Agriculture, Ecosystems and Environment 95 (2003) 465–479

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

Modelling the spatial distribution of agricultural land use at the regional scale

M.D.A. Rounsevell [a,*], J.E. Annetts [b], E. Audsley [b], T. Mayr [c], I. Reginster [a]

[a] *Department of Geography, Université Catholique de Louvain, Place Louis Pasteur 3, Louvain-La-Neuve B-1348, Belgium*
[b] *Silsoe Research Institute, Wrest Park, Silsoe, Bedford MK45 4HS, UK*
[c] *National Soil Resources Institute, Cranfield University, Cranfield MK45 4DT, UK*

Received 17 September 2001; received in revised form 10 October 2002; accepted 1 November 2002

Abstract

Agriculture is the most important land use in Europe in geographic terms and because of this it plays a central role in the quality of the wider environment. Whilst the spatial patterns of agricultural land use have changed considerably in recent times, further changes are likely as a result of the influences of policy reform, socio-economics and climate change. Understanding, therefore, how agricultural land use might respond to global environmental change drivers is a research question of considerable importance. The first step, however, in projecting potential future changes in agricultural land use is to be able to understand and represent in models both the socio-economic and physical processes that control current land use distributions.

Thus, this paper presents an approach to modelling the spatial distribution of agricultural land use at the regional scale. The approach is based on the simulation of farm-scale decision making processes (based on optimisation) and the response of crops to their physical environment. Regional scale applications of the model are undertaken through the use of spatially-variable, geographic data sets (soils, climate and topography) combined with economic data. Examples of the application of the model are given for two regions of England: the north-west and east Anglia. These regions were selected to give examples of contrasting land use systems within the context of northern European agriculture. The model results are compared statistically with observed distributions of agricultural land use for the same regions in a quasi-validation exercise. The comparison suggests that the model is very good at representing land use that is aggregated at the regional level, and at representing general spatial trends in land use patterns. Some differences were observed, however, in land use densities between the modelled and observed data.

The results suggest that the basic hypothesis of the model: that farmers are risk averse, profit maximisers, is a reasonable assumption for the regions studied. However, further study of decision making processes would be likely to improve our ability to model agricultural land use distributions. This includes, for example, the role of farmer attitudes to risk, differing views on future prices and profitability, and the effect of time lags in the decision process.

Keywords: Land use; Spatially-explicit models; Agricultural decision making; Farm model

1. Introduction

In spatial terms, agriculture is one of the most important land uses in Europe. For example, the CORINE land cover map classifies about 53% of the European land surface as agriculture. The management of this land has profound impacts on the

* Corresponding author. Tel.: +32-10-472872; fax: +32-10-472877.
E-mail address: rounsevell@geog.ucl.ac.be (M.D.A. Rounsevell).

PII: S0167-8809(02)00217-7

quality of the wider environment through, for example, nutrient dynamics, water resources and biological diversity. European landscapes have experienced rapid changes in agricultural land use throughout the second half of the twentieth century arising from developments in technology and management driven by socio-economic and political forces. These trends are anticipated to continue into the future through the effect of reforms to the Common Agricultural Policy (CAP), enlargement of the European Union, globalisation, technological change and climate change.

Scientific interest in the issue of land use change has been stimulated internationally by the Land Use and Cover Change initiative created by the International Geosphere Biosphere Programme (IGBP) and the International Human Dimensions Programme (IHDP). LUCC has produced a science/research plan (Turner et al., 1995), which provides a framework for the development of LUCC research in the future, world-wide. The science/research plan implies that the study of LUCC requires a multi-disciplinary perspective and methodologies that approach the problem from a non-traditional viewpoint. For example, integrating understanding of both the socio-economic and biophysical drivers of change is an important research objective for LUCC, and how this can be achieved in contrasting environments remains a research problem of considerable importance (Riebsame et al., 1994). The development and implementation of integrated models is central to LUCC research because models provide an opportunity to improve understanding of the processes and process interactions (feedback) within complex systems (e.g. see Lambin et al., 2000). Models can be used to evaluate the sensitivity of land use systems to different drivers of change, and identify which drivers and processes are most important. However, to effectively address land use change processes, models need to operate explicitly within space at different spatial and temporal scales. This requires the use of appropriate GIS technology, as an interface between models and spatial data, but also raises technical problems in the 'spatialisation' of models that future research needs to address (e.g. Wassenaar et al., 1999).

Agricultural land use decisions are made by farmers managing farms. In other words farm level decisions mediate the impact of market and policy change on land use. Thus, any attempt to assess agricultural land use change needs to explicitly incorporate descriptions of the goals and constraints that affect farmers and their management strategies. In terms of the development of models of agricultural land use change, this suggests an approach focused on decisions made at the farm level. It is also important to recognise, however, that whilst farmers are operating within a broadly similar economic environment (at the regional scale), but subject to their own preferences opinions and experiences, they are faced with considerable differences in the physical characteristics of the landscape, such as soil types and climates. This creates large observed differences in land use decisions, and so extrapolation of farm scale models to wider geographical regions must account for the spatial heterogeneity of the physical environment. The physical environment affects not only the potential productivity of land (i.e. crop yields), but also the constraints arising from soil tillage opportunities (Rounsevell and Jones, 1993) and pests and diseases.

This paper describes the development and implementation of a model of agricultural land use at the regional scale. The approach is based on the integration of a farm level optimisation model with spatial data sets of the biophysical environment (soils and climate). The model is demonstrated in two contrasting regions of England, east Anglia and the north-west of England (see Fig. 1), by a comparison of the model outputs against current, observed land use. In this way, we have sought to test the hypothesis that the key determinant of regional agricultural land use within the study region is 'profit maximisation'. Whilst the model has been tested against current land use data, its potential long-term value is the study of land use change arising from biophysical or policy trends extrapolated to the future. The paper describes the farm scale model and physical data sets used in the analysis, as well as explaining the methods used to integrate the model and data, and the statistical analyses used in the quasi-validation of the model results.

2. Theory

2.1. The approach

The approach presented here was based on two core models. The SFARMOD whole farm model is used

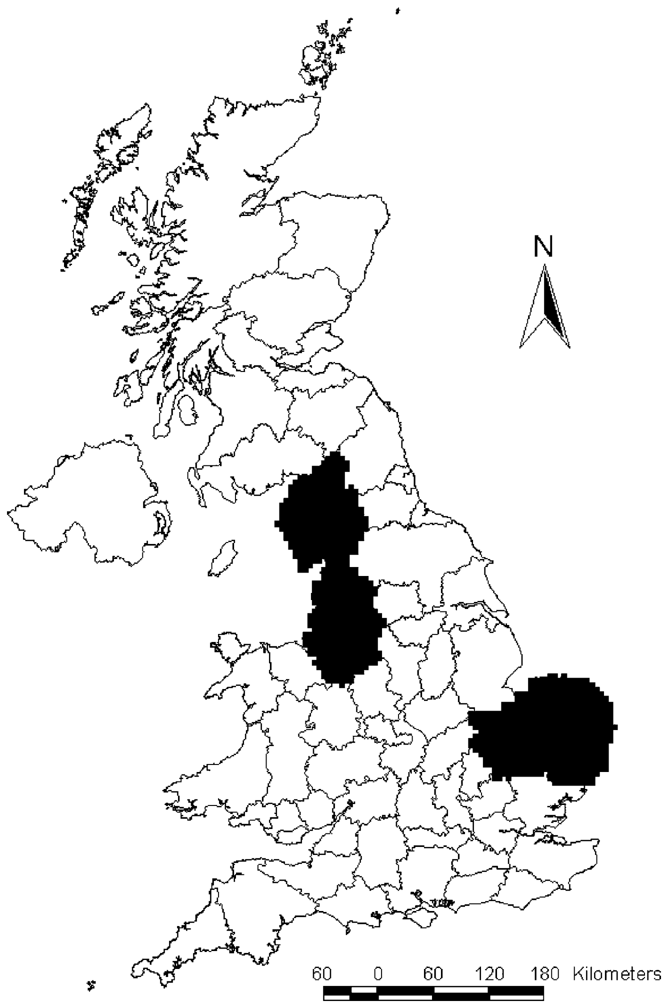

Fig. 1. The two study regions: north-west England and east Anglia.

to simulate farm decisions using an optimisation approach (Audsley, 1993). The ACCESS crop growth model (Mayr et al., 1996) is used to simulate the response of different crops to the biophysical environment (soils and climate). The relationships between these models are shown in Fig. 2, and the models themselves are described in more detail later.

2.2. Description of the farm level model

The underlying hypothesis of the farm scale modelling approach is that farmers are 'profit maximisers'. The main differences in the outcomes, in other words the actual crops grown, by different farmers is due to the soil type, climate and scale of operation. The whole farm model was previously developed for individual farms to determine the profit maximising labour, machinery and cropping (Audsley, 1993). In the application reported here, representative individual farms were created for a region in which soil type is determined by spatial location, and crop yield and soil workability are determined by the crop growth model, as a function of the soil and climate.

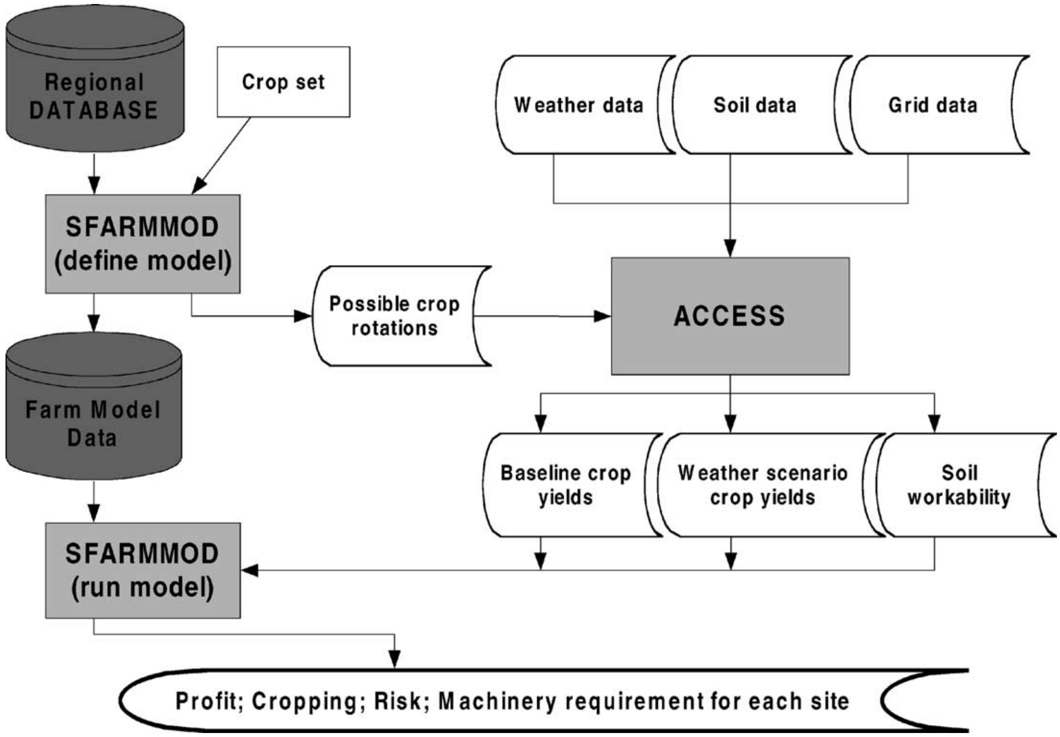

Fig. 2. Schematic of the agricultural sector model, incorporating ACCESS and SFARMOD.

The whole farm model can be viewed as a core optimisation model around which there are a number of mechanisms for providing the necessary data. The model selects the optimum steady-state farm system according to a series of constraints with an objective that is a weighted sum of the profit and a measure of risk. Data is provided either directly as a value, or indirectly as a formula. A formula expresses the value as a function of data such as soil type, crop yield, machine size, nitrogen applied, etc. A formula or data item can be technical, the work rate is a function of the tractor power and soil type, or derived from other models such as the nitrate leaching being a function of the soil type and the nitrogen not taken off in the crop.

Workable hours are an important part of the model, and these are operation specific. For example, there are fewer workable hours for spraying than for cereal harvesting and still fewer hours than for ploughing. The year is divided into periods of 2 weeks and a sub-model calculates an expected number of available workable hours in the seventh year in 10, based on the soil type and average annual rainfall. This approach has been found to be suitable for annual farm planning (Audsley, 1981). A heavy soil is never workable in the winter whereas a light soil is workable most of the winter.

A major feature of the model is the constraints arising from timeliness and rotational penalties. *Timeliness penalties* reduce the yield of a crop if, for example, it is planted late. However, because in general it would be far too expensive to plant all the crops at exactly the right time, there is an optimum (minimum) level of timeliness penalty. This level is different for different farms depending on the other crops chosen and workable hours. The model determines a work schedule, which gives the optimum amount of each operation planned for each time period. *Rotational penalties* are the reduction in yield of one crop following another, for example wheat after wheat relative to wheat after oilseed rape, due to pest and disease incidence. A database of rotational penalties has been constructed for all the possible crop sequences and the model determines the optimum rotation from this database, including the influence of the consequent effects on timeliness, labour and machinery.

The model has been designed to examine the effect of changes on a farm such as crop gross margins, new crops, machinery or techniques. A mathematical description of the model follows.The area of operation j on crop i in period k is denoted x_{ijk}. The area of crop i is a_i. The objective is to find the values of these variables that maximise the steady state profit z and the corresponding resources n_m, the number of men and machines, m:

$$z = \sum G_i a_i - \sum C_{ijk} x_{ijk} - C_m n_m \tag{1}$$

where G_i is the basic gross margin of crop i, C_{ijk} the cost of the operation, including any adjustment to the gross margin such as the reduction in yield due to carrying out the operation late, C_m the annual cost of resource m and,

$$a_i = \sum_k x_{i1k} \tag{2}$$

The resource constraints are:

$$\sum_{i,j} R_{ijkm} x_{ijk} \leq H_{mkn} n_m \quad \forall m, k, n \leq N \tag{3}$$

where, R_{ijkm} is the amount of resource m of type N required to carry out x_{ijk}, H_{mkn} the amount of resource m of type n available in period k.

Types are successively more restrictive workable hours such as for ploughing, harvesting (must be dry), spraying (must be dry and not windy) which are assumed to be sub-sets. Thus, the hours when spraying is possible are also the hours when harvesting is possible, but not vice-versa.

Sequencing constraints ensure that a sequential operation is not carried out before its preceding operation.

$$\sum_{k \leq K} x_{ijk} \leq \sum_{k \leq K} x_{i(j-1)k} \quad \forall i, j > 1, k \tag{4}$$

where $K \in (P_{ij} \cap P_{ij-1})$, P_{ij} is the set of periods in which operation j on crop i can be carried out. If the intersection K is null, the constraint is simply the sum over all periods.

For non-sequential operations,

$$a_i = \sum_k x_{ijk} \tag{5}$$

When $j = 1$, the above sequence constraints refer to the previous crops in the rotation. Define r_{ick} to be the area of crop i following crop c in period k.

$$\sum_i r_{ick} = x_{cJk}, \qquad \sum_{k \leq K} x_{i1k} \leq \sum_c r_{ick} \tag{6}$$

Each crop is a member of a disease class, P^d, which affects the rotation possibilities. There is a loss of yield penalties (from zero to 'not allowed') for crops following particular disease class crops. The annual build up of a disease is offset by the growth of crops not encouraging that disease. The build-up value B_d is the minimum number of years between crops of that disease class. There may also be an associated disease class, which does not provide a break from the build-up of the disease. It is assumed that for an associated disease, d' the build-up value is the smaller of B_d or $B_{d'}$. Then the constraint for disease d is:

$$\sum_{i \in P^d} a_i B_d + \sum_{i \in P^{d'}} a_i \min[B_d, B_{d'}] - \sum_{i \notin P^d \cup P^{d'}} a_i \leq 0 \quad \forall d \tag{7}$$

Rotational penalties are also subtracted from the objective (Eq. (1)).

In simple terms, the sum of the crops must be less than the land available. However this constraint must take into account the possibilities of more than one crop per year and more than 1 year per crop. Thus, the area of land occupied by a crop or between crops, at any time, must be less than or equal to the area of land available for crops. Crops include permanent crops such as grazing, perennial crops such as forage, annual crops such as wheat, rape and set-aside, and catch crops.

Let t_{ic} be the total area of land transferring from crop i to crop c. Define the start of transfer be the first period of the last operation of crop i and the end of transfer be the last period of the first operation of crop c. Note that land must either be in crop or being transferred; there is no overlap. At any particular period of the year, one can calculate the land use as the sum of crops and transfers between crops. An arbitrary, but useful, period is the year-end. Let N_i be the number of year ends an annual, perennial or catch crop i crosses, P be the set of permanent crops and Δ_{ic} have value one if the transfer from crop i to c crosses

Table 1
Simple numerical example of a farm model

	a (crop)		x (crop) (operation) (period)								r (from) (to) (per)			n (machine)		
	a1	a2	x111	x112	x113	x122	x123	x134	x211	x225	r114	r124	r215	n1	n2	
Objective—Eq. (1)																
(E1)	982	1526	−12	−12	−12	−14	−54	−13	−12	−17	−150	0	0	−18000	−4152	
Area of crop (crop)—Eq. (2)																
(E2)1	−1		1	1	1											=0
(E2)2	−1								1							=0
Total area of land—Eq. (8)																
(E8)	1	1														=200
Machine hours (machine) (period) (type)—Eq. (3)																
(E3)111			0.5						1.5					−68		≤0
(E3)121				0.5		0.8								−122		≤0
(E3)122						0.8								−95		≤0
(E3)131					0.5		0.8							−134		≤0
(E3)132							0.8							−105		≤0
(E3)141								1.6						−142		≤0
(E3)151										3.2				−106		≤0
(E3)241								0.8							−142	≤0
(E3)251										1.6					−106	≤0
Sequencing (crop) (operation) (version)—Eq. (4)																
(E4)111			1	1	1						−1		−1			≤0
(E4)121			−1	−1	−1	1	1									≤0
(E4)122				−1	−1	1	1									≤0
(E4)123					−1		1									≤0
(E4)131						−1	−1	1								≤0
(E4)211									1			−1				≤0
(E4)221									−1	1						≤0
Rotational sequencing (crop) (period)—Eq. (6)																
(E6)14								−1			1	1				≤0
(E6)25										−1			1			≤0
Rotational constraint (disease)—Eq. (7)																
(E7)1	−2	1														≤0

(En) refers to equation n in the text. The example has two crops (hence the subscript i takes two values 1 and 2), one with three operations ($j = 1, 2, 3$) and the other with two ($j = 1, 2$). There are five time periods considered ($k = 1, \ldots, 5$). The first and second operations of the crop 1 overlap (Eq. (4)). The second operation requires better workability ((E3)122) and there is a penalty if it is carried out in period 3 (x1231 and (E1)). There are two machines (n1 and n2), one used for all operations (cf a tractor) and the other used only for the final operation on each crop (cf a harvester). Crop 1 can follow itself (penalty 150), but crop 2 can only follow crop 1 (the r columns). Crop 2 is limited to 1 year in 3.

the year end, otherwise zero. Then Eq. (8) defines the land use.

$$\sum_{i\in P} a_i + \sum_{i\notin P} N_i a_i + \sum_{ic} \Delta_{ic} t_{ic} \le T \quad (8)$$

Additional constraints can be added to represent features such as sugar beet quotas, which limit the amount of the operation that can be carried out in any period. Modifications can allow alternative methods of carrying out operations, such as different sizes of tractor or contractors. This is a linear programme, which can be rapidly solved on modern computers.

Table 1 shows a small numerical example.

3. The database

The whole farm model has a comprehensive database, which covers a wide range of common crops. The database consists of two main levels: a level of general farm data and an operational level.

3.1. Farm level input data

The location of a farm has a big influence on its performance and profitability. Environmental factors, such as soil texture and climate influence crop yield, the soil workability (Rounsevell and Jones, 1993; Rounsevell and Brignall, 1994), the risk of diseases and the fertiliser requirement. In the farm model, nine soil textures are defined, varying from light soil (sandy) to heavy clay. The climate is described by Annual Average Rainfall (AAR). High rainfall means the soil returns to field capacity earlier in the autumn, after which work is rarely possible, which results in fewer workable days during the year. A heavy soil is never workable in the winter whereas a light soil is workable most of the winter. Thus, these two parameters have a strong influence on the scheduling of farm operations. For each 2 weekly period a certain number of hours are available to work the land, defined by soil moisture, temperature, daylight and available labour hours.

3.2. Operation level input data

For every crop, input data specify crop parameters, such as yield, fertiliser rates, and information about the required operations that have to be performed to grow the crop. In addition, available machinery and labour capacity are important features for the calculation of the crop distribution over the available arable area, optimised over farm profit. A list of the model-input variables, on the operational level is given in Table 2.

Table 2
Summary of operational input variables for the farm model

	Variables
Crop	Gross margin: yield (primary and secondary), prices, seed rate, fertiliser rates, sprays and costs Husbandry operations: list of operations and their feasible timing e.g. plough, cultivate, drill, spray, fertilise, harvest, bale Timeliness penalties: extra cost or loss of yield of doing operation at other than the optimal timing Rotational penalties: reduction in yield from one crop following another, including impossible
Operations	Work rate as a function of size of machinery and amount of e.g. yield or fertiliser or soil type Workable hours as a function of soil type, rainfall, time of the year and labour availability Machinery needed (i.e. size, power)
Machinery and labour	Cost: capital, repairs, fuel, resale value Replacement interval

4. Running the model over regions

The whole farm model was initially developed to calculate the optimal cropping for one farm. However, an approach has been developed to apply the model on a regional basis using gridded soil and climate data at a resolution of 5 km × 5 km (although finer resolutions would also be possible). The basis of the approach is the assumption that a generic, model farm is representative of the sum of the farms within each grid cell. The cropping can then be estimated for each generic farm/grid, and these values mapped to show spatial distributions of land use or aggregated to larger spatial units (e.g. administrative regions) if required.

On an individual farm a farmer would select a sub-set of the large number of possible crops. Over a region however the average individual farm can include a small percentage of a crop. Thus, when calculating the distribution of crops over a grid, it is important to take all the crops into account. For one farm it might not be profitable to allocate 2% of the land to a certain crop, but over larger regions this area could be relevant. However, many crops are grown in very small areas for very specific customers. Thus, the following major arable crops and grass were considered: wheat, winter barley, spring barley, spring oats, winter oilseed rape, spring oilseed rape, linseed, winter beans, spring beans, dried peas, potatoes, potatoes (100 mm irrigation), potatoes (200 mm irrigation), sugar beet, sugar beet (100 mm irrigation), sugar beet (200 mm irrigation), maize, sunflower, soybean, grass, permanent grass and forage maize.

Sugar beet is a special case in a regional study as its profitability is partly determined by the existence of a local factory. To examine the full range of possible futures, this restriction was removed to determine the level of need for a factory.

The following, spatially-variable inputs are required by the model:

- soil textural types;
- primary yield;
- workable hours as a function of soil texture and climate.

4.1. Soil textural types

The soils data were based on the range of soil series that occur within each 5 km × 5 km grid square of the rasterised National Soil Map of England and Wales at a scale of 1:250,000 (Avery, 1980). These data are held within the Land Information System (LIS) of the Soil Survey and Land Research Centre, Cranfield University. In a single 5 km × 5 km grid square, several soil-types can occur, and the model was run for each of these types.

4.2. Primary yield

Primary crop yields are influenced by the location of the grid for several reasons including the soil type and climate. Spatial crop yield data do not exist within England at the resolution required by the model. It is necessary, therefore, to model the spatial variability of crop yields. This was undertaken using an existing crop growth simulation model named ACCESS (Mayr et al., 1996; Wassenaar et al., 1999, etc.). The model was applied to the soils and climate data within each 5 km × 5 km grid square and these outputs scaled according to a comparison of the model results with available crop yield values (for specific sites). In this way, the crop model provided a representation of the spatial heterogeneity of crop yields, but the yield values were referenced to real observations. Crop models are an important part of the approach presented here because they allow responses to climate change (e.g. arising from temperature, precipitation and atmospheric CO_2 changes) to be assessed.

The crop model simulates the soil moisture content, the soil nitrogen content, the temperature and the radiation, as the level of each affects the crop's growth. Thirty years were simulated for all of the soil types occurring within each 5 km × 5 km grid square. The model outputs comprised:

- the crop dry matter yield each year;
- the crop maturity date—if at the end of the year the crop has not matured, there is no crop yield;
- the nitrogen used on the crop;
- the water used on the crop, which is applied as required up to the maximum specified—if there is insufficient need for water it is not used;
- the daily water draining from or running off the soil;
- the nitrate-N content of the soil each day.

4.3. Workable hours

The workable hours for each 5 km × 5 km grid were estimated from the soil texture and climate using the simulated soil moisture content for 30 years. The results were calibrated by comparing the estimates with an ADAS survey of workable days, the estimates in Nix (1999) and ABC (1999). The hours in the seventh best year in 10 were then calculated.

4.4. Other spatially variable inputs to the farm level model

The rotational penalties influence the primary yield, but the penalties themselves were considered to stay constant over all of the grids. The inputs for fertiliser and irrigation were also considered constant.

A grid has a surface of 5 km × 5 km, or 2500 ha. The area of a grid, therefore, is larger than the area of a single farm. The cropping solution on the representative farm is assumed to estimate the aggregate cropping of the farms in the grid. As the whole surface of the grid is not available for arable crops, because a certain amount is used for housing, some area can be water, etc. the total area available to agriculture was corrected as follows:

$$\text{Available area grid} = 2500 \times \left(\frac{100\% - \text{urban}\,\% - \text{woodland}\,\% - \text{sea}\,\% - \text{lake}\,\%}{100} \right)$$

In addition some grids have no soil type recorded and are assumed to be non-agricultural, some areas are at best marginal for all crops and are also assumed to be non-agricultural, though some could be included in census data as low quality rough grazing. Some areas are only suitable for grass and are assumed to be grass. The model does not estimate forestry.

The model was run for the year 1995 in the regions of east Anglia and north-west England (see Fig. 1). Crop data such as prices, seed rate, fertiliser rates, spraying costs and extra variable costs per crop were derived from ABC (1999). These prices were based on the average values (corrected for inflation) for the 4 years before the model run year, i.e. 1991–1994. The economic scenario was different prior to 1991. This attempts to simulate farmer decision-making, which is based on previous experience and not the prices in a current year (which will not be known at the start of the year). Thus, for example, a farmer will select a crop in 1995 if it has performed well (economically) in previous years. A 4-year period, although largely arbitrary, was chosen because it was thought to be representative of the past period that contributes to decision-making. Differences in the length of time for price data are unlikely to be important unless the prices for a particular crop are especially volatile, or especially poor in the most recent year. However, when a change in price policy occurs within this time period the data may need to be adjusted.

The variability of a farmer's perception has been included in the model to better estimate the expected distribution of cropping within a region. Conventionally in farm economics, the optimum cropping on a farm is assumed to be conservative in order to smooth out peaks in profit and loss from variable crops. This can be optimised using techniques such as MOTAD (Hazell and Norton, 1986), which minimise the total absolute deviation at the same time as maximising profit. Whilst this is undoubtedly the case on an individual farm, on a collection of farms the variability causes some crops to be grown, even though on estimated profit and variability grounds they should not be grown. Put simply, the prices and yields that a farmer expects are different from the mean values, due to past experience and personal prejudices. Thus, the optimal decisions that farmers reach will be different. To simulate this variability in expected net crop incomes, due to different perceptions of prices and yields, a series of solutions have been determined with randomly generated crop gross margins, using the coefficient of variation derived from historical series of prices and yields of crops. The regional cropping is thus the sum of these different optimal cropping plans. The method gave an improved estimate of the distribution of cropping in the east Anglian region.

5. Quasi-validation

A *quasi-validation* of the land use model outputs was undertaken by comparing statistically the model results for 1995 with observed statistics provided by the UK Ministry of Agriculture, Fisheries and Food (MAFF, 1996). The MAFF data were based on the annual census data that have been interpolated to the same 5 km × 5 km resolution grid used for the model, and were provided from MAFF by the UK Climate Impacts Programme (UKCIP) (Iain Brown, personal communication). The analysis is referred to as a quasi-validation because the MAFF data are not truly an 'observed validation set', but an 'estimated' land use distribution, the data being interpolated to the 5 km × 5 km grid from observations collected on a different spatial basis. In addition, the MAFF data refer to a single year (1998) that is different from the modelled baseline (1995). There may not be large differences between the land use distributions for these two periods, but some differences are possible. The analysis reported here, therefore, should only be considered as a 'comparison' and not a complete validation of the model outputs. This comparison does, however, provide some useful information concerning the behaviour of the model under current conditions, which affects the way the model results can be interpreted.

Because the land use model generates large amounts of output (the surface area of each of the crops considered in the analysis), it is necessary to simplify some of these data when comparing the model and observed distributions statistically. Therefore, the following variables were used in the statistical analysis: wheat (ha), cereals (wheat, barley and oats, ha), grassland (permanent and rotation, ha) arable (i.e. non-grassland, ha). The statistical comparisons between model and observed data were based on the following tests. Before applying these tests, the data distributions were tested for normality.

1. A visual comparison of the mapped observed and modelled spatial distributions.
2. A simple matching coefficient was derived from frequency tables. The coefficient is calculated from the number of grid squares that are common between the compared distributions (1–1, 0–0) divided by the total number of grid squares in the region. The coefficient is, thus, a measure of the degree of agreement between the compared land use.
3. A comparison of the means and standard deviations for the variables over the whole of the region.
4. The comparison of the means was based on (a) the difference between the means, and (b) a paired *t*-test. The paired *t*-test computes the difference between each grid square of the compared distribution and tests if the average differs from 0. The probabilities of the *t*-test are also determined: a probability value of 0.001 means that there is only a 0.1% chance that there is no difference between the measured and computed values, i.e. there is a 99.9% chance that there is statistically significant difference between the two. This is a test on the regional means that compares the different distributions in a global way. A probability level of 0.001 is a severe threshold, indicating that there is a very high degree of confidence that there is a difference between the two means. Probabilities of 0.05 (equivalent to 5 %), however, still demonstrate significant differences. The following discussion is based on the probability threshold of 0.001.
5. The Pearson correlation coefficient is a measure of the degree of association between the distributions of the values being compared. It is calculated from the covariance of the compared scenarios divided by the product of the standard deviations of the two distributions. Pearson correlation coefficients vary between -1 and $+1$, indicating either negative or positive linear relationships, respectively. A value of 0 indicates that neither of the two distributions can be predicted from the other using a linear equation. A value of 1 indicates a perfect match.

6. Results and discussion

Visual comparison of the mapped results for the modelled and observed land use distributions (see Fig. 3) shows relationships that, generally, are good. That is to say the spatial trends observed in the two distributions appear to have similar patterns, and the geographic concentration of each land use class is located in similar places on each map. There are also clear differences between the two regions, with arable agriculture being more prevalent in east Anglia compared with the north-west.

It is interesting to note from Fig. 3 that the model produces a land use pattern that is more spatially-distributed than the MAFF data. This is consistent with a well-recognised observation that actual land use tends to become concentrated in certain areas, reflecting the influence of neighbouring land uses on farmer choices (e.g. White and Engelen, 1993, 1997). In some cases this can result in actual land use distributions being significantly different from those which might be expected from a knowledge of the physical conditions (soils and climates). Classic geographic studies (e.g. Hägerstrand, 1968) have shown the influence of this so-called *neighbourhood effect* on land use distributions, but how this type of behaviour might change in the future is unknown.

A comparison of the regional means for the different land use classes in each grid square supports the observation of a good correspondence between modelled and observed data (see Table 3). The *t*-test further corroborates this conclusion (see Tables 4 and 5), with the exception of the grassland distributions, which show significant differences from the MAFF data for both regions and cereals in the north-west (indicated by the values 0.000 in brackets in Table 5). Table 3 also shows, however, that there are large differences in the standard deviations between the modelled and observed distributions, implying a greater variability in land use classes in the MAFF data compared

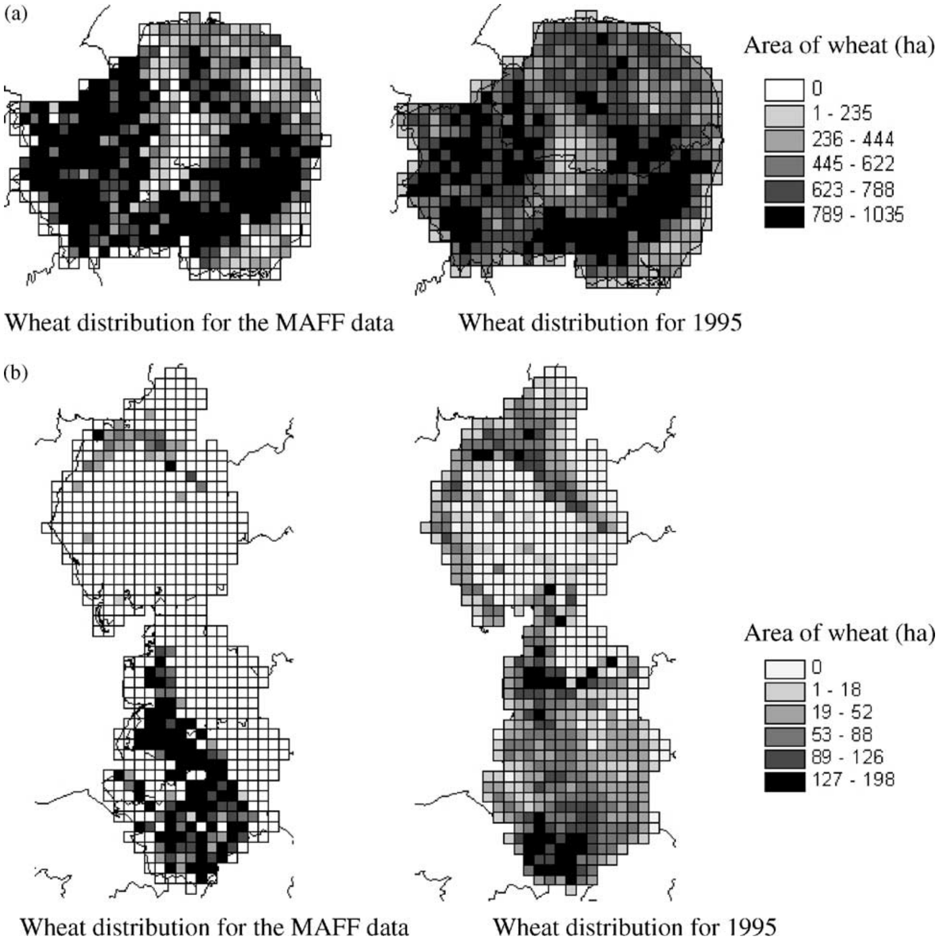

Fig. 3. The distribution of wheat in (a) east Anglia, and (b) the north-west for the MAFF data and the 1995 modelled baseline.

Table 3
Comparison of the regional means and the S.D. (ha) in east Anglia and the north-west region

	East Anglia		North-west	
	Wheat MAFF cropping statistics	Wheat Modelled (1995)	Wheat MAFF cropping statistics	Wheat Modelled (1995)
Mean	627	612	41	39
S.D.	444	238	101	44
	Cereals MAFF cropping statistics	Cereals modelled (1995)	Cereals MAFF cropping statistics	Cereals modelled (1995)
Mean	864	881	104	72
S.D.	501	322	187	78
	Grassland MAFF cropping statistics	Grassland modelled (1995)	Grassland MAFF cropping statistics	Grassland modelled (1995)
Mean	161	52	864	750
S.D.	160	145	692	545

Table 4
Tests on the mean differences (t-test) between the baseline and the MAFF data in east Anglia

	Wheat MAFF cropping statistics	Cereal MAFF cropping statistics	Grassland MAFF cropping statistics
Wheat modelled (1995)	15.062 (0.332)		
Cereal modelled (1995)		17.580 (0.351)	
Grassland modelled (1995)			−108.610 (0.000)

The number in brackets represent the level of statistical significance of the test.

Table 5
Tests on the mean differences (t-test) between the baseline and the MAFF data in the NW region

	Wheat MAFF cropping statistics	Cereal MAFF cropping statistics	Grassland MAFF cropping statistics
Wheat modelled (1995)	−2.145 (0.567)		
Cereal modelled (1995)		−31.256 (0.000)	
Grassland modelled (1995)			−114.316 (0.000)

The number in brackets represent the level of statistical significance of the test.

Table 6
Pearson correlation coefficient between the Baseline and the MAFF data in east Anglia

	Wheat MAFF cropping statistics	Cereal MAFF cropping statistics	Grassland MAFF cropping statistics
Wheat modelled (1995)	0.58		
Cereal modelled (1995)		0.50	
Grassland modelled (1995)			0.27

with those simulated. The Pearson correlation coefficients, which are small in each case, support this observation, because the correlation depends strongly on the standard deviation (see Tables 6 and 7). This suggests, therefore, that the model is very good at the level of regional means and at representing general spatial trends, but is less good at representing the full variability of land use density that is observed in the MAFF data.

This conclusion is emphasised by Figs. 4 and 5, which show the aggregated cropping outputs for the entirety of each region. For east Anglia the results are largely good, although the areas of grass, potatoes and beans are underestimated and the areas of oilseed rape and oats are overestimated. For the north-west, the modelled area of grass is 83% against the observed area of 87%. Thus, the model estimates the area of arable cropping quite well, with the exception of sugar beet (see Fig. 5). Sugar beet, however, is a special case. In the simulations presented here there were no restrictions imposed on sugar beet cropping as a result of the lack of a nearby sugar beet factory (e.g. arising from additional transport costs). The model correctly simulates that there is insufficient area of sugar beet

Table 7
Pearson correlation coefficient between the Baseline and the MAFF data in the NW region

	Wheat MAFF cropping statistics	Cereal MAFF cropping statistics	Grassland MAFF cropping statistics
Wheat modelled (1995)	0.34		
Cereal modelled (1995)		0.41	
Grassland modelled (1995)			0.15

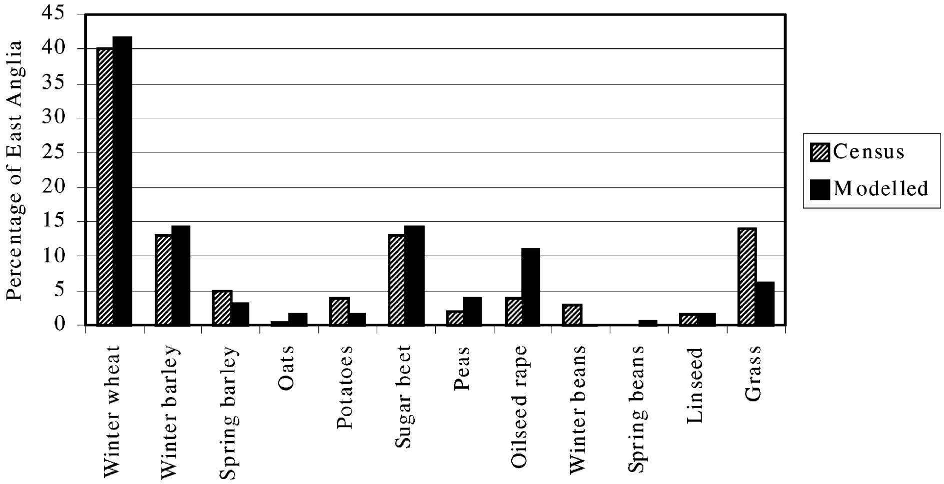

Fig. 4. Actual versus modelled agricultural land use in the east Anglian region in 1995.

to justify a nearby factory and thus the crop is not profitable in the north-west region. As the model only predicts an area of 3.2% of sugar beet, there is little impact on the overall simulation of the other crops. As for east Anglia, the areas of potatoes and beans are underestimated and the area of oats is overestimated. In this case however the area of oilseed rape is underestimated.

Any differences between the observed and modelled land use distributions may either be a function of the land use model itself, or alternatively because of the MAFF data interpolation procedure. It is not possible to say which. The results of the modelling exercise suggest, however, that the hypothesis that farmers in the two regions are 'profit maximisers' is a good assumption. It is, however, apparent that in attempting to predict farmer cropping choices that neighbouring farmers, in identical situations do not make identical choices. There are two alternative explanations for this. One is that the farmers have different attitudes

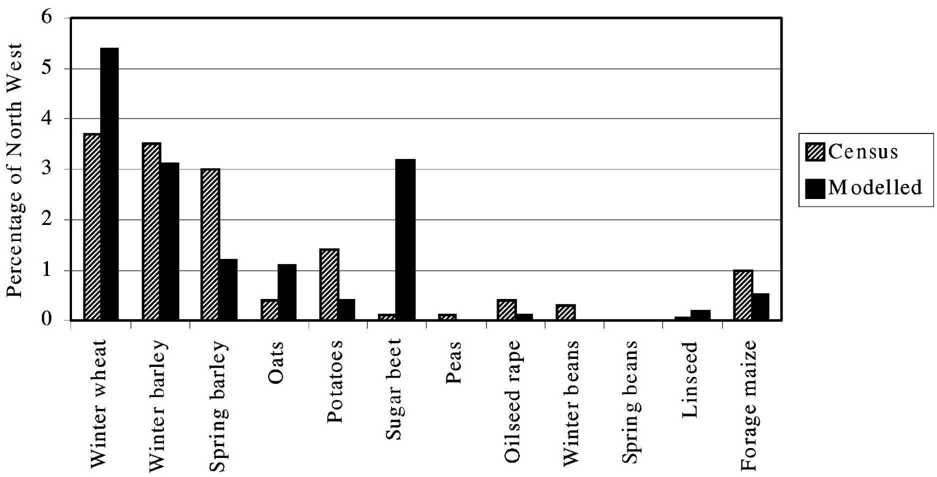

Fig. 5. Actual versus modelled agricultural land use in the north-west region in 1995 (note: sugar beet is modelled as if there was a local factory in order to examine future possibilities).

to risk and the second that they have different views on the likely future profitability (yield and prices) of crops. Both are due to the variability in yields and prices. Thus, we believe that risk aversion is a pivotal question in the study of the decision-making process of the farmer (e.g. Hazell and Norton, 1986). Generally speaking, risk aversion is revealed in the choice of crop division and techniques through crop diversification and the adoption of crops with low profit variability. However, estimating the role of this parameter is fairly difficult as crop diversification also occurs for other reasons, such as agronomic and organisational needs. It is, therefore, difficult to establish unequivocally if taking this risk into account when constructing programming models is appropriate, and further research is required to evaluate whether modelling risk perception would be useful to the model reported here. An alternative approach is to consider that farmers have a personal view of the likely future value of a crop, which differs from farmer to farmer and there is thus more variety of choice between farmers. For example, potato is currently a poor crop to grow and also very risky and so, no farm model should select it. Yet some farmers do grow potatoes. This is due to past experience of shortage years and very high prices, which cause some farmers to have higher expectations from potatoes than most other farmers.

A further possible explanation of the differences between the observed and modelled crop areas is the effect of rates of change (or time lags) in the decision-making process. As the price of a crop changes through time, there is not an instantaneous response from the farmer to this new price level. It may take several years of higher/lower prices for a farmer to be convinced that it would be profitable to switch to a different crop. Such an effect might be exaggerated if the change involved a crop that was new to an area and for which, there was no or little collective experience within the local farming community. This situation might reflect an aversion to risk, or the effect of existing short or long duration capital equipment (not included in the model reported here). As an example of the short case, a farmer with a potato planter and harvester would be more likely to continue growing potatoes than a farmer would be to start growing potatoes who does not have the necessary equipment. As an example of the long case, a farmer with a grain store, but no dairy parlour, will most probably continue to cultivate cereals, and not change to dairy farming. The effect of long duration capital equipment could be modelled as the probability of a change given the pressure to change which is the difference in the profitability of the alternative systems. Further work would be required to develop such approaches.

7. Conclusions

The basic hypothesis that underpins the work presented here is that farm level decisions mediate the impact of market and policy change on land use. The model has been developed, therefore, in considering this basic hypothesis. Testing the performance of the model against observed land use statistics suggests that the approach has merits that warrant further investigation and application. Amongst such applications are the capacity to use the model to explore future changes in agricultural land use arising from socio-economic and climate change. In this respect, the model appears appropriate not only to address short-term changes in market conditions, e.g. arising from European policy reform, but also as a tool to analyse the implications of longer-term climatic change, e.g. over the next 50 years. This is because the basic concept of modelling farmer decisions as the basis of understanding agricultural land use is unlikely to change during this period of time. Climate change, for example, will affect both crop yields and the types of management constraints, such as soil workability, that are explicitly treated by the model. In predicting future land use, however, much greater uncertainty exists in the values of the economic inputs to the model, e.g. prices, to which the model (like reality) is very sensitive. The capacity to 'predict' future agricultural land use change will depend strongly, therefore, on an understanding of the economic processes and changes arising from a range of causes: policy change, technological development, consumer preferences, etc.

As a contribution to the understanding of land use change, the model has important strengths in terms of its ability to represent management and decision processes. These approaches are based, however, on the assumption that a certain range of land use choices are available to farmers. This does not take account of potential future changes to these possible options. Already in Europe, for example, agricultural policy has the stated aim of encouraging diversification of rural land use, and this policy objective might be expected to continue into the future. Models of agricultural land use change need to consider, therefore, non-agricultural land use options such as forestry, conservation and residential development.

Acknowledgements

The authors would like to thank the European Commission (DGXII Framework IV Programme) and the UK government (Ministry of Agriculture Fisheries and Food/Department of the Environment, Transport and Regions) who funded this work within the IMPEL (an Integrated Model to Predict European Land use) and RegIS (Regional Climate Change Impact and Response Studies in east Anglia and north-west England) projects, respectively.

References

ABC, 1999. ABC The Agricultural Budgeting and Costing Book. Agro Business Consultants Ltd.

Audsley, E., 1981. An arable farm model to evaluate the commercial viability of new machines or techniques. J. Agric. Eng. Res. 26 (2), 135.

Audsley, E., 1993. Labour, machinery and cropping planning. In: Annevelink, E., Oving, R.K., Vos, H.W. (Eds.), Farm planning. Labour and labour conditions. Computers in agricultural management, Proceedings of the XXV CIOSTACIGR V Congress, Wageningen, Netherlands, pp. 83–88.

Avery, B.W., 1980. Soil Classification for England and Wales (higher categories). Soil Survey Technical Monograph no. 14, Harpenden.

Hägerstrand, T., 1968. Innovation Diffusion as a Spatial Process. University of Chicago Press, Chicago.

Hazell, P.B.R., Norton, R.D., 1986. Mathematical Programming for Economic Analysis in Agriculture. Macmillan, New York.

Lambin, E., Rounsevell, M.D.A., Geist, H., 2000. Are agricultural land use models able to predict changes in land use intensity? Agric. Ecosyst. Environ. 82, 321–331.

MAFF, 1996. The Digest of Agricultural Census Statistics, London, United Kingdom 1995.

Mayr, T.R., Rounsevell, M.D.A., Loveland, P.J., Simota, C., 1996. Agroclimatic change and European soil suitability: regional modelling using monthly time-step data. Int. Agrophys. 10, 155–170.

Nix, J., 1999. Farm Management Pocketbook. Imperial College, Wye.

Riebsame, W.E., Meyer, W.B., Turner, B.L., 1994. Modelling land-use and cover as part of global environmental change. Climatic Change 28, 45–64.

Rounsevell, M.D.A., Jones, R.J.A., 1993. A soil and agroclimatic model for estimating machinery work days: the basic model and climatic sensitivity. Soil Tillage Res. 26 (3), 179–191.

Rounsevell, M.D.A., Brignall, A.P., 1994. The potential effects of climate change on autumn soil tillage opportunities in England and Wales. Soil Tillage Res. 32, 275–289.

Turner, B.L., II, Skole, D., Sanderson, S., Fisher, G., Fresco, L., Leemans, R., 1995. Land Use and Land Cover Change: Science/Research Plan. IGBP Report no. 35 and HDP Report no. 7.

Wassenaar, T., Legacherie, P., Legros, J.-P., Rounsevell, M.D.A., 1999. Modelling wheat yield responses to soil and climate variability at the regional scale. Climate Res. 11, 209–220.

White, R., Engelen, G., 1993. Cellular dynamics and GIS: modelling spatial complexity. Geograph. Syst. 1, 237–253.

White, R., Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. Environ. Plan. B 24, 235–246.