The benefits of more spatial detail in regional bioenergy models and the problems of disaggregating agricultural census data

van der Horst D.

in

Rozakis S. (ed.), Sourie J.-C. (ed.).
Comprehensive economic and spatial bio-energy modelling

Chania : CIHEAM / INRA
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 48

2002
pages 131-138

The benefits of more spatial detail in regional bioenergy models and the problems of disaggregating agricultural census data

Dan van der Horst
Centre for Environmental Strategy, University of Surrey, Guildford, UK

Abstract: Regional bioenergy models serve to match areas of high biomass production with suitable locations for plants to convert this biomass to energy. Increasingly, digital data for such models is becoming available at finer spatial scales. One important category of data still poses a problem. Agricultural statistics are collected at the farm level but only made available to researchers at the agricultural district level. This paper investigates the economic and environmental need for more spatially detailed models and identifies a number of obstacles in the disaggregation of district-level agricultural statistics. It concludes that various options for disaggregation exist but disaggregation comes at a price. The real benefits of disaggregation can only be assessed if an analysis is repeated at different levels of disaggregation.

Keywords: Agricultural statistics, disaggregation, spatial modelling, biomass energy.

Introduction: the benefits of more geographical detail in regional biomass energy models

Bioenergy systems can be modelled by using the aggregate statistics of a large geographical area (e.g. state level). The actual planning of bioenergy plants which provide heat and/or electricity, requires a more geographically detailed approach to match areas of (potentially) high biomass production with suitable locations for biomass power plants (e.g. Rozakis *et al*., 2001a; Towers *et al,* 1997; Dagnall *et al*., 2000). Models which are designed to assist in such planning are termed 'regional bioenergy models' in this paper. As with many other mixed-data, computer-based land-use models, these models are recent developments, stimulated and enabled by the rapid advances in IT technology with respect to hardware, software and data capture. Conceptually however, these models are not new. In quantitative Human Geography, such models are generically known as 'location-allocation models', or more generally 'spatial interaction models', which aim to increase efficiency of a system by matching the spatial distribution of locations of demand with the spatial distribution of locations of supply (e.g. Bailey and Gatrell, 1995; Birkin *et al.,* 1996). But what is the actual benefit of developing such regional bioenergy models, and how much spatial detail should such models ideally have?

The most important benefit, and the main reason for developing these models, lies in the fact that both the potential availability of biomass, and the potential demand for electricity or heating, are highly heterogeneous in space. If a bioenergy system is to compete with fossil fuel, then sufficient biomass must be produced in locations of low (opportunity) cost and transported over relatively short distances to the plant. Three generic stages of spatial analysis for bioenergy systems can be recognised (Figure 1):

1. Estimating the regional distribution of land suitability or potential crop yields. This can be based on existing suitability maps, a customised analysis of the relevant bio-physical variables, or observations of land use.

2. Building on this physical information, the likelihood of uptake of an energy crop or availability of crop residues must be estimated across that region. This expected land use may be based on current conditions or a change in market conditions such as a policy intervention (e.g. subsidies for energy crops).
3. Using the spatial distribution of the likely amount of available energy crops or crop residues to establish the best location for the power plant. This stage of analysis also requires the consideration of planning constraints and infrastructural necessities such as roads and substations of the national electricity grid in the vicinity (e.g. Towers *et al.* 1997, Dagnall *et al.* 2000).

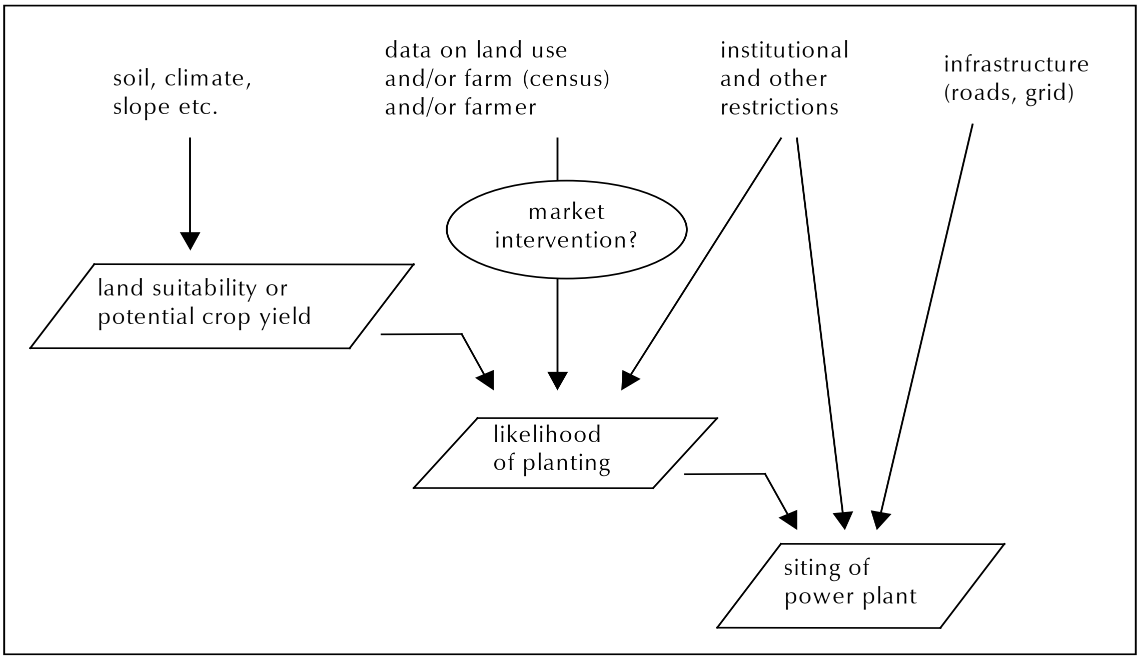

Figure 1. Three stages in spatial analysis for bioenergy systems. Parallelograms represent maps resulting from each of the three stages of the analysis.

GIS-based spatial analysis requires the availability of a number of georeferenced digital data sets. Dependent on the available data sets, the analysis can take place at a variety of spatial scales. Developers of bioenergy plants in the UK do not use GIS for plant siting. They simply select sites by comparing paper maps of roads, power lines and biomass potential [1]. Their 'eyeball' method of spatial analysis does not require a high level of spatial detail. However, maps of biomass potential (spatial analysis stage 1 or 2) are typically created in GIS, combining different spatial data-sets such as soil and land-use data. The need for such maps underlines the importance of access to detailed spatial databases. In addition, there are at least two environmental reasons to strive for a more detailed level of spatial analysis.

First of all, many of the environmental impacts of a regional bioenergy system are site dependent. It is possible to identify a range of site dependent environmental impacts or externalities of the plant, the biomass source or the transport element of the bioenergy system. These impacts are not limited to tangible emissions to air, soil and water, but can also include less tangible issues such as impacts on the landscape and biodiversity. Some of these impacts of the biomass source may even be positive. Sensitively designed and located fields of biomass crops may for example reduce the risk of erosion, provide habitat to species or cover up 'eyesores' in the landscape such as quarries and factories . In the UK guidelines have been developed for short rotation coppice (SRC) to minimise the negative impacts and stimulate the positive impacts (British Biogen, 1996). The site-dependency of environmental impacts has been acknowledged environmental economic literature which offers a number of methods to quantify these externalities in monetary values (Saez *et al.*, 1996; Groscurth *et al.*, 2000; van der Horst, 2002).

Secondly, the emissions inventory of a bioenergy system is sensitive to the mode of transport and the distance over which the fuel is transported. More detailed maps allow a better estimation of these. Since most bioenergy schemes are sponsored by the government with (amongst others) the aim to reduce green house gas (GHG) emissions, there is an implicit need to identify the 'GHG optimal' geographical catchment area of the system. This implies a need to investigate the environmental trade-offs between plant size and transport distance. LCA studies which compare different scales of production are rare. Andersson and Ohlsson (1998) did such a study for bakeries and they concluded that "there is a limit beyond which the increased efficiency that can be obtained on a larger scale production is outweighed by the environmental loads from the distribution". Hart (1997) provides an example of this trade-off between optimal economic efficiency and optimal environmental performance. He calculates the energy loss embodied in the transport of barbecue charcoal from the place of production to UK retail outlets, comparing small scale UK producers with large scale producers in Brazil or South-Africa. The transport from the economically competitive, large scale producers in Brazil or South-Africa requires 14-17% of the energy embodied in the charcoal. Small scale regional UK producers, who can not compete with the South African and Brazilian imports, requires less than 3% of the charcoal energy for transport. Börjesson and Gustavsson (1996) present a relative measure for defining the catchment area of a plant in terms of the impact of transport distance (for different modes of transport) on environmental performance. They calculate the transport distances at which the various emissions or the energy use of transport become as big as those from the actual production and harvest of the biomass crops.

The studies of Hart (1997) and Andersson and Ohlsson (1998) indicate that the environmental performance of the system is more sensitive to transport distance than the economic performance of the system would be (although the transport cost can be highly relevant for competition with fossil fuel). Transport costs are indeed only a minor component of the total cost of biomass fuel (e.g. Kallivroussis *et al.*, 1996). And because an important amount of the economic cost of transport is fixed in the (un)loading stage, the marginal cost of increased transport distance is small. This issue of optimal economic and environmental scale would merit further research. Such research would require a closer look at the performance of small scale bioenergy systems and alternative modes of transport. This in turn would require more detailed spatial data.

It is clear from the above text that spatial detail in regional bioenergy models can be desirable. This raises the inevitable question of data availability. In general, detailed spatial databases of roads or power lines are widely available these days, if not always for free. This situation can be quite different for spatially detailed agricultural data. Agricultural data can be obtained from (widely available) remote sensing imagery, but the most detailed information must be collected from the farmer directly. Agricultural statistics are collected annually from each farmer (farm census), aggregated over the whole farm over the course of a whole year. In order to 'protect the privacy of the individual farmers', this data is further aggregated to the level of an agricultural district or parish before it is made available to researcher.

The aim of this paper is to present an overview of the main issues involved in the attempts to make better use of district level agricultural statistics when modelling bioenergy systems.

Models and data requirements for estimating the distribution of available biomass

Prior to investigating how this spatial aggregation can be undone, it is essential to understand why this aggregation can be a problem for the accurate estimation of available biomass. Three conceptually different types of models can be identified for this estimation. These models can be listed in order of increasing need of different types of data: models focused on land, models focused on the farm and models focused on the farmer.

Land based models are basically the same as the first stage in Figure 1, using a land suitability classification or a crop yield model as a proxy measure of the likely amount of a crop to be produced. Bibby *et al.* (1982) provide an example of a general suitability maps for agriculture, ranking the agricultural potential of the land according to a range of biophysical variables such as soil characteristics, hydrology, slope etc.. Most of these variables have been measured in the field and converted to map form. More sophisticated are maps of estimated crop yield, wich combine these data with a crop growth model (e.g. Bateman and Lovett, 1998). Finally it is also possible to simply map the actual cultivation of the relevant crop or other crops with similar biophysical requirements. This can be done through remote sensing or the use of farm census data. Observed or modelled suitability for alternative crops can be used as a rough estimate of the opportunity cost of the land.

Farm based models place the likelihood of a crop being grown in the wider context of farm management and require the kind of data which is typically collected in the annual farm census. The simulation of decision making in farm based models is much more realistic than that of land based models. A farmer makes the choice of land use for each field as a part of an overall farm strategy. For example, issues such as crop rotation or the need of bedding material for life stock will clearly influence the farmer's choice of land use in ways that can not be anticipated with the land model only. Farm-based models for 'traditional' micro-economic analysis are typically run at the regional or national level without taking an interest in the spatial distribution of farms and their statistics within the area of aggregation (e.g. Rozakis *et al.,* 2001b). More spatially explicit farm based models require the geo-referencing (i.e. attaching x,y coordinates) of the agricultural statistics (e.g. Skop and Schou, 1999).

Farm based models will typically assume that the farmer's behaviour is that of a profit maximiser. In contrast, *Farmer* based models try to explain farmer behaviour in terms of social and demographic variables as well as economic ones. For example a farmer nearing retirement age may be more interested in SRC's low labour input (i.e. good for extensification). The long life cycle of an SRC plantation (up to 25 years) may in turn make the crop less attractive to farmers who will pass the farm on to their heirs within that time period (i.e. bad for flexibility). They may also consider landscape or wild life impacts, the benefits of wind breaks etc. Farmer based models require the same data as farm based models, plus socio-demographic data which may be derived from the population census or separate interview-based surveys (e.g. Wynn *et al*., 1998).

It is clear that the modelling of uptake of energy crops should in principle be more accurate and realistic if it takes place at the farm level and adopts a broader view of the farmer's motivation, including the possible importance of non-economic motives. This means that the farm census data which is central to the farm(er) based models, but too much spatially aggregated, needs to be linked to the type of map based data used in land based models at finer scales than the district level.

How can the access to disaggregated farm statistics be improved?

A number of solutions for improved access could be proposed. One (legal) solution would be the provision of full access after signing a confidentiality agreement. The trend towards more transparency and accountability in public policy, and the level of support to the farming sector provided by the Common Agricultural Policy (about half the EU budget), could be an argument in favour of more public access to farm data.

More technical solutions include various options which do not reveal the full amount of information to the researcher. By far the best technical solution would be the development of intelligent query tools for minimal aggregation, guaranteeing a pre-defined level of privacy. Such a system would aggregate the farm statistics on the basis of the researcher's expressed need for information, providing an (interactively operated?) aggregation which will have the lowest possible impact on the accuracy of the proposed research. While it will take time and money to develop such a system, it should be noted that the benefits would not be solely for the researchers. The current 'rigid' practice of aggregating within the often historical boundaries of agricultural districts, does not automatically guarantee privacy, and certainly does not guarantee it in equal measures. This is well illustrated by Gimona *et al.* (2000) who note that the smallest agricultural district in Scotland is only 1.3 ha (an island with one farm), more than 86,000 times smaller than the largest district.

But for the time being, researchers must come up with their own solutions to the aggregation problem of farm census data. Ideally the researcher would like access to an integrated and geo-referenced database which contains for every farm in the region at least the following information: Field boundaries, potential and current land use of each field, all fields belonging to the farm, other physical and economic farm data, other socio-demographic data about farmer and the farming household.

In some European countries, the various separate items on this shopping list do already exist in digital format. There is a growing range of digital maps of most biophysical phenomena, and derived suitability maps. The amount of remotely sensed data is also ever growing, allowing a better identification and analysis of land use and spatio-temporal land use change patterns. Digitised boundaries of administrative areas are now often in the public domain, including those of agricultural districts to which agricultural census data have been traditionally assigned. Many European countries have digitised cadastral maps which display the boundaries of land ownership at the parcel level, while population census data have been made more GIS-friendly (Martin, 1997) [2].

What are the opportunities and disadvantages of attempts to use these separate databases to disaggregate the farm census data currently available at the district level?

Using spatial databases to disaggregate district level farm census data.

Two basic types of complementary disaggregation methods can be recognised. The simplest method is to seek to reduce the size of part of the district to which the farm statistics can be assigned. This is done by identifying non-relevant areas such as water, forests or urban land. These areas can be identified from existing maps or remotely sensed imagery and removed from the map of agricultural districts.

The next step is to assign specific farm census statistics to parts of the (remaining) agricultural areas within the district. This entails the mapping of smaller sub-district areas and classifying these areas on the basis of those farm statistics which are required at a finer spatial disaggregation level. The district level statistics can then be disaggregated and assigned to the appropriate classes of the new sub-district areas. Classification of sub-district areas is mostly based on sievemapping, which is in turn based on Boolean expert systems. The most sophisticated of these classifications utilise fuzzy theory (Zhang and Stuart, 2001) or even artificial neural networks (Wang, 1994). The variety of examples in the literature is great, both in terms of the databases and the methods used. Probably the most straight forward applications are the use of land suitability maps or the use of satellite imagery to identify patters, validated by control points on the ground (e.g. Walker and Mallawaarachchi, 1998). Skop and Schou (1999) provide an example of a more experimental approach, using an expert-based association between farm type and soil class to disaggregate district level farm statistics to the relevant soil class by creating thiessen polygons based on farm house location (from a post-code map) and sized according to farm size. They acknowledge the need to extend their spatial farm-based (GIS) approach to include farmer behaviour, but this modelling gap seems to be wide open at the moment. Gimona *et al.* (2000) provide and example of a statistical approach, combining an existing land cover map which was interpreted from high resolution aerial photographs with a grid of land suitability observations on the basis of a cluster analysis of joint observations.

Problems with disaggregation

Disaggregation of farm census data takes time and effort, but that is not the only disadvantage. Additional databases may be hard to find or expensive to purchase. The overlay of different spatial databases may result in the creation of sub-district areas of unfavourable sizes (too small, too diverse) or shapes (too elongated) to disaggregate district level farm statistics to. Different databases may also have a poor compatibility because of differences in scale, mismatch of boundaries and uncertainties about data quality. Data quality may not even be consistent across a single spatial database. Gimona *et al.* (2000) demonstrate this by investigating the correlation between the district level farm statistics and the categories on a land use map derived from aerial photographs. They notice a difference in correlation between the large districts of marginal agriculture on the mountainous Scottish west coast and the more agriculturally diverse lowlands where aerial photographs were more difficult to interpret.

Another well known problem, the Modifiable Areal Unit Problem (MAUP) may occur:

1. The outcome of any spatial analysis may be scale dependent. If the same analysis is carried out on the same dataset at a different level of spatial aggregation, the result of the analysis might be different.
2. The outcome of any spatial analysis may also be dependent on the configuration of areal units (e.g. the shape of the agricultural districts). If the same analysis if carried out on the same dataset with differently shaped (boundaries of) areal units, the results of the analysis might be different.

Unlike the effects of an irregular areal unit configuration, the effects of scale can be anticipated to some extent. This anticipation is based on the comparison between the spatial scale of the analysis and the spatial scales at which the phenomena under investigation actually take place. It is advisable to identify the basic units and derive optimal scales and zonal configurations for the phenomena being studied (Openshaw, 1984). A farm(er) level model of land use change for example, should ideally be measured at the spatial scale and areal configuration of the individual field. It is also advisable to carry out a sensitivity analysis (Fotheringham and Wong, 1991), for example by comparing the outcomes of the analysis carried out at the best achievable disaggregated level, as well as at the level prior to disaggregation, i.e. the district level.

Conclusions

Most datasets can be characterised by their level of aggregation. The data collected during the annual farm census is aggregated by the farmer over a spatial area (the farm) and a time period (a year), and then further spatially aggregated to the district level. The level of 'acceptable' aggregation depends on the purposes for which the data is to be used. Improved access to more sub-district level agricultural data, may be primarily relevant from environmental perspectives rather than economic perspectives. The number of ways to disaggregate farm statistics may be as big as the number of available spatial databases with boundaries which cut through the farm district. The most advanced methods allow for a targeted disaggregation based on statistical description of the data sets or fuzzy membership functions. In practice the objectives of a study may have to be watered down because of insufficient or poor quality data. The level of dis-aggregation is therefore often a practical and pragmatic decision, taking into account the trade-off between the preferred level of spatial detail on one hand and the effort of disaggregation and reliability of the results on the other. Outcomes of spatial analysis are known to be affected by the level of aggregation. Best practice should be based on explicit descriptions of the datasets and processes under consideration. The extent to which a regional bioenergy model is actually improved by increasing the level of spatial detail through disaggregation of agricultural statistics, can only be assessed by comparing the outcomes of analyses by the model at different levels of spatial disaggregation.

Notes

[1] Personal communications with Keith Pitcher, head of First Reneables Ltd. (the developer of the ARBRE plant in Yorkshire, a 10MW wood gasification plant) and Peter Billins, chairman of British Biogen (the UK trade organisation for the bioenergy sector). It must be noted that within the UK context, competition over biomass resources such as SRC has not yet occurred. Fuel competition from existing plants will increase the need for a more sensitive siting strategy for new plants. This is likely to require a more complex GIS-based model which utilises more data at finer resolutions.

[2] In addition to the traditional farm census, a far more spatially detailed database is emerging. Under the EU's Integrated Administration and Control Scheme (IACS), farmers must

provide land use data at the field level for the administration of agricultural subsidy payments. Monitoring and control of these payments necessitates the accurate mapping of IACS field parcel data (e.g. ADAS, 2000). However this database may not have national coverage, is still off-limit to most researchers and lacks integration with the more extensive 'normal' farm census data (Gimona *et al*., 2001).

References

[1] ADAS (2000) www.adas.co.uk/env/environment/workforgov/env_govsc.htm

[2] Andersson K. and Ohlsson T. (1998) Life cycle assessment of bread produced on different scales. In: Andersson K., *Life cycle assessment (LCA) of food products and production systems*. Doctoral thesis, Chalmers University of Technology and Swedish Institute for Food and Biotechnology (SIK). AFR-report 203, Göteborg.

[3] Bailey T.C. and Gatrell A.C 1995. *Interactive spatial data analysis.* Longman Group Ltd., Harlow, England.

[4] Bateman I.J. and Lovett A.A. (1998) Using geographical information systems (GIS) and large area databases to predict Yield Class: a study of Sitka spruce in Wales. *Forestry*, 71(2), 147-165.

[5] Bibby J.S., Douglas H.A., Thomasson A.J. and Robertson J.S. (1982) *Land Capability for Agriculture*. Macaulay Land Use Research Institute, Aberdeen.

[6] Birkin M., Clarke G., Clarke M. and Wilson A. 1996. *Intelligent GIS Location decisions and strategic planning*. Geoinformation International, Cambridge.

[7] Börjesson P. and Gustavsson L. (1995) Regional production and utilization of biomass in Sweden. *Energy* 21(9), 747-764.

[8] British Biogen (1996) Short Rotation Coppice for energy production. Good practice guidelines. (pdf) www.britishbiogen.co.uk/gpg/srcgpg/srcgpgfront.htm

[9] Dagnall S., Hill J. and Pegg D. (2000) Resource mapping and analysis of farm livestock manures – assessing the opportunities for biomass-to-energy schemes. *Bioresource Technology* Vol. 71, 225-234.

[10] Fothermingham A. and Wong D. (1991) The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, Vol. 23(6), 1025-1044.

[11] Gimona A., Geddes A. and Elston D.A. (2000) Localised areal dis-aggregation for linking agricultural census data to remotely sensed land cover data. In: *Innovations in GIS 7, computational Issues,* Taylor and Francis.

[12] Groscurth H.M., de Almeida A., Bauen A. et al. (2000) Total costs and benefits of biomass in selected regions of the European Union. *Energy* Vol. 25, 1081-1095.

[13] Hart A. (1997) Charcoal for barbecue consumption within the United Kingdom: A transport energy analysis of charcoal supplied from UK coppiced woodlands and from Brazil and South African importation. BSc thesis. Queen Mary and Westfield College, London.

[14] Kallivroussis L., Rozakis S., Vassilatos V., Mellides V., Petouni D. and Kyritsis S. (1996) Crop residues as a source for decentralised heeat and power production in rural areas. The case study of Thrace. *Biomass for energy and the environment*. Proceedings of the 9th European Bioenegy Conference, Copenhagen 24-27 June 1996. Pergamon, UK.

[15] Martin D. (1997) *Census geography 2001: designed by and for GIS?* Proceedings of the GIS research UK 5th national conference, Leeds.

[16] Openshaw S. (1984) The Modifiable Areal Unit Problem. CATMOG Series, No. 38, Institute of British Geographers, London.

[17] Rozakis S., Soldatos P., Kalivroussis L. and Nikolaou E. (2001a) Multiple criteria decision-making assisted by G.I.S.: Evaluation of Bio-Electricity Production in Farsala Plain, Greece*. Journal of Geographic Information and Decision Analysis* Vol. 5(1), 49-64.

[18] Rozakis S., Sourie J.C. and Vanderpooten D. (2001) Integrated micro-economic modelling and multi-criteria methodology to support public decision making: The case of liquid bio-fuels in France. *Biomass and Bioenergy*, Vol. 20(5).

[19] Saez R., Leal J. and Linares P. (1996) An assessment of the externalities of biomass for electricity production in Spain. In: *Biomass for energy and the environment*. Proceedings of the 9th European Bioenegy Conference, Copenhagen 24-27 June 1996. Pergamon, UK.

[20] Skop E. and Schou J.S. (1999) Modelling the effects of agricultural production. An integrated economic and environmental analysis using farm account statistics and GIS. *Ecological Economics*, Vol. 29, 427-442.

[21] Towers W., Morrice J., Birnie R., Dagnall S and Aspinal R. (1997) Assessing the potential for shor rotation coppice in Scotland. Macaulay Land Use Research Institute (MLURI), Aberdeen.

[22] Van der Horst D. (2002) Spatial targeting of multi-purpose forestry. Doctoral thesis. Aberdeen University, UK.

[23] Walker P. and Mallawaarachchi T. (1997) Disaggregating agricultural statistics using NOAA-AVHRR NDVI. *Remote Sensing of the Environment* Vol. 63(2), 112-125.

[24] Wang F. (1994) The use of artificial neural networks in a geographical information system for agricultural land-suitability assessment. *Environment and planning A* Vol. 26, 265-284.

[25] Wynn, G. Crabtree, R. and Potts, J. (2001) Modelling farmer entry into the ESA scheme in Scotland. *Journal of Agricultural Economics* Vol. 52 (1): 65-82

[26] Zhang J. and Stuart N. (2001) Fuzzy methods for categorical mapping with image-based land cover data. *International Journal of Geographical Information Science*, Vol 15(2) 175-195.