Statistical Modeling of
Hunting Success Using Hunter Surveys

Jose Carlos Martinez Avila[1], Peter Filzmoser[1] and Neyko M. Neykov[2]

[1] Vienna University of Technology, Vienna, Austria

[2] Bulgarian Academy of Sciences, Sofia, Bulgaria

Abstract: Different methods are available to estimate the wildlife population development. In this paper a method based on hunter surveys of a Spanish game preserve is proposed. A Poisson model is fitted to estimate the relationship between hunting success and other independent variables. In order to find the most appropriate model, a model selection procedure based on repeated cross validation is used to determine which explanatory variables describe the hunting success better in terms of prediction ability. The estimated model is successful in the case of the red legged partridge (*Alectoris rufa*). The model can be used for applying different management criteria in a game preserve. Some possibilities are demonstrated for particular scenarios.

Zusammenfassung: Verschiedene Methoden zur Schätzung der Entwicklung einer Tierpopulation stehen zur Verfügung. In diesem Artikel wird eine Methode basierend auf Aufzeichnungen von Jägern in einem spanischen Wildgehege vorgeschlagen. Ein Poisson Modell wird angepasst, um die Beziehung zwischen Jagderfolg und anderen unabhängigen Variablen zu schätzen. Um das am besten geeignete Modell zu finden wird ein Modellauswahlverfahren basierend auf wiederholter Kreuzvalidierung verwendet, um festzustellen, welche erklärenden Variablen den Jagderfolg besser in Bezug auf Prognosefähigkeit beschreiben. Das geschätzte Model ist erfolgreich im Falle des roten Rebhuhns (*Alectoris rufa*). Das Modell kann zur Anwendung verschiedener Management-Kriterien in einem Wildgehege verwendet werden. Einige Möglichkeiten werden für bestimmte Szenarien erläutert.

Keywords: Model Selection, Poisson Model, Hunting Success, Statistical Modeling, Sustainable Management.

1 Introduction

A correct assessment of the number of animals in a wildlife population is the first step for a sustainable management of this population. There are between species, within species, and environment species interactions. Particularly relevant is the between species interaction such as predator-prey (Virgós and Travaini, 2005). This predator-prey interaction is in many cases critical if the predator is highly specialized and depending on the prey. An example is the Iberian Lynx (*Lynx pardinus*) and the Spanish imperial eagle (*Aquila adalberti*), which both depend on the European rabbit (*Oryctolagus cuniculus*) (Delibes-Mateos, Delibes, Ferreras, and Villafuerte, 2008). However, their interactions on the environment are most relevant because their effects spread out over the ecosystem. An

extremely dry period or heavy rains (Palomares, 2003) lead to a change in the amount of resources available and a dynamic process starts until a new equilibrium is reached. Intra species interactions such as diseases, like the Rabbit Hemorrhagic Disease (RHD), modify the population dynamics (Calvete, 2006).

When the human factor plays a determinant role as in the case of hunting preserves, the population management is more complicated. Hunting is an outdoor activity which has two key features, economic (Bernabéu, 2000) and social. Hunting creates an income in areas where agriculture and livestock are not profitable. The market associated to hunting, like weapons, ammunition, catering, accommodation, guides and assistants, means a principal amount in rural areas in order to fix human population (Martínez, Viñuela, and Villafuerte, 2002; Mozumder, Meghan Starbuck, Berrens, and Alexander, 2007). Fees and hunting taxes are important financial sources for regions and towns. Lands have a great value for hunters that sometimes pay more than the farmers receive for cereal or meat production.

One of the most widely used approaches to manage a population is to carry out a census. Such a survey needs to be related to other explanatory variables, like crops, rainfall, predators, and it needs to be repeated for different seasons, mainly before and after the breeding season. A census should extract a representative sample from a population, and there is a large number of techniques to minimize the sampling error or the bias due to heterogeneity of the population distribution (Borralho, Rego, and Vaz Pinto, 1996). In a practical case of a hunting game preserve the most important measure is the hunting success or number of hunted animals, made by hunters during a hunting day or several days during the hunting season. The DeLury method (DeLury, 1947) allows the prediction of population size by graphing kills per gun-hour as a function of cumulative kill. These methods are indicated to estimate population size but not the hunting success.

The ideal statistical model to describe the hunting success and compute all factors, independent variables and their interactions will be unreachable. First, measuring all the independent variables (e.g. grams of pesticides per hectare, hectares of grains and sunflowers, rainfall in mm, etc.) and factors (e.g. season, use of dog, dog breed, type of weapon) involved results in an undetermined model with more unknowns than records. Second, after a certain boundary, the relation between the variables and the response shows no or a negative effect. This happens for example in case of heavy rainfall, which leads to a poor breading season for rabbits (Palomares, 2003).

The focus of this paper is to establish a statistical model of "hunting success" defined as the number of hunted animals, which can be considered as one of the most interesting source of information in a game preserve. Since hunting success is directly related to counting the number of hunted animals, a Poisson model seems to be appropriate, using other external explanatory variables. One main contribution of this paper is the selection of an appropriate model among all different possible Poisson models. For this purpose we are proposing a repeated cross validation procedure based on the deviance measure of the goodness of fit. The resulting best model can be used to manage the game preserve by just integrating it during the days of hunting with different management criteria. We will investigate and test different statistical models to explain the hunting success, using the raw hunters´ surveys during different days and seasons, and distinguishing among different groups of hunters. An alternative way for identifying management criteria in a game

preserve using statistical evidence is to measure various influential variables, like weather, area of wheat or barley, grams of pesticides per hectare used by farmers, etc. This would lead to a much more complex treatment of the problem. However, our goal is to stick to a simple scheme that is only employing basic information of the hunters and their hunting success. This allows for an applicability of the resulting model with standard hunting survey data.

2   Material and Methods

2.1   Study area and data collection

The data were collected in the game preserve of Puebla de Almenara ($39° 47'$ N $2° 48'$ E), Cuenca, Spain. The game preserve consists of about 2500 hectares of steppe, mostly flat, at an altitude of about 800 m. The boundaries to the West are the hills of Sierra Jarameña which have a maximum altitude of 1054 m, following the direction North-South, with *Quercus ilex* and olive trees *Olea europaea*. At the East, also in North-South direction, there are the hills of Las Lomas and Los Picorzos. These are scrublands with around 900 m altitude. In the middle of the game preserve, there is a stream called Arroyo de la Vega. The climate is Continental Mediterranean. The main crops are grains (wheat and barley), sunflower and legumes with an intensive mechanization due to a land reploting in the 90's.

The game preserve of Puebla de Almenara had no management at the time when the records were collected. This means that no supplementary feed or water troughs were provided. Water troughs are an important management tool in game preserves where extremely dry periods are usual (Gaudioso Lacasa et al., 2010). No predator control was done.

The hunters were allowed to hunt free across the area, without limitations on the number of animals to hunt and places where to hunt. The records were collected during five seasons and with 13 different groups of hunters. One hunter per group was responsible to fill in a survey at the end of each hunting day. Every survey was personally hand delivered and collected to assess confidentiality of the results. At the end of the season each hunting group received a leaflet with an evaluation, where the (anonymized) groups were compared, and including a comparison with the previous season. Three species were hunted: red legged partridge (*Alectoris rufa*), rabbit (*Oryctolagus cuniculus*) and Iberian hare (*Lepus granatensis*).

The way of hunting consists in a front of hunters with or without dogs, which move forward. The idea behind this strategy is that a prey which tries to run away comes into the shooting area of one or more hunters, increasing the probability to be hunted. The speed of the hunters is variable: the hunters move faster in a flat terrain when the hunters are looking for the red legged partridge, and they are slower if the target is rabbits or hares.

From the surveys, 274 observations of the following variables are available:

- *Season*, contains the information of the season (corresponding to the hunting year) when data were collected. Five seasons were recorded: 1999/2000, 2000/2001, 2001/2002, 2002/2003 and 2003/2004.

- *Amigos*, the group of hunters that normally hunt together: 13 groups of hunters were involved but not everyone could participate from the first to the last season; labeled as A, B, ..., M.

- *Dia*, day of hunting, measured from the beginning of the legal small game season; typically from middle October to middle February; labeled as 1, 2, 3, etc., for day 1, 2, 3, etc..

- N, number of hunters of a group of hunters (*Amigos*) which hunted at one day (*Dia*).

- *Mano*, class variable derived from N: The class code is *N.3+* if three or more hunters are in one group, and *N.1-2* otherwise.

- *Perd*, the number of red legged partridges hunted by a number of N hunters of one group of hunters on one day.

- *Cone*, the number of rabbits hunted by a number of N hunters of one group of hunters on one day.

- *Lieb*, hunting success as number of hares hunted by a number of N hunters of one group of hunters on one day.

- *SumPCL*, hunting success as the sum of *Perd*, *Cone* and *Lieb* for one day for one group on hunters.

Table 1 shows statistical summary information of the numeric variables.

Table 1: Statistical summary information of variables

Variable	Mean	Standard deviation	minimum	Maximum
Perd	2.81	3.5	0	22
Cone	0.94	1.58	0	9
Lieb	0.35	0.71	0	4
SumPCL	4.1	4.13	0	25
N	2.31	1.11	1	8
Dia	5.56	3.34	1	15

2.2 Statistical analysis

The relationship between the explanatory and the response variables is modeled by a generalized linear model (GLM). The theory of GLMs is well known and described in many textbooks (McCulloch and Searle, 2001; Green and Silverman, 1994). GLMs can be seen as a derivation of a linear regression with the particularity of a general function which links the mean of the observed variable to the linear predictor. This link function depends on the probability distribution of the dependent variable. The dependent variables considered here are *Perd*, *Cone*, *Lieb* and *SumPCL*, i.e. variables containing count data. Accordingly, the probability distribution of the response variables is assumed as a Poisson distribution. The assumption of a Poisson distribution transforms the original GLM in a Poisson regression.

The formulation of the model is as follows: Suppose that the values $y_1, y_2, \ldots, y_n$ of a response are given. In our case, the response is either *Perd*, *Cone*, *Lieb*, or *SumPCL*. The random variable representing the ith value of the response variable is supposed to follow a Poisson distribution, and its expectation μ_i, or more precisely, the logarithm of μ_i, is modeled by

$$\log(\mu_i) = \beta_0 + x_i^T \beta \,, \tag{1}$$

where β_0 is the intercept, x_i is the vector with the values of the ith observation of the explanatory variables, and β the vector of the regression coefficients.

A challenge for Poisson regression to model the hunting success was the number of zeros in the data set, which had a high importance in pointing out days of null success. Table 2 gives the proportion of zeros for each species (*Perd*, *Cone* and *Lieb*) and their sum (*SumPCL*).

Table 2: Percentages of zeros included in the response variables

Variable	% of zeros
Perd	23
Cone	58.8
Lieb	74.1
SumPCL	11.3

In order to avoid this and other problems, different models were exhaustively tried and checked, like zero inflated Poisson and Negative binomial models (McCulloch and Searle, 2001). However, these models turned out to be not more useful than Poisson regression and we did not find either over-dispersion or zero inflated problems, therefore the Poisson distribution of the independent variable was retained.

2.3 Model selection

Model selection is frequently based on information criteria like the Akaike information criterion, AIC (Akaike, 1973). Here we decided to evaluate different models with repeated cross validation (Varmuza and Filzmoser, 2009). The $n = 274$ observations of the data set are randomly split into $k = 5$ groups of about equal size. A model is fitted to 4 groups (training data), and the response of the 5th group (test data) is predicted using the estimated coefficients. Thus, the observations of the 5th group can be considered as "future observations" for which we obtain predictions with our model. Each of the 5 groups is once playing the role of the test data, and thus finally n test-set predictions are available, coming from 5 fitted models. Since one random split may not give a very realistic picture for the prediction quality of the model, the whole procedure is repeated $m = 100$ times with different random splits into 5 groups, resulting in a matrix of predicted values $\hat{y}_{ij}$, $i = 1, \ldots, n; j = 1, \ldots, m$.

This procedure allows for a realistic model evaluation, because the actually observed values y_i of the response can be compared with their predictions. Note that a simple correlation coefficient would not be useful for this comparison, since the response variable is categorical. Therefore we use the *deviance* as a quality measure of the model, which is

defined as

$$
d_{ij} = \begin{cases} \delta_{ij} \sqrt{2 \left[y_i \cdot \log \left(\dfrac{y_i}{\hat{y}_{ij}} \right) - (y_i - \hat{y}_{ij}) \right]} & \text{if } y_i \neq \hat{y}_{ij} \\ \delta_{ij} \sqrt{2 \hat{y}_{ij}} & \text{if } y_i = 0 \,, \end{cases} \tag{2}
$$

where δ_{ij} is $+1$ if $y_i > \hat{y}_{ij}$ and -1 otherwise (McCulloch and Searle, 2001). The measure $D_j = \sum_{i=1}^{n} d_{ij}^2$ evaluates the jth prediction, and as overall prediction quality criterion we use the median D_{med} over all values D_j, for $j = 1, \ldots, m$. In addition, the MAD (Median Absolute Deviation) D_{MAD} of the values D_j serves as a measure of spread, and $S.E. (D_{med}) = D_{MAD}/\sqrt{m}$ is the standard error of D_{med}. Thus, $D_{med} \pm 2 \cdot S.E. (D_{med})$ is an approximate 95% confidence interval for D_{med}. The use of these robust measures for the model evaluation protects against outliers in the deviance measure, which can result from single cross validation runs.

For each response variable, *Perd*, *Cone*, *Lieb*, and *SumPCL*, the predictor variables N, *Dia*, *Mano*, *Amigos*, and *Season* are considered. In addition, a lag variable for the each specific response is used, describing the number of hunted animals from the previous day. For the first day of the season, the value of the lag variable is assigned the value of the corresponding first day from the previous season for a specific group of hunters. Thus, for each considered response variable, up to 6 explanatory variables can be used in the model. Since we are interested in obtaining a good prediction model, all combinations of the 6 explanatory variables (these are 64 combinations) are tested, using the above repeated cross validation procedure. This is feasible and it takes only a couple of minutes on a standard computer for each response variable.

The statistical analysis and the data management were done using the statistical software environment R (R Development Core Team, 2012).

3 Results

3.1 Results for *Perd*

The response variable *Perd* gave the best results. After analyzing all 64 different models with the repeated cross validation procedure we sorted according to the resulting dispersion measure D_{med}. Table 3 lists the best ten models. 0/1 in this table refers to exclusion/inclusion of the respective variable in the model. It can be seen that the quality of the listed models is somehow comparable.

For example, using the previously defined confidence interval for D_{med}, the best model is not significantly better than the second best. Nevertheless, it is interesting that some explanatory variables, like *Dia*, are consistently appearing in the models. In the following we will focus on the best model, but one has to keep in mind that also other models could be considered for the explanation of hunting success of *Perd*. The deviance measure is useful for comparing different models, but the number itself does not tell as much as for example an R^2 measure as used for a usual linear model. In the context of a GLM, the R^2 measure is not recommended because of the categorical response. Instead, a pseudo R^2 measure can be used, which results again in a number between 0 and 1 (Long

Table 3: Best ten models for *Perd*: 0 and 1 refer to exclusion and inclusion of the variable in the model. $C.I.\,(D_{med})$ refers to the approximative 95% confidence interval for D_{med}.

N	*Dia*	*Mano*	*Amigos*	*Season*	*Perd*-lag	D_{med}	$C.I.\,(D_{med})$	Pseudo R^2
1	1	0	1	0	0	508.71	$[505.85, 511.57]$	0.551
1	1	0	1	1	0	511.35	$507.07, 515.63]$	0.566
0	1	1	0	1	1	511.82	$[509.80, 513.84]$	0.504
1	1	1	0	1	1	514.61	$[512.79, 516.43]$	0.510
1	1	0	1	0	1	514.92	$[510.74, 519.10]$	0.556
1	1	1	1	0	1	515.56	$[510.22, 520.90]$	0.558
1	1	0	0	1	1	517.08	$[515.42, 518.74]$	0.502
1	1	0	1	1	1	517.84	$[512.96, 522.72]$	0.570
1	1	1	1	0	0	518.56	$[514.58, 522.54]$	0.551
1	1	1	1	1	0	519.11	$[514.57, 523.65]$	0.567

and Freese, 2006). The pseudo R^2 is defined as $1 - D_{Res}/D_{Null}$, where D_{Res} is the deviance of the residuals and D_{Null} is the null deviance using only an intercept model, see Long and Freese (2006) for details. The last column in Table 3 reports the pseudo R^2 values for the best models. The difference in the values for the different models is only marginal, but the overall size of the pseudo R^2 tells us that the models are useful.

According to Table 3, the best model for the response variable *Perd* includes the explanatory variables N, *Dia*, and *Amigos*. This outcome is logical because the number and group of hunters, as well as the day in the season might be important for explaining the success to hunt this animal. To some extent, the day of the season might reflect the information of the lagged variable *Perd*-lag, containing the number of hunted *Perd* from the previous day. Also the variable *Mano* contains similar information as N and *Amigos*. The inference statistics for this best model is shown in Table 4.

Table 4: Inference statistics for the best model of the response *Perd*

Variable	Estimate	Std. Error	z value	p-value
Intercept	0.727	0.170	4.280	< 0.001
N	0.433	0.051	8.463	< 0.001
Dia	-0.203	0.014	-14.158	< 0.001
Amigos-B	-0.050	0.148	-0.338	0.735
Amigos-C	-1.441	0.392	-3.671	< 0.001
Amigos-D	0.725	0.131	5.521	< 0.001
Amigos-E	0.342	0.137	2.484	0.013
Amigos-F	0.400	0.185	2.133	0.033
Amigos-G	0.292	0.185	1.578	0.115
Amigos-H	-0.489	0.181	-2.697	0.007
Amigos-I	-0.908	0.407	-2.233	0.026
Amigos-J	0.886	0.190	4.658	< 0.001
Amigos-K	-0.121	0.203	-0.595	0.552
Amigos-L	-0.208	0.256	-0.812	0.417
Amigos-M	-0.788	0.333	-2.364	0.018

According to Equation (2), the residual deviance for this model is 431.81, compared

to the null deviance 961.43. This shows that our model is a considerable improvement compared to an "empty" model, but it also shows that with the evaluation done above the median residual deviance D_{med} of 508.71 shows a clearer and more realistic picture. Both N and *Dia* are numerical variables, and they are highly significant in the model. N has a positive coefficient, meaning that on average the larger groups are more successful. The coefficient for *Dia* is negative, i.e. in the first days of the season the hunting success is higher than later in the season. Several categories of Amigos are significant, meaning that the whole factor has a significant contribution. Obviously, there are more successful or less successful groups. Moreover, some of the groups are larger, and they are expected to be more successful – which they often cannot fulfill. For example, group *Amigos-C* was hunting with 7 or 8 people, and their success was 14 *Perd* in total. The other groups were much more successful, even with a smaller number of hunters.

An analysis of variance (ANOVA), shown in Table 5, reflects the same findings of Table 4. All variables in the models are highly significant.

Table 5: ANOVA test for the best model of the response *Perd*

Name	Df	Deviance	Resid Df	Resid Dev	p-value
NULL			273	961.43	
N	1	89.959	272	871.47	< 0.001
Dia	1	294.211	271	577.26	< 0.001
Amigos	12	145.446	259	431.81	< 0.001

Since the model deviance does not really give a clear impression of the quality of the model in terms of "correct" prediction, we want to compare the "observed" number of hunted *Perd* according to the given data with a "predicted" number using our best model from above. However, we do not directly predict from the available data, but we use "test set" predictions to obtain less optimistic but more realistic numbers. In more detail, from the above repeated cross validation we already generated 100 test sets. Each of these sets is taken for prediction with the best model, resulting in 100 predictions of the number of hunted *Perd* for each of the $n = 274$ observations. Finally, the average over the 100 predictions is used in the presentation of Figure 1, where the predicted values are rounded to integers. This so-called sunflower plot visualized multiple observations on one point with special symbols, the sunflowers. The number of leaves of the sunflowers represents the frequencies of the observations in one point (Murrel, 2005). A clear relation between observed and predicted values is visible, indicating that the model is well suitable for prediction purposes.

Although the statistical model for *Perd* seems to be satisfactory, it is not clear if it is suitable for a management of the population. In this case it is important that the total number of hunted animals can be estimated at a specific time point. So far it is not clear whether the statistical model is suitable for this purpose. However, the day *Dia* of the season is included in the model, and thus the time axis is considered. In order to check the validity of the model for the management purpose, we sort the observation of the number of hunted and predicted *Perd* according to time. Note that for the predictions we use again the average test-set predictions as before. Figure 2 shows the comparison of estimated and predicted cumulative numbers of *Perd*, considering the correct time axis for
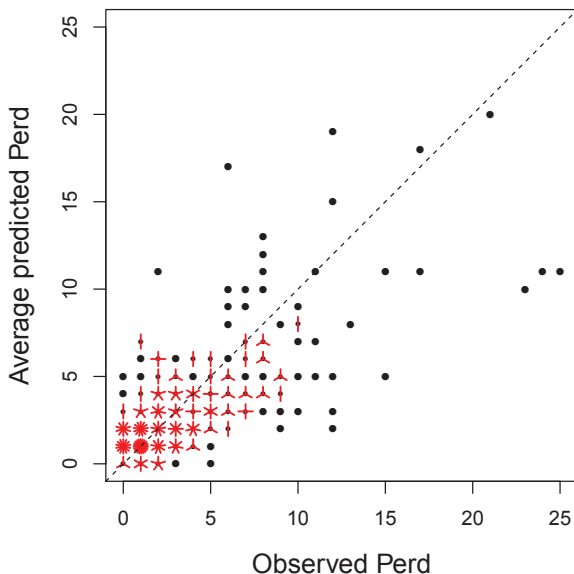
Figure 1: Number of hunted *Perd* (observed) versus average number of *Perd* predicted from the best model.

the observations. For an effective management system we would like to have a reliable prediction of the total hunted animals at each time point. Obviously, the model is very well suitable for this purpose. The maximum difference is 34 animals (overestimation), occurring in the season 2001/2002. Here, the prediction was done for the overall period. It could easily be changed to seasonal prediction, where the precision would be similarly successful.

3.2   Results for *Cone*, *Lieb*, and *SumPCL*

The models for the response variables *Cone* and *Lieb* are not as useful as that for *Perd*. We performed the same procedure for model selection as above. The final best model for *Cone* selects the explanatory variables *Amigos* and the lagged *Cone* variable. The next best model selects only *Amigos*, and it is not significantly different from the previous. For *Lieb* we obtain for the best model the explanatory variables N and *Lieb*-lag. Taking these best models for prediction, we obtain – in analogy to Figure 1 – the plots in Figure 3. While the model for *Cone* still seems to be useful, the model for *Lieb* is not of any use.

Indeed, from a management point of view, the model for *Cone* is still suitable, as is shown in Figure 4. The maximum difference between observed and predicted values is 19 animals. The results for *Lieb* are not useful and thus not shown here.

Since the number of hunted *Lieb* is rather small compared to the other species, the sum of these hunted animals will not be dominated by *Lieb*. Consequently, it is possible to model the response *SumPCL*, but the resulting model will not be more effective than a single model for one species. Therefore we do not report the results for *SumPCL*, and recommend using the specific models for *Perd* and *Cone*.
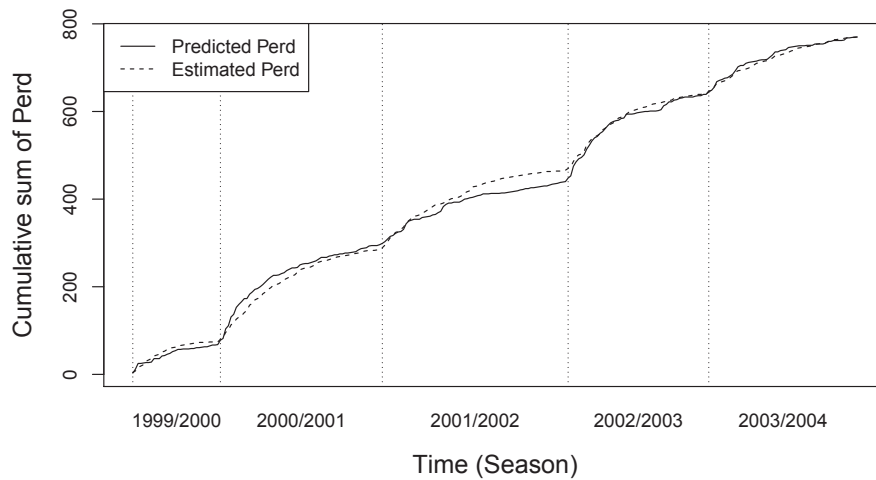
Figure 2: Number of hunted *Perd* (observed) versus average number of *Perd* estimated from the best model.
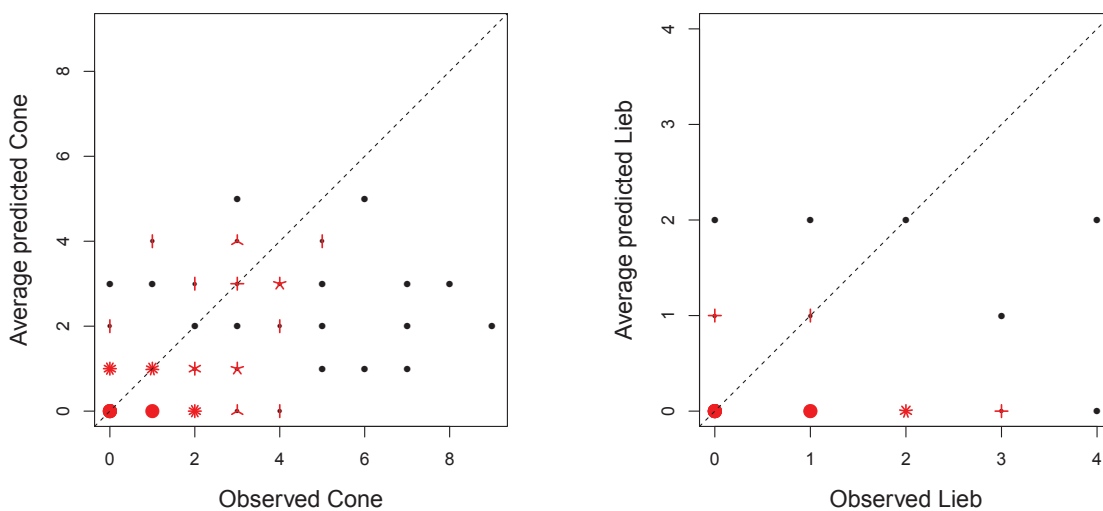

Figure 3: Observed versus predicted values of *Cone* (left) and *Lieb* (right).

4   Discussion

4.1   Red legged partridge (*Alectoris rufa*)

The models proposed to describe the hunting success in this article showed two important properties, first the description of the process in an easy way, and second the ability to predict the hunting success in the future. The model uses N (the number of hunters), *Dia* (the day of hunting), and *Amigos* (the group of hunters that normally hunt together), variables that are somehow in the common understanding of the red legged partridge hunting techniques. These variables are easy to compute and consider. This is an advantage in the practical application of the model, because for a game preserve manager this information is easily available. The ability of the model to predict future hunting success of *Perd* is
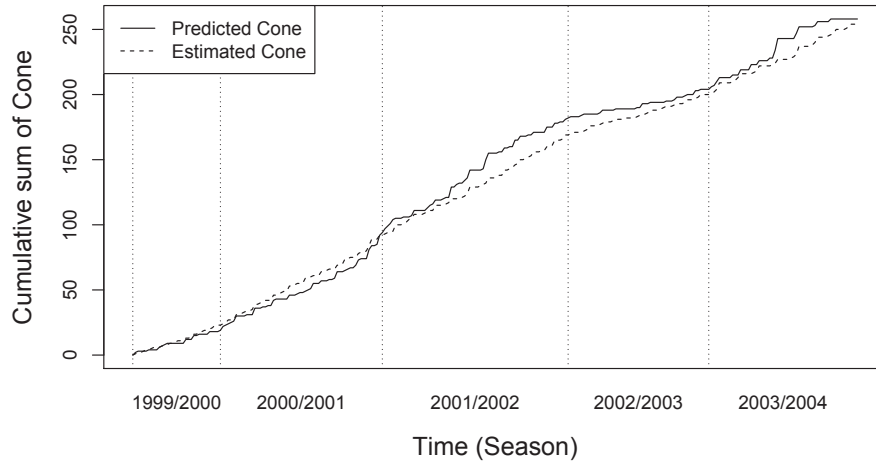
Figure 4: Number of hunted *Cone* (observed) versus average number of *Cone* estimated from the best model.

shown in Figure 1 and Figure 2. The model appears as a reliable tool for a sustainable management of a game preserve. Using the estimated values of the variables, a formula which depends only on time (*Dia*) will describe the hunting success.

Based on the information provided for the estimated model, we propose some scenarios to manage the population relying on hunting success. Different criteria could be applied to find a threshold moment, t_m, when the hunt will not be sustainable any more. We have chosen three of them as follows, when the hunters will not hunt anything, when the number of hunted animals is equal to the remained ones and when the hunters keep alive a certain proportion, α, of the total possible *Perd*. Looking at equation (1), we can write the best model as

$$\mu_i = \exp\left(\beta_0 + \beta_1 \cdot N_i + \beta_2 \cdot Dia_i + \beta_3 \cdot Amigos_i\right) \tag{3}$$

for an observation with index i. Since the estimated coefficient for *Dia* is negative (see Table 4), the predicted value is getting smaller and smaller as the day increases. In this point the probability to have values close to zero increases drastically. In particular, it may be interesting to know when the expected value is below 1, which means that less than one animal is to be expected. The day when this occurs can be estimated from the model easily, using the estimated coefficients from Table 4. For example, for the group of *Amigos E* with two or three hunters we get the expected values shown in Figure 5 as curves (the observed data are shown as points). In the case of two hunters (Figure 5, left) the boundary of 1 is reached after day 9 (t_m), and for three hunters (Figure 5, right) it is reached after day 11 (t_m). Stopping after day 9 the hunting season in this game preserve will thus be reasonable.

The criterion can also be used to find a day, t_m, where the number of hunted animals is equal to the remaining ones,
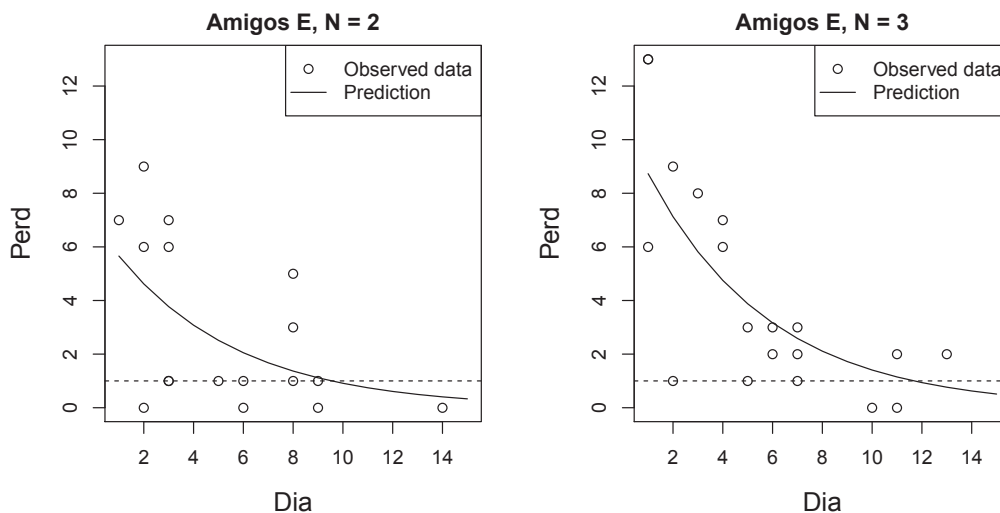
$$\sum_{i=1}^{t_m} Perd_i = \sum_{t_m+1}^{T} Perd_i \,,$$

Figure 5: Number of predicted *Perd* (using the best model) along time (*Dia*) for two different numbers of hunter in group *Amigos E*. The points are the observed values.

e.g. with a total number of hunting days, $T = 15$ days. The estimated results from Figure 5 give a value of $t_m = 3$. At this day the hunters will keep alive the same number of *Perd* as they have hunted. The use of the third management criterion is straight-forward: Using a proportion α, a number between 0 and 1, we can be interested in $\sum_{i=1}^{t_m} Perd_i = \alpha \sum_{t_m+1}^{T} Perd_i$. This will calculate the day when the hunters keep alive a certain proportion, α, of the total possible *Perd*. For the example in Figure 5, a proportion $\alpha = 1/3$ will result in $t_m = 6$. The three proposed criteria point a range of threshold moments, t_m, between day 3 and day 11, after that the hunt will not be sustainable.

4.2 Wild rabbit (*Oryctolagus cuniculus*) and Iberian hare (*Lepus granatensis*)

The models for these species were not as successful as in the case of the red legged partridge, mainly because the way of hunting of these two mammals is totally different. The use of dogs could be a key factor that was not recorded in the data and thus it is not included in the model. The variable *Amigos* could take into account a part of this information, but detailed information about the use of dogs could be more successful for modeling.

5 Conclusions

Our results show that it is possible to develop powerful models to predict hunting success. The use of a Poisson model gives important information for the management of a game preserve. Based on a careful model selection procedure, useful models in terms of accuracy and prediction were obtained for the red legged partridge. The variables *Amigos*, the group of hunters, N, the number of hunters and *Dia*, day of hunting, explain a

large amount of variability in the models. The presented approach has a great predictive ability and it is easy to implement. The difficulties found with hare and rabbit should be interpreted as a lack of important variables in the model which determine the population dynamics. On the other hand, this study provides guidance for other reserves with respect to variables to be considered as possible predictors. Different threshold moments, t_m, when the hunt will not be sustainable any more are provided based on the proposed model. Further research should contribute to use the result of a census in the model to explain more variability, which will increase the accuracy and prediction ability. Another research line could be to search for variables that improve the models for rabbit and hare. Potentially interesting variables could be pesticides per hectare, agriculture intensification methods, which account for the human impact in the environment.

Acknowledgments

References

Akaike, H. (1973). Information theory as an extension of the maximun likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory* (p. 267-281). Budapest: Akademiai Kiado.

Bernabéu, R. (2000). *Evaluación económica de la caza en Castilla La Mancha*. Unpublished doctoral dissertation, IREC. Universidad de Castilla La Mancha, Ciudad Real, Spain.

Borralho, R., Rego, F., and Vaz Pinto, P. (1996). Is driven transect sampling suitable for estimating red-legged partridge *alectoris rufa* densities? *Wildlife Biology*, 2, 259-268.

Calvete, C. (2006). Modeling the effect of population dynamics on the impact of rabbit hemorrhagic disease. *Conservation Biology*, 20, 1232-1241.

Delibes-Mateos, M., Delibes, M., Ferreras, P., and Villafuerte, R. (2008). The key role of European rabbits in the conservation of the western mediterranean basin hotspot. *Biodiversity Conservation*, 22, 1106-1117.

DeLury, D. (1947). On the estimation of biological populations. *Biometrics*, 3, 145-167.

Gaudioso Lacasa, V., Sánchez García-Abad, C., Prieto Martín, R., Bartolomé Rodríguez, D., Pérez Garrido, J., and Alonso De La Varga, M. (2010). Small game water troughs in a spanish agrarian pseudo steppe: visits and water site choice by wild fauna. *European Journal of Wildlife Research*, 56, 591-599.

Green, P., and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman and Hall.

Long, J. S., and Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata* (2nd ed.). College Station, TX: Stata Press.

Martínez, J., Viñuela, J., and Villafuerte, R. (2002). *Socio-economic aspects of game-bird hunting, hunting bags, and assessment of the status of gamebird populations in reghab countries. part 1: Socio- economic and cultural aspects of gamebird hunting*. Ciudad Real, Spain: Instituto de Investigaci ón en Recursos Cinegéticos IREC-CSIC.

McCulloch, C., and Searle, S. (2001). *Generalized, Linear and Mixed Models*. New York: Wiley.

Mozumder, P., Meghan Starbuck, C., Berrens, R., and Alexander, S. (2007). Lease and fee hunting on private lands in the U.S.: A review of the economic and legal issues. *Human Dimensions of Wildlife*, *12*, 1-14.

Murrel, P. (2005). *R Graphics*. Boca Raton, FL: Chapman and Hall/CRC Press.

Palomares, F. (2003). The negative impact of heavy rains on the abundance of a mediterranean population of european rabbits. *Mammal Biology*, *68*, 224–34.

R Development Core Team. (2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org/` (ISBN 3-900051-07-0)

Varmuza, K., and Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics*. Boca Raton, FL: CRC Press.

Virgós, E., and Travaini, A. (2005). Relationship between small-game hunting and carnivore diversity in central Spain. *Biodiversity Conservation*, *14*, 3475-3486.

Authors' addresses:

Jose Carlos Martinez Avila and Peter Filzmoser
Department of Statistics and Probability Theory
Vienna University of Technology
Wiedner Hauptstr. 8-10
1040 Vienna, Austria

E-mail: `jcmartinezavila@gmail.com`, `P.Filzmoser@tuwien.ac.at`

Neyko M. Neykov
National Institute of Meteorology and Hydrology
Bulgarian Academy of Sciences
66 Tsarigradsko chaussee
1784 Sofia, Bulgaria

E-mail: `Neyko.Neykov@meteo.bg`