

Chapter

A Synthesis of the Information Given by Temporal Data Series: The Representative Day

Tiziano Tirabassi and Daniela Buske

Abstract

The recording of air pollution concentration values involves the measurement of a large volume of data. Generally, automatic selectors and explicators are provided by statistics. The use of the Representative Day allows the compilation of large amounts of data in a compact format that will supply meaningful information on the whole data set. The Representative Day (RD) is a real day that best represents (in the meaning of the least squares technique) the set of daily trends of the considered time series. The Least Representative Day (LRD), on the contrary, it is a real day that worst represents (in the meaning of the least squares technique) the set of daily trends of the same time series. The identification of RD and LRD can prove to be a very important tool for identifying both anomalous and standard behaviors of pollutants within the selected period and establishing measures of prevention, limitation and control. Two application examples, in two different areas, are presented related to meteorological and SO_2 and O_3 concentration data sets.

Keywords: air pollution, daily trends, data set, temporal series, air pollution management, representative day

1. Introduction

In recent years, environmental management and a suitable development have assumed great importance [1, 2]. Air quality management and protection presuppose knowledge of the state of the environment. Such knowledge involves a properly cognitive and interpretative ability.

Local or regional air pollution control is usually achieved through air quality monitoring networks. These networks are a useful tool for the protection of human health and the environment, and allow both to evaluate the benefit of remediation actions and to prepare specific interventions in case of exceeding the threshold levels considered dangerous. For economic and managerial reasons, the number of measuring points in a network is limited and, especially if their arrangement has not been carefully studied, the detection units risk being unrepresentative of the entire territory that is to be monitored. In this regard, the mathematical models that simulate the transport and diffusion of pollutants in the atmosphere constitute a valid integration to the measurements, allowing to have estimates of concentrations over the entire territory for which it is interesting to know the evolution of concentrations. Once the good quality of the answers provided by a model has been ascertained, it allows us to trace the contribution of the different sources to the

overall pollution, and therefore to correctly address any actions to limit emissions. Furthermore, only with the models is it possible to make forecasts or simulate concentration scenarios in connection with emission limitation policies as part of the preparation of recovery plans. The analyzers network, together with the inventory of emissions sources, is of fundamental importance for the construction of the cognitive framework, but not the interpretive one. In reality, air quality control requires an instrument interpretative capable of extrapolating in space and time the values measured in the of the analyzers, while the improvement of the air quality can be obtained only with plans that reduce emissions and then with instruments (such as the air pollution mathematical model) capable of linking the cause (source) of pollution with the effect (the concentration of the pollutant) [3].

The introduction of mathematical modeling produces a qualitative leap in the management of atmospheric pollution compared to that possible through measurements alone, because the models allow functions that are not accessible to the latter [4].

Mathematical models are capable of:

- describe and interpret the experimental data;
- control in real time and/or analyze air quality;
- manage accidental releases and assess risk areas;
- identify pollution sources;
- evaluate the contribution of a single source to the pollution load;
- manage and plan the territory.

For the above consideration, they turn out to be a technical instrument indispensable for environmental management. In fact, it is of considerable importance the description of the processes that govern the transport and diffusion of pollutants. They are generally represented by meteorological preprocessors able to describe the transport operated by the wind and the variables useful for the different models to calculate the diffusion of pollutants [3]. They are extremely useful when local phenomena such as land-sea and/or upstream-downstream breezes have to be described.

While the complete dataset of the measured data is useful (also from a legal point of view) for an environmental control and the data can also be considered as alarm signals for particularly dangerous situations, on the contrary, for the use of the models it is generally necessary to limit the number of the data to simulate. To this regard it is useful to have techniques that allow to filter the data or identify subsets. Generally, in order to summarize information automatic selectors and explicators are used, many of which are provided by statistics as probability density function, mean standard deviation, median, quantiles. Moreover, time-series trends, spectral analysis, principal component analysis and cluster analysis are usually used. In support of good decision-making, the use of statistic has become widespread in air pollution assessment. In addition, collected data are used to define specific typical periods that may be of particular interest in a study of pollutant diffusion. We can also mention, for example, a typical working day, a typical holiday, a typical seasonal day, etc. The purpose of a such typifying is that of outlining characteristic scenarios for a given period under investigation. Afterwards, mathematical models make it possible to attempt simulation of a typical period trend.

It is very useful to identify from the set of data periods that can represent the peculiarity of the area under control and at the same time extraordinary events, in

particular if they represent critical situations from the point of view of environmental pollution, so as to be able to study which meteorological situations and emissions conditions that cause them.

To meet these needs, it is proposed to use a methodology capable of identifying the most representative day of a series of daily data so that the simulation of that day allows to understand the diffusion and pollution situation typical of the study area. At the same time, the methodology should also identify the most anomalous day or the most anomalous days that correspond to situations of major pollution on the ground. Naturally, the methodology must identify real days, not fictitious as the typical day [5] constructed with time series of data composed of the average values of the concentration averaged (at the same time) over the whole dataset.

Identifying real days is also important because only in this way can be identified the meteorological and emission data that characterized that day and the measured ground level concentrations. This methodology will be presented below.

2. The representative day methodology

What we want to select in an annual, seasonal or monthly dataset of daily time series is the one that best represents the set of those stored. This can be achieved with the RD technique.

What we call RD technique is a daily data set, actually recorded at a field station, which is characterized by the minimal differences with respect to all the daily measurements series of that station's temporal series: that is, the daily series whose sum of the squared differences over one day turns out to be the smallest compared to all the other days of the period under consideration [6].

That is, if A is the matrix:

$$A_{ij} = \sum_{k=1}^{24} (c_{ki} - c_{kj})^2, \quad i, j = 1, 2, \dots, N \quad (1)$$

where N is the number of days in the time period for which the representative day is calculated, and c_{ki} is the pollutant concentration of the i -th day at the k -th time period.

We adopt A_i to indicate the sum of all the squared residuals of the i -th line (or column, A_{ij} being a symmetrical matrix with all zeros in the diagonal):

$$A_i = \left(\sum_{j=1}^N A_{ij} \right) \quad (2)$$

The RD is the one with the lowest sum, i.e. the i -th day where A_i is the smallest of the quantities obtained. The purpose of such typifying is that of outlining characteristic scenarios for a given period under investigation. That is, identify the day of the data series that is closest (in the meaning of the least squares technique) to all the time series included in the data set under examination.

2.1 The least representative day

The shown above approach also allows the identification of the "least representative day" (LRD), i.e. the daily series that maximizes the mean sum of squared residuals. The LRD identifies an anomalous situation of pollutant dispersion.

As LRD is a real day, it allows us to identify the meteorological characteristics and air pollution emissions related to that day and, therefore, giving us the possibility to study the phenomena and conditions that contributed to the realization of that air pollution diffusion and that distribution situation of air pollution concentrations on the ground.

Of course, by eliminating the data of that day from the original series and repeating the procedure on the remaining data, it is possible to highlight the second less representative day. Proceeding in the same way, the third, fourth, etc., LRD can be highlighted.

2.2 Results normalization

To compare the degree of representativity of the most or least representative days with that obtained for other time periods and/or at other measurement stations or stations in different areas, a normalization is required in order to make the day independent of the length of the measurement series, sampling period and characteristics of the area under study.

The “representativity” of an RD can be quantified by introducing the adimensional index DI:

$$DI = \frac{\sum_{i=1}^N \sum_{k=1}^{24} (\Gamma_k - c_{ik})^2}{\sum_{i=1}^N \sum_{k=1}^{24} (\bar{c}_k - c_{ik})^2} \quad (3)$$

where \bar{c}_k is the time mean concentration of the RD at the $k - th$ data, and Γ_k is the time mean concentration at the $k - th$ data calculated over the period under consideration, i.e.:

$$\bar{c}_k = \frac{1}{N} \sum_{i=1}^N c_{ik}, \quad k = 1, 2, \dots, 24 \quad (4)$$

N is the number of days making up the time interval and c_{ik} is the time mean concentration of the pollutant of the $i - th$ day at the $k - th$ data of the daily sequence.

DI is an adimensional quantity greater than or equal to 1, which is closer to unit the more the RD is representative of the period under consideration.

The LRD can also be normalized in the same way: one simply substitutes in Eq. (3) the time mean concentration of the RD (Γ_k) with the least representative ones. In this case, the value of DI will always be greater than one, providing an indication of the low degree of representativity of the day obtained; the more DI is greater than 1, the more the LRD is “anomalous”, compared to the trend of RD.

The normalization procedure described above it is independent of the size of the measured concentrations and of the number of days (N) included in the time period considered.

2.3 A fictional day: the typical day

We introduce the typical day (TD), that could be defined as a “fictional” day, whose concentrations (the time concentrations that form the daily sequence) are given by the concentration means, calculated, time by time over all the days of the period of study. The daily sequences can belong to periods of a month, a season, a year, or grouping of particular days that share the features one wishes to study. This form of data representation is widely used in Italy [5].

The TD, with the notation already introduced, can be mathematically expressed by:

$$\bar{C}_k = \frac{1}{N} \sum_{i=1}^N c_{ik}, \quad k = 1, 2, \dots, 24 \quad (5)$$

However, since the TD is not a real day, this form of evaluation provides a presentation which cannot take account of the variations characterizing the actual behavior of the quantity under examination. Furthermore, since it is not a real day, it cannot be associated with any meteorological or emission parameters and therefore is of little interest for a possible air pollution diffusion model applied in the area.

The TD can be considered as an extreme case: for an infinite series of data, the RD tends toward the TD. Therefore, the TD can be considered an asymptotic limit for the RD.

3. Application examples

We applied the method to SO_2 concentrations measurements recorded in the Ravenna area (Italy). The town of Ravenna is 10 Km from the sea, while his industrial area is situated between the sea and the town. The climate is basically continental but more temperate by the proximity of the coast. The entire area is subject to a series of weak local wind circulation, frequent inversion phenomena and high relative humidity [7, 8]. The methodology was applied to a time series of hourly SO_2 concentration data measured by a station of the automatic monitoring network located in the city of Ravenna. In **Figure 1** it is shown the RD and LRD of hourly SO_2 record. There is a big difference between the RD and LRD from 10 am onwards.

Such behavior can be explained by the fact that SO_2 is mainly emitted from point sources located in the industrial area (located between the sea and the town) and therefore the measurements are influenced by wind direction. The LRD correspond to the SO_2 concentration recorded on the January 3 where, as it is shown in **Figure 2**, at 10 am wind change direction with a corresponding doubling in the speed.

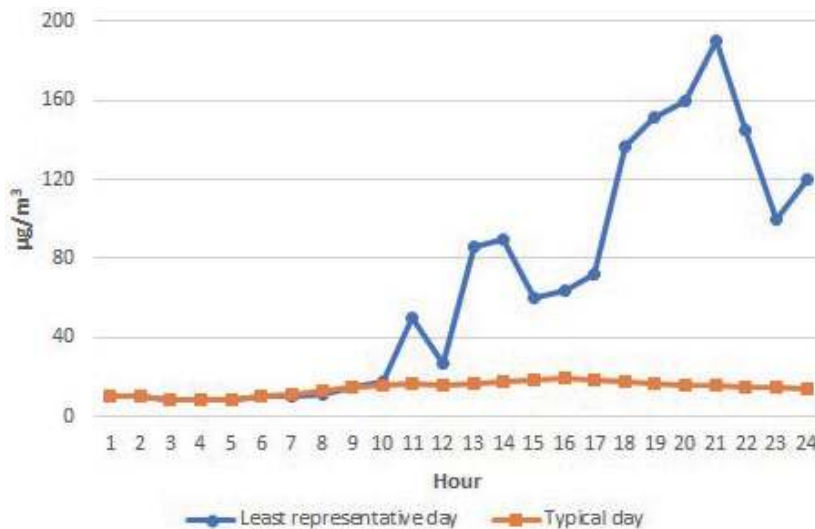


Figure 1. Representative day (RD) and least representative day (LRD) of SO_2 .

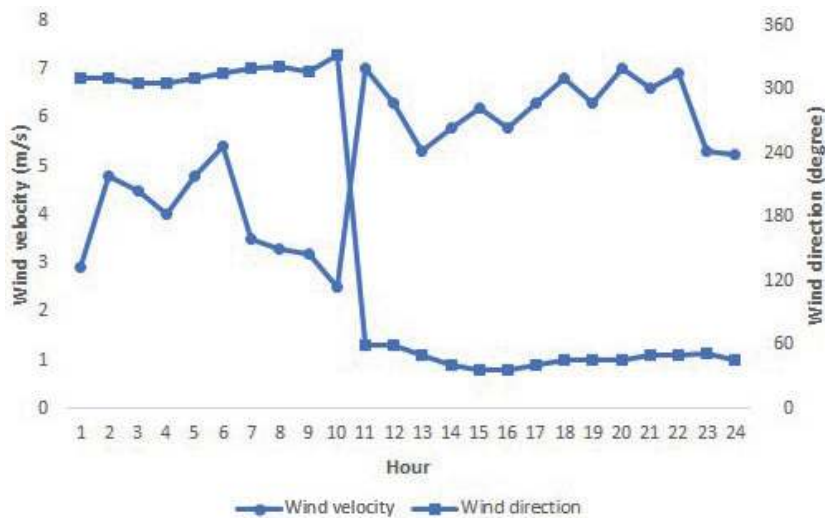


Figure 2.
Wind speed and direction during the day to which LRD of **Figure 1** refers.

Another interesting example of application of the methodology is that of concentration time series detected in the Falconara industrial area, where we have a high ozone production due to the characteristics of the local conditions and the emission sources [9]. In addition, there is an automatic monitoring network in the area which has collected a large amount of data.

Falconara is an urban center located north of Ancona on the Adriatic coastline of Italy. The Falconara area faces air pollution problems, mainly during summertime. Preliminary studies [10] showed that the most important factor contributing to urban air pollution in Falconara and its surroundings is the amount of emissions from mobile sources and industries.

The Falconara area can be roughly divided into two parts: a coastal area and an inland area. The coastal area is characterised by the presence of a large oil-refinery. The inland area comprises the main urban area surrounded by hills. The description of the microclimate and landform of this area can be found in [9].

The accumulation of photochemically produced Ozone depends strongly on the prevailing meteorological conditions. In fact, meteorological conditions observed on days with high Ozone mixing ratios are often quite different from those when Ozone concentrations are low [11, 12].

In the first example we present, the Ozone shows the same trend on the RD and LRD, although with lower concentration values in the second part of the day.

Regarding **Figure 3**, the explanation can be found by analyzing the solar energy data for the day corresponding the LRD, June 21st., **Figure 4** shows lower solar radiation values in the middle of the day. The explanation for this case is very obvious because of the direct correlation between Ozone and solar radiation [13].

Another example, where on the contrary the Ozone shows the same trend in the RD and LRD, although with higher concentration values in the second part of the day (see example **Figure 5**). In this case the main wind direction during the last afternoon and the evening is North-East, indicating that the wind came from the sea (see **Figure 6**). Land breeze/sea breeze phenomenon seems to predominate in the Italian Adriatic area, so that the air pollution produced over the urban-industrial coast is transported as plumes over the sea and, subsequently, due to sea breezes, transported back to the coastal areas. The evening sea breezes transport the masses

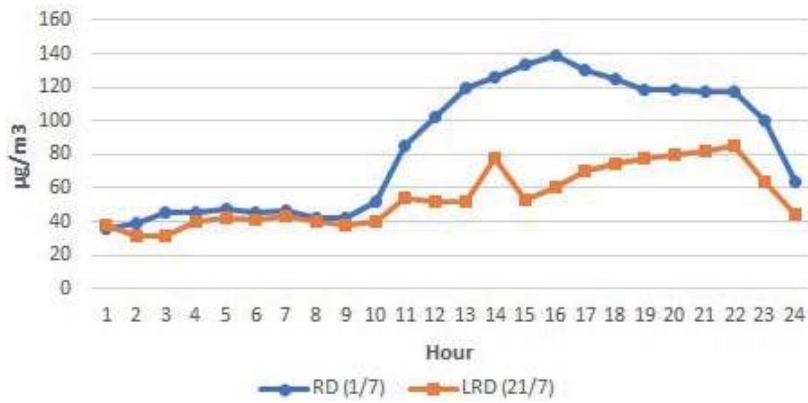


Figure 3.
 Representative day (RD), least representative day (LRD) of O₃.

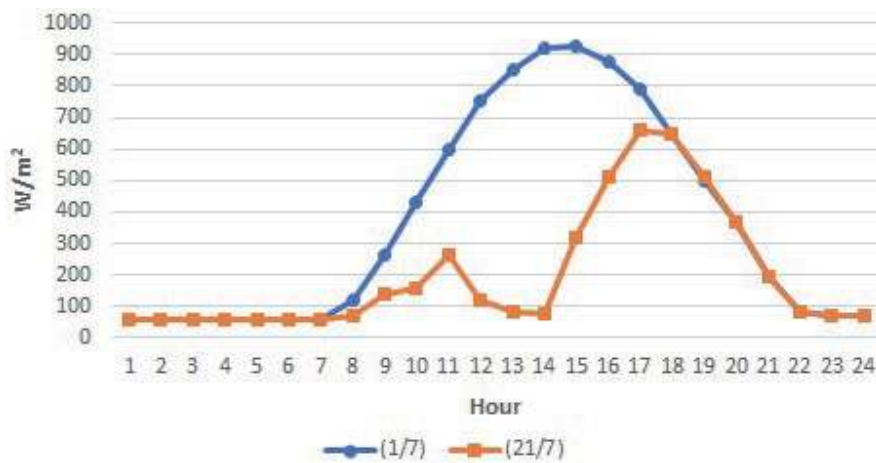


Figure 4.
 Solar radiation measured during the RD and LRD of Figure 3.

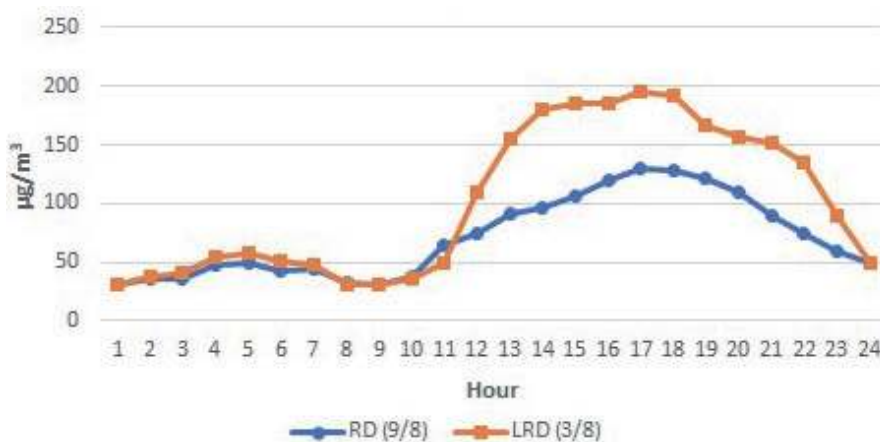


Figure 5.
 Representative day (RD), least representative day (LRD) of O₃.

of air offshore, where the deposition rate is so slow that the ozone accumulates and is transported back to the coast when the daytime sea breezes resume, thus setting in motion mechanisms of photo-oxidant re-circulation [8, 9, 14].

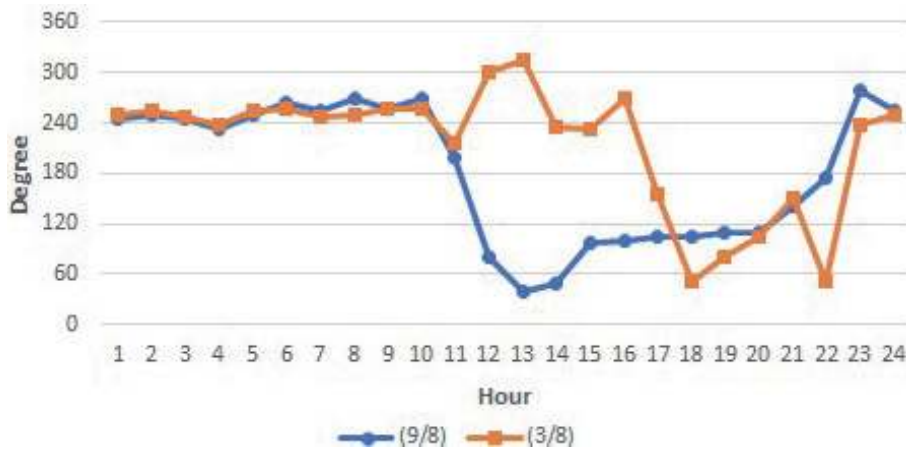


Figure 6. Wind direction measured during RD and LRD of Figure 5.

DI / gas	SO ₂	O ₃ - Figure 3	O ₃ - Figure 5
RD	1.01	1.07	1.07
LRD	6.34	2.17	2.26

Table 1. Representativity adimensional index DI of RD and LRD.

Table 1 provides the representative adimensional index DI of RD and LRD. It is possible to see the different behavior of SO₂ and O₃. In the case of O₃ the indices show that the RD are less representative and the LRD is less anomalous.

An explanation of this different behavior can be explained by the fact that SO₂ is emitted mainly from point industrial sources, so that at the beginning very concentrated plumes are composed. Therefore, the resulting concentrations on the ground are more subject to weather conditions and wind direction. On the contrary, ozone is a secondary gas (it is not directly emitted into the atmosphere) and therefore forms less concentrated plumes and offers greater inertia (compared to SO₂) to the production of different concentration scenarios on the ground.

4. Conclusions

Automatic monitoring networks are often used for study, control and management of local environmental problems. As a result, over time, a large mass of data is collected. While the individual data are very useful for real-time control and to report any alarms, it is necessary, for the study of the territory, for the understanding of the phenomena present in the area, to obtain a synthetic set of the measured data. Moreover, the processes that govern the transport and diffusion of pollutants are numerous and of such complexity that it is not possible to describe them without using mathematical models. Both the interpretation of the phenomena governing pollutant diffusion and the use of mathematical model requires a synthesis of the information given by temporal data series.

For this purpose, the most representative day constitutes a simple and immediate method through which to characterize the temporal structure of daily trends. We have called the “representative day” the day which, in a set of data composed of daily series, best represents the whole set of series. In mathematical terms, the one

which minimizes the sum of squared differences with respect to all the daily trends in a temporal series. The attention is focused on the “day as a unit”, without losing, however, the particular temporal structure, as partly occurs in the case of the “typical day”.

Moreover, RD being an actual day, it allows the identification of the date on which it occurred and, thus, a knowledge of the meteorological and emission parameters which characterized it. Among other things, this allows it to be simulated with air pollution diffusion models.

The same approach also allows the identification of the “least representative day”, that is, the day on which an anomalous, nearly always critical situation occurred, compared to the average trend recorded for that period. LRD is also an actual day. The study of the meteorological and emission parameters relating to this day, will allow a preliminary interpretation of the phenomena which brought about the situation.

By eliminating the data of that day from the original series and repeating the procedure on the remaining data, it is possible to highlight the second less representative day. Proceeding in the same way, the third, fourth, etc. LRD can be highlighted.

Both RD and LRD can be normalized so that the degree of representativity can be compared independently of the length of the measurement series, sampling period and characteristics of the area under study.

Author details

Tiziano Tirabassi*[†] and Daniela Buske^{2†}


1 Federal University of Santa Maria, Santa Maria (RS), Brazil

2 Federal University of Pelotas, Pelotas (RS), Brazil

*Address all correspondence to: tiziano.tirabassi@gmail.com

[†] These authors contributed equally.

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Khan S A R, Zhang Y, Kumar A, Zavadskas E, Streimikiene D. Measuring the impact of renewable energy, public health expenditure, logistics, and environmental performance on sustainable economic growth. *Sustainable Development*. 2020; 28: 833–843.
- [2] Khan S A R, Sharif A, Golpîra H, Kumar A. A green ideology in Asian emerging economies: From environmental policy and sustainable development. *Sustainable Development*. 2019; 27: 1063–1075.
- [3] Tirabassi T. Operational advanced air pollution modeling. *Pageoph*. 2003; 160: 5–16.
- [4] Seinfeld J H, Pandis S N. *Atmospheric chemistry and physics*. New York: Wiley; 2016. 1152 p.
- [5] Tirabassi T, Rizza U. Flux parameterization for atmospheric-diffusion models in an industrialized area. *Il Nuovo Cimento*. 1993; 16C: 187–192.
- [6] Tirabassi T, Nasseti S. The representative day. *Atmospheric Environment*. 1999; 33:2427–2434.
- [7] Tirabassi T, Fortezza F, Vandini W. Wind circulation and air pollutant concentrations in the coastal city of Ravenna. *Energy and Buildings*. 1991; 1 (2): 699–704.
- [8] Fortezza F, Strocchi V, Giovanelli G, Bonasoni P, Georgiadis T. Transport of photochemical oxidants along the northwestern Adriatic coast. *Atmospheric Environment*. 1993; 27A: 2393–2402.
- [9] Latini G, Cocci Grifoni, R, Passerini G. Influence of meteorological parameters on urban and suburban air pollution. In: Brebbia C A, Latini G, editors. *Air Pollution IX*. Southampton (GB): WIT Press; 2002. p. 753–762.
- [10] Latini G, Cocci Grifoni, R, Passerini G, Tirabassi T. Evaluation of Mixing Height over Complex Coastal Terrain. In: Brebbia C A, Jacobson M, Power H, editors. *Air Pollution 99*. Southampton and Boston: WIT Press; 1999. p. 1073–1081.
- [11] Bower J, Broughton G, Stedman J, Williams M. A winter NO_2 episode in the UK. *Atmospheric Environment*. 1994; 28: 461–475.
- [12] Ziomas I, Melas D, Zerefos C, Bais A, Paliatsos A. Forecasting peak pollutant concentrations from meteorological variables. *Atmospheric Environment*. 1995; 29 (24): 1352–2310.
- [13] Colbeck, I, Mackenzie, A R. *Air Pollution by Photochemical Oxidants*. Amsterdam: Elsevier; 1994. 232 p.
- [14] Latini G, Cocci Grifoni, R, Passerini G, Fava G. Applicability of a Photochemical Box Model over Complex Coastal Areas. In: Brebbia C A, Jacobson M, Power H, editors. *Air Pollution 99*. Southampton and Boston: WIT Press; 1999. p. 697–706.