

Modelling spatial and temporal variability in a zero-inflated variable: The case of stone pine (*Pinus pinea* L.) cone production

Rafael Calama^{a,*}, Sven Mutke^a, José Tomé^b, Javier Gordo^c, Gregorio Montero^a, Margarida Tomé^b

^a Dpto. Silvicultura y Gestión Forestal, CIFOR – INIA: Ctra. Coruña km 7.5, 28040 Madrid, Spain

^b Dpto. de Engenharia Florestal, Instituto Superior de Agronomia, Tapada da Ajuda, 1349 – 017 Lisboa, Portugal

^c Servicio Territorial de Medio Ambiente, Junta de Castilla y León, Duque de la Victoria, 5, 47071 Valladolid, Spain

ARTICLE INFO

Article history:

Received 26 March 2010

Received in revised form 8 September 2010

Accepted 14 September 2010

Available online 9 October 2010

Keywords:

Zero-inflated

Seed production

Non-wood forest product

Cycling masting

Weather tracking

Resource depletion

ABSTRACT

Modelling *masting habit*, i.e. the spatial synchronized annual variability in fruit production, is a huge task due to two main circumstances: (1) the identification of main ecological factors controlling fruiting processes, and (2) the common departure of fruit data series from the main basic statistical assumptions of normality and independence. Stone pine (*Pinus pinea* L.) is one of the main species in the Mediterranean basin that is able to grow under hard limiting conditions (sandy soils and extreme continental climate), and typically defined as a masting species. Considering the high economical value associated with edible nut production, the masting habit of stone pine has been a main concern for the forest management of the species. In the present work we have used annual fruit data series from 740 stone pine trees measured during a 13 years period (1996–2008) in order: (a) to verify our main hypothesis pointing out to the existence of a weather control of the fruiting process in limiting environments, rather than resource depletion or endogenous inherent cycles; (b) to identify those site factors, stand attributes and climate events affecting specific traits involved in fruiting process; and (c) to construct a model for predicting spatial and temporal patterns of variability in stone pine cone production at different spatial extents as region, stand and tree. Given the nature of the data, the model has been formulated as zero-inflated log-normal, incorporating random components to carry out with the observed lack of independence. This model attains efficiencies close to 70–80% in predicting temporal and spatial variability at regional scale. Though efficiencies are reduced according to the spatial extent of the model, it leads to unbiased estimates and efficiencies over 35–50% when predicting annual yields at tree or stand scale, respectively. In this sense, the proposed model is a main tool for facilitating decision making in some management aspects such as the quantification of total amount of cones annually supplied to nut industry, design of cone harvest programs or the optimal application of seedling felling.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

At present, natural regeneration is one of the main topics of interest in forest ecology research. In this regard, the spatial and temporal dynamics of flowering, fruiting and seeding can be considered key to controlling natural recruitment of plant populations. Seed availability and abundance is also important for many species of granivore (Herrera et al., 1998), and subsequently for the maintenance of forest biological diversity. Edible fruits and seeds from some forest species can represent an important non-timber forest product (Scarascia-Mugnozza et al., 2000) for forest owners or local dwellers. Additionally, where fruit production is one of the main

objectives in forest management planning, adequate estimates of fruit production at spatial and temporal scales are often required.

The variability in fruit production at different spatial extents is largely related to site and stand conditions as well as management, and is subject to genetic control over processes involved in fruiting (Karlsson, 2000; Sirois, 2000; Calama et al., 2008). Although no consensus exists on the main factors controlling the synchronized annual variability in fruit production (*masting habit*), several theories have been postulated (Sork, 1993; Kelly, 1994; Herrera et al., 1998; Koenig and Knops, 2000; Kelly and Sork, 2002). Species-specific studies have determined proximate causes for mast seeding such as climatic factors (*weather tracking*), acting through resource-constraint to provide a synchronising cue (García et al., 2000; Sarker et al., 2003), or defining cycles of the population levels of insect pests and potential seed dispersers or predators (Koenig et al., 1994; Despland and Houle, 1997; García, 2001). This hypothesis is supported by the observed synchronisations (with

* Corresponding author. Tel.: +34 91 347 68 68; fax: +34 91 347 67 67.

E-mail address: rcalama@inia.es (R. Calama).

possible delays) between fruit production and different environmental conditions or biotic factors. According to this hypothesis, masting would be a more common habit in environments where the availability of resources is periodically limited rather than in non-limiting environments.

The reduction in the number of flowers induced after good crop years as a result of resource depletion was also proposed as another mechanism for masting (*resource depletion masting*) particularly in species with relatively high nutrient and carbon costs of fruit production (Davis, 1957; Sork et al., 1993; Tapper, 1996). Under this hypothesis, fruit production is more than a mere trade-off with available resources, and although some environmental control may exist, more complex processes are involved (Lalonde and Roitberg, 1992; Isagi et al., 1997). Finally, based on meta-database analyses and on theoretical models focusing on resource balance as well as on economies of scale considerations, e.g. wind pollination efficiency, predator swamping or animal disperser attraction, several authors have suggested that the cyclic masting patterns (*cycling masting*) may have evolved as demographic strategies over the plants' life-histories (Isagi et al., 1997; Herrera et al., 1998; García, 2001; Obeso, 2002). Some of the evidence for this evolutionary hypothesis is based on the occurrence of inherent cycles of fruit production (Sork et al., 1993). However, despite a large amount of empirical results supporting the various hypotheses, none of them are universally accepted today, the likelihood being that several of the processes proposed in the different theories work concomitantly (Woodward et al., 1994; Abrahamson and Layne, 2003).

Most of the previous studies on fruit production have focused on the analysis of the synchronized masting patterns at both large (Mencuccini et al., 1995; Koenig and Knops, 2000; Schaubert et al., 2002) or small (Tapper, 1996; Masaka and Sato, 2002) spatial extents and only a few purely predictive temporal models have been derived (Masaka and Sato, 2002; Abrahamson and Layne, 2003). Average fruit production has been modelled at tree, stand and regional extent for a number of different forest species (Masaka and Sato, 2002; Ihalainen et al., 2003; Mutke et al., 2005a); using tree, stand, site and management attributes as predictors, with little consideration being given to annual variability. Interannual variability is either not considered at all in these spatial models (since the response variables are either the average or the cumulative fruit production for a series of years (Calama et al., 2008)), or is simply considered by including a fixed or random year effect (Mencuccini et al., 1995; Calama and Montero, 2007; Miina et al., 2009). However, modelling both temporal and spatial variability for trees and stands can be of great value in forest management, particularly when studying processes such as natural regeneration or seed-predator population dynamics (Mutke et al., 2005a). To model masting at smaller spatial extents, long term fruit production data are generally required, which are rarely available and costly to obtain.

The species examined in this study, the Mediterranean stone pine (*Pinus pinea* L.), has generally been described as a masting species (Montero et al., 2008). Given the economical importance of the edible stone pine nuts, cone production in the species has been extensively studied. Long-term patterns of high interannual variability in cone production has been described for this species, with weather conditions and negative autocorrelations having been proposed as the main factors behind temporal variability in cone production at regional extent (Mutke et al., 2005a). Unlike most *Pinus* species, the stone pine fruiting process covers a four-year period between reproductive bud differentiation, cone opening and subsequent seed dispersal. Due to the length of the process, cone ripening in stone pine can be considered highly sensitive to external factors such as pests or extreme climatic events, which constitute an additional factor defining the specific patterns of interannual variability (Mutke et al., 2005b; Calama and Montero, 2007).

Several studies have focused on explaining and modelling spatial variability in cone and nut production at different spatial scales for different regions (Castellani, 1989; Cañadas, 2000; Nanos et al., 2003; Calama and Montero, 2007; Calama et al., 2008). The models developed have been integrated into existing growth and yield models for the species, such as PINEA2 (Calama et al., 2007). However, as tends to be the case with other forest species, none of the research carried out to date has attempted to describe, explain and model the masting pattern for the species.

The main objective of the present work is to describe, analyse and model the spatial and temporal patterns of variability in cone production in stone pine forests located in one of the most representative growing regions: Spain's Northern Plateau. We shall identify the factors controlling spatial and temporal variability in order to develop models which allow us to predict the annual cone production for a given tree or stand. Our key hypothesis is that weather conditions are the main determinant of the masting behaviour (*weather tracking hypothesis*), so special emphasis will be placed on testing this hypothesis and identifying the possible complementary role of other theories such as the resource depletion or cycling hypotheses. Given the nature of the data, it is important to examine the non-normal distribution, lack of independence among the data and the zero inflation of the observations.

2. Materials

2.1. Ecological conditions of the study region

Stone pine is the main forest tree species in Spain's Northern Plateau, a wide flat sedimentary area defined by the central basin of the river Duero and located at a mean altitude between 650 and 900 m a.s.l. Most of the 60,000 ha of stone pine forests in the region are pure, even-aged stands, which have been managed since the end of XIXth century for multifunctional purposes, namely; timber, firewood and pine nut production, with particular importance being given to soil protection to avoid erosion. In recent decades, other objectives such as landscaping and recreational uses have been taken into consideration. Within the above defined area, the pine forests are mainly distributed over a flat area covering about 80 km east-west by 40 km north-south (Fig. 1). From an ecological perspective, Calama et al. (2008) proposed an ecological stratification for the stone pine forests of the region (Fig. 1), dividing them into ten groups, denominated Natural Units (Table 1). These Natural Units are defined according to the geology and soil attributes (origin, main texture, water retention), as well as physiographic characteristics (altitude and position in the drainage basin) or microclimate differences.

The study area is a typical Mediterranean-Continental region with hot, dry summers and long, harsh winters. Average annual temperatures in the area range from 10.1 °C to 13.5 °C, with absolute values from −22 °C to 44 °C. Frosts may occur from September until May. Yearly rainfall ranges from 275 to 700 mm (average 460 mm) but with a very irregular distribution, not only between years but also from one season to another. The climate is homogeneous over this limited zone and in any case, no long-term weather data series were available for each natural unit. Hence, mean monthly temperature and precipitation data from the first-order meteorological station of the area (Valladolid, 04° 46' 27" W; 41° 38' 40" N; 735 m) were used.

2.2. Permanent plots for cone production

In the autumn of 1995, a network of 140 sample plots was installed in even-aged stone pine stands in the region (Fig. 1), in

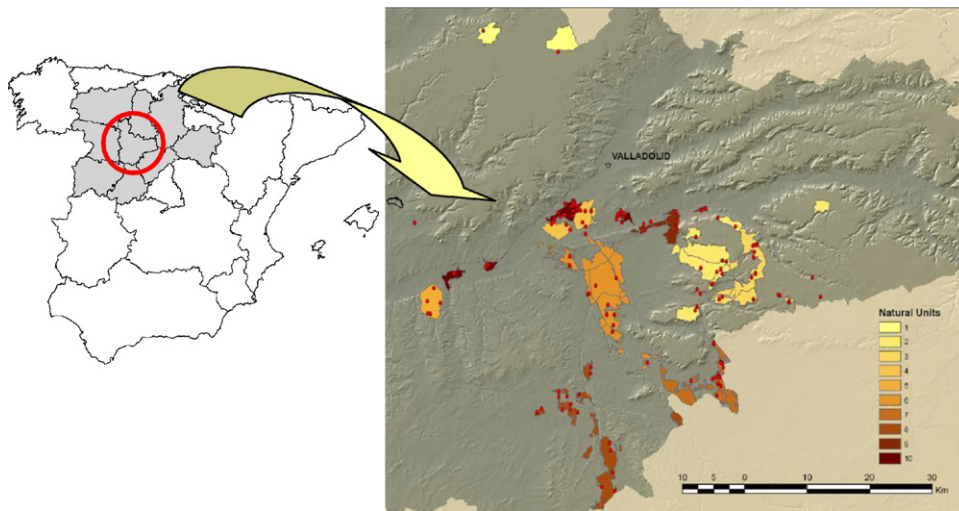


Fig. 1. Study region, natural unit's stratification and plots location (dots).

order to develop growth and yield models (for both timber and cones) for the species (see Calama et al., 2007, for more details of the current version of PINEA2). The plot selection attempted to provide a balanced representation of all the possible age, site quality and density classes identified in the region. The plots are circular, of variable size, and include 20 trees. Breast height diameter, crown diameter, total height and height to crown base were recorded for each tree at plot installation. Individual tree positions were georeferenced with respect to the centre of the plot. Plot age was estimated by averaging individual ages of three to five trees, computed after counting the rings on cores taken at stump height using a Pressler borer.

Between 1996 and 2005, cones from the five trees nearest to the centre of the plot were manually harvested every autumn (except in a few plots where cones were collected from the ten nearest trees). The cones cropped from each tree were classified as sound or damaged (those attacked by the larvae of the moth *Dioryctria mendacella* Stgr (Lepidoptera, Pyralidae) or the weevil *Pissodes validirostris* Gyll. (Coleoptera, Curculionidae)). Both fractions were counted and weighed, although the present study focuses exclusively on sound cone production. At the beginning of the trial in 1996, 741 trees were sampled annually. However, ten plots were lost completely as a result of fire or unauthorized logging. Twenty-eight trees in other plots died of natural causes, although the remaining trees in these plots were kept in the trial. By 2005, a total of 663 trees in 130 plots were still sampled. Since 2006, the cone production inventory has been maintained in a sample of 150 trees located in 30 of the remaining plots, selected in order

to cover the different Natural Units in the territory. The original data base was split into three data sets, one for fitting and two for validation. The fitting data set includes cone production sampled between 1996 and 2005 in 123 of the original plots, containing 656 trees and 6194 observations, while validation data set 1 includes data collected between 1996 and 2005 in the remaining 17 plots (85 trees and 804 observations), and validation data set 2 includes additional cone production data obtained between 2006 and 2008 (450 observations from 150 trees).

3. Methods

3.1. Response variable

Prediction of the tree annual production of sound cones was considered the main objective of the study. Annual cone yield per tree can be expressed either by the number of sound cones (nc) or by their fresh weight (wc). Both variables are highly related, therefore only the weight (wc) was used as a response variable, since this variable takes into account both the phenomena related to the initial processes of female floral induction and pollination (determining nc) and the cone growth, which mainly occurs during the third year of development and significantly depends on the amount of rainfall in that maturation year (Mutke et al., 2005b; Calama et al., 2007). Cone weight is a better indicator of the total amount of resources allocated to the reproductive effort, ensuring the natural regeneration of the stands and indirectly assuring the sustainability of the populations of seed predators. Finally, to evaluate the eco-

Table 1
Main environmental characteristics of the natural units.

Unit	Name	Total area (ha)	Altitude (m)	SI (m)	Geology and texture	Soil origin	Water retention (mm)	N° plots
1	Torozos	1834	847	14.0	Limestone – Marl	Tertiary sediments	>300	2
2	Limestone plain W	4757	845	14.9	Limestone – Marl	Tertiary sediments	200–300	17
3	Limestone plain E	2284	854	14.4	Limestone – Marl	Tertiary sediments	200–300	18
4	Valladolid	1908	687	13.1	Quartz sands	Wind deposits	100	17
5	Nava del Rey	1480	710	11.8	Quartzitic gravels	Wind deposits	100	6
6	Viana de Cega	7211	707	15.1	Quartz sands	Wind deposits	100–150	12
7	Iscar	2945	745	16.7	Quartz sands – clays	Alluvial – Wind	>300	19
8	Medina	4044	746	12.5	Quartz sands	Wind deposits	75–150	17
9	Tudela de Duero	1401	712	12.6	Quartz sands	Wind deposits	150	10
10	River terraces	1664	688	16.8	Alluvial meadow soils	Alluvial deposits	150–300	22

Where SI is site index, as stated by Calama et al. (2003); total area is uniquely referred to public owned forests.

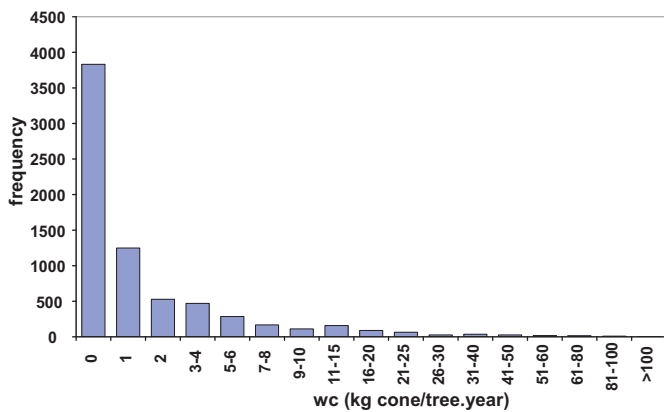


Fig. 2. Histogram of observed frequencies of annual cone production of single trees ($\text{kg tree}^{-1} \text{ year}^{-1}$) including all the observations from the three data sets.

conomic importance of stone pine nut production it is also necessary to express the data in terms of weight. However, the distribution of frequencies for both the wc and nc variables did not fulfill the standard normality assumption, displaying:

- Asymmetry: empirical distribution is significantly skewed towards the higher values of the variable, with a massive number of observations showing smaller values of cone production, and only a small number of trees in a few years giving very large crops and forming a long tail to the right (Fig. 2).
- Zero inflation: the distribution displays a strong mode at zero (corresponding to null production by sampled trees), comprising 54.49% of the observations in the fitting data set (Table 2).
- Truncation: given the nature of the response variable, negative values are not possible.

Furthermore, the hierarchical structure of the data (repeated observations from trees nested in sample plots within Natural Units) implies a lack of independence among observations, which prevented us from using estimation methods based on ordinary least squares minimization.

3.2. Testing for cyclic production and crop depression

Our main working hypothesis is that weather conditions control the fruiting processes rather than cyclical masting patterns or depressions in subsequent crops due to depletion following bumper crops. In order to test for the existence of cyclical patterns or depressions in cone production, we selected 371 trees (from the original set of 741) which had a series of observations from more than six consecutive years of cone production and at least five years with non-null values to assure reasonably long series without interruptions or an excess of zero observations. The number of observations per series ranges from seven to thirteen.

Cyclical patterns (*cycling masting hypothesis*) were statistically analysed by calculating the Spearman rank correlation coefficients for the observed 1–4 year lagged responses of each of the 371 series, as proposed by Greene and Johnson (2004). If a cyclical pattern exists, a larger than expected quantity of significant positive (or negative) correlations would be detected for any of the evaluated lag periods, whereas in the absence of cyclical behaviour, a spurious correlation would lead to an expected value of 10 significant positive correlations (approximately 2.5% of 371) for each analysed lag period.

The possible depression in crops during the years following a bumper production year (evidence of the *resource depletion*

hypotheses) was checked by examining the effect of the biggest annual production of each series on the crop production over the three following years (Greene and Johnson, 2004). To do this, we first assigned an ordinal value i for every observation of each series (considering $i = 2$ for the second biggest annual production in the series s , $i = 3$ for the third biggest annual production, etc.). Under the null hypothesis of no depression effect, the expected value (averaging all the series) for the ordinal value i in a subsequent year, lagged 1, 2 or 3 years after the bumper production year, should be close to the expected median value $\Sigma [0.5N_s + 1]/s$, where N_s is the total number of observations in the series and s is the total number of series. In this respect, if there is no depression effect, approximately half of the 1, 2 and 3 year lagged observations from the bumper production year of each series should be larger than the median value for all the data. To test the hypothesis that there is a lag in the depression effect of one, two and three years respectively, the series with biggest annual production occurring in the last, second last and third last observation in the series were not taken into consideration. Hence, the corresponding numbers of available series in these cases were, 352, 348 and 311.

If neither cycles nor clear patterns of resource depletion are detected, the identification of weather variables driving the inter-annual production variability would support the hypothesis of environmental control over masting habit (*weather tracking*).

3.3. Modelling approach

3.3.1. Zero-inflated log-normal distribution

Although it is possible to approach normality in the case of a skewed distribution of mean fruit production data following a log-transformation and a translation by including a small constant to replace any zero data (Masaka and Sato, 2002; Ihalainen et al., 2003; Mutke et al., 2005a), this method is not viable for zero-inflated observed distributions (Mayer et al., 2005; Li et al., 2008). Other approaches for dealing with skewness and zero inflation are based on more flexible, Poisson mixture distributions (Affleck, 2006), such as the negative binomial and the generalized Poisson, although these generally tend to underestimate the number of zeros. A third approach is provided by the zero-inflated models (Lambert, 1992), which deal with the zero observations and the positive outcomes separately. In this approach, distributional particularities are dealt with by combining a distribution which explains the binary nature of absence or occurrence of zeros, with a second distribution which models the response variable, conditioned by its occurrence, with a non-null value. In this case, the phenomenon under examination is considered the result of two different processes; the first with a binary outcome, the occurrence of the event (in our case, fruiting) and the second, determining the intensity of the event (in our case, total weight of the tree cones), conditioned by the occurrence of the event.

The binary part is commonly modelled using a logistic function assuming a binomial distribution. A large range of distributions is available for modelling the intensity of the event. For discrete count data, the more common distributions for the component of the mixture are either Poisson or Negative binomial, leading to the Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models, which have attracted much attention in forest and ecological literature in recent years (Hall, 2000; Rathbun and Fei, 2006; Fortin and DeBlois, 2007). A complete review of zero-inflated models for count data can be found in Welsh et al. (1996) or Ridout et al. (1998). In contrast, much less attention has been paid to modelling continuous data with an excess of zeros. In this case, the log-normal distribution is widely used, since it avoids zero-truncation (meaning that zero responses can only occur in the binomial part of the model) and offers greater potential for

Table 2
Main annual cone production statistics for (2.1) fitting data set (123 plots, covering 1996–2005); (2.2) validation data set 1 (17 plots, covering 1996–2005) and (2.3) validation data set 2 (30 plots, covering 2006–2008).

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	Total
(2.1)											
Number of plots	123	123	120	120	115	118	118	115	116	115	
Number of trees	656	655	631	630	606	620	613	593	599	591	6194
Number of trees with wc=0	212	267	314	427	185	406	446	397	497	228	3379
% trees with wc=0	32.0	40.7	49.7	67.8	30.5	65.5	72.5	66.6	82.8	38.6	54.5
$\overline{wc}$ (kg tree ⁻¹ year ⁻¹)	6.2	3.4	1.4	0.7	5.5	2.8	1.0	0.9	0.2	1.7	2.4
Max (wc) (kg tree ⁻¹ year ⁻¹)	135.0	99.8	49.5	28.5	92.5	89.5	49.8	40.8	5.6	55.0	135.0
STD (wc) (kg tree ⁻¹ year ⁻¹)	13.0	8.9	4.2	2.3	11.3	9.4	4.3	3.3	0.6	4.5	7.7
$\overline{nc}$	22.4	13.4	4.1	2.7	17.7	9.1	4.5	2.9	0.6	8.5	8.7
Max (nc)	473	297	178	134	238	300	172	135	16	205	473
STD (nc)	44.2	33.2	13.2	10.0	32.6	28.4	16.5	10.0	2.0	19.2	25.6
(2.2)											
Number of plots	17	17	17	17	17	17	17	16	15	15	
Number of trees	85	85	84	83	83	82	79	78	73	72	804
Number of trees with wc=0	28	33	37	53	24	44	51	51	57	14	392
% trees.wc=0	32.9	38.8	44.0	63.9	28.9	53.7	64.6	65.4	78.1	19.4	48.8
$\overline{wc}$ (kg tree ⁻¹ year ⁻¹)	7.4	3.9	1.8	1.1	4.0	3.1	0.6	0.4	0.1	2.5	2.6
Max (wc) (kg tree ⁻¹ year ⁻¹)	68.5	53.9	30.8	20.5	48.9	31.1	7.6	4.9	2.4	34.0	68.5
STD (wc) (kg tree ⁻¹ year ⁻¹)	12.5	7.3	4.3	3.6	7.1	6.5	1.4	0.9	0.4	4.8	6.5
$\overline{nc}$	28.4	17.1	5.3	4.2	14.5	11.5	3.4	1.5	0.6	13.9	10.2
Max (nc)	234	215	112	88	161	97	36	20	8	202	234
STD (nc)	49.2	32.4	14.3	13.6	26.3	22.1	7.5	3.4	1.5	27.5	25.8
(2.3)											
Year	2006			2007			2008				
Number of plots	30			30			30				
Number of trees	150			150			150				
Number of trees with wc=0	45			119			91				
% trees with wc=0	31.0			79.3			58.7				
$\overline{wc}$ (kg tree ⁻¹ year ⁻¹)	4.9			1.2			1.5				
Max (wc) (kg tree ⁻¹ year ⁻¹)	87.7			26.7			24.4				
STD (wc) (kg tree ⁻¹ year ⁻¹)	11.4			4.4			3.6				
$\overline{nc}$	20.2			5.2			4.9				
Max (nc)	325			106			89				
STD (nc)	42.9			18.0			12.1				

$\overline{wc}$ and $\overline{nc}$ are the mean weight and the mean number number of healthy cones *per tree* and *per year*, max(wc) and max(nc) the maximum values of weight and number of healthy cones *per tree* and *per year*, STD(wc) and STD(nc) standard deviation of the weight and the number of cones *per tree* and *per year*.

modelling highly skewed distributions (Tu, 2002). The so-called Zero-Inflated Log-Normal mixture distribution (hereafter referred to as ZILN; p.d.f. in Eq. (1)) was first introduced in statistical literature by Aitchinson (1955), who defined it as *delta distribution*, and although not widely used, it has been applied to modelling phenomena such as fishery catch rates (Mayer et al., 2005), cattle production (Belasco and Ghosh, 2008) or air contaminant levels (Tu, 2002).

$$P(Y = y | \mathbf{x}, \mathbf{z}) = \begin{cases} \theta & \text{if } y = 0 \\ (1 - \theta)f(y) & \text{if } y > 0 \end{cases} \quad (1)$$

where θ represents the probability of the occurrence of a non-event (zero fruiting) and $f(y)$ indicates the probability density function for a log-normal distribution with expected value μ and variance σ^2 ; $\mathbf{x}$ and $\mathbf{z}$ represent vectors of known possible explanatory covariates for modelling θ and μ , respectively; $\ln$ represents the natural logarithm. Regression models can then be parameterized as:

$$\logit \theta = \ln \left(\frac{\theta}{1 - \theta} \right) = \mathbf{x}\alpha \quad (2a)$$

$$\mu = E[\ln(y)|y > 0] = \mathbf{z}\beta \quad (2b)$$

where α and β represent vectors of unknown but estimable parameters. Simultaneous estimation of both models is attained by maximising the overall likelihood function L , which is given by the product of the two independent components of the likelihood

(Welsh et al., 1996; Li et al., 2008):

$$L(\alpha, \beta, \sigma^2) = \prod_{i=1}^n \left[\frac{1}{1 + \exp(\mathbf{x}_i\alpha)} \right]^{1-r} \times \left[\left(\frac{\exp(\mathbf{x}_i\alpha)}{1 + \exp(\mathbf{x}_i\alpha)} \right) \frac{1}{y_i\sigma\sqrt{2\pi}} \exp \left(-\frac{1}{2\sigma^2} (\ln(y_i) - \mathbf{z}_i\beta)^2 \right) \right]^r \quad (3)$$

with r representing a dummy binary variable, whose value equals 0 if the observation y_i is null and 1 if it is positive and n is the number of observations.

3.3.2. Mixed modelling in ZILN

The hierarchical structure of the data defines different levels of correlation among observations from the same tree, plot or production year that are taken into account by introducing correlated random effects into the intercepts of the linear predictors for both the binary and positive outcomes of the zero-inflated model (Hall, 2000; Tooze et al., 2002). Assuming these Eqs. (2a) and (2b) should be re-expressed as:

$$\logit \theta = \ln \left(\frac{\theta}{1 - \theta} \right) = \mathbf{x}\alpha + u \quad (4a)$$

$$\mu = E[\ln(y)|y > 0] = \mathbf{z}\beta + v \quad (4b)$$

where u and v denote the random components acting at the proposed level of variability (tree, plot or year) for the logistic and the log-normal parts of the ZILN model. Random intercepts for the

logistic and the lognormal part of the model corresponding to measurement unit i , u_i and v_i , were assumed to follow a bivariate normal (BVN) distribution:

$$(u_i, v_i)^T \sim \text{BVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \right]$$

According to the above considerations, the overall likelihood function for a y_{ij} observation under a ZILN mixed model with m levels of random variability and n_i observations measured at the i th level is given by the following expression (Tooze et al., 2002; Li et al., 2008):

$$L[\alpha, \beta, \sigma^2, \sigma_u^2, \sigma_v^2, \rho] = \prod_{i=1}^m \int_{u_i} \int_{v_i} \prod_{j=1}^{n_i} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}\alpha + u_i)} \right)^{1-r_{ij}} \\ \times \left(\frac{\exp(\mathbf{x}_{ij}\alpha + u_i)}{1 + \exp(\mathbf{x}_{ij}\alpha + u_i)} \right)^{r_{ij}} \left[\left(\frac{1}{y_{ij}\sigma\sqrt{2\pi}} \right) \exp \left(-\frac{1}{2\sigma^2} (\ln(y_{ij}) - (\mathbf{z}_{ij}\beta + v_i))^2 \right) \right]^{r_{ij}} \\ \times \left(\frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho^2}} \right) \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{u_i^2}{\sigma_u^2} + \frac{v_i^2}{\sigma_v^2} - \frac{2\rho u_i v_i}{\sigma_u\sigma_v} \right) \right] du_i dv_i \quad (5)$$

The objective of the maximization is to obtain simultaneous estimates for the vectors of parameters α and β , associated with explanatory covariates for both logistic and log-normal linear predictors, and for the variance components σ_u^2 , σ_v^2 , ρ and σ^2 . Maximization was carried out using quasi-Newton optimization approximated through adaptive Gaussian quadrature. This method is implemented in the SAS procedure NLMIXED, and model fitting was carried using a routine based on the MIXCORR macro by Tooze et al. (2002). Given the complexity of Eq. (5), we decided to compare the three possible alternatives of additional level of random variability (plot, tree or year), and then select the best in terms of Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and minus twice log-likelihood ($-2LL$). Plot $\times$ year iteration term was also evaluated as a random source of variability, but after detecting severe problems with the convergence of the model and parameter estimation we decided not to consider it.

3.3.3. Covariate selection: two-step mixed model

The explanatory covariates may or may not be common to both the occurrence and intensity models. In modelling stone pine cone production, we evaluated different groups of possible explanatory variables:

- Tree size: basal area at breast height (g) and diameter at breast height (d), total height (h) and crown ratio (cr). Crown width was available, but it was not evaluated since this covariate is rarely measured in forest inventories.
- Stand attributes: number of stems per ha (N), basal area (G), Stand Density Index (SDI), stand age (T), quadratic mean diameter (dg) and dominant height ($hdom$)
- Distance-independent competition indices: basal area of trees larger than the subject tree (BAL), d/dg , g/G
- Site: site index (SI ; Calama et al., 2003) and Natural Unit classification (see Table 1)

The above variables are assumed to account for spatial variability in cone production and can be considered constant over periods of at least five years. Temporal variability in cone production was explained by evaluating different weather attributes over the course of the study period: monthly rainfall, average, maximum and minimum monthly temperatures and number of days with frost. Given the three-year duration of cone development, the weather variables up to four years prior to cone maturation were evaluated.

The selection of covariates was carried out by first, independently fitting the binomial regression model (Eq. (4a)) to the dichotomized data for fruiting occurrence and then fitting the log-normal model (Eq. (4b)) using only the non-null intensity data (weight of cones), as proposed by Heilbron (1994) and Woollons (1998). Independent fitting of these generalized linear models can be accomplished using maximum likelihood estimation methods. Information criteria such as $-2LL$ and AIC were used to define the best independent model for each component. In a subsequent step, simultaneous fitting using ZILN was carried out with these pre-selected covariates, testing the level of significance of the parameters and removing those which were non-significant.

3.4. Evaluation of the model

3.4.1. Predictive ability

The accuracy of the selected model was checked against both the evaluation and the validation data sets. Two alternatives were considered to predict the production of cones from a tree using the fitted model:

- If the value predicted by the occurrence part of the model was greater than the cut-off value of 0.46 (proportion of fruiting trees in the fitting dataset), then the production of cones was predicted using the intensity part of the model, otherwise the predicted production was zero
- The production of cones from a tree equals the product of the expected probability of occurrence, as estimated using the logistic part, and the expected value of cone production estimated by the log-normal part. In this case, no zero values are predicted.

Approaches (a) and (b) were compared using the fitting data set and validation data set 1, considering the mean error (E), root mean squared error (RMSE) and modelling efficiency (EF) (Eqs. (6a)–(6c)):

$$E = \frac{\sum_{ij} (y_{ij} - \hat{y}_{ij})}{n} \quad (6a)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{ij} (y_{ij} - \hat{y}_{ij})^2}{n - 1}} \quad (6b)$$

$$\text{EF} = 1 - \left[\frac{\sum_{ij} (y_{ij} - \hat{y}_{ij})^2}{\sum_{ij} (y_{ij} - \bar{y})^2} \right] \quad (6c)$$

where n represents the total number of observations, y_{ij} and $\hat{y}_{ij}$ are the observed and predicted values of cone production for observation j of unit i , and $\bar{y}$ is the mean value for the response variable. The evaluations were done using the estimated marginal mean (in real scale of $\text{kg tree}^{-1} \text{ year}^{-1}$) as the predicted values for the estimated nonzero intensities, assuming $v_i = E(v_i) = 0$ and the following antilogarithmic transformation:

$$E(y_{ij}|y_{ij} > 0) = \exp \left[\mathbf{z}_{ij}\beta + \frac{\sigma_v^2}{2} + \frac{\sigma^2}{2} \right] \quad (7)$$

In the case of approach (a), specificity (rate of correctly classified non-events) and sensitivity (rate of correctly classified events) were also evaluated. The models were evaluated at tree and plot extent on an annual basis and histograms of observed and predicted values were used to evaluate the ability of the model to predict null or very low and very large values.

3.4.2. Predicting pure temporal variability

The proposed models were also evaluated in terms of their accuracy in predicting the pattern of interannual variability in cone production at regional extent. For this purpose, comparisons

between the estimated and predicted average annual values of tree cone production were carried out using pairwise *t*-testing together with graphical analysis. The comparisons were undertaken using validation data sets 1 and 2, including observations from the period 2006 to 2008, with no observations from the original fitting data set.

3.4.3. Predicting pure spatial variability

The capacity of the model for explaining pure spatial variability in cone production was checked by evaluating, in terms of the statistics *E*, RMSE and EF, the ability of the model to predict the total sum of cones produced by each tree or plot in the fitting data set and the validation data set 1 over the studied ten-year period 1996–2005, thus dampening the effect of interannual variability.

4. Results

4.1. Cone production measured in permanent sample plots

In the studied period, the annual mean cone yield varied from 0.2 to 6.3 kg cones per tree (0.6–22.4 cones per tree). This represents a 45-fold difference between the best and the worst year, which can be even larger at individual tree level. The maximum registered value in terms of weight for a tree in a single year is 135 kg, corresponding to a total of 473 cones. Although 54% of the 7448 observations were null, only 9% of the trees never produced cones during the whole study period. The percentage of null yields per year varied from 32% to 82%. Table 2 shows the main cone production statistics for both the fitting and validation data sets. As regards the distribution of frequencies for observed annual values of cone production at tree level (Fig. 2), it is possible to detect a clear pattern of zero abundance and asymmetry, with pronounced right-tail behaviour indicating that the main part of the production is limited, even in good years, to a few trees.

4.2. Testing for cyclic production and crop depression

Using Spearman's rank correlation coefficient to check for significant temporal correlation of the selected 371 cone production tree series, we identified 10, 3, 3 and 10 significant positive correlations for lags of 1, 2, 3 and 4 years, respectively. Since the expected number of significant but spurious positive correlations for each lag (2.5% of the cases) would be around 10, even if records were uncorrelated, it can be concluded that our results reveal a non-cyclic pattern, at least for periods up to 4 years.

As regards a depression in future crops following a year in which the crop is particularly large, the average median value for the ordered response in the analysed series was 6.3, while the average values for the observed ordered response $E(i)$ for the first, second and third year following the large crop were 4.9, 6.1 and 7.0, respectively. In 27%, 48% and 63% of the series, the production values for the first, second and third year after the large crop, respectively, were greater than the median value (expected value in the absence of depression would be 50%). These results point to the existence of a three-year delayed depression in crop production after a bumper crop year. However, the most significant finding is that there is a high positive relationship between the best observed production value in each series and the crop of the following year. In fact, in almost 20% of the analysed series, the second best year in the series is that which immediately follows the best production year, while in another 20%, the best production year is followed by the third best year, and in 16% of the occasions, it is followed by the fourth best (with random draws, the expected values would always be around 9%). Possible reasons for this phenomenon will be discussed below.

4.3. Selection of covariates and random effects

After separately fitting the logistic (Eq. (4a)) and the log-normal parts (Eq. (4b)), the following covariates were selected:

Logistic equation:

- Stand:
 - Log(N): natural logarithm of number of stems per ha
 - T20: categorical variable equal to 1 if stand age is <20 years, 0 otherwise
 - T50: categorical variable equal to 1 if stand age is 20–50 years, 0 otherwise
- Tree:
 - d/dg : breast height diameter/mean square diameter (dimensionless)
- Site:
 - SI: Site Index (m)
 - Natural unit (categorical variable)
- Climate:
 - Pp.my.jn.3: precipitation (mm) during May–June, three years before cone maturity
 - Pp.oc.nv.3: precipitation (mm) during October–November, three years before cone maturity
 - Nhel: number of days with severe frost (minimum temperature < −5 °C) during the winter period (December to March) of the first year after flowering

Log-normal equation:

- Stand:
 - Log(N): natural logarithm of number of stems per ha
- Tree:
 - d : breast height diameter
 - d/dg : breast height diameter/mean square diameter (dimensionless)
- Site:
 - SI: Site Index (m)
 - Natural unit (categorical)
- Climate:
 - Pp.my.jn.3: precipitation (mm) during May–June, three years before cone maturity
 - Pp.oc.nv.3: precipitation (mm) during October–November, three years before cone maturity
 - Pp.summ.2: precipitation (mm) during the summer period – July to September – just after flowering (two years before maturation)
 - Pp.feb.my.0: precipitation (mm) during the period from February to May, the year of cone maturity
 - Nhel: number of days with severe frost (minimum temperature < −5 °C) during the winter period (December to March) of the first year after flowering

As can be seen, most of the covariates are common to both parts of the model, indicating that the occurrence and abundance of fruiting are mainly governed by the same factors. The selected covariates were then used to fit the different alternatives of the ZILN mixed model, considering random intercepts at plot, tree or year level (Table 3). The results show that the best alternative in terms of likelihood was to consider random intercepts acting at plot level for both the logistic and log-normal components of the model.

Table 4 includes the estimates for the components of vectors of parameters α and β associated with fixed effects covariates x and z , together with the variance components σ_u^2 , σ_v^2 , ρ (indicating random variability at plot level) and σ^2 (variance for the error in the log-normal element).

Table 3

Fitting statistics and information criteria for the different alternatives of random structure (random plot, tree or year) for the ZILN mixed model.

Random level	No	Tree	Plot	Year
–2LL	15296	14951	14684	15199
AIC	15372	15033	14766	15281
BIC	15628	15216	14882	15293
N° parameters	38	41	41	41
$p < \text{LRT}$	–	<0.0001	<0.0001	<0.0001

Where –2LL: minus twice logarithm of likelihood function; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion; $p < \text{LRT}$ indicates the probability for a likelihood ratio test among evaluated alternative and the alternative not including random components.

Most of the common covariate parameters in both parts of the model share the same sign, indicating that the effect of the variables over the analysed processes of occurrence and abundance is similar. The main exceptions are age, which only influences the occurrence processes and rain events after flowering, which only affect the total final weight of cones. To carry with the detected nonlinear relation between tree maturity and fruiting event probability, we decided to enter age into the model as a categorized variable. The coefficient of correlation between random effects in the two parts of the model is highly significant (Wald test, p -value < 0.0001) and positive (0.6364), indicating that the unobserved factors acting at plot level are also the same for both processes. These results again suggest that the same factors govern the processes of fruiting occurrence and abundance.

4.4. Model evaluation

4.4.1. Predictive ability of the model at annual tree production level

Table 5 presents the values of the statistics evaluating the predictive ability of the model for predicting annual tree production, using the fitting and the validation data set 1. Using the fitting data set, approach (a) with a fixed cut-off, tended to overestimate the average value, leading to biased estimates, while approach (b),

Table 4

Fixed parameters and variance components estimates for the selected ZILN mixed model including correlated random parameters at plot level.

	Covariate	Logistic part (α)	Log normal Part (β)
Climate covariates	Intercept	–1.7062	–2.8930
	pp.my.jn.3	0.0131	0.0089
	pp.oc.nv.3	0.0111	0.0055
	pp.summ.2	–	0.0030
	pp.feb.my.0	–	0.0036
	Nhel	–0.1003	–0.0425
Natural unit	1	2.1916	0.7427
	2	0	0
	3	–0.3720	–0.1823
	4	–1.8417	–0.7850
	5	–1.3570	–0.6053
	6	–1.4608	–0.3152
	7	–0.7816	–0.1937
	8	–1.7267	–0.4149
	9	–0.7981	–0.6504
	10	–0.4319	0.2744
Stand covariates	Log(N)	–0.5664	–0.2673
	SI	0.0754	0.0421
	$T < 20$	–2.7243	–
	$20 < T < 50$	–0.3298	–
Tree covariates	d/dg	2.2421	0.5895
	d	–	0.0454
Random components	$\sigma_u^2(\text{plot})$	0.6490	–
	$\sigma_e^2(\text{plot})$	–	0.2142
	$\sigma^2(\text{residual})$	–	1.0608
	ρ (correlation term)	0.6364	–

Table 5

Ability of the model for predicting annual production of cones at tree level over fitting data set and validation data set 1. The ZILN approaches used were “a” (fixed cut-off = 0.46 for defining fruiting-non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production).

Approach	Fitting data set (n = 6194)		Validation data set (1) (n = 804)	
	(a)	(b)	(a)	(b)
E (kg tree ^{–1} year ^{–1})	–0.339	0.062	0.361	0.669
p-value (t-test)	<0.0001	0.4254	0.0599	0.0006
wc.pred (kg tree ^{–1} year ^{–1})	2.760	2.359	2.213	1.913
wc.obs (kg tree ^{–1} year ^{–1})	2.422	–	2.574	–
RMSE (kg tree ^{–1} year ^{–1})	6.263	6.176	5.447	5.498
EF (%)	33.44%	35.27%	28.71%	27.28%
Sensitivity (%)	66.64%	–	60.92%	–
Specificity (%)	74.96%	–	80.61%	–
Events/no events obs	2815/3379	–	412/392	–
Events/no events pred	2722/3472	–	477/327	–

Where E: mean error, wc.pred and wc.obs: predicted and observed mean cone production (kg tree^{–1} year^{–1}); RMSE: root mean squared error; EF: model efficiency; sensitivity: rate of correctly classified events; specificity: rate of correctly classified non-events.

weighting the estimated production with the estimated probability of the event, lead to unbiased estimates, with smaller RMSE and larger EF. In the case of the validation data set, both approaches led to underestimation of the model, with the best and less biased results attained using approach (a). Nevertheless, the results in terms of EF and RMSE are quite similar for both the validation and the fitting data sets.

With respect to the capacity for correctly classifying the observations as events or non-events, approach (a) correctly classifies 75% of the non-events (trees with null production in a given year) in the case of the fitting data and 80% for the validation data set, although the sensitivity (percentage of events correctly classified) drops to 67% for the fitting data set and to 61% for the validation data set. As approach (b) is not capable of estimating a null production, no values were obtained for specificity or sensitivity.

Table 5 shows the specificity (percentage of zeros correctly classified) and the sensitivity (percentage of events correctly classified), together with the ratio of the number of predicted events/number of non-events for prediction approach (a). As regards the closeness of the model predictions to the real distribution of observations, Fig. 3 presents the histograms for observed and predicted frequencies in annual cone production for both the fitting and validation data sets. As in approach (b) no zeros are predicted, the first two weight classes were merged into a single observation class < 2 kg tree^{–1} year^{–1}. Approach (a) tends to underestimate the frequency in the lower weight classes, while approach (b) more accurately predicts the observed frequency values, especially for classes below 4 kg tree^{–1} year^{–1}.

4.4.2. Predictive ability of the model at annual plot production level

From a practical management point of view, it is far more important to be able to determine annual cone production for a given plot or stand than for a single tree. Table 6 presents the values of the statistics used to evaluate the predictive ability of the models for both the fitting and validation data set 1 at plot level (considering the total weight of cones produced in a plot within a given year). Once more, approach (a) tends to overestimate cone production for the fitting data set, although in this case, bias is less important than

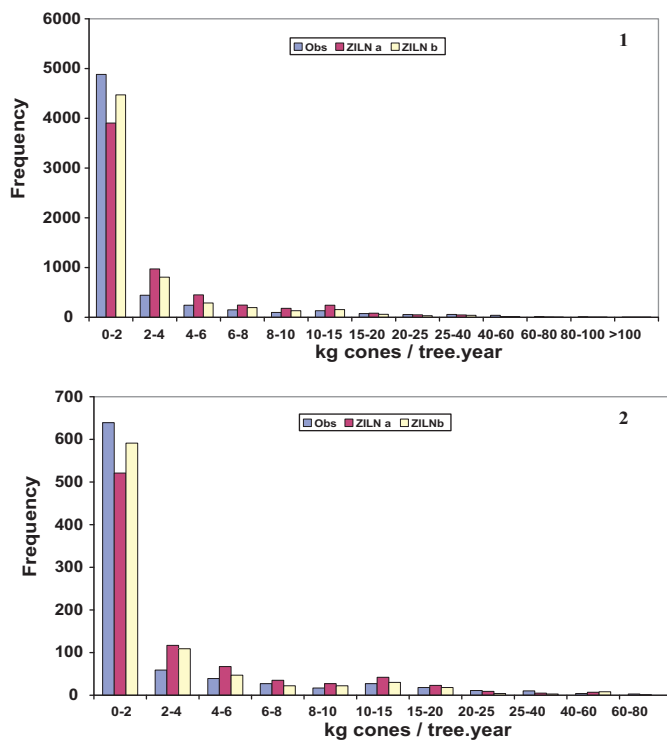


Fig. 3. Histograms of observed and predicted frequencies using ZILN approaches “a” (fixed cut-off = 0.46 for defining fruiting–non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production) for single tree annual cone production ($\text{kg tree}^{-1} \text{ year}^{-1}$) for fitting (3.1) and validation (3.2) data set.

at the tree scale. With regard to the validation data set, the best results are again attained using approach (a). Regarding RMSE and EF, both approaches lead to similar results, attaining efficiency values close to 50% for the fitting data set and 38% for the validation data set.

4.4.3. Predicting pure temporal variability

Table 7 and Fig. 4 show the pattern of predicted and observed interannual variability in mean cone production for the fitting and validation data sets 1 and 2. The model mimics the global pattern of interannual variability at regional extent, being able to correctly discern good and bad crop years in both the fitting and validation data set, and even correctly predicting those observations in the three years not included in the original data set, especially the 2006 bumper crop. The model greatly

overestimates production in 2000 and underestimates cone production in 1997 and 2001, for both the fitting and validation data set. The proposed model explains about 72–81% of the interannual variability in the fitting data set and 61% in the validation data set. The latter result is mainly due to large prediction residuals for 2000 and 2001; years which display a somewhat unusual pattern in the validation data set.

4.4.4. Predicting pure spatial variability

The spatial accuracy of the model was evaluated by comparing the amount of variability among plots (Table 8) and among trees (Table 9) in accumulated cone production explained by the model with the corresponding observed variability. The model explains 75% of the variability in ten years accumulated cone production detected among plots (fitting data set), reaching 80% in the validation data set. As regards the variability among trees, the model explains about 50% of detected variability in the fitting data set, reaching 54–63% in the validation data set. However, again there were no large differences detected between approaches (a) and (b), although as in the cases described above, approach (b) tends to better explain the fitting data set while approach (a) performs better with the validation data set.

5. Discussion

5.1. Mast seeding

The annual production of stone pine cones at tree level showed a skewed and zero-inflated distribution, together with a large interannual variability, the regional average ranging between years from 0.2 to 6.3 kg tree^{-1} . This variation was largely synchronized among the trees and stands throughout the study area, confirming the masting habit for the species. Apparently, climatic factors are the main determinants of the masting behaviour, explaining about 60–80% of the temporal variation in cone production at regional scale in the fitting and validation data sets for a series of 13 years, and correctly discriminating between good and bad crop years. The weather covariates entering the model are in close concordance with main phenological stages of the flowering and fruiting processes, such as floral bud induction (spring precipitation year –3 before maturation) and bud differentiation (fall precipitation year –3); cone setting (extreme drought or frost during the summer and winter periods of year –2) and final cone growth and ripening (precipitation during maturation year) (Mutke et al., 2005a). Evidence of weather control over fruit production has mainly been identified in less productive and limiting environments (Despland and Houle, 1997; Koenig et al., 1996; Sarker et al., 2003; Kelly and Sork, 2002; Abrahamson and Layne, 2003), such as those associated with Mediterranean forests, indicating that lack of water, light or adequate temperatures at crucial stages, influences reproductive success.

Our analyses of the masting pattern suggest that there are no fixed cyclic patterns of annual production, at least up to a four-year period. There is no evidence from these results pointing to cycling masting as an evolved reproductive strategy in stone pine, as has been postulated for some species (e.g. Sork et al., 1993). According to our data, stone pine cone production, though basically governed by weather factors, was also influenced by the occurrence of bumper crop years, which appear to lead to lower production over the following three years. Taking into account that the female floral primordia are initiated in late spring of the year prior to pollination, just at the time when the ripening and main growth of the bumper crop is occurring, it is likely that this flower formation will be inhibited, either due to resource depletion as a result of the allocation of resources to the larger than usual crop (Sork

Table 6

Ability of the model for predicting annual production of cones at plot level over fitting data set and validation data set 1. The ZILN approaches used were “a” (fixed cut-off = 0.46 for defining fruiting–non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production).

Approach	Fitting data set (n = 1180)		Validation data set 1 (n = 165)	
	(a)	(b)	(a)	(b)
E (kg)	–1.775	0.327	1.760	3.221
p-value (t-test)	0.0449	0.7143	0.3694	0.1046
wc_pred ($\text{kg plot}^{-1} \text{ year}^{-1}$)	14.455	12.354	10.786	9.325
wc_obs ($\text{kg plot}^{-1} \text{ year}^{-1}$)	12.681		12.546	
RMSE ($\text{kg plot}^{-1} \text{ year}^{-1}$)	30.466	30.770	25.181	25.552
EF (%)	51.34%	50.36%	38.80%	36.99%

Where E: mean error, wc_pred and wc_obs: predicted and observed mean cone production ($\text{kg tree}^{-1} \text{ year}^{-1}$); RMSE: root mean squared error; EF: model efficiency.

Table 7

Observed and predicted values and level of significance of the pair wise difference test for the average annual cone yield per tree for both fitting and validation data sets 1 and 2. The ZILN approaches used were “a” (fixed cut-off = 0.46 for defining fruiting-non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production).

Year	Fitting data set (6194 obs)			Validation data sets 1 (804 obs) and 2 (150 obs)		
	Observed (kg tree ⁻¹)	ZILN a	ZILN b	Observed (kg tree ⁻¹)	ZILN a	ZILN b
1996	6.199	6.485*	5.550*	7.438	5.664*	4.686***
1997	3.402	2.385	1.872	3.937	1.971***	1.466
1998	1.435	1.994***	1.550*	1.809	1.584*	1.172*
1999	0.656	1.186	1.033	1.109	0.695**	0.720*
2000	5.489	8.330	7.258	3.994	7.148	5.988**
2001	2.772	2.087**	1.850***	3.147	0.996	1.261***
2002	1.039	1.219*	1.026*	0.632	0.808	0.748*
2003	0.985	1.437**	1.286*	0.413	0.791*	0.993
2004	0.154	0.139*	0.345	0.129	0.000	0.249***
2005	1.736	2.121**	1.643*	2.469	1.933	1.420**
2006				4.959	6.276*	5.257*
2007				1.205	1.458*	1.308*
2008				1.526	2.832	2.170**
Explained interannual variability (EF %)		72.33%	81.70%		62.25%	61.14%

No significant differences between observed and predicted with p -value: * $p > 0.05$ level ** $p > 0.01$ *** $p > 0.001$.

et al., 1993; Woodward et al., 1994; Piovesan and Adams, 2001) or because of a direct hormonal inhibition caused by the ripening seeds (Lee, 1979). This explanation is also supported at shoot level, where both vegetative shoot vigour (length, number of lateral shoot buds) and reproductive effort (number of female conelets initiated) in a given year suffer from both insufficient water availability and the number of mature cones ripened during bud formation (Mutke et al., 2005c).

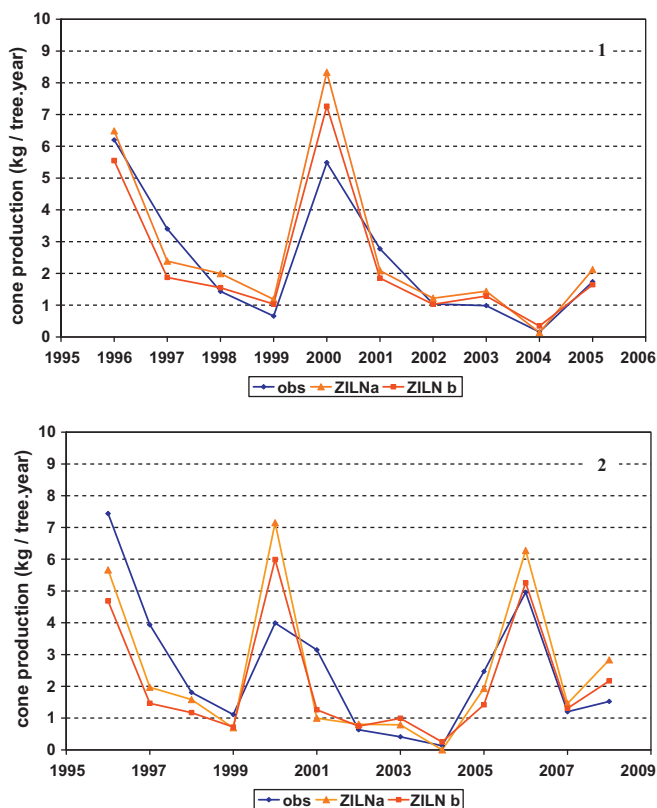


Fig. 4. Observed and expected regional average cone production (kg tree⁻¹ year⁻¹) using ZILN approaches “a” (fixed cut-off = 0.46 for defining fruiting-non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production) for fitting (4.1) and validation (4.2) data set.

Table 8

Ability of the ZILN mixed model for predicting spatial variability among plots, evaluating ten years accumulated cone production (kg plot⁻¹ 10 years) over fitting and validation data set. The ZILN approaches used were “a” (fixed cut-off = 0.46 for defining fruiting-non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production).

ZILN approach	Fitting data set (n = 123)		Validation data set (1) (n = 17)	
	(a)	(b)	(a)	(b)
E (kg plot ⁻¹)	17.072	3.150	17.082	31.256
p-value	0.078	0.4012	0.1106	0.0316
RMSE (kg plot ⁻¹)	133.8	139.5	56.5	72.2
EF (%)	75.83%	74.84%	88.27%	80.85%

Where E: mean error; p-value: level of significance associated with E; RMSE: root mean squared error; EF: model efficiency.

In contrast to the simpler version of the resource depletion theory (in which the demands on resources of the bumper crop leaves the tree with insufficient stocks to support a normal crop in the following year), we detected a positive correlation between a very large cone production and the following year's crop, with 73% of the studied series producing crops which were larger than the expected median value in the year immediately following a bumper crop. This positive correlation indicates that the exogenous factors promoting a large crop in a given year – e.g. spring rainfall – also favour the crop of the following year. In terms of resource tracking, the effect of an environmentally favourable year on cone production can last several years, for example, by increasing leaf area or photosynthetic efficiency (Woodward et al., 1994).

Table 9

Ability of the ZILN mixed model for predicting spatial variability among trees, evaluating ten years accumulated cone production (kg tree⁻¹ 10 years) over fitting and validation data set. The ZILN approaches used were “a” (fixed cut-off = 0.46 for defining fruiting-non fruiting) and “b” (the production of cones equals the product of the expected probability of occurrence and the expected value of cone production).

ZILN approach	Fitting data set (n = 656)		Validation data set (1) (n = 85)	
	(a)	(b)	(a)	(b)
E (kg tree ⁻¹)	3.201	0.596	3.416	6.251
p-value	0.0054	0.3144	0.0511	0.0030
RMSE (kg tree ⁻¹)	32.3	31.3	19.4	21.4
EF (%)	48.39%	51.44%	63.64%	54.39%

Where E: mean error; p-value: level of significance associated with E; RMSE: root mean squared error; EF: model efficiency.

These positive one-year (resource induction) or negative three-year (resource depletion) delayed effects of bumper crops cannot be easily incorporated into our proposed multi-level model. Nevertheless, they might explain a good deal of the interannual variation not explained by the model: bumper crops were identified in 1996 or 2000, thus significant underestimates of production the subsequent years (1997 and 2001) and small overestimates in 1999 and 2003 may be due to the delayed effect of those crops. On the other hand, the significant overestimates by the model for the year 2000 can be related to the extremely wet spring (165 mm) and autumn (325 mm) in 1997, and might indicate that an upper threshold can be reached in the strictly increasing relationship proposed between rainfall and flower induction in those years where the water availability is not the main limiting factor. Finally, those small overestimates detected in years such as 2003 or 2008, may also be related to nonlinear responses to consecutive extreme frost days (December 2001, with 13 days of minimum temperatures below -5°C), or extremely dry summers (2005, with only 7 mm of rainfall between July and August).

5.2. Modelling approach

Despite numerous attempts at characterizing, identifying and describing the main causes of masting in forest species, little attention has been paid to modelling the interannual variability of forest fruit production, and less still to modelling this variability simultaneously at various hierarchical levels. Although a number of studies have focused on the development of theoretical models for masting, which do not rely on real data (e.g. Lalonde and Roitberg, 1992; Isagi et al., 1997), the present study constitutes, as far as we know, one of the few attempts to jointly model spatial and temporal variability in fruit production at tree scale for a forest tree species. Two circumstances made this possible: Firstly, the availability of a long data series (up to 13 years) for cone production from an extensive network of plots and trees (140 plots and 740 trees) covering all the possible classes of stand density, age and site quality across a large area. Several previous efforts at modelling fruit production failed due to insufficient availability of temporal statistical series, preventing the accurate identification of the factors governing inter-annual variability (Karlsson, 2000; Van der Wall, 2002). In other cases, in spite of having long series, the research was hampered by the fact that data came from a limited number of trees under similar environmental conditions (Eis, 1973, 1976; Koenig et al., 1994; Tapper, 1996). The second factor facilitating the modelling attempt described in this paper is the use of zero-inflated log-normal mixed models, which allows us to avoid the violation of basic statistical assumptions derived from zero abundance, non-normality and inherent correlation among observations, factors which have been identified as the main impediments to modelling annual fruit production (Ihalainen et al., 2003; Calama et al., 2008).

Traditional approaches based on logarithmic transformation after adding a constant value (e.g. Ihalainen et al., 2003; Mutke et al., 2005a; Calama and Montero, 2007) are not suitable in this case since the presence of a secondary mode at the zero transformed value does not satisfy the assumption of normality. Compared with other possible statistical approaches, based on Mixture-Poisson distributions, such as the negative binomial or the generalized Poisson (Miina et al., 2009), zero-inflated distributions allow a more accurate estimation of zero events (Welsh et al., 1996) and the consideration of fruiting as two different processes (presence/abundance), even assuming different distributional assumptions for the zero non-events and the non-zero events (Fortin and DeBlois, 2007). Finally, if compared with two-step modelling procedures (Woollons, 1998), zero-inflated modelling allows a simultaneous and correlated estimation for the parameters explaining both processes.

5.3. Model structure and spatial variability

The proposed final structure of the model incorporates 41 parameters, including four variance components, two intercepts for both the logistic and the log-normal parts of the model, and thirty-five parameters associated with fixed effects (seventeen in the logistic part and eighteen in the lognormal part) including continuous and categorical variables. Despite this large number, most of the covariates are common to both parts of the model, indicating that those factors which determine presence or absence of cones are the same ones which affect the final total amount of cones produced. One main exception to this is age, a factor which determines the absence of cone formation in reproductively immature young trees. This parameter is of particular relevance in stands less than 20 years old although the age factor loses importance from then onwards (Thornley and Johnson, 1990; Montero et al., 2008). Moreover, once the tree starts to produce cones, the total yield ceases to be related to the age of the tree but rather to the size. However, although 54% of observations are zeros, only 9% of trees actually produce no crops at all over the whole period of the study, which suggests that the presence or absence of cones is more dependant on temporal variables (climate) than on “switching on/off” processes (such as tree maturity). A second exception to the aforementioned shared covariates has to do with the weather covariates related to events following female flowering, which only govern the abundance process. In this case, severe droughts during the summer and extreme frosts during winter, cause significant loss of recently opened female flowers, while the level of rainfall during the year of maturation governs the final size and average weight of the existing cones (Mutke et al., 2005b).

Unsurprisingly, the variables explaining spatial variability between trees and between plots indicate that lowest probability of null crops and larger crops can be expected for trees which are large, dominant and growing in lightly stocked stands with the highest site quality (within a given natural unit). At regional extent, the greatest probability of high cone production and absence of crop failure is expected in those areas which have greater soil water availability, i.e. soils with larger proportions of clay and hence, higher capacity for water retention (such as units 1, 2 and 7) or soils with additional water supply due to high water table level (unit 10). These results again confirm the role of water and soil water retention capacity as the main limiting factors in fruit production in a dry Mediterranean environment (Sarker et al., 2003).

5.4. Predictive ability and model application

The model developed is capable of explaining much of the spatial variability between plots, attaining levels of efficiency close to 80%; a figure much larger than those previously obtained using shorter statistical series at plot level for the same (or other) regions (Castellani, 1989; Piqué, 2003). The importance of an ecologically based stratification of the territory suggests spatial dependence of the process at large scales, with a secondary control related to management practices (stand density). Soil-specific attributes which define differences within Natural Units may also contribute to random non-explained variability at this scale. As regards purely temporal variability between years at regional scale, efficiency levels reach values of 70–80%, a figure similar to that attained by Mutke et al. (2005a) using a regional statistical series covering a period of 40 years. The above results indicate that the model adequately identifies the factors governing the fruiting process at larger spatial or temporal scales, while the efficiency of the model is lower when explaining the variability at smaller ranges. In the latter case, the model explains 50–60% of the purely spatial variability between-trees, the rest of the variability probably being associated with either genetic or microsite factors. Nevertheless, even at tree

extent, our results are a significant improvement on those reported in Calama et al. (2008), in which a purely spatial model explained 39% of between trees variability in cone production. Finally, when considering the patterns of interannual variability at plot or tree extent, the efficiency of the model drops to values of 45–50% and 30–35% respectively, which may be due to factors such as the use of a single weather series for the whole of the studied region instead of more accurate weather data at natural unit level (not available at present), different response to rainfall depending on the soil properties, or annual events such as tree damage, pests or cone predation.

Few differences were detected between the two proposed approaches for model application, with similar goodness-of-fit statistics for both the fitting and the validation data set. The main advantage of approach (a) – fixed value of 0.46 as the threshold probability – is that it allows the percentage of non cone-bearing trees to be estimated, providing useful information for managing both commercial cone collection and natural regeneration (effective population size, etc.). On the other hand, approach (b) – where expected yield equals the product of the expected probability for non-zero process and the expected cone weight – shows better results in matching frequencies distribution – especially for small non-zero values – and in explaining temporal variability at regional scale.

However, both of the abovementioned approaches allow us to accurately predict the cone production at different temporal and spatial scales, especially at regional and stand level, which are more relevant to forest planning and management and to the pine nut processing industry. In this sense, the implementation of the proposed model as the annual cone production module within the PINEA2 model and its software application will enable forest managers to carry out annual planning activities such as identifying less viable areas for cone production or predicting the amount of cones suitable for industry at local and regional scale. It will also help them to define the crops which can be expected under different climatic scenarios or to propose more suitable silvicultural practices for promoting cone production at each location.

Acknowledgements

Authors thank Carlos García-Güemes, Ángel Bachiller and Enrique Garriga for their dedication to plot installation, maintenance and data collection. Authors also wish to thank Janet Tooze, for providing MIXCORR macro; Marta Pardos, Felipe Crecente and Susana Barreiro for their helpful comments on earlier versions of the manuscript, and Adam Collins for English edition. Finally, authors would like to express their gratitude to two anonymous reviewers whose comments and suggestions largely improved the text. Present work was mainly developed during a short term scientific mission carried by the corresponding author in the Instituto Superior de Agronomia de Lisboa, under the framework of COST Action FP-0603. Research has been carried out in the context of INIA funded projects CPE-03-001-C5.2 and RTA-00044-2007.

References

- Abrahamson, W.G., Layne, J.N., 2003. Long-term patterns of acorn production for five oak species in xeric Florida uplands. *Ecology* 84 (9), 2476–2492.
- Affleck, D.L.R., 2006. Poisson mixture models for regression analysis of stand level mortality. *Can. J. Forest Res.* 36, 2994–3006.
- Aitchinson, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. *J. Am. Stat. Assoc.* 50, 901–908.
- Belasco, E.J., Ghosh, S.K., 2008. Modeling censored data using zero-inflated regressions with an application to cattle production yields. American Agricultural Economics Association Annual Meeting, Florida, July 2008. Available on line at <http://ageconsearch.umn.edu/bitstream/6341/2/456273.pdf>.
- Calama, R., Cañadas, N., Montero, G., 2003. Inter-regional variability in site index models for even-aged stands of stone pine (*Pinus pinea* L.) in Spain. *Ann. Forest Sci.* 60, 259–269.
- Calama, R., Montero, G., 2007. Cone and seed production from stone pine (*Pinus pinea* L.) stands in Central Range (Spain). *Eur. J. Forest Res.* 126 (1), 23–35.
- Calama, R., Sánchez-González, M., Montero, G., 2007. Integrated management models for Mediterranean multifunctional forests: the case of stone pine (*Pinus pinea* L.). *EFI Proc.* 56, 57–69.
- Calama, R., Mutke, S., Gordo, J., Montero, G., 2008. An empirical ecological-type model for predicting stone pine (*Pinus pinea* L.) cone production in the Northern Plateau (Spain). *Forest Ecol. Manage.* 255 (3/4), 660–673.
- Cañadas, M.N., 2000. *Pinus pinea* L. en el Sistema Central (valles del Tiétar y del Alberche): desarrollo de un modelo de crecimiento y producción de piña. Ph.D. Thesis. Universidad Politécnica de Madrid.
- Castellani, C., 1989. La produzione legnosa e del frutto e la durata economico delle pinete coetanee di pino domestico (*Pinus pinea* L.) in un complesso assestato a prevalente funzione produttiva in Italia. *Ann. ISAF* 12, 161–221.
- Davis, L.D., 1957. Flowering and alternate bearing. *Proc. Am. Soc. Hortic. Sci.* 70, 545–556.
- Despland, E., Houle, G., 1997. Climate influences on growth and reproduction of *Pinus banksiana* (Pinaceae) at the limit of species distribution in eastern North America. *Am. J. Bot.* 84, 928–937.
- Eis, S., 1976. Association of western pine crops with weather variables. *Can. J. Forest Res.* 6, 6–12.
- Eis, S., 1973. Cone production of Douglas-fir and grand fir and its climatic requirements. *Can. J. Forest Res.* 3, 61–70.
- Fortin, M., DeBlois, J., 2007. Modelling tree recruitment with zero-inflated models: the example of hardwood stands in Southern Québec. *Can. Forest Sci.* 53 (4), 529–539.
- García, D., Zamora, R., Gómez, J.M., Jordano, P., Hódar, J.A., 2000. Geographical variation in seed production, predation and abortion in *Juniperus communis* throughout its range in Europe. *J. Ecol.* 88, 436–446.
- García, D., 2001. Effects of seed dispersal on *Juniperus communis* recruitment on a Mediterranean mountain. *J. Veg. Sci.* 12, 839–848.
- Greene, D.F., Johnson, E.A., 2004. Modelling the temporal variation in the seed production of North American trees. *Can. J. Forest Res.* 34, 65–75.
- Hall, D.B., 2000. Zero-inflated, Poisson and binomial regression with random effects: a case study. *Biometrics* 56, 1030–1039.
- Heilbron, D., 1994. Zero-altered and other regression models for count data with added zeros. *Biom. J.* 36, 531–547.
- Herrera, C.M., Jordano, P., Guitián, J., Traveset, A., 1998. Annual variability in seed production by Woody plants and the masting concept: reassessment of principles and relationships to pollination and seed dispersal. *Am. Nat.* 152, 576–594.
- Ihalainen, M., Salo, K., Pukkala, T., 2003. Empirical prediction models for *Vaccinium myrtillus* and *V. vitis-idaea* berry yields in North Karelia, Finland. *Silva Fenn.* 37 (1), 95–108.
- Isagi, Y., Sugimura, K., Sumida, A., Ito, H., 1997. How does masting happen and synchronize? *J. Theor. Biol.* 187, 231–239.
- Karlsson, C., 2000. Seed production of *Pinus sylvestris* after release cutting. *Can. J. Forest Res.* 30, 982–999.
- Kelly, D., 1994. The evolutionary ecology of mast seeding. *Trends Ecol. Evol.* 9, 465–470.
- Kelly, D., Sork, V.L., 2002. Mast seeding in perennial plants: why, how, where? *Annu. Rev. Ecol. Syst.* 33, 427–447.
- Koenig, W.D., Mumme, R.L., Carmen, W.J., Stanback, M.T., 1994. Acorn production by oaks in Central Coastal California: variation within and among years. *Ecology* 75, 99–109.
- Koenig, W.D., Knops, J.M.H., Carmen, W.J., Stanback, M.T., Mumme, R.L., 1996. Acorn production by oaks in central California: influence of weather at three levels. *Can. J. Forest Res.* 26, 1677–1683.
- Koenig, W.D., Knops, J.M.H., 2000. Patterns of annual seed production by Northern Hemisphere trees: a global perspective. *Am. Nat.* 155 (1), 59–69.
- Lalonde, R.G., Roitberg, B.D., 1992. On the evolution of masting behaviour in trees: predation or weather? *Am. Nat.* 139 (6), 1293–1304.
- Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34, 1–14.
- Lee, K.J., 1979. Factors affecting cone initiation in pines: a review. Research Report 15. Institute of Forest Genetics, Soweon (Korea).
- Li, N., Elashoff, D.A., Robbins, W.A., Xun, L., 2008. A hierarchical zero-inflated log-normal model for skewed responses. *Stat. Methods Med. Res.*, doi:10.1177/0962280208097372.
- Masaka, K., Sato, H., 2002. Acorn production by Kashiwa oak in a coastal forest under fluctuating weather conditions. *Can. J. Forest Res.* 32, 9–15.
- Mayer, D., Roy, D., Robins, J., Halliday, I., Sellin, M., 2005. Modelling zero-inflated fish counts in estuaries – a comparison of alternate statistical distributions. In: Zerger, A., Argent, R.M. (Eds.), MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 2581–2587, ISBN: 0-9758400-2-9. <http://www.mssanz.org.au/modsim05/papers/mayer.pdf>.
- Mencuccini, M., Piussi, P., Zanzi Sulli, A., 1995. Thirty years of seed production in a subalpine Norway spruce forest: Patterns of temporal and spatial variation. *Forest Ecol. Manage.* 76, 109–125.
- Miina, J., Hotanen, J.P., Salo, K., 2009. Modelling the abundance and temporal variation in the production of bilberry (*Vaccinium myrtillus* L.) in Finnish mineral soil forests. *Silva Fenn.* 43 (4), 577–593.
- Montero, G., Calama, R., Ruiz-Peñado, R., 2008. Selvicultura de *Pinus pinea* L. In: Montero, G., Serrada, R., Reque, J. (Eds.), *Compendio de Selvicultura de Especies*. Madrid, INIA – Fundación Conde del Valle de Salazar, pp. 431–470.

- Mutke, S., Gordo, J., Gil, L., 2005a. Variability of Mediterranean stone pine cone production: yield loss as response to climatic change. *Agric. Forest Meteorol.* 132, 263–272.
- Mutke, S., Gordo, J., Gil, L., 2005b. Characterization of a stone pine (*Pinus pinea* L.) clone bank. *Silvae Genet.* 54 (4–5), 189–197.
- Mutke, S., Siev nen, R., Nikinmaa, E., Perttunen, J., Gil, L., 2005c. Crown architecture of grafted stone pine (*Pinus pinea* L.): shoot growth and bud differentiation. *Trees* 19 (1), 15–25.
- Nanos, N., Calama, R., Ca nadas, N., Garc a, C., Montero, G., 2003. Spatial stochastic modelling for cone production from stone pine (*Pinus pinea* L.) stands in Spanish Northern Plateau. In: Amaro, A., Reed, D., Soares, P. (Eds.), *Modelling Forest Systems*. CABI Publishing, Wallingford, pp. 131–141.
- Obeso, J.R., 2002. The cost of reproduction in plants. *New Phytol.* 155, 321–348.
- Piovesan, G., Adams, J.M., 2001. Masting behaviour in beech: linking reproduction and climatic variation. *Can. J. Bot.* 79, 1039–1047.
- Piqu , M., 2003. Modelos de producci n para las masas de *Pinus pinea* L. en Catalunya: orientaciones para la gesti n y el aprovechamiento sostenible de madera y pi a. Ph.D. Thesis. Universidad de Lleida.
- Rathbun, S.L., Fei, S., 2006. A spatial zero-inflated Poisson regression model for oak regeneration. *Environ. Ecol. Stat.* 13, 409–426.
- Ridout, M., Demetrio, C.G.B., Hinde, J., 1998. Models for count data with many zeros. In: *Proceedings of the XIXth International Biometrics Conference*, Cape Town, pp. 79–192.
- Sarker, A., Erskine, W., Singh, M., 2003. Regression models for lentil seed and straw yields in Near East. *Agric. Forest Meteorol.* 116, 61–72.
- Scarascia-Mugnozza, G., Oswald, H., Piussi, P., Radoglou, K., 2000. Forests of the Mediterranean region: gaps in knowledge and research needs. *Forest Ecol. Manage.* 132, 97–109.
- Schauber, E.M., Kelly, D., Turchin, P., Simon, C., Lee, W.G., Allen, R.B., et al., 2002. Masting by eighteen New Zealand plant species: the role of temperature as a synchronize cues. *Ecology* 83 (5), 1214–1225.
- Sirois, L., 2000. Spatiotemporal variation in black spruce cone nad seed crops along a boreal forest-tree line transect. *Can. J. Forest Res.* 30, 900–909.
- Sork, V.L., 1993. Evolutionary ecology of mast seeding in temperate and tropical oaks (*Quercus* spp). *Vegetatio* 107/108, 133–147.
- Sork, V.L., Bramble, J., Sexton, O., 1993. Ecology of mast fruiting in three species of North American deciduous oaks. *Ecology* 74 (2), 528–541.
- Tapper, P.G., 1996. Long-term patterns of mast fruiting in *Fraxinus excelsior*. *Ecology* 77 (8), 2567–2572.
- Thornley, J.H.M., Johnson, I.R., 1990. *Plant and Crop Modelling*. Blackburn and Press, Caldwell.
- Tooze, J.A., Grunwald, G.K., Jones, R.H., 2002. Analysis of repeated measures data with clumping at zero. *Stat. Methods Med. Res.* 11, 341–355.
- Tu, W., 2002. Zero-inflated data. In: El-Shaarawi, A.H., Piegorisch, W.W. (Eds.), *Encyclopedia of Environmetrics*. John Wiley and Sons, Chichester, pp. 2387–2391.
- Van der Wall, S.B., 2002. Masting in animal dispersed pines facilitates seed dispersal. *Ecology* 83 (12), 3508–3516.
- Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B., 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecol. Model.* 88, 297–308.
- Woodward, A., Silsbee, D.G., Schreiner, E.G., Means, J.E., 1994. Influence of climate on radial growth and cone production in subalpine fir (*Abies lasocarpa*) and mountain hemlock (*Tsuga mertensiana*). *Can. J. Forest Res.* 24, 1133–1143.
- Woollons, R.C., 1998. Even-aged stand mortality estimation through a two-step regression process. *Forest Ecol. Manage.* 105, 189–195.