Use of Multiple Linear Regressions to Evaluate the Influence of $O_3$ and PM$_{10}$ on Biological Pollutants


Abstract—Exposure to ambient air pollution has been linked to a number of health outcomes, starting from modest transient changes in the respiratory tract and impaired pulmonary function, continuing to restrict activity/reduce performance and to the increase emergency room visits, hospital admissions or mortality. The increase of allergic symptoms has been associated with air contaminants such as ozone, particulate matter, fungal spores and pollen.

considering the potential relevance of crossed effects of non-biological pollutants and airborne pollens and fungal spores on allergy worsening, the aim of this work was to evaluate the influence of non-biological pollutants ($O_3$ and PM$_{10}$) and meteorological parameters on the concentrations of pollen and fungal spores using multiple linear regressions.

The data considered in this study were collected in Oporto which is the second largest Portuguese city, located in the North. Daily mean of $O_3$, PM$_{10}$, pollen and fungal spore concentrations, temperature, relative humidity, precipitation, wind velocity, pollen and fungal spore concentrations, for 2003, 2004 and 2005 were considered. Results showed that the 90th percentile of the adjusted coefficient of determination, $P_{90}$ ($R^2_{aj}$), of the multiple regressions varied from 0.613 to 0.916 for pollen and from 0.275 to 0.512 for fungal spores. $O_3$ and PM$_{10}$ showed to have some influence on the biological pollutants. Among the meteorological parameters analysed, temperature was the one that most influenced the pollen and fungal spores airborne concentrations. Relative humidity also showed to have some influence on the fungal spore dispersion.

Nevertheless, the models for each pollen and fungal spore were different depending on the analysed period, which means that the correlations identified as statistically significant can not be, even so, consistent enough.

Keywords—Air pollutants, meteorological parameters, biological pollutants, multiple linear correlations.
Pollen is the male gametophyte of seed plants and is produced as part of the sexual reproduction cycle. Fungal spores, uni- or multicellular, are reproductive or distributional structures produced during the life cycle of fungi. About 100,000 species of fungal spores are known today, being about 80 related with allergic pathologies, whereas allergenic proteins have been identified in 23 fungal genera [8]. Pollen grains and fungal spores are inert particles being seasonal air pollutants. Their presence and dispersion in the atmosphere depends on a wide range of factors including meteorological (temperature, rain, humidity, wind, etc), biological (physiological state of plants, plant distribution, pollinators, etc) and topographical issues.

In the last decades, epidemiological studies have showed that those particles may be responsible for various pathologies in the respiratory tract [9] and an increasing number of aerobiological studies has been conducted around the world [10]-[14].

Also the relation between meteorological parameters and the biological particles has been worldwide documented [12], [15]-[21]. However, studies concerning the relations between non-biological pollutants and biological particles are still few. The relationships between inhalable airborne pollen and fungal spores with some chemical air pollutants were examined by [22] using basic linear correlations and linear regression models for different pairs of variables. [23] analyzed correlations between fungal spores (not considering pollens), non-biological pollutants and meteorological factors using correlation coefficients and multiple linear regressions; they concluded that the correlations were complex, showing that further studies are needed on these topics. Furthermore, as far as the authors know, the influence of multiple parameters on pollen was not yet studied.

Considering the potential relevance of crossed effects of non-biological pollutants and pollens and fungal spores on allergy worsening [24], the aim of this work was to evaluate the influence of ozone, PM$_{10}$ and meteorological parameters on the concentrations of pollen and fungal spores using multiple linear regressions.

II. METHODS
A. Data

The data considered in this study were collected in Oporto which is the second largest Portuguese city, located in the North. With about 263 thousand inhabitants and a population density of 5787 inhabitants per square kilometre, Oporto is limited on the west by the Atlantic Ocean and on the south by the Douro River. The annual average temperature is around 15°C and the difference between warmer and colder monthly averages is less than 10°C. Annual air humidity is between 75% and 80%, and the total annual mean precipitation varies between 1000 mm and 1200 mm, with about 40% in the winter season, and with more than one hundred days per year with precipitation equal to or higher than 1.0 mm. Prevailing winds are from the W and NW in summer and from the E and SE in winter [25], [26].

Samplers were located at one urban site with traffic influences. O$_3$ and PM$_{10}$ samplers were located in the area covered by pollen and fungal spores sampling. Ornamental and nonornamental trees, shrubs and herbaceous species can be found in the surroundings of the sampling site (some of this species are considered to be allergenic pollen producers and can act as an inoculation source for fungal).

The study here reported considered the daily mean of ozone (O$_3$) and PM$_{10}$ concentrations, temperature (T), relative humidity (RH), precipitation (PP), wind velocity (WV), pollen and fungal spore concentrations, for 2003, 2004 and 2005. The non-biological data were recorded by the Air Quality Monitoring Network of Oporto-MA, managed by the Regional Commission of Coordination and Development of Northern Portugal (Comissão de Coordenação e Desenvolvimento Regional do Norte), under the responsibility of the Ministry of Environment.

O$_3$ measurements were performed through UV-absorption photometry using the equipment 41MUV Photometric Ozone Analyser from Environment S.A., according to ISO 13964, Portaria n° 623/96 and EU Directive 2002/3/CE. PM$_{10}$ concentrations were obtained through the beta radiation attenuation method, considered equivalent to the one advised by EU Directive 1999/30/CE and by Decreto-Lei n° 111/2002, using the equipment MPSI 100 I et E from Environment S.A. All the equipment was submitted to a rigid maintenance program being periodically calibrated.

The meteorological parameters were continuously measured by the Geophysical Institute of Oporto University (Instituto Geofísico da Universidade do Porto) at Serra do Pilar on the left edge of the Douro River at an approximate altitude of 90 m. The analysis of the wind patterns in different places of the city, allowed concluding that the wind pattern at Serra do Pilar can be considered representative of the wind patterns at the site here considered.

Airborne pollen and fungal spores were continuously monitored, using a seven day Hirst type volumetric spore trap manufactured by Burkard Manufacturing Company Limited UK. This sampler has a 2x14 mm intake orifice through which the sampled air is impacted onto a drum, rotating once every 7 days. A vane tail keeps the intake orifice facing the wind and a vacuum pump allows a suction of ten litres of air per minute that is equivalent to the human inhalation. Airborne particles were trapped on a Melinex tape coated with silicone oil, which was cut in seven pieces of 48 mm (each one corresponds to one day) and mounted on the slides with a mounting media of glycerol jelly. Pollen grains and fungal spores were identified and counted under an optical microscope (400X) using four and two traverses lines, respectively, evenly spread over the glass slide.

The pollinic types considered in this study were: Acer spp., Alnus spp., Asteraceae, Betulaceae, Caryophyllaceae, Castanea spp., Chenopodiaceae/Amaranthaceae, Corylus spp., Cupressaceae,Ericaceae, Fraxinus spp., Myrtaceae, Olea europaea, Pinaceae, Plantago spp., Platanus spp., Poaceae,
Quercus spp., Rumex spp., Salix spp., Ulmus spp. and
Urticaceae. Due to the difficulty on distinguishing some
pollen of plant species belonging to the same family, because
of its similar morphology, it is common to aggregate them by
the family names.

The fungal spore types considered were: Alternaria spp.,
Aspergillaceae, Botrytis spp., Cladosporium spp., Coprinus
spp., Corynespora spp., Didymella spp., Drechslera spp.,
Epiconium spp., Rusts, Fusarium spp., Ganoderma spp.,
Leptosphaeria spp., Oidium spp., Periconia spp., Pithomyces
spp., Pleospora spp., Polytrichium spp., Rhizopus stolonifer,
Torula spp. and Ustilago spp.. The sampling method used in
this study does not allow the distinction between Aspergillus
spp. and Penicillium spp. spores due to their reduced size
(Aspergillus spp.: 2-10 µm; Penicillium spp.: 3-5 µm), being
both spore types included in the family Aspergillaceae.

B. Statistical Model

Multiple linear regression (MLR) was used to model pollen
and fungal spores, considering non-biological pollutant
concentrations (O₃ and PM₁₀) and meteorological parameters
(T, RH, PP and WV) as predictors.

MLR is a commonly used method in environmental
sciences. The measured variable y is given by:

\[ y = P_0 + \sum_{i=1}^{k} P_i x_i + \varepsilon \]  

where \( x_i \) are the predictor variables, \( P_i \) the regression
coefficients and \( \varepsilon \) the error associated to the regression
assumed to have expectation of zero. The predicted variable
given by the regression model \( \hat{y} \) can be written as:

\[ \hat{y} = P_0 + \sum_{i=1}^{k} P_i x_i \]  

The regression parameters \( P_i \) are calculated by minimizing
the sum of square errors, through:

\[ P_i = \arg \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  

The statistical significance of the regressions obtained was
also analysed calculating the critical correlation coefficient,
\( R_{crit} \):

\[ R_{crit} = \pm \frac{1}{\sqrt{DF + t_{crit}^2}} \]  

with a significance level of 0.05 (two-tailed test), being DF=n-
k degrees of freedom, k the number of independent variables
and n the number of data.

The correlation is statistically valid if the absolute value of
\( R_{crit} \) is lower than the absolute value of the correlation coefficient which is given by:

\[ R = \pm \sqrt{\frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}} \]  

The behaviour of MLR was evaluated calculating the
adjusted correlation of determination (\( R^2_{adj} \)), and the root
mean squared error (RMSE) according to the following
equations:

\[ R^2_{adj} = R^2 - \frac{k-1}{n-k} \times (1 - R^2) \]  

\[ RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \]  

where \( R^2 \) is the coefficient of determination.

\( R^2_{adj} \) is lower than \( R^2 \), adjusting it for the number of explanatory terms in a model. Unlike \( R^2 \), the \( R^2_{adj} \) increases
only if the new term improves more the model than would be
expected by chance. As this test does not give the model
accuracy, RMSE was also reported (7).

This statistical parameter measures residual errors, which
provides a global idea of the difference between the observed
and modelled values [27], [28].

III. RESULTS

The multiple linear regressions were performed using as
input variables the minimum, average and maximum values of
O₃ and PM₁₀ concentrations, T, RH, PP and WV. For each
period, regressions were performed for all types of pollen and
fungal spores. Tables I and II show the statistically valid
multiple linear regressions, for pollen and fungal spores,
respectively, with statistically valid parameters. These
parameters were calculated using genetic algorithms [29], [30].

The 90th percentile of \( R^2_{adj} (P_{90} (R^2_{adj})) \) of multiple linear
regressions were 0.609 and 0.274, respectively, for pollen and
fungal spores. Tables I and II show the statistically valid

The 90th percentile of \( R^2_{adj} (P_{90} (R^2_{adj})) \) of multiple linear
regressions were 0.609 and 0.274, respectively, for pollen and
fungal spores. Tables I and II show the statistically valid

Results showed that \( P_{90} (R^2_{adj}) \) for pollen multiple linear
regressions varied from 0.613 to 0.916. The statistically valid
multiple linear regressions with \( R^2_{adj} \) superior to \( P_{90} (R^2_{adj}) \)
were obtained for: i) Acer spp. (2003 and 2005); ii) Alnus spp.
(2003); iii) Caryophyllaceae (2005); iv) Castanea spp. (2003);
v) Pinaceae (2005); and vi) Ulmus spp. (2004). The pollen
multiple regression models for 2003-2005 presented \( R^2_{adj} \)
between 0.039 and 0.544.
Table II shows that the $P_{90}$ ($R_{aj}^2$) for fungal spores multiple linear regressions varied from 0.275 to 0.512. The statistically valid multiple linear regressions with statistically meaningful, indicating perhaps a coincidence in these particles pattern of occurrence rather than a cause-effect relationship.

IV. DISCUSSION

For pollen, the multiple linear regressions with statistically valid parameters varied with the year and the type of pollen. However, it was possible to perceive that $O_3$ influenced these types of pollen, with exception of Caryophylaceae and Pinaceae. The influence of PM$_{10}$ was only observed for Acer spp. (2005), Caryophylaceae and Castanea spp.; $O_3$ was present on Acer spp. (2005), Caryophylaceae, and Pinaceae. $O_3$ showed significant negative influence on the concentrations of fungal spores. The influence of PM$_{10}$ was not very significant.

Table I

<table>
<thead>
<tr>
<th>Period</th>
<th>Regression $^a$</th>
<th>n $^b$</th>
<th>$R_{aj}^2$</th>
<th>RMSE $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Aceraceae $=0.369+0.221\times O_3-0.191\times O_{3hr}-0.288\times RH_{hr}$</td>
<td>26</td>
<td>0.721</td>
<td>0.231</td>
</tr>
<tr>
<td>2005</td>
<td>Aceraceae $=0.644-0.216\times O_{3hr}+0.198\times PM\text{<em>{10hr}}+0.907\times T-0.594\times T</em>{hr}-0.173\times PP_{hr}+0.214\times WV_{hr}$</td>
<td>20</td>
<td>0.916</td>
<td>0.133</td>
</tr>
<tr>
<td>2003</td>
<td>Alnus spp. $=0.201-0.079\times O_3+0.142\times T_{hr}-0.062\times PP_{hr}$</td>
<td>36</td>
<td>0.680</td>
<td>0.124</td>
</tr>
<tr>
<td>2005</td>
<td>Caryophylaceae $=0.028-0.061\times O_{3hr}+0.120\times O_3-0.128\times O_{3hr}-0.066\times PM\text{<em>{10hr}} + +0.081\times PM\text{</em>{10hr}}-0.063\times HR_{hr}+0.086\times HR_{hr}+0.026\times PP_{hr}+0.057\times WV_{hr}+0.067\times WV_{hr}$</td>
<td>33</td>
<td>0.631</td>
<td>0.042</td>
</tr>
<tr>
<td>2003</td>
<td>Castanea spp. $=0.328-0.172\times O_{3hr}+0.510\times T+0.168\times HR+0.225\times WV_{hr}$</td>
<td>38</td>
<td>0.634</td>
<td>0.236</td>
</tr>
<tr>
<td>2005</td>
<td>Pinaceae $=0.601-0.154\times PM\text{<em>{10hr}}+0.393\times T</em>{hr}$</td>
<td>52</td>
<td>0.613</td>
<td>0.314</td>
</tr>
<tr>
<td>2004</td>
<td>Ulmus spp. $=0.655+0.447\times O_3+0.461\times PM\text{<em>{10hr}}+0.220\times T</em>{hr}-0.280\times PP_{hr}$</td>
<td>26</td>
<td>0.629</td>
<td>0.329</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Aceraceae $=0.568+0.189\times O_3-0.139\times PM_{10}+0.118\times PM\text{<em>{10hr}}+0.356\times T-0.132\times PP</em>{hr}+0.141\times WV_{hr}$</td>
<td>75</td>
<td>0.544</td>
<td>0.354</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Alnus spp. $=0.441-0.093\times PM_{10hr}-0.153\times T_{hr}+0.180\times T_{hr}-0.075\times PP_{hr}-0.072\times WV_{hr}$</td>
<td>83</td>
<td>0.378</td>
<td>0.267</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Betulaeae $=0.225+0.199\times AvgO_3-0.094\times MinPM_{10}+0.192\times AvgPM_{10}-0.086\times MaxPM_{10}-0.103\times MinWV$</td>
<td>130</td>
<td>0.254</td>
<td>0.319</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Caryophylaceae $=0.107-0.047\times O_{3hr}+0.105\times O_3-0.056\times O_{3hr}-0.035\times PM\text{<em>{10hr}}+0.076\times PM\text{</em>{10hr}}-\sim-0.101\times T_{hr}+0.095\times T$</td>
<td>119</td>
<td>0.385</td>
<td>0.139</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Castanea spp. $=0.264+0.251\times T+0.119\times HR+0.114\times WV_{hr}$</td>
<td>116</td>
<td>0.235</td>
<td>0.344</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Olea europaea $=0.348+0.221\times O_3+0.131\times PM_{10}-0.135\times WV$</td>
<td>68</td>
<td>0.269</td>
<td>0.320</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Platanaceae $=0.463+0.305\times O_{3hr}+0.144\times PM\text{<em>{10hr}}+0.153\times PP</em>{hr}$</td>
<td>79</td>
<td>0.223</td>
<td>0.582</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Pinaceae $=0.642+0.195\times O_3+0.106\times PM\text{<em>{10hr}}+0.188\times T</em>{hr}-0.196\times RH+0.177\times RH_{hr}$</td>
<td>142</td>
<td>0.365</td>
<td>0.388</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Ulmus spp. $=0.398+0.155\times PM\text{<em>{10hr}}-0.099\times T</em>{hr}-0.196\times HR_{hr}$</td>
<td>91</td>
<td>0.265</td>
<td>0.410</td>
</tr>
<tr>
<td>2003-2005</td>
<td>Urticaceae $=0.809+0.102\times O_3+0.086\times O_{3hr}+0.138\times PM_{10}-0.068\times PM\text{<em>{10hr}}+0.191\times T</em>{hr}+\sim+0.111\times HR+0.048\times WV_{hr}$</td>
<td>277</td>
<td>0.385</td>
<td>0.313</td>
</tr>
</tbody>
</table>

*Logarithm of pollen concentrations were used to develop the regression models; $^b$: number of data; $^c$: adjusted determination coefficient; $^d$: RMSE: root mean squared error.
T and RH were the parameters that showed to have more influence. The spores without T influence were Aspergillaceae, Coprinus spp. and Oidium spp.; and those not influenced by RH were Aspergillaceae, Corynespora spp., Drechslera spp., Oidium spp., Ustilago spp and the total fungal spores. In general, the other meteorological parameters did not show a great influence.

Similar results were achieved by [23]: multiple linear regressions of total fungal spores, with contributions of temperature, PM10 and O3 on the fungal spores. For Aspergillaceae, Cladosporium spp. and Ganoderma spp., the results were different; even so, the influence of temperature was frequent on both studies.

Nevertheless, the models for each pollen and fungal spore were different depending on the year, i.e., different influences were found for each year and for 2003-2005, which makes the models somehow inconsistent.

Accordingly, it should be remarked that the results previously published supported in only one period, may not be safe enough, which means that the comparison of the results here presented with those obtained by other authors should be made carefully (the differences can also be due to regional specificities, such as climate, vegetation, species diversity and fungal growth substrates).

However, in spite of the inconsistent trends between air pollutants and pollen and fungal spores, further research should be made concerning their crossed interaction with health. In fact, pollen and fungal spore allergens may reach peripheral airways, leading to airway reactivity and symptoms exacerbation, due to its transference to other small particles such as PM [31], especially in regions such as Oporto, where its concentrations exceeded the standard levels (Directive 1999/30/EC). Additionally, the concentrations of other pollutants, such as O3, can produce an inflammatory effect on the airways causing increased permeability and easier penetration of pollen and fungal spore allergens [2], [33].

V. CONCLUSION

In general, statistically significant correlations were observed randomly. The parameters that influenced the multiple linear regressions, performed for each pollen and fungal spore, depended on the analysed period, which means that the correlations identified as statistically significant can be inconsistent. Thus, it seems that the concentrations of ozone and PM10 do not influence most of the airborne pollen and fungal spore concentrations. According to the different conclusions reached, that depended on the periods analysed, it should be remarked that the results previously published supported in only one period, may not be safe enough.

Therefore, the comparison of the results here presented with those obtained by other authors should be made carefully, because the differences can also be due to regional
specificities, such as climate, vegetation, species diversity and fungal growth substrates.

Nevertheless, in spite of the inconsistent trends between air pollutants and pollen and fungal spores, further research should be made concerning their crossed interaction with health.

ACKNOWLEDGMENTS

Authors are grateful to Comissão de Coordenação da Direcção Regional-Norte and to Instituto Geográfico da Universidade do Porto for kindly providing the air quality and meteorological data. The authors also thank to Fundação Calouste Gulbenkian (project 77161). S. I. V. Sousa also thanks the FCT for the fellowship SFRD/BD/38843/2007.

REFERENCES


