Recovery of Incomplete Data of Statistical Livestock Number Applying an Entropy Approach

António Xavier
PhD student, Algarve University and CEFAGE (Center for Advanced Studies in Management and Economics) amxav@sapo.pt
Maria de Belém Martins
Assistant Professor, Algarve University and CEFAGE (Center for Advanced Studies in Management and Economics) mbmartins@ualg.pt
Rui Fragoso
Assistant Professor, Évora University and CEFAGE (Center for Advanced Studies in Management and Economics) rfragoso@uevora.pt

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

RECOVERY OF INCOMPLETE DATA OF STATISTICAL LIVESTOCK NUMBER APPLYING AN ENTROPY APPROACH

1. INTRODUCTION

Disaggregated statistical information is necessary to the correct analysis of spatial patterns inside each country. However, data on agricultural and forest occupation and production is frequently found only at national and sub-national level (You et al. 2006; You et al. 2007; Chakir 2009; Fragoso et al. 2008), and many times this problem doesn't have an appropriate and accurate solution.

In Portugal, beside the general lack of data on agricultural and forest occupation, there is a need for up-to-date data on livestock numbers (Fragoso et al. 2008). Only the Agricultural Census (AC), conducted every 10 years, features information at disaggregated level by sub-region, municipality and parish. The other information sources display information merely according to Agrarian Regions and NUTS II (INE 2006, 2007, for instance).

The planning and devising of a clear and sustainable rural development policy calls for the availability of disaggregated information (You et al. 2007).

In Alentejo Region at South Portugal the importance of livestock breeding activity through the years is unquestionable (DRAPAl, 2007). However, this region suffered the influence of different policies and as a consequence there are several rural areas with problems and an extensification of agricultural activities (Carvalho and Godinho, 2004).

A particular municipality where there is a tendency towards demographic decrease is Castelo de Vide located, in the NUTS II Alto Alentejo. Here, there is a need for data on extensive livestock for breeding to enable a correct analysis of the current state of affairs and the implementation of development policies.

A methodology which aims to obtain disaggregated information on livestock numbers would be relevant and it is what we propose in this paper, focusing especially on the forecast of the main categories of breeding livestock and total livestock numbers.

The key question here is to be able to adapt land use disaggregation studies, based on entropy, to allow, not only a good added value to data calculation, but also a dynamic calculation process of the livestock data, which we think will provide good results for solving the problem and will allo the possibility of predicting future trends.

The study was applied to Alentejo Region being livestock disaggregated and recovered, in a first step for the sub-region of Alto Alentejo and then in a second step for its municipality of Castelo de Vide (Fig.1).

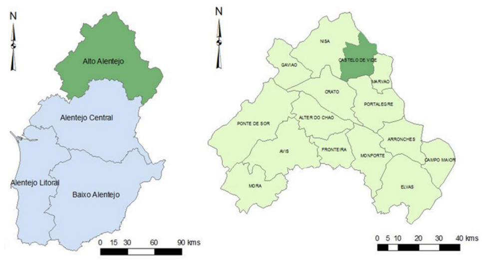

Fig. 1- The model's application area

2. THE DISAGGREGATION PROBLEM

The objective of the model resented in this paper is to disaggregate data on the main livestock categories at municipality level for 2005 and also to obtain a historical sequence of data.

So, for addressing this problem it must be taken into account that in Alentejo the livestock number is closely linked with the agricultural and forest area, as the livestock is raised mainly in an extensive way, fed by pastures, seeded or not, under trees or not.

The information from the disaggregated units (municipalities and sub-regions) is available only for the first r periods (r < T), that is 1989 and 1999. Therefore, the information disaggregation problem can be formulated as follows:

$$E^{x}{}_{P}(t) = \sum_{B_i \subset A_x} E^{i}_{P}(t) \tag{1}$$

The aim is to obtain from the variable $E^{x}{}_{p}(t)$, which is known, the data in E^{i}_{p} (t), which is the target variable (p) at disaggregate level in unit (i). The matrix A_x represents the aggregated unit x, in which $A_{x,\ x=1...X}$. It is therefore necessary to determine $E^{i}_{p}\ \forall p,\ i\ e\ t$, for the sub-units B_i at the moment (t), of which there is no available information at disaggregate level.

As it was said before, this problem is also connected to the agricultural and forest occupation at disaggregate level ($S^{i}_{k}(t)$) or, in its absence, to the data at aggregate level ($S_k(t)$). The aim is to figure out the livestock distribution at disaggregate level, knowing that it is a function of agricultural and forest occupation in that moment:

$$E^{i}{}_{P}(t) = f\,[S^{i}_{k}(t)1 \ldots S^{i}_{k}(t)n] \tag{2}$$

where $E^{i}{}_{P}(t)$ represents the livestock numbers p in unit i, at the moment t and $S^{i}_{k}(t)$ is the agricultural or forest occupation k in unit i, at the moment t. Additionally it is assumed that this is a direct relation, with the land uses, namely pastures, being the possibility of a complex relation excluded.

These livestock numbers are determined by the following:

$$E^{i}{}_{P}(t) \leq Cv\ ^{i}_{k}(t)\ \forall i, k, t = r+1,\ldots,T \tag{3}$$

that is, the composition of the livestock must respect the pre-defined rules of heading (Cv) as a function of agricultural and forest occupation k at the moment t, i.e. for the total physical area necessary to raise these livestock.

These values will also be subjected to a restriction described as follows:

$$E_P(t) = \sum_{i=i}^{I} E^{i}_{P}(t),\ \forall p, t = r+1,\ldots,T \tag{4}$$

This restriction demands that the numbers of a certain livestock category at regional /aggregate level equal the sum of that category in each disaggregated territorial unit. That is, the sums of livestock numbers p in units i at the moment (t) must equal the respective aggregated value.

3. THE ESTIMATION OF TOTAL LIVESTOCK NUMBERS IN NH

Admitting that livestock numbers are a function of farms' land occupation and considering that it is possible to estimate the coefficients that indicate this relation, to determine the total livestock numbers it is necessary that data resulting from a model for agricultural and forest occupation estimation is available.

We should assume that each livestock category has different requirements in terms of feeding and, therefore, the necessity for a wider or smaller area to that effect. The conversion table of different livestock categories into normal heads (NH) is therefore a useful tool. As an example, a sheep over 1 year will equal 0.15 NH, while a bovine over 2 years will equal 1 NH. In order to better understand this measure, table one presents the main conversion indexes for the different species according to their age.

Table 1- The indexes of conversion into normal heads

	Livestock type	Conversion index
	Bovines over 2 years	1
Bovine	Other bovines	0,6
	Reproducer females>50kg	0,5
Suines	Other suines >20kg	0,3
	Sheep	0,15
	Goats	0,15
	Equines	1

(source: GPAA)

We also assume that each municipality has a determined relation between NH and the agricultural and forest occupation. So,the number of breeding livestock p in NH, in unit i, at the moment t can be calculated in the following way:

$$ENH_p^i(t) = E_p^i(t) \times INH_{\text{p}} \tag{5}$$

where INH is the conversion index from livestock p into NH (the NH equivalent). On the other hand, the relation between livestock numbers in NH and agricultural and forest occupation is determined by:

$$\text{R}_{\text{pk}} = (\sum_{p=1}^{P} ENH_p^i(t)) \div S_k^i \tag{6}$$

in which R is the relation between total livestock numbers p and k agricultural and forest occupation in territorial unit i. These values can then be transferred to a period t+1 according to the following formula:

$$sENH^i(t+1) = \text{R}_{\text{pk}} \times S_k^i(\text{t}+1) \tag{7}$$

It is also possible to define sustainability limits on the number of NH per agricultural and forest occupation, with specific additional information.

4. THE ESTIMATION OF LIVESTOCK PERCENTAGE WEIGHT

Until this point, the methodology proposed allow us to estimate the total number of breeding livestock in (NH), but it does not allow us to calculate the percentage weight of each category. This leads us to the need of finding a methodology which estimates this kind of data. International studies were therefore considered, namely Howitt and Reynaud (2002, 2003) and Fragoso et al. (2008), which applied an entropy methodology considering Markov processes in order to estimate the land uses weight. Since it was a study which implied the consideration of these methodologies, one considered also the developments that took place regarding maximum entropy, namely Good (1961) and Golan et al. (1996).

Therefore, and because we are adapting the methodology proposed by Howitt and Reynaud (2002, 2003) and Fragoso et al. (2008), we have to convert the data of each livestock breeding categories into NH, for the years in which information is available; disaggregate the data from a database created at aggregate level, based on the theory of maximum entropy and calculate the livestock weight in t+1…T; redistribute livestock numbers in NH according to estimated proportions and convert into animal numbers.

This way we assume that a sequence of livestock can be characterized by a first-order Markov process, considering P^r decision states corresponding to P^r possible strategies, indexed by $j \in \{1,...J\}$ with $J=P^r$. Assuming a second-order Markov process, the probability of moving from any decision state $j \in \{1,...,p^r\}$ in the year t-1 to any decision state $j' \in \{1,...,p^r\}$ in the year t can be calculated by $ENHpr_p^i(t-1) \times ENHpr_p^i(t)$. This probabilities' product can be associated to a $T_{jj'}^i$ matrix of ($P^r \times P^r$) dimension, which is called probability transition matrix (Howitt and Reynaud 2002, 2003; Fragoso et al. 2008).

Therefore, the creation of the information prior at the aggregated level is presented now. The dynamic process of livestock percentual weight estimation in NH, at aggregate level, is obtained based on the estimation of the probability transition matrix at aggregate level $T_{jj'}$, using the generalized maximum entropy (GME) approach (Golan et al. 1996). To fulfil this objective the following ME model, inspired in Howitt and Reynaud (2002) and Fragoso et al. (2008), was adapted to the aims of this work and the Alentejo conditions:

$$\underset{T,e}{Max}\, \mathrm{H(T,e)} = -\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{m=1}^{M} T_{jj'm}.\log(T_{jj'm}) - \sum_{j'=1}^{J}\sum_{n=1}^{N}\sum_{t=r}^{T-1} e_{j'n}(t).\log(e_{j'n}(t)) \quad (8)$$

Subject to:

$$Q_{j'}(t+1) = \sum_{j=1}^{J} Q_j(t).\sum_{m=1}^{M} w_m.T_{jj'm} + \sum_{n=1}^{N} v_n.e_{j'n}(t), \quad \forall \text{j' and t} \tag{9}$$

$$\sum_{j'=1}^{J} \sum_{m=1}^{M} w_m.T_{jj'm} = 1 \qquad \forall \text{j} \tag{10}$$

$$\sum_{m=1}^{M} T_{jj'm} = 1 \quad \text{and Tjj'} \in [0,1] \quad \forall \text{j and j'} \tag{11}$$

$$\sum_{n=1}^{N} e_{j'n}(t) = 1 \quad \text{and } \mathrm{e}_{\mathrm{j'n}} \in [0,1] \quad \forall \text{j' and t} \tag{12}$$

This optimization program aims to maximize the entropy of the probability distribution $\{T_{jj'1},...,T_{jj'M}\}$ $\forall$j and j' and $\{\mathrm{e}_{j'1},...,e_{j'N}\}$ $\forall$j' and t , taking into account the conditionals determined by restrictions. Equation (9) defines the dynamic process of livestock numbers in NH. Equation (10) determines that the sum of transition probabilities in any Markov state is equal to 1. In the same way, equations (11) and (12) guaranty that the values of the variables of $\{T_{jj'1},...,T_{jj'M}\}$ and $\{\mathrm{e}_{j'1},...,e_{j'N}\}$ are defined as a probability distribution, and between 0 and 1, and so respecting the formalism of Golan et al. (1996).

According to Golan *et al.* (1996) it is necessary to define a parameter $w' = \{\mathrm{w}_1,...,\mathrm{w}_\mathrm{M}\}$ of M≥ 2 points with z_1=0 and z_M=1 and with a $\{\mathrm{T}_{jj'1},...\mathrm{T}_{jj'M}\}$ probability distribution. For errors $e_{j'}(t)$ we proceed in the same way re-parametrizing the error through a support vector $\mathrm{v}' = \{\mathrm{v}_1,...,\mathrm{v}_\mathrm{N}\}$ with N ≥ 2 points, so that errors are then defined in $\{\mathrm{e}_{j'1}(\mathrm{t}),...,\mathrm{e}_{j'N}(\mathrm{t})\}$. So, an example of such re-parameterization is presented next:

$$e_{j'}(t) = \sum_{n=1}^{N} v_n.e_{j'n}(t) \tag{13}$$

Knowing the T_{jj} probability distribution we can easily create a database of information at aggregate level, which will later be used in the estimation process of probability transition matrixes at disaggregate level for recovery livestock numbers.

The formalization of the first step in the disaggregation process can be given through the following cross-entropy minimization program where the previously estimated transition matrix is used as an information prior:

$$\underset{T^i,e}{Min}\, \mathrm{H}(\mathrm{T}^i,\mathrm{e}) = \sum_{\mathrm{i}=1}^{\mathrm{I}} \sum_{j=1}^{J} \sum_{j'=1}^{J} T_{jj'}^i.\log(T_{jj'}^i / \widehat{T}_{jj'}) + \sum_{\mathrm{p}=1}^{\mathrm{P}} \sum_{n=1}^{N} \sum_{t=r+1}^{T} e_{pn}(t).\log(e_{pn}(t)) \tag{14}$$

Subject to:

$$ENH_p(t+1) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{j'\in\psi} \mathrm{q}_\mathrm{j}^\mathrm{i}(t).T_{jj'}^i(t).sNH^i + \sum_{n=1}^{N} \zeta_n.e_{pn}(t) \forall \mathrm{p} = 1,\ldots \mathrm{P} \tag{15}$$

and

$$\sum_{j'=1}^{J} T_{jj'}^{i} = 1 \text{ and } T_{jj'}^{i} \in [0,1] \quad \forall \text{j and j'} \tag{16}$$

$$\sum_{n=1}^{N} e_{pn}(t) = 1 \quad \text{and } e_{pn}(t) \in [0,1] \quad \forall \text{p and t} \tag{17}$$

in which, $\{\zeta_1,...,\zeta_N\}$ with N ≥ 2 points is the support vector associated to the $\{e_{p1},...,e_{pN}\}$ probabilities.

This optimization model minimizes the cross-entropy of transition probability distribution and entropy of probability distribution of errors (14) subjected to restrictions (15)-(17).

Equation (15) shows a data consistency restriction. This equation guaranties the compatibility of information at aggregate level with disaggregated information, and simply states that the sum of all the disaggregated values plus the error must equal the aggregate. The last two equations assure us that $T^{i}_{jj'}$ and e_{pn} respect the properties of a probability distribution, that is, their sum is equal to 1. The last one is the error re-parameterized, according to the ideas expressed by Golan et al. (1996).

To overcome the enormous lack of data in almost disaggregated units a direct disaggregation of data regarding the Castelo de Vide municipality can be done, simply rewriting equation (15) as follows:

$$ENHpr_p(t+1) = \sum_{j=1}^{J} \sum_{j' \in \psi} q_j^i(t).T_{jj'}^{i} + \sum_{n=1}^{N} \zeta_n.e_{pn}(t) \quad \forall \text{p and t} \tag{18}$$

where $ENHpr_p(t+1)$ is the livestock percent weight of the aggregate p. This is useful because it enables us to address situations in which we only have data on some territorial units comprising an aggregate or because the errors resulting from the disaggregation process of agricultural and forest occupation don't estimation of livestock numbers in other municipalities into question. Therefore, this materializes a second variant of the model.

From the estimations of matrix $\widehat{T}_{jj'}^{i}$ one may reproduce the percentage of livestock numbers p, in NH at the moment t+1:

$$ENHpr_p^i(t+1) = \sum_{j=1}^{J} \sum_{j' \in \psi} q_j^i(t).T_{jj'}^{i}(t) \tag{19}$$

5. THE ESTIMATION OF LIVESTOCK NUMBERS

Assuming that the number of NH in relation to land agricultural and forest occupation has already been calculated, one must simply make its redistribution into the NH percentage weight calculated:

$$ENH_p^i(t+1) = ENHpr_p^i(t+1).sNH^i \tag{20}$$

Afterwards, the number of NH can easily be converted into real number of animals by means of the inverted use of each conversion index. So:

$$E_p^i(t+1) = ENH_p^i(t+1) / INH_p \tag{21}$$

in which INH is the conversion index of livestock effectives p into NH.

Finally, if we want to estimate total livestock numbers, the data conversition in NH cannot be made due to the NH coeficients limits. If a set of explanatory variables is available it can be used and will solve this problem. If not, one may take the premise that the year variation rate follows the livestock intended for breeding rate , and so:

$$\vec{\nabla} Et_p^i = \left(\left(\frac{ENH_p^i(t+x)}{ENH_p^i}\right)^{(\frac{1}{x})} \right) - 1 \tag{22}$$

in which $\vec{\nabla} Et_p^i$ is the year variation rate of total livestock p, x the number of years.

This method is not needed in our context, since we have a stable weight of each of the units to be disaggregated. Therefore we may assume the premise that the model may be applied in a simultaneous disaggregation process, and so there is only the need of using the entropy process presented above and calculate the livestock numbers of specie p for unit i:

$$E_p^i(t+1) = Epr_p^i(t+1).E_p^{\;i} \tag{23}$$

6. RESULTS AND DISCUSSION

The model proposed was implemented in two stages, admitting the use of only one direct disaggregation procedure in the agricultural and forest occupation for the different Alto Alentejo's municipalities. Nevertheless, its application to Portuguese data implied some adaptations. The first is linked with the lack of data on the classes of conversion into NH in the AC of 1989. Therefore, the application of this methodology to the total amount of breeding livestock determines that it can only be done after 1999. The second point has to do with the existence of limited data in certain classes of land occupation, which determines the impossibility of a correct simultaneous disaggregation process for the Alto Alentejo municipalities and allows only the application of a direct disaggregation process to breeding livestock. A third consideration is related to the fact that conversion into NH cannot be made when calculating the total number of livestock (including younger animals), due to the conversion limits. However, similar methodologies can be used to calculate the percentage weight of each animal and, supposing that the total number of animals evolution follows the same pattern than breeding livestock, to calculate the total number of estimated livestock.

The following variants of the model were used: 1) Direct disaggregation of breeding livestock for Castelo de Vide considering 2 years (2003 and 2005), 2) Direct disaggregation of breeding livestock for Castelo de Vide considering a historical sequence, 3) Simultaneous disaggregation of the total number of livestock in the sub-regions and 4) Direct disaggregation of the total number of animals' proportions and their further conversion into number of animals for Castelo de Vide.

In the application of the model's different variables the support limits of parameters and error had to be defined. Golan *et al.* (1996) recommended the use of Pukelsheim's three-sigma rule to establish the limits of error components. So, considering previous studies, it was established that w can be defined to M = 3 points, being therefore considered $w = \{0, 0.5, 1\}$ for all models.

In the estimation process for NH, regarding breeding livestock in 2005, the limits assumed to the disaggregation level of Alto Alentejo and Castelo de Vide were $v = \{-1,0,1\}$. For the other breeding livestock disaggregation model $v = \{-2,0,2\}$ was considered at the first disaggregation level and $v = \{-5,0,5\}$ at the second disaggregation level.

On the procedure of recovering historical series and total livestock numbers of all categories the limits $v = \{-1,0,1\}$ in the direct disaggregation process and $v = \{-0.7,0.7\}$ in the simultaneous disaggregation process were considered.

The main livestock categories considered in Castelo de Vide for the first application variant were bovine cattle, sheep and goats (Table 2). However, the need to convert breeding livestock into NH has led to the consideration of different divisions regarding bovine cattle: bovines of 1 to 2 years of age and bovines over 2 years of age.

Table 2-Estimated livestock intended for breeding in Castelo de Vide

Animal / year	2003	2005
BO 1 to 2 yrs	342	371
BO over 2 yrs	2833	2862
Sheep	7111	6625
Goats	857	808

(source: model results)

For partial estimations of breeding livestock aiming the historical reconstruction of predominant livestock, one considered t=7 years and only cows over 2 years of age (COT) in regard to bovine cattle. Livestock numbers were calculated, in NH, partially using data from 1989 (Table 3).

Table 3- Recovery Historical Data of livestock intended for breeding in Castelo de Vide

Animal / year	1993	1995	1997	1999	2003	2005
COT	2223	2337	2395	2498	2529	2585
Sheep	8763	8364	7641	7888	7389	6659

(source: model results)

The variant of the model which allowed the disaggregation of data simultaneously for the different sub-regions considered the following livestock types: cows over 2 or more years of age (COT), other bovines (OB), sheep (SH) and goats (GOA) and a series of t=7 years (Table 4). It allowed a correct recover the numbers of the different types of livestock.

Table 4-The number of livestock estimated for each one of the Alentejo's sub-regions

Sub-reg	Lvst//year	1993	1995	1997	1999	2003	2005
AL	COT	28268	31141	35006	41437	42275	46558
	OB	21644	25202	21911	29734	36036	35631
	SH	179618	180408	181575	200462	185408	165832
	GOA	17372	15101	15929	14918	10777	8314
AA	COT	37397	41563	46753	55957	58030	64842
	OB	27587	32405	28190	38676	47641	47793
	SH	304412	308444	309809	345219	323374	291845
	GOA	39252	34429	36343	34420	25274	19780
AC	COT	49929	55082	62253	74023	76205	85054
	OB	39271	45789	40022	54548	66705	66846
	SH	421675	424054	426898	471786	437254	392328
	GOA	25420	22124	23458	22061	16071	12555
BA	COT	23052	25764	29119	35162	37077	42378
	OB	18182	21479	18776	25993	32561	33415
	SH	445558	454103	457018	513064	486988	447089
	GOA	34769	30672	32535	31097	23226	18603

AL- Alentejo litoral sub-region; Alto Alentejo sub-region; Alentejo Central sub-region; Baixo Alentejo sub-region.

(source: model results)

On the other hand, the variant of the model intending to recover the total proportions of livestock considered cows over 2 or more years of age (COT), other bovines (OB), sheep (SH) and goats (GOA) and a series of t=7 years (Table 5 and Table 6).

Table 5- Proportion of total livestock numbers estimated in Castelo de Vide

Animal/year	1993	1995	1997	1999	2003	2005
COT	0.128	0.134	0.148	0.15	0.154	0.177
OB	0.075	0.083	0.073	0.085	0.105	0.111
Sheep	0.655	0.665	0.658	0.667	0.669	0.653
Goats	0.142	0.117	0.121	0.097	0.071	0.059

(source: model results)

Table 6- Livestock numbers in Castelo de Vide

Animal /year	t2003	t2005
COT	1934	2128
OB	1319	1334
Sheep	8402	7849
Goats	892	709

(source: model results)

The results' validation was made based on cross reference with statistical data, using the opinion of experts and technicians with a good knowledge of the area as well as other sources of information. Of all the measures to analyze deviations from real statistics data, the Weighted Prescription Absolute Deviation (WPAD) stands out:

$$\mathrm{WPAD}_{\mathrm{p}}^{\mathrm{i}} = y_p^i \left| \frac{y_p^i - \widehat{\mathrm{y}}_{\mathrm{p}}^{\mathrm{i}}}{\mathrm{y}_{\mathrm{p}}^{\mathrm{i}}} \right| \quad \text{and} \quad \mathrm{WPAD}^{\mathrm{i}} = \sum_{p=1}^{p} WPAD_p^i \tag{24}$$

and at aggregate level, according to the aggregation of disaggregated units:

$$\mathrm{WPAD} = \sum_{\mathrm{i}=1}^{\mathrm{I}} \frac{\mathrm{S}^{\mathrm{i}}}{\mathrm{S}} . \mathrm{WPAD}^{\mathrm{i}} \tag{25}$$

The $\mathrm{W}PAD_{\mathrm{p}}^{\mathrm{i}}$ expresses the true importance of the percentual deviation in each livestock category regarding the observed values ($\mathrm{PAD}^{\mathrm{i}}_{\mathrm{p}}$), by weighting these by the true values and the $\mathrm{WPAD}^{\mathrm{i}}$ corresponds to the sum of these values giving the idea of the real total deviation for the values of the unit i. The WPAD corresponds to the weighted sum of the $\mathrm{WPAD}^{\mathrm{i}}$ by the weight of each unit i regarding the total value or aggregate.

The procedures for the estimation of breeding livestock in Castelo de Vide (first application variant) shows tendencies that can be observed in the municipality. On the other hand, the model aiming to estimate part of the breeding livestock (second application variant) was validated in 1999, in comparison to the data from the AC of 1999. This analysis showed that the model produced satisfactory data, since we obtained a $\mathrm{WPAD}^{\mathrm{i}}$ of around 14%, with only the data on breeding goats being not valid.

The simultaneous disaggregation process of total livestock effectives was only applied to the Alentejo's sub-regions, as presented before. However, these values are very close to the reality, since the WPAD is only 6%. In terms of $\mathrm{WPAD}^{\mathrm{i}}$ only Alentejo Central stands out with a 8.11 percentage and the percentage in Alto Alentejo is of about 3.16. Since it is a simultaneous disaggregation procedure it was also possible to measure the amount of heterogeneity of the information recovered through the "Disaggregation Informational Gain" (DIG) used by Howitt and Reynaud (2002).

This measure is based on the cross entropy between the observed values of breeding livestock at aggregated and disaggregated level and on the cross entropy between the breeding livestock estimated by the model and the observed values. In case of a perfect disaggregation, the DIG is equal to 1. The formula is presented next:

$$DIG = 1 - \frac{\sum_{i=1}^{\mathrm{I}} \sum_{k=1}^{K} \widehat{y}_k^i . \ln \frac{\widehat{y}_k^i}{y_k^i}}{\sum_{i=1}^{\mathrm{I}} \sum_{k=1}^{K} y_k . \ln \frac{y_k}{y_k^i}} = 1 - \frac{\widehat{\mathrm{CE}}}{\mathrm{CE}} \, or DIG = 1 - \frac{\sum_{i=1}^{\mathrm{I}} \sum_{k=1}^{K} \widehat{y}_k^i . \ln \widehat{y}_k^i - \sum_{i=1}^{\mathrm{I}} \sum_{k=1}^{K} \widehat{y}_k^i . \ln y_k^i}{\sum_{i=1}^{\mathrm{I}} \sum_{k=1}^{K} y_k . \ln y_k - \sum_{i=1}^{\mathrm{I}} \sum_{k=1}^{K} y_k . \ln y_k^i} \tag{26}$$

Therefore, the DIG obtained revealed a percentage of heterogeneity of recovered information around 0.603 which means the level of recovered information was very satisfactory over passing the levels obtained by Fragoso et al. (2008) for the Alentejo area.

The direct disaggregation procedures have shown a reliable recovery of the proportions of total livestock effectives in Castelo de Vide. The PAD^i_p values for Castelo de Vide are very low for sheep, which is the predominant livestock category. The biggest percentage of PAD^i_p regards bovine cattle: cows with 2 or more years of age (30%). All the other categories show PAD^i_p values under 15%. This leads to acceptable $WPAD^i$ values for Castelo de Vide around 14.6%.

7. CONCLUDING REMARKS

The methodology used enables us to solve the problem of lack of data regarding the main categories of breeding livestock with satisfactory results. It was possible to demonstrate that dynamic methodologies based on Howitt and Reynaud (2002are flexible enough to be applied, with adaptations, to other types of data and still obtain satisfactory results.

The above model was only applied using a direct disaggregation process for the Castelo de Vide county, in result of the lack of data of land uses. The availability of this information, linked with livestock, would be useful, as it would imply a better development of the land use disaggregation model, but it also may give the chance of developing a combined one considering simultaneous disaggregation of livestock raised in an extensive way and the main land uses.

With new data available, we envisage the development of this methodology in a more complex way, with the inclusion of new databases, since it allows obtaining relevant information regarding animal production management and thus better decisions. This methodology has therefore an important role, considering the fact that regulations may influence the livestock breeding activity leading to the subsequent decrease of pastures and to the increase of shrubs and forests. Knowing better the real situation will provide us the ability of better understanding the consequences according to different scenarios.

8. References

DRAPAl- Direcção Regional de Agricultura e Pescas do Alentejo (2007): Programa de Desenvolvimento Rural-Alentejo (Alentejo Rural Development Programme). DRAPAL.

Carvalho, M. and Godinho, M. (2004). A Nova reforma da Política Agrícola Comum e suas consequências num sistema agrícola mediterrâneo de Portugal. Available at http://www.sober.org.br/ (accessed 15 January 2009).

Chakir R. (2009): Spatial downscaling of agricultural land use data: an econometric approach using cross–entropy". Land Economics, 85(2): 238–251.

Fragoso, R., Martins, M.B., and Lucas, M.R. (2008): Generate disaggregated soil allocation data using a Minimum Cross Entropy Model. WSEAS Transaction on Environment and Development, 9 (4): 756-766.

Golan, A., Judge, G. and Miller, D. (1996): Maximum Entropy Econometrics: Robust Estimation with Limited Data. John Wiley & Sons, NewYork.

Good, I. (1963): Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. The Annals of Mathematical Statistics, 34 (3): 911-934.
Howitt, R and Reynaud, A. (2002): Spatial Disaggregation of production data by maximum entropy. Proccedings of X Congress of the European Association of Agricultural Economists "Exploring the diversity in the European Agri-Food-System", 28-31 of August 2002. European Association of Agricultural Economists, Saragoça, Spain.
INE-Instituto Nacional de Estatística (2001): Recenseamento geral da agricultura de 1999 (General Agriculture Census). INE, Lisbon.
INE-Instituto Nacional de Estatística (2006): Inquérito à estrutura das explorações agrícolas. INE, Lisbon.
INE-Instituto Nacional de Estatística (2007): Estatísticas agrícolas. INE, Lisbon.
Jongeneel, R. (2002): An Analysis of the Impact of Alternative EU Dairy Policies on the Size Distribution of Dutch Dairy Farms: an Information Based Approach to the Non- Stationary Markov Chain Model. Proceedings of X Congress of the European Association of Agricultural Economists "Exploring the diversity in the European Agri-Food-System" 28-31 of August, 2002. European Association of Agricultural Economists, Saragoça, Spain.
Karantininis, K. (2001): Information Based Estimators for the Non–stationary Transition Probability Matrix: An Application to the Danish Pork Industry. Unit of Economics Working Papers, The Royal Agricultural University of Copenhagen.
Martins, M. B., Fragoso, R., Xavier, A. (2010): Spatial disaggregation of agricultural data in Castelo de Vide, Alentejo, Portugal: an approach based on maximum entropy. J.P. Journal of Biostatistics (paper accepted for publication).
Tonini, A. and Jongeneel, R. (2007): Modelling the Dairy Farm Size Distribution in Poland Using an Instrumental Variable Generalized Cross Entropy Markov Approach. Paper prepared for presentation at IAAE - 104th EAAE Seminar Agricultural Economics and Transition: "What was expected, what we observed, the lessons learned" September 6-8, 2007. European Association of Agricultural Economists, Corvinus University of Budapest (CUB), Budapest.
You, L. and Wood, S. (2004): Assessing the spatial distribution of crop production using a cross-entropy method. EPTD discussion paper Nº 126, International Food Policy Research Institute (IFPRI). Washington DC.
You, L., Wood, S. and Wood-Sichra, U. (2006): Generating Global Crop Distribution Maps: From Census to Grid. Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference. August 11-18, International Association of Agricultural Economists, Gold Coast, Australia.
You, L., Wood, S., Wood-Sichra, U. and Chamberlin, J. (2007): Generating Plausible Crop Distribution Maps for Sub-Saharan Africa Using a Spatial Allocation Model. Information Development, 23 (2/3): 151-159.