We analyze the influence of wind distribution in probability density function of ground concentrations of material emitted in the atmosphere. We know that the wind distribution follows a Weibull function and we demonstrate that the ground concentrations follows a Weibull distribution too.

The Weibull SAS program aims at representing the trend of positions measurements (quantiles and percentiles) relative to concentrations distribution of airborne material using the Weibull distribution. In particular we analyze the statistical distribution of airborne pollen grains, but the same program can be used to evaluate the density distribution of air pollution.

We illustrate the program structure, in particular we treat in detail how is possible to give rise a graphic output that represents a straight line if the concentrations follow a Weibull distribution; the problem is to transforme the variables to obtain a straight line and to print the values of the no-transformed variables on the respective axes.

Since these operations turn out to be recurring for each of the input files to be acquired it was decided to use a Macro which is able to parametrize the control of three Data Set and their variables.

The Weibull Distribution

Over recent years, the Weibull distribution has frequently been applied in investigations of wind speed distribution. A review of approaches for estimating the Weibull distribution of wind data can be found in Condradsen et al. (1984a; 1173-1183). The Weibull distribution is expressed by:

\[ f(x) = \frac{b}{a} \left(\frac{x}{a}\right)^{b-1} \exp\left[-\left(\frac{x}{a}\right)^b\right] \]

where \(a\) and \(b\) are the scale and slope parameters respectively.

The cumulative distribution is given by:

\[ F(x) = 1 - \exp\left[-\left(\frac{x}{a}\right)^b\right] \]  \hspace{1cm} (1)

Mean and variance are:

\[ \bar{x} = a \Gamma\left(1 + \frac{1}{b}\right) \]  \hspace{1cm} (2)

\[ \sigma^2 = a^2 \left[ \Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right) \right] \]

where \(\Gamma\) is the natural gamma function.

An important characteristic of the Weibull distribution is that, if \(x\) follows it with slope parameter \(b\) and scale parameter \(a\), then \(x^2\) follows a Weibull distribution with slope parameter \(2\) and scale parameter \(a^2\).

We now attempt to show the influence of wind speed on the turbulent transport of material in the atmosphere. This turns out to be far simpler if we limit ourselves to analytical formulations, i.e. to the Gaussian and Analytical models mentioned in the introduction, where the functional dependence of wind velocity can be analysed directly from the formulae used. In the case of the Gaussian model, the ground level concentration \(C\) can be written:

\[ C = A(u) \exp\left[-F\left(H/H\right)\right] \]  \hspace{1cm} (3)

where

\[ A = \frac{Q}{\sqrt{2\pi} \sigma \Delta \rho \theta} x \]

\[ F = 1/(2\sigma)^x \]

where \(Q\) is the emission, \(\sigma\) the vertical standard deviation, \(x\) the distance from the source and \(\Delta \theta\) the angular width of the sectors in radiants.

We illustrate the program structure, in particular we treat in detail how is possible to give rise a graphic output that represents a straight line if the concentrations follow a Weibull distribution; the problem is to transforme the variables to obtain a straight line and to print the values of the no-transformed variables on the respective axes.

Since these operations turn out to be recurring for each of the input files to be acquired it was decided to use a Macro which is able to parametrize the control of three Data Set and their variables.
lae, such as the sigma in the Gaussian model or the alpha in equation 4, depend, albeit not explicitly, on wind velocity. Due to the characteristics of the abovementioned Weibull distributions, the more the functional expression of C with u is of power type, the more the ground concentration values will follow a Weibull-type distribution. Thus, in the case of Gaussian and Analytical approaches, the implicit dependence on wind velocity in the parameters of expressions 4 and 5 is an element of disturbance. In view of this, the more direct the dependence of C on wind velocity, the more reasonable it seems that diffusion can be accurately expressed by a Weibull distribution. These conditions appear to be best expressed in equation 6 relative to the Box model. Thus, in the case of areal sources and measurements in proximity of or within the same area as the source, one is likely to obtain probability density functions of ground level concentrations which follow a Weibull distribution with good agreement.

The probability density functions of the data were studied in order to establish whether they could be represented by a Weibull function and if so, with what degree of accuracy. In fact, the aim of our investigation was to ascertain the influence of wind on ground level pollen concentrations; bearing in mind the considerations made in the previous section, this influence may be understood through a study of the probability density function: Weibull distributions suggest a significant influence of wind velocity.

The data from a Weibull distribution can be plotted as a straight line. By examining the natural logarithm of the cumulative distribution equation (Eq. 1) and then using logarithm to base 10 on both sides after some arithmetic, it can be seen that a straight line is obtained. Two graphic applications are reported in Figures 1 and 2.

The former illustrates the trend of the daily mean data of three kinds of pollen grains. In the latter the cumulative distribution for one year of SO2 is plotted.

This demonstrates the great adaptability of the SAS program to distinct cases. The only one difference in abovementioned applications is that in the first one we have used a Macro-procedure to control three data sets.

Examining the figures, it can be seen how the data does not generally deviate from the straight line. As well as this visualization, in order to evaluate quantitatively the reliability of the representation of pollen data by means of a Weibull distribution, Chi-squared and Kolmogorof tests were applied, here no carried.

Programme Structure

The WEIBULL SAS3 programme aims at representing the trend of position measurements (quantiles and percentiles) relative to 3 pollen concentration distributions using the WEIBULL distribution.

The programme can be divided into three parts:
A) Data acquisition
B) Percentile computation
C) Transformation into the Weibull and graphic scale

A) Data acquisition
This first stage of the programme allows the definition of the Work Data Set containing the basic information to be processed and the Annotate Data Set required to modify the graphic representation.

In practice, point A can in turn be divided into two subphases:
A.1 Annotate Definition
A.2 Definition of Work Data Set

A.1) Annotate Definition
The Annotate Data Set is defined in a temporary form with the name (WORK).ANNO and defines the LABELS to substitute those produced by default by the graphic procedure.

The procedures defining the Annotate Data Set are as follows:

<table>
<thead>
<tr>
<th>data anno;</th>
<th>length position $ 3;</th>
<th>input function $ x y color $ style $ size position $ text char$;</th>
</tr>
</thead>
<tbody>
<tr>
<td>card;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 10.0 white . 1.2 * 5 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 10.0 white . 1.2 * 10 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 15.0 white . 1.2 * 20 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 16.0 white . 1.2 * 50 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 18.0 white . 1.2 * 75 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 20.0 white . 1.2 * 95 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 12.5 22.0 white . 1.2 * 95 -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 20.0 04.5 white . 1.0 * 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>label 22.5 04.5 white . 1.0 * 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>run;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A.2) Definition of Work Data Set
In this phase, 3 SAS Data Sets are defined which contain the information to be processed. The base data are acquired from 3 external files:
CMS FI DAT1 DISK BOB84 DATI A
CMS FI DAT2 DISK BOB85 DATI A
CMS FI DAT3 DISK BOB86 DATI A

For each of these external files, a SAS temporary data set is generated taking the name of WBN (n = 1..3) and containing the variable Xn (n = 1..3) for processing. The variable Xn represents the pollen concentration distribution collected in sample n.

After the acquisition of the SAS Data Set WBN, a sort procedure (PROC SORT) is carried out on this SDS, with output in the temporary SDS WBSORT (n = 1..3) which therefore contains the values of variable Xn sorted in decreasing order.

The latter SDS subsequently contributes to the creation of a third temporary SDS, called WBCMn (n = 1..3), which as well as containing the variable under analysis, allows the generation of its cumulative function, called CUMn (n=1..3).
Since these operations turn out to be recurring for each of the three files to be acquired, it was decided to use a MACRO called %FIRSTOS which is able to parameterize the control of the three SOS and their variables. The instructions utilized to set up the macro are the following:

```sas
proc sort data=wb1 out=wb1sort; by x1; run;
data wb1; set wb1sort;
run;
```

After the preliminary acquisition of data for processing with the generation of SOS, we pass on to the second phase which relates to the computation of the quantiles and percentiles required for the representation of the sampled pollen concentration values.

**B) Percentile computation**

Percentile computation is carried out by means of the (PROC) UNIVARIATE procedure, acquiring the SAS data set WBCUMn in input and producing in output (NOPRINT option) a new SDS containing the quantile and percentile values in distinct variables:

- Q3=q75
- MEDIAN=q50
- Q1=q25
- PS=PERC5
- PI0=PERC10
- Q9=Q75
- P95=PERCS
- P99=PERC99

The UNIVARIATE procedure is performed twice: a first time on the SDS WBUCUM, using the variable Xn and therefore producing the PERCX SDS; subsequently, it is performed on the same SDS (WBUCUM) using the variable CUMn and producing the PERCC SDS. The two SDS PERCX and PERCC are merged to form a single SDS UNIONE by means of a SET instruction. An example of the content of the SDS UNIONE is given below:

```sas
proc sort data=wb1; by x1; run;
data wb1; set wb1; run;
```

The PROC TRANSPOSE gives rise to the NEWPERC SDS, a transpose of the UNIONE SDS. Subsequently, a new data set called GRPWbn (n = 1..3) is generated which, starting from NEWPERC, redefines the names of the 2 variables (COL1 and COL2); thereupon, by means of a SELECT instruction, it defines a new variable Y which contains the numerical values of the percentiles.

As these operations turn out to be recurring in each of the three data sets under exam, here too, a MACRO-LANGUAGE structure was used.

In practice, the following macro called %GENOS was used:

```sas
%macro firstos(xn, datn, x1, wbsortn, wbcumn, cumn);
data wb;
    input x1 6.1;
    proc sort data=xn out=wbsort;
        by x1;
    run;
    data wb;
        set wbsort;
        accumulate x1;
    run;
    %end firstos;
```

The macro were run as follows:

```sas
%firstos(x1, dat1, x1, wb1sort1, wbcum1, cum1);
%firstos(x2, dat2, x2, wb1sort2, wbcum2, cum2);
%firstos(x3, dat3, x3, wb1sort3, wbcum3, cum3);
```

Since it is necessary to treat cases as if they were variables, a (PROC) TRANSPOSE procedure is used to transform the variables (columns of the UNIONE SDS) into cases and, similarly, cases (lines of the UNIONE SDS) are transformed into variables.

```sas
proc sort data=wbgbkp1 by _name_;
    merge grpb1 grpb2 grpb3;
    by _name_;`
As shown in the figure above, before embarking on the transformation into the WEIBULL scale, a single SDS called TUTTI is generated which contains the MERGE of all previously generated data sets.

C) Transformation Into Weibull and graphic scale

The transformation into Weibull scale is performed during the phase generating the SDS TUTTI, according to the following form:

```plaintext
data tutti;
merge grpwb1
   grpwb2
   grpwb3
by _name_
;
  lxl=log10(x1);
  lx2=log10(x2);
  lx3=log10(x3);
  dif=1-y;
  ldif=log10(dif);
  ly=log10(-ldif);
```

Once a final SORT has been performed on SDS TUTTI to produce the SDS GRAF containing the percentiles sorted in decreasing order of the variable Y, we can proceed with the effective graphic representation of the pollen grain concentration values and relative linear best-fits, thanks to the high graphic potential of the SAS/GRAPH® Software.

The instructions to produce the required graph are as follows:

```plaintext
proc sort data=tutti out=graf;
by descending y;
run;

proc gplot data=graf;
symbol1 c=white v=star i=r1;
symbol2 c=red v=square i=r1;
symbol3 c=cyan v=hash i=r1;
title1 ' ';
title2 c=white "con annotate";
axis1 value " none
     label = (c=white 'percentiles')
     minor = none
     major = none
;
axis2 value = none
     label = (c=white 'pollen per m**3')
plot ly*lxl=1
    ly*lx2=2
    ly*lx3=3 / overlay
      vaxis=axis1
      haxis=axis2
      annotate
      caxis=white
;run;
```

Conclusions

The SAS programme processed has permitted us to demonstrate graphically the linear best-fit of airborne material concentrations on the Weibull scale. In this way, if a data set follows a Weibull distribution, the data will produce a straight line. Thanks to remarkable flexibility, this SAS programme can also be used in the statistical evaluation of any numbers of data set, in particular of pollutant concentrations relative to air quality monitoring networks.

SAS and SAS/GRAPH are registered trademarks of SAS Institute Inc., Cary, NC, USA.

References


Figure 1  Cumulative distribution of daily mean pollen grain in Bologna (Gramineae *,Oleaceae () ,and Urticaceae**). Weibull cumulative probability distributions are represented by Straight line.
RAVENNA CITY
Station N. 15

Figure 2  Cumulative distribution of daily mean concentration of sulphur dioxide relative to station 15 of the air quality monitoring network of RAVENNA (ITALY).