Inter-annual variability in *Prosopis caldenia* pod production in the Argentinean semiarid Pampas: A modelling approach

Lucia Risio[1, 2,*], Rafael Calama[2, 3] Stella M. Bogino[4] and Felipe Bravo[1, 2]

[1] Sustainable Forest Management Research Institute University of Valladolid-INIA, Av. Madrid 44, 34004 Palencia, Spain.

[2] Departamento de Producción Vegetal y Recursos Forestales, E.T.S. de Ingenierías Agrarias, Universidad de Valladolid, Palencia, Spain.

[3] Departamento de selvicultura y Gestión Forestal, CIFOR-INIA. Madrid, Spain.

[4] Departamento de Ciencias Agropecuarias, Universidad Nacional de San Luis, Argentina.

[*] *Corresponding author: Tel: +34 979108427, Fax: +34 979108440,*
E-mail address:luciarisio@gmail.com.

Abstract

The driest part of the Argentinean pampas is occupied by semiarid woodlands dominates by *Prosopis caldenia* Burkart (Calden). Calden pods are a highly valuable fodder supplement for livestock but its production is highly variable. Our objective was to analyze and model the temporal pattern in inter-annual variability of Calden pod production. Our key hypothesis is that weather conditions are the main determinant of the pod masting behavior. Tree size and climatic variables were evaluated as explanatory covariates using a zero-inflated log-normal modelling approach. The proposed final model structure incorporated 25 parameters, including four variance components, two intercepts for both the logistic and the log-normal parts of the model, and nineteen parameters associated with fixed effects. Climate had a strong influence on the flowering-fruiting Calden process and on the inter-annual variability of the final pod production at the tree level. Temperatures during bud breaking, flowering and fruit shedding, together with the precipitation from the final month of fruit shedding and the total amount of the prior vegetative cycle, were the main weather covariates that affect the processes.

Keywords

Abbreviation list

g per crown area square metre grams per square metre of crown area; ZILN zero inflated log normal

1. Introduction

Calden forests (Caldenales) are semiarid woodlands covering about 170,000 km^2 of central Argentina. This xerophytic open woodland is a transitional ecosystem between the Pampas grasslands and the dry Monte shrublands, dominated by calden trees (*Prosopis caldenia* Burkart) an endemic species of Argentina. Two opposing processes simultaneously operate in the Caldenales. Firstly, there is a high rate of deforestation across the area of its distribution (0.86% per year), leaving only 18% of the original area (SAyDS, 2007). Secondly, as the original area shrinks, adjacent grasslands not occupied by crops are increasingly encroached upon by calden, leading to very high density secondary woodland (Dussart et al., 1998).

The invasion of pristine grasslands by calden and the increased densities of this species in savannas are well-known vegetation changes in the semiarid region of central Argentina (Dussart et al., 1998). Woody-plant encroachment has long been of concern to a broad range of stakeholders, from pastoral farmers to ranchers, because of the subsequent negative impact on livestock production (Anadon et al., 2014), the main economic activity in these woodlands after beekeeping. The non-wood forest products are a major source of income for the Caldenales woodland owners since wood extraction, mostly firewood production, is only a marginal activity (SAyDS, 2007).

Fruits of *Prosopis* spp. have served as a food source for humans and domestic cattle in rural communities of arid and semiarid environments around the world since ancient times (Burkart, 1952). Calden pods are a highly valuable fodder supplement for livestock due to the pod's nutritional characteristics: 15% raw protein, 2.2 Mcal of metabolizable energy per kg of dry matter and a 52% of dry matter *in situ* digestibility (Privitello et al., 2001; Menvielle, 1985), during its ripening period (June-July). Due to its availability, the pods can provide the sole source of food at certain times of the year or supplement grass for cattle in the winter months (Privitello and Gabutti, 1988). Calden flowers are also recognised as a source of nectar (Genise et al., 1990) and pollen for honey bees (Andrada et al., 2005). Also, the spatial and temporal dynamics of flowering, fruiting and seeding can be considered as a key to controlling natural recruitment of plant populations (Calama et al., 2011).

Most *Prosopis* species produce abundant flowers at a predictable time of the year (spring and/or summer depending on the species) (Simpson et al., 1977). Despite the predictability of the *Prosopis* blooming, high variability in fruit production has been observed (Salvo et al., 1988). Calden pod production is highly variable from year to year and from tree to tree, even with trees from the same stand (Peinetti et al., 1991). In years of high fruit production, it is common to find trees that allocate a high proportion of photoassimilates into pod production growing close to neighboring trees of a similar size that do not bear fruits (Peinetti et al., 1991). Many long-lived plant species exhibit strong synchronized annual variability in fruit production, this phenomenon is known as masting or mast fruiting (Ostfel and Keesing, 2000). The interval between consecutive masting and the degree of periodicity is species specific and varies depending on endogenous control factors, weather conditions and resource availability (Thomas and Packham, 2007; Han et al., 2008).

The mast seeding has important ecological consequences, not only on the recruitment of the species that exhibit this reproductive behaviour (e.g. seedling establishment may be limited to mast years, (Negi et al., 1996)), but also on a myriad of organisms in other trophic levels, whether directly or indirectly related to seed and fruit availability: e.g. direct consumers of seeds (insects, birds, small and large mammals), and predators of seed consumers and parasites (Espelta et al., 2008; Kelly et al., 2008). Edible fruits and seeds from some forest species can represent an important non-timber forest product (Scarascia-Mugnozza et al., 2000). Additionally, when fruit production is one of the main objectives in forest management planning, adequate estimates of fruit production on spatial and temporal scales are often required.

Our objective is to describe, analyze and model the temporal pattern in inter-annualvariability of *Prosopis caldenia* pod production in woodlands located in the northern limit of its natural distribution area. We shall identify the factors controlling the temporal variability in order to develop models that allow us to predict the annual pod production at tree level. Our key hypothesis is that weather conditions are the main determinant of the masting behaviour. Given the data structure, a Zero-Inflated Log-Normal mixture distribution will be evaluated and fitted as the modelling approach for our data.

4

2. Material and Methods

2.1.1 Study area

Prosopis caldenia woodlands thrive at the edge of the driest area of the Argentinean Pampas, across 34-36ºS and 64-66ºW (Fig. 1) (Anderson et al., 1970). Across its natural distribution area, the total annual precipitation varies from 450 to 620 mm, and it is concentrated in the spring and summer months (78%, from October to March). Temperature ranges from the annual isotherms of 16 to 18 ºC. The area is a well-drained plain with moderate slopes produced by wind and fluvial processes (SAyDS, 2007). Soil types are mainly poorly developed and well drained, with scarce horizon differentiation and low water holding capacity and with 1.5 to 3% organic matter. The area is severely affected by wind and water erosion due to its poor soil structure (Peña Zubiate et al., 1998).

Five sampling sites were located on privately-owned properties in the San Luis province, near the northern edge of the natural distribution area of calden woodlands. These lands are used for cattle ranching, and there is no active silvicultural management practices on them.

2.1.2 *Prosopis* phenology, flowering and fruiting.

Prosopis species sprout at the beginning of spring (average temperature around 16ºC) and stay in leaf until autumn. The initiation of leaf production and cambium activity appears to be rather independent of rainfall (Mooney et al., 1977; Villalba, 1985). The flowers are hermaphrodite, entomophilous and they depend on pollinating insects for seed setting (Fig. 2).

Most *Prosopis* species produce abundant flowers at a predictable time of the year, since they bloom regardless the yearly rainfall fluctuation (Simpson et al., 1977), responding rather to photoperiod as well as the length of the growing season (Solbrig and Cantino 1975). High variability in *Prosopis* fruit production has been observed (Salvo et al., 1988); close observation of native stands of *Prosopis* indicated that only between 0.05 and 0.25% of the flowers buds initiated fruits, and only 20–45% of the initiated fruits reached full size (Cariaga et al., 2005). Similar patterns have been observed in *P. flexuosa* and *P. chilensis* (Mooney et al., 1977). Flower mortality takes into accounts either inflorescence abortion and/or intra-inflorescence flower abortion

5

Cariaga et al., 2005). Peinetti (1991) reported that for *P. caldenia* only 5×10^{-4}% of flowers reached fruit maturity, following the flower and fruit abortion patterns of the genus.

2.1.3 Data

Tree data

We sampled *P. caldenia* pod production in five different sites during the months of June and July (when the ripe pods dehiscence occurred) of seven non-consecutive years (1982, 2000, 2004, 2011, 2012, 2013, 2014). A total of 400 trees were sampled, and the number of sampled trees varied among years and sites (Table 1). Only the trees of site 1 (64) were measured twice (1982 and 2000). Different trees were measured at the remaining sampling sites. At each tree, diameter at breast height (at 1.3 m) in cm, diameter in the base of the stem (at 0.3 m) in cm, tree height (m), and crown area (m) were recorded (Table 2). Once a year when the dehiscence of the annual pod production occurred , four samples per tree of 1 m^2 in each cardinal point below the tree crown, (in the middle of their projection above the ground), were collected and taken to the laboratory, where they were oven dried at 80º C until they reached a constant weight. Wild animal predation or pod redistribution before sample was not taken in account. In order to evaluate the morphological variability (weight and length) a random sub-sample (n=50) of the 2014 pod production were measured.

Climate data

The climate data of the EEA INTA Villa Mercedes meteorological station was used due to its proximity to the samples sites (40 kilometers from the farthest site). Data was organized according to the southern hemisphere vegetative cycle, from May to April. Climate data is available at: http://siga2.inta.gov.ar/en/datoshistoricos/

2.2 Methods

2.2.1 Response and explanatory variables

Annual pod-yield per tree can be expressed either by the number of pods or by their dry weight. In our case we decided to use the weight as the response variable since its takes into account both the phenomena related to the initial processes of floral induction and pollination (determining pod numbers) and the pod growth. Furthermore,

weight is a better indicator of the total amount of resources allocated to the reproductive effort than the number of pods, since it reflects the observed variability in pod length and weight. Tree pod productivity was expressed in grams per square metre of crown area (g per crown area square metre) because production per crown unit area is the most objective way to measure productivity and compare between different stands and locations (Gea-Izquierdo et al., 2006).

For both variables (pod number and weight), the distribution of frequencies 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 pod 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 45% of the observations in the fitting data set (Table 1).

- 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.

The tree size group variables; diameter at breast height, crown radio, basal areas at breast height and crown width were evaluated to explain spatial variability in pod production. The temporal variability in pod production was explained by evaluating different characteristics of the weather over the course of the study period: monthly rainfall (mm), mean, maximum and minimum temperatures (ºC), monthly sum of chill hours, monthly sum of effective sunlight hours, frost free period (days), monthly sum of the days with precipitation and with a mean wind velocity over 18 km.h^{-1}. The last two variables were included due to the possible effect they can have on insect pollination activity. Because flower bud induction is produced the year prior to the flowering-fruiting year, weather variables from the induction year were also evaluated. Finally, due to the southern hemisphere location of the study area, the climate variables were considered according to the vegetative year from May to April, involving two different calendar years.

2.2.2 Modeling approach.

We used zero-inflated models for our modelling approach. The explanatory covariates selection was carried out by first independently fitting a binomial regression model for the dichotomized data for fruiting occurrence and, after, fitting a log-normal model using only the non-null intensity data (weight of pods), as proposed by Heilbron (1994) and Woollons (1998). The 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 significance level of the parameters and removing those that were non-significant. The explanatory covariates may or may not be common to both the occurrence and intensity models.

We compared the three possible alternatives of additional level of random variability (site, year and tree), and then selected the best according to Akaike´s Information Criterion (AIC), Bayesian Information Criterion (BIC), and minus two likelihood (-2LL). Site per year, site per tree and tree per year iteration terms were also evaluated as random sources of variability, but problems in the model convergence were detected and, thus, it was discarded.

The accuracy of the selected model was checked using the fitting data set, since not more data was available. Two alternatives were evaluated to predict the pod production from a tree using the fitted model: a) a cut-off value of 0.55 was set (proportion of fruiting trees in the data set). If the value predicted by the occurrence part of the model was greater than the cut-off value, the pod production was predicted using the intensity part of the model, otherwise the predicted production was zero; b) The pod production from a tree was equal to the product of the expected probability of occurrence, as estimated using the logistic part, and the expected value of pod 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, considering the mean error (E), root mean squared error (RMSE) and modelling efficiency (EF) (see the Electronic Appendix for details). In the case of approach (a), specifity (rate of correctly

8

classified non-events) and sensitivity (rate of correctly classified events) were also evaluated.

All the statistical analyses were carried out with SAS/ETS 9.2 software. The independent fitting of the binomial regression model was implemented in the NLMIXED procedure, the independent fitting of log-normal model in the MIXED procedure and the simultaneously fitting of the ZILN model in the NLMIXED procedure.

3 **Results**

3.1. Pod production

During the study, the annual mean tree pod yield varied from 0 to 124.18 g per crown area square metre and 45.3% of the observations were null. The difference in fruit production can be even bigger at the individual tree level; the maximum registered value for a tree in a single year was 1730 g per crown area square metre. The variability of the annual pod production could be observed at Fig. 4, the production in the year 2000 was nil, so it is not presented. In order to simplify the comparison among years, two trees were not considered due their high production values. The first (1982) with a pod production of 908.6, and the second (2014) with the highest production of the data set (1730 g per square metre of crown area).

The percentage of nil yields per year per site varied from 0 to 100%. Table 1 shows the main fruit production statistics for all sites and years. Regarding the distribution of frequencies for observed annual values of pod production at tree level (Fig. 3), it is possible to detect a clear pattern of zero abundance and asymmetry, with pronounced right-tail behavior indicating that the main part of the production is limited, even in good years, to a few trees. Differences in the length and size of fruits were also observed (Fig. 2). Pod mass ranged from 6.6 to 1.6 g, with a mean of 3.3 and standard deviation (SD) of 1.0. Pod length ranged from 24 to 9.5 cm (mean 16.5, SD 3.4).

3.2 Selection of covariates and random effects

After separately fitting the logistic and the log-normal models, the following covariates were selected:

Logistic equation:

- DBH: breast height diameter (cm)

- Pp^{-1}: is the total precipitation during the previous flowering-fruiting cycle of the sampled flowering-fruiting cycle (mm)

-mTO: is the month mean temperature of October of the sampled flowering-fruiting cycle (ºC)

- mMinTS: is the lowest minimum temperature registered in the September month of the sampled flowering-fruiting cycle (ºC)

- mMinTN: is the lowest minimum temperature registered in the November month of the sampled flowering-fruiting cycle (ºC)

- mMaxTD: is the mean of the maximum temperatures registered in the December month of the sampled flowering-fruiting cycle (ºC)

- mMinTD: is the lowest minimum temperature registered in the December month of the sampled flowering-fruiting cycle (ºC)

- maxMaxTJ: is the highest maximum temperature registered in the January month of the sampled flowering-fruiting cycle (ºC)

Log-normal equation:

- DBH: breast height diameter (cm)

- CA: Crown area (m^2)

- PpApr: is the total precipitation registered in the month of April of the sampled flowering-fruiting cycle (mm)

- mTO: is the month mean temperature of October of the sampled flowering-fruiting cycle (ºC)

- mMinTS: is the lowest minimum temperature registered in the September month of the sampled flowering-fruiting cycle (ºC)

- mMinTO: is the lowest minimum temperature registered in the October month of the sampled flowering-fruiting cycle (ºC)

- mTD: is the month mean temperature of December of the sampled flowering-fruiting cycle (ºC)

- maxMaxTJ: is the highest maximum temperature registered in the January month of the sampled flowering-fruiting cycle (ºC)

- mTF: is the month mean temperature of February of the sampled flowering-fruiting cycle (ºC)

- mTM: is the month mean temperature of March of the sampled flowering-fruiting cycle (ºC)

Most of the covariates are related to temperature and 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 alternative of the ZILN mixed model. Different alternatives of random structures (year, tree and site) were considered (Table 3).

The results show that the best alternative was to consider random intercepts at the year level for both the logistic and the log-normal component of the model (showed the lowest AIC and BIC values). The random intercepts at tree and site levels were not significantly different statistically than the model without random intercepts (P values 0.0524 and 0.0229 respectively). The model does not improve its behavior by partitioning the data according to the tree or site.

Our analyses (Electronic Appendix) suggest that the effects of the tree and site on annual variability of *P. caldenia* pod production were not significant and that the variability of the fructificationis were mainly governed by the year. The common covariate parameters in both parts of the model (Electronic Appendix, Table 1) have the same sign, indicating that the effect of the selected variables on occurrence and abundance is similar. The results also show that temperature is the dominant factor influencing both occurrence and abundance. Total precipitation of the prior cycle influences the occurrence, and April precipitation only affects total final pod mass. The correlation coefficient between random effects in the two parts of the model was highly

significant (Wald test, $P < 0.0001$) and positive (0.5612), indicating that the unobserved factors acting at the year level are also the same for both processes.

3.3 Model evaluation

In the Table 2 of the Electronic Appendix the values of the statistics evaluating the predictive ability of the model for predicting annual tree pod production per m^2 of crown area (g), using the fitting data set are presents. Approach (a) with a fixed cut-off, tended to underestimate the average value, leading to biased estimates, approach (b) instead showed better statistical parameters, leading to unbiased estimates with smaller RMSE and larger EF, but also underestimate the average value. The capacity for correctly classifying the observations as events or non-events approach (a) correctly classifies 67.5% of the non-events (trees with null production), in the case of the sensitivity (percentage of events correctly classified) approach (a) correctly classified the 59% of the data set. As approach (b) is not capable of estimating a null production, no values were obtained for specificity or sensitivity.

4. Discussion

Most *Prosopis* species produce an abundance of flowers at a predictable time of the year, since they bloom regardless of yearly rainfall fluctuation (Simpson et al., 1977), responding, on the other hand, to photoperiod and to the length of the growing season (Solbrig and Cantino, 1975). Despite the predictability of blooming, high variability in fruit production of *Prosopis caldenia* and other *Prosopis* has been observed (Salvo et al. 1988; Peinetti, 1991; Privitello et al., 2001; Cariaga, 2005). Resource limitation, the time of rainfall, frost and wind occurrence, and soil water content have been suggested as physical causes of different *Prosopis* pod abortion rates (Solbrig and Cantino, 1975; Cariaga et al., 2005). The absence of an adequate pollinator system, and predation by bruchids (Coleoptera) might be among the ecological causes (Simpson et al., 1977; Toro et al., 1993).

According to our data, climatic factors are the main determinants of the calden masting behavior at the northern limit of its natural distribution. Evidence of weather control over fruit production has mainly been identified in less productive and resource limiting environments (Sarker et al., 2003), such as those associated with Mediterranean

forests, indicating that a lack of water, light or adequate temperatures at crucial stages influences reproductive success (Calama et al., 2011)

The proposed final model structure incorporates 25 parameters, including four variance components, two intercepts for both the logistic and the log-normal parts of the model and nineteen parameters associated with fixed effects (nine in the logistic part and ten in the log-normal part). Diameter at breast height was only included in the logistic part of the model because data on tree age was unavailable. Diameter could be used as a proxy for age, a factor that determines the absence of pod formation in reproductively immature young trees.

Although 45% of the observations in our data set were zeros, only 17% of the sampled trees had a diameter at breast high < 13 cm. According to Privitello (2001), a 13 cm DBH (in our sampling site area) corresponds to a calden tree approximately 20-25 years old. *Prosopis* peak pod production occurs at 15-20 years of age (Sawal et al., 2004), which suggest that the presence or absence of pods is more dependent on other factors, such as climate, than on "switching on/off" processes like tree maturity.

The weather covariates entering in the model accord closely with the main phenological stages of the flowering-fruiting processes described for the species by Peinetti (1991), such as bud break (temperatures in September, October and November), flowering (temperatures in November and December), fruiting (temperatures in December and January) and fruit shedding (temperatures in March as well as April precipitation). Even though calden fruits always come from flower buds formed during the previous season (Peinetti, et al. 1991), only total precipitation of the prior vegetative cycle was statistical significant and included in the model. The amount of precipitation in the last month of fruit shedding (April) was also included but only affects the final weight of pod production.

Eleven of the fifteen selected covariates were temperature related. Our results suggest that calden fruit occurrence and production are mainly governed by temperature and not by precipitation, even though precipitation does have an effect, as would be expected in a semiarid environment (Cariaga et al., 2005). The independence of the flowering-fruiting process observed in our data could be explained by the access the trees in the sampling sites had to the underground water table.

13

Prosopis trees and shrubs typically develop an extensive dimorphic root system with a vertical tap root several metres long which gives them access to deep soil water (Jobbágy et al., 2011). This led to a relatively high and seasonally stable predawn water potentials (Nilsen et al., 1984), higher transpiration and photosynthesis rates than shallow-rooted associated species (Nilsen et al., 1984), and higher net primary productivity than that which was predicted by precipitation (Sharifi et al., 1982). In the five sampling sites, the ground water table was not deeper than 15 metres (data not shown). In the sampling area the water table has risen over the past 40 years with an average rate of increase at 0.15 m yr^{-1} and total elevation increments reaching up to 10 m in 35 years (Contreras et al., 2013). In one of the sampling sites, stable oxygen isotopes (O_{18}) in ring wood and underground water were analyzed and the use of this source of water by the trees was proved (data not showed).

Felker at al. (1984) suggested that for *Prosopis* species, a sustained temperature threshold requirement exists in order for pollen shed to attract bees. They observed that if the hot weather ceased before the flowers are in full blossom, little bee activity is observed on the flowers, and few pods are observed in the *P. vetulina* trees. This phenomenon is well documented for pines (Boyer, 1978). Without pollen shed and bees, insect-pollinated, self-incompatible plants such as *Prosopis* seriously reduced its fruit production (Felker, et al., 1984).

According to our data, climate has a strong influence on the flowering-fruiting *P. caldenia* process and on the inter-annual variability of the final pod production at tree level. However, we must also take into account that our data set has limitations. The moderate length of data series (7 years), the lack of continuity (only the last four years were consecutive), the trees have had access to the groundwater (which could mask the climate signal), and we only had one set of production data from each tree (only 16% of our trees were sampled twice). Thus, evaluating the existence of a fixed cyclic pattern of annual pod production was not possible. Nevertheless, our pod production results are a significant improvement on the correlation and multiple regression models presented by Privitello et al. (2001). The use of zero-inflated models allows 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 events and non-events (Fortin and Deblois, 2007). Finally, zero-inflated

modeling allows a simultaneous and correlated estimation of the parameters explaining both processes (Calama et al., 2011). It also prevents us from violating 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 (Calama et al., 2008). Both of our modelling approaches (Electronic Appendix, Table 2) allowed us to accurately predict pod production at the tree level.

Weather is not the only controlling factor over the masting behavior of *P. caldenia*. A deeper analysis is needed that includes; a) physiological factors (e.g., hormonal inhibitions caused by the ripening seeds (Lee, 1979)), b) intraspecific genetic variability. Extrapolation of *Prosopis* pod production data from one "natural stand" to another would be extremely difficult, even under similar climatic and moisture conditions because of the genetic variability in *Prosopis*; Felker et al. (1984), c) dendrochronological analysis for evaluating the resource depletion theory whereby the demands on resources in a bumper crop leave the tree with insufficient resources to support a normal crop the following year, d) the incorporation of stand attributes to take into account spatial variability such as stem density, stand age and site index, and e) predation by bruchid beetles (Vir, 1996).

5. Conclusions

Climate has strong influence on the flowering-fruiting of *P. caldenia* and on the inter-annual variability of the final pod production at the tree level on the northern limit of its natural distribution. Temperatures in the months of bud break, flowering and fruit shedding, together with precipitation from the final month of fruit shedding and the total amount of the prior vegetative cycle, are the main weather covariates that affect the processes.

Zero-inflated models allow us to take into consideration the idiosyncrasies of data from plant flowering-fruiting studies without violating standard assumptions or using data transformation. Our statistical approach allows an accurate prediction of *P. caldenia* annual pod production, which will enable forest managers to carry out annual planning activities such as predicting the amount of pods to include in their livestock

management plan or estimating the crops which can be expected under different climatic scenarios.

6. Acknowledgments

7. References

Affleck, D.L.R., 2006. Poisson mixture models for regression analysis of stand level mortality. Canadian Journal of Forest Research 36, 2994–3006.

Aitchinson, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of American Statistical Association 50, 901–908.

Anderson, D.L., Del Aguila, J.A., Bernardon A.E., 1970. Las formaciones vegetales en la provincia de San Luis. RIA, Serie 2, Vol. VII, N°3, 153-183.

Anadón, J.D., Sala, O.E., Turner, B.L., Bennett, E.M., 2014. Effect of woody plant encroachment on livestock production in North and South America. Proceedings of the National Academy of Sciences of the United States of America 111   DOI: 10.1073/pnas.1320585111

Andrada, A., Tellería, M.C., 2007. Pollen collected by honey bees (*Apis mellifera* L.) from south of Caldén district (Argentina): botanical origin and protein content. Grana, 44, 115-122

Belasco, E., 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.

Burkart, A., 1952. Las leguminosas Argentinas silvestres y cultivadas. Acme Agency.

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 Ecology and Management 255, 660-673

Calama, R., Mutke, S., Tomé, J., Gordo, J., Montero, G., Tomé, M., 2011. Modelling spatial and temporal variability in a zero-inflated variables: The case of stone pine (*Pinus pinea* L.) cone production. Ecological Modelling 222, 606-618.

Cariaga, R.E., Aguero, P.R., Ravett, D.A., Vilela, A.E., 2005. Differences in production and mortality of reproductive structures in two *Prosopis* L. (Mimosaceae) shrub species from Patagonia, Argentina. Journal of Arid Environments 63, 696–705.

Contreras, S., Santoni, C., Jobbágy, E., 2013. Abrupt watercourse formation in a semiarid sedimentary landscape of central Argentina: the roles of forest clearing, rainfallvariability and seismic activity. Ecohydrology 6, 794-805.

Dussart, E., Lerner, P., Peinetti, R., 1998. Long-term dynamics of 2 populations of *Prosopis caldenia* Burkart. Journal of Rangeland Management 51, 985-991.

Espelta, J.M., Cortés, P., Molowny-Horas, R., Sánchez-Humanes, B., Retana, J., 2008. Masting mediated by summer drought reduces acorn predation in mediterranean oak forest. Ecology 89, 805-817.

Felker, P., Clark, P., Osborn, J., Cannel, G.H., 1984. *Prosopis* pod production-comparison of North American, South American, Hawaiian and African germplasm in young plantations. Economic Botany 38, 36-51

Fortin, M., DeBlois, J., 2007. Modelling tree recruitment with zero-inflated models: the example of hardwood stands in Southern Québec. Forest science 53 , 529-539.

Gea-Izquierdo, G., Cañellas, I., Montero, G., 2006. Acorn production in Spanish holm oak woodlands. Investigaciones Agrarias: Sistemas de Recursos Forestales 15, 339-354.

Genise, J., Palacios, R.A., Hoc, P., Carrizo, R., Mofffat, L., Mom, M. P., Agullo, M.A., Picca, P., Torregrosa, S., 1990. Observaciones sobre la biología floral de *Prosopis* (Leguminosae, Mimosoidae) II. Fases florales y visitantes en el Distrito Chaqueño Serrano. Darwiniana 30, 71-85

Han, Q., Kabeya, D. Iio, A., Kakubari,Y., 2008. Masting in *Fagus crenata* and its influence on the nitrogen content and dry mass of Winter buds. Tree Physiology 28, 1269-1276.

Heilbron, D., 1994. Zero-altered and other regression models for count data with added zeros. Biometrical Journal 36, 531–547.

Jobbágy, E., Nosetto, M., Villagra, P., Jackson, R., 2011. Water subsides from mountains to desert: Their role in sustaining groundwater-fed oases in a Sandy landscape. Ecological applications 21, 678-694.

Kelly, D., Koenig, W.D., Liebhold, A.M., 2008. An intercontinental comparison of the dynamic behavior of mast seeding communities. Population ecology 50, 329-342.

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).

Ostfeld, R. Keesing, F., 2000. Pulsed resources and community dynamics of consumers in terrestrial ecosystems. Tree 15 , 232-237.

Peinetti, R., Martinez, O., Balboa O., 1991. Intraespecific variability in vegetative and reproductive growth of a *Prosopis caldenia* Burkart population in Argentina. Journal of Arid Environment 21, 37-44.

 Peña Zubiate C. A., Anderson, D. L., Demmi, M.A., Saenz, J. L., D´ Hiriart, A., 1998. Carta de suelos y vegetación de la provincia de San Luis. Secretaría de Agricultura, Ganadería, Pesca y Alimentación, INTA and Gob. de la prov. de San Luis. 115 pp.

Privitello, L. Gabutti, E., 1988. Producción de vainas de caldén (*Prosopis caldenia* BURKART) y análisis de la calidad forrajera. VI congreso forestal argentino. Tomo I: 169-171. Santiago del Estero, Argentina.

Privitello, L., Gabutti, E., Leporati, J., 2001. Chauchas de caldén. Factores ambientales que afectan su producción. Actas en la XVII reunión de la Asociación Latinoamericana de producción animal. La Habana. Cuba.

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: 05-9758400-2-9.
http://www.mssanz.org.au/modsim05/papers/mayer.pdf.

Menvielle, E.E, Hernandez, O.A., 1985. El valor nutritivo de las vainas de caldén (*Prosopis caldenia* Burk.) Rev- Arg. Prod. Animal. Vol. 5 Nº 7-8, 433-439.

Mutke, S., Gordo, J., Gil, L., 2005. Variability of Mediterranean Stone pine cone production: yield loss as response to climatic change. Agricultural and forest meteorology 132, 263-272.

Mooney, H.A., Simpson, B.B., Solbrig, O.T., 1977. Phenology, morphology, physiology. In: Simpson BB (ed), Mesquite. Its biology in two desert scrub ecosystems. U.S./ibp synthesis series 4. Hutchinson & Ross, Dowden, pp 26–43

Negi, A.S., Negi, G.C.S., Singh, S.P., 1996. Establishment and growth of *Quercus floribunda* seedlings after a mast year. Journal of vegetation science 7, 559-799.

Nilsen, E.T, Sharifi, M.R., Rundel, P.W., 1984. Comparative water relations of phreatophytes in the Sonoran Desert of California. Ecology 65, 767-776.

Sarker, A., Erskine, W., Singh, M., 2003. Regression models for lentil seed and straw yields in Near East. Agricultural and forest meteorology 116, 61-72.

Sharifi, M.R, Nilsen, E.T., Rundel, P.W., 1982. Biomass and net primary production of *Prosopis glandulosa* (*Fabaceae*) in the Sonoran Desert of California. American journal of botany 69, 760-767.

Sawal, R.K, Ratan, R, Yadav, S., 2004. Mesquite (*Prosopis juliflora*) pods as a Feed Resource for Livestock –A review- Asian Australasian Journal of Animal Science Vol 17, 719-725.

Scarascia-Mugnozza, G., Oswald, H., Piussi, P., Radoglou, K., 2000. Forests of the Mediterranean region: gaps in knowledge and research needs. Forest ecolology and Management 132, 97–109.

Secretaría de ambiente y desarrollo sustentable de la Nación., 2007. Informe Regional Espinal. Segunda Etapa. ANEXO I. Estado de Conservación del Distrito Caldén.

Salvo, B., Botti, C., Pinto, M., 1988. Flower induction and differentiation in *Prosopis chilensis* (Mol.) Stuntz and their relationship with alternate fruit bearing. In: Habitat MA (ed), The current state of knowledge of *Prosopis juliflora*. FAO, Rome, pp 269–275.

Simpson, B.B., Neff, J.L., Moldenke, A.R., 1977. *Prosopis* flowers as a resource. In: Simpson BB (ed), Mesquite. Its biology in two desert scrub ecosystems. US/IBP Synthesis Series 4. Hutchinson & Ross, Dowden, pp 84–105

Solbrig, O.T., Cantino, P.D., 1975. Reproductive adaptations in *Prosopis* (*Leguminosae*, *Mimosoideae*). Journal of the Arnold arboretum 56,185–210

Thomas, P.A., Packham, J.R., 2007. Ecology of woodlands and forest. Cambridge University Press, Cambridge.

Tooze, J.A., Grunwald, G.K., Jones, R.H., 2002. Analysis of repeated measures data with clumping at zero. Statistical Methods in Medical Research 11, 341–355.

Toro, H., Chiappa, E., Covarrubias, R., Villaseñor, R., 1993. Interrelaciones de polinización en zonas áridas de Chile. Acta Entomologica Chilena 18, 20–29

Tu, W., 2002. Zero-inflated data. In: El-Shaarawi, A.H., Piegorsch, W.W. (Eds.), Encyclopedia of Environmetrics. John Wiley and Sons, Chichester, pp. 2387–2391.

Villalba, R., 1985. Xylem structure and cambial activity in *Prosopis flexuosa* D.C. IAWA Bulletin n.s. 6, 119–130.

Vir, S., 1996. Bruchid infestation of leguminous trees in the Thar desert. Tropical Science 36, 11-13.

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. Ecological modelling 88, 297–308.

Woollons, R.C., 1998. Even-aged stand mortality estimation through a two-step regression process. Forest Ecology and Management 105, 189–195.

Highlights:

We model the *Prosopis caldenia* pod production at tree level.

Zero-Inflated Log-Normal mixture distribution with random components was fitted.

Climate has strong influence on the flowering-fruiting *Prosopis caldenia* process.

Temperatures and precipitation are the main covariates that affect the processes.
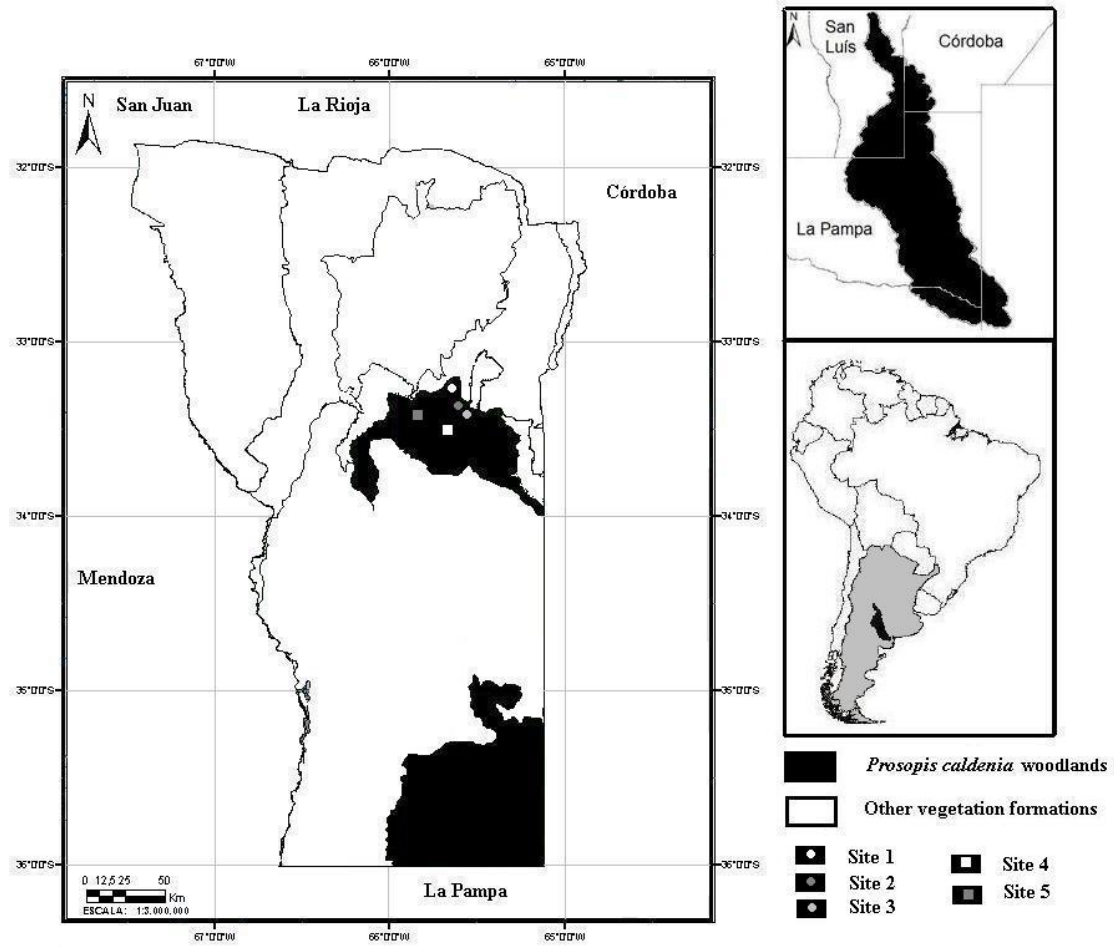
Figure

Figures.

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure legends
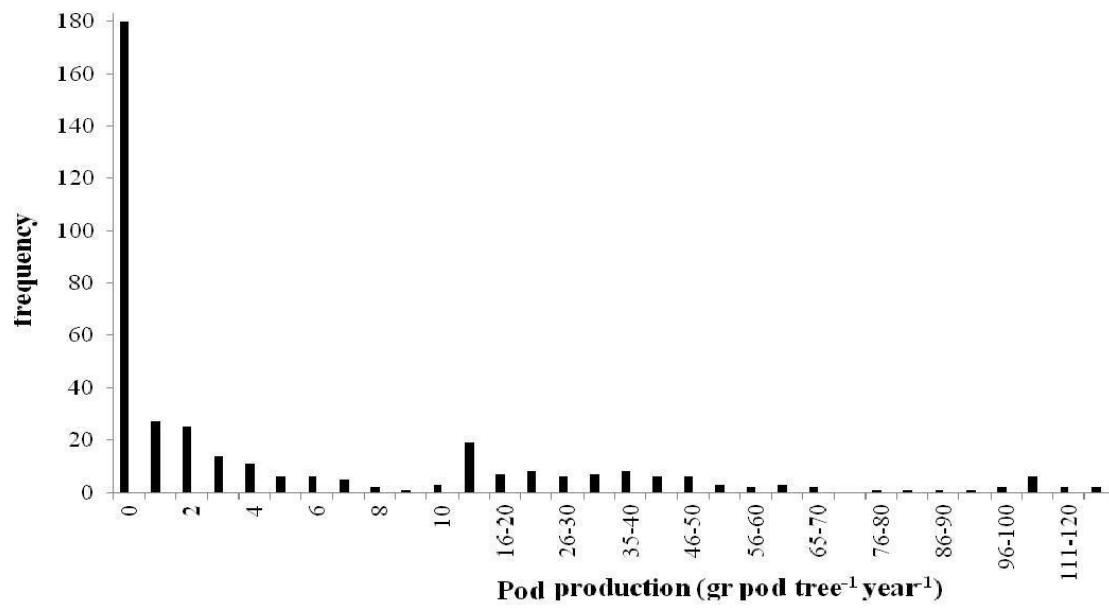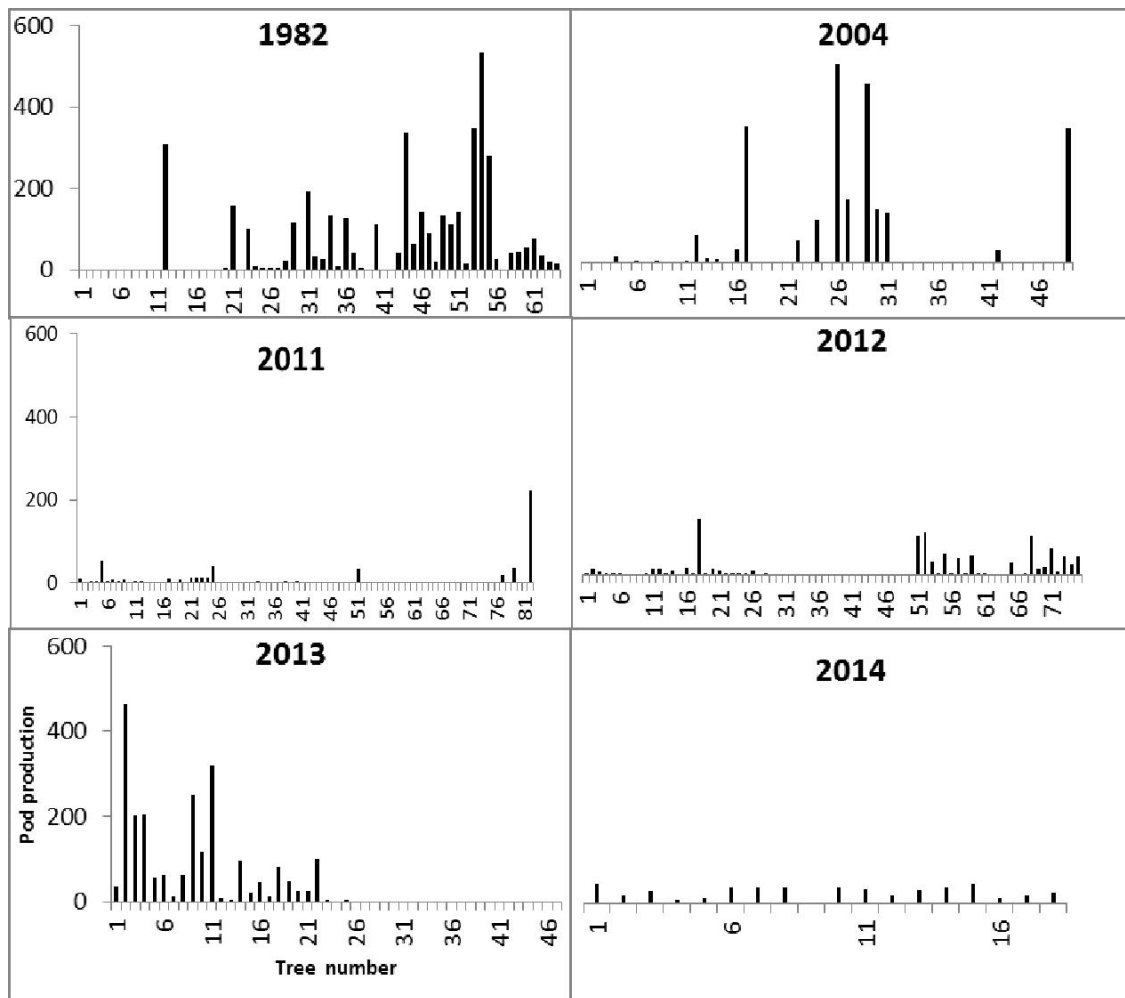
Figure Legends.

Figure 1. Location of the sampling sites and natural distribution area of *Prosopis caldenia* in the San Luis province and in the Argentinean Pampas (based on Anderson, 1970).

Figure 2. *Prosopis caldenia* tree with a high pod production growing along side trees without production and flower and pod details.

Figure 3. Annual pod production frequency of single trees (g tree^{-1}year^{-1}) of *Prosopis caldenia* at the five sampling sites (all sites and years combined).

Figure 4. *Prosopis caldenia* annual pod production (g m^{-2} canopy) per tree and per year. Numbers on the x axis represent individual trees (no tree was sampled more than once).

Tables

Table 1. *P. caldenia* pod production at each sampling site and year (n=400).

Site	Year	n	% with no pods	Mean pod mass (g m^{-2} canopy)	Max pod mass (g m^{-2} canopy)	SD pod mass (g m^{-2} canopy)
1	1982	64	35.91	76.34	908.67	148.08
	2000	64	100	0.00	0.00	0.00
2	2004	49	67.4	47.67	490.75	114.22
	2011	31	33.3	9.51	222.21	40.14
3	2011	26	0.00	3.21	3.20	6.23
	2012	26	16.67	10.49	137.15	26.93
	2013	25	8.34	21.85	115.63	28.59
4	2011	25	0.00	9.00	53.85	12.12
	2012	25	95.33	0.47	10.88	2.18
	2013	22	100	0.00	0.00	0.00
5	2012	25	16	29.00	104.5	32.81
	2014	18	0.00	124.18	1726.59	400.08

Table 2. Main tree level variables of the sampled trees of *Prosopis caldenia* (n=400)

Tree attributes	Tree diameter (cm)		Height (m)	Crown		Pod yield (g m^{-2} canopy)
	Stem base	Breast height		Diameter (m)	Area (m^2)	
Maximum	133.7	96.3	14.5	18.5	265.0	1726.6
Minimum	8.3	3.2	2.6	1.2	1.2	0.0
Mean	38.9	28.9	6.7	8.0	52.0	30.2
SD	21.9	17.0	2.6	3.4	47.0	115.1

Table 3. Fitting statistics and information criteria to the different alternatives of random structures (random site, tree and year) for the ZILN mixed model.

Random level	No	Tree	Site	Year
-2LL	17281	18813	17701	16996
AIC	18342	18547	16904	16403
BIC	18221	17947	17041	16637
p<LRT	-	0.0524	0.0229	0.0004

-2LL: minus 2 logarithm of likelihood function; AIC: Akaike´s Information Criterion; BIC: Bayesian Information Criterion; R^2-adj: adjusted R square coefficient; p<LRT: indicates the probability for a likelihood ratio test among evaluated alternatives with the

1. Zero-inflated log-normal distribution (ZILN)

The zero inflated models (ZI) (Lambert, 1992) was used for the zero observations, the skewed distribution and the positive outcomes with no transformation of the data, including a small constant to replace the zero data (Mutke et al., 2005) or using some Poisson mixture distribution that generally tends to underestimate the number of zeros (Affleck, 2006). In the zero-inflated models, distributional particularities are dealt with by combining a distribution which explains the binary nature of the absence or occurrence of zeros, with a second distribution which models the response variable conditioned by its occurrence, with a non-null value. The binary part is commonly modeled using a logistic function with a binomial distribution. For modelling the intensity of the event, the log-normal distribution is widely used with continuous data, since it avoids zero-truncation and offers greater potential for modelling highly skewed distribution (Tu, 2002).

The so-called Zero-Inflated Log-Normal mixture distribution (ZILN eq.(1)) was first introduced by Aitchinson (1955), and it has been applied to modelling phenomena such as fishery catch rates (Mayer et al., 2005), cattle production (Belasco and Ghosh, 2008), air contaminant levels (Tu, 2002) or stone pine cone production (Calama et al., 2011).

$$P\,(Y=y|\mathbf{x},\mathbf{z}) \begin{cases} \theta & \text{if } y = 0 \\ (1-\theta)f(y) = (1-\theta)\dfrac{1}{y\sigma\sqrt{2\pi}}\,e^{-(\ln y - \mu)^2/2\sigma^2)} & \text{if } y > 0 \end{cases} \qquad (1)$$

Where θ represents the probability of the occurrence of a non-event (zero fruiting) and f(y) indicates the probability density function of 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 modeling θ and μ, respectively; ln represents the natural logarithm.

The regression models can be parameterized as:

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

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

Where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent vectors of unknown but estimable parameters.

Simultaneous estimation of both models is attained by maximizing the overall likelihood function L, a more detailed explanation of this methology can be found at (Welsh et al., 1996, Calama et al., 2011). Different data correlation levels were 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 (Tooze et al., 2002) assuming these Eqs. (2a) and (2b) should be re-expressed as:

$$\text{logit } \theta = \ln\left(\frac{\theta}{1-\theta}\right) = \mathbf{x}\boldsymbol{\alpha} + \mu \tag{3a}$$

$$\mu = E\{\ln(y)|y > 0\} = \mathbf{z}\boldsymbol{\beta} + \nu \tag{3b}$$

Where μ and ν denote the random components acting at the proposed level of variability (tree, site and year) for the logistic and the log-normal parts of the ZILN model. Random intercepts for the logistic and the log-normal part of the model were assumed to follow a bivariate normal distribution (more information about the likelihood function under a ZILN mixed model with different levels of random variability can be found at Calama et al., 2011). The objective of the maximization is to obtain simultaneous estimates for the vectors of parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, associated with explanatory covariates for both logistic and log-normal linear predictors, and for the variance components $\sigma_\mu^2, \sigma_\nu^2, \rho$ and σ^2 (random variability at year, site and/or tree level), ρ (correlation term between the two parts of the model) and σ^2 (the variance for the error in the log-normal part).

2. Evaluation of the different modeling approaches

The different modelling approaches (a) and (b) were compared using the fitting data set, considering the mean error (E), root mean squared error (RMSE) and the modelling efficiency (EF).

$$E = \frac{\sum_{ij}(y_{ij} - \hat{y}_{ij})}{n} \tag{1}$$

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

$$EF = 1 - \left[\frac{\Sigma_{ij}(y_{ij} - \hat{y}_{ij})^2}{\Sigma_{ij}(y_{ij} - \bar{y}_{ij})^2} \right] \qquad (3)$$

Where n represents the total number of observations, y_{ij} and $\hat{y}_{ij}$ are the observed and predicted values of pod 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 as the predicted values for the estimated nonzero intensities, assuming $v_i = E$ $(v_i)=0$ and the following antilogarithmic transformation:

$$E(Y_{ij}|y_{ij} > 0) = exp\left[z_{ij}\beta + \frac{\sigma_v^2}{2} + \frac{\sigma^2}{2} \right] \qquad (4)$$

3. Results

Table 1. Fixed parameters and variance component estimates for the selected ZILN mixed model including correlated random parameters at plot level.

	Covariate	Logistic part (α)	Log normal part (β)
	Intercept	-1.5123	-1.9832
Tree covariates	d	1.9401	0.9524
	CA	-	0.0897
Climate covariates	mTO	0.0021	0.0073
	mMinTS	-0.0725	-0.1876
	mMinTN	-1.1045	-
	mMaxTD	0.3875	-
	maxMaxTJ	-0.0540	-0.7854
	mMinTD	-0.4532	-
	mTA	1.1201	-
	mMinTO	-	-0.4875
	mTD	-	0.7589
	mTF	-	-0.0037
	mTM	-	0.8751
	Pp-1	0.1701	-
	PpApr	-	0.7986
Random components	σ^2_u (year)	0.5107	-
	σ^2_v (year)	-	0.3574
	σ^2(residual)	-	1.5019
	ρ (correlation term)	0.5612	

Table 2. Ability of the model for predicting annual production of *P. caldenia* pod at tree level over fitting data set. The ZILN approaches used were "a" (fixed cut-off=0.55 for defining fruiting-non fruiting) and "b" (the production of pod equals the product of the expected probability of occurrence and the expected value of pod production).

Approach	(a)	(b)
E (pp.CA m^2 year^{-1})	9.52	1.67
p-value (t-test)	0.0006	0.0012
pw_pred (pp.CA m^2 year^{-1})	51.02	58.87
pw_obs (pp.CA m^2 year^{-1})	60.54	-
RMSE	6.987	6.782
EF (%)	29.84%	31.02%
Sensitivity (%)	59.32%	-
Specificity (%)	67.54%	-
Events/no events obs	220/180	-
Events/no events pred	130/270	-

E: mean error; pw_pred and pw_obs: predicted and observed mean pod production (pp.CA m^2, pod production per m^2 of crown area (g)); RMSE: root mean squared error; EF: model efficiency; sensitivity: rate of correctly classified events; specificity: rate of correctly classified non-events.