RECOVERY OF INCOMPLETE AGRICULTURAL LAND USES AND LIVESTOCK NUMBERS BY ENTROPY

Maria Martins[1,3], **Rui Fragoso**[2] and **António Xavier**[1]
[1]Sciences and Technology Faculty, University of Algarve, Portugal
[2]Management Department, University of Évora, Portugal

[3]Corresponding author-E-mail:mbmartins@ualg.pt

Paper prepared for presentation at the 114th EAAE Seminar
'Structural Change in Agriculture', Berlin, Germany, April 15 - 16, 2010

RECOVERY OF INCOMPLETE AGRICULTURAL LAND USES AND LIVESTOCK NUMBERS BY ENTROPY[1]

Maria de Belém Costa Freitas Martins (mbmartins@ualg.pt)

Assistant Professor, Sciences and Technology Faculty and CEFAGE-UE (Center For Advanced Studies in Management and Economics), University of Algarve

Rui Manuel de Sousa Fragoso (rfragoso@uevora.pt)

Assistant Professor, Management Department, University of Évora

António Manuel de Sousa Xavier (amxav@sapo.pt)

Master, Sciences and Technology Faculty, University of Algarve

Abstract

This paper proposes a combined use of previous developed models to estimate agricultural land use and livestock intended for breeding at disaggregated level, aiming to solve the lack of data problem between the Agricultural Census studies in Portugal for all statistical levels. The proposed model estimates incomplete information at disaggregated level through an entropy approach using an information prior. We applied this model in Nisa County and its parishes. The results showed a small error of the estimated data for the counties and a satisfactory one for the parish level.

Keywords: agricultural land use, livestock, data disaggregation, Nisa, maximum entropy.

[1]**Acknowledgments:** The authors gratefully acknowledge partial financial support from FCT, Program POCTI.

1. INTRODUCTION

Investigation concerning agricultural and rural development policies is greatly dependant on agricultural production statistics (You et al., 2007a) but this is a problem both worldwide (You and Wood, 2006; You et al., 2007a), in Europe (Chakir, 2007) and in Portugal (Fragoso et al., 2008, Martins et al. 2009a).

In Portugal, agricultural data at NUTS III, county and parish levels is almost impossible to obtain, except in the Agricultural Census, occurring from 10 to 10 years. Apart from the general lack of data, there is a special need for up-to-date data on agricultural land use, and on livestock numbers (Fragoso et al., 2008; Martins et al., 2009a and Martins et al. 2009b). Therefore, this fact is a problem at decision level for sustainable rural development and agricultural policies' options, especially for rural areas of low density. The planning and devising of a clear and sustainable rural development policy calls for the availability of disaggregated information (You et al., 2007a), at least when it comes to the numbers of breeding livestock, agricultural land use or other relevant variables in order to analyse the changes in the agricultural sector .

In Alentejo Region at South Portugal the importance of the agricultural and livestock breeding activity through the years is unquestionable (DRAPAl, 2007). However, with Portugal's entry into the European Union, this region has come under the influence of different policies and as a consequence there are several rural areas with problems (Carvalho and Godinho, 2004?).

However, with all the changes that took place, there is a need of data regarding agricultural land uses and livestock for several levels of disaggregation, especially for the rural counties where there is a tendency towards demographic decrease and decline of the agricultural activity.

To solve these problems, data disaggregation models, such as those developed by Chakir (2007), Fragoso et al. (2008), Martins et al. (2009a), Martins et al. (2009b), Howitt and Reynaud (2002), You and Wood (2004, 2006), You et al. (2004, 2006, 2007a, 2007b), are interesting solutions, as they supply data to support decision making in an inexpensive way, depending the choosing of one of them on the objectives of the work, i.e. the problem formulation.

Previous studies by the authors (Martins et al. 2009a) allowed obtaining valid data of agricultural land use for the several Alto Alentejo's counties, and of livestock intended for breeding for Castelo de Vide (Martins et al. 2009b). However, these authors did not: 1) developed a combined use of these models; 2) developed an application of the created livestock estimation models to other areas; 3) the models were not applied, with precision, to a third disaggregation level.

The proposed model estimates agricultural land use at a disaggregated level using aggregated data. It is built upon two fundamental steps: in the first one an information prior is built at aggregated level using a generalized maximum entropy (GME) model; in the second one, the data is disaggregated based on the previously estimated prior with restrictions such as the soil capacity. The model has additionally two different paths: simultaneous disaggregation and direct disaggregation, in order to solve problems that may arise in the previous one and in order to address situations of lack of data (Martins et al. 2009a).

For livestock the model developed is based on the approach by Martins et al. (2009b), which assumes that livestock's numbers are a function of the territory's agricultural and forest occupation, since the livestock is raised mainly in an extensive way, fed by pastures seeded or not, under trees or not, and so the livestock number is closely linked with the agricultural and forest area.

This model comprises two steps: in a first step data of the agricultural and forest occupation (resulting from the previous model) as well as the conversion of livestock numbers into normal heads (NH) are needed in order to determine the relation between NH and agricultural and forest occupation; in a second step, a model similar to the first one is followed in order to calculate the weight of each livestock category in normal heads (NH). Finally these results are converted into animal numbers.

These models were applied for the County of Nisa, in the NUTS II Alentejo following three disaggregation levels: 1) NUTS II Alentejo → Alentejo NUTS III regions; 2) NUTS III Alto Alentejo → Counties 3) Nisa county→parishes. The following figure (Fig. 1) represents the application area and the 3 levels of disaggregation.

Fig. 1 - **The model's application area**

Nisa is a county in the NUTS III Alto Alentejo, where there is a tendency to demographic decrease. For this county, data is needed on extensive breeding livestock and agricultural and forestry occupation (at parish level) to enable a correct analysis of the county's current state and the changes that took place last years.

The remainder of this paper is organized as follows. In section 2 the mathematical formulation of the disaggregation problem, including the necessary restrictions to have into account is presented. Section 3 presents the disaggregation model, in section 4 the results are presented and in section 5 the model's validation is presented. Finally, section 6 stresses the main conclusions of this work.

2. THE DISAGGREGATION PROBLEM

There are four main aspects that the disaggregation problem being studied must consider. In the first place, there is available data at aggregated level for **t** periods that should be considered (t, t+1…T); there is also disaggregated data for the first periods (1989 and 1999); one can incorporate some co-variables or restrictions; and finally the livestock is raised mainly in an extensive way.

So, the disaggregation problem for agricultural land use can be formulated as:

$$S_K^x(t) = \sum_{B_i \subset A_x} S_k^i(t) \tag{1}$$

Through the known variable $S^x_k(t)$, the objective is to obtain data for S_k^i (t), which represents the target variable **k** at disaggregated level **i**. The A_x matrix represents the aggregated unit **x**, in which $A_{x,\, x=1\dots X}$. So, one has to determine $S_k^i\ \forall k, i \text{ and } t$ for the sub-units B_i at the moment (t), of which there is no available information at disaggregate level.

Considering all the unities to disaggregate belong to a region (or aggregate) and that the soil occupation each year is given by $S_k(t)$, where k=1,...,K corresponds to the observed agro-forestry activities and t=1,...T corresponds to the year they occur, then the probability of the activity **k** in the year **t** is:

$$Y_k(t) = \frac{S_k(t)}{\sum_k S_k(t)}, \forall k, t = 1,...r \tag{2}$$

At disaggregated level the information in what concerns soil occupation for each agro-forestry activity $s_k^i(t)$ is available only for the first r periods (r < T) and is given by the probability of finding the **k** activity in year **t** and **i** unit:

$$Y_k^i(t) = \frac{s_k^i(t)}{\sum_k s_k^i(t)}, \forall k, t = 1,...r \tag{3}$$

When the objective is to estimate soil occupation for each unit (i) and years r+1,...T it is necessary to combine complete aggregated information at aggregated level for t=1,...,T with incomplete information disaggregated for each unit and concerning years t=1,...r < T.

It is also necessary to guarantee that soil usage with a certain activity at aggregated level equals the sum of the activity areas at disaggregated level. This is done considering that the s_k^i estimation or, alternatively the $y_k^i \ \forall \mathrm{k, i\ and\ t = r+1,...T}$ also satisfies the following restriction:

$$S_k(t) = \sum_{i=i}^{I} s_k^i(t), \ \forall k, t = r+1,...,T \tag{4}$$

Finally, a last restriction should impose that the used area for each activity must not exceed the biophysical adequate area:

$$B^i{}_k(t) \geq s_k^i(t), \ \forall k, t = r+1,...,T \tag{5}$$

Equation (5) reports to situations in which there is data of the biophysical area existing in each unit i for a occupation k, and ensures that this area is respected. It must be also respected that the sum of several areas for k activities, with the same biophysical requirements (some uses may be developed in the same biophysical conditions), when all combined must not be larger than the referred biophysical area.

In what concerns breeding livestock we want to obtain disaggregated data on the main livestock categories at municipality level. To address this problem it must be taken into account that in Alentejo the livestock's number is closely linked with the agricultural and forest area ($S_k^i(t)$), as it is raised mainly in an extensive way, fed by pastures, seeded or not, under trees or not. 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:

$$E^i_P(t) = f\,[S_k^i(t)1 \ldots S_k^i(t)n] \quad (6)$$

where $E^i_P(t)$ represents the livestock numbers p in unit i, at the moment t and $S_k^i(t)$ is the agricultural or forest occupation k in unit i, at the moment t. These livestock numbers are determined by:

$$E^i_P\,(t) \leq Cv\;{}^i_k(t)\;\Box\, i, k, t = r+1, \ldots, T \quad (7)$$

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. the total necessary physical area.

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

$$E_P(t) = \sum_{i=i}^{I} E_P^i(t),\ \Box p, t = r+1, \ldots, T \quad (8)$$

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.

3. THE PROPOSED MODEL

In what concerns soil occupation, the variables levels (each agro-forestry activity) depend only on their values for precedent years - the dynamic process of soil occupation for a county in a given year **t** depends only on its occupation at the previous year **t-1**. So, the inter-temporal existing relations can be characterized by a first degree Markov process, which allows the use of all the existing information.

The probability of passing from any state of decision $j \in \{1,\ldots,K^r\}$ in the year **t-1** to a state of decision $j' \Box \{1,\ldots,K^r\}$ in year **t**, assuming a second order Markov process, can be given by $y_k^i(t-1) \times \mathrm{y}_\mathrm{k}^\mathrm{i}(t)$. This probabilities product can be associated to a matrix $T_{jj'}^i$ with dimension ($K^r \times K^r$), which is the transition probabilities matrix.

3.1. The agricultural land use estimation

3.1.1.The estimation of the transition probabilities matrix at an aggregated level

The dynamic process of agro-forestry activities distribution at aggregated level is obtained based on the estimation of the transition probabilities matrix at aggregated level $T_{jj'}$, using the maximum entropy theory (ME). It is a problem of information recovering which also estimates the distribution error ($e_{j'}(t)$). To do so, considering the Alentejo conditions and based in Howitt and Reynaud (2002) and Fragoso et al. (2008), Martins et. al (2009a), the following model was developed:

$$\underset{T,e}{Max}\ \mathrm{H(T,e)} = -\sum_{\mathrm{j}=1}^{\mathrm{J}} \sum_{j'=1}^{J} \sum_{m=1}^{M} T_{jj'm}.\log(T_{jj'm}) - \sum_{\mathrm{j'}=1}^{\mathrm{J}} \sum_{n=1}^{N} \sum_{t=r}^{T-1} e_{j'n}(t).\log(e_{j'n}(t)) \tag{9}$$

Subject to:

$$Q_{j'}(t+1) = \sum_{j=1}^{J} \sum_{m=1}^{M} Q_j(t).w_m.T_{jj'm} + \sum_{n=1}^{N} v_n.e_{j'n}(t), \quad \forall\ \mathrm{j' \ and \ t} \tag{10}$$

$$\sum_{j'=1}^{J} \sum_{m=1}^{M} w_m.T_{jj'm} = 1 \qquad \forall\ \mathrm{j} \tag{11}$$

$$\sum_{m=1}^{M} T_{jj'm} = 1 \quad \mathrm{and\ Tjj'} \in [0,1] \quad \forall\ \mathrm{j\ and\ j'} \tag{12}$$

$$\sum_{n=1}^{N} e_{j'n}(t) = 1 \quad \mathrm{and\ e_{j'n}} \in [0,1] \quad \forall\ \mathrm{j'\ and\ t} \tag{13}$$

This optimization problem maximizes the entropy of the probabilities distribution $\{T_{jj'1},...,T_{jj'M}\}$ $\forall \mathrm{j\ e\ j'}$ and $\{\mathrm{e}_{j'1},...,e_{j'N}\}$ $\forall \mathrm{j'\ e\ t}$, considering the conditions imposed by the restrictions. Equation (10) defines the dynamic process of soil occupation. Equation (11) determines that the sum of the transition probabilities in any Markov state is equal to 1. Equations (12) and (13) guaranty that the variables values $\{T_{jj'1},...,T_{jj'M}\}$ and $\{\mathrm{e}_{j'1},...,e_{j'N}\}$ are defined between 0 and 1 and that its sum is 1.

As the matrix $T_{jj'}$ is defined between 0 and 1 it is necessary to define, according to Golan et al. (1996), a parameter $w' = \{w_1,...,w_M\}$, of $M \geq 2$ points with $w_1=0$, $w_M=1$ and a probabilities distribution $\{T_{jj'1},...T_{jj'M}\}$ such as:

$$Tjj' = \sum_{m=1}^{M} w_m . T_{jj'm} \tag{14}$$

For the unknown disturbances $e_{j'}(t)$ estimation, the same procedure is taken, reparameterizing the error trough a support vector $v' = \{v_1,...,v_N\}$, with $N \geq 2$, so that the residues may be defined as $\{e_{j'1}(t),...,e_{j'N}(t)\}$.

This problem is convex, which means it has a unique optimal solution for both transition probabilities matrixes $\widehat{T}_{jj'm}$ and $\widehat{e}_{j'n}(t)$ - so, we have the distribution of agro-forestry activities at the aggregated level. The information prior built for the aggregated level will then be used in the transition probabilities matrixes at disaggregated level's process of estimation. To restore the agro-forestry distributions at NUTS III and counties levels the transition probabilities matrix at disaggregated level should be estimated each year, by solving a generalized cross entropy (GCE) minimization problem concerning the estimated prior at aggregated level (transition matrixes) and using the transition probabilities estimates the land use at NUTS III and county levels should be computed.

3.1.2. The disaggregation process

According to Golan et al. (1996), Howitt and Reynaud (2002), Fragoso et al. (2008), Martins et. al (2009a), the first step of this disaggregation process can be translated by the following cross entropy minimization problem:

$$\underset{T^i, e}{Min}\, \mathrm{H}(\mathrm{T}^i, \mathrm{e}) = \sum_{\mathrm{i}=1}^{\mathrm{I}} \sum_{j=1}^{J} \sum_{j'=1}^{J} T_{jj'}^{i} . \log(T_{jj'}^{i} / \widehat{T}_{jj'}) + \sum_{\mathrm{k}=1}^{\mathrm{K}} \sum_{n=1}^{N} \sum_{t=r+1}^{T} e_{kn}(t) . \log(e_{kn}(t)) \tag{15}$$

Subject to:

$$S_k(t+1) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{j' \in \psi} \mathrm{q}_{\mathrm{j}}^{\mathrm{i}}(t) . T_{jj'}^{i}(t) . s^i + \sum_{n=1}^{N} \zeta_n . e_{kn}(t) \forall \mathrm{k} = 1, \ldots \mathrm{K} \tag{16}$$

$$\sum_{j'=1}^{J} T_{jj'}^{i} = 1 \text{ and } T_{jj'}^{i}{}' \in [0,1] \ \forall \mathrm{j} \text{ and } \mathrm{j}' \tag{17}$$

$$\sum_{n=1}^{N} e_{kn}(t) = 1 \quad \text{and } \mathrm{e_{kn}}(\mathrm{t}) \square [0,1] \ \square \mathrm{j' and\ t} \tag{18}$$

$$yhlim_k^i \geq \sum_{j=1}^{J} \sum_{j' \square \psi} \mathrm{q_j^i}(t).T_{jj'}^i(t) \forall \mathrm{i\ and\ k} = 1, \ldots \mathrm{K} \tag{19}$$

and

$$y\,\mathrm{slim}_k^i \geq \sum_{j=1}^{J} \sum_{j' \square \psi} \mathrm{q_j^i}(t).T_{jj'}^i(t) \forall \mathrm{i\ and\ k} = 1, \ldots \mathrm{K} \tag{20}$$

or

$$\mathrm{B}_k^i \geq \sum_{j=1}^{J} \sum_{j' \square \psi} \mathrm{q_j^i}(t).T_{jj'}^i(t).S^i \forall \mathrm{\ i\ and\ k} = 1, \ldots \mathrm{K} \tag{21}$$

where $\{\zeta_1,...,\zeta_N\}$ with N≥ 2 points is the support vector associated to the probabilities $\{\mathrm{e_{k1}},...,e_{KN}\}$.

The objective is to minimize the cross entropy of the transition probabilities distribution and the entropy of the errors probabilities distribution (15) subjected to the exposed restrictions.

Equation (16) guarantees information compatibility between aggregated and disaggregated levels. Equations 17 and 18 ensure that $T^i_{jj'}$ and e_{kn} sum is equal to 1.

The equations (19) and (20) respectively refer (as Martins et. al, 2009a stated) to the percentual value in relation to the total area of the farm (which is known) and imply that the probabilities of occupation must not exceed the historical maximum limits of the probability of each occupation k in each unit i ($yhlim_k^i$) or the biophysical limits for each occupation ($y\,\mathrm{slim}_k^i$), when the data complemented with experts' opinions does not allow the establishment of the first restriction.

Supposing, in other situation, which is frequent in this area, where we have a biophysical area of a unit i referring to a county or parish, and so, that unit is larger than the farms total land use. If we do have not specific information about its distribution in the farm's area, one may take advantage of this data through the equation (21), instead of (20). In spite of not being used it must always be considered when applying data to others Alto Alentejo's counties in order to take the utmost advantage of the existing information.

After this minimization problem, one may simply use the $\widehat{T}^{i}_{jj'}$ to calculate $\widehat{y}^{i}_{k}(t+1)$, by the following equation:

$$\widehat{y}^{i}_{k}(t+1)=\sum_{j=1}^{J}\sum_{j'\in\psi}\mathrm{q}^{\mathrm{i}}_{\mathrm{j}}(t).T^{i}_{jj'}(t) \tag{22}$$

We may also calculate the agricultural land use area S for activity k in unit i, by the following:

$$\hat{S}^{i}_{k}\ (t+1)=\widehat{y}^{i}_{k}(t+1).s^{i} \tag{23}$$

3.2.The estimation of livestock numbers in Normal Heads (NH)

3.2.1. The establishment of the relation between NH and agricultural land use

The approach used was developed by Martins et. al (2009b) and admits that livestock numbers are a function of farms' land occupation. Considering that it is possible to estimate the coefficients indicating this relation it is necessary that data resulting from a model for agricultural and forest occupation estimation is available.

It is assumed that each livestock category has different requirements in terms of feeding, which means it needs to be linked with a wider or smaller area. It is also necessary to convert the different livestock categories into NH. A NH is a livestock measure that converts the different livestock categories in function of the species and ages, based on a legal table of conversion (Portaria n.º 229-A/200 of 6 March 2008; and INE, 2006c). As an example, a sheep over 1 year will equal 0.15 NH, while a bovine over 2 years will equal 1 NH.

Each county/ territorial unit has a determined relation between NH and the agricultural and forest occupation. The procedures to estimate total livestock numbers in NH consist in calculating the number of NH for each predominant livestock category in the area and establishing the relation between the number of NH and agricultural and forestry occupation, namely forage crops and permanent pastures. So, we consider that the number of effective breeding livestock p in NH, in unit i, at the moment t can be calculated in the following way:

$$ENH^{i}_{p}(t)=E^{i}_{p}(t)\times INH_{\mathrm{p}} \tag{24}$$

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:

$$\mathrm{R}_{\mathrm{pk}}^{\mathrm{i}} = (\sum_{p=1}^{P} ENH_{p}^{i}(t)) \div S_{k}^{i} \tag{25}$$

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:

$$ENH^{i}(t+1) = \mathrm{R}_{\mathrm{pk}}^{\mathrm{i}} \times S_{k}^{i}(\mathrm{t}+1) \tag{26}$$

3.2.2. The estimation of livestock percentage weight

With the methodology proposed the total number of breeding livestock (NH) is estimated, but not the percentage weight of each category. Therefore, data of each livestock breeding categories should be converted to NH, for the years in which information is available; then, the data from a database created at aggregate level should be disaggregated, based on the theory of maximum entropy and the livestock weight in t+1…T should be calculated; and finally livestock numbers in NH should be redistributed according to estimated proportions and convert them into animal numbers.

When there is not enough or precise data about the agricultural and forestry occupation (such as in this case), instead of going for the simultaneous disaggregation of information in all counties in Alto Alentejo’s NUT III described before, one may also choose the direct disaggregation of animals’ data (second variant of the model defined by Martins et. al, 2009b) regarding the Nisa County, by rewriting equation (16) in the following way:

$$Epr_{p}(t+1) = \sum_{j=1}^{J} \sum_{j' \in \psi} \mathrm{q}_{\mathrm{j}}^{\mathrm{i}}(t).T_{jj'}^{i} + \sum_{n=1}^{N} \zeta_{n}.e_{kn}(t) \quad \forall \mathrm{k} \ \text{ and t} \tag{27}$$

3.2.3. 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:

$$ENH^{i}_{p}(t+1) = Epr^{i}_{p}(t+1).ENH^{i} \tag{28}$$

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:

$$E^{i}_{p}(t+1) = ENH^{i}_{p}(t+1) / INH_{p} \tag{29}$$

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

To estimate total livestock numbers one may suppose, as refferred by Martins et al. (2009b) that the year variation rate follows the livestock intended for breeding rate, and so:

$$\vec{\nabla}Et^{i}_{p} = \left(\left(\frac{E^{i}_{p}(t+x)}{E^{i}_{p}}\right)^{(\frac{1}{x})} \right) - 1 \tag{30}$$

in which $\vec{\nabla}Et^{i}_{p}$ is the year variation rate of total livestock p, x the number of years. If the years to estimate are the same ones, we may use this procedure in a simple way.

3.2.4. Empirical adaptations of the model

As there is not enough data to consider a second order Markov process at the transition matrixes level, since at disaggregated level the data is only available to 1989 and 1999, taken as validation year, to calculate the probability of occupation at time t+1, the transition matrix only considers the previous year and not the year before (is not possible to calculate $\mathrm{q}^{\mathrm{i}}_{\mathrm{j}}(t)$ for t=1989) (Martins et al., 2009a).

To solve the optimization problem exposed regarding agricultural land use, it is necessary to define the limits for the errors support values and for parameters. According to (Martins et al. 2009a) **w** was defined with M = 3 points and then $v = \{\text{-}5,0,5\}$ for the first level of disaggregation, $v = \{\text{-}1,0,1\}$ for the second level and $v = \{\text{-}0.04,0,0.04\}$ for the third level of disaggregation were used for the simultaneous disaggregation model.

In order to obtain livestock data in Nisa, it was necessary to implement the disaggregation process in two stages, admitting the use of only one direct disaggregation and the adaptations proposed by Martins et al. (2009b) given the existing data for Portugal.

In the application of the model's different variables, the support limits of parameters and error had to be defined. So, considering previous studies, it was established the w can be defined to M = 3 points, being therefore considered $w = \{0, 0.5, 1\}$ for all models.

In the estimation process of the number of NH, regarding the livestock intended for breeding in 2005, the following limits $v = \{-1,0,1\}$ to the disaggregation level of Alto Alentejo and Nisa were assumed. For the other breeding livestock disaggregation model $v = \{-2,0,2\}$ was considered at the first disaggregation level and $v = \{-1.2, 0, 1.2\}$ 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 for all levels of disaggregation.

4. THE RESULTS

In the simultaneous agricultural land use model, 7 agro-forestry activities (K=7) were considered, which consistently express the representative land uses in the studied area, and also T=7 years, since only the essential information was taken into account. The activities were: cereals (CC), fallows (FF), pastures and forages (PF), permanent crops (PC), permanent pastures, (PP), shrubs and forest without pastures (SF) and other uses (OU).

The results are shown for NUTS III Alto Alentejo's counties (table 1) for 2005 and Nisa (table 2), considering the obtained sequence of years (which was created for all counties). The results are also presented for the several Nisa's parishes, regarding the third level of disaggregation proposed (table 3).

Table 1 - Estimated land use probabilities of the several Alto Alentejo's counties in 2005

Table 2 - Land use probabilities in Nisa

Table 3 - Land use probabilities for Nisa's parishes in 2005

For the direct disaggregation model of breeding livestock, the categories considered in Nisa were bovine cattle, sheep and goats (Table 4). 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 4 - Estimated breeding livestock in Nisa

For partial estimations of breeding livestock aiming the historical reconstruction of predominant livestock, t=7 years were considered and only cows over 2 years of age (COT) in regard to bovine cattle. So livestock numbers were calculated, in NH, according to the occupation registered in year 2005, but also partially using data from 1989 (Table 5).

Table 5 - Recovery Historical Data of livestock intended for breeding in Nisa

On the other hand, the variant of the model intending to recover the total proportions of livestock considered, and the livestock number at the time in the years that exists information on the livestock intended for breeding: 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 6 and Table 7).

Table 6 - Proportion of total livestock numbers estimated in Nisa

Table 7 - Livestock numbers in Nisa

5. THE MODELS' VALIDATION

Validation was made using the comparison between estimated and real data (for 1999), the experts' opinions and other cartographical sources and inquiries.

The comparison between estimated and real data was based in the analysis of several variation indicators, such as (Martins et al., 2009a):

$$PAD_k = \left| \frac{\mathrm{Y_k} - \widehat{\mathrm{Y}}_\mathrm{k}}{\mathrm{Y_k}} \right| \times 100 \quad \mathrm{WPAD}^\mathrm{i}_\mathrm{k} = y^i_k \left| \frac{y^i_k - \widehat{\mathrm{y}}^\mathrm{i}_\mathrm{k}}{\mathrm{y}^\mathrm{i}_\mathrm{k}} \right|, \quad \text{and} \quad \mathrm{WPAD}^\mathrm{i} = \sum_{k=1}^{K} \mathrm{WPAD}^i_k \tag{31}$$

and, at the aggregated level:

$$\mathrm{WPAD} = \sum_{i=1}^{I} \frac{S^i}{S} . \mathrm{WPAD}^i \qquad (32)$$

The Prescription Absolute Deviation (PAD_k) is the absolute percentual variation of the estimated ($\hat{Y}_k$) occupation relating the observed values (Y_k) and can also be applied to each unit i. The WPAD_k^i (Weighted Prescription Absolute Deviation) is the deviation in each land use category in unit i weighed by its true importance or probability of occupation and WPAD^i corresponds to the sum of the WPAD_k^i 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 WPAD^i by the weight or importance of each unit i regarding the total value.

The WPAD analysis of the simultaneous disaggregation results revealed good values for land use at NUTS III and Counties level (Fig. 2). For the various counties of Alto Alentejo it is 26%[2] and for the various NUTS III regions of Alentejo it is 16.746%, having the Alto Alentejo sub-region a WPAD^i of only 9.608%, with very precise predictions in all agricultural land uses. For the several Nisa's parishes the results were satisfactory since the calculated WPAD was of 27.426. The counties that more contributed to the DAPP were Alter do Chão followed by Ponte de Sor, Mora and Fronteira (fig. 3).

Fig. 2 - WPAD^i distribution in the various counties of Alto Alentejo

Fig. 3 - The contribution for the WPAD by the various counties of Alto Alentejo

The median of the PAD^i_k land uses at the several counties reveals very satisfactory results of some occupations, such as permanent crops (7.57%) or permanent pastures (17.6%). Only the cereals and fallows (31.7 and 40.7% respectively) reveal values higher than 30%, which means that a considerable number of counties has satisfactory values.

---

[2] We may also obtain satisfactory results for the several counties with the consideration of no additional restrictions, since its values were 12.4% for Nisa, and the WPAD for the counties was 30.2%.

The analysis of the several Nisa's parishes revealed that, in spite of the total WPAD being 27.426, there are several parishes with values higher than 35%: Amieira do Tejo, Santana and São Matias (fig.4). On the other hand, the parishes that revealed better results of this indicator were S. Simão and Montalvão. The parish that more contributed to the WPAD was Espírito Santo, with a contribution of more than 4.8%.

Fig. 4 - WPAD^{i} distribution in the parishes of Nisa

Fig. 5 - The contribution for the WPAD by the various parishes of Nisa

The analysis of the of the PAD$^{i}_{k}$ land uses at the several parishes reveals satisfactory values for permanent crops (21.5%), permanent pastures (12.7%) and shrubs and forest without pastures (15.7%).

To measure the information gains the Disaggregation Informational Gain (DIG) was used (Howitt and Reynaud, 2002). DIG is based in the cross entropy between the observed values for land use at aggregated level y_k and at disaggregated level y_k^i and in the cross entropy between the land uses estimated by the disaggregation process $\hat{y}_k^i$ and the observed at disaggregated level y_k^i.

DIG had satisfactory results at the different disaggregation levels considered. The NUTS III regions have a DIG of 0.443, which means 44.3% of the information heterogeneity was recovered. In spite of the low DIG in comparison with studies made by the authors before (Martins et al. 2009a), the good reliability of the data obtained for the Alto Alentejo justifies the acceptance of the referred value. As regards the Alto Alentejo Counties, the value was 0.658, which means 65.8% of the information heterogeneity was recovered and for the Nisa's parishes the value was unsatisfactory.

The experts' opinions were taken using the results of the direct disaggregation process only for the year 2005 and only for Nisa county (and not its parishes individually). Even if they corroborate some of the estimated land use for the 1990's, there are some variations that are not justified (for permanent pastures between 1997 and 1999). In 2005 values were also appointed to be probably in excess, such as permanent pastures. There are also other aspects that resulted of the forest fires that the model didn't handle. This information will be taken into account in a final version of this study.

Other sources were used to survey the land use obtained, such as the land use map COS 90 and the Corine Land Cover (CLC) map 2000.

The procedures for the estimation of breeding livestock in Nisa for 2003 and 2005 were corroborated by the experts consulted in what concerns bovine cattle and some tendencies that can be observed in the municipality, but there is an improvement necessity on the final numbers of sheep and goats that result from the land uses.

On the other hand, the model aiming to estimate part of the breeding livestock was validated in 1999, in comparison to the data from the RGA of 1999. This analysis showed that the model produced satisfactory data, since we obtained a $WPAD^i$ of around 9.899%, with only the data on breeding goats being not valid.

The direct disaggregation procedures have shown a reliable recovery of the proportions of total livestock effectives in Nisa. The biggest percentage of PAD^i_p regards bovine cattle: cows with 2 or more years of age (22.5%). All the other categories show PAD^i_p values under 4%. This leads to acceptable $WPAD^i$ values for Nisa around 3.6%, and also shows a clear improvement of the previous developed model (Martins, 2009b).

6. CONCLUSIONS

The proposed models create a valid basis for designing strategies and to analyze possible impacts of the agricultural policy changes both in Alto Alentejo and Nisa, which contributed to a greater extensification and led to a raise in the permanent pastures area and to a decline of traditional productions. The agro-forestry occupation for Alto Alentejo and Nisa after 1999 tends to be more homogeneous and tends now, according to the model, to maintain the predominance of permanent pastures. On the other hand, the livestock sheep and goats revealed a decline, and there is a stabilization and growth of the bovine cattle, following some of the region's tendencies.

These models can also be applied in other locations, with differentiated agro-ecological conditions, since the analysis demonstrated that this approach is flexible and can be adapted to various kind of data and differentiated situations, allowing generating data in order to analyse the structural changes that took place in agriculture. Nevertheless, there are still some problems and aspects that could be improved.

There are some changes that could not be captured, such as forest fires and other aspects. The experimental study for the disaggregation of agricultural land uses values on the several Nisa's parishes reveals some problems that will be solved in the future with the full development and optimization of the model and insertion of more expert knowledge.

In the near future the aim is to optimize and completely develop this model in order to overcome these problems in order to apply it with the maximum accuracy to all the areas of the region, and others of Portugal. The results will also be enhanced by the possibility of using more accurate data, namely the 2009 Agricultural Census, and by the development of data disaggregation methodologies that can use the Soil Occupation Maps which will soon be available. The results would be more precise with the inclusion of other information priors, coming from experts' opinions or available cartographic information.

REFERENCES

Cai, X, De Fraiture, C., Hejazi, M., (2006). Retrieval of irrigated and rainfed crop data using a general maximum entropy approach. *Irrig. Sci* 25: 325-338.

Chakir, R. (2007). Spatial downscaling of Agricultural Land Use Data: An econometric approach using cross-entropy method. Toulouse: INRA. (www2.toulouse.inra.fr/lerna/projet/ADD%20WP%203%20(RChakir)%20v2.pdf). Accessed 20 August 2008.

DRAPAl- Direcção Regional de Agricultura e Pescas do Alentejo (2007). *Programa de Desenvolvimento Rural-Alentejo*. DRAPAL. (http://www.draal.min-agricultura.pt/). Accessed at 20 January 2008.

Carvalho, M. and Godinho, M. (2004?). A Nova reforma da Política Agrícola Comum e suas consequências num sistema agrícola mediterrâneo de Portugal. (http://www.sober.org.br/palestra/2/1099.pdf). Accessed 15 January 2009.

Fragoso, R., Martins, M. and Lucas, R. (2008). A Minimum Crossed Entropy Model to Generate Information to the Disaggregated Level. Proceedings of the 4th IASME/WSEAS international Conference on Energy, Environment and Sustainable Development. 11 to 13 June, Faro, Algarve, Portugal: WSEAS Press, 540-544.

Good, I. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *The Annals of Mathematical Statistics* 34 (3): 911-934.

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

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 August 2002. European Association of Agricultural Economists, Saragoça, Spain: EAAE.

INE-Instituto Nacional de Estatística (1989). *Recenseamento geral da agricultura de 1989*. Lisbon, Portugal: INE.

INE-Instituto Nacional de Estatística (1993-2007a). *Estatísticas agrícolas*. Lisbon, Portugal: INE.

INE-Instituto Nacional de Estatística (1994-2007b). *Anuários estatísticos da Região do Alentejo*. Lisbon, Portugal: INE.

INE-Instituto Nacional de Estatística (1995-2006c). *Inquérito à estrutura das explorações agrícolas*. Lisbon, Portugal: INE.

INE-Instituto Nacional de Estatística (2001d). *Recenseamento geral da agricultura de 1999. Lisbon*, Portugal: INE.

INE-Instituto Nacional de Estatística (2002e). *Recenseamento geral da população -Censos 2001*. Lisbon, Portugal: INE.

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: EAAE.

Just, R. and Pope, R., (1999). The agricultural producer: theory and statistical measurement. In Gardner, B., Rauser, C. (Eds), *Handbook of Agricultural Economics*. Elsevier.

Martins, M. B., Fragoso, R., and Xavier, A. (2009a). Spatial disaggregation of agricultural data in Castelo de Vide, Alentejo, Portugal: an approach based on maximum entropy. *Paper in peer review*.

Martins, M. B., Fragoso, R., and Xavier, A. (2009b). Recovery of incomplete data of statistical livestock number applying an entropy approach. *Paper in peer review*.

Miller, D., Plantinga, A. (1999). Modelling land use decision with aggregate data. *American Journal of Agricultural Economics* 81 (February): 180-194.

Pukelsheim, F. (1994). The three sigma rule. *American statistician* 48(4): 88-91.

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, Hungry: EAAE.

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: IFPRI.

You, L., Wood, S., and Wood-Sichra, U. (2004). Generating Plausible Crop Distribution Maps for Sub-Sahara Africa Using Spatial Allocation Model. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting. August 1-4, Denver, Colorado.

You, L. and Wood, S. (2006). An entropy approach to spatial disaggregation of agricultural production. *Agricultural Systems* 90 (2006): 329–347.

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, Gold Coast, Australia.

You, L., Wood, S., Wood-Sichra, U. and Chamberlin, J. (2007a). Generating Plausible Crop Distribution Maps for Sub-Saharan Africa Using a Spatial Allocation Model. *Information Development* 23 (2/3): 151-159.

You, L., Wood, S. and Wood-Sichra, U. (2007b). Generating Plausible Crop Distribution and Performance Maps for Sub-Saharan Africa Using a Spatially Disaggregated Data Fusion and Optimization Approach. IFPRI Discussion Paper 00725. International Food Policy Research Institute (IFPRI). Washington DC: IFPRI.

Table 1– Estimated land use probabilities of the several Alto Alentejo's counties in 2005

Ocup./conc.	A. do Chão	Arronches	Avis	Campo Maior	Crato	Elvas	Fronteira	Gavião	Marvão	Monforte	Mora	Nisa	P. de Sor	Portalegre	C. Vide
CC	0.127	0.163	0.124	0.305	0.012	0.182	0.282	0.028	0.012	0.12	0.139	0.011	0.024	0.013	0.014
FF	0.287	0.23	0.137	0.142	0.039	0.148	0.267	0.028	0.02	0.135	0.375	0.036	0.038	0.046	0.063
PF	0.108	0.091	0.051	0.032	0.066	0.036	0.081	0.05	0.057	0.035	0.136	0.098	0.068	0.068	0.04
PC	0.137	0.122	0.126	0.275	0.072	0.107	0.222	0.154	0.152	0.061	0.079	0.171	0.05	0.115	0.057
PP	0.268	0.319	0.444	0.158	0.739	0.458	0.075	0.221	0.685	0.61	0.179	0.567	0.591	0.607	0.75
SF	0.027	0.025	0.059	0.012	0.067	0.019	0.012	0.471	0.042	0.002	0.051	0.087	0.205	0.115	0.031
OU	0.046	0.05	0.059	0.077	0.006	0.051	0.061	0.048	0.032	0.037	0.042	0.029	0.025	0.037	0.045

(source: model results)

Table 2- Land use probabilities in Nisa

Land use/year	1993	1995	1997	1999	2003	2005
CC	0.024	0.022	0.025	0.017	0.015	0.011
FF	0.131	0.106	0.118	0.068	0.06	0.036
PF	0.137	0.127	0.142	0.088	0.094	0.098
PC	0.236	0.206	0.227	0.17	0.172	0.171
PP	0.303	0.37	0.33	0.513	0.531	0.567
SF	0.133	0.131	0.138	0.106	0.101	0.087
OU	0.036	0.039	0.021	0.038	0.028	0.029

(source: model results)

Table 3- Land use probabilities for Nisa's parishes in 2005

	Alpalhão	Amieira do Tejo	Arez	Espírito Santo	Montalvão	Nossa Senhora da Graça	Santana	São Matias	São Simão	Tolosa
CC	0.019	0.021	0.001	0.054	0.018	0.011	0.009	0.022	0.019	0.008
FF	0.038	0.010	0.039	0.055	0.012	0.060	0.088	0.092	0.062	0.029
PF	0.097	0.090	0.128	0.113	0.032	0.157	0.155	0.162	0.016	0.139
PC	0.145	0.162	0.147	0.114	0.122	0.195	0.327	0.235	0.242	0.281
PP	0.728	0.617	0.639	0.677	0.674	0.327	0.148	0.254	0.342	0.476
SF	0.010	0.112	0.072	0.025	0.127	0.224	0.190	0.170	0.148	0.081
OU	0.018	0.041	0.012	0.016	0.079	0.010	0.058	0.068	0.188	0.017

(source: model results)

Table 4-Estimated livestock intended for breeding in Nisa

Animal / year	2003	2005
BO 1 to 2 yrs	409	470
BO over 2 yrs	3039	3633
Sheep	18267	16972
Goats	2643	2637

(source: model results)

Table 5- Recovery Historical Data of livestock intended for breeding in Nisa

Animal / year	1993	1995	1997	1999	2003	2005
COT	923	1245	1248	2080	2381	2874
Sheep	12297	14575	12364	18473	17800	16907

(source: model results)

Table 6- Proportion of total livestock numbers estimated in Nisa

Animal/year	1993	1995	1997	1999	2003	2005
COT	0.042	0.044	0.049	0.051	0.052	0.060
OB	0.058	0.063	0.055	0.063	0.077	0.080
Sheep	0.708	0.736	0.737	0.760	0.780	0.787
Goats	0.193	0.156	0.159	0.126	0.091	0.073

(source: results from the MGCE model-livestock numbers)

Table 7- Livestock numbers in Nisa

Animal /year	2003	2005
COT	1638	1839
OB	2425	2452
Sheep	24563	24126
Goats	2866	2238

(source: results from the MGCE model-livestock numbers)

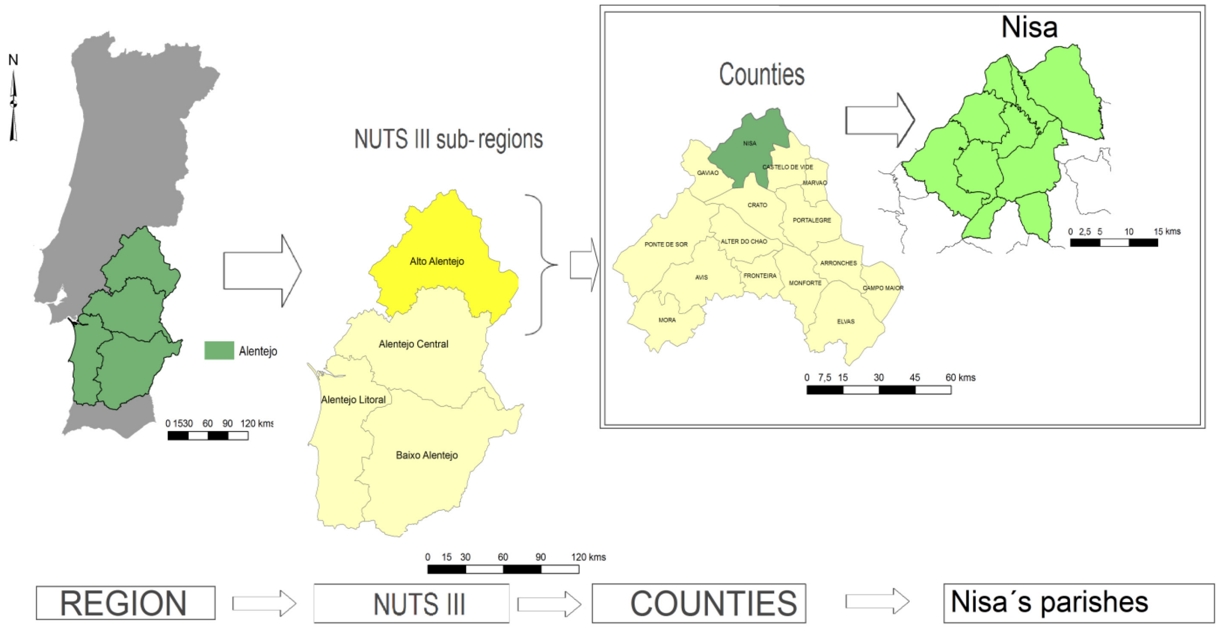

Fig. 1- The model's application area

(source: Carta administrativa de Portugal)

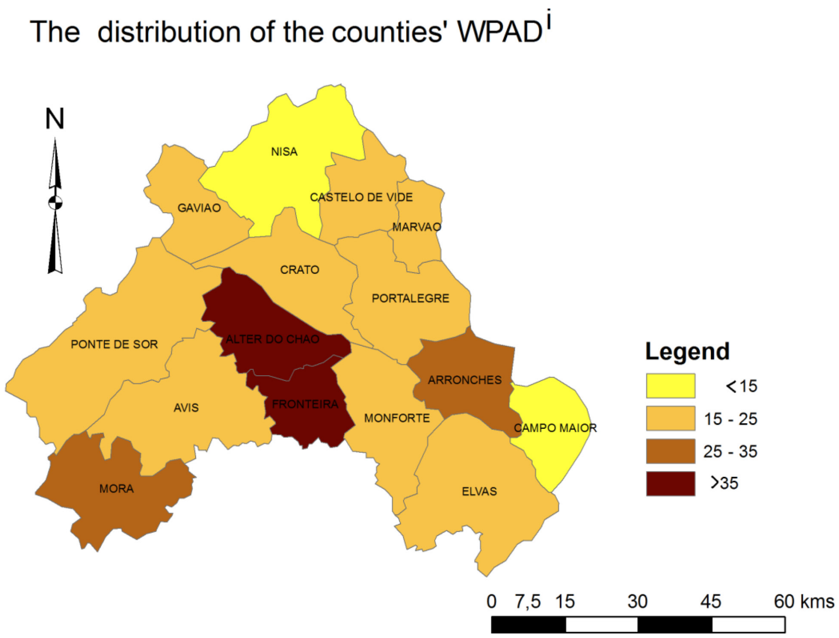

Fig. 2- WPAD^{i} distribution in the various counties of Alto Alentejo

(source: model results)

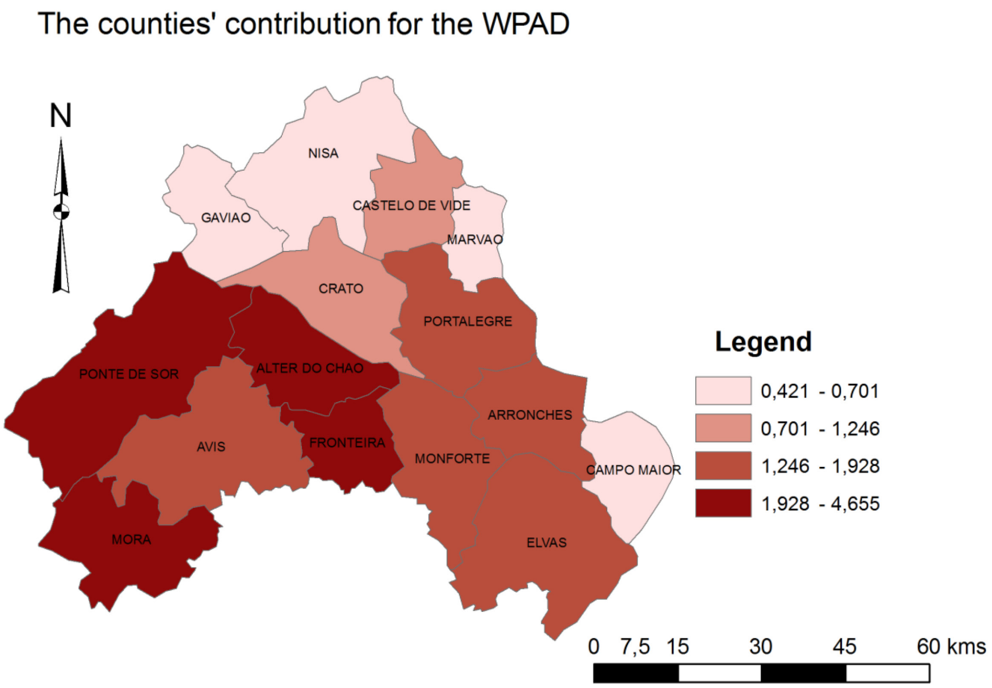

Fig. 3- The contribution for the WPAD by the various counties of Alto Alentejo

(source: model results)

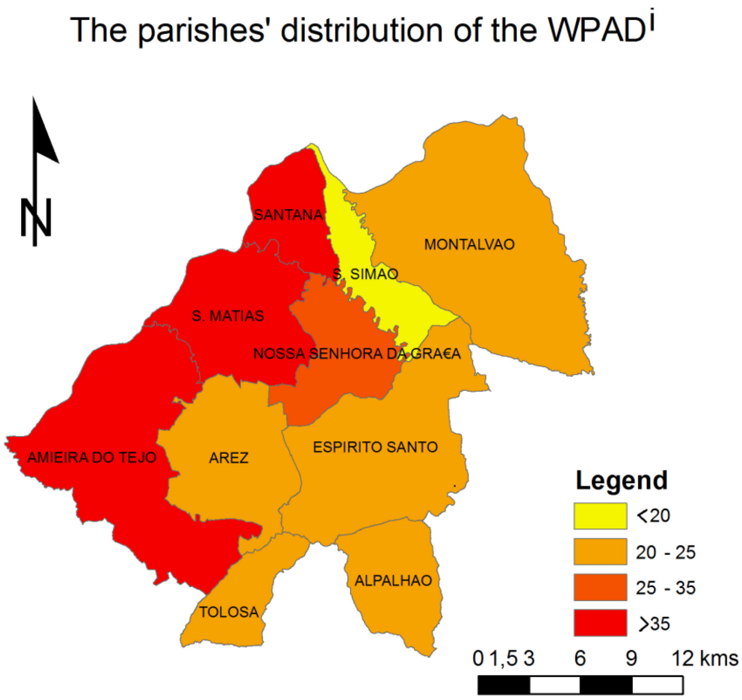

Fig. 4-- WPAD^i distribution in the parishes of Nisa

(source: model results)

The parishes' contribution for the WPAD

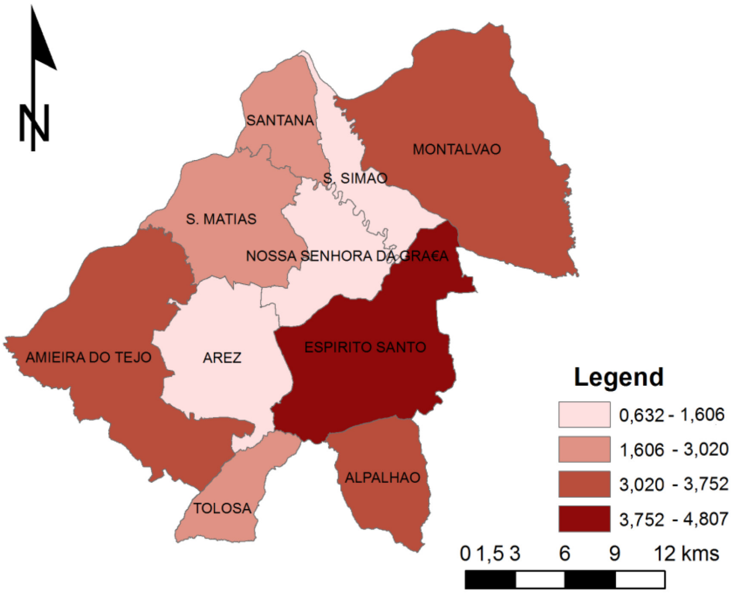

Fig. 5- The contribution for the WPAD by the various parishes of Nisa

(source: model results)