient for visualization and interpretation. A side-effect of the "doubling" used in coding is that the total of values of all variables associated with each case is constant. When this happens, the chi-squared distance is reduced to a weighted Euclidean distance. Furthermore, each case is now identically weighted in the correspondence analysis, which would be naturally the case in a principal components analysis.

The second reason for using the "doubled" form of coding is that we are led to a particular form of interpretation. Having projections on the planar maps produced by correspondence analysis which are associated with good seeing and bad seeing is advantageous for our purposes. Having, additionally, projections associated with low and high humidity (for example) means that we will more easily draw conclusions with respect to such fuzzy variables. This "doubled" form of fuzzy coding is therefore helpful for interpretation purposes.

Complementary couples of variables (e.g. low humidity and high humidity) are symmetrically positioned vis-à-vis the origin in the space of full dimensionality. The origin itself represents an average profile, so that the nearer a complementary couple of variables are positioned to the origin, the less discriminating they are.

2.5 Correspondence Analysis Results

Table 1 lists the fuzzily coded variables used for each case. "High" and "low" were specified with respect to the 33rd and 67th percentile values within a season. Each case referred to a median of night-time recorded values, and was therefore associated with an observing night. In Table 2, it can be seen how many such observing nights were retained in each season. The maximum of about 90 nights in each season was lessened because of missing data. The definition of conditional possibility of good seeing (let us call this A_i for case i) given low wind speed and low wind speed standard deviation (let us call these, respectively, B_i and C_i) has been discussed in Section 2.3. The "possibility of variables happening" is

$$\bigcap_i (A_i, B_i, C_i) \quad , \; 1 \leq i \leq n$$

where n is the number of cases.

Fig. 1 shows nowcasting interrelationships as provided by 11 correspondence analyses. Using the heuristic value of these plots, we looked at various relationships which are potentially of interest. These are overviewed in Table 2. Two fuzzy criteria are used for each such relationship, which involves a trade-off: how good is the relationship; and even if very good, how relevant in practice is the relationship?

Some of the relationships summarized in Table 2 are very good indeed. One sees for instance that in autumn 1992, the meteorological variables `s1L`, `s1dL` and `pH` (respectively, low wind speed, low wind speed standard deviation, and high pressure) had a 90% association with good (i.e. low seeing); and that such a "constellation" of meteorological variables arose 31% of the time.

By seasonally-referencing our data, Table 2 presents an appealing list of useful findings. It indicates that variables `s1L` and `s1dL` often co-occur with good seeing. But there are nonetheless differences between the same season in different years, which limits the usefulness of this finding. Our problem is to make these relationships useful in a practical and consistent framework.

Within a season, it appears thus that there is information available which would allow good forecasts. The most relevant meteorological variables are difficult to specify, though, because we find that they differ in different seasons.

3 A Dynamic and Recurrent Neural Network

4 The Network Architecture

This section describes the connectionist model we used to assess the predictability of seeing. A network that is to achieve sequence recognition tasks such as prediction must know about the past. In this perspective, we concentrate on a fully recurrent structure in which the synapses are expressed as autoregressive filters to endow the model with an internal dynamic. By virtue of this capability, the network can identify dynamical systems without having to be explicitly fed with past inputs and outputs and discover contingencies spanning unknown temporal intervals. Note that this architecture encompasses most of the earlier networks introduced for temporal processing tasks [1, 2, 12, 16].

The DRNN node equations are:

$$
\begin{aligned}
o_i(t) &= f(net_i(t)) \\
net_i(t) &= bias_i + \sum_{k=0}^{M} \sum_{j=1}^{N} w_{ji}^k o_j(t-k) \\
f(h) &= \frac{1}{1+\exp(-2\beta h)}
\end{aligned}
\tag{1}
$$

where $o_i(t)$ is the output of node i at time t, w_{ji}^k is a weight from k-th delay of node j to node i, N is number of nodes and M is the order of the autoregressive linear filters.

The sigmoid performs a smooth mapping from $(-\infty, +\infty) \rightarrow (0, 1)$. Note that the steepness β can be absorbed into weights and biases without loss of generality. Bear in mind that there is a set of terminal units whose only role is to feed input patterns into the rest of the network. There are no connections leading to the input units, since these units are clamped to the values specified in the input pattern. Similarly one unit is taken as the output from which the result of the computation is read off. Note that the desired output is considered as a scalar, without loss of generality, since this study deals with univariate time series. Units having neither input nor output connection to the outside world, are called *hidden* units. In addition, a bias input is achieved by fixing a particular unit activation to 1.0. For notational convenience, we did not distinguish between input and non-input units.

The motivation for using neurons as internal state coordinates is that certain features of the time series can be encoded and help us characterize some of the essential properties that lie behind the observation sequence.

Formally define $\mathbf{x}_t = (x_t, \ldots, x_{t-M})$ where x_t denotes the input value at time t. Let $\mathbf{o}_t^j = (o_t^j, \ldots, o_{t-M}^j)$ for $j = 1, \ldots, N'$ where N' is the number of hidden units. By virtue of the internal dynamic behavior, the DRNN offers the hope to approximate any discrete-time system defined by the following system of deterministic equations:

$$\mathbf{o}_{t+1}^j = g(\mathbf{x}_t, \mathbf{o}_t^1, \ldots, \mathbf{o}_t^{N'}) \quad \text{for} \quad j = 1, \ldots, N' \tag{2}$$

and for which the observable x_{t+1} is given by the combination of the internal parameters $\mathbf{o}_t^j$ according to:

$$x_{t+1} = h(\mathbf{x}_t, \mathbf{o}_t^1, \ldots, \mathbf{o}_t^{N'}) \tag{3}$$

The promise of the DRNN model is to capture the behavior of such dynamical systems, insofar as g and h are "well behaved" functions, accurately approximated by Eq. (1).

4.1 A Time-Recurrent Error Propagation Algorithm

This section succinctly reviews the main features of the gradient-based training algorithm that is to deal with the time and space recurrency introduced by delay lines. The efficiency/complexity dilemma of the learning procedures is a legitimate concern for connectionist network practitioners. A true gradient following training algorithm, extending the RTRL algorithm [16] to the DRNN, involves exceedingly large resource requirements [2, 5, 16]. Computational requirements scales as M^2N^5 and the derivative bookkeeping entails a storage capacity of order MN^3. In addition, such undesirable features are compounded by the loss of locality due to the matrix inversion required to compute the gradient terms. Fortunately, a much more attractive algorithm can be derived provided we depart a little from the true gradient trajectory.

One way to circumvent these troublesome issues is to use a technique referred to as *unfolding in time*. The strategy is to remove all delays by expanding the network into several static ones, each associated with a time slice. Units and links are replicated backward in time until all delays have been removed. Thus, the error propagation algorithm we describe is a generalization of the back-propagation-through-time algorithm [5, 14] in which one stores the activations of the units while going forward in time. The backward phase is also backward in time and recursively uses these activations to compute the required gradients. Besides, it benefits from the desirable simplicity of the recurrent-backprop algorithm [1, 5, 12] by relaxing the transposed error network in order to compute the weight changes.

However, for the unfolded version to be strictly equivalent to the original network, the structure has to be stretched up to the time slice associated with the first pattern. In this case, the memory requirements involved by the algorithm grow proportionally with the number of patterns. This is clearly inhibitory from an implementation standpoint. Hopefully, the solution of the problem lies in the decay rate of the error terms. It is demonstrated in [2] that the error terms, for a stable network, drop off exponentially when backpropagated in time, rapidly becoming negligible. This exponential decay rate is an intricate function of the network parameters, and thus varies with time. In this respect, the maximum time interval, say $L \geq M$, spanned by the unfolded network is taken large enough, in view of this decay rate, not to corrupt the true gradient trajectory. On the other hand, L is tailored in the course of the training process to minimize the bookkeeping and the number of computational operations required by the algorithm.

Let the objective of training be the minimization of

$$E = \frac{1}{2} \sum_{k} E^2(k) \tag{4}$$

$E(k)$ denotes the difference between the target and the network output at time k. At each time step, a new pattern is presented to the network. The network is then unfolded backwards in time, up to the limit discussed below.

Suppose the network is run up to some time T and that the weights remain fixed throughout this trajectory. Assume moreover that each unit i keeps track of its L past activation and output values, namely net_i and o_i. Adjusting the weights of the system boils down to backpropagating the error terms, Y, in the unfolded error-propagation network as for the static recurrent-backprop algorithm [1, 5, 12]. Its topology is the same as that of the original network to the extent that it is unfolded L times.

Let I and O denote the sets of input and output units respectively (in this study, O is a singleton). When a fixed-point solution of the transposed error network, at some intermediate time $t_0 \leq T$, exists and is a stable attractor [2], it is given by

$$Y_i(t) = \sum_{d=0}^{M} \sum_{j \notin I} w_{ji}^{d\,\star}(t+d) Y_j(t+d) \mathbf{1}_{\{t+d \leq t_0\}} + E_i(t) \mathbf{1}_{\{t=t_0; i \in O\}} \quad \text{for} \quad t_0 - L \leq t \leq t_0 \tag{5}$$

where

$$w_{ji}^{d\,\star}(t) = f'(net_j(t)) w_{ij}^{d}(t) \tag{6}$$

Note that the weights w_{uv}^{d} (and $w_{uv}^{d\,\star}$) are duplicated in time and considered as distinct by the gradient descent algorithm in the unfolded network. In this study, L is taken proportional to M, the maximum connection order, and adapted at each time step t_0 to satisfy

$$\mid \frac{\mathrm{Max}_{i \notin I}(Y_i(t_0 - L))}{\mathrm{Max}_{i \notin I}(Y_i(t_0))} \mid \leq \epsilon \tag{7}$$

where ϵ is an arbitrary predefined value, equal to 10^{-3} in our implementation.

The temporal delta-rule at time t_0 is then given by

$$\Delta w_{uv}^{d}(t_0) = \eta \sum_{n=0}^{L-d} \delta_v(t_0 - n) o_u(t_0 - d - n) \mathbf{1}_{\{t_0 - d - n \geq 0\}} \quad \forall (u, v) \tag{8}$$

where

$$\delta_j(t) = f'(net_j(t)) Y_j(t) \tag{9}$$

Classically, the gradient terms are accumulated over the trajectory, and afterwards recombined to compute the overall weight change, such that

$$\Delta w_{uv}^{d} = \eta \sum_{t=0}^{T} \Delta w_{uv}^{d}(t) \tag{10}$$

Instead, we deviate from this approach and actually adapt the weight at each time increment to avoid any storage requirement proportional to the trajectory length. We call this algorithm the *time-recurrent error propagation* training algorithm. While it is not guaranteed to follow the true gradient, it is known to work well in practice, presumably because the use of a small learning rate leads to weight updates whose direction is a close enough approximation to the true gradient.

From a computational standpoint, the procedure exhibits very attractive features. Order M^2N^2 operations are needed and storage requirements scale as MN. The stability and the existence of a fixed point solution of the DRNN are addressed in [2].

5 Fuzzy Nowcasts of Seeing

5.1 Nearest Neighbor Nowcasts

The nearest neighbor approach is extended to deal with the fuzzy coding. Given a set of meteorological variables, we seek among all those stored in our database, the fuzzy seeing value associated with the nearest neighbor vector as the estimate of the desired value. In order to be consistent with the coding used, a fuzzy Jaccard measure [8, 10] is used to find the best match. The Jaccard measure is well-known in information retrieval, and in some tests we carried out, it outperformed a fuzzy Euclidean distance. The definition, for vectors $\{A_i\}$ and $\{B_i\}$ is:

$$\bigcap_i (A_i, B_i) / \bigcup_i (A_i, B_i) = \sum_i \min(A_i, B_i) / \sum_i \max(A_i, B_i) \qquad (11)$$

When equivalued similarities were found, we averaged the seeing values in an *ad-hoc* way, using the ordinary definition of mean. The experimental "leave one out" strategy consists of searching for the most similar set of meteorological variables in the entire database which also contains data collected at time beyond the case currently investigated. The seeing associated with the most similar set of meteorological variables provided the estimate. The bigger our database, the more chance we have of capturing every meteorological combination of interest.

The results obtained are presented in Table 3 and may be summarized as follows:

- Conditional probability of good and bad seeing, given that good seeing was estimated, is 48.2 % and 22.0 % respectively.

- Actual and estimated seeing in the same category: 40.2 % of cases.
- Actual and estimated seeing in the opposite category (good-bad or bad-good): 9.6 % of cases.

Here, one may remark that an uninformed policy based on the definition of good seeing would lead to 33% of cases.

5.2 A Cost-Adaptive Connectionist Approach

In this section, we report some preliminary simulation results obtained with a connectionist network. We note that an interest in this paper, beyond the quest for reliable forecasts, is in optimizing the scheduling observation time assignment procedure at ESO. A preliminary cost analysis has been carried out given that the queue policy amounts to selecting the observation that best fits the estimated seeing value, see Table 4. While the fuzzy nearest neighbor method provides unequivocal estimates of seeing, the neural net output interpretation is dependent on this particular performance objective.

Since the network output $o(t)$, at any time t, fell mostly inside the interval [0,1] , two thresholds (s_1, s_2) were required to interpret the output value. The prediction was assumed to be low whenever $o(t) < s_1$, a moderate value when $s_1 \leq o(t) \leq s_2$ and a high value when $o(t) > s_2$. These thresholds were optimally estimated from the performance achieved on the training set and then applied to the prediction set to appraise the predictive performance of the model. The thresholds could be adjusted on-line in a practical implementation as time progresses, among the chronologically most recent tuples. Consequently, the results presented are a lower bound of the expected performance. We emphasize that the choice of the cost function can be tailored in a convenient way to meet the changing requirements of the observatory engineers and astronomers.

With a view to presenting the performance of the connectionist model with the above objective in mind, three pairs of thresholds were used, related to the following performance measures (or policies) of interest:

- P1 - maximum hit rate, when actual and estimated seeing are in the same category.
- P2 - maximum conditional probability to have good seeing when good seeing was predicted.
- P3 - minimum of the cost function given by Table 4.

The results obtained on the prediction set are cross-tabulated in Tables 5 to 7. Various measures of interest can be derived from these contingency tables. They are described below in a convenient way if one wishes to diminish the number of "false alarms", i.e. bad seeing when good seeing is expected, minimize a given cost function, etc.

5.3 Neural-Based Nowcast Results

In view of the autocorrelation function, Fig. 2, short-term correlations are present in the coded seeing sequence despite a few sequence disruptions. We use a DRNN to capture the dynamical dependencies between local meteorological conditions and the fuzzily coded seeing. The model is a four-layer network with $5 \times 10 \times 10 \times 1$ nodes. The layers are fully connected to each other, except the input layer which doesn't receive feedback signals. All connections are made of zero-order filters (undelayed links) except those connecting the input units to the first layer and those connecting the output unit with the third layer, which are of order 1.

The network had 351 links including bias. Overfitting was prevented by a weight elimination method [2, 15]. In performing these simulations, we have made no effort to find an optimal architecture. Further investigations are being conducted at present and will be reported in due course.

Cumulative prediction histograms are plotted in Fig. 3 and neural network outputs in Fig. 4.

- For policy P1, the conditional probabilities of good and bad seeing, given that good seeing was predicted, are 69.5 % and 8.3 % respectively. The possibility that good seeing is nowcast, that is, the number of good nowcasts / the total number of nowcasts, is 34.3 %.

- For policy P2, the conditional probabilities of good and bad seeing, given that good seeing was predicted, are 84.0 % and 4.0 % respectively. The possibility that good seeing is nowcast is 23.8 %.

- For policy P3, the conditional probabilities of good and bad seeing, given that good seeing was predicted, are 100 % and 0 % respectively. The possibility that good seeing is nowcast is 2.9 %.

As an indication, the three parameter configuration is related to the costs +9.5, +10.0 and −16.0 respectively. The cost associated with the nearest neighbor nowcasts is +16.8 (this cost was normalized to be consistent with the 105 predictions carried out on the prediction set).

5.4 Discussion

Notwithstanding the remarkably encouraging figures presented here, we noticed that the measurement quality and the data regularity were not uniform over the dataset. Indeed, Table 2 indicates that taking appropriate meteorological variables into account yields slightly higher good seeing conditional probabilities at Cerro NTT, than at Cerro Paranal. Since our prediction set was set up from Cerro NTT measurements, we have been wondering if the results would be as satisfactory at Paranal. Consequently, the same methodology was adopted with a prediction set taken from data at Cerro Paranal; the results were found to be very similar to those presented here confirming the adequacy of our approach. Further discussions and detailed results are presented in [3, 10].

6 Conclusion

We have described a connectionist approach for nowcasting fuzzily coded astronomical seeing at Cerro Paranal, site of the Very Large Telescope. A fuzzy correspondence analysis was carried out to demonstrate the predictability of seeing, given the present environmental condition parameters. Remarkably reliable estimates were obtained, in regard to fuzzy seeing, by a dynamic and recurrent neural network. Such an approach was found to outperform a k-nearest neighbors method. The flexible neuronal methodology discussed for fuzzy seeing nowcasts enables the astronomer and the observatory engineer to select appropriately the performance objective function most suited to their requirements. The adequacy of the nowcasting approach for seeing prediction is confirmed and future work should aim to obtain encouraging results for the prediction of fuzzily coded seeing one hour in advance.

The assessment figures for nowcasting of seeing which are quoted in this article, and the connectionist model which has been prototyped, are a contribution to a future telescope decision support system.

References

[1] L.B. Almeida, Backpropagation in perceptrons with feedback, in: R. Eckmiller and Ch. von der Malsburg, ed., *Neural Computers* (Springer-Verlag, Berlin, 1987) 199-208.

[2] A. Aussem, Dynamical and recurrent neural networks towards time series prediction, in preparation.

[3] A. Aussem, F. Murtagh and M. Sarazin, Dynamical recurrent neural networks and pattern recognition methods for time series prediction: application to seeing and temperature forecasting in the context of ESO's VLT Astronomical Weather Station, *Vistas in Astronomy* (1994), submitted.

[4] N.A. Gershenfeld and A.S. Weigend, The future of time series: learning and understanding, in: A.S. Weigend and N.A. Gershenfeld, ed., *Time Series Prediction: Forecasting the Future and Understanding the Past* (Addison-Wesley, New Jersey, 1993) 1-70.

[5] J. Hertz, A. Krogh and R. Palmer, *An Introduction to the Theory of Neural Computation* (Addison-Wesley, New Jersey, 1991).

[6] M.B. Jorgenson, G.J.M. Aitken and E.K. Hege, Evidence of a chaotic attractor in star-wander data, *Optics Letters* 16 (1991) 64-66.

[7] A. Lapedes and Farber, Non linear signal prediction using neural networks: prediction and system modelling, Preprint, (Los Alamos National Laboratory LA-UR-87-2662, 1987).

[8] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis* (Kluwer, Dordrecht, 1990).

[9] F. Murtagh, M. Sarazin and H.M. Adorf, Statistical prediction of astronomical seeing and of telescope thermal environment, in: M.H. Ulrich, ed., *ESO Conference on Progress in Telescope and Instrumentation Technologies* (ESO, Garching, 1992) 377-380.

[10] F. Murtagh, A. Aussem and M. Sarazin, Nowcasting astronomical seeing: towards an operational approach, *Publications of the Astronomical Society of the Pacific* (1994), submitted.

[11] F. Murtagh, M. Sarazin, Nowcasting astronomical seeing: a study of ESO La Silla and Paranal, *Publications of the Astronomical Society of the Pacific* 105 (1993), 932-939.

[12] F.J. Pineda, Generalization of back-propagation to recurrent neural networks, *Physical Review Letters* 59 (1987) 2229-2232.

[13] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, New York, 1988).

[14] D.E. Rumelhart, G.E. Hinton and R.J. Williams, Learning internal representations by error propagation, in: D.E. Rumelhart and J.L. McClelland, ed., *Parallel Distributed Processing Explorations in the Microstructure of Cognition*, Vol 1 (MIT Press, Cambridge, 1986) 318-362.

[15] A.S. Weigend, D.E. Rumelhart and B.A. Huberman, Predicting the future: a connectionist approach, *International Journal of Neural Systems 1* (1990) 193-220.

[16] R.J. Williams and D. Zipser, Experimental analysis of the real-time recurrent learning algorithm, *Connection Science 1* (1989) 87-111.

s1H	wind speed – High
s1L	wind speed – Low
s1dH	wind speed standard deviation – High
s1dL	wind speed standard deviation – Low
rhH	relative humidity – High
rhL	relative humidity – Low
t1H	temperature – High
t1L	temperature – Low
pH	pressure – High
pL	pressure – Low
seeH	seeing – High
seeL	seeing – Low

Table 1: Fuzzy variables used.

Season	No. of cases	Good seeing given ...	Conditional possibility of good seeing	Possibility of variables happening
Spr89	81	s1L	0.68	0.35
		s1dL	0.61	0.27
		s1L and s1dL	0.62	0.20
Sum89	77	pH	0.64	0.32
		s1dL	0.69	0.38
		s1L	0.68	0.35
		pH, s1dL and s1L	0.75	0.24
Aut90	48	rhL	0.76	0.35
Win90	69	rhL	0.63	0.32
		t1H	0.63	0.31
		pH	0.63	0.33
Spr90	47	s1L	0.58	0.29
		s1dL	0.67	0.36
		s1L and s1dL	0.67	0.27
Sum90	38	rhL	0.52	0.27
		t1H	0.62	0.31
		rhL and t1H	0.60	0.21
Aut91	61	s1L	0.61	0.29
		s1dL	0.70	0.41
		s1L and s1dL	0.60	0.25
Sum92	55	s1L	0.76	0.40
		s1dL	0.77	0.38
		pH	0.68	0.35
		s1L, s1dL and pH	0.88	0.24
Aut92	66	s1L	0.78	0.40
		s1dL	0.80	0.43
		pH	0.67	0.37
		s1L, s1dL and pH	0.90	0.31
Win92	44	s1L	0.89	0.46
		s1dL	0.87	0.42
		rhL	0.72	0.38
		s1L, s1dL and rhL	0.96	0.27
Spr92	69	s1L	0.74	0.38
		s1dL	0.68	0.33
		s1L and s1dL	0.72	0.29

Table 2: Nowcasting results.

estimated \ actual seeing	good	moderate	bad
good	68	42	31
moderate	109	117	104
bad	27	49	58

Table 3: Contingency table exemplifying estimated versus actual seeing values with the k-nearest neighbor method in the overall data set - Same category: 40.2 %, Partially wrong: 50.2 % and Incorrect: 9.6 %

estimated \ actual seeing	good	moderate	bad
good	-1	+1	+1
moderate	+1/2	0	+1/4
bad	+1/2	+1/4	-1/4

Table 4: Costs associated with each possibility. Positive and negative values express failure and success respectively.

estimated \ actual seeing	good	moderate	bad
good	25	8	3
moderate	8	10	10
bad	1	15	25

Table 5: 1992 - Contingency table exemplifying estimated versus actual seeing values in prediction set - Output thresholds $(0.33; 0.65)$ have been optimally determined from the training set to maximize the hit rate (P1) - Same category: 57.2 %, Partially wrong: 39.0 % and Incorrect: 3.8 %

estimated \ actual seeing	good	moderate	bad
good	21	3	1
moderate	12	15	12
bad	1	15	25

Table 6: 1992 - Contingency table exemplifying estimated versus actual seeing values in prediction set - Output thresholds $(0.33; 0.74)$ have been optimally determined to maximize the good-good conditional probability (P2) - Same category: 58.1%, Partially wrong: 40.0% and Incorrect: 1.9 %

estimated \ actual seeing	good	moderate	bad
good	3	0	0
moderate	31	25	22
bad	0	8	16

Table 7: 1992 - Contingency table exemplifying estimated versus actual seeing values in prediction set - Output thresholds $(0.28; 0.9)$ have been optimally determined to minimize the cost function (P3) - Same category: 41.9%, Partially wrong: 58.1% and Incorrect: 0%

Figure 1: Correspondence analysis for 11 seasons.

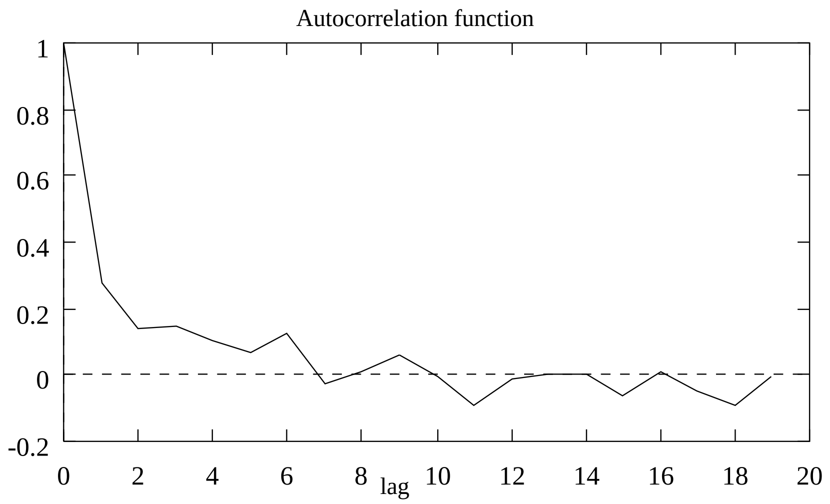

Figure 2: Autocorrelation function of fuzzy seeing sequence.

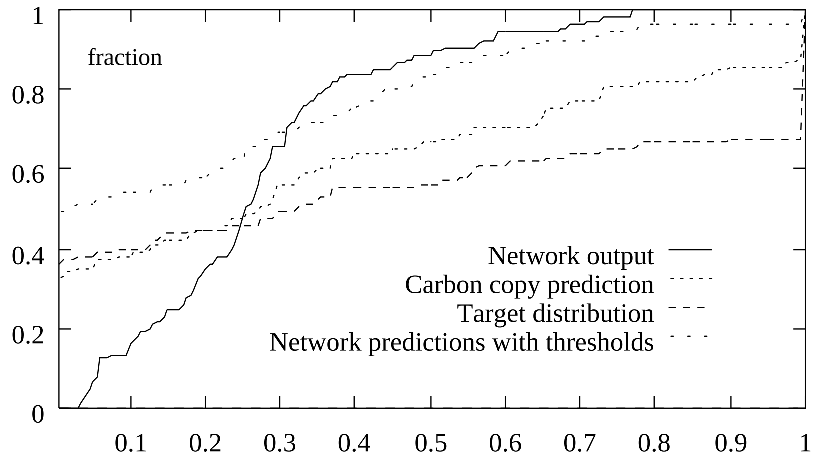

Figure 3: Fuzzily coded seeing cumulative histograms during winter and spring 1992. The plain line corresponds to the raw output of a dynamical recurrent neural network. The dotted line corresponds to the output with respect to two thresholds optimally determined to maximize the hit rate.

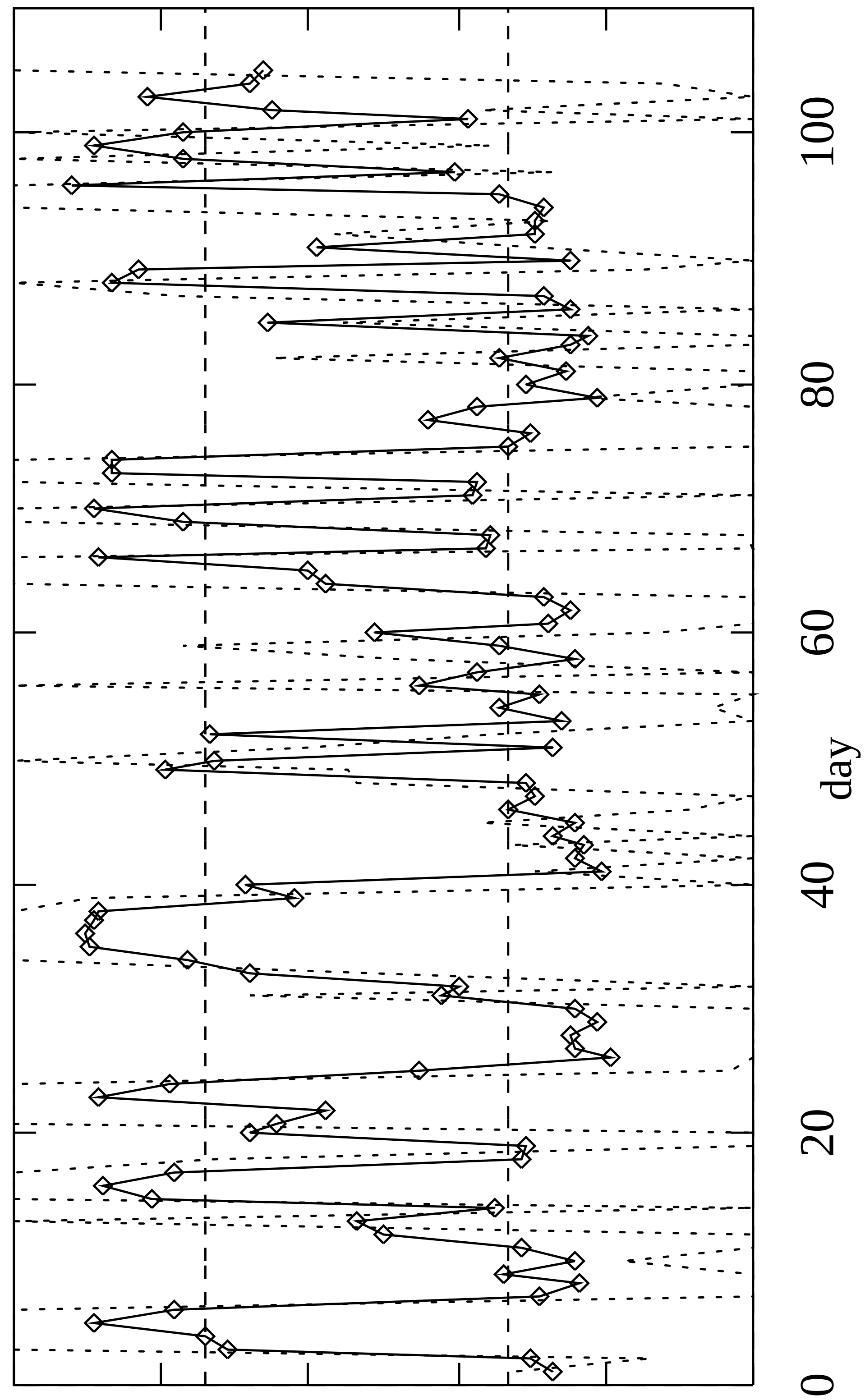

Figure 4: Fuzzy coded seeing (dotted line) during winter and spring 1992 and single-step forecasts (plain line) in the prediction set. Thresholds 0.33 and 0.74 are shown.

Alex Aussem was born in Paris in 1968. He received the engineer degree from the French Institut National des Télécommunications in 1992. At present, he is working toward a Ph.D. degree at the Mathematical and Computer Science Department of University René Descartes in Paris. He is currently pursuing his thesis in the Very Large Telescope Division at the European Southern Observatory. His research interests concern the design of training algorithms for dynamical and recurrent neural networks, and their applications to resource management in ATM networks and more generally to time series forecasting.

Fionn Murtagh (PhD University of Paris VI, Habilitation University of Strasbourg I) is a Senior Scientist with the European Space Agency. He is an elected member of the Board of Directors of the Classification Society of North America, and of the Board of Directors of the European Regional Society of the International Association of Statistical Computing. He is President of Technical Committee 13 of the International Association for Pattern Recognition. He is editor of the annual Classification Literature Automated Search Service, and is a member of the editorial boards of the Journal of Classification, Les Cahiers de l'Analyse des Données, and Neurocomputing. He has authored or edited about 10 books and about 150 papers.

Marc Sarazin was born in 1955 and he obtained a Physics Engineering degree from Ecole Nationale Supérieure de Marseille (France) in 1978. He is Applied Physicist at the European Southern Observatory since 1984 where he took an active part in the site survey for the Very Large Telescope Project. He obtained a Ph.D in Physics from the University of Aix-Marseille III in 1992. He is a member of the Optical Society of America and is currently working on the monitoring and prediction of observing conditions for large astronomical telescopes.