MODELLING AIRBORNE TREE POLLEN DATA WITH THE GAMMA DISTRIBUTION

J. Belmonte & M. Canela***

(*) Unitat de Botànica, UAB
(**) Departament de Matemàtica Aplicada i Análisi, UB

ABSTRACT

We examine how well the gamma distribution fits airborne tree pollen daily concentrations, using data extracted from the Catalonian Aerobiological Network. Although this paper is restricted to arboreous pollen types, the same methods have being used for shrub and herbaceous pollen data, and even for fungal spore data.

We first discuss the advantages of the gamma distribution with respect to other non-symmetric models, like the lognormal. The gamma distribution has two parameters, shape and scale. The interpretation of the parameters and the relationship with the usual statistics are also discussed.

For each pollen type (22), location (7) and year (1994-2002), we fit a gamma model, so that we can check the stability of the parameters across time, for each site. The fit is assessed both graphically and numerically. The rationale of this approach is that we expect the scale to change from year to year, depending on the climatic conditions, but the shape to be fairly stable. The results obtained confirm this hypothesis, although the scale does not fluctuate independently of the shape.

A benefit of this approach is that we can use the parameters to classify pollen types in a few generic ones. Such a classification is less subjective than other approaches based on a visual inspection of line plots, smoothed or not, and simplifies the management of an aerobiological database. We present a simple proposal for this classification, based on the shape, which could reduce the study of the tree pollen data of the network to three selected types.

MODELLING POLLEN DATA

- Time series:
  - ⋆ Trend analysis
  - ⋆ Autocorrelation

- Probability distributions:
  - ⋆ Normal distribution
  - ⋆ Lognormal distribution
  - ⋆ Gamma distribution

Figure 1. Cupressaceae (BCN/2002)

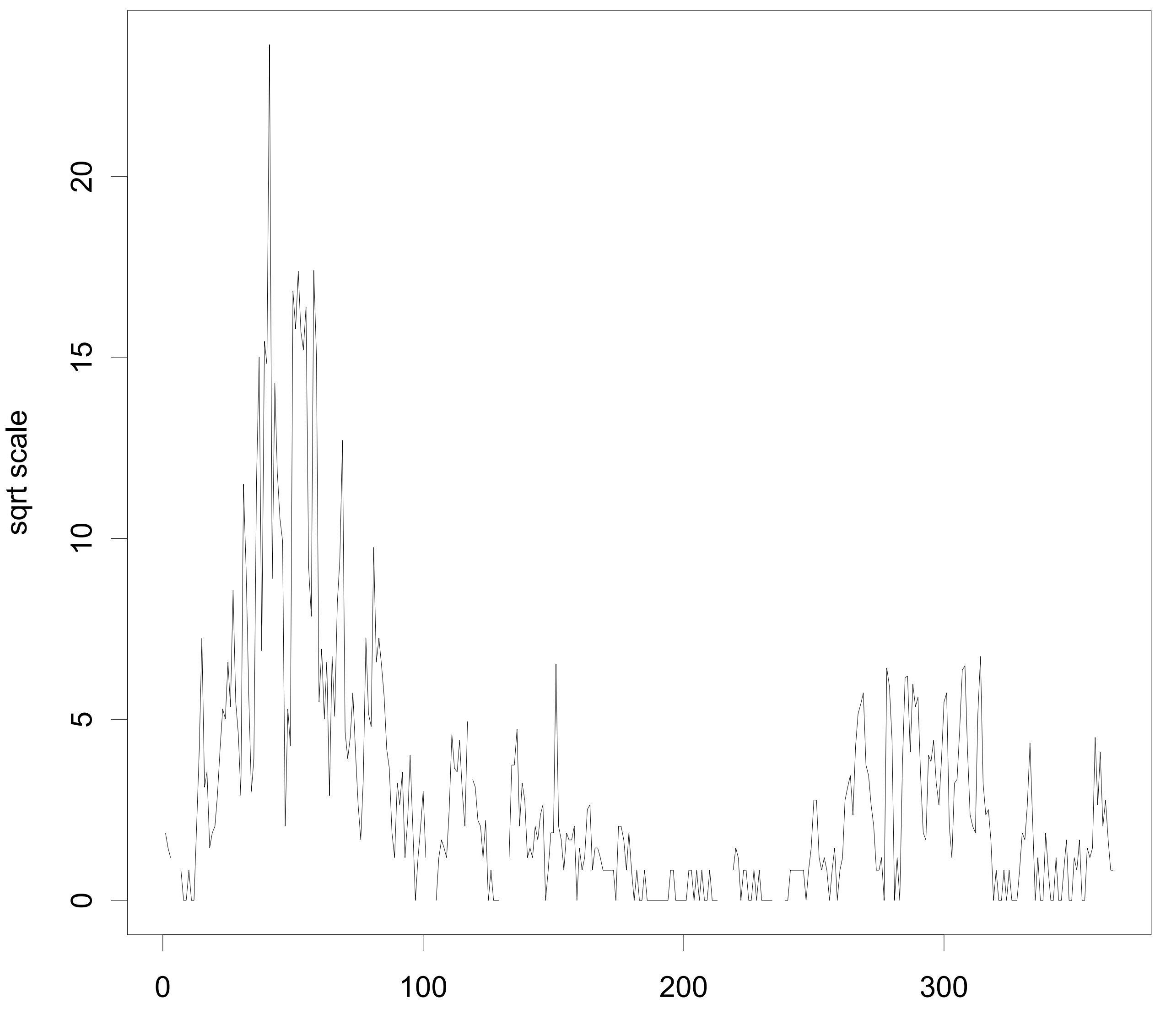
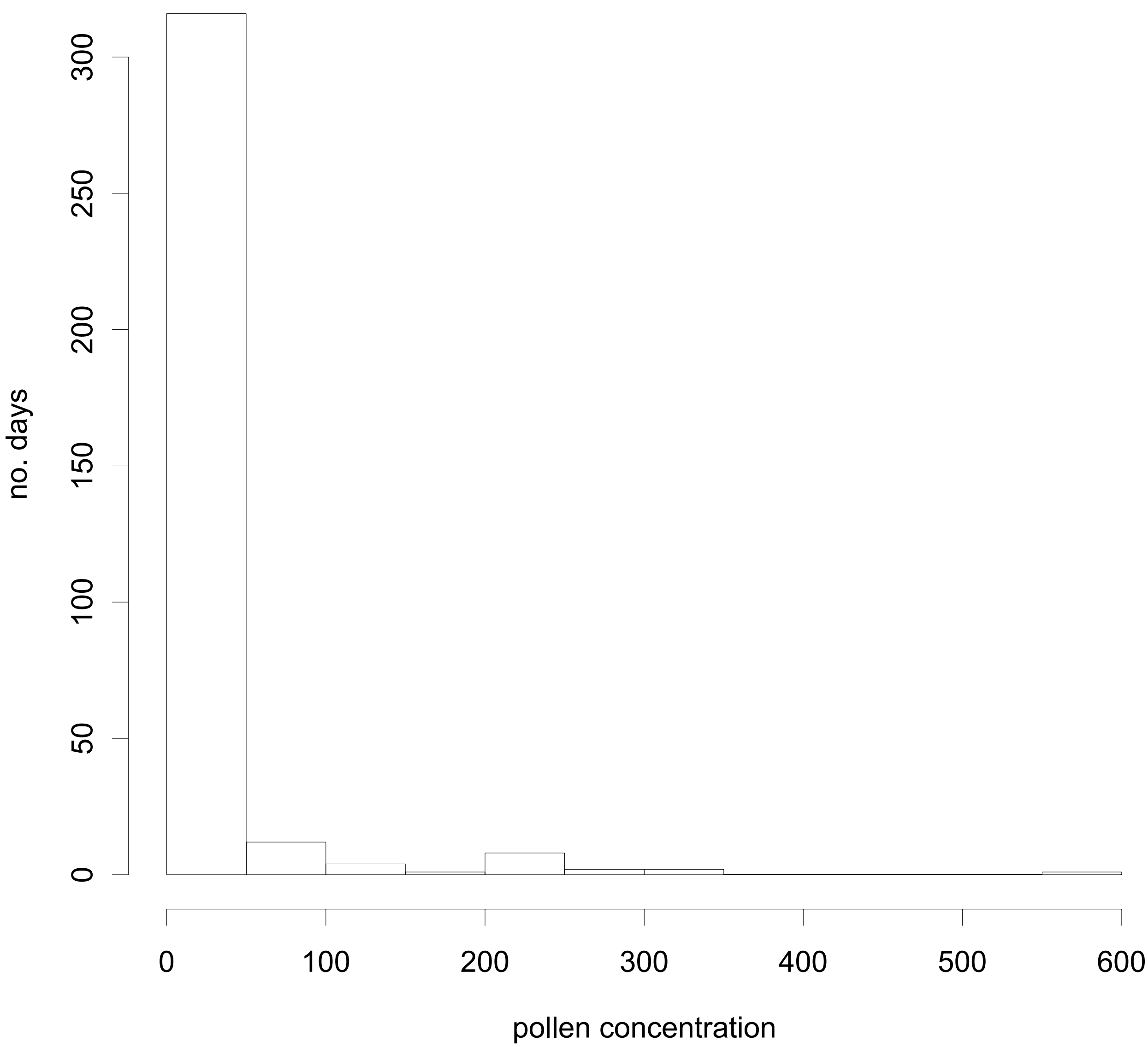


Figure 2. Cupressaceae (BCN/2002)



GAMMA PARAMETERS

- 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:

  $\star$ α is the shape parameter, no units, scale-free.

  $\star$ β is the scale parameter, same units as the concentration.

- Relationship with the mean and variance:

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

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

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

Figure 3. Lognormal distribution

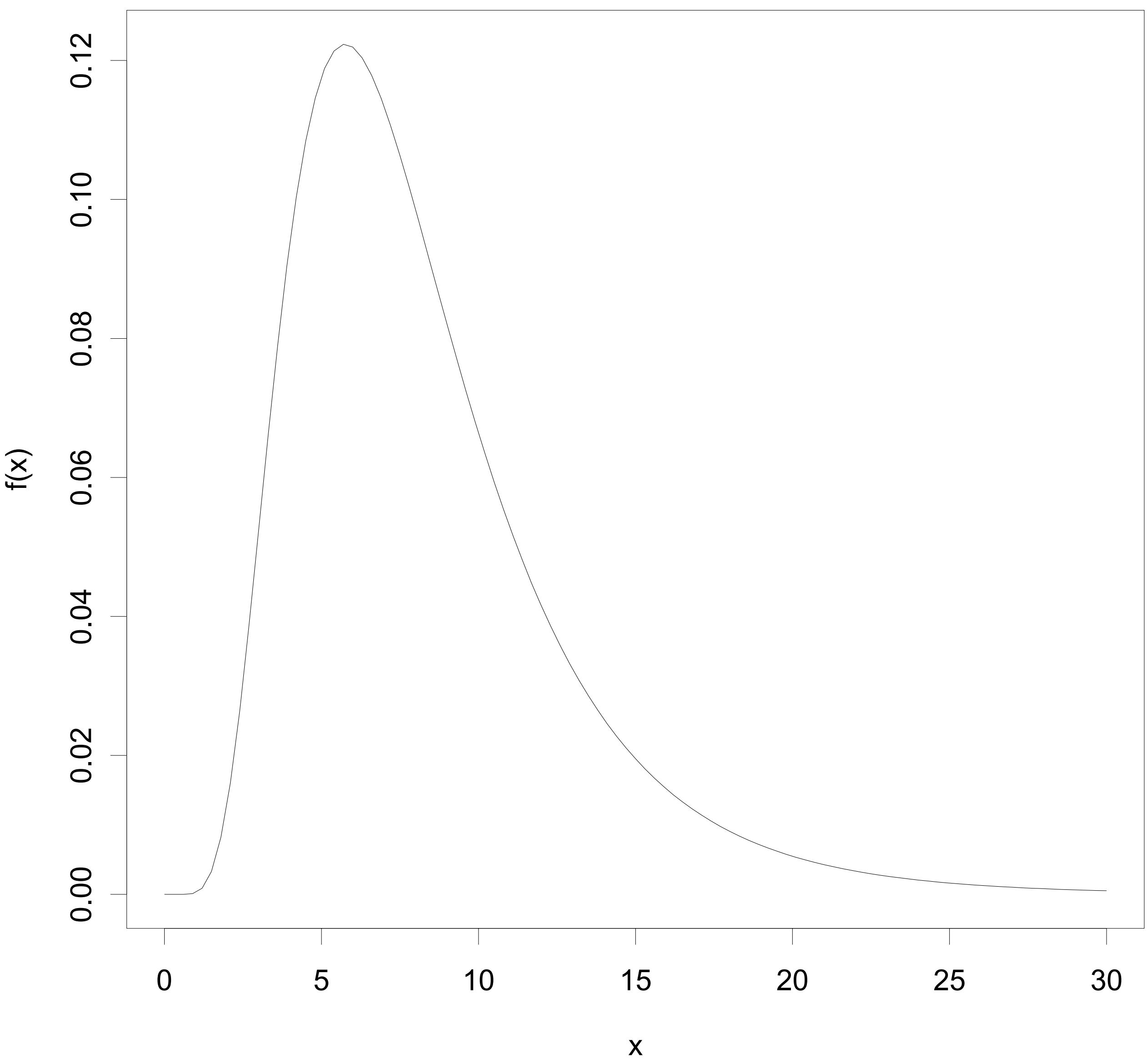


Figure 4. Gamma distribution, alpha < 1

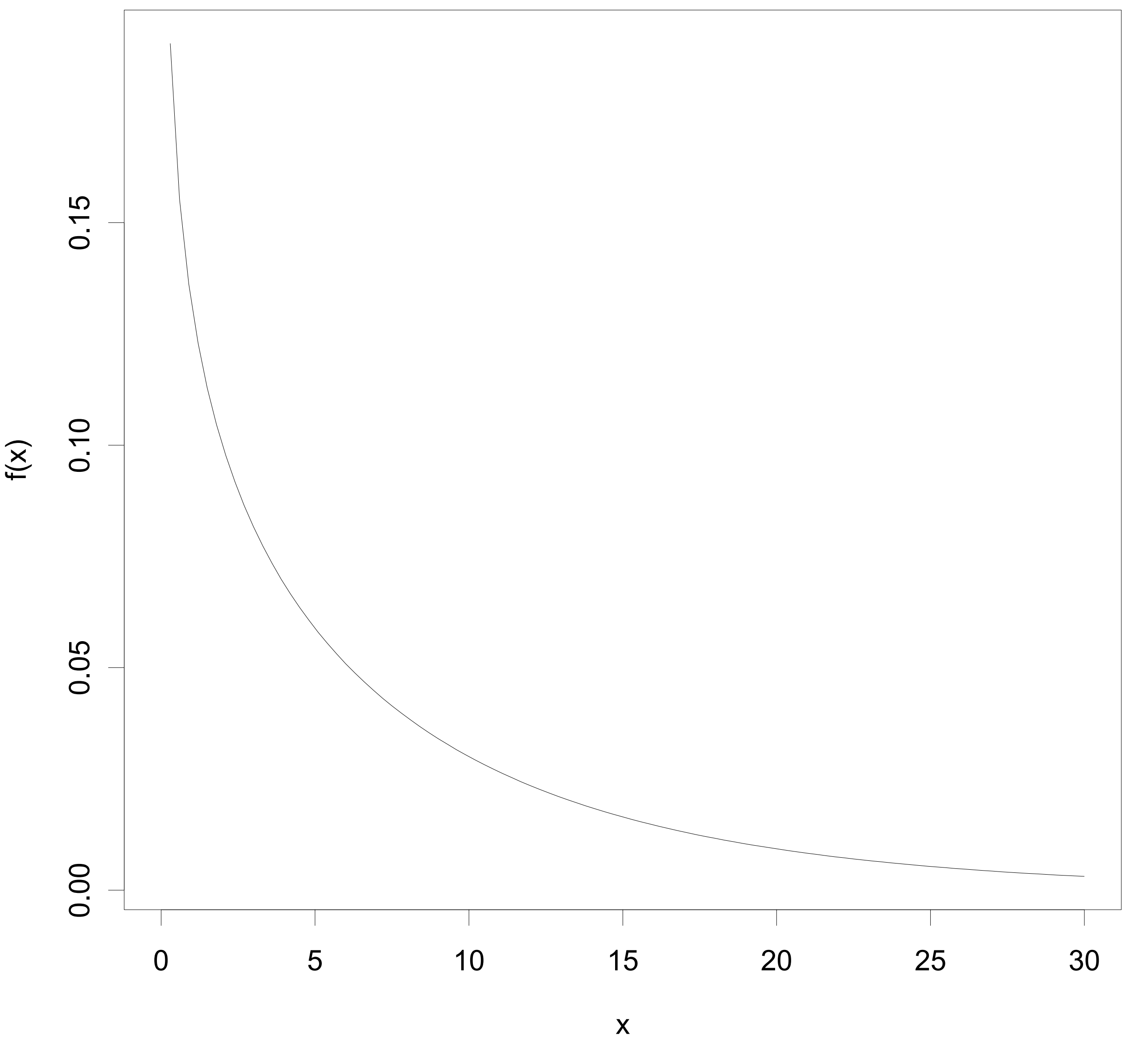


Figure 5. Gamma distribution, alpha = 1

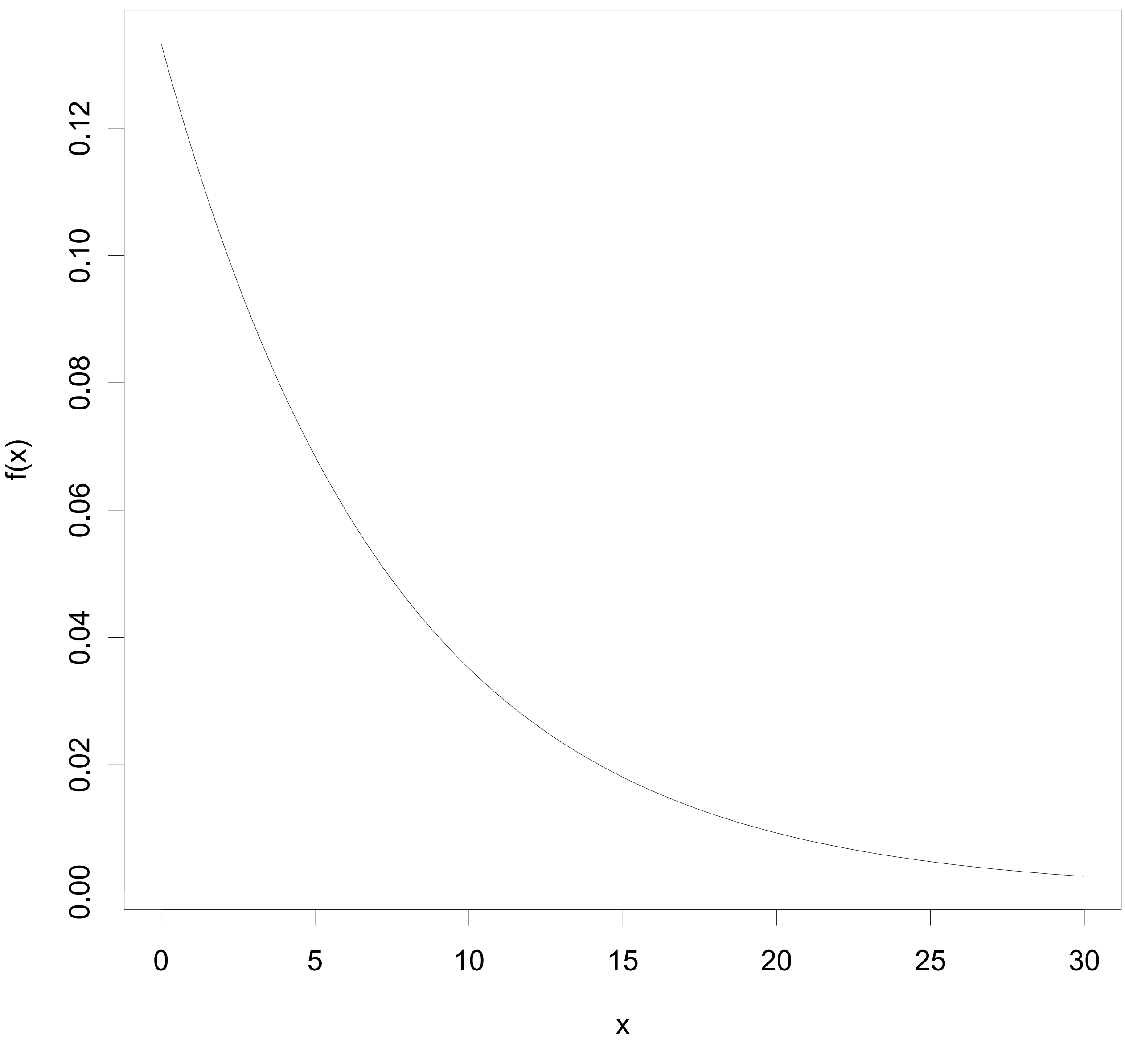


Figure 6. Gamma distribution, alpha > 1

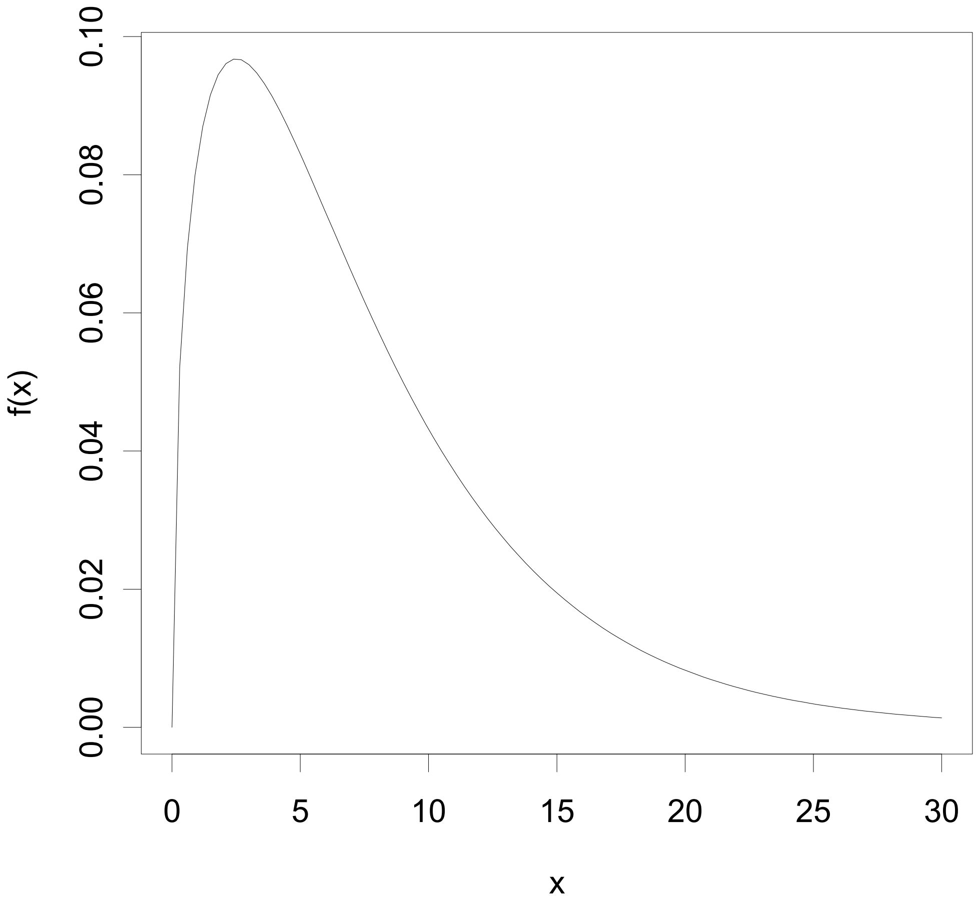


Figure 7. Cupressaceae (BCN/2002)

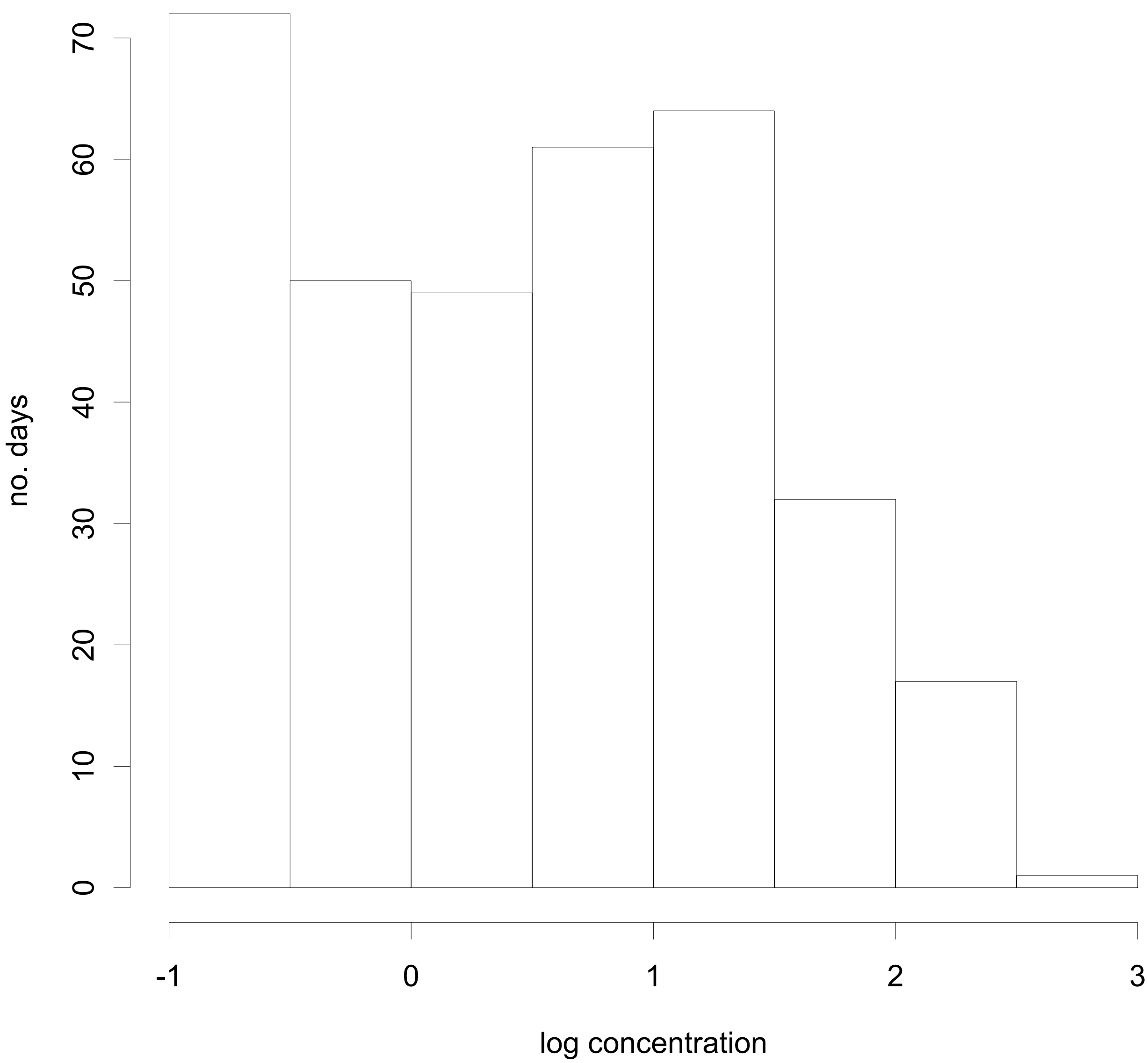
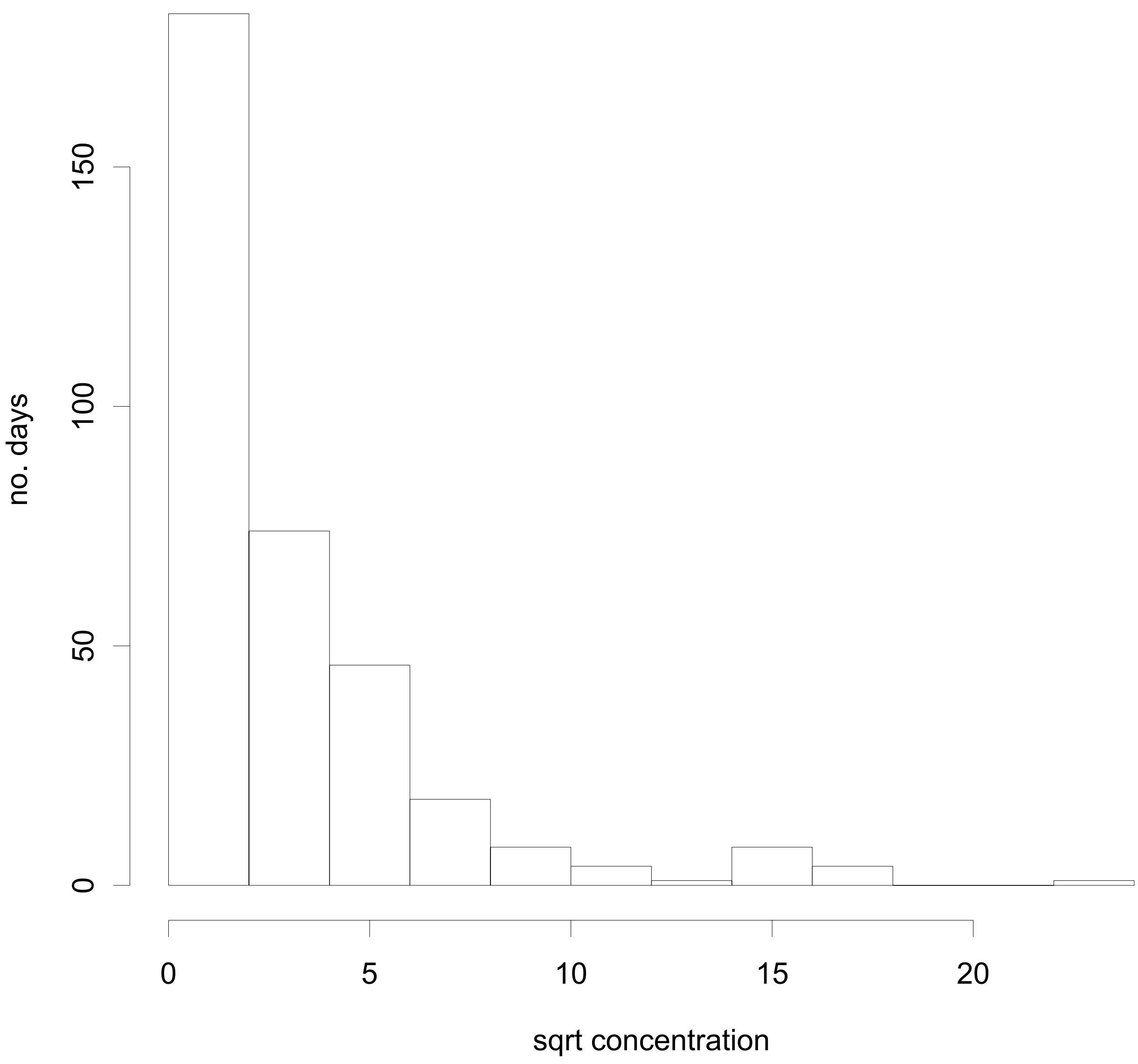


Figure 8. Cupressaceae (BCN/2002)



EXAMPLE

For the Cupressaceae mean daily pollen concentrations measured in the Barcelona station in 2002, the mean and the standard deviation are

$$\bar{x} = 22.09 \text{ pollen/m}^3, \qquad s = 57.33 \text{ pollen/m}^3,$$

and the coefficient of variation is $\mathrm{CV} = 2.595$.

Then the shape and scale parameters are

$$\alpha = \frac{(\bar{x})^2}{s^2} = \frac{1}{\mathrm{CV}^2} = 0.1485,$$

$$\beta = \frac{s^2}{\bar{x}} = 148.8 \text{ pollen/m}^3.$$

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*.

  It is always 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.

Figure 9. QQ plot for Cupressaceae (BCN/2002)

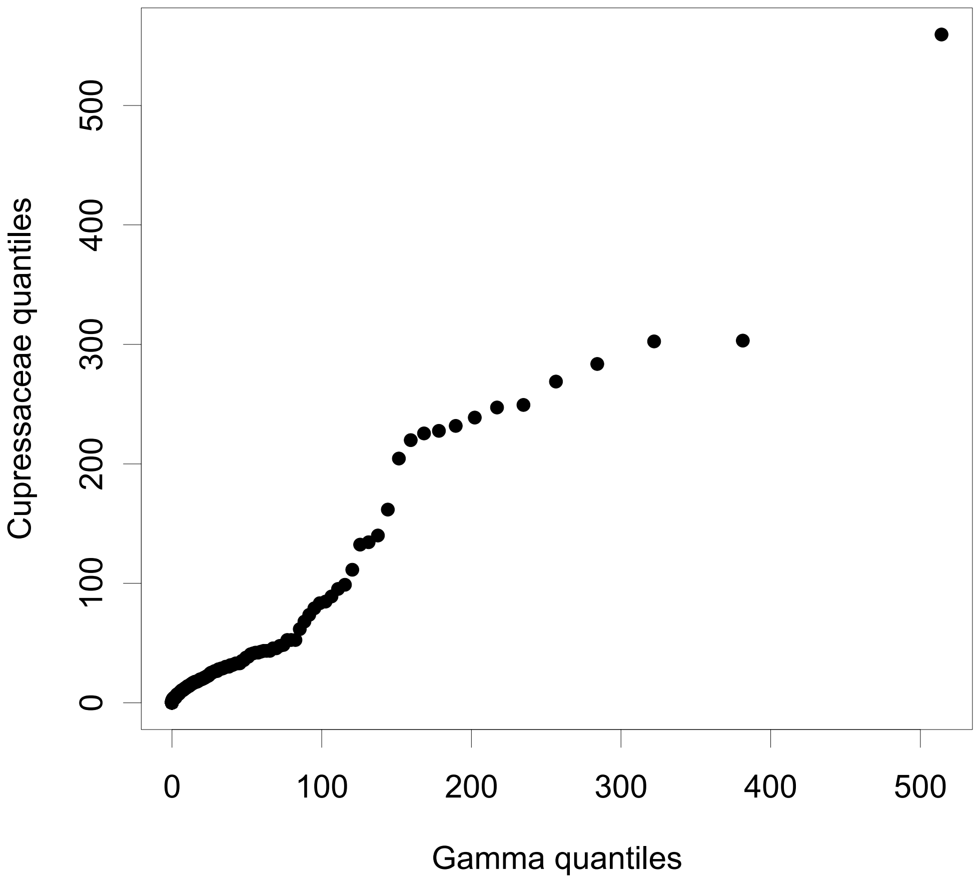
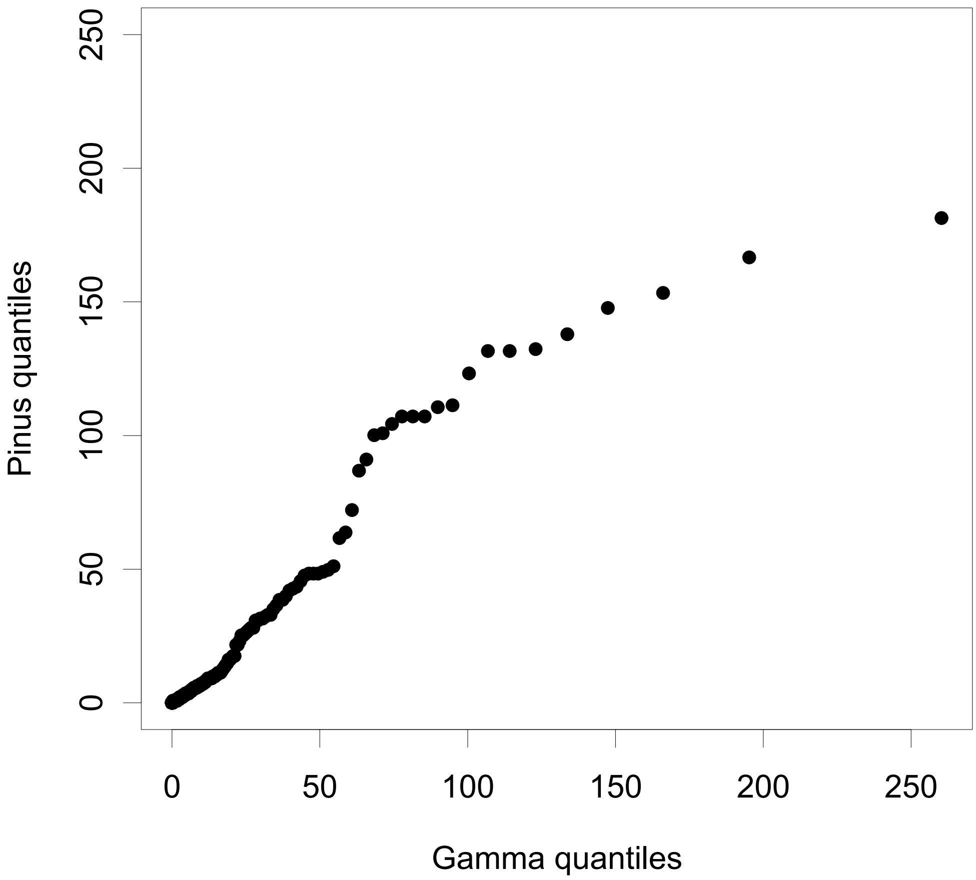


Figure 10. QQ plot for Pinus (BCN/2002)



CHI SQUARE STATISTICS

TABLE 1. Chi square testing for tree pollen types (Barcelona 2002)

Statistics	Cupressaceae	*Olea*	*Pinus*	*Platanus*	*Quercus*
<u>Observed</u>					
Null	59.0%	90.2%	71.7%	79.2%	77.4%
Low	19.4%	5.5%	12.4%	8.7%	8.1%
Medium	13.0%	2.3%	9.2%	2.0%	4.1%
High	3.5%	1.2%	1.7%	1.7%	4.1%
Very high	5.2%	0.9%	4.9%	8.4%	6.4%
<u>Expected</u>					
Null	64.4%	90.3%	67.4%	75.3%	75.6%
Low	13.7%	4.8%	15.9%	6.6%	8.4%
Medium	9.3%	2.7%	10.0%	4.5%	5.7%
High	6.1%	1.4%	4.8%	3.5%	4.0%
Very high	6.4%	0.9%	2.5%	10.1%	6.3%
Chi square	19.43	0.64	18.68	11.94	1.75

TABLE 2. Generic pollen types

Pollen type	Stations	No. days	Mean	Shape
Total pollen	All	345	110	0.40
Pinus	All	152	13.3	0.13
Quercus	All	131	16.7	0.12
Cupressaceae	All	200	19.2	0.11
Fraxinus	All	46	1.46	0.08
Platanus (1)	BCN	158	62.1	0.07
Populus	All	47	2.76	0.07
Olea	All	62	3.37	0.06–0.07
Palmae	BCN/TAU	56.5	0.563	0.06–0.07
Moraceae (1)	Except GIC	35	0.933	0.06
Ulmus	BCN/BTU/MAN/TAU	28	0.464	0.06
Salix	All	23	0.284	0.06
Moraceae (2)	GIC	62	25.7	0.05–0.06
Platanus (2)	Except BCN	53	6.53	0.05
Castanea	BCN/BTU/GIC/TAU	33.5	0.849	0.05
Alnus	All	32	0.752	0.05
Acer	LLE	24	0.625	0.04
Casuarina	BCN	13	0.165	0.04

REFERENCES

J. Belmonte, M. Canela & R.A. Guàrdia (2000). Comparison between categorical pollen data obtained by Hirst and Cour sampling methods. *Aerobiologia* **16**, 177–185.

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

J. Belmonte & M. Canela (2003). 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.

W.N. Venables & B.D. Ripley (2002). *Modern Applied Statistics with S*. Springer.