GEOPHYSICAL RESEARCH LETTERS, VOL. 20, NO. 12, PAGES 1179-1182, JUNE 18, 1993

A METHOD OF COMPARISON OF TWO CLOSE BATCHES DATA: APPLICATION TO ANALYSIS OF FOG FORMATION CAUSES

J. J. Boreux
Fondation Universitaire Luxembourgeoise

J. Guiot
Laboratoire de Botanique Historique et Palynologie

Abstract. Given suitable conditions of air temperature and humidity, the density of a fog and its microphysical properties depend mainly on the availability of cloud condensation nuclei (CCN) and their nature. Fogs become particularly dense near certain industrial plants because of high concentration of hygroscopic combustion particles in the air. Their role in dense fog formation is estimated by comparing the local climates and CCN concentrations at two similar sites, the first being more subject to air pollution and dense fogs than the second. Orthogonal regression is applied to three meteorological variables (air temperature, relative humidity, wind speed) and CCN concentration. As we compare very close variables, bootstrap provides precise confidence intervals independent of Gaussian assumptions. Two sites are compared : they are located in the Meuse valley (Belgium) at a distance of about 15 km. We found that the local climate of the polluted site is not only colder and wetter but also richer in CCN that the control site. These results suggest interactions of natural and anthropogenic causes in dense fog formation at industrial site. This method is useful in various domains of geophysics when correlated time series have to be compared.

1. Introduction

Fogs and aerosols are intimately associated. The role of hygroscopic aerosol particles in fog formation has long been studied. The main result is that the fog formation, fogwater chemical composition and fog duration are strongly correlated with the ambient air quality (Munger et al. 1990, Hudson et al. 1991). It is well known that topography has an effect on air quality, especially in case of ground inversion.

In the Meuse valley (Belgium) the fog nuisances are numerous. We shall mention the notorious case of the first week in December 1930 when numerous deaths have been correlated with a blocking anticyclone (Jaumotte, 1931). Today, fog causes serious road accidents.

To reduce accident risks we can forecast fog formation (Boreux and Guiot, 1992) but also act on the smoke emissions, all the more as they play a role in fog formation.

Beez and Tailfer are two sites having a "V shaped valley" but the first, industrial by nature, only is frequently subject to dense fogs (Boreux, 1988). Tailfer, very close to Beez, can be considered as a control site. In order to distinguish between natural and anthropogenic causes of dense fogs it is necessary to compare the local climates of both sites. In the case there is no systematic and significant differences between the two local climates, the frequent dense fog occurrences at Beez can likely be connected with human activities. In the opposite case, additional investigations on the cloud condensation nuclei, i.e. CCN, concentration are absolutely necessary for decision-making. It has been pointed out by a simple method (and it is confirmed in this paper) that the industrial site is richer in CCN concentration than the control site (Boreux and Serpolay, 1990). The difference should be due to the industries producing magnesian limestone, lime and naphthalene. This explanation is in agreement with many investigations which have shown that the physico-chemical properties of aerosol particles have a direct influence on the life cycle of radiation fogs (Twomey, 1968; Bott, 1991).

The local climates comparison as that of the local charge in CCN are based on the time series pairwise comparisons of selected physical parameters measured during a sufficiently long monitoring period. The orthogonal regression is the most appropriate method when the two variables to be compared have a symmetric role. The confidence intervals for each regression parameter (i.e. intercept, slope and correlation coefficient) are computed by bootstrap (Efron, 1979; 1981). The main difficulty stems from autocorrelations which reduce the number of degrees of freedom to much less than the number of observations.

This close time series comparison technique should provide a valuable assistance in many domains of geophysics when time series observations are to be compared.

2. Data

In the Meuse valley (Belgium) we have compared the local climates of BEEZ (50°28'N, 4°56'E, 80 m) and TAILFER (50°24'N, 4°52'E, 80 m). The neighborhood of Beez, industrial by nature, is often submitted to very localized dense fogs while Tailfer is not. Both sites have been equipped with an automatic meteorological station which supplies every 30 minutes the values of three meteorologic variables: screen air temperature, air humidity and wind speed averaged over this period. The monitoring period was the first five months in 1988 (Table 1).

Table 1. Meteorological and microphysical variables studied at Beez and Tailfer between January and May 1988.

Variable	Symbol	Unity	Size
air temperature	AT	°C	4544
relative humidity	WS	%	4545
wind speed	WS	m/s	3603
CCN concentration	N	cm^{-3}	226

To lay emphasis on the role of Beez human activities in fog formation and duration, we used a thermal diffusion chamber (TDC). In a TDC a steady state flux of water vapor is maintained between two wet horizontal surfaces having different temperatures (Fig.1). The chamber being filled with sampled air, water vapor diffuses from the warm surface with temperature T_W toward the cold surface with temperature T_C. At each level in the chamber, the vapor pressure is larger than the saturation pressure, i.e. a

Paper number 93GL00075
0094-8534/93/93GL-00075$03.00

supersaturation S is created. The maximum supersaturation S_m occurring nearly midway between the two surfaces can be estimated as a function of the temperature difference $\Delta T = T_w - T_c$ (Götz et al, 1991):

$$S_m \cong \Delta T^2 / 25 \qquad (1)$$

where S_m is given in percent. By illuminating the chamber at a level where S_m occurs, the droplets being in the light beam may be photographed to determine the concentration of CCN having a critical supersaturation S_c less then or equal to S_m . This critical supersaturation is related to critical radius of the developping droplet, the lowest size from which the droplet can grow bigger. The lower is the critical supersaturation of a particle, the greater is its critical radius and the higher is the probability for its participation in droplet formation.

As it was first proposed by Twomey (1959), supersaturation spectra follow a power law of type

$$N = c\, S^k \qquad (2)$$

where N is the total number concentration of CCN active at a value of supersaturation S, while c and k are constant determined by solving the logarithm-transform of Eq. (2) for each experiment.

A important fact is that the thermal diffusion chambers activate all particles having a radius larger than a critical radius corresponding to the supersaturation created into the TDC.

We carried out 54 comparative experiments between Beez and Tailfer including 226 measurements of supersaturations S ranging from 0.1 % to 1.8 %. The 226 values of c and k obtained by solving Eq. (2) characterize the local aerosol of both sites. We found that, at the 90 % level, k is not significantly different for both sites (with a value about 0.60) while c is 3250 cm^{-3} to 3360 cm^{-3} at Tailfer (the control site) against 3630 cm^{-3} to 3760 cm^{-3} at Beez (the industrial site). In other words, it is confirmed that a systematic increase of CCN concentration of order of magnitude 10^2 cm^{-3} appears in the polluted site.

3. Methodology

Let us consider two time-series to be compared. Air temperature (Fig. 2) does not seem clearly different in both

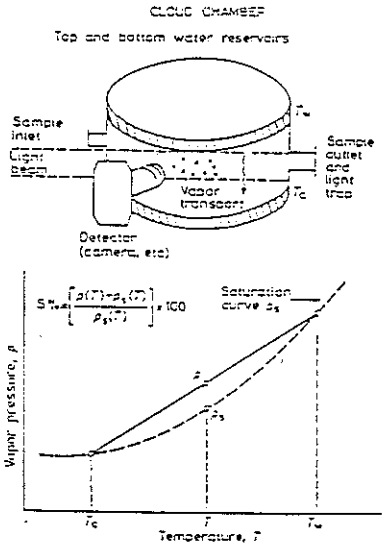

Fig.1. Principle of a thermal diffusion chamber (from Götz et al, 1991, pp 94). The gradient temperature between the upper and the lower plates induces a supersaturation which is maximum half-way up. All particles having a critical supersaturation lower are activated.

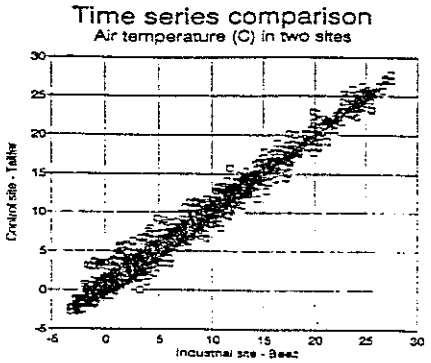

Fig. 2. Scatterplot of the temperature at Tailfer in function of the temperature at Beez.

sites while CCN concentration is (Fig. 3). Of course it is necessary to give a precise and rigorous sense to this statement. Because there is no causal relationship between the time series, i.e. the variables play a symmetric role, the appropriate technique for analysis is orthogonal regression which minimizes the sum of the residual deviations products (Dagnélie, 1973).

Let X and Y represent the respective time series, the general equation is then written either

$$\hat{Y} = a + bX \quad \text{or} \quad \hat{X} = b^{-1}(Y - a) \qquad (3)$$

where $\hat{Y}$ represents the estimate of Y given X and $\hat{X}$ the estimate from Y.

For a given set of observations (i = 1,....n) the quantity Q to be minimized is defined as

$$Q = |x_i - b^{-1}(y_i - a)|\,|y_i - a - bx_i| \qquad (4)$$

Using the residuals defined as $u_i = y_i - \hat{y}_i = y_i - a - bx_i$, Eq. (4) becomes:

$$Q = |b|^{-1} \sum_{i=1}^{n} u_i^2 \qquad (5)$$

and the corresponding normal equations can be written as

$$\sum_{i=1}^{n} u_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} u_i^2 + 2b \sum_{i=1}^{n} u_i x_i = 0 \qquad (6)$$

The orthogonal regression line is easily derived from these two latter

$$\hat{Y} = (S_y/S_x)(X - \bar{x}) + \bar{y} \qquad (7)$$

where S_y , $\bar{y}$, S_x and $\bar{x}$ are, respectively, the standard

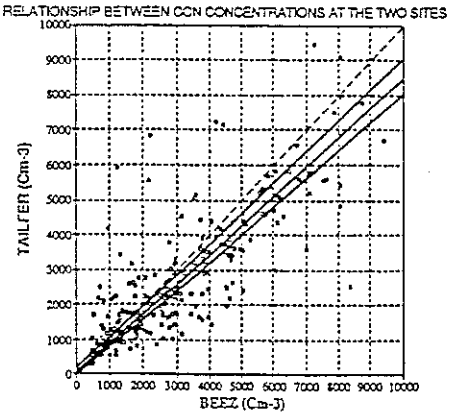

Fig. 3. Scatterplot of CCN concentrations measured at Tailfer against this measured at Beez. The dashed line represents the identity relationship and the three heavy lines are the orthogonal regression straight lines with its 95 % confidence level.

deviation and the mean value of Y and X while the sign in Eq. (7) is given by that of their correlation coefficient r.

Both variables will be considered as not significantly different if the straight line given by Eq. (7) has a slope close to 45° and an intercept close to zero while the correlation coefficient r tends to 1. The bootstrap technique gives us the possibility to compute a robust confidence interval for each parameter estimated without particular assumptions on the residuals distribution (Efron, 1979; 1981). It has been successfully applied in paleoclimatology by Guiot (1990) and Till and Guiot (1990) and in fog forecasting by Boreux and Guiot (1992).

The idea of the bootstrap method is to resample the original observations in a suitable way to construct pseudo data sets from which the estimations are performed. The variability of the estimates from one pseudo data set to another is the key for their reliability evaluation. The resampling is done by random drawing with replacement. A pseudo data set contains as many observations as the original one and consequently some observations are present several times and others are not.

For each pseudo data set we calibrate an orthogonal regression which gives us a particular value for each parameter in Eq.(3): hence for the k th pseudo data set we have a_k , b_k and r_k where $1 \leq k \leq K$. The K regressions (i.e. the K slopes, intercepts and correlations) are summarized by their mean values and standard deviations. The range [mean value - 2*standard deviation, mean value + 2*standard deviation] provides a 95%-confidence interval if we assume that these K calculated parameters have a Gaussian distribution (the example in Fig. 4 shows that this assumption is reasonable). To avoid the normality assumption, we prefer to use the 2.5, 50 and 97.5 percentiles which provide the parameters with their confidence interval.

4. Methodological discussion

Each observation of the data set must have the same probability of selection by random drawing. This probability is given by the inverse of the number of observations (i.e. 1/N). Because of large autocorrelations due to the sampling time step (half an hour), each observation is over-represented : even when observation x_t is not taken, its correlation with x_{t-1}, x_{t-1}, x_{t+2}... makes it to be indirectly present in the pseudo dataset. These high autocorrelations make underestimate the bootstrap errors.

A solution to avoid these biases is to resample the data at larger time intervals before beginning further calculations. Fig. 5 presents the variations of the bootstrap errors on the intercept (a) as a function of temperature autocorrelations at

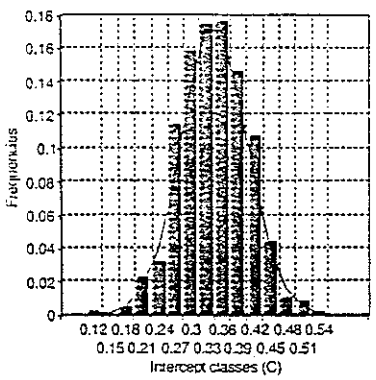

Fig. 4. Histogram of the 500 bootstrap estimates of the intercept (Tailfer versus Beez) and fitted Gaussian law with mean 0.35 and standard deviation 0.062. The hypothesis of normality cannot be rejected at the 0.05 significance level.

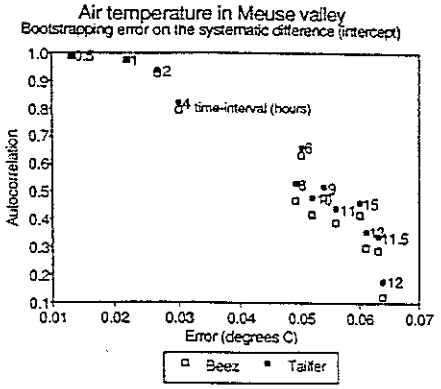

Fig. 5. Relationship between the autocorrelation at various lags (from 0.5 hour to 15 hours) for the Beez and Tailfer air temperatures and the bootstrapping error for the temperature intercept.

Beez and Tailfer. In this figure temperature have been resampled each half hour, each hour and so on until 15 hours. The autocorrelation decreases from 0.99 to 0.12 (for an half day) and increases thereafter until 0.42. In the meantime, the bootstrap error increases from 0.013 to 0.064. There is then a clear relationship between both statistics. The maximum error is found out for an half-day resampling, and between 6 and 15 hours, the error is close to 0.06. We select a resampling value of 12 hours which maximizes the error and which is representative of the daily cycle.

Another parameter to be determined is the number of replications K. Fig. 6 shows the value of the intercept cumulated from 1 to 500 replications. It appears that the intercept varies from 0.11 to 0.52 and, when averaged, tends to 0.33. The stabilization around this value is reached after about 100 replications. We adopt this value for K.

The histogram of the 500 values computed for the temperature intercept is displayed in Fig.4. It shows that we may approximate the distribution with a Gaussian law of mean equal to 0.35 °C and standard deviation equal to 0.06 °C. The confidence interval at the 95%-level (i.e. mean • 2 * standard deviation) is [0.22, 0.46], which means that the temperature at Beez is 0.35 °C significantly higher than in Tailfer. Nevertheless to avoid normality assumptions, we prefer to use confidence intervals defined by the 2.5 and 97.5 percentiles of K replications and the central value defined by the 50 th percentile: [0.21, 0.33, 0.45], which is here very close to the values provided by the mean and standard deviation.

5. Results

As already mentioned, three meteorological variables are considered to compare the local climate of Beez and

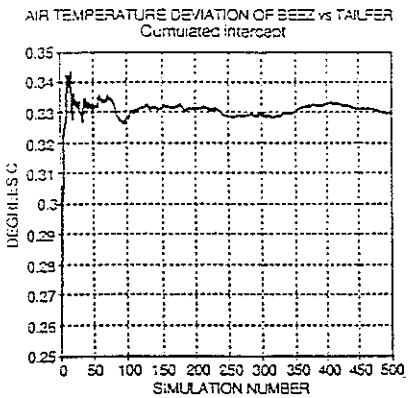

Fig. 6. Cumulated mean intercept for air temperature (Tailfer versus Beez) from replication 1 to 500.

Tailfer: air temperature, relative humidity and wind speed. Table 2 gives the 2.5, 50 and 97.5 percentiles of the intercept (a), the slope (b) and the squared correlation (r^2) established on 100 replications (a value proven to be sufficient in Fig.6).

Table 2. The 2.5, 50 and 97.5 percentiles of the intercept (a), slope (b) and squared correlation (r^2) calculated by bootstrap (100 replications) orthogonal regression between Beez and Tailfer. The variables are defined in Table 1.

variable	a	b	r^2
AT	0.23 0.32 0.46	0.97 0.98 1.00	0.98 0.98 0.99
RH	-5.85 -4.26 -1.91	0.99 1.02 1.04	0.94 0.95 0.97
WS	0.43 0.48 0.54	1.14 1.27 1.47	0.11 0.21 0.37
N	-30 130 270	0.76 0.84 0.90	0.75 0.80 0.87

It appears that two meteorological variables, air temperature and relative humidity, are very correlated from site to site ($r^2 > 0.95$) and that the slope of two orthogonal regression lines is not significantly different from one (at the 95%-level). On the other hand, the intercepts are significantly different from zero, which implies a systematic bias, independent on the values themselves. In other words, the industrial site (i.e. Beez) is significantly less hot (- 0.3 °C) and more moist (+ 4%) then the control site (ie. Tailfer).

As expected, wind speeds are not very strongly correlated between the two sites. This correlation is nevertheless significant at the 95%-level as the 2.5th percentile (0.11) is markedly greater than 0. In contrast with the two other variables, the slope is significantly higher than 1, meaning that the wind speed at the control site has a greater variability than at industrial site. The intercept of 0.48 m/s confirms that usually the wind speed is slightly stronger at Tailfer than Beez.

The temperature, humidity and wind speed conditions are more favorable to fog formation in the industrial site. This effect is likely emphasized by the higher concentration of CCN. Indeed Table 2 shows a significant slope of 0.84 and a negligible intercept. While at high concentration of CCN ($N > 2000$ cm^{-3}) the industrial site has a significantly larger amount of CCN (20 %), at lower concentrations there is no significant difference. Therefore it must be emphasized that the measurements errors are larger when the thermal diffusion chamber operates with lower supersaturation and therefore we cannot distinguish both sites at lower concentrations.

6. Concluding remarks

The Sambre river runs from the industrial basin of Charleroi into the Meuse river at Namur. After the confluence the river surface temperature increases by about 0.7°C (Boreux, 1988). The polluted site, Beez, is situated downstream of Namur where the valley is oriented NNE-WSW, while the control site, Tailfer, is located upstream, where the valley is oriented N-S. Consequently, the south riverside of the industrial site does not store much heat, which explains the systematically lower air temperature at Beez (-0.3°C). If we add the difference of 0.7°C found for the water temperature, the evaporation which is proportional to the difference between water and air temperatures (here 1°C) is stronger at the polluted site than at the control site. This is confirmed by the systematic bias found for the humidity (4.3% less in Tailfer than in Beez). As the polluted site also has higher capacity for CCN activated in a thermal diffusion chamber, i.e. this site contains more smaller particles than the control site, we assume that these particles are emitted by the factories chimneys. To relate the presence of these particles (too small to act as CCN in the atmosphere where the supersaturations are always less than 0.1 %) to the dense fog formation, we can express an hypothesis: the very small particles act as CCN inside the chimneys where temperature gradients exist, i.e. with sufficient supersaturations. With regard to wind speed, it is difficult to conclude because the correlation between both sites is rather poor. Nevertheless we can conclude that the control site, less rich in CCN by nature, is also slightly more windy.

Additional investigations with isothermal haze chamber, which operates with very low supersaturations, would be useful to separate natural and anthropogenic sources.

Acknowledgements: This work has been financed by the Belgian ministry of the "Travaux Publics pour la Région Wallonne". R. Serpolay (Université Blaise Pascal, Clermont-Ferrand, France) has lent the thermal diffusion chamber and Degreane Co (avenue Font-Pré, Toulon, France) has provided the automatical meteorological stations. L. Duckstein (University of Arizona, Tucson, USA) and F. Rogers (Desert Research Institute, Reno, USA) have kindly revised the manuscript.

References

Boreux, J.J., Etude des brouillards locaux denses dans la vallée de la Meuse, FUL Scientific Report, Arlon, 48p, 1988.

Boreux, J.J., and R. Serpolay, Comparative CCN concentration measurements between two sites in the Meuse valley and occurence of localized dense fogs in relation with industrial activities, Proc. Third Intern. Aerosol Conf., 1, 515-518, 1990.

Boreux, J.J., and J. Guiot, A fog forecasting method in a deeply embanked valley, Atmospheric Environment., 26, 759-764, 1992.

Bott, A., On the influence of the physico-chemical properties of aerosols on the life cycle of radiation fogs, Boundary-Layer Meteorology, 56, 1-31, 1991.

Dagnélie, P., Théorie et méthodes statistiques, Vol. 1, Les Presses Agronomiques de Gembloux, Gembloux, 378 p, 1973.

Efron, B., bootstrap methods : Another look at the jacknife, Annals of Statistics, 7, 1-26, 1979.

Efron, B., Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods, Biometrika, 68, 589-599, 1981.

Götz, G., Mészaros, E., and G. Vali, Atmospheric particles and nuclei, Akadémiai Hiado és Nyomda Vallalat, Budapest, 274 p, 1991.

Guiot, J., Methods of calibration, verification and reconstruction. Cook and Kairiukstis (eds): Methods of dendrochronology, application in the Environmental Sciences, IIASA - Kluwer, Dordrecht, 163-217.

Hudson, J. G., Observations of antropogenic cloud condensation nuclei, Atmospheric Environment, 25, 2449-2455, 1991.

Jaumotte, J., Sur le brouillard meurtrier de la vallée de la Meuse, Bull. Soc. Belge d'Astr., de Mét. et Phys. Globe, 3, 4 et 5, 100-106, 1931.

Munger, J.W., Collet, J., Daube, B. and M. R. Hoffmann, Fogwater chemistry at Riverside, California, Atmospheric Environment, 24B, N°2, 185-205, 1990.

Till, C., and J. Guiot, Reconstruction of precipitation in Morocco since A. D.1100 based on Cedrus Atlantica tree-ring widths, Quaternary Research, 33, 337-351, 1990.

(Received: 22April 1992;
Revised: 28 September 1992;
Accepted: 7 December 1992)

(continued from outside back cover)

(continued on facing page)