Mixed Models Approach to Estimate Time Series: Modeling the Environmental Effect on the Honeybee Egg Laying Activity

M. Salomé Cabral[1], M. Helena Gonçalves[2], P. Duarte-Rodrigues[3] and M. Manuela R. Branco[4]

[1] Centro de Estatística e Aplicações, Departamento de Estatística e Investigação Operacional, Faculdade de Ciências da Universidade de Lisboa, Portugal.
[2] Centro de Estatística e Aplicações da Universidade de Lisboa, Departamento de Matemática, FCT,Universidade do Algarve, Portugal.
[3] Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Portugal.
[4] Departamento de Engenharia Florestal, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Portugal.

Abstract: In many ecological studies the main interesting is to model the linkage between environmental conditions and population growth. When the data are collected at different locations or groups over time (longitudinal data) we may have observations taken at unequal intervals. In this situation autocorrelation coefficients can only be obtained directly by fitting a smoothed approximation to the time series. In this paper we purpose fitting time series using the mixed models approach. This method is illustrated using the results from longitudinal measurements of the broad area in honeybee colonies over a two-year period and the atmospheric conditions, temperature, humidity, rainfall and rain days. The second year was used to validate this model. The methodology is general and may be useful in many other applications.

Keywords: mixed models; longitudinal data; time series; cross-correlation function; auto-regressive model.

1 Introdução

In some communities the dynamic of the populations assume a well defined pattern; nevertheless the values achieved are not constant through the years. Environmental variability and population properties are important components in the environment-population interaction, which is the cause of those fluctuations. In such studies longitudinal data are collected. These data have more than one observation to each time in each location or group and are frequently unbalanced and the observations taken at unequal intervals. In this situation autocorrelation coefficients can only be obtained directly by fitting a smoothed approximation to the time series (Chatfield, 1992). In this paper we purpose a method for fitting the time series using the mixed models approach (Diggle et al., 1994; Pinheiro and Bates, 2000). These models incorporate the variability between subjects by means of the expression of the models parameters and in terms of both fixed and random effects and allows to model covariance structures that need to be considered. They have also the possibility to deal with unbalanced data. Based on the estimated time series that gives us the growth and population dynamics we can related it with environmental conditions and to built a prediction model. A real example is used to illustrate the methodology. All the analysis were performed in R 2.1.1 and the library *nlme* (Pinheiro and Bates) was used.

1.1 Data

For two years (1988/89, 1989/90) the queen egg laying of honeybees colonies (*Apis mellifera* L.) was sampled from 44 colonies, distributed by 4 apiaries (Cabração, Inferno, Labruja and Portela), located at the North of Portugal. The data comprised the area (sq. dm) of total brood (eggs, larvae and pupae) present in the colony, sampled every 21 days. This coincide to the total queen egg laying during the last 21 days, considering that honeybee immature stages last for 21 days since the time the egg is laid until the adult worker bee emerges. The data analyzed for the first and second years are unbalanced and consisted of observations divided into 17 time periods not equally spaced. For the first year the 7th time period is missing and for the second year the time periods from 7th to 10th are missing to apiaries Cabração, Inferno and Portela, and from 8th to 10th to apiary Labruja. To both years temperature, humidity, rainfall and rain days were collected.

1.2 Methods

To the data was fitted the mixed model

$$\begin{aligned} y_{ij} &= (\beta_0 + \gamma_{02}A_{2i} + \gamma_{03}A_{3i} + \gamma_{04}A_{4i} + b_{0i}) \\ &+ (\beta_1 + \gamma_{12}A_{2i} + \gamma_{13}A_{3i} + \gamma_{14}A_{4i} + b_{1i}) \times time_{ij} \\ &+ (\beta_2 + \gamma_{22}A_{2i} + \gamma_{23}A_{3i} + \gamma_{24}A_{4i}) \times time_{ij}^2 \\ &+ (\beta_3 + \gamma_{32}A_{2i} + \gamma_{33}A_{3i} + \gamma_{34}A_{4i}) \times time_{ij}^3 \\ &+ (\beta_4 + \gamma_{42}A_{2i} + \gamma_{43}A_{3i} + \gamma_{44}A_{4i}) \times time_{ij}^4 + e_{ij} \end{aligned} \quad (1)$$

$$\mathbf{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \cap N(\mathbf{0}, \mathbf{\Sigma}); e_i \cap N(\mathbf{0}, \sigma^2\Lambda_i)$$

where y_{ij} is the egg laying activity of the colony i at time j; A_{2i} to A_{4i} are binary indicator variables for apiaries Inferno, Labruja and Portela, respectively. $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 are, respectively, the intercept, linear, quadratic, cubic and quartic fixed effects averaged over the levels of A; γ_{li} is the egg-laying activity associated with the fixed effect $\beta_l, l = 0, \ldots, 4$; $\mathbf{b}_i$ is the vector of random effects and $\mathbf{e}_i$ is the within-group error vector with a spherical Gaussian distribution, where the scaled variance-covariance matrix Λ_i depends on i only through its dimensions. The estimated mean value of each apiary is a time series whose order of dependence was studied as well as its relation with the environmental time series (Chatfield, 1992). An auto-regression model, whose covariates were those selected by the cross-correlation function, was fitted to the data, using generalized least square. In both fittings graphical techniques were used to check for normality and independence of the residuals.

2 Results

The mixed model indicates that there is no differences between the apiaries Cabração and Inferno ($P = 0.1575$). Therefore, the colonies of these two apiaries were treated together; the common data is thereafter called (CI). The structure of the matrix Σ is diagonal ($P = 0.3595$). The within-groups errors were assumed to be heteroscedastic with variance model $var(\epsilon_{ij}) = \sigma^2 exp(2 \times 0.01805\hat{y}_{ij})$ and be correlated according to an auto-regressive process. Accurate estimated of velocity and acceleration was also performed. From the interpretation of the autocorrelation function between temperature and the several estimated time series we conclude that the series are related when delayed by 2 time periods. The same conclusion was obtained to humidity. No relations were found between rainfall and the several estimated growth curves as well as to rain days.
To the several apiaries of the first year was fitted the auto-regressive model

$$\begin{aligned} y_i(t) &= \beta_1 y_i(t-1) + \beta_2 temperature(t-2) + \\ &+ \beta_3 humidity(t-2) - 1 + e_i(t), \quad i = 1, 2, 3. \end{aligned} \tag{2}$$

where $y_i(t)$ is the estimated mean value of apiary i at time t, $temperature(t-2)$ and $humidity(t-2)$, are respectively the *temperature* and *humidity* at time $(t-2)$ for the first year and, $e_i(t)$ the random error with the usual normal conditions (Table 1). To all the apiaries the graphical analysis of the residuals leads us to assume the normality, independence and homoscedasticity. Data from the second year was used to validate the model and the four apiaries were considered separately (Table 2).

3 Discussion and Conclusion

3.1 The Dynamics of Honeybee egg laying Activity

In the first year the maximum estimated egg laying activity was reached in May, for the estimated velocity the minimum was observed in September

TABLE 1. R^2 for the first year for model (2).

Apiary	CI	Labruja	Portela
R^2	0.984	0.987	0.995

TABLE 2. R^2 for the second year to validate the model (2).

Apiary	Cbração	Inferno	Labruja	Portela
R^2	0.804	0.748	0.815	0.840

and the maximum in March. The maximum estimated acceleration was observed in December. Therefore, a delay of 63 days between the maximum velocity and the maximum egg laying and a delay of 160 days between the maximum acceleration and the maximum egg laying were observed. A biological interpretation for those delays is given and it is shown the practical relevance of our findings.

3.2 The Autoregressive Model

In Mediterranean climates drought stress is probably the major constraint to nectar production by plants. Therefore it seems most likely that the positive correlation with humidity and the negative correlation with temperature might reflect the influence of nectar supply in the queen egg laying activity. Studying the relation between the time series weight of each apiary and humidity we concluded that they are related when delayed by 2 time periods, which agrees with our hypothesis.

3.3 Conclusion

The use of mixed models gives us estimative of time series that allowed us to relate it to environmental variables. Based on those relationships prediction models can be built.

References

Chatfield, C. (1992). *The Analysis of Time Series. An Introduction.* 4th ed. Chapman and Hall. London.

Diggle, P.J., Liang, K-Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* Springer-Verlag. New York.