Forecasting maximum hourly ozone concentrations on a daily basis in Belgium by means of the model SMOGSTOP.

Gilbert Lissens*, Walter Debruyn*, Gerwin Dumont**
**VITO, CETAP, Boeretang 200, 2400 Mol, Belgium*
Email:lissensg@vito.be, dbruynw@vito.be
***IRCEL, Kroonlaan 310, 1050 Brussel,Belgium*
Email: dumont@grafair.ihe.be

Abstract

Every summer, ground level ozone concentrations rise in Belgium and cause episodes of photochemical summer smog. This phenomenon is the cause of well recognised public health distress especially for people suffering from respiratory diseases. In order to warn groups of sensitive people against forthcoming smog episodes, VITO (the Flemish Institute for Technological Research) and VMM (the Flemish Environmental Agency) have joined forces to create an ozone pollution forecasting model, called SMOGSTOP (Statistical Models Of Groundlevel Short Term Ozone Pollution).

Ozone pollution levels, as well as concentrations of other air pollutants are monitored in Belgium by the telemetric air quality measuring networks of the 3 Belgian Regions: Flanders, Wallonia and Brussels. The major meteorological variables such as the windvector, temperature, pressure, humidity and precipitation, are also monitored by the same networks. The historical time series of those parameters, generated by the networks are the source of input data for SMOGSTOP. Photochemical ozone pollution is the result of complex non-linear chemical reactions and meteorology. As a consequence it is extremely complex to determine relationships between source emissions (ozone precursors) and ambient pollutant immissions (ozone). To deal with this complexity, SMOGSTOP was constructed as an empirical model, linking meteorological and precursor information on a statistical basis.

In this paper, an overview of the process of ozone formation is given, followed by the definition of the explanatory variables which will be used in the model. Then the methodology behind the model is reported and finally the results of the forecasting efforts in the period of 1.5.1995 till 31.8.1995 are presented.

1 The ozone problem in Belgium

In the beginning of the fifties, chemical precursors (hydrocarbons and nitrogen oxydes) of photochemical smog and the adjoint increased ozone concentrations in the troposphere, were identified by a.o. Haagen-Smit [1]. Today, four decades later, photochemical smog is causing more and more problems for public health. Exposure to ozone also induces damage to agricultural crops, forests and ecosystems, as well as to materials such as rubber and paints. In addition to the increase in the number of ozone episodes as the years go by, an increase in the tropospheric background levels of ozone concentrations is perceived [2].

Where a century ago background levels of ozone concentration were about 10ppbv, they have increased to 20 - 40ppbv at the moment. On top of increasing background concentration levels the typical seasonal peaks are superimposed.

The background immisions are at their maximum level in the Spring at about 90 µg/m³, corresponding with an increased vertical turbulence in the atmosphere, which brings down ozone and its precursors from higher layers down to ground level in this season. Ozone episodes peak on top of these background immissions which drive ozone concentrations over the 180 µg/m³ "warning" treshold. For Belgium, there is an increase in the number of times the treshold values for ozone concentration are surpassed [3,4]. Belgium is a country that is situated on the crossroads of a number of very important industrial areas emitting precursors for ozone. The transport sector, through emissions of exhaust gases, and emissions due to refilling of fuels is the major contributor to the total emission of precursors (NO_x, VOC's and CO) in Belgium (49% of NMVOC, 58% of NO_x and 87% of CO in Belgium). Other emissions relate to activities such as the use of solvents (VOC's), electricity generation with gas and coal fired power plants (NO_x),...

Additional sources of VOC's are biogenic like the emissions of terpenes by pine trees. Summertime meteorological conditions (high temperatures, low wind speeds) invoke the formation of ozone as a secondary pollutant from these precursor molecules.

2 Precursors of ozone

Details on the chemistry of ozone formation can be found in literature. For this paper it is enough to determine the major possible pollutants having an influence on the formation of ozone. These are identified as :

- NO_2
- NO
- VOC's

3 Meteorology

Weather patterns patterns play an important role in establishing conditions conducive to ozone formation and accumulation, as well as conditions terminating ozone episodes. Meteorological conditions determine whether ozone and/or its precursors will be retained locally, (stagnant air masses) or will be transported downwind. High temperatures will enhance the rate of ozone formation. Some processes such as moving high pressure systems, can disperse ozone and transport it to or from higher layers in the atmosphere. Episodes are usually terminated with the frontal passage of cool clean air. Clouds play an important role in the vertical distribution of ozone and its precursors. Meteorological parameters commonly referred to in literature as being important are :

- Daily Maximum temperature, daily temperature range and other temperature related variables are crucial parameters for identifying episodes of ozone pollution [5,6,7,8,9,10]
- Weather patterns bringing stratospheric ozone down to ground level, have sometimes a significant impact on the background concentration of ozone. A low air humidity usually corresponds to this situation. [11,12]
- Ozone episodes are usually related to the occurrence of slow moving high pressure zones [13,14]

4 Input data for SMOGSTOP

Meteorological parameters

- Daily maximum temperature (TMAX)
- Difference between daily maximum temperature and daily minimum temperature (TRANGE) which mimics vertical turbulence.
- Average daily windspeed (AVGWSP)

Immission parameters

- Morning concentration of ozone (and possibly NO_2 if measured at that station) (MORNOX).
- Maximum ozone concentration of the previous day (O3_1)

A number of other parameters could be included, but are left out either because no sufficient data was available (e.g. hydrocarbon concentration) or because they add little to the forecasting power of the model (e.g. precipitation, pressure,...), or they could not be made available in time for the operator to include them in the real time model run.

5 SMOGSTOP : The models

A set of 8 models are fitted every time SMOGSTOP is run. They can be grouped into 3 categories :

- Regression models (2)
- “Nearest neighbour”-like models (2)
- Stratified pattern matching models (or SP-models) (4)

Methodology of the stratified pattern matching approach.

First of all, all explanatory variables are normalised in the traditional way, so whenever an explanatory variable is mentioned, actually, the normalised explanatory variable is addressed.

The process starts off with the grouping of the historical data according to the measured maximum daily ozone concentrations. This gives us a number of classes k with $k=1,...,K$. For each of the explanatory variables X_j the average for each class k $\overline{X}_{jk}$ is computed, as well as the standard deviations $s_{X_{jk}}$.

Out of the vector of explanatory variables $\hat{X}_j$ for the forecast, the following derived variables (n_{jk}) are computed :

$$\forall_{j,k}: n_{jk} = \frac{\hat{X}_j - \overline{X}_{jk}}{s_{X_{jk}}} \qquad (1)$$

The association measure A_k for membership probability of the observed pattern of meteorological and immision parameters to ozone concentration class k is calculated based on a subset of the explanatory variables. In this case, they include TMAX, TRANGE and MORNOX, three variables that show a monotonous behaviour over the different classes, sorted in ascending order. We denote this subset of the total number of explanatory variables as X_p, with $p=1,...,P$, and $P<T$, T being the total number of explanatory variables.

$$\forall k: A_k = \frac{\prod_{p=1}^{P} \frac{w_p \cdot \left(1 - \int_{-\infty}^{|n_{pk}|} \frac{1}{\sqrt{2p}} \cdot e^{-\frac{z^2}{2}} dz\right)}{\sum_{p=1}^{P} w_p}}{\sum_{k=1}^{K} \left(\prod_{p=1}^{P} \frac{w_p \cdot \left(1 - \int_{-\infty}^{|n_{pk}|} \frac{1}{\sqrt{2p}} \cdot e^{-\frac{z^2}{2}} dz\right)}{\sum_{p=1}^{P} w_p} \right)} \quad (2)$$

with

$$\sum_{k=1}^{K} A_k = 1.0 \quad (3)$$

The weights w_p in this equation are defined as the correlation coëfficients of the corresponding variables with the ozone concentration :

$$\forall p: w_p = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{[O_3]_i - [\overline{O}_3]}{s_{[O_3]}} \right) \cdot \left(\frac{X_{ip} - \overline{X}_p}{s_{X_p}} \right) \quad (4)$$

From hereon we move to the determination of the historical pattern that matches most closely the observed pattern of meteorological and immission data within each class k of ozone concentration levels. For every class a distance measure is calculated, representing the match between all historical measurements within a class of ozone concentration levels and the observed pattern. From hereon, all the explanatory variables are used instead of the subset of the explanatory variables. The distance measure is calculated in the following way :

$$\forall i,k: D_{ik} = \sum_{j=1}^{T} w_{jk} \cdot \left(\hat{X}_j - X_{ijk} \right)^2 \quad (5)$$

The weights w_{jk} represent the correlation coëfficiënts of the respective variables X_j within each class k of ozone concentration levels.

The ozone concentration which corresponds with the minimum of these distance measures is taken as the expected value of the ozone concentration in this class ($[\hat{O}_3]_k$).

The forecasted ozone concentration is then defined as the sum of the expected values of the ozone concentration in each class, multiplied by the association measure A_k for all classes.

$$\left[\hat{O}_3\right] = \sum_{k=1}^{K} \left[\hat{O}_3\right]_k \cdot A_k \tag{6}$$

Four models are run which differ in the way the expected maximum ozone $[\hat{O}_3]_k$ concentration within each class matches is calculated

- Model 1 (*SP:Best Match*): takes the ozone concentration corresponding to the best match in every class as $[\hat{O}_3]_k$ (the ozone concentration which corresponds to the minimum of the distance measures)
- Model 2 (*SP: Avg(2)*): takes the (distance weighted) average of the ozone concentration of the two best matches in every class as $[\hat{O}_3]_k$
- Model 3 (*SP: Med(3)*): takes the median of the ozone concentration of the three best matches in every class as $[\hat{O}_3]_k$
- Model 4 (*SP: Overestim*): takes the highest ozone value of the three best matches in every class as $[\hat{O}_3]_k$ (this is in fact a manipulated overestimation)

6 The model output

- The model produces :
  - a forecast of the daily 1 hour maximum ozone concentrations at each measuring station
  - a map of geographically interpolated predicted maxima over Belgium:
    - in a 5 x 5 km grid
    - devided over smog classes: (classes can be defined by the user : defaults are reported below)

      0 - 120 µg/m³ (no smog)
      120 - 180 µg/m³ (minor smog)
      180 - 240 µg/m³ (moderate smog)
      240 - 360 µg/m³ (major smog)
      > 360 µg/m³ (alarm level)
- During the summer of 1996 the model was adapted in order to predict ozone maxima for day+0, day+1 and also day+2. The [O_3-1] input values for day+1 and day+2 are the model's own output for day+0 and day+1 repectively. The [MORNOX]-input for day+1 and for day+2 are replicated from the input for day+0. Temperature and windspeed input are taken from the different forecast reports for day+0, day+1 and day+2, issued by a professional forecasting bureau which has access to European weather forecasting systems.

7 Program environment

The computer program SMOGSTOP (Statistical Models Of Groundlevel Short Term Ozone Pollution) is embedded in a user friendly "Windows-like"shell. It is programmed in Visual Basic and currently runs on a Pentium PC (100 Mhz, 32 Mb RAM). A complete forecast running the 8 models and producing the ozone maps for Belgium takes about 15 minutes.

Several add-ons like a file manager for transforming the raw data from the measurement sites to the file format used by the SMOGSTOP program and a graphical interface which projects the forecasts onto a map of Belgium are available.

SMOGSTOP is fairly easily adaptable to other countries, regions or cities, provided the necessary basic material (time series of ozone measurements and meteorological parameters) is available for the concerned area.

8 Results

Forecasting maximum daily ozone concentrations is a tedious problem, where conventional methods often fail to perform well. The objectives are to forecast extreme ozone concentrations, which occur relatively little. Contrary to normal statistical procedures, which often minimise the influence of extremes, the interest of this research is on the extremes. Mix this in with a highly non-linear behaviour with the explanatory variables when the maximum daily ozone concentrations reach a certain level, and what we have got is an interesting case for trying alternative non-parametric or semi-parametric models. Added to this general statistical problem, there is a problem of insufficiency of data. In Belgium ozone measurements up to the beginning of 1995 have been performed at 9 sites since 1986. In 1995, 11 sites were added, and forecasts for these stations were clearly suffering from lack of data. Another problem having to do with the operational requirements of the model, and the precursors and meteorological parameters measured in Belgium forced us to limit the number of explanatory variables. Several interesting variables pictured in sections 2 and 3 of this paper could not be used in the model, either because they were not measured on a regular basis, could not be made available in time to the operator, or were measured at too little a number of sites to be included. In what follows we will focus on some of the features and forecasting power of the stratified pattern matching models compared to other, more general, techniques, serving as benchmarks for the new approach.

The first benchmark is the frequently used inertia benchmark. This benchmark consists of taking the maximum ozone level of the previous day as a forecast for the maximum ozone level of the current day. A regression model, with the ozone concentration as dependent variable and the same independent variables as the other models, was fitted and will be used as a second benchmark. Two other models, very comparable to "nearest neighbour" methods serve as a third benchmark. Nearest neighbour models calculate a “distance measure” between the set of explanatory variables for the forecasted day and the set of explanatory variables for all other observations in the database. The forecast of the ozone concentration for this day is the maximum ozone concentration measured in the past for which the observed explanatory variables match most closely the pattern of explanatory variables for the forecasted day (i.e. the maximum ozone concentration of the day for which the distance measure was minimal). In this report the “best match” maximum ozone concentration and the distance weighted average of the maximum ozone concentration of the five best matches are reported.

The tests for the models will be carried out on the summer of 1995, a hot summer with a relatively large number of exceedances (29) of the 180µg/m³ treshold in Belgium.

The crucial test for the model is the contingency table which crosstabulates the observed ozone and non-ozone days against the predicted ozone days and non-ozone days (table 1). A particular day is called an ozone day if at least one of the 20 stations has measured an hourly ozone concentration higher than the 180 µg/m³ treshold.

From table 1, we can immediately see the inappropriateness of regression models for forecasting high ozone concentration levels. Only 2 out of 29 ozone days were predicted. What regression models do predict well are the non-ozone days. For this purpose, the regression models outscore the other models. The inertia model scores better in predicting ozone days than the regression model, but stays well below the scores of the nearest neighbour and the stratified pattern matching models. The nearest neighbour models score about the same on predicting ozone days, but produce significantly more “false ozone warnings” (an ozone day was predicted, but not observed). Within the class of stratified pattern matching models, the differences in performance are small.

Due to the relatively small size of Belgium and therefore the close proximity of the stations to one another (the average distance from one station to its nearest neighbour is 25 km), the demands for very accurate estimates of immission levels of maximum ozone concentration levels at the point of measurement were considered less important than the prediction of the very occurrence of an ozone day. In other words, less attention was given to the problem of **where** exceedances occurred than to the problem of **if** exceedances occurred.

Table 1 : Contingency tables for ozone days

	No O_3 day obs. No O_3 day pred.	No O_3 day obs. O_3 day pred.	O_3 day obs. No O_3 day pred.	O_3 day obs. O_3 day pred.
INERTIA	81	13	13	16
REGRESSION	93	1	27	2
NN*: Best Match	79	15	5	24
NN: Avg.(5)	83	11	8	21
SP**: Best Match	86	8	4	25
SP: Avg(2)	87	7	6	23
SP: Med(3)	87	7	6	23
SP: Overestim.	84	10	4	25

* NN = Nearest Neighbour
** SP = Stratified Pattern matching

Table 2 gives the contingency table for all stations together. It is the sum for all the stations of the correct and incorrect predictions of the ozone days **at the point of measurement**.

Table 2 : Contingency tables for ozone days at the point of measurement

	No O_3 day obs. No O_3 day pred.	No O_3 day obs. O_3 day pred.	O_3 day obs. No O_3 day pred.	O_3 day obs. O_3 day pred.
INERTIA	1065	72	74	44
REGRESSION	1132	5	115	3
NN: Best Match	1062	75	69	49
NN: Avg.(5)	1095	42	85	33
SP: Best Match	1072	65	75	43
SP: Avg(2)	1077	60	72	46
SP: Med(3)	1084	53	71	47
SP: Overestim.	1051	86	57	61

Key words

Ozone pollution forecasting, pattern matching, modelling, photochemical smog

References

[1] Haagen-Smit A.J. Chemistry and physiology of Los Angeles smog, *Ind. Eng. Chem.*, 1952, 44, 1342-1346.

[2] Dumont G. en Roekens E. Photochemical air pollution (in Dutch), Leren om te keren, Milieu- en Natuurrapport Vlaanderen, 1994.

[3] De Muer D. Recente ontwikkelingen in de ozonlaag en de ultraviolette straling boven België en Nederland, Een rapportage van het KMI, het RIVM en het KNMI, 1993.

[4] Roekens E. Study of the photochemical air pollution in Belgium, 1979-1990 (in Dutch), IHE, Afdeling Lucht, 1992.

[5] Feister U. and Balzer K. Surface O_3 and meteo predictors on a subregional scale, *Atmospheric environment*, 1991, 25A, 9, 1781-1790.

[6] Korsog P.E. and Wolff G.T. An examination of urban ozone trends in the Northeastern U.S. (1973-1983) using a robust statistical model, *Atmospheric environment*, 1991, 25B, 1, 47-57.

[7] Stoechenius T.E. A multivariate analysis technique for assessing the influence of meteorological conditions on ozone concentration trends, paper 91-119.5, *8th Annual air and waste management meeting*, Vancouver, 1991.

[8] Kelly N.A. et al. The chemical and meteorological conditions associated with high and low ozone concentrations in Southeastern Michigan and nearby areas of Ontario ,paper 85-22.4, *78th Annual meeting of the air pollution control association*, Detroit, 1985.

[9] Liu et al. Important meteorological parameters for O_3 episodes experienced in the Taipei basin, *Atmospheric environment*, 1994, 28, 1, 159-173.

[10] Yap D. Relationship of O_3 to meteorological conditions in Southern Ontario, *70th Annual meeting of the air pollution control association*, Toronto, 1977.

[11] Gay M.J. Meteorological and altitudinal influences on the concentration of ozone at Great Dun Fell, *Atmospheric environment*, 1991, 25A, 9, 1767-1779.

[12] Kato H. et al. Mechanism of spring high oxidant episodes, a meteorological analysis in and around the Hokuriku district, Japan, *Atmospheric environment*, 1990, 24A, 8, 2023-2033.

[13] Rethinking the ozone problem in regional and urban air pollution, National Research Council, National Academy Press, 93-107, 1992.

[14] Ludwig F.L. and Sheler E. Effects of weather fronts on O_3 transport, Air quality, meteorology and atmospheric O_3. ASTM STP653, 389-406, 1978.