

Simplified reading of one-year air pollution: ranking of chemical and physical variables

M. C. Mura

*Department of Environment and Primary Prevention,
Istituto Superiore di Sanità, Rome, Italy*

Abstract

This study offers a simple tool to simplify reading of one-calendar-year pollution. Average-concentration and monthly-persistence distributions of chemical pollutants in the troposphere and of meteorological variables are ranked, analysed by means of a non-parametric technique and depicted in a non-parametric space. Sign test is used for ensembles of paired monthly-data related to nitrogen total oxides and ozone, temperature and pressure, continually detected by a survey station of the Network for air-quality control in Rome. Ranks describe atmospheric pollution in one calendar year; pollution is represented in one single graph, where mutual behaviours of variables can be analysed by application of the relational model depending on results of testing paired variables.

As a consequence of this non-parametric combination (ranking chemical and physical atmospheric variables + test) an easy graphic model of pollutant behaviour in one-year climatic oscillations is available.

The current dataset dates back to 1994 since each series was almost complete and meteorological variables were available; moreover, the survey station remained fixed in one position all year long.

Keywords: metropolitan area, Rome, hourly concentration of pollutants, behavioural model.

1 Introduction

Analysis and representation of pollution phenomena in non-parametric spaces is not usual in management of chemical and physical air data; even so, non-parametric tests prove essential for small samples; these tests are common in



contexts such as psychology [1] and economy and generally in time series analysis [2, 3]. They are well-documented in scientific literature and their powerfulness and efficiency are measured with respect to corresponding parametric tests.

Here a methodological attempt to treat air datasets non-parametrically is accomplished, in order to identify possible significant presences between trends of pollutant concentrations and meteorological variables in one calendar year and, at the same time, to represent such trends in one single graph. This pictorial representation is convenient since graphs play a role of powerful research instruments, from empirical analysis of trends up to choice of forecasting methods [4].

The transformation procedure of quantitative variables into qualitative and, then, the application of the test will be described. Tested samples are small and they are the result of an elaboration of continuous one-calendar-year hourly data acquired by a survey station of the Network for air-quality control in Rome, placed in an area which is characterized by a medium/high density of vehicular traffic, the main source of urban pollution.

2 Materials and methods

2.1 Databases

The database consists of hourly-concentration variables with respect to a) nitrogen oxides and ozone ([concentration], ppb) and meteorological hourly parameters such as b) temperature (T, °C) and pressure (p, hPa). Each variable describes a time series of hourly values. Chemical monitoring data were collected through an automated integration process in one-hour interval (mediation time). The reliability of the chosen dataset has been validated by the Network Administrator; hence, data have been accepted and accounted as believable [5], in step with current epistemological literature [6].

NO_x represents the total amount of NO, NO₂ and other nitrogen oxides.

High-temperature combustion processes, like those occurring in piston-engines, are the main NO_x anthropic-source. NO_x play a key-role in O₃ formation. Both of them are reactive pollutants.

2.2 Preprocessing

First Student's t-test [7] was applied to check normality of each distribution relating to the corresponding time series of chemical-pollutants hourly-concentrations; all distributions proved normal at 95%. The initial size of each pollutant distribution turned out not to be smaller than 85% of the expected yearly one on average.

Anyway, corresponding time series showed gaps caused by bad working of analysers; in order to complete them, data-filling was needed. To perform it, suitable working matrices, transposed with respect to the original sample, were used. Hourly-concentration values of all pollutants in one month (720X1) were analysed within 24 h x 30 days matrices to which interpolation and forecasting

techniques [8] were opportunely applied. This way more than 10% of missing chemical data in pollutant distributions were recovered. The size of meteorological-variable distributions was about 95% of theoretical one, with no need of reconstruction.

2.3 Statistical description of monthly series of hourly values

Each time series of hourly-concentration values and meteorological variables with respect to each month in one year was described statistically [9] by means of its arithmetic mean, median, confidence interval for the mean, quartiles, standard deviation and variance (s^2). Since time series were under examination, monthly means and variances for every variable were chosen as working parameters. As known, variance describes statistical dispersion (spreading) around the expected value and allows the experimenter to know how one variable depends on another (autoregression) [10]. By means of variance the persistence or stationarity of the series can be determined; in this context this parameter is used to monitor the one-year trend of each monthly variable since it is a measure of how probable is a process in a particular time interval (one month in this case), if the same is observed in the previous one.

Monthly values of each mean and persistence distribution were normalized with respect to their one-year averages. Results are shown in Table 1.

Table 1: Parametric monthly statistics of chemical and physical parameters continually monitored by a survey station of the Network for air-quality control in Rome

	NO _x	NO _x	NO ₂	NO ₂	O ₃	O ₃	T	T	p	p
units	ppb	ppb	ppb	ppb	ppb	ppb	°C	°C	hPa	hPa
	m	s ²	m	s ²	m	s ²	m	s ²	m	s ²
Jan	1,14	19,94	0,75	0,84	0,65	0,07	0,56	0,66	1,0002	2,06
Feb	0,98	11,40	0,89	0,77	0,28	0,09	0,52	0,99	0,9993	1,27
Mar	1,02	9,59	1,06	1,29	0,69	0,53	0,79	0,81	1,0044	0,71
Apr	0,80	4,86	0,95	1,04	0,77	0,28	0,78	1,06	0,9949	1,48
May	0,94	3,62	1,10	0,61	1,26	1,06	1,16	1,28	0,9971	0,45
Jun	0,75	4,41	1,12	1,07	1,48	1,75	1,34	1,30	0,9992	0,27
Jul	0,63	3,56	1,08	1,21	1,79	3,17	1,68	1,06	0,9961	0,10
Aug	0,53	2,43	0,96	0,98	2,02	3,29	1,68	1,08	0,9969	0,31
Sep	0,76	5,03	0,96	1,05	1,38	1,11	1,38	1,05	0,9988	0,41
Oct	1,04	13,25	0,94	1,18	0,93	0,46	1,03	1,08	1,0005	1,39
Nov	1,82	22,54	1,09	1,06	0,74	0,08	0,51	0,85	1,0062	1,51
Dec	1,60	19,36	1,12	0,90	0,46	0,10	0,57	0,78	1,1341	2,05

1-yr values 118,8 10532,2 41,3 333,3 9,9 92,0 16,6 1,00 1032,2 1,00

m = mean ; *s*² = variance

Monthly values are normalized with respect to one-year averages



2.4 Procedure for transforming and constructing the sample

Each quantitative value in each distribution, of means and variances in sequence, is assigned a rank [1] 1 to 12 (since months are 12 in one year), according to its magnitude. Assignment can proceed either in increasing or decreasing order. In case of *ex-aequo* values, first value is assigned its suitable rank, second one the following rank and so on; all assigned ranks are added and the sum is divided by the number of equivalent cases which were found, so getting a final rank to be assigned to the whole of *ex-aequo* values. Ranks of each monthly variable over one calendar year are shown in Table 2 (means and variances in sequence).

Table 2: Non-parametric statistics of chemical and physical variables.

	NO _x		NO ₂		O ₃		T		p	p
	r _m	r _{s2}	r _m	r _{s2}	r _m	r _{s2}	r _m	r _{s2}	r _m	r _{s2}
Jan	11	11	1	3	3	1	3	1	8	12
Feb	10	8	2	2	1	3	2	5	7	7
Mar	7	7	7	12	4	7	6	3	10	6
Apr	8	5	4	6	6	5	5	7,5	1	9
May	5	3	10	1	8	8	8	11	4	5
Jun	6	4	12	9	10	10	9	12	6	2
Jul	3	2	8	11	11	11	11	7,5	2	1
Aug	2	1	5,5	5	12	12	12	9,5	3	3
Sep	1	6	5,5	7	9	9	10	6	5	4
Oct	4	9	3	10	7	6	7	9,5	9	8
Nov	9	12	9	8	5	2	1	4	11	10
Dec	12	10	12	4	2	4	4	2	12	11

r_m = rank for mean ; r_{s2} = rank for variance

In order to identify significant differences between variable distributions, the latter have to be paired and members in each pair must be mutually classified, taking account of sign-differences. Sign distributions (either > or <) are obtained; possible *ex-aequo* distributions are removed. This way binomial distribution suitably describes probability of a plus (>) or a minus (<) occurrence. Sum over all signs is a measure of sample dimension, while sum over less frequent sign is used as a row-value in the probability-table.

2.5 Sign test

In sign test [1] the cumulative function of statistical distribution is independent of observation one; if observations in two ensembles are distributed according to the same cumulative function (not specified), null hypothesis (H_0) proves true



[11]. Sign-test powerfulness and efficiency are 95% for small samples, but they decrease at increasing of sample dimension, until they reach the asymptotic value of 63% for large ones [12].

In statistical hypothesis testing, the p-value (significance) is the probability of obtaining a result at least as extreme as that obtained, assuming the truth of the null hypothesis that the finding was the result of chance alone.

Null hypothesis (H_0) corresponds to:

$$p(X_A > X_B) = p(X_A < X_B),$$

i.e. probability p of having distribution X_A larger than X_B equals that concerning the inverse case.

The significance level (α) is the probability that the null hypothesis will be rejected in error when it is true; generally, one rejects the null hypothesis if the p-value is smaller than or equal to the significance level. Here $\alpha = 0.05$ (significance at 5%); when $\alpha > p$, H_0 is rejected and the finding is said to be 'significant at the 5% level'. In other words, rejecting H_0 is equivalent to say a significant difference exists in the behaviour of the two distributions.

2.6 Sampling

From one-year distributions, both for ranks of mean values and of monthly-persistence of pollutant-concentrations and meteorological values, the corresponding distributions relating to periods:

- spring–summer (April–September); sample dimension = 6
- fall–winter (October–March); sample dimension = 6
- 'school-time' (October–June); sample dimension = 9

were extracted.

The first two periods were chosen since the maximum (minimum) of one of meteorological variables describing climatic oscillations in the Mediterranean area falls in one of them and the minimum (maximum) in the other. The third period was chosen as a marker of anthropic activity in the urban area, decreasing July to August since didactics is interrupted and people are on vacation. In each period all paired variables were checked under the hypothesis H_0 was true.

Table 3 holds mean and persistence pairs of those variables whose distributions proved significant under test application, together with associated probabilities.

Fig. 1 shows a graphic model of parameter interrelation, which was carried out taking account of sign-differences among distribution-pairs, i.e. rank distributions.

One-year trends of monthly-means of NO_x , NO_2 , O_3 , T and p are drawn in Fig. 2.



Table 3: Results of applying sign test to pollutant and meteorological variable pairs collected in one calendar year in Roman urban area.

Variables	spring–summer period		fall–winter period	
	April–September		October–March	
	p_m	p_s^2	p_m	p_s^2
$\text{NO}_x \leftrightarrow \text{O}_3$	0,016	0,031	0,016	0,031
$\text{NO}_x \leftrightarrow \text{NO}_2$	0,031	0,109	0,031	0,344
$\text{NO}_x \leftrightarrow \text{T}$	0,016	0,031	0,016	0,109
$\text{NO}_x \leftrightarrow \text{p}$	0,181	0,656	0,031	0,344
$\text{NO}_2 \leftrightarrow \text{p}$	0,016	0,344	0,016	0,344
$\text{O}_3 \leftrightarrow \text{p}$	0,016	0,109	0,016	*
$\text{p} \leftrightarrow \text{T}$	0,016	0,109	0,016	0,656

Variables: pairs showing significant differences in r_m and/or r_s trends are reproduced.

p_m = p-value for mean, p_s^2 = p-value for variance.

Significant differences are written in bold.

Probabilities are calculated according to tables available in [1].

*-symbol stands for 'no value', when sample dimension is smaller than 5.

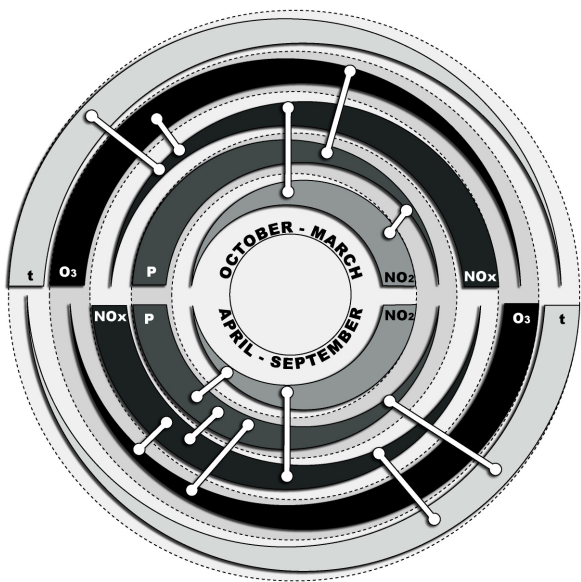


Figure 1: Behavioural model of reactive pollutants in the climatic oscillations.



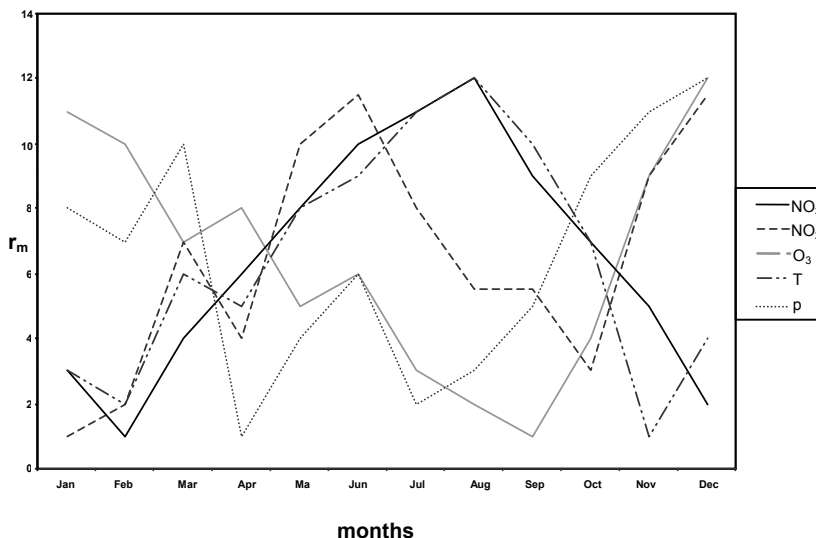


Figure 2: Trends of pollutants and meteorological variables in one calendar year.

3 Results and discussion

Statistical hypothesis testing tells apart two different situations: either the absence of significant differences or a marked dominance of one sign on the other; this means a different behaviour among pollutants and between pollutants and the other physical parameters.

Looking at Table 3, it is evident how physical-parameter behaviours (temperature and pressure) change significantly during climatic oscillations.

Temperature and pressure control processes relating to photo-chemical pollutants such as O₃ and NO_x both in spring-summer period and in fall-winter one. In both periods O₃ and NO_x are significantly and inversely connected: in the former interval NO_x decreases at increasing of temperature and ozone-concentration [13–16], while it increases at decreasing of the same chemical-physical parameters in the latter one; in both periods the corresponding trends prove not mutually stationary.

Pressure controls NO₂ and O₃ behaviours in both terms; no significant difference exists between respective stationarities, but just between trends of mean-values. Note the relation between NO_x and p is only significant in fall-winter period ($p = 0.031 < \alpha = 0.05$), not since the pollutant increases at decreasing of pressure, but because of oscillations in trends of NO₂, which show the same stationarity tendency ($p = 0.344 > \alpha = 0.05$). It is reasonable to assert climatic oscillations control reactive processes, which prove specular in the two cycles. Trends mutually cancel out in these two periods, as it was pointed out by test results, which found no significant differences between 12-month trend pairs.



All pollutants show the same expected behaviour: no significant differences, not only within one calendar year, but also during 'school-time'. This suggested the three summer months, removed from analysis because of reduced traffic, are not able to generate a different behaviour with respect to the entire year.

NO_x-trend (Fig. 2) increases September up to November, when absolute maximum is reached, and decreases June to August (absolute minimum) [17]. NO₂ doesn't follow oxide-trend, but rather pressure, even if such a behaviour is reversed in June-July. In fact, from this period up to December, NO₂ mean-concentration increases more rapidly than pressure, while January to June the reverse happens. O₃ follows temperature curve.

4 Conclusions

This procedure, quickly automatable, allows the researcher to compare pollution events in different temporal situations at different latitudes. Moreover, according to data-management finalities, it can be extended to other chemical variables such as benzene and polycyclic aromatic hydrocarbons, but also vehicular-traffic fluxes, emissions and so on; in short, to all variables contributing to atmospheric pollution in urban and industrial areas (provided the sample dimension is smaller than 25). Of course, the present scientific results are nor exhaustive, but they're not in contradiction with known outcomes from parametric statistical techniques.

Acknowledgements

The author sincerely thanks Gianfranco Bielli, past supervisor of Environment Department of ARPA LAZIO, and his collaborators for providing data to be examined.

References

- [1] Siegel S., *Non Parametric Statistics: for the Behavioral Science*, McGraw – Hill Book Company, Inc., New York, 1956.
- [2] Anirvan Banerji Economic Cycle Research Institute, *The lead profile and other non-parametric tools to evaluate survey series as leading indicators*, 24th CIRET Conference: Wellington, New Zealand; 1999 17-20, March.
- [3] Moore G. H., Wallis W. A., 1943, *Time series significance tests based on signs of difference*, J. Amer. Statist. 1943; 38: 153-164.
- [4] Goodrich R. L., 1989, *Applied Statistical Forecasting*, Business Forecast Systems, Inc.
- [5] ARPAV - ARPA, Percorsi Formativi Finalizzati all'Adozione di Omogenei Programmi di Garanzia di Qualità per la Gestione delle Reti di Rilevamento per la Qualità dell'Aria. Linea Progettuale 3b GQ "Programmi di garanzia per la qualità delle reti di rilevamento della qualità dell'aria", Modulo A "Qualità e Processo di rilevamento"; 2004; Roma 10-11 giugno.



- [6] MARRADI A., *Fedeltà di un dato, affidabilità di una definizione operativa*, Rassegna Italiana di Sociologia, XXXI, 1990; 1: 55-96.
- [7] Aivazian S., *Etudes statistique de dépendance*, Editions MIR, Moscow.
- [8] Valero F., Rodriguez R. M., Garcia-Miguel J. A., *Filling Data Algorithms in Urban Air Pollution Monitoring Networks*, J. Environ. Engin. 1997; 395-404.
- [9] Mura M. C., *Statistical elements in calculation procedure for air-quality control*, ISTISAN report 01/2, Rome: Istituto Superiore di Sanità 2001.
- [10] Hamilton J. D., *Time Series Analysis*, Princeton University Press, Chichester, West Sussex, 1994.
- [11] Hoel P. G., *Introduction to Mathematical Statistics*, IV edition, Wiley, New York, 1971.
- [12] Mood A. M., *On the Asymptotic Efficiency of Certain Non Parametric Two-sample Test*, Annals of Mathematic Statistics 1954; 25: 514-522.
- [13] Parrish D. D., Trainer M., Williams E. J., Fahey D. W., Hubler G., Eubank C. S., Liu S. C., Murphy P. C., Albritton D. L., Fenseld F. C., *Measurements of NO_x-O₃ Photostationary State at Nitwot Rodge, Colorado*, J. Geophys. Res. 1986; 91: 5361-5370.
- [14] Kuang-Jung H., *Time Series Analysis of the Interdependence Among Air Pollutants*, Atmos. Environ. 1992; 268: 491-503.
- [15] Gotoh T., *Relation between heat islands and nitrogen dioxide pollution in some Japanese cities*, Atm. Envir. Part B: Urban Atmosphere, 1993, 27B (1): 121-128.
- [16] Mura M. C., Fuselli S., Garcia Miguel J. A., Valero F., *Benzene, Monossido di Carbonio ed Ozono nell'atmosfera di un'area di Roma-Est. Relazioni Statistiche in uno studio preliminare*, Boll. Geof. 1995; 18: 39-46.
- [17] Menichini E., Merli F. Monfredini F., *The temporal variability of the profile of carcinogenic polycyclic aromatic hydrocarbons in urban air: a study in a medium traffic area in Rome, 1993-199*, Atmos. Environ. 1999; 33: 3739-3750.

