

## Research Article

# Disaggregation of Statistical Livestock Data Using the Entropy Approach

António Xavier,<sup>1</sup> Maria de Belém Costa Freitas,<sup>1</sup> and Rui Fragoso<sup>2,3,4</sup>

<sup>1</sup> Faculdade de Ciências e Tecnologias, Universidade do Algarve, Edifício 8, 8005-139 Faro, Portugal

<sup>2</sup> Universidade de Évora (UE), Centro de Estudos e Formação Avançada em Gestão e Economia Tecnologias, 7000-809 Évora, Portugal

<sup>3</sup> Department of Management, Universidade de Évora, 7000-809 Évora, Portugal

<sup>4</sup> Instituto de Ciências Agrárias e Ambientais Mediterrâneas, Universidade de Évora, 7000-809 Évora, Portugal

Correspondence should be addressed to António Xavier; [amxav@sapo.pt](mailto:amxav@sapo.pt)

Received 30 January 2014; Accepted 2 May 2014; Published 3 June 2014

Academic Editor: Konstantina Skouri

Copyright © 2014 António Xavier et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A process of agricultural data disaggregation is developed to address the lack of updated disaggregated data concerning main livestock categories at subregional and county level in the Alentejo Region, southern Portugal. The model developed considers that the number of livestock units is a function of the agricultural and forest occupation, and data concerning the existing agricultural and forest occupation, as well as the conversion of livestock numbers into normal heads, are needed in order to find this relation. The weight of each livestock class is estimated using a dynamic process based on a generalized maximum entropy model and on a crossentropy minimization model, which comprises two stages. The model was applied to the county of Castelo de Vide and their results were validated in cross reference to real data from different sources.

## 1. Introduction

Disaggregated statistical information is necessary to have a correct analysis of spatial patterns inside each country, but data on agricultural and forest occupation and production are frequently found only at national and subnational level [1–4], and oftentimes this problem does not have an appropriate and accurate solution.

In Portugal, southern Europe, apart from the general lack of data on agricultural and forest occupation, there is a need for up-to-date data on livestock numbers [2]. Only the General Agricultural Census (GAC), conducted by the National Institute of Statistics (NIS) every 10 years, features information at disaggregated level by subregion, county, and parish. The other information sources display information merely according to agrarian regions and NUTS II [5, 6]. Even the different agents operating in the territory do not have more detailed data, and only the health veterinary entities have some more accurate data for nowadays, which is not accessible to all.

However, planning and devising of a clear and sustainable rural development policy call for the availability of disaggregated information [4], at least when it comes to the numbers of livestock intended for breeding, mainly in regions where these variables have a great importance for the farmers' income.

After the entrance of Portugal into the European Union, the Alentejo region, southern Portugal, has come under the influence of different policies and as a consequence there are several inland rural areas with problems and in decline, as well as the extensification of agricultural activities [7]. In this region, the importance of livestock breeding activity is unquestionable [8].

A particular county where there is a tendency towards demographic decrease is Castelo de Vide located in the NUTS II of Alto Alentejo. Several teams from the University of Évora have performed depth studies and identified this county as a case study about the decline process of the inland counties of the NUTS II of Alto Alentejo. Here, there is a need for data on extensive livestock for

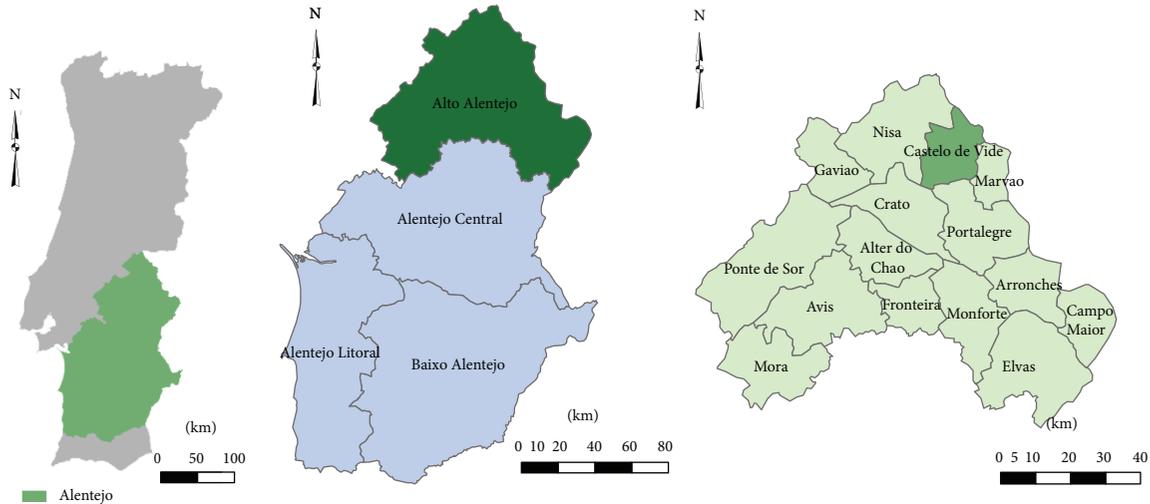


FIGURE 1: The model's application area.

breeding to enable a correct analysis of the current state of affairs and the implementation of development policies, which may contribute to the increase of the farms' income.

Therefore, methodologies for estimating livestock data are relevant, namely, in rural areas where there is a tendency towards the decline of agricultural activity, livestock breeding and local economy. Thus, a methodology to obtain disaggregated information on livestock units is proposed in this paper, focusing especially on the forecast of the main categories of livestock units intended for breeding, as well as total livestock units.

In order to develop such a methodology, one has considered the previous studies carried out and different methodologies that may be adapted to this area and which are quite few. Fragoso et al. [2] presented a way of calculating the main livestock species intended for breeding. The authors used a conversion process of the livestock into normal heads and followed the evolution of the aggregate based on the historical pattern. Thus, they simply made the data follow the aggregate variation rate with some adaptations. This procedure is very simple but may lead to some mistakes. Other studies carried out for the Alentejo region, namely, by the state entities, do not use a suitable disaggregation level or present incomplete information.

In spite of the above referred, there are some studies carried out for land use disaggregation based on entropy in the Alentejo Region, which allowed a good disaggregation of information [2, 9]. There are also some international studies that use methodologies based on the entropy approach for estimating livestock data. Some of these are presented next. Jongeneel [10] used a nonstationary Markov model approach, where the transition probabilities are explained by a set of exogenous (policy) variables. Karantininis [11] applied a generalized cross entropy instrumental variables estimator to recover the nonstationary transition probability matrix for the Danish pork industry. Tonini and Jongeneel [12] analyzed the evolution of the dairy farm structure of Poland during the postsocialist period and applied a generalized

cross entropy Markov chain approach which incorporated prior information for estimation.

Therefore, the key question is to be able to adapt such studies in order to allow not only a good added value to data calculation, but also a dynamic calculation process of the livestock data, since there are some indicators of a possibility of this methodology providing good results for solving the problem, and allowing the possibility of predicting future trends.

The study was applied to the Alentejo Region where livestock were disaggregated, in the first step for the subregion of Alto Alentejo and then in the second step for its county of Castelo de Vide. The following maps show the location of the application area (Figure 1).

In addition to this introduction, the paper is organized in more three parts. The second part includes the material and methods, which describe the disaggregation problem and the methodological approach developed for estimating the number of livestock units. The third part is concerned with the presentation and discussion of results. Finally, the concluding remarks are included in the last part.

## 2. Material and Methods

*2.1. The Disaggregation Problem.* The intent is to obtain disaggregated data on the main livestock categories at county level for 2005 and also to obtain a historical sequence of data. So, for addressing this problem it must be considered that in the Alentejo region the number of livestock units is closely linked with the agricultural and forest areas, since the livestock are breeding mainly in extensive systems based on pastures under Mediterranean forest, seeded or spontaneous.

The information from the disaggregated units (counties and subregions) 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:

$$E_p(t) = \sum_{B_i \subset A_x} E_p^i(t). \quad (1)$$

The aim is to obtain from the known variable  $E_p^x(t)$  the data in  $E_p^i(t)$ , which is the target variable related to the livestock activities ( $p$ ) at the disaggregated level ( $i$ ) in a given moment ( $t$ ). The matrix  $A_x$  represents the aggregated unit  $x$  and it is necessary to determine  $E_p^i \forall p, i$  and  $t$ , for the subunits  $B_i$  at the moment ( $t$ ), for which there is no available information at the disaggregated level ( $i$ ).

As it was referred before, the number of livestock units is connected to the agricultural and forest occupation at the disaggregated level ( $S_k^i(t)$ ) or, in its absence, to the data at the aggregate level ( $S_k(t)$ ). The aim is to find the livestock distribution at disaggregated level, knowing that it is a function of agricultural and forest occupation in that moment:

$$E_p^i(t) = f[S_k^i(t)] \quad \text{with } t = 1, 2, \dots, T, \quad (2)$$

where  $E_p^i(t)$  represents the number of livestock units  $p$  in territorial unit  $i$  at the moment  $t$  and  $S_k^i(t)$  is the agricultural or forest occupation  $k$  in territorial unit  $i$  at the moment  $t$ . Therefore, it is assumed that this is a direct relation with the land uses, namely, pastures. However, the number of livestock units is bounded by the following relation:

$$E_p^i(t) \leq Cv_k^i(t) \quad \forall i, k, t = r + 1, \dots, T. \quad (3)$$

According to that, the composition of livestock must respect the predefined rules of heading ( $Cv$ ) as a function of agricultural and forest occupation  $k$  at the moment  $t$ , that is, for the total physical area necessary to breeding livestock. These values should be also subjected to the following restriction:

$$E_p(t) = \sum_{i=i}^I E_p^i(t), \quad \forall p, t = r + 1, \dots, T. \quad (4)$$

This restriction demands that the number of livestock units from a certain class at an aggregate level must be equal to the sum of that class in all disaggregated territorial units. That is, the sum of the number of livestock units  $p$  in all disaggregated units  $i$  at the moment  $t$  must be equal to the respective aggregated value.

**2.2. The Estimation of Livestock's Normal Heads.** The number of livestock units is a function of farms' land occupation and hence there are coefficients that indicate this relation considering that each livestock class has different requirements in terms of feeding and, therefore, the necessity for a wider or smaller area. The conversion of different livestock categories into normal heads (NH) is a useful tool for considering those aspects.

Normal head is a common livestock measure that allows dealing in the same units with different categories of livestock in function of species and ages. There is a legal table of conversion (Portaria no. 229-A/200 of 6 March 2008, [5]). For instance, according to that table a sheep over 1 year is equal to 0.15 NH and a bovine over 2 years is equal to 1 NH. In order to better understand this measure, Table 1 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
Bovine	
Bovines over 2 years	1
Other bovines	0.6
Suines	
Reproducer females >50 kg	0.5
Other suines >20 kg	0.3
Sheep	0.15
Goats	0.15
Equines	1

(Source: GPAA).

Each territorial unit (region, subregion, and county) has a well-determined relation between NH and the agricultural and forest occupation. This relation is obtained calculating the number of NH for each predominant livestock class in the area and establishing the relation between the number of NH and agricultural and forest occupation, namely, forage crops and permanent pastures. So, the number of effective livestock intended for breeding  $p$  in NH, in territorial unit  $i$ , at the moment  $t$  can be calculated as follows:

$$ENH_p^i(t) = E_p^i(t) \times INH_p, \quad (5)$$

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

$$R_{pk} = \left( \sum_{p=1}^P ENH_p^i(t) \right) \div S_k^i, \quad (6)$$

where  $R$  is the relation between total livestock and  $k$  agricultural and forest occupation in territorial unit  $i$ . These values can then be transferred to a period  $t + 1$  as follows:

$$sENH^i(t + 1) = R_{pk} \times S_k^i(t + 1). \quad (7)$$

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

**2.3. The Estimation of Livestock's Percentage Weight by Class.** Until this point, the methodology proposed allows us to estimate the total number of breeding livestock in (NH), but it does not allow us to calculate the percentage weight of each class. This leads us to search for a methodology that is able to estimate this type of data. Howitt and Reynaud [13, 14] and Fragoso et al. [2] applied entropy methodology considering Markov processes in order to estimate the land uses weight. The studies about maximum entropy of Good [15] and Golan et al. [16] and the ones related to Markov processes and maximum entropy previously referred to livestock were also considered.

Based on the methodology of Howitt and Reynaud [13, 14] and Fragoso et al. [2], we propose an analytical framework developed in the following three steps.

- (1) Convert the data of each livestock breeding class into NH, for the years in which information is available.
- (2) Disaggregate the data from a database created at aggregate level, based on the theory of maximum entropy and calculate the livestock weight in  $t + 1, \dots, T$ .
- (3) Redistribute livestock numbers in NH according to estimated proportions and convert into animal numbers.

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  $j \in \{1, \dots, J\}$  with  $J = P^r$ . Assuming a second-order Markov process, the probability of moving from any decision state  $j \in \{1, \dots, P^r\}$  in the year  $t - 1$  to any decision state  $j' \in \{1, \dots, P^r\}$  in the year  $t$  can be calculated by multiplying the respective probabilities  $ENHpr_p^i(t - 1) \times ENHpr_p^i(t)$ . This product of probabilities can be associated with a  $T_{jj'}^i$  matrix of  $(P^r \times P^r)$  dimension, which is called probability transition matrix [2, 13, 14].

Therefore, the creation of the prior information at the aggregated level is presented now. The dynamic process of livestock's percentage 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 [16]. To fulfil this objective the following maximum entropy model, inspired in Howitt and Reynaud [13] and Fragozo et al. [2], was adapted to the aims of this work and the Alentejo region conditions:

$$\begin{aligned} \max_{T,e} \quad & H(T, e) = - \sum_{j=1}^J \sum_{j'=1}^J \sum_{m=1}^M T_{jj'm} \cdot \log(T_{jj'm}) \\ & - \sum_{j'=1}^J \sum_{n=1}^N \sum_{t=r}^{T-1} e_{j'n}(t) \\ & \cdot \log(e_{j'n}(t)), \end{aligned} \quad (8)$$

$$\begin{aligned} \text{Subject to:} \quad & Q_{j'}(t + 1) = \sum_{j=1}^J Q_j(t) \cdot \sum_{m=1}^M w_m \cdot T_{jj'm} \\ & + \sum_{n=1}^N v_n \cdot e_{j'n}(t), \quad \forall j', t, \end{aligned} \quad (9)$$

$$\sum_{j'=1}^J \sum_{m=1}^M w_m \cdot T_{jj'm} = 1 \quad \forall j, \quad (10)$$

$$\sum_{m=1}^M T_{jj'm} = 1, \quad T_{jj'} \in [0, 1] \quad \forall j, j', \quad (11)$$

$$\sum_{n=1}^N e_{j'n}(t) = 1, \quad e_{j'n} \in [0, 1] \quad \forall j', t. \quad (12)$$

This optimization program aims to maximize the entropy of the probability distribution  $\{T_{jj'_1}^i, \dots, T_{jj'_M}^i\} \forall j$  and  $j'$  and  $\{e_{j'_1}, \dots, e_{j'_N}\} \forall j'$  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, (11) and (12) guaranty that the values of the variables of  $\{T_{jj'_1}^i, \dots, T_{jj'_M}^i\}$  and  $\{e_{j'_1}, \dots, e_{j'_N}\}$  are defined as a probability distribution, and between 0 and 1, thus respecting the formalism of Golan et al. [16].

According to Golan et al. [16] it is necessary to define a parameter  $w' = \{w_1, \dots, w_M\}$  of  $M \geq 2$  points with  $z_1 = 0$  and  $z_M = 1$  and with a  $\{T_{jj'_1}^i, \dots, T_{jj'_M}^i\}$  probability distribution. For errors  $e_{j'}(t)$  we proceed in the same way reparameterizing the error through a support vector  $v' = \{v_1, \dots, v_N\}$  with  $N \geq 2$  points, so that errors are then defined in  $\{e_{j'_1}(t), \dots, e_{j'_N}(t)\}$  as follows:

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

Therefore, 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 disaggregated level for calculating the number of livestock units.

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 prior information:

$$\begin{aligned} \min_{T^i, e} \quad & H(T^i, e) \\ & = \sum_{i=1}^I \sum_{j=1}^J \sum_{j'=1}^J T_{jj'}^i \cdot \log\left(\frac{T_{jj'}^i}{\hat{T}_{jj'}}\right) \\ & + \sum_{p=1}^P \sum_{n=1}^N \sum_{t=r+1}^T e_{pn}(t) \cdot \log(e_{pn}(t)), \end{aligned} \quad (14)$$

$$\begin{aligned} \text{Subject to:} \quad & ENH_p(t + 1) \\ & = \sum_{i=1}^I \sum_{j=1}^J \sum_{j' \in \Psi} q_j^i(t) \cdot T_{jj'}^i(t) \cdot sNH^i \\ & + \sum_{n=1}^N \zeta_n \cdot e_{pn}(t) \quad \forall p = 1, \dots, P, \end{aligned} \quad (15)$$

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

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

in which  $\{\zeta_1, \dots, \zeta_N\}$  with  $N \geq 2$  points is the support vector associated with the  $\{e_{p1}, \dots, e_{pN}\}$  probabilities of the error vector.

This optimization model minimizes the cross-entropy of transition probability distribution and the entropy of probability distribution of errors (14) subjected to restrictions (15)–(17). Equation (15) is a data consistency restriction that guaranties the compatibility of information at aggregate level with disaggregated information and states that the sum of all the disaggregated values plus the error must be equal to the aggregate. The last two equations assure that  $T_{jj'}^i$  and  $e_{pn}$  respect the properties of a probability distribution; that is, their sum is equal to 1.

In our case, at the county disaggregation level, the aim is to obtain livestock data specifically for the county of Castelo de Vide. Thus, instead of going for the simultaneous disaggregation of information in all counties of Alto Alentejo's NUT III, one may also choose the direct disaggregation of data regarding the county of Castelo de Vide. In order to do that, one may rewrite (15) as follows:

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

where  $ENHpr_p(t+1)$  is the livestock's percentage weight of the aggregate  $p$ . This is useful because it enables addressing situations in which we only have data on some territorial units comprising an aggregate and can materialize a second variant of the model proposed.

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

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

**2.4. The Estimation of Livestock Number.** Assuming that the number of NH in relation to land agricultural and forest occupation has already been calculated for having livestock number, one may simply make its redistribution into the NH percentage weight calculated as follows:

$$ENH_p^i(t+1) = ENHpr_p^i(t+1) \cdot sNH^i. \quad (20)$$

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

$$E_p^i(t+1) = \frac{ENH_p^i(t+1)}{INH_p}, \quad (21)$$

where  $INH$  is the conversion index of livestock effective  $p$  into NH.

Finally, if we want to estimate total livestock number, the data conversion in NH cannot be made due to the NH

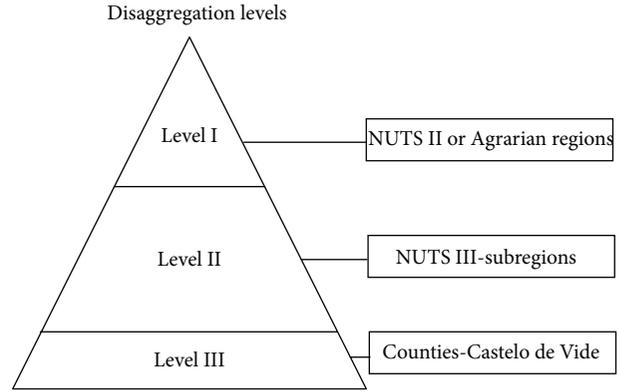


FIGURE 2: The territorial disaggregation levels.

coefficient limits. If a set of explanatory variables is available, it can be used and it will solve this problem. If not, one may take the premise that the year variation rate follows the livestock intended for breeding rate as follows:

$$\vec{\nabla} E_t^i = \left( \left( \frac{ENH_p^i(t+x)}{ENH_p^i} \right)^{(1/x)} \right) - 1, \quad (22)$$

in which  $\vec{\nabla} E_t^i$  is the year variation rate of total livestock  $p$  and  $x$  is the number of years after  $t$  year.

This method is based on strong assumptions and its application requires the opinion of experts. It should be applied only when we do not have a stable weight for each of the units to be disaggregated. In our case we may assume that the model may be applied in a simultaneous disaggregation process, where there is only the need to use the entropy process presented above and calculate the livestock numbers of specie  $p$  for unit  $i$ , as follows:

$$E_p^i(t+1) = Epr_p^i(t+1) \cdot E_p^i. \quad (23)$$

### 3. Results and Discussion

In order to obtain data on livestock for the county of Castelo de Vide, it was necessary to implement the disaggregation process in two stages. Figure 2 presents the disaggregation levels from I (for which there is available information), to II (subregions), and to III (the counties' level).

The application of the proposed methodology to Portuguese data implied also some adaptations.

The first one is due to the almost complete lack of data on the classes of conversion into NH in the GAC of 1989, which makes the application of the methodology be only done correctly after 1999, despite the fact that its application to the 1990s can be made with partial inclusion of livestock intended for breeding.

The second point is the existence of limited data in certain classes of land occupation. This makes it impossible to carry out a correct simultaneous disaggregation process for the Alto Alentejo's counties and allows for only the application of a direct disaggregation process to the breeding livestock.

TABLE 2: Estimated livestock intended for breeding in Castelo de Vide.

Animal/year	2003	2005
Bovines of 1 to 2 years	342	371
Bovines over 2 years	2833	2862
Sheep	7111	6625
Goats	857	808

(Source: model results).

A third consideration is related to the conversion into NH. The relation between livestock breeding and the agricultural and forest occupation cannot be made when calculating the total number of livestock units (including younger animals) due to the conversion limits. However, similar methodologies can be used to calculate the percentage weight of each animal, supposing that the evolution of the total number of animals follows the same pattern more than breeding livestock.

Given these considerations, the following implementation lines of the model were used: (1) direct disaggregation of breeding livestock for the county of Castelo de Vide considering 2 years (2003 and 2005), (2) direct disaggregation of breeding livestock for the county of Castelo de Vide considering a historical sequence, (3) simultaneous disaggregation of the total number of livestock in the subregions (NUTS III) of the Alentejo region, and (4) direct disaggregation of the total number of animals' proportions and their further conversion into number of animals for the county of Castelo de Vide.

In the application of the model different support limits of parameters and error had to be defined. Based on previous studies, it was established the parameter  $w$  can be defined with  $M = 3$  points, being therefore considered as  $w = \{0, 0.5, 1\}$  for all models. In order to establish the limits of error components, the Pukelsheim's three-sigma rule recommended by Golan et al. [16] was used and we selected the value of the limits that presented better improvements in the results during the calibration process, following the empirical findings of previous studies [2, 9].

In the estimation process of the number of NH, regarding the livestock intended for breeding in 2003 and 2005, the following limits  $\nu = \{-1, 0, 1\}$  to the disaggregation level of the subregion of Alto Alentejo (NUTS III) and the county of Castelo de Vide were assumed. For the other breeding livestock disaggregation model,  $\nu = \{-2, 0, 2\}$  was considered at the first disaggregation level and  $\nu = \{-5, 0, 5\}$  at the second disaggregation level. On the procedure of reconstructing historical series and total livestock numbers of all categories, the limits  $\nu = \{-0.7, 0.7\}$  in the simultaneous disaggregation process and  $\nu = \{-1, 0, 1\}$  in the direct disaggregation process were considered.

The results are presented next, following the lines of implementation presented before.

The main livestock classes considered in the county of Castelo de Vide for the first model's implementation line were bovine cattle, sheep, and goats (Table 2). However, the need to convert breeding livestock into NH has led to consider

different divisions regarding bovine cattle, such as bovines of 1 to 2 years of age and bovines over 2 years of age.

In the second model's line of implementation aimed to calculate partial estimations of breeding livestock and to obtain the historical reconstruction of predominant livestock, one considered  $t = 7$  years and only cows over 2 years of age in regard to bovine cattle, due to lack of data. Livestock numbers were calculated, using data from 1989 (Table 3).

The third implementation line, which allowed the disaggregation of data simultaneously for the different subregions, includes the following livestock types: cows over 2 or more years of age; other bovines, sheep and goats; and a series of  $t = 7$  years (Table 4). It allowed disaggregating correctly the numbers of the different types of livestock, which were obtained directly from the model of cross-entropy minimization (14) to (17).

The fourth line of implementation of the model, which aimed to calculate the total proportions of livestock, considered cows over 2 or more years of age; other bovines, sheep and goats; and a series of  $t = 7$  years (Tables 5 and 6).

**3.1. Validation.** The results' validation was made based on crossreference with statistical data, using the opinion of experts and technicians from the Portuguese Ministry of Agriculture with a good knowledge of the area of application and on the calculation of the weighted prescription absolute deviation (WPAD) which measures the estimation deviations from real statistics data from the 1999 GAC. The WPAD was calculated considering the absolute deviation at the disaggregated level for each livestock specie and class, considering the total deviation for the disaggregated territorial units:

$$WPAD_p^i = y_p^i \left| \frac{y_p^i - \hat{y}_p^i}{y_p^i} \right|, \quad WPAD^i = \sum_{p=1}^p WPAD_p^i \quad (24)$$

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

$$WPAD = \sum_{i=1}^I \frac{s^i}{S} \cdot WPAD^i, \quad (25)$$

where  $\hat{y}_p^i$  is the estimated livestock number  $p$  at disaggregated territorial level  $i$ ,  $y_p^i$  is the respective observed value from the available statistics, and  $s^i/S$  is a ratio that allows weighing the  $WPAD^i$  according to the representativeness of disaggregated unit  $i$  in the aggregate.

The  $WPAD_p^i$  expresses the true importance of the percentage deviation in each livestock class regarding the observed values (PAD) weighed by its true importance at disaggregated level  $i$ . The  $WPAD^i$  corresponds to the sum of  $WPAD_p^i$  values by livestock number  $p$ , which allows giving the idea of the real total deviation for the values of the unit  $i$ . The WPAD corresponds to the weighted sum of the  $WPAD^i$  by the weight of each unit  $i$  regarding the total value or aggregate.

In the first implementation line, the procedures for the estimation of livestock intended for breeding in the county of

TABLE 3: Estimated historical data of livestock intended for breeding in Castelo de Vide.

Animal/year	1993	1995	1997	1999	2003	2005
Cows over 2 years	2223	2337	2395	2498	2529	2585
Sheep	8763	8364	7641	7888	7389	6659
Goats	2306	2420	2498	2609	2678	2819

(Source: model results).

TABLE 4: The number of livestock estimated for each of the Alentejo's subregions.

Sub-reg	Animal/year	1993	1995	1997	1999	2003	2005
AL	Cows over 2 years	28268	31141	35006	41437	42275	46558
	Other bovines	21644	25202	21911	29734	36036	35631
	Sheep	179618	180408	181575	200462	185408	165832
	Goats	17372	15101	15929	14918	10777	8314
AA	Cows over 2 years	37397	41563	46753	55957	58030	64842
	Other bovines	27587	32405	28190	38676	47641	47793
	Sheep	304412	308444	309809	345219	323374	291845
	Goats	39252	34429	36343	34420	25274	19780
AC	Cows over 2 years	49929	55082	62253	74023	76205	85054
	Other bovines	39271	45789	40022	54548	66705	66846
	Sheep	421675	424054	426898	471786	437254	392328
	Goats	25420	22124	23458	22061	16071	12555
BA	Cows over 2 years	23052	25764	29119	35162	37077	42378
	Other bovines	18182	21479	18776	25993	32561	33415
	Sheep	445558	454103	457018	513064	486988	447089
	Goats	34769	30672	32535	31097	23226	18603

AL-Alentejo litoral subregion; Alto Alentejo subregion; Alentejo Central subregion; Baixo Alentejo subregion.

(Source: model results).

Castelo de Vide are close to the opinion of experts consulted, showing also trends that can be observed in this county.

In the second implementation line, the model aiming to estimate part of the breeding livestock was validated for 1999 by comparison with the data from the GAC of 1999. This analysis showed that the model produced satisfactory data, since the  $WPAD^i$  is below 15%, which is a threshold usually recommended by some authors, such as Fragoso et al. [2] or Hazell and Norton [17]. Following this criteria for the  $PAD_p^i$ , only the data on breeding goats are not valid.

The simultaneous disaggregation process of livestock developed in the third line of implementation was only applied to the Alentejo's subregions and presented values that are very close to the reality, since the  $WPAD$  is only 6% in 1999. In terms of  $WPAD^i$  only Alentejo Central stands out with an 8.11 percentage and the percentage in Alto Alentejo is about 3.16.

Finally, in the last line of implementation of the model presented, the direct disaggregation procedures have shown a reliable disaggregation of the proportions of total livestock effective in Castelo de Vide. The  $PAD_p^i$  values for Castelo de Vide are very low for sheep (12%), which is the predominant livestock category. All the other categories show  $PAD_p^i$  values under 15%, with exception of the cows with 2 or more years

of age that present a 30%  $PAD_p^i$ . It was obtained an acceptable  $WPAD^i$  for Castelo de Vide with 14.6%, above the thresholds recommended by Fragoso et al. [2].

*3.1.1. The Disaggregation Informational Gain.* The heterogeneity of the information calculated was measured through the disaggregation informational gain (DIG) used by Howitt and Reynaud [13, 14]. According to the authors the measure of information gain from disaggregation should have the following properties: "(i) the measure of potential gain increases monotonically with the heterogeneity of the disaggregated sample; (ii) the gain from disaggregating a uniform set of samples is zero; (iii) the measure is invariant to changes in the number of disaggregated samples and the variability of the aggregated sample; (iv) the measure has an information-theoretic interpretation" [14].

The DIG 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. The DIG measure increases as the units' posterior distributions become closer to the true distributions. In the case of perfect disaggregation, the DIG is equal to 1. This measure may, however, be subjected to some critics,

TABLE 5: Proportion of total livestock numbers estimated in Castelo de Vide.

Animal/year	1993	1995	1997	1999	2003	2005
Cows over 2 years	0.128	0.134	0.148	0.15	0.154	0.177
Other bovines	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
Cows over 2 years	1934	2128
Other bovines	1319	1334
Sheep	8402	7849
Goats	892	709

(Source: model results).

considering the fact that a model with considerable errors might cause negative values. The formula can be presented as follows:

$$\text{DIG} = 1 - \frac{\sum_{i=1}^I \sum_{p=1}^P \hat{E}_p^i \cdot \ln(\hat{E}_p^i / E_p^i)}{\sum_{i=1}^I \sum_{p=1}^P E_p \cdot \ln(E_p / E_p^i)} = 1 - \frac{C\hat{E}}{CE}, \quad (26)$$

where  $\hat{E}_p^i$  are the estimated shares of livestock  $p$  in unit  $i$  and  $E_p^i$  are the real shares on the disaggregated unit  $i$ , and  $E_p$  is the aggregated share of livestock  $p$ . The same informational measure can be defined for each of the disaggregated units:

$$\text{DIG} = 1 - \frac{\sum_{p=1}^P \hat{E}_p^i \cdot \ln(\hat{E}_p^i / E_p^i)}{\sum_{p=1}^P E_p \cdot \ln(E_p / E_p^i)}. \quad (27)$$

Due to the nature of DIG and its line of use in previous studies [14], it was only applied to the third line of implementation of the model: the simultaneous disaggregation process. Here, the DIG revealed a percentage of recovered information of around 60.3%. This is a very satisfactory level of recovered information which overcomes the levels obtained by Fragoso et al. [2] for the Alentejo area. The  $\text{DIG}^i$  revealed also very satisfactory results: the best one was obtained in Alentejo Litoral with 78%, which is followed closely by Baixo Alentejo with 75%. The other two subregions revealed values near the 60%: Alto Alentejo and Alentejo Central presented 59% and 57%, respectively.

#### 4. 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 [13, 14], which initially were applied to the data disaggregation of agricultural occupation, are flexible enough to be applied with adaptations, to other types of data and still obtain satisfactory results. However, there are some recommendations that should be made for future studies in this area.

The proposed model was applied using a direct disaggregation process for the county of Castelo de Vide due to the lack of data about land uses. However, if there are valid data regarding several land uses, including the relation between them and livestock, they should be inserted in future studies. This will imply better development of the land use disaggregation model, but it also may give the chance of developing a combined approach, considering the simultaneous disaggregation of the livestock breeding in extensive systems and the main land uses. This is a considerable challenge, since the studies carried out until now showed unsatisfactory results in several cases.

A future application of the methodology proposed should also consider variations in the relation between agricultural and forest occupation and the number of livestock NH through time, as well as the application to a set of counties or territorial units. Taking into account these issues, we envisage the development of this methodology in a more complex way, with the inclusion of new databases, since in spite of its simplicity it allows us to obtain relevant information regarding animal production and management.

This methodology could have an important role, considering that some regulations and policy measures may influence the livestock breeding activity, hence leading to change its relation with agricultural and forest occupation. A good knowledge of the real situation will allow providing the ability of better understanding the consequences under different scenarios.

#### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

#### Acknowledgment

The authors gratefully acknowledge partial financial support from FCT, program POCTI.

#### References

- [1] R. Chakir, "Spatial downscaling of agricultural land-use data: An econometric approach using cross entropy," *Land Economics*, vol. 85, no. 2, pp. 238–251, 2009.
- [2] R. Fragoso, M. D. B. Martins, and M. R. V. Lucas, "Generate disaggregated soil allocation data using a Minimum Cross Entropy Model," *WSEAS Transactions on Environment and Development*, vol. 4, no. 9, pp. 756–766, 2008.

- [3] L. You, S. Wood, and U. Wood-Sichra, "Generating Global Crop Distribution Maps: From Census to Grid," Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference, International Association of Agricultural Economists, Gold Coast, Australia, August 2006.
- [4] L. You, S. Wood, U. Wood-Sichra, and J. Chamberlin, "Generating plausible crop distribution maps for sub-saharan Africa using a spatial allocation model," *Information Development*, vol. 23, no. 2-3, pp. 151-159, 2007.
- [5] INE-Instituto Nacional de Estatística, *Inquérito à Estrutura das Explorações Agrícolas*, INE, Lisbon, Portugal, 2006.
- [6] INE-Instituto Nacional de Estatística, *Estatísticas Agrícolas*, INE, Lisbon, Portugal, 2007.
- [7] M. Carvalho and M. Godinho, A Nova reforma da Política Agrícola Comum e suas consequências num sistema agrícola mediterrâneo de Portugal (The New Common Agricultural Policy Reform and its implications in a mediterranean agricultural system of Portugal), 2004, <http://www.sober.org.br/>.
- [8] DRAPAL, Direcção Regional de Agricultura e Pescas do Alentejo, Programa de Desenvolvimento Rural-Alentejo (Alentejo Rural Development Programme), DRAPAL, 2007.
- [9] M. B. Martins, R. Fragoso, and A. Xavier, "Spatial disaggregation of agricultural data in Castelo de Vide, Alentejo, Portugal: an approach based on maximum entropy," *JP Journal of Biostatistics*, vol. 5, no. 1, pp. 1-16, 2011.
- [10] R. Jongeneel, "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," in *Proceedings of the 10th Congress of the European Association of Agricultural Economists, Exploring the Diversity in the European Agri-Food-System*, European Association of Agricultural Economists, Saragoça, Spain, August 2002.
- [11] K. Karantininis, *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, 2001.
- [12] A. Tonini and R. Jongeneel, "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, European Association of Agricultural Economists, Corvinus University of Budapest (CUB), Budapest, Hungary September 2007.
- [13] R. Howitt and A. Reynaud, "Spatial Disaggregation of production data by maximum entropy," in *Proceedings of the 10th Congress of the European Association of Agricultural Economists, Exploring the diversity in the European Agri-Food-System*, European Association of Agricultural Economists, Saragoça, Spain, August 2002.
- [14] R. Howitt and A. Reynaud, "Spatial disaggregation of agricultural production data using maximum entropy," *European Review of Agricultural Economics*, vol. 30, no. 3, pp. 359-387, 2003.
- [15] I. Good, "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 911-934, 1963.
- [16] A. Golan, G. Judge, and D. Miller, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley & Sons, NewYork, NY, USA, 1996.
- [17] P. Hazell and R. Norton, *Mathematical Programming for Economic Analysis in Agriculture*, Macmillan, New York, NY, USA, 1986.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

