3.5

# MIXING HEIGHT SHORT RANGE FORECASTING THROUGH NEURAL NETWORK MODELING APPLIED TO RADON AND METEOROLOGICAL DATA

Antonello Pasini[1,*], Fabrizio Ameli[2], Massimo Loré[2]

[1] *CNR, Istituto sull'Inquinamento Atmosferico, Rome, Italy*
[2] *Università "La Sapienza", Dipartimento di Fisica, Rome, Italy*

## 1. INTRODUCTION

The properties of the atmospheric boundary layer were extensively investigated in the last decades and the nonlinear nature of its physics (and chemistry) is now well known: see, for instance, Stull (1988), Seinfeld and Pandis (1998). A particular attention was recently paid to the nocturnal stable layer, whose diagnostic and prognostic modeling represents a major challenge for physicists of the atmosphere and meteorologists (Mahrt 1998).

In general, the problem in searching for a detailed characterization of the boundary layer is twofold. First of all, it is difficult to represent all the influences of concern on a certain process or phenomenon, because they are a lot and have approximately the same magnitude. Furthermore, as far as the evolution of the system is concerned, all these elements interact in a highly nonlinear manner, giving rise to some drawbacks in modeling this evolution, like difficulties in the balance among parametrizations for different processes and the birth of deterministic chaos in the long range.

In such a framework, in this paper we firstly identify a useful physical quantity (the radon concentration at the surface), which leads to an index of the diffusion properties of the low atmosphere that is able to summarize our knowledge of the boundary layer state. Then, in a situation where we are not able to completely determine the dynamics of the system, we perform short range forecasts of this quantity *via* neural network modeling. As final result we obtain the prediction of the nocturnal stable layer depth up to 6 hours ahead through the application of a box model to the data coming from the neural network forecasting activity.

In what follows, after a brief discussion about the relevance of the radon progeny measurements for the boundary layer characterization and the use of a simple box model, a closer attention is paid to the neural network modeling for radon concentration forecasting. In particular, in the framework of feed forward networks with backpropagation training, two strategies are adopted: a time series approach using only inputs from time-delayed radon data and a synchronous pattern approach where inputs are given by both meteorological and radon detections at a certain time $t_0$. Their respective forecasting results are compared and the peculiar roles of preprocessing and pruning are stressed.

## 2. RADON DETECTION AND MODELING

During the last decades, many studies about indoor and outdoor radon concentration were performed, mainly for epidemiological purposes. Since the late 70s, French researchers recognized the relevance of radon detections also for a characterization of the boundary layer properties (see, for instance, Guedalia et al. 1980). These detections are usually based on the counts of beta radioactivity coming from the decay of the $^{222}$Rn short-lived daughters, which are attached to particulate matter.

Under some simplifying assumptions (highly confirmed in stable situations and for limited spacetime intervals), radon exhalation from the ground can be considered constant in time and spatially homogeneous, the attached fraction of radon daughters also constant and the radon concentration directly proportional to the number of beta counts detected. A brief description of the instrument developed and used at CNR-IIA can be found in Pasini et al. (2002).

The typical patterns of beta counts show maxima during the night and minima during the day in conditions of nocturnal stability and well developed diurnal boundary layers. Otherwise, low quasi-constant values are found in advective and turbulent situations. This fact leads to suppose that the concentration of radon (which undergoes no chemical reaction) mainly depends on the volume available for its dilution. Furthermore, old evidences (Lopez et al. 1974) and more recent results (Vinod Kumar et al.

* *Corresponding author address:* Antonello Pasini, CNR, Istituto sull'Inquinamento Atmosferico, I-00016 Monterotondo Stazione (Roma), Italy; e-mail: pasini@iia.mlib.cnr.it

1999) show that radon and radon daughters concentrations are approximately homogeneous with altitude in the nocturnal stable layer and that they undergo a rapid transition to background values above the mixing height (in the so called residual layer).

All these evidences induce us to use a box model of the nocturnal stable layer endowed with a homogeneous radon concentration in the vertical. The top of this box is called equivalent mixing height $h_e$ and was found to be a good index of the dispersion properties of the boundary layer (Guedalia et al. 1980): low (high) values of $h_e$ are related to low (high) dispersion power and high (low) values of primary pollution. Allegrini et al. (1994) found a physical interpretation of this index and showed that, under some hypotheses for the application of the box model, $h_e$ well represents the depth of the urban stable layer $h_u$ as estimated by purely meteorological methods. At present, we know that in general $h_e$ and $h_u$ differ from the real value of the mixing depth, the equality being valid at the limit of null mechanical turbulence, when the physical features of the boundary layer are completely driven by the vertical thermal state of the low atmosphere. Furthermore, while $h_u$ has no chance to be sensitive to nonthermal factors, $h_e$ represents an index whose value is determined by all the factors which have influence on the dilution properties of the boundary layer.

More recently, work is in progress for improving the cited box model, taking radon decay and mass exchanges between stable and residual layers into account: a preliminary formulation can be found in Pasini et al. (2002).

## 3. A NEURAL NETWORK MODEL FOR RADON FORECASTING

Once identified $h_e$ as a single index which summarize our knowledge of the boundary layer state as far as its diffusion properties are concerned, forecasting its future values should be a major goal for air pollution meteorologists. Because of the cited difficulties in a general and reliable dynamical treatment of the system, it is worthwhile to apply a more phenomenological, adaptive and fully nonlinear forecasting method to this problem and analyse its short range performance.

In doing so, we apply neural modeling: our aim is to obtain $h_e$ values in the near future with a time frequency close enough to the characteristic time step of change for primary pollutants. Due to the lack of an instrument like RASS, that could allow us to obtain high frequency data about the vertical thermal profile, our sampling time for an accurate estimation of $h_u$ is 6 hours, determined by the radiosounding frequency. On the other hand, the sampling periods for radioactive detections and meteorological observations are 2 hours and 1 hour, respectively.

These considerations, added to the fact that $h_u$ is insensitive to mechanical turbulence, dissuade us from performing a direct forecast of $h_u$ and suggest to arrive at a forecast of $h_e$ in two steps: application of a neural network model, which leads to a radon concentration forecast, and the following run of the cited box model, which leads to obtain an estimation of $h_e$ in the future, starting from these forecast radon data.

While the neural model used is quite standard (it will be described in this section) and previously applied to similar forecasting problems, we want to stress that two distinct strategies will be investigated: a time series approach using only inputs from time-delayed radon data and a synchronous pattern approach where both meteorological and radon detections (chosen as inputs) approximately represent the "initial state" of the system at a certain time $t_0$. Available data, distinct strategies, forecasting results and further investigations will be presented in the following sections.

The neural networks considered in this work are feed forward and endowed with a backpropagation training rule. The original model and a particular modified version were extensively described elsewhere (Pasini and Potestà 1995; Pasini et al. 2001b), where they were applied to meteorological problems concerning short range forecasting in the boundary layer. Recently, a preliminary brief note on the problem of radon prediction also appeared (Pasini et al. 2001a). With reference to the notations shown in Fig. 1, in this paper we stress those features of our model structure and data handling which differ from more usual applications.

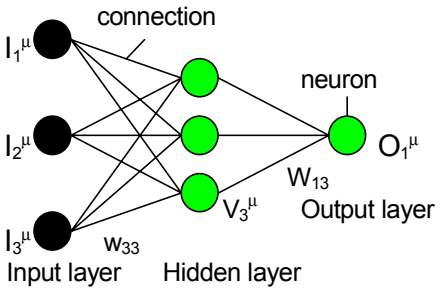


Fig. 1. Notations used in this paper for a feed forward neural network with one hidden layer.

Once noted that several proofs revealed that a single hidden layer is sufficient to grasp the transfer function to be determined for our problem, we stress

that the backpropagation training is endowed with both gradient descent and momentum terms. With reference to Pasini et al. (2001b), where a weighted cost function was adopted, here a simpler quadratic cost function is chosen. In short, cost function and updating rules for the weights in our pattern-based learning mode (Hertz et al. 1991; Bishop 1995) read as follows:

$$E^{\mu} = \frac{1}{2}\sum_{i}\left(T_i^{\mu} - O_i^{\mu}\right)^2 \tag{1}$$

$$\begin{aligned} W_{ij}(t+1) &= W_{ij}(t) - \eta\frac{\partial E^{\mu}}{\partial W_{ij}(t)} + m\left[W_{ij}(t) - W_{ij}(t-1)\right] \\ W_{ij}(t+1) &= W_{ij}(t) + \eta g_i'\left(h_i^{\mu}\right)\left(T_i^{\mu} - O_i^{\mu}\right)V_j^{\mu} + m\left[W_{ij}(t) - W_{ij}(t-1)\right] \end{aligned} \tag{2}$$

$$\begin{aligned} w_{jk}(t+1) &= w_{jk}(t) - \eta\frac{\partial E^{\mu}}{\partial w_{jk}(t)} + m\left[w_{jk}(t) - w_{jk}(t-1)\right] \\ w_{jk}(t+1) &= w_{jk}(t) + \eta\, g_j'\left(h_j^{\mu}\right)\sum_i W_{ij}g_i'\left(h_i^{\mu}\right)\left(T_i^{\mu} - O_i^{\mu}\right)I_k^{\mu} \\ &\quad + m\left[w_{jk}(t) - w_{jk}(t-1)\right] \end{aligned} \tag{3}$$

Here $h_j^{\mu}$ and $h_i^{\mu}$ are the weighted sums converging to the neurons of the hidden and output layers, respectively, $g_j$ and $g_i$ are the nonlinear transfer functions (sigmoids) calculated at hidden and output neurons, $g'$ are the sigmoid derivatives, $\eta$ and $m$ are the so called learning rate and momentum term.

In our model, a major attention is paid to the form of transfer functions. In particular, we choose sigmoids which are normalized with respect to the number of connections ($n_{hl}$ or $n_{ol}$) converging to a single neuron of the hidden and output layer, respectively. For instance:

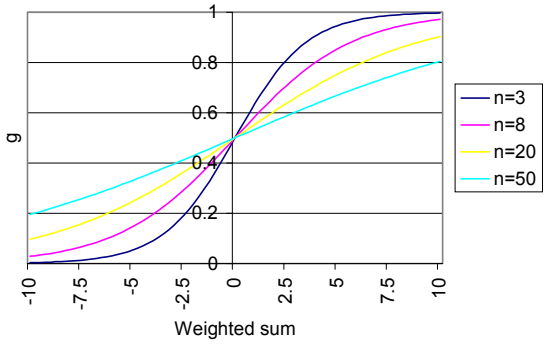


Fig. 2. Form of our normalized sigmoids.

$$g_i\left(h_i^{\mu}\right) = \frac{1}{1 + \exp\left(-\dfrac{h_i^{\mu}}{\sqrt{n_{ol}}}\right)} \tag{4}$$

When networks are endowed with many inputs and connections, the weighted sums can achieve large positive or negative values, corresponding to the asymptotic regions of $g$. In these cases, the derivatives assume values $g'(...) \cong 0$, the weight corrections could not occur efficiently and the system could become stuck in a local minimum or a very flat *plateau* (Hertz et al. 1991). As shown elsewhere (Pasini et al. 2001b), the choice of the form (4) for the sigmoids helps to avoid these problems. From a complementary point of view, one can look at this choice as a method to avoid overfitting, which can be considered quite alternative to the so called weight decay method: see the discussion in Marzban (2002).

The particular normalization performed in (4) leads to transfer functions which are less nonlinear when one moves from networks endowed with few connections to bigger ones (see Fig. 2). In this case a problem can arise with big networks ($n_{hl}, n_{ol} > 30 \div 50$), when the nonlinear behavior of the system under study could not be recognized. In this work $n_{hl}, n_{ol} \cong 10$ and that problem is avoided.

Our model is also endowed with an early-stopping method in order to furtherly prevent overfitting. Usually, the iterative application of (2) and (3) is stopped when a certain threshold in training convergence is achieved, namely when

$$\frac{\sum_{\mu,i=1}^{M_1,N}\left(T_i^{\mu} - O_i^{\mu}\right)^2}{M_1 N} < \text{MSS} \tag{5}$$

Here MSS (Mean Square Sum) is a chosen threshold, $M_1$ is the number of training patterns and $N$ is the number of output units. In some cases, when consistent differences in learning for distinct periods of the year are recognized, the alternative strategy to fix the number of epochs is considered. Anyway, MSS or the number of epochs are chosen in correspondence with the point when a decrease in performance in the validation set begins to be shown.

## 4. FORECASTING STRATEGIES AND PRELIMINARY DATA HANDLING

The data analyzed in this study comes from a 1-year extensive campaign performed in 1997 at the airport of Pratica di Mare, a site near Rome and very

close to the Tyrrhenian sea. Here beta counts from short-lived radon daughters decay were detected with a sampling period of 2 hours by the cited automatic instrument, located *in situ* by CNR-IIA. Weather parameters on a 1-hour basis are available from the local meteorological station.

While in a previous brief note we showed preliminar results about a time-series approach to radon concentration neural forecasting, here a more complete treatment and a comparison among the forecasting performance of two main strategies and their improvements are presented. Basically, while the quantity to be forecast and considered as target is always the number of beta counts (directly proportional to radon concentration), we choose two distinct sets of variables as inputs. In the first approach we use a time-delayed series of beta counts as input series, in the second one we try to give to the network a synchronous "measure" of the initial state of the atmosphere by considering several variables at a certain time $t_0$, numbered as follows: (1,2) hour of the day (expressed in two inputs as $[sin(\pi t/12)+1]/2$ and $[\cos(\pi t/12)+1]/2$), (3) beta counts, (4) beta counts time derivative with respect to 2 hours before, (5) sky covering and (6) height of the lowest cloud layer, (7) temperature, (8) dew point, (9) pressure, (10) horizontal wind speed, (11) visibility. Because of evidences of periodic signals in the beta counts time series a preprocessing activity is also performed in the framework of the first approach. On the other hand, a bivariate analysis inputs ÷ output shows some cases of very low linear and nonlinear correlations in the synchronous pattern approach and induces us to perform a pruning activity.

After preliminar proofs with some training and validation sets in order to establish the optimal thresholds for the early-stopping, the somewhat original method of "moving window" training is adopted, as in previous papers (Pasini and Potestà 1995; Pasini et al. 2001a,b). Here we consider a training set of about 2 months and update it every 2 hours, inserting the latest detected data and discarding the oldest ones, then training again the networks and performing a new forecast. The motivation for adopting this method has to be searched in the evidence, already shown in previous papers, that data which belong to another "season" have negative effects on short range forecasts in the boundary layer: of course, this is a peculiarity of historical data, like radon and meteorological ones.

## 5. INDICES OF PERFORMANCE

Even if the monitoring activity at Pratica di Mare was performed for an entire year, in this paper we choose to concentrate on the summer period, when we have no holes in radon data and we are also able to correctly study the series *via* Fourier analysis and other statistical methods which require contiguous data. Our time horizons for the forecasts are $t_0$+2, $t_0$+4 and $t_0$+6 hours: for each of them and for every approach a new network (endowed with a single output neuron) is trained and tested.

The performance of the neural modeling is assessed by one continuous index and several discrete ones. First of all we find the linear correlation coefficient $R$ of detected vs. forecast data; furthermore, as usual in the analysis of forecasting performance, we can divide detections and forecasts in classes, build contingency tables and assess performance in a dichotomic form. We choose to divide our data range in 100 classes, by means of 100 thresholds, and analyze 100 contingency tables. With reference to Table 1, where $a$ = number of nonevents predicted as nonevents, $b$ = number of nonevents predicted as events, $c$ = number of events predicted as nonevents, $d$ = number of events predicted as events, for every threshold we calculate the following indices:

BIAS = $f/h$;
POD (Probability Of Detection) = $d/h$;
FAR (False Alarm Ratio) = $b/f$;
HR (Hit Rate) = $(a+d)/n$;
EFF (EFFiciency) = $(a/g)\times(d/h)$;
CSI (Critical Success Index) = $d/(b+h)$;
HSS (Heidke's Skill Statistics) = $[2(ad-bc)]/(gf+he)$.

| DET\FOR | No | Yes | Sum |
|---|---|---|---|
| No | $a$ | $b$ | $g$ |
| Yes | $c$ | $d$ | $h$ |
| Sum | $e$ | $f$ | $n$ |

Table 1. Contingency table at a defined threshold distinguishing between events and nonevents.

## 6. FORECASTING RESULTS

In the time series approach we consider only time-delayed radon data as inputs. Once recognized in the data a clear 24-hour periodicity due to the day-night cycle, we perform an autocorrelation analysis, both in linear and nonlinear terms. For the nonlinear correlation analysis we use the so called correlation ratio $R_{nl}$, whose square can be written as (Marzban et al. 1999; Pasini et al. 2001b):

$$R_{nl}^2 = \frac{\sum_i q_i (\bar{y}_i - \bar{y})^2}{\sum_i \sum_\alpha (y_{i\alpha} - \bar{y})^2} \qquad (6)$$

In general, if the target of our problem (the dependant variable) is the value of beta counts at $t_0$ and the input (the independant variable) is formed by the time-delayed series of radon data or by the meteorological parameters at $t_0$, depending on the approach followed, we can group the values of the input into classes. Then the formula (6) defines $R_{nl}$ in terms of the average of the target for every specific *i*th class of the chosen input.

The results of our autocorrelation analysis are shown in Fig. 3: here, the difference between $R$ and $R_{nl}$ is quite sensible only at low correlation values. Furthermore, also anticorrelated points (around $\Delta t = 12$) do not achieve high absolute values. Therefore, even if some nonlinearities are recognized in the data (*via* the difference between $R$ and $R_{nl}$), what reveals the best correlations with data at time $t_0$, both in linear and nonlinear terms, are the data at t: $\Delta t = t_0 - t = 2, 4, 6, 20, 22, 24, 26, 28$.

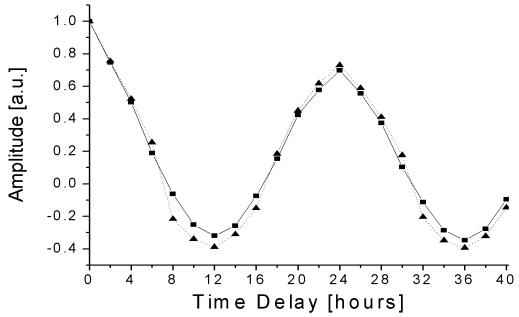


Fig. 3. Autocorrelation analysis in terms of $R$ (square) and $R_{nl}$ (triangle) for our time series.

When we choose the inputs to be inserted into the networks for forecasts at $t_0$+2, $t_0$+4 and $t_0$+6 hours, other independant empirical tests confirm that the inputs identified by the autocorrelation analysis lead to the best prediction performance: therefore, our inputs are determined accordingly. In short, we use networks with one hidden layer and the following topologies: 8-8-1, 7-8-1 and 6-8-1 for forecasts at $t_0$+2, $t_0$+4 and $t_0$+6 hours, respectively.

Due to the evidence of periodicities in the radon time series, clear in the amplitude spectrum shown in Fig. 4, it is worthwhile to search for a preprocessing method in order to model, or at least to extract, these periodicities and to leave to the networks a residual time series which represents the hidden dynamics to be modeled by neural modeling. In doing so, we use Seasonal Differencing (SD): see, for instance, Masters (1995). By SD we build a new "residual" series as a series of differences between the actual value and the value detected 24 hours before.

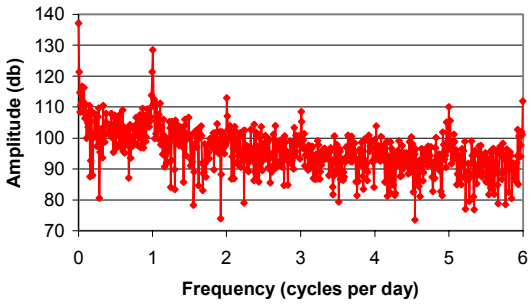


Fig. 4. Fourier spectrum of our time series.

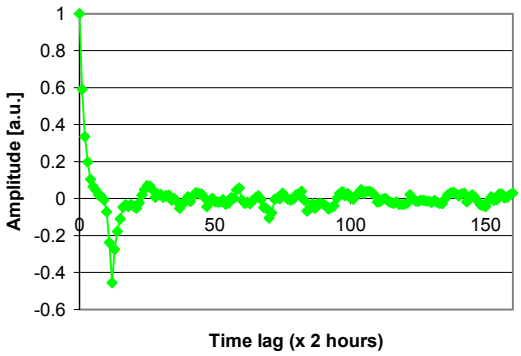


Fig. 5. Linear autocorrelation analysis for the residual time series after application of the SD.

The SD allows us to eliminate the periodicities in the Fourier graph. In the meantime, it leads to a peculiar form for the autocorrelation function of the new residual series, shown in Fig. 5. This form is discussed in details in a work in progress (Ameli and Pasini, in preparation). Anyway, together with several empirical proofs, it leads to choose the inputs in an analogous way to what discussed before. Of course, the final results in the use of preprocessing comes from the combination of the network forecasts on the residual series and the values "detrended" by SD.

As far as the synchronous pattern approach is concerned, an optimal value of epochs and the optimal topology 11-10-1 are found in preliminar proofs. Then we apply *tout court* the cited approach. However, linear and nonlinear statistical bivariate analyses, between any test set of a single input variable and the set of detected beta counts at $t_0$+2, $t_0$+4 and $t_0$+6 hours, reveal that some inputs are correlated with the output in a very poor manner (see Fig. 6 for the $t_0$+2 hours case). This induces us to perform a pruning of some inputs. Advantages of a

pruning activity are many (see, for instance, the discussion in Pasini et al. 2001b). Here we want to stress that, due to our particular model structure (see the discussion about (4) and Fig. 2), the shape of our sigmoids changes with the number of connections, leading in short to different models. In particular, this reduction could even lead to an increase in forecasting performance, when pruning is applied.

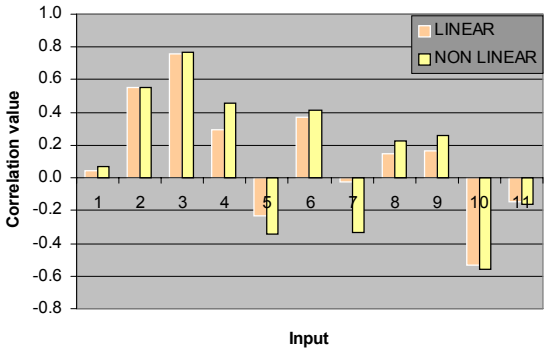


Fig. 6. Bivariate analysis in terms of $R$ and $R_{nl}$. The input variables are numbered as in section 4.

Several runs with pruned inputs are performed and good results come from networks endowed with topology 5-8-1, obtained by retaining the following inputs: hour of the day (two inputs), horizontal wind speed, beta counts and beta counts time derivative.

In what follows we present forecasting results on the period 7 August - 11 October, which represents our test set. In Table 2 we use the continuous performance index $R$ for 3 steps ahead and with reference to the four methods used in the forecasting activity: here the results are shown for TS (time series approach with no preprocessing), TS-PP (time series approach after SD preprocessing), SP (synchronous pattern approach) and SP-PR (synchronous pattern approach after pruning).

Note that:

(1) in the TS approaches we forecast beta counts starting from detected data at $t_0$ even for the second and third steps: attempts at considering forecast values coming from the previous step as a part of input patterns for the following forecast lead to poorer results in our case;

(2) the error bars associated with the neural forecasts are derived from different runs of the model (10 for each time horizon and class of data) with different random weights, so that the model is able to widely explore the landscape of the cost function: these error bars indicate ±2 standard deviations.

| Method\Hour | $R$ ($t_0$+2) |
|---|---|
| TS | 0.812 ± 0.010 |
| TS-PP | 0.894 ± 0.004 |
| SP | 0.849 ± 0.003 |
| SP-PR | 0.854 ± 0.002 |
| Method\Hour | $R$ ($t_0$+4) |
| TS | 0.735 ± 0.017 |
| TS-PP | 0.870 ± 0.004 |
| SP | 0.821 ± 0.005 |
| SP-PR | 0.820 ± 0.005 |
| Method\Hour | $R$ ($t_0$+6) |
| TS | 0.672 ± 0.017 |
| TS-PP | 0.861 ± 0.005 |
| SP | 0.788 ± 0.006 |
| SP-PR | 0.773 ± 0.007 |

Table 2. Linear correlation coefficient (forecast vs. detected).

The results shown in Table 2 allow us to recognize that the synchronous pattern approach always over-performs the original time series approach in a significant statistical way. This bears witness to the relevance of meteorological information in order to better characterize the boundary layer state with the goal of short range forecasting. As far as the pruning activity is concerned, this leads to a little improvement at $t_0$+2 hours, while the results at $t_0$+4 and $t_0$+6 hours are comparable and worse, respectively, with respect to the runs without pruning.

On the other hand, the recognition of periodicities and the following time series preprocessing activity lead to even better performance results, especially at $t_0$+4 and $t_0$+6 hours. This bears witness to the importance of the use of dynamical information, if available, and to the efficiency of the network learning on the dynamics hidden in the residual series.

This last consideration is of course very important, because it shows the potentialities of a neural network model, which is able to recognize a hidden dynamics. Nevertheless, we would like to stress that, if we want to perform an operational forecasting activity by means of neural modeling, the specific SD preprocessing used in this work is no more applicable, because it needs data about the target in the near future in order to build the series of differences. Other preprocessing methods which are not endowed with this negative feature are now under study.

Moving towards an analysis in terms of contingency tables, a discrete assessment of performance, through the indices previously defined, is shown in Figs. 7-13 for the case $t_0$+2 hours. Even here, we compare the four distinct approaches to the forecast of beta counts.

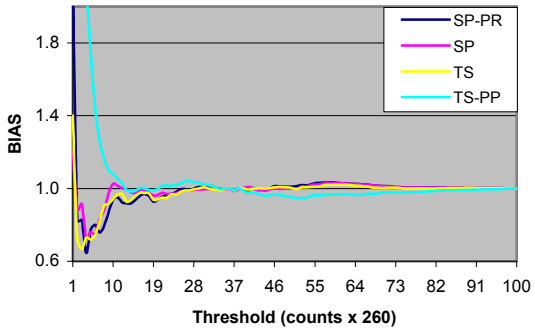


Fig. 7. BIAS

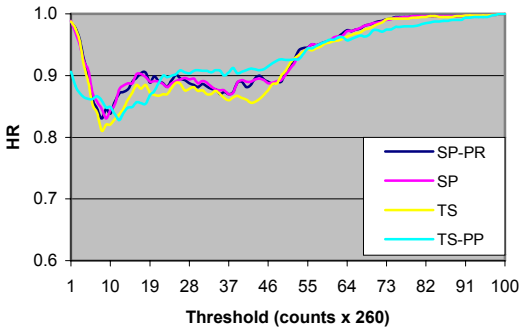


Fig. 10. Hit Rate

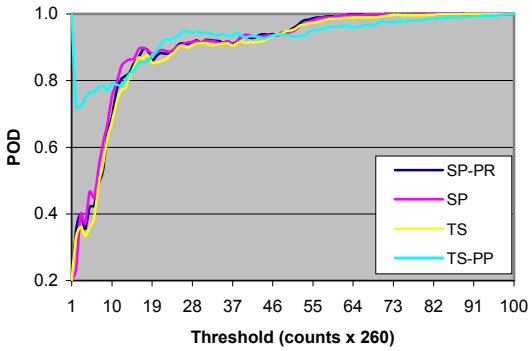


Fig. 8. Probability Of Detection

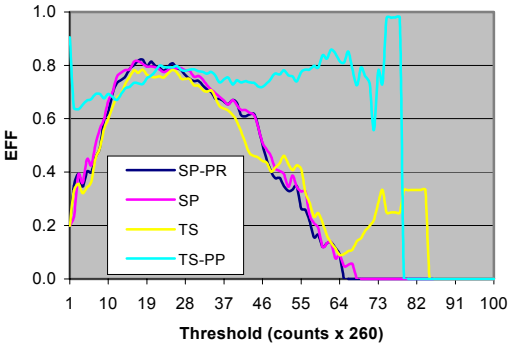


Fig. 11. EFFiciency

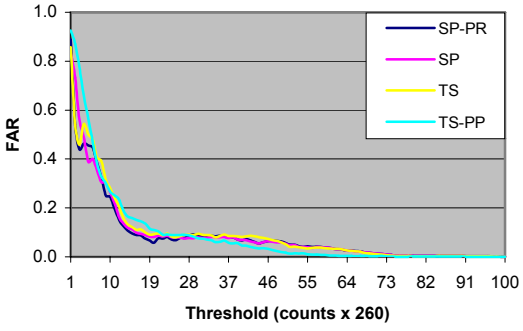


Fig. 9. False Alarm Ratio

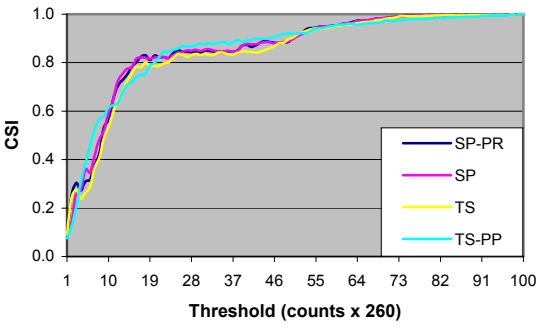


Fig. 12. Critical Success Index

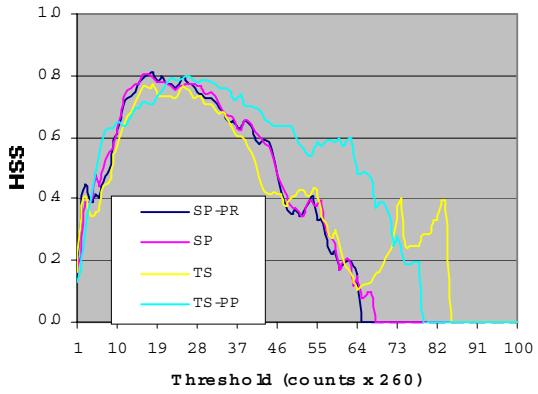


Fig. 13. Heidke's Skill Statistics

An analysis of Figs. 7-13 leads us to understand that:

- in FAR, HR and CSI the differences of performance among the various modeling strategies are nearly absent.
- In BIAS and POD we recognize a characteristic feature of the TS-PP approach, that is to say the under-estimation of the low values of beta counts detected. This is probably due to the transition periods between situations driven by the thermal cycle and advective situations, when the SD modulation leads to very low final forecast values.
- In EFF the time series approach seems to over-perform the synchronous pattern approach at high thresholds, especially in the TS-PP case.
- We think that the most interesting results come from the analysis of the graph related to HSS. In fact, first of all HSS is a very good measure of performance (see, for instance, the discussion in Marzban 1998). Furthermore, Fig. 13 shows that the time series approach reveals very good features for high thresholds; in particular, TS-PP is always better than other approaches for thresholds > 6000 beta counts. If we exclude TS-PP from the operational forecast methods, we can affirm that the SP and SP-PR approaches over-perform TS till a threshold of about 12000 beta counts. Finally, we stress that all the approaches lead to a maximum of performance for a threshold of about 5000 beta counts: this is very important for us, because this value can be considered as the threshold which allows us to distinguish between advective situations and maxima due to the presence of nocturnal stable layers driven by the vertical thermal state of the atmosphere.

After these global statistical considerations, it is interesting to look at a little subset of our time series of radon data and to see how these values are forecast by neural models that use the various strategies adopted here. Original series and forecasting results of this case study are shown in Fig. 14.

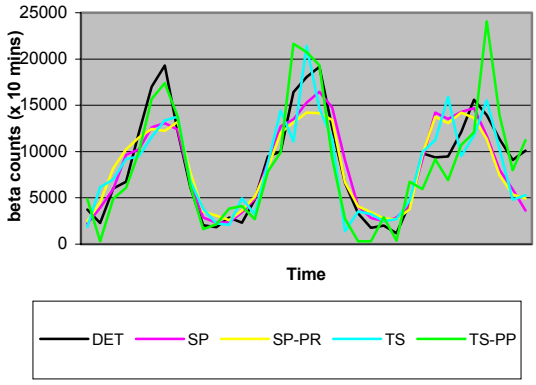


Fig. 14. Beta counts forecast in a case study.

Fig. 14 shows the general satisfying results of our forecasting activity. Here we note the quite irregular forecast of the TS approach which leads to some large oscillations and counter-tendency derivatives in the beta counts evolution. Of course, the preprocessing smooths this effect and, in the meantime, leads to more accurate forecasts. The synchronous pattern approach gives more regular pattern tendency, with some case of underestimation of the peaks.

Finally, we can perform an estimation of the nocturnal $h_e$ by means of the cited box model for data coming from the detected beta counts and the four forecast series. Results concerning the case study just shown in Fig. 14 are presented in the following Fig. 15.

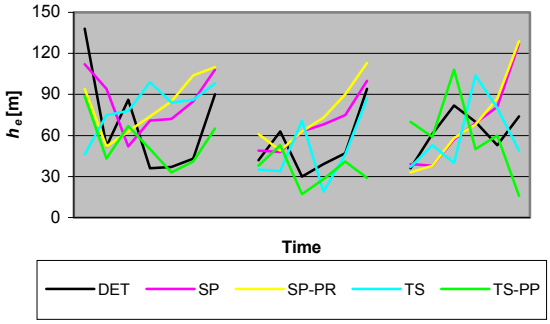


Fig. 15. Mixing height forecast in a case study.

In Fig. 15 we note again counter-tendency derivatives, now in the evolution of the nocturnal mixing height, if we consider the pure time series approach TS. Here the preprocessing activity leads to very accurate forecasts and avoids this problem. If we compare TS-PP with SP and SP-PR, we note that TS-PP is superior. This is more evident now than in Fig. 14, because the very small and smooth increase in beta counts in cases of SP and SP-PR applications leads to small increases (and not decreases) of $h_e$: this is due to the box model, which includes also radon decay (see Pasini et al. 2002).

## 7. CONCLUSIONS AND PERSPECTIVES

In this paper a short range neural forecast of beta counts, coming from radon progeny measurements, is performed. The use of four distinct strategies in data handling and neural network applications is presented and discussed. The results show the concrete possibility to successfully handle this kind of short range forecasting. The further application of a box model leads to obtain reliable forecasts of the nocturnal stable layer depth, at least in terms of the best approaches.

The perspectives of further study mainly concern the possibility of extending the forecast range, in order to obtain useful information about the future radon concentrations and mixing height values for the day after. It seems to us this approach requires a post-processing activity on the outputs coming from a dynamical meteorological model.

## REFERENCES


Allegrini, I., A. Febo, A. Pasini, and S. Schiarini, 1994: Monitoring of the nocturnal mixed layer by means of particulate radon progeny measurement, *J. Geophys. Res.*, **99 (D9)**, 18765-18777.

Bishop, C.M., 1995: *Neural networks for pattern recognition*, Oxford University Press, 482 pp.

Guedalia, D., A. Ntsila, A. Druilhet, and J. Fontan, 1980: Monitoring of the atmospheric stability above an urban and suburban site using sodar and radon measurements, *J. Appl. Meteor.*, **19**, 839-848.

Hertz, J., A. Krogh, and R.G. Palmer, 1991: *Introduction to the theory of neural computation*, Addison-Wesley, 327 pp.

Lopez, A., D. Guedalia, J. Servant, and J. Fontan, 1974, Advantages of the use of radiactive tracers 222 Rn and 212 Pb for the study of Aitken nuclei within the low troposphere, *J. Geophys. Res.*, **79**, 1243-1252.

Mahrt, L., 1998: Stratified atmospheric boundary layers and breakdown of models, *Theoret. Comput. Fluid Dynamics*, **11**, 263-279.

Marzban, C., 1998: Scalar measures of performance in rare-event situations, *Weather Forecasting*, **13**, 753-763.

Marzban, C., 2002: Neural networks for regression analysis, *Proc. Short Course on Neural Network Applications to Environmental Sciences*, V. Krasnopolsky, Ed., Orlando, FL, American Meteorological Society (available at www.nhn.ou.edu/~marzban).

Marzban, C., E.D. Mitchell, and G.J. Stumpf, 1999: On the notion of "best predictors": An application to tornado prediction, *Weather Forecasting*, **14**, 1007-1016.

Masters, T., 1995: *Neural, Novel & Hybrid Algorithms for Time Series Prediction*, Wiley, 514 pp.

Pasini, A., and S. Potestà, 1995: Short-range visibility forecast by means of neural-network modelling: a case study, *Nuovo Cimento*, **18C**, 505-516.

Pasini, A., F. Ameli, and V. Pelino, 2001a: A neural-network approach to radon short-range forecasting from concentration time series, *Nuovo Cimento*, **24C**, 331-337.

Pasini, A., V. Pelino, and S. Potestà, 2001b: A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables, *J. Geophys. Res.*, **106 (D14)**, 14951-14959.

Pasini, A., F. Ameli, and A. Febo, 2002: Estimation and short-range forecast of the mixing height by means of box and neural-network models using radon data, *Physical Chemistry 2002: Proceedings volumes* , S. Anić, Ed., Belgrade, Jugoslavia, Society of Physical Chemists of Serbia, 607-614.

Seinfeld, J.H., and S.N. Pandis, 1998: *Atmospheric chemistry and physics: from air pollution to climate change*, Wiley, 1326 pp.

Stull, R.B., 1988: *An introduction to boundary layer meteorology*, Kluwer, 666 pp.

Vinod Kumar, A., V. Sitaraman, R.B. Oza, and T. Krishnamoorthy, 1999, Application of a numerical model for the planetary boundary layer to the vertical distribution of radon and its daughter products, *Atmos. Environ.*, **33**, 4717-4726.