MODELLING AIRBORNE FUNGAL SPORE DATA WITH THE GAMMA DISTRIBUTION

J. Belmonte* & M.A. Canela**

Unitat de Botànica, UAB

(**) Departament de Matemàtica Aplicada i Análisi, UB

ABSTRACT

We present here a discussion on the performance of the gamma distribution as a model for airborne fungal spore daily concentrations, using data extracted from the Catalonian Aerobiological Network (Xarxa Aerobiològica de Catalunya, XAC). This is the sequel of a previous study, dealing with similar models for pollen concentrations (Worcester conference).

The gamma distribution has two parameters, a shape parameter and a scale parameter. The interpretation of these parameters and the relationship between them and the usual statistics are also discussed.

The study includes the eight spore types that are more abundant in the area covered by the network (*Alternaria*, *Arthrinium*, *Aspergillus/Penicillium*, *Cladosporium*, Coprinaceae, *Leptosphaeria*, *Ustilago* and Total spores).

For four of these spore types, the statistics are enough regular to allow a systematic statistical analysis. We examine here the distributions resulting fro this analysis. The fit is assessed by means of a chi square statistic.

Figure 1. Mean daily concentration of Cladosporium (BCN/1999)

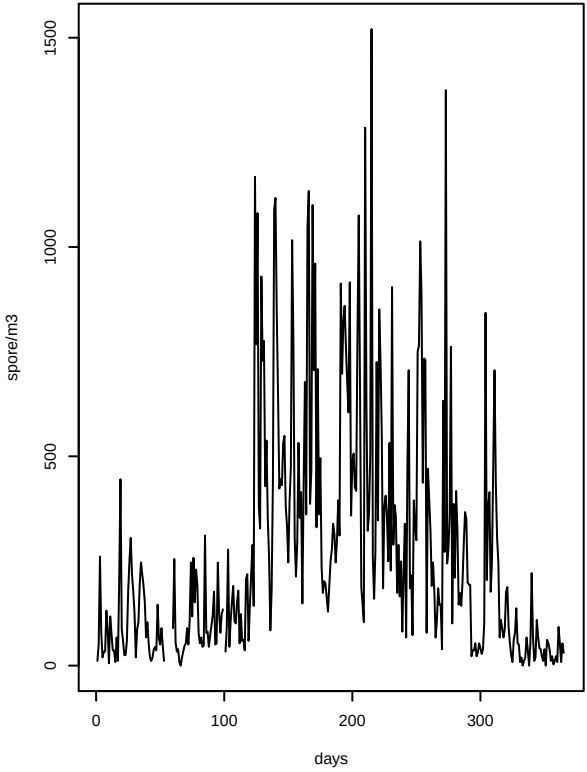
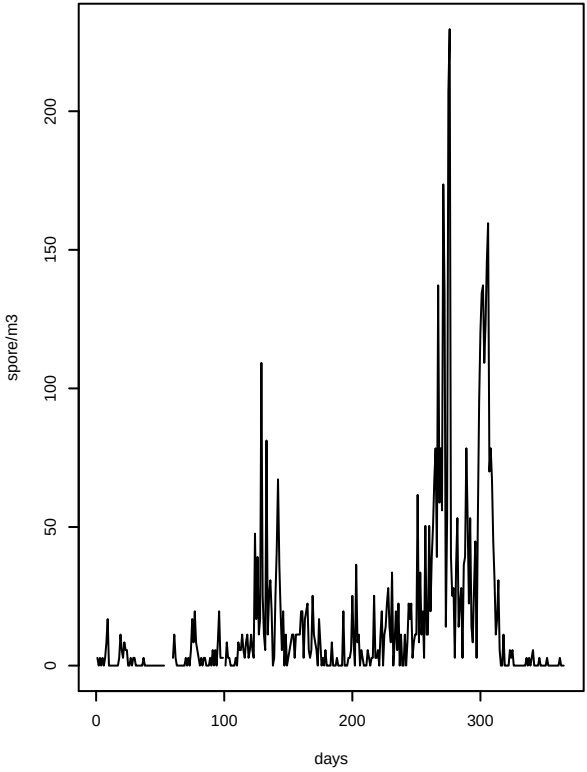


Figure 2. Mean daily concentration of Coprinaceae (BCN/1999)



GAMMA DISTRIBUTION

- Probability density function:

$$f(x) = C\, x^{\alpha-1}\, e^{-x/\beta}, \qquad x > 0,$$

  where C is a constant, which depends on α and β.
- Parameters:
  * α is the shape parameter, no units, scale-free.
  * β is the scale parameter, same units as the concentration.
- Relationship with the mean and variance:

$$\bar{x} = \alpha\beta,$$

$$s^2 = \alpha\beta^2,$$

$$CV = \frac{s}{\bar{x}} = \alpha^{-1/2}.$$

Figure 3. Gamma distributions with equal mean but different shape

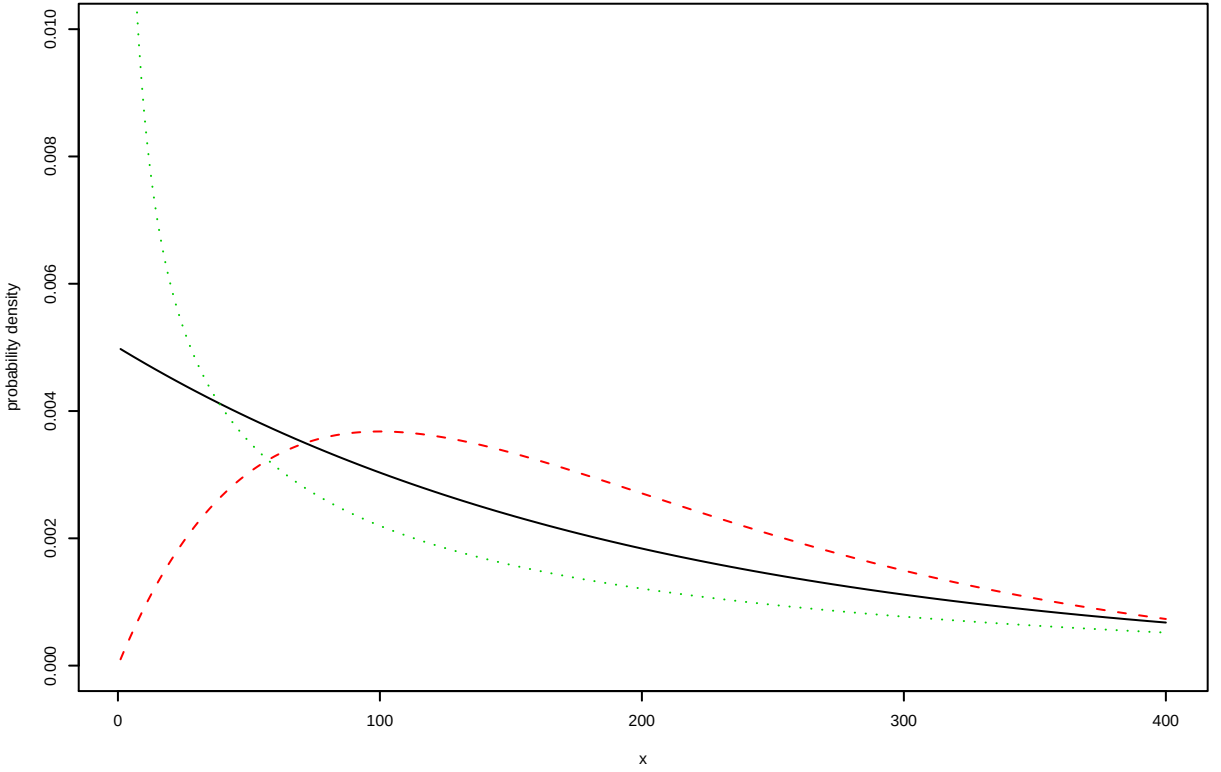


Figure 4. Mean daily concentration of Cladosporium (BCN/1999)

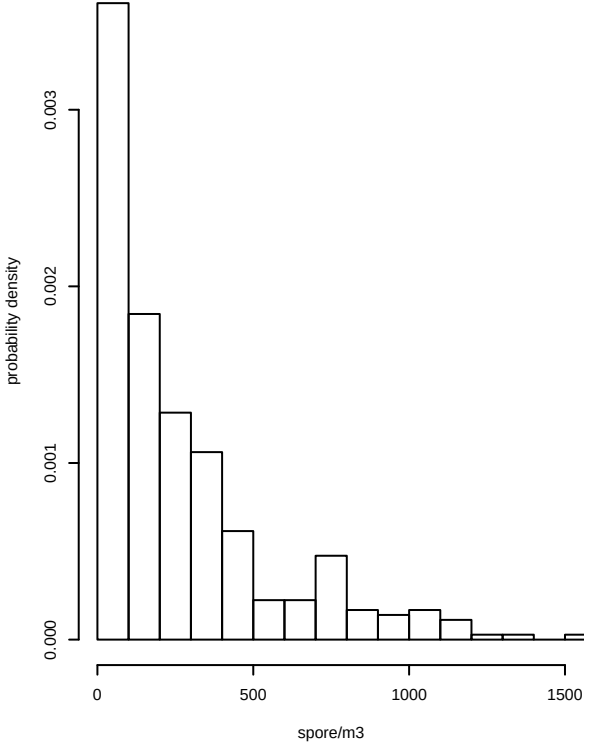


Figure 5. Gamma distribution fitted to Cladosporium data

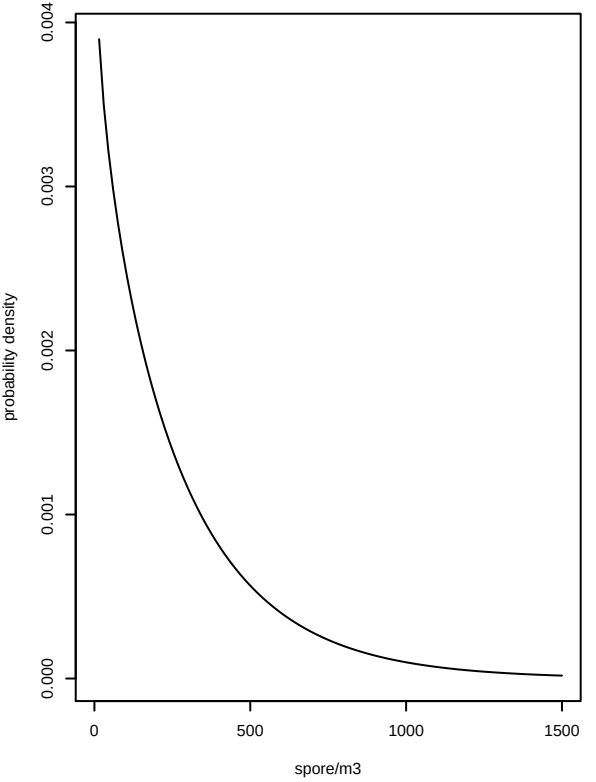


Figure 6. Mean daily concentration of Total spores (BCN/1999)

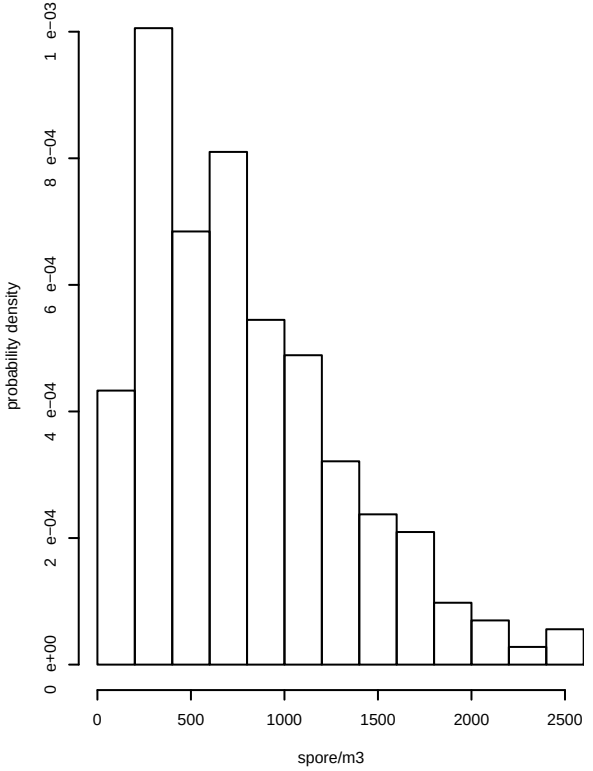
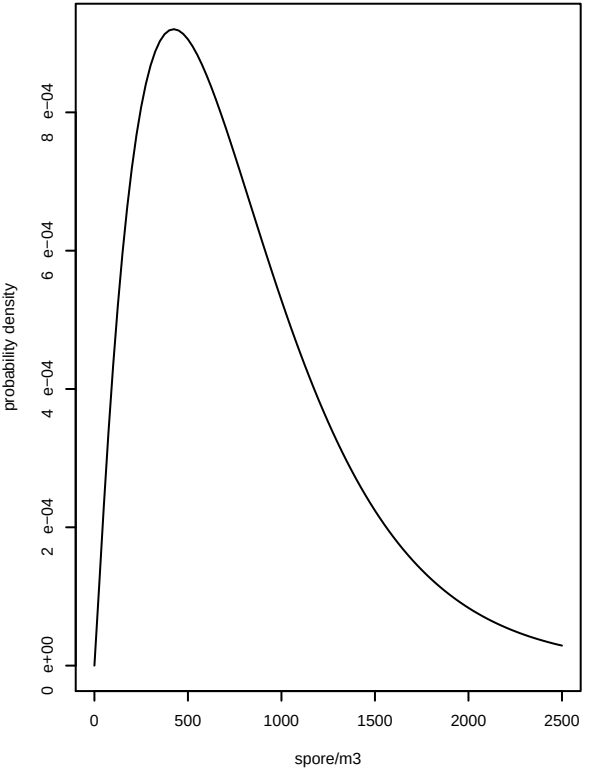


Figure 7. Gamma distribution fitted to Total spores data



EXAMPLE

For *Cladosporium* mean daily spore concentrations measured in the Barcelona station in 1999, the mean and the standard deviation are

$$\bar{x} = 271.69 \text{ spore/m}^3, \qquad s = 283.85 \text{ spore/m}^3,$$

and the coefficient of variation is $CV = 1.045$.

Then the shape and scale parameters are

$$\alpha = \frac{\left(\bar{x}\right)^2}{s^2} = \frac{1}{CV^2} = 0.9162,$$

$$\beta = \frac{s^2}{\bar{x}} = 296.6 \text{ spore/m}^3.$$

Note that α is dimensionless, while β, being a scale parameter, has the same units as the pollen concentrations.

ASSESSING THE FIT

- QQ plot:

  Two-dimensional graphical display of the data set against the quantiles of a theoretical probability distribution.

  The perfect fit corresponds to a plot where all the points belong to the line $y = x$.

- Chi square statistic:

  Measure of the agreement between the probabilities given by a theoretical model, called the *expected proportions* and the proportions actually found in the data, called the *observed proportions*.

  Based on a partition of the range of the variable into a set of intervals.

  It can be used in a significance test, since it follows (approximately) a chi square distribution when the model fitted to the data is valid.

  With 5 intervals, the critical value are $\chi_{0.05} = 7.81$ and $\chi_{0.01} = 11.34$.

Figure 8. Gamma QQ plot for Cladosporium

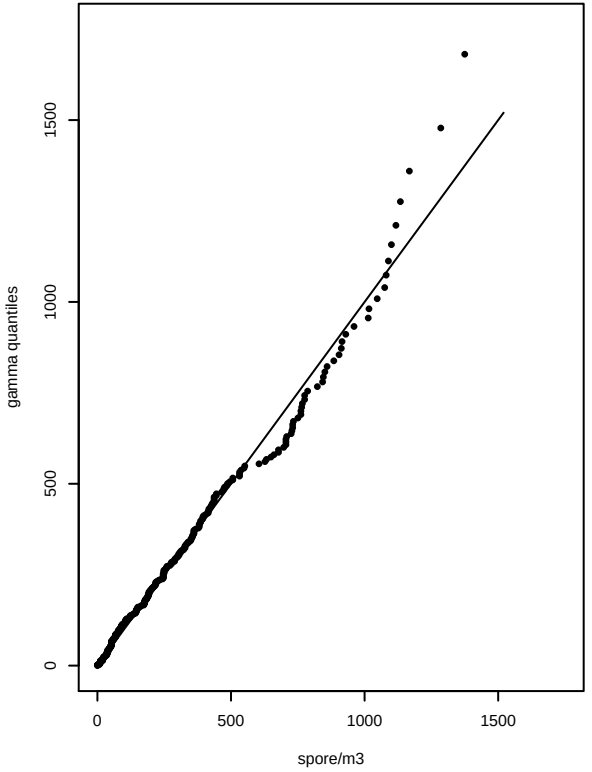


Figure 9. Gamma QQ plot for Cladosporium (sqrt scale)

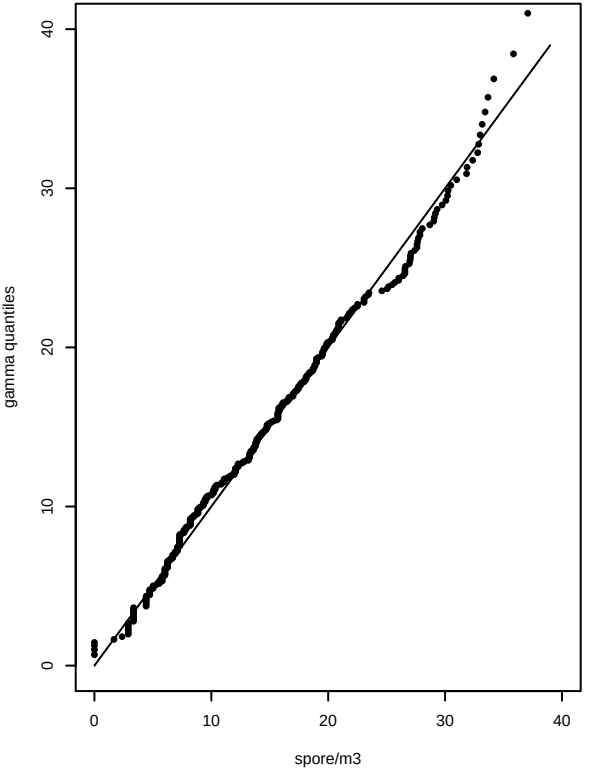
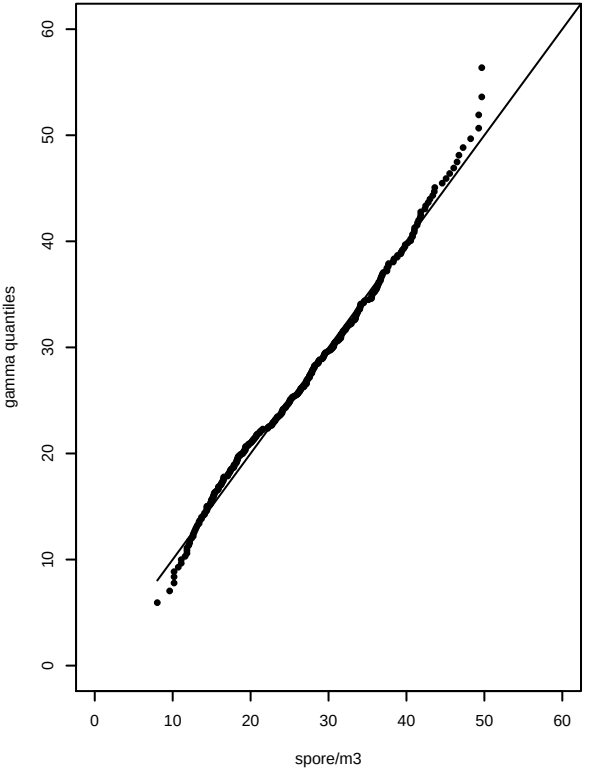


Figure 10. Gamma QQ plot for Total spores (sqrt scale)



CHI SQUARE STATISTICS (1)

TABLE 1. Chi square testing for *Cladosporium* (Bellaterra)

Statistics	1995	1996	1997	1998	1999	2000	2001	2002
Observed (%)								
sp/m3<50	2.6	10.1	4.9	10.7	12.8	11.6	6.8	1.6
50<s/m3<100	4.1	11.3	7.4	9.6	13.1	9.4	16.7	6.9
100<s/m3<250	26.5	28.7	20.6	24.8	21.5	20.5	22.7	24.2
250<s/m3<500	23.6	23.5	16.4	19.6	22.6	18.3	17.5	29.5
s/m3>500	43.2	26.4	50.7	35.3	29.9	40.2	36.2	37.7
Expected (%)								
s/m3<50	10.1	13.3	9.5	19.9	14.3	9.9	17.3	8.7
50<s/m3<100	7.3	11.0	7.2	9.6	10.3	8.3	9.3	8.4
100<s/m3<250	16.7	24.9	16.9	18.4	22.5	19.8	18.6	21.5
250<s/m3<500	19.2	24.4	19.7	17.9	22.6	22.4	18.7	24.5
s/m3>500	46.7	26.3	46.6	34.2	30.3	39.5	36.2	36.8
Chi square	48.8**	4.66	14.3**	24.0**	3.41	4.49	48.4**	26.9**

CHI SQUARE STATISTICS (2)

TABLE 2. Chi square testing for Total spores (Bellaterra)

Statistics	1995	1996	1997	1998	1999	2000	2001	2002
Observed (%)								
s/m3<500	20.7	25.8	10.7	12.4	25.7	36.0	52.9	30.9
500<s/m3<1000	28.5	19.1	17.8	30.3	28.2	26.3	20.3	24.8
1000<s/m3<1500	15.0	21.4	11.5	20.9	19.0	15.5	12.6	19.0
1500<s/m3<2000	13.3	13.0	17.5	13.2	9.8	7.8	4.1	10.5
s/m3>2000	22.5	20.6	42.5	23.1	17.3	14.4	10.1	14.9
Expected (%)								
s/m3<500	22.5	24.3	13.3	24.9	27.3	37.8	50.2	30.4
500<s/m3<1000	24.0	24.0	17.1	20.2	25.5	22.7	21.0	26.4
1000<s/m3<1500	18.1	17.6	15.4	15.1	17.7	14.3	11.6	17.3
1500<s/m3<2000	12.5	12.0	12.8	11.1	11.4	9.1	6.8	10.7
s/m3>2000	22.9	22.1	41.4	28.7	18.1	16.1	10.4	15.3
Chi square	5.46	7.28	12.3**	54.3**	2.69	4.09	4.78	1.01

REFERENCES

J. Belmonte & M. Canela (2003). Modelling aerobiological data with the gamma distribution — Part I: Pollen data. Technical Report.

J. Belmonte & M. Canela (2004). Modelling aerobiological data with the gamma distribution — Part II: Spore data. Technical Report.

G. Casella & R.L. Berger (1990), *Statistical Inference*. Duxbury Press, Belmont.

P. Comtois (2000). The gamma distribution as the true aerobiological probability density function (PDF). *Aerobiologia* **16**, 171–176.