Neural Networks for Characterization and Forecasting in the Boundary Layer *via* Radon Data

13

Antonello Pasini

13.1 Introduction

The complexity of air-pollution physical-chemical processes in the boundary layer (BL) is well known: see, for instance, Stull (1988) and Seinfeld and Pandis (1998). In this framework, we do not make any attempt at reviewing the manifold use of neural networks (NNs) for air-pollution assessments and forecasting. Instead, we focus just on the (complex) physics of the BL and discuss the coupled use of an original index of the BL properties (radon concentration) and of NN modeling in order to obtain interesting results for characterizing and/or forecasting important variables in the BL, like the concentration of a dangerous primary pollutant (benzene) and the 2-h evolution of stable layer depth. In this scenario, the particular strategies for applying a NN model are described, showing how they lead to important original results, for grasping the BL physical behavior. In doing so, one can discover the usefulness of an empirical AI data-driven approach to investigating a complex system that is very difficult to deal with in terms of dynamical models.

In the next section, a brief introduction to fundamentals of radon detection will be presented and the qualitative and quantitative relevance of radon concentration for summarizing the physical state of the BL will be discussed. In particular, we will present the structure of a box model (based on radon data) for estimating the nocturnal stable layer depth.

In Section 13.3, after having introduced simple linear indices that characterize days and nights in terms of meteorological predisposition to a primary pollution event, generalized nonlinear indices are built up by a NN model: they show the ability of NNs at capturing nonlinearities in the data and achieving better modeling results.

In Section 13.4, by applying NN modeling and a particular preprocessing activity, we will show how we are able to achieve reliable short-range forecasts of radon concentration and stable layer depth in an urban environment. These forecasts represent fundamental information for assessing dispersion properties and their related influence on pollutant concentrations.

Finally, in the last section, brief conclusions will be drawn and prospects of future developments in this field will be envisaged.

13.2 Relevance of Radon in Studies of the Boundary Layer

As well known, radon is an important factor that can lead to lung cancer (see www.epa.gov for further information). Due to this fact, many epidemiological studies on radon have been performed during the last decades, especially in indoor environments. Less known is the role of radon as a "tracer" of the physical characteristic features of the lower layers of the atmosphere.

Despite this fact, it is worthwhile to stress that the first studies on the role of radon in characterizing BL

Antonello Pasini (✉)
CNR – Institute of Atmospheric Pollution, Via Salaria Km 29.300, I-00016
Monterotondo Stazione (Rome), Italy
Phone: + 39 06 90672274; fax: + 39 06 90672660;
email: pasini@iia.cnr.it

S. E. Haupt et al. (eds.), *Artificial Intelligence Methods in the Environmental Sciences*

dispersion properties dated back to the late 1970s, when French researchers began to consider this noble gas, which undergoes no chemical reaction, as a perfect tracer of the BL dilution features. They found that counts of beta radioactivity, coming from the decay of short-lived radon daughters, represent a simple index of the stability state of the BL (see Guedalia et al. 1980, and references therein).

In general, the search for an index which is able to summarize the characteristic features of a system is seen merely as a simplification (sometimes used for working out a conceptual model of the behavior of the system) that is unnecessary and can be overcome by a dynamical description of the system itself. Nevertheless, in a highly nonlinear system, the knowledge of a suitable index whose time development mimics the behavior of some intrinsic property of the system itself gives us key information, certainly in a qualitative way and, hopefully, also in a more quantitative manner.

13.2.1 Semi-quantitative Information by Estimating an Equivalent Mixing Height

Without discussing the instrumentation for detecting beta counts from the decay of short-lived radon progeny, we just stress that the fraction attached to particulate matter is usually detected (see Allegrini et al. 1994, for details). As shown in Fig. 13.1, the typical time patterns of beta counts are maxima during the night in conditions of nocturnal stability and minima during the day when the mixed layer is well developed (the more enhanced is the stability of the nocturnal stable layer, then the higher are the maxima in radioactivity counts). Otherwise, low quasi-constant values are found in advective situations characterized by mechanical turbulence. This qualitative analysis suggests that, in general, the number of beta counts can be inversely proportional to the "degree" of stability of the lower layers.

Especially interesting cases are nocturnal stable situations over towns, when anticyclones at synoptic scale and local physiographic and emissive features can create conditions for the development of strong stable layers and peak events of primary pollution. In these cases, rain is absent, relative humidity and pressure are quite constant and the spacetime interval to be analyzed is limited (a night and a town). Thus, if we limit our study to these situations, then 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.

Further evidence (see Lopez et al. 1974, for pioneering work and Vinod Kumar et al. 1999, for more recent results) shows 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. This fact induced Guedalia et al. (1980) to use a box model of the nocturnal stable layer endowed with a homogeneous radon concentration in the vertical. The top of this box, named equivalent mixing height h_e, has been found to be a good index of the dispersion properties of this layer in a semi-quantitative way, because low (high) values of h_e are related to low (high) dispersion power and high (low) concentrations of primary pollutants.

In Guedalia et al. (1980) the calculation of the top of the box has been performed by means of the following equation:

$$h_e(t) = \frac{\Phi \Delta t}{C(t) - C_0} \tag{13.1}$$

Here Φ is the radon flux at the surface, Δt is the time interval from the start of accumulation, $C(t)$ is the radon concentration at time t and C_0 is the radon concentration at the beginning of accumulation (evening).

13.2.2 A Physical Interpretation of the Equivalent Mixing Height

Guedalia et al. (1980) wondered what h_e should represent from a physical point of view. They correctly asserted that it is different from the inversion layer thickness, and that instead, it should be identified with the base of the inversion, although even this option is not the precise one (in particular, for ground-based inversions, the box model is no longer applicable).

The correct solution to this puzzle has been given by Allegrini et al. (1994). They show that h_e can be

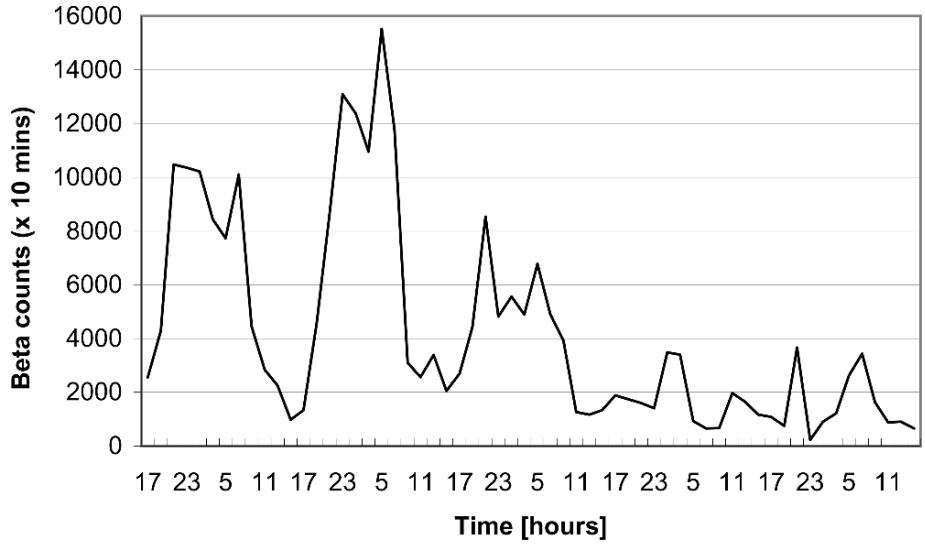

Fig. 13.1 A typical time series of beta counts: the first three maxima refer to stable nights and then one must appreciate a transition to conditions of moderate advection

identified with the height at which a parcel of air coming from the ground halts its free convection, at least in nocturnal stable situations dominated by the thermal factor. They confined their research to some experimental campaigns in a town and used the fundamental phenomenon of the urban heat island that, even in the presence of ground inversions in suburban sites, permits the creation of a shallow mixed layer where convection is not suppressed.

In order to quantify the nocturnal stable layer depth over a town, they used a thermal profile from a radiosonde station in the suburbs and the surface air temperature at the radon detection site inside the town. With a potential temperature method or, equivalently, by means of a simple graphical representation (see Fig. 13.2), they were able to estimate an urban mixing height, h_u, from meteorological data. A statistical analysis show that h_e and h_u are highly correlated, so inducing to think that the box model output represents a correct estimation of the urban mixing height, at least in situations of high nocturnal stability.

In general, we must be aware that h_e and h_u can differ from the real value of the stable layer depth, the equality being valid at the limit of null mechanical turbulence, when the physical features of the BL are completely driven by the vertical thermal state of the atmosphere. Furthermore, while h_u has no chance to be sensitive to nonthermal factors, h_e represents an

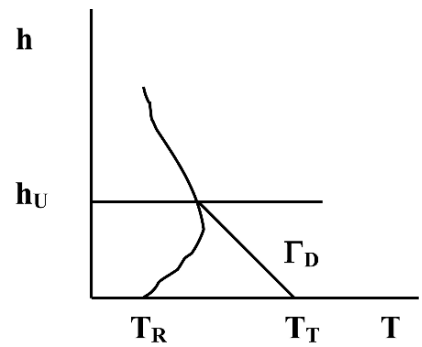

Fig. 13.2 Estimation of the urban mixing height: T_R is the temperature at the radiosonde station, T_T is the temperature inside the town, Γ_D is a dry adiabat drawn from T_T that intercepts the vertical profile

index whose value is determined by all the factors that influence the dilution properties of the BL.

As a final remark, we note that, if direct measurements of Φ are not available, its value can be estimated by inverting equation (13.1) in which we substitute h_e with its meteorological estimation h_u at a certain hour in the evening, at the beginning of accumulation.

13.2.3 An Improved Box Model

Thus, in strong stable situations during the night, we are able to monitor the height of the stable layer

over a town by means of radon detection and the application of a simple box model. This estimation furnishes important physical information about the volume available for the dilution of pollutants emitted at the ground. In particular, this method allowed us to explain critical peak events of primary pollution during the night due to negative fluctuations (reductions) of the stable layer depth. Sometimes, however, the nocturnal fluctuations in h_e, shown by the box model, assume wide unphysical values: then, in order to achieve more realistic modeling of the nocturnal BL behavior and to avoid these problems, a new version of the model has been recently worked out and preliminarily presented in Pasini et al. (2002).

As a matter of fact, the structure of the box model described above is too simple for at least two reasons: first, radon decay is neglected; secondly, entrainment of air with different radon concentrations is not allowed from the top of the box (this is critical just in situations characterized by nocturnal fluctuations of the stable layer depth). A new model structure, which includes these elements, is briefly presented here.

In Fig. 13.3 "compressions" are the situations in which the stable layer depth decreases (i = 1, 2, 3, 6) and "expansions" are the cases when h_e increases (i = 4, 5). In what follows, λ represents the decay constant of radon, Δt is our sampling rate (usually 2 h), C^a is the calculated concentration in the residual layer and we adopt the symbolic form $\Delta h_e(\mathrm{n}, \mathrm{m}) = h_e(\mathrm{n}) - h_e(\mathrm{m})$ for the difference between equivalent mixing heights at time n and m, respectively.

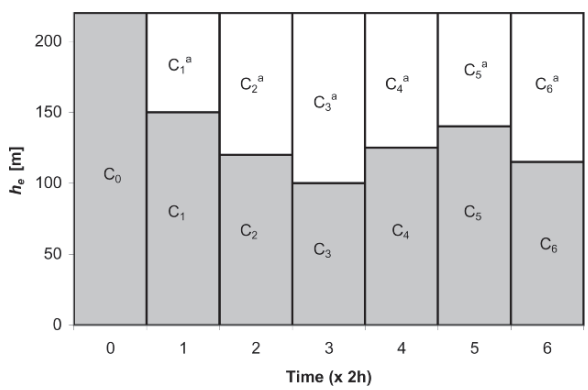

Fig. 13.3 Time evolution of the nocturnal stable layer (gray box). The white area above represents the residual layer

In compression cases the generalization of equation (13.1) reads as follows:

$$h_e(i) = \frac{\Phi}{\lambda} \frac{1 - \exp(-\lambda \Delta t)}{C_i - C_{i-1} \exp(-\lambda \Delta t)} \tag{13.2}$$

The concentration left in the residual layer after the ith compression is:

$$C_i^a = \frac{C_{i-1}^a \exp(-\lambda \Delta t) \cdot \Delta h_e(0, i-1) + C_{i-1} \exp(-\lambda \Delta t) \cdot \Delta h_e(i-1, i)}{\Delta h_e(0, i)} \tag{13.3}$$

If the stable layer depth increases and overlying air is included in the box, i.e. in cases of expansions, the equivalent mixing height can be calculated as:

$$h_e(i) = \frac{(\Phi/\lambda)[1 - \exp(-\lambda \Delta t)] + h_e(i-1)(C_{i-1} - C_{i-1}^a) \exp(-\lambda \Delta t)}{C_i - C_{i-1}^a \exp(-\lambda \Delta t)} \tag{13.4}$$

and the concentration over the top of the box is of course

$$C_i^a = C_{i-1}^a \exp(-\lambda \Delta t) \tag{13.5}$$

This enhanced structure of the box model, tested on past results, leads to a little increase (about 1%) in the value of the linear correlation coefficient between the total set of h_e estimated by the new model and the total set of h_u. The actual, relevant and statistically significant improvement is obtained in critical situations such as nocturnal fluctuations of the stable layer depth, when the application of equations 13.2–13.5 prevents the estimation of wide unphysical oscillations.

Thus, data from radon progeny measurements well represent (in a synthetic way) the dilution properties of the lower layers of the atmosphere and the analysis performed here has led to characterizing these features both qualitatively and quantitatively. In the next section, after having presented another useful application of radon data to primary pollution characterization in an urban environment, we will begin to "handle" these data with a NN model.

13.3 Linear Stability Indices and Their Nonlinear Generalization by NNs

As we have seen, knowledge of radon concentration, or simply of the beta counts coming from the decay of short-lived radon daughters, gives us information about the physical state of the BL. If we consider the so-called primary pollutants, their concentration in the lowest layers of the atmosphere is driven by two main factors: the emission flux and the mixing properties of the BL. In this situation, our knowledge of these physical properties can lead to studying the role of the BL in the accumulation of these pollutants.

13.3.1 Linear Stability Indices

Some years ago, Italian researchers (Perrino et al. 2001) developed stability indices based on radon data in order to characterize days and nights (in every season) in terms of their meteorological predisposition to a primary pollution event.

Substantially, these indices are scalars: they are the results of equations (coming from multiple linear regressions) that aim at reconstructing the mean concentration of benzene on a 12-h interval. The multiple linear regression considered 2-h beta count data and their time derivatives as predictors and benzene concentration during night or day as predictand. For these calculations the year has been divided in three periods, according to different duration and intensity of the solar radiation (winter period: October to February; summer period: May to July; intermediate period: March, April, August, September). A total of six indices has been therefore obtained.

The correlation between the values of these indices and the concentration of benzene is generally good, but it must be highlighted that these indices take into account only one of the two driving forces in determining primary pollutant concentration (the mixing properties of the BL), so that a one-to-one correlation could be possible only in the theoretical case of a constant emission flux. Thus, the scope of indices determination must not be estimating the correct value of observed benzene concentrations, but, more properly, these indices can become a tool for uncoupling the roles of meteorology and emissions for determining the concentration of a primary pollutant.

In fact, these indices allow us to estimate the concentration of a primary pollutant uniquely due to the contribution of the meteorological factor. Therefore, the environmental applications of these indices are especially important in the study of the differences between this estimation and the primary pollutant concentration actually observed. This is of help in identifying days when the atmospheric pollution is heavier (lighter) than predictable on the basis of the atmospheric mixing and allows us to evaluate, for example, the real effect of traffic restriction measures. Also, at a longer range, we are able to understand if a decrease in the air concentration of a given primary pollutant from one year to another is due to a real improvement of the air quality or, instead, only to a lighter meteorological situation.

For further details on the calculation of these linear indices and their application, see Perrino et al. (2001).

13.3.2 Nonlinear Stability Indices via NN Modeling

As stated in the previous subsection, the aim in using the stability indices is not the accurate reconstruction of benzene values in every situation, since these indices take only the BL dilution factor into account and neglect the contribution of emission variations to benzene concentration. With a change of perspective, we recently asked if this meteorological contribution to benzene behavior is correctly modeled (Pasini et al. 2003c). In fact, if we consider that the linear stability indices are obviously not able to fully capture the complex nonlinear relationships among different variables in the BL, we can suggest the use of a more complex nonlinear regression method and hope that its use can lead to an increase of the amount of variance explained by the linear regression. Therefore, after a preliminary statistical analysis of the data available, in order to discover nonlinearities hidden therein, a NN model is applied to the problem of reconstructing benzene concentrations by means of radon data, even in cases not included in the training set.

As far as the neural model is concerned, the NN development environment briefly described at Subsection 12.3.3 of the previous chapter is used even in this case study. We briefly remind that our NNs are multi-layer perceptrons endowed with one hidden

layer and backpropagation training. Furthermore, it is worthwhile to stress that some specific tools are present in our development environment in order to handle historical data and to train NNs starting from quite short time series. We have seen a first example of these training tools in Chapter 12 of this book and we will meet another example in the next section, where a sketch of our normalized sigmoids will be also be presented, together with a discussion of their influence on the model structure.

The application of a nonlinear NN model is suggested by the results of an *a priori* bivariate statistical analysis that is able to estimate linear and nonlinear correlations between each predictor/input (a 2-h radon detection or its time derivatives) and the predictand/target (mean benzene concentration during day or night). By this analysis we compare values of the linear correlation coefficient R and values of the so-called correlation ratio, a nonlinear generalization of R (see Pasini et al. 2001, 2003a for technical details on the fundamentals of such an analysis). Differences are found between linear and nonlinear correlation values for the same input-target sets and, sometimes, inputs showing low linear correlation with the target assume high nonlinear correlation values. In short, even if the correlation ratio does not measure all types of nonlinearity, for our problem it allows us to understand that some nonlinearities are hidden in the relationships among the variables. In particular, as a consequence of this statistical analysis, the inputs considered for an optimal nonlinear regression can be generally different from the variables chosen for an optimal linear regression.

As in the development of linear stability indices, we consider six records of cases (diurnal and nocturnal situations for winter, summer and intermediate periods). Each record is divided into three sets: the first months represent the training set and include the validation set (useful in order to establish the threshold for early stopping and chosen as 15 random days inside these first months), the last month of each record is the test set on which we assess the networks' ability to generalize. When comparing linear and NN models' performance on the test set, we obtain a statistically significant improvement in "simulating" the behavior of day and night mean benzene concentrations: our NN model explains 77% of the variance in benzene data, while the linear model achieves a performance of 68%.

Table 13.1 Performance in the modeling of benzene concentrations on the test sets

Period	Linear model	Neural model
Winter – morning	0.786	0.809 ± 0.038
Winter – evening	0.740	0.812 ± 0.020
Summer – morning	0.558	0.576 ± 0.037
Summer – evening	0.639	0.725 ± 0.031
Intermediate – morning	0.664	0.877 ± 0.012
Intermediate – evening	0.698	0.795 ± 0.038

Source: Adapted from Pasini et al. (2003c). With kind permission of Società Italiana di Fisica.

More specifically, the values of several indices of performance have been calculated and consistent results are obtained on the test sets in the six cases cited above. In Table 13.1 the results of both the linear model and the NN model are shown in terms of the linear correlation coefficient (detected vs. modeled). Here, as in applications described in the previous chapter, the error bars associated with the NN performance come from ensemble runs of the model with different initial random weights, so that each network is able to widely explore the landscape of its cost function, and represent ± 2 standard deviations.

As one can see, the majority of improvements obtained by the application of the fully nonlinear NN model is statistically significant. Furthermore, the calculation of the bias allows us to appreciate that, even in the few cases when the statistical significance of performance improvement is not sure, the systematic error in the results of the NN model is lower than that of the linear model: this gives us more reliable results.

In short, the application of NNs to this problem leads to better modeling the meteorological contribution to the behavior of mean benzene concentration over days and nights in distinct periods of the year, if compared with a multiple linear regression.

13.4 Neural Forecasting of the Radon Concentration and Short-Range Estimation of the Nocturnal Stable Layer Depth

Once the ability of a NN model for characterizing the meteorological-induced behavior of a primary pollutant is shown, it would be interesting to test this model in forecasting BL physical features. In this framework, using the knowledge of radon behavior and applying

the box model previously developed, we concentrate on stable nocturnal situations in the BL.

As a preliminary remark, it is worthwhile to note that the physics of the nocturnal stable layer has been recognized as very complex and modeling in this domain represents a major challenge for physicists of the atmosphere (see, for instance, Mahrt 1998). Thus, once again, an AI empirical data-based method (NN modeling) could help in a framework where dynamical modeling shows drawbacks.

In what follows, several forecasting strategies will be considered and particular attention will be paid to didactical aspects rather than technical details: see Pasini and Ameli (2003), Pasini et al. (2003a, b) for a more detailed treatment.

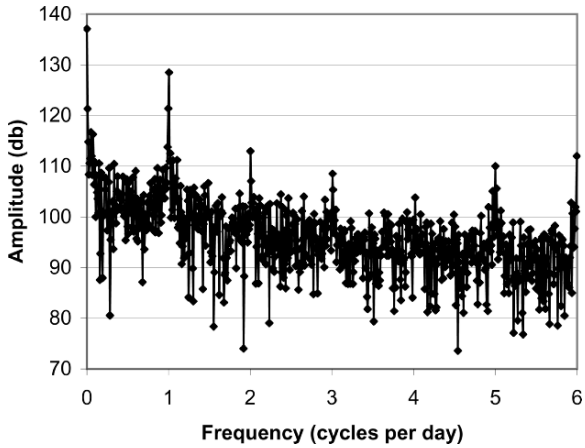

Fig. 13.4 Fourier spectrum of our time series of beta counts

13.4.1 A Time Series Approach: Black Box vs. Preprocessing

The Institute of Atmospheric Pollution of the Italian National Research Council has sponsored several extended duration monitoring campaigns to detect beta counts from radon progeny decay. Thus, long time series at the time resolution of 2 h are available: here, we analyze data from the entire year 1997, detected in a site near Rome, Italy.

Since the development of the model of multi-layer perceptron, NNs have shown their ability to forecast time series data for some steps in the future and they often beat other methods in intercomparison studies: see, for instance, Weigend and Gershenfeld (1993). Therefore, it is natural to apply NNs to a short-range forecast of beta-count time series in order to obtain accurate estimations for the values of this index of the BL dilution capacity.

Usually, when the global dynamics of the system under study is unknown, a NN is used as a black box that outputs future values of the time series, given input of a sequence of its past values. Of course, we can follow this approach even in the treatment of our forecasting problem in the BL: as we will see, this has actually been done with quite good results. Nevertheless, here a part of the dynamics is known: for instance, the influence of a day-night cycle is very clear in situations characterized by nocturnal stability and well developed diurnal mixed layers (look at the first 3 days in Fig. 13.1). Furthermore, a 24-h periodicity and its sub-harmonics are evident in a Fourier spectrum of our time series (see Fig. 13.4). In this situation, this part of dynamics can be modeled or, at least, simply described, leading to using this knowledge before applying a NN for studying the system.

In practice, we chose to preprocess our data. We filtered out the known periodicities (by means of the so-called Seasonal Differencing (SD) method[1]), subtracted this signal from the detected data and modeled the residuals with NNs. In this manner we left the network model the hidden (unknown) dynamics. Of course, the forecasting results of the neural modeling must be added to the preprocessing function in order to obtain final forecasts of the original time series.

The NN tool cited above and briefly described in the previous chapter has also been used in this case study. The choice of inputs is widely discussed in Pasini and Ameli (2003) and the reader can refer to that paper for details. Here we just stress that a particular training-test iterative procedure has been applied. Due to the recognition of a negative effect of old data on forecasting results, we fix a "50-day memory" of training cases and update it for every new forecast, thus limiting the training to the same "season" of any forecast case. Figure 13.5 shows this "moving window" strategy: the window is the 50-day memory (training set), it is updated every 2 h by inserting the latest detected

[1] In our application, SD consists in subtracting from the time series a replica of the series itself delayed by a 24-h time lag. In doing so, we obtain a residual series that is practically lacking in those periodicities shown by the original data.

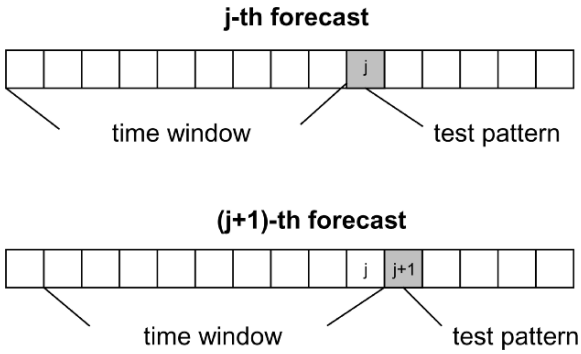

Fig. 13.5 The "moving window" procedure for training and test (forecast) of historical data

data and discarding the oldest ones, then allowing new training and a new forecast.

Once our data is preprocessed by SD, we obtain a residual time series that looks like noise, so that a man-made analysis does not allow us to glean additional clear dynamics in these data. In this situation, the application of a NN model to forecasting this residual time series and to evaluating its forecasting performance could lead to determining whether nonlinear dynamics is hidden in these data, or rather, if they are actually random.

As a matter of fact, we found quite good performance on the test set of residuals when forecasts were extended to 2, 4 and 6 h after the most recent data in the time series. In particular, we calculated the linear correlation coefficient (modeled vs. detected) and obtain $\mathrm{R}(t_0+2) = 0.675 \pm 0.006$, $\mathrm{R}(t_0+4) = 0.500 \pm 0.008$ and $\mathrm{R}(t_0+6) = 0.431 \pm 0.009$. Even if these values are not very high (remember that the time series looks like noise), they still indicate a clear signal of nonlinear dynamics hidden in the residuals. In this way the NNs show their ability in capturing this dynamics and can add information to modeling the system. In particular, they allow us to improve our forecasting performance on the system itself.

Furthermore, a black box strategy and a preprocessing one can be compared in NN modeling on this case study. Table 13.2 shows the performance of the direct application of NNs to forecasting the original time series (TS) and of the joint application of SD and NN forecasting of the residuals (SD-TS): the SD-TS approach outperforms the black box one in all cases and the increment in performance is particularly high at the longer ranges.

Table 13.2 Performance on the test set for the black box (TS) and the preprocessed (SD-TS) strategies in a time series approach

Period	$\mathrm{R}(t_0+2)$	$\mathrm{R}(t_0+4)$	$\mathrm{R}(t_0+6)$
TS	0.812 ± 0.010	0.735 ± 0.017	0.672 ± 0.017
SD-TS	0.894 ± 0.004	0.870 ± 0.004	0.861 ± 0.005

In general, the shape of the signal for beta detections/radon concentrations is well forecasted by SD-TS on all the 24-h interval (nights, days, intermediate periods). In particular, SD-TS prevents the forecasting of counter-tendency behaviors in radon evolution (e.g., increasing beta counts for situations of a detected decrease of beta counts), which are quite often present in the TS forecasts.

If we concentrate on nights with stable conditions in the lower layers, we can apply the box model cited above and obtain forecasts for the depth of the nocturnal stable layer. In Fig. 13.6 an example of h_e 2-h forecasts is presented for three nights: one can recognize the very good results obtained by the joint application of preprocessing and NN model (SD-TS). Due to the very impact of nocturnal BL depth fluctuations on primary pollutants' behavior, a useful feature of this approach is that it generally leads to a correct forecast for the sign of derivatives in h_e records, while the black box approach often presents counter-tendency derivatives in its forecasts.

13.4.2 A Synchronous Pattern Approach with Meteorological Data

In the previous subsection a typical time series approach has been adopted for NN application by using delayed values of beta counts data as input. When the behavior of an index-variable (here beta counts) is influenced by the state of a physical system (here the BL), however, it is a good idea to estimate this state at a time t_0 through the values of certain critical variables and to search for a relationship linking this estimation at time t_0 with future values of the index-variable by means of NNs.

An estimation of the state of the BL can be given through monitoring performed by a standard weather station *in situ*. In our case study, 1-h weather parameters were available from the local meteorological station, so that we considered the following synchronous

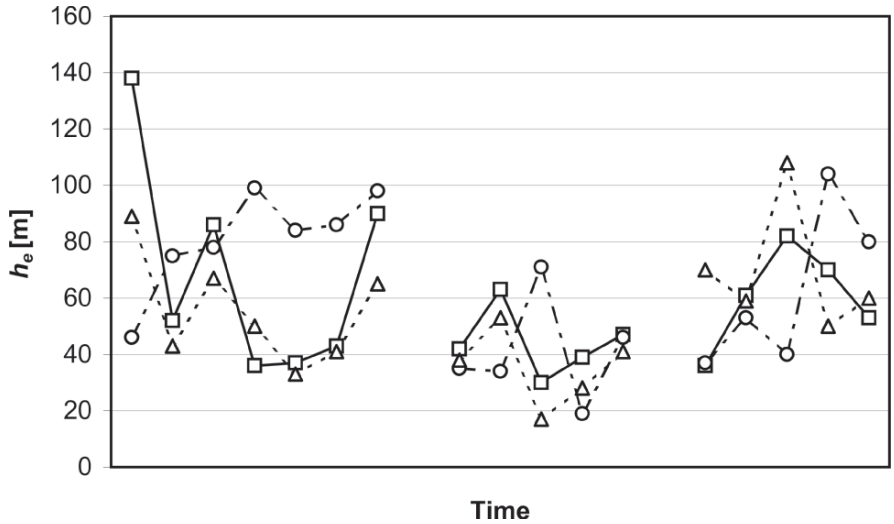

Fig. 13.6 The depth of the nocturnal stable layer calculated by the box model for detected data (square), non-preprocessed forecast values by TS (circle) and preprocessed forecast values by SD-TS (triangles)

pattern of variables which describes the BL state at time t_0: hour of the day (expressed in two inputs as $[\sin(\pi t/12)+1]/2$ and $[\cos(\pi t/12)+1]/2$), beta counts, time derivatives of beta counts with respect to 2 h before, sky covering and height of the lowest cloud layer, temperature, dew point, pressure, horizontal wind speed, and visibility.

Thus, the first attempt can be to apply this approach and to analyze the related NN forecasting performance. 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 h, however, show that some inputs are poorly correlated with the output. This induces prune some of the inputs.

Pruning is a consolidated technique and the description of its advantages can be easily found in the literature. Here we just stress that, as cited in Subsection 12.3.3 of the previous chapter, the transfer functions in our tool are sigmoids in which the arguments of the exponential function are normalized with respect to the number of connections converging to a single neuron of the hidden and output layer, respectively. For the hidden layer, for instance, we have:

$$g_j\left(h_j^{\mu}\right)=\frac{1}{1+\exp\left(-\frac{h_j^{\mu}}{\sqrt{n_{hl}}}\right)} \tag{13.6}$$

where n_{hl} is the number of connections converging to a single neuron of the hidden layer.

Figure 13.7 shows the consequences of this normalization on the shape of sigmoids: in practice, it leads to transfer functions which are less nonlinear when one moves from networks with few connections to bigger ones. In short, this leads to different models for different networks, with more nonlinear transfer functions for small NNs: this could even lead to an increase in forecasting performance when pruning is applied in a strongly nonlinear system.

The best performance with a pruned network came from considering: the hour of the day (two inputs), the horizontal wind speed, beta counts, and time derivative of beta counts. Thus two networks with 11 and 5 inputs, respectively, have been applied to this forecasting activity. The performance results are shown in

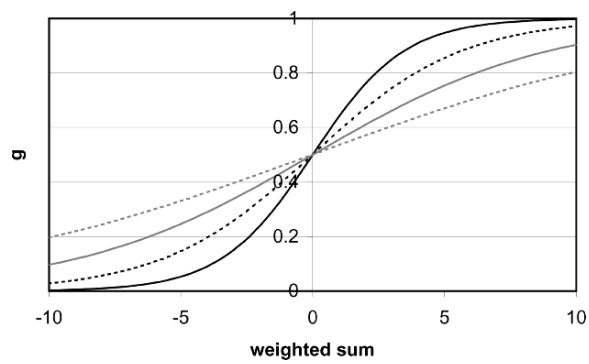

Fig. 13.7 Form of sigmoids in our NN tool for different numbers n of connections converging to a single neuron of the hidden or output layer: solid curve, $n=3$; heavy dashed curve, $n=8$; shaded curve, $n=20$; shaded dashed curve, $n=50$ Source: Pasini et al. (2001). Copyright AGU

Table 13.3 Performance on the test set in a synchronous pattern approach for the full set of inputs (SP) and in the pruned case (SP-PR)

Period	R($t_0 + 2$)	R($t_0 + 4$)	R($t_0 + 6$)
SP	0.849 ± 0.003	0.821 ± 0.005	0.788 ± 0.006
SP-PR	0.854 ± 0.002	0.820 ± 0.005	0.773 ± 0.007

Table 13.4 Contingency table at a fixed threshold distinguishing between events and nonevents

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

Table 13.3 in terms of the linear correlation coefficient between modeled (forecasted) and detected beta counts.

A comparison of these results with those presented in Table 13.2 for the time series approach allows us to recognize that the synchronous pattern approach always outperforms the original time series approach (TS) in a statistically significant way. This bears witness to the relevance of meteorological information to better characterize the BL state for short range forecasting. Pruning leads to a little improvement at $t_0 + 2$ h, while the results at $t_0 + 4$ and $t_0 + 6$ h are comparable and worse, respectively, with respect to the runs without pruning.

On the other hand, the recognizing periodicities and preprocessing the time series (SD-TS in Table 13.2) leads to even better performance results, especially at $t_0 + 4$ and $t_0 + 6$ h. This bears witness to the importance of using dynamical information when available.

In Subsection 13.4.4 we will combine the strong points of these two approaches in order to build a hybrid approach that allows us to obtain better forecasting performance. Now, we would like to briefly analyze performance through indices calculated on contingency tables.

13.4.3 Some Measures of Performance

As discussed in Chapter 3 of this book, the problem of performance assessment is manifold. Here we do not enter into details of which index is more appropriate to "measure" the performance in our case. Nevertheless, we show that useful information comes from calculating some indices on contingency tables of events and nonevents related to detected and forecasted beta counts.

As usual in analyzing forecasting performance, we divide detections and forecasts into classes, build contingency tables and assess performance in a dichotomic form. By limiting ourselves to the analysis of the $t + 2$ forecasting performance, we choose 100 equidistant thresholds and divide our range in 100 classes, so obtaining 100 contingency tables. Our notations are referred to Table 13.4, 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 each threshold we calculated 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) $= [2(ad - bc)]/(gf + he)$.

Once the values of these indices for each threshold are calculated, one can plot them on graphs built with the value of the threshold as abscissa and the value of the index as ordinate, thus obtaining pictures of global performance for every range of data. This can be done for every forecasting strategy adopted in our case study. Due to the fact that the performance of the synchronous pattern approach with and without pruning are very similar, just SP-PR, TS and SD-TS are considered in these graphs in order to make them more clearly readable. Furthermore, calculations of FAR, HR and CSI reveal that the differences in performance among these three modeling strategies are nearly absent in these indices, so that the related graphs are not shown here. In Fig. 13.8 the results for BIAS, POD, EFF and HSS are shown: they imply some considerations.

The high values of the SD-TS curves for the first low thresholds in BIAS and POD plots account for an overestimation of the low values of beta counts detected. This is primarily due to the transition periods from situations driven by the thermal cycle to

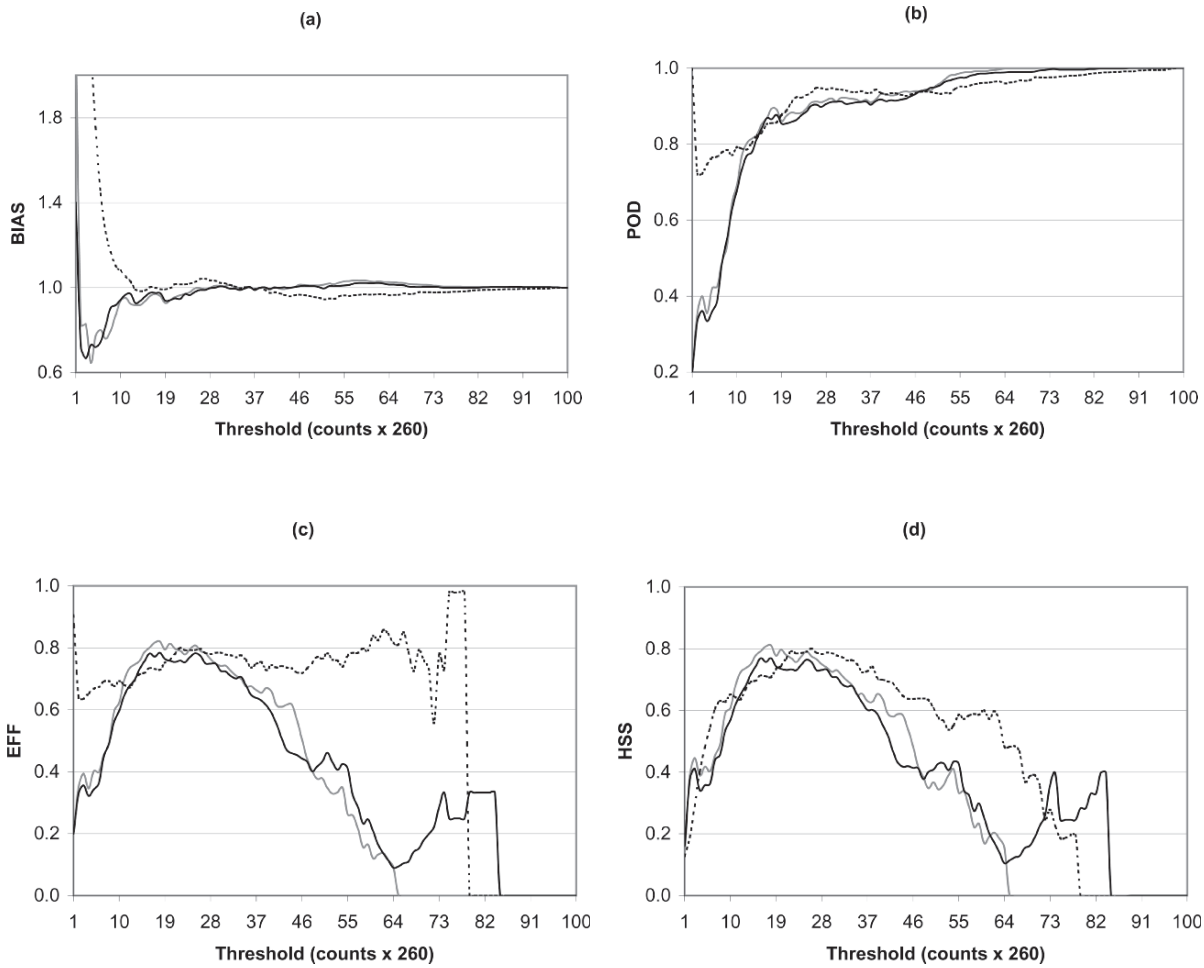

Fig. 13.8 Forecasting performance on the test set of beta counts for the black box time series approach (TS: black curves), the preprocessed time series approach (SD-TS: dashed curves) and the synchronous approach with pruning (SP-PR: gray curves). The four plates refer to BIAS (a), POD (b), EFF (c) and HSS (d)

advective situations when the SD modulation leads to very low final forecast values. Of course, this represents a negative feature of this preprocessing method and suggests adopting some different preprocessing methods during these transition periods.

The graph of EFF (Fig. 13.8c) indicates that the two time series approaches outperform SP-PR at high thresholds. The difference between SD-TS and SP-PR above abscissa 30 (in threshold units) is particularly large.

Undoubtedly, the most interesting results come from analyzing the graph related to HSS (Fig. 13.8d). First of all, HSS is a very good measure of performance (see, for instance, the discussion in Chapter 3 and in Marzban 1998). Furthermore, Fig. 13.8d shows that the time series approach reveals very good features for high thresholds; in particular, SD-TS is always better than other approaches for thresholds >23 abscissa units (about 6,000 beta counts). On the other hand, SP-PR outperforms TS until a threshold of about 46 abscissa units. Finally, we stress that all the approaches lead to a maximum of performance for a threshold around 19 abscissa units (about 5,000 beta counts): this is very important for us, because this value can be considered as the threshold that 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 lower atmosphere.

13.4.4 A Hybrid Approach and Its Forecasting Results

The results discussed above showed that the using a synchronous pattern of meteorological parameters as inputs to the neural model leads to better results than a standard (non-preprocessed) time series approach. This bears witness to the importance of having information about the initial physical state of the BL. On the other hand, using our *a priori* knowledge of day-night periodicities, and the consequent preprocessing activity *via* SD, outperforms both TS and SP strategies.

Taking these results into account, we now wish to explore the possibility that meteorological conditions could give improvements in capturing nonlinearities when periodic contributions to the BL dynamics are subtracted by SD. Therefore, once we recognize that the day-night cycle affects the radon time series as well as time series of meteorological parameters, even if sometimes at a different degree, we apply SD preprocessing to available meteorological data, as well.

Therefore, first we apply SD to the time series of beta counts and to the records of meteorological parameters. Then we concentrate on the following three different choices of inputs for the networks to be trained on the residual series:

- SD-MET: The inputs are the residuals coming from the application of SD to several meteorological variables (wind speed, pressure, temperature, dew point, sky covering and height of the lowest cloud layer, meteorological visibility) and to values of beta counts and the time derivative of beta counts at t_0.
- SD-TS + V: The inputs are the residuals coming from the application of SD to both the delayed values of the time series and the wind speed. Wind is considered very important because it characterizes the mechanical turbulence in the BL and, in particular, the situations of advection, when the periodicity in the time series (essentially due to the day-night cycle) is broken.
- SD-TS + MET: The inputs are the residuals coming from the application of SD on both the delayed values of the time series and the same meteorological variables considered also in SD-MET.

In this subsection our aim is to compare these approaches with the SD-TS approach previously discussed in order to see if the insertion of residuals of meteorological variables into the input layer leads to better results. In doing so, we compare the ability of the distinct networks in capturing the hidden dynamics on just the residual series.

Table 13.5 Forecasting performance on the residual series when SD is applied also to records of meteorological variables

Method	**R($t_0 + 2$ h)**
SD-TS	0.675 ± 0.006
SD-MET	0.571 ± 0.008
SD-TS+V	0.687 ± 0.004
SD-TS+MET	0.679 ± 0.005
Method	**R($t_0 + 4$ h)**
SD-TS	0.500 ± 0.008
SD-MET	0.297 ± 0.006
SD-TS+V	0.528 ± 0.007
SD-TS+MET	0.544 ± 0.006
Method	**R($t_0 + 6$ h)**
SD-TS	0.431 ± 0.009
SD-MET	0.151 ± 0.012
SD-TS+V	0.456 ± 0.008
SD-TS+MET	0.496 ± 0.007

Source: Pasini et al. (2003b). Copyright IEEE.

The results on the test set are presented in a concise form in Table 13.5, where the calculation of the linear correlation coefficient between targets and outputs is reported. Note that, contrary to Tables 13.2 and 13.3, here the forecasting results refer to the NN performance on only the residual series, without any *a posteriori* composition with the SD signals.

A brief analysis of these results indicates that the information on the meteorological parameters alone in the SD-MET approach (even if data on beta counts and its derivatives is included) is not able to capture a satisfying hidden dynamics or to improve forecasting results on the residual series. On the other hand, the same meteorological parameters contribute (in a statistically significant manner) to improving the results of the SD-TS approach when inserted into the input layer together with data on the time series itself. In a certain sense, one could say that the meteorological parameters act as second-order correctors to the forecast obtained by the SD-TS approach. Nevertheless, we must stress again that in a non-linear system we are not able to accurately separate the contributions of each "influence factor" to the final result.

The increase in performance obtained through application of the SD-TS + V and SD-TS + MET approaches is more evident at the longest ranges, so that we can envisage a shift of the predictability horizon for forecasting radon from observations beyond 6

h. Furthermore, we want to stress that networks fed with TS and all the meteorological parameters in the input layer are obviously larger than networks with TS and wind only. The particular structure of our sigmoids allows us to obtain better results in the very short-range with a little network and wind only, which is correlated (linearly and non-linearly) quite well with the target. When, at the following time steps, this correlation decreases, other meteorological inputs become so important to invert the situation of performance scores between SD-TS + V and SD-TS + MET.

13.5 Conclusions and Prospects

In this chapter, the role of NNs has been analyzed for modeling some features of a complex system such as the BL, whose dynamical modeling is very critical (especially in nocturnal stable situations) due to the many interactions and feedbacks that occur therein. In doing so, the BL physical dispersion properties have been summarized by means of a suitable index, the beta counts coming from the decay of short-lived radon progeny. Once this index is identified as a critical variable for describing BL behavior, several approaches have been presented for NN processing its data, with the aim of both BL diagnostic characterization and forecasting.

In this framework, interesting goals have been achieved: we stress the improved forecasting results from jointly applying our knowledge of the day-night cycle's periodicities *via* SD and the NN forecasting of the residual series. Here, in particular, NNs are able to find a hidden dynamics in what appears as a noise signal, leading to a substantial improvement in forecasting performance when compared with a black box NN application. Furthermore, the results of the hybrid approach described in the last subsection appear quite promising.

Once more, the applications described in this paper imply that, at present, the identification of key variables in a complex system and its data-driven modeling by NNs can represent a valid alternative to dynamical modeling.

Of course, although these investigations are preliminary, the scope of this paper was merely to introduce the reader to specific applications of NN modeling in another complex system, the BL, after the previous chapter dedicated to climate. This preliminary research leaves several open questions and directions for further development, considering that, for instance, only very standard NNs have been used and a very simple kind of preprocessing has been adopted. Nevertheless, in an introductory book to techniques and applications, I believe that this is not a fault, but rather a spur for the reader to further research this field.

References

Allegrini, I., Febo, A., Pasini, A., & Schiarini, S. (1994). Monitoring of the nocturnal mixed layer by means of particulate radon progeny measurement. *Journal of Geophysical Research*, *99*(D9), 18765–18777.

Guedalia, D., Ntsila, A., Druihlet, A., & Fontan, J. (1980). Monitoring of the atmospheric stability above an urban and suburban site using sodar and radon measurements. *Journal of Applied Meteorology*, *19*, 839–848.

Lopez, A., Guedalia, D., Servant, J., & Fontan, J. (1974). Advantages of the use of radioactive tracers 222 Rn and 212 Pb for the study of Aitken nuclei within the low troposphere. *Journal of Geophysical Research*, *79*, 1243–1252.

Mahrt, L. (1998). Stratified atmospheric boundary layers and breakdown of models. *Theoretical and Computational Fluid Dynamics*, *11*, 263–279.

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

Pasini, A., & Ameli, F. (2003). Radon short range forecasting through time series preprocessing and neural network modeling. *Geophysical Research Letters*, *30*(7), 1386. Doi:10.1029/2002GL016726.

Pasini, A., Pelino, V., & Potestà, S. (2001). A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *Journal of Geophysical Research*, *106*(D14), 14951–14959.

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

Pasini, A., Ameli, F., & Lorè, M. (2003a). Mixing height short range forecasting through neural network modeling applied to radon and meteorological data. *Proceedings of the 3rd Conference on Artificial Intelligence Applications to Environmental Sciences, 83rd Annual Meeting of the American Meteorological Society*. AMS: Long Beach, CA, CD ROM.

Pasini, A., Ameli, F., & Lorè, M. (2003b). Short range forecast of atmospheric radon concentration and stable layer depth by neural network modelling. *Proceedings of the IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications* (pp. 85–90). Lugano, Switzerland: IEEE.

Pasini, A., Perrino, C., & Žujić, A. (2003c). Non-linear atmospheric stability indices by neural-network modelling. *Nuovo Cimento*, *26C*, 633–638.

Perrino, C., Pietrodangelo, A., & Febo, A. (2001). An atmospheric stability index based on radon progeny measurements for the evaluation of primary urban pollution. *Atmospheric Environment*, *35*, 5235–5244.

Porstendörfer, J. (1994). Properties and behaviour of radon and thoron and their decay products in the air. *Journal of Aerosol Science*, *25*, 219–263.

Seinfeld, J. H., & Pandis, S. N. (1998). *Atmospheric chemistry and physics: From air pollution to climate change* (1326 pp.). New York: Wiley.

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

Vinod Kumar, A., Sitaraman, V., Oza, R. B., & Krishnamoorthy, T. (1999). Application of a numerical model for the planetary boundary layer to the vertical distribution of radon and its daughter products. *Atmospheric Environment*, *33*, 4717–4726.

Weigend, A. S., & Gershenfeld, N. A. (Eds.) (1993). *Time series prediction: Forecasting the future and understanding the past* (664 pp.). New York: Addison-Wesley.