Spatial Disaggregation of Agricultural Production Data: An Econometric Approach using Minimum Cross Entropy

Raja Chakir

Introduction
Context
The Disaggregation model
Land use model estimation
Land use Disaggregation
Application to the Picardie region
Data
Results
Conclusion and research perspectives

Spatial Disaggregation of Agricultural Production Data: An Econometric Approach using Minimum Cross Entropy

Raja Chakir

February 21th 2006

Outline

1. Introduction
2. The Disaggregation model
   - Land use model estimation
   - Land use Disaggregation
3. Application to the Picardie region
   - Data
   - Results
4. Conclusion and research perspectives

Context

- The latest reform of the Common Agricultural Policy (CAP) aims to encourage environmentally friendly farming practices in order to preserve the quality and the diversity of rural areas.

Context

- The latest reform of the Common Agricultural Policy (CAP) aims to encourage environmentally friendly farming practices in order to preserve the quality and the diversity of rural areas.
- It is important to evaluate very precisely the impacts of this policy on : Agricultural land use, space management and environment.

Context

- To study precisely the behavior of the farmers it is necessary to carry out microeconomic analysis with quite localized individual data

Context

- To study precisely the behavior of the farmers it is necessary to carry out microeconomic analysis with quite localized individual data
- It is also necessary to take into account the physical variables (soil, climate and altitude) in economic modelling for better analyzing the effects of public policies

Context

- To study precisely the behavior of the farmers it is necessary to carry out microeconomic analysis with quite localized individual data
- It is also necessary to take into account the physical variables (soil, climate and altitude) in economic modelling for better analyzing the effects of public policies
- Example: an accurate study of non-point water nitrogen pollution generated by agriculture needs to use precise individual data on land use and to take into account spatial variance in physical variables (soil, climate) and farming systems.

Why do we need a spatial disaggregation (1)

- To deal with the lack of data at the disaggregated level.

Why do we need a spatial disaggregation (1)

- To deal with the lack of data at the disaggregated level.
- Just(2000): "...agricultural economics research must focus on decision making at the farm level rather than continue to demonstrate points and methodology with aggregate data simply because they are available..."

Why do we need a spatial disaggregation (1)

- To deal with the lack of data at the disaggregated level.
- Just(2000): "...agricultural economics research must focus on decision making at the farm level rather than continue to demonstrate points and methodology with aggregate data simply because they are available..."

- $\Longrightarrow$ **A valid disaggregation method that helps to face the lack of data problem would be very useful.**

Why do we need a spatial disaggregation (2)

- To better study the interaction between economic models and biophysical models.

Why do we need a spatial disaggregation (2)

- To better study the interaction between economic models and biophysical models.
  - It is necessary to take into account the pedoclimatic variables (soil characteristics, the climate and altitude) in economic modeling to better analyze farmer behavior and the effects of public policies.

Why do we need a spatial disaggregation (2)

- To better study the interaction between economic models and biophysical models.
  - It is necessary to take into account the pedoclimatic variables (soil characteristics, the climate and altitude) in economic modeling to better analyze farmer behavior and the effects of public policies.
  - Usually, economic data concerning farms are available at the aggregated scale whereas the pedoclimatic data are provided on a finer level.

Why do we need a spatial disaggregation (2)

- To better study the interaction between economic models and biophysical models.
  - It is necessary to take into account the pedoclimatic variables (soil characteristics, the climate and altitude) in economic modeling to better analyze farmer behavior and the effects of public policies.
  - Usually, economic data concerning farms are available at the aggregated scale whereas the pedoclimatic data are provided on a finer level.
- $\Longrightarrow$ **A disaggregation method is necessary to define a compatible scale between different sources of data.**

Spatial Disaggregation: Literature Survey

- Disaggregation methods have been widely used in other fields : climate science (*downscaling*), political science (*ecological inference*)...

Spatial Disaggregation: Literature Survey

Spatial Disaggregation of Agricultural Production Data: An Econometric Approach using Minimum Cross Entropy

Raja Chakir

Introduction
Context
The Disaggregation model
Land use model estimation
Land use Disaggregation
Application to the Picardie region
Data
Results
Conclusion and research perspectives

- Disaggregation methods have been widely used in other fields : climate science (*downscaling*), political science (*ecological inference*)...
- Only few recent papers have dealt with spatial disaggregation in agricultural economics.

Spatial Disaggregation: Literature Survey

- Disaggregation methods have been widely used in other fields : climate science (*downscaling*), political science (*ecological inference*)...
- Only few recent papers have dealt with spatial disaggregation in agricultural economics.
- Howitt and Reynaud(2003) have proposed a dynamic spatial disaggregation model of land use of a Californian farmers sample observed at aggregated level (region) to a disaggregated (district) level. The land use model was defined as a Markov process and the outcome of the regional model was disaggregated using the Maximum Entropy (ME) approach.

Spatial Disaggregation: Literature Survey

- Disaggregation methods have been widely used in other fields : climate science (*downscaling*), political science (*ecological inference*)...
- Only few recent papers have dealt with spatial disaggregation in agricultural economics.
- Howitt and Reynaud(2003) have proposed a dynamic spatial disaggregation model of land use of a Californian farmers sample observed at aggregated level (region) to a disaggregated (district) level. The land use model was defined as a Markov process and the outcome of the regional model was disaggregated using the Maximum Entropy (ME) approach.
- You and Wood (2004) have proposed a spatial disaggregation model for crop production statistics based on a cross-entropy approach. They used various information sources (satellite, FAO/IIASA database biophysical crop suitability data, population density) to disaggregate Brazilian crop production data to a pixel level.

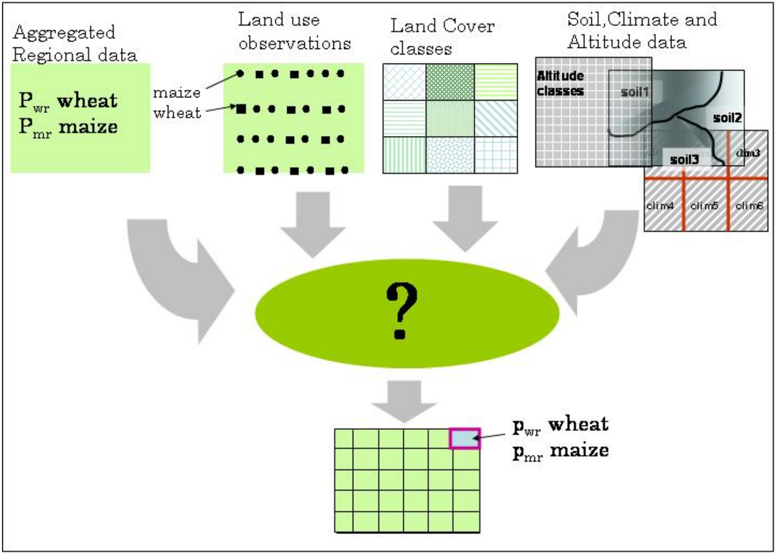
Spatial Disaggregation: Literature Survey

- Disaggregation methods have been widely used in other fields : climate science (*downscaling*), political science (*ecological inference*)...
- Only few recent papers have dealt with spatial disaggregation in agricultural economics.
- Howitt and Reynaud(2003) have proposed a dynamic spatial disaggregation model of land use of a Californian farmers sample observed at aggregated level (region) to a disaggregated (district) level. The land use model was defined as a Markov process and the outcome of the regional model was disaggregated using the Maximum Entropy (ME) approach.
- You and Wood (2004) have proposed a spatial disaggregation model for crop production statistics based on a cross-entropy approach. They used various information sources (satellite, FAO/IIASA database biophysical crop suitability data, population density) to disaggregate Brazilian crop production data to a pixel level.
- Kempen et al. (2005) have used a spatial disaggregation procedure combining a logit model with posterior density estimators to break down production data available at the regional level to a homogeneous spatial mapping unit (HSMU) level for the entire EU.

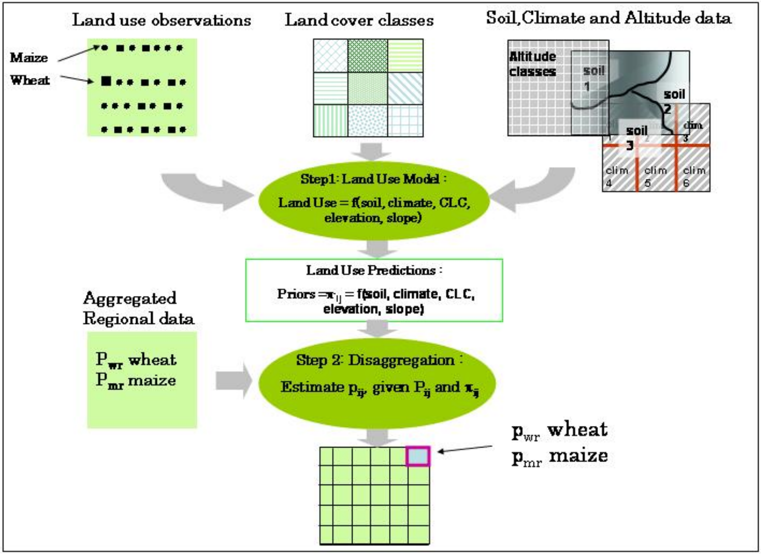
The objective of the paper

- Our paper shows how partial land use data available at the disaggregated level can be combined with biophysical data to constitute more complete information at the disaggregated level.
- The objective of the disaggregation is not to perfectly match the real world, rather to derive a fairly more informative idea of the spatial distribution of production of individual crops than available regional data allow for.

The problem illustrated for a FADN region



The proposed solution



The Disaggregation model: 2-step procedure

Introduction
Context
The Disaggregation model
Land use model estimation
Land use Disaggregation
Application to the Picardie region
Data
Results
Conclusion and research perspectives

- Step 1: Land Use Model Estimation
  - Use data on land use available at the disaggregated level combined with biophysical data to estimate "priors"
  - Land use=f(price, subsidies, soil, rain, temperature, altitude, slope, time)

The Disaggregation model: 2-step procedure

- Step 1: Land Use Model Estimation
  - Use data on land use available at the disaggregated level combined with biophysical data to estimate "priors"
  - Land use=f(price, subsidies, soil, rain, temperature, altitude, slope, time)
- Step 2 : Land use disaggregation
  - Estimate a consistent solution of the land use allocation given the priors and the aggregated data

Land use model : Economic specification

- Land use models have been widely used in agricultural production economics.
- We use a simple static profit maximization model under risk neutrality (Wu and Segerson, 1995).
- Consider a farmer that has L_i acres of land of type i $(i = 1, ..., I)$ $(L = \sum_i L_i)$.
- Land types are distinguished by a number of different biophysical characteristics that affect soil productivity (soil type, altitude, slope,etc).
- For each land type, the farmer must decide how to allocate the L_i acres to each land use c $(c = 1, ..., C)$.

Land use model : Economic specification

- We assume that, for each land type, the farmer chooses the land allocation that maximizes total profit:

$$\max_{l_{ic}} \sum_{c=1}^{C} \pi_{ic}(x, l_{ic}) \tag{1}$$

subject to:

$$\sum_{c=1}^{C} l_{ic} = L_i \tag{2}$$

where x is a vector of exogenous input prices, crop prices and other economic decision variables.

- The solution to this problem gives the optimal land allocation $l^*_{ic} = l_{ic}(x, L_i)$ for land of type i .
- Then the optimal share of land of type i allocated to crop c is

$$s^*_{ic}(x) = \frac{l^*_{ic}}{L_i} \tag{3}$$

Land use model : Econometric specification

- We assume here that the share equations take the logistic form (Considine and Mount, 1984)

Land use model : Econometric specification

- We assume here that the share equations take the logistic form (Considine and Mount, 1984)
- The multinomial logit MNL model estimate the probability outcome associated with each category of land use depending on a set of explanatory variables.

Land use model : Econometric specification

- We assume here that the share equations take the logistic form (Considine and Mount, 1984)
- The multinomial logit MNL model estimate the probability outcome associated with each category of land use depending on a set of explanatory variables.
- The share of the crop c at the location i which is equivalent to the probability of observing the crop c at location i can be expressed as:

$$s^*_{ic}(x) = P_{ic} = \frac{exp(\beta^{'}_{c} x_{ic})}{\sum_{c=1}^{C} exp(\beta^{'}_{c} x_{ij})}, \quad (4)$$

Land use model : Econometric specification

- We assume here that the share equations take the logistic form (Considine and Mount, 1984)
- The multinomial logit MNL model estimate the probability outcome associated with each category of land use depending on a set of explanatory variables.
- The share of the crop c at the location i which is equivalent to the probability of observing the crop c at location i can be expressed as:

$$s^*_{ic}(x) = P_{ic} = \frac{exp(\beta'_c x_{ic})}{\sum_{c=1}^{C} exp(\beta'_c x_{ij})}, \tag{4}$$

- The log-likelihood function of a sample of size N is given by:

$$\ln(L(\beta)) = \sum_{i=1}^{N} \sum_{c=1}^{C} lnP_{ic}^{y_{ic}}, \tag{5}$$

where y_{ic} is a dummy variable such that $y_{ic} = 1$ if crop c is observed at location i and $y_{ic} = 0$ otherwise.

Disaggregation: Cross Entropy approach

- The estimation of the land use model by the Multinomial Logit will provide us with some "priors" about the land share of each crop c in each pixel i.

$$\hat{\pi}_{ic} = \frac{exp(\hat{\beta}_c' x_{ic})}{\sum_{c=1}^{C} exp(\hat{\beta}_c' x_{ic})}, \tag{6}$$

Disaggregation: Cross Entropy approach

- The prior probabilities $\hat{\pi}_{ic}$ will enable us to derive probabilities p_{ic} to observe crop c in pixel i by solving the linear optimisation program:

$$\min_{p} \; CE(p_{ic}, \hat{\pi}_{ic}) = \sum_{i=1}^{I} \sum_{c=1}^{C} p_{ic} \; ln(p_{ic}/\hat{\pi}_{ic}) \tag{7}$$

subject to:

$$\sum_{i=1}^{I} p_{ic} \times s_i = S_c^{FADN}, \forall c = 1, ..., C \tag{8}$$

$$\sum_{c=1}^{C} p_{ic} = 1, \forall i = 1, ..., I \; and \; p_{ic} \in [0, 1], \tag{9}$$

where

- s_i is the area of pixel i
- S_c^{FADN} is the area allocated to crop c at the regional level according to the FADN observations.

Data description

Table: Data Description

Data	Description	Resolution
FADN	Accountancy data of professional farms	Regional level
Climate	Climate data	50 km x 50 km
Soil	Soil characteristics	1km x 1 km
DEM	elevation	90m x 90m
CLC	Land cover	250m x 250m
LUCAS	land use/ land cover	1 PSU each 18 km and 10 SSU in each PSU

- All data layers are converted and overlaid on a 100m x 100m pixel grid basis.
- Land use estimation and desegregation are conduced within this geo-referenced database.

FADN (Farm Accountancy Data Network)

- The FADN data is collected every year and concerns all EU member states. The annual sample covers approximately 90 % of the total utilized agricultural area (UAA) and accounts for more than 90 % of the total agricultural production of the Union.

FADN (Farm Accountancy Data Network)

- The FADN data is collected every year and concerns all EU member states. The annual sample covers approximately 90 % of the total utilized agricultural area (UAA) and accounts for more than 90 % of the total agricultural production of the Union.
- The collected data concern physical and structural data (crop areas, labor force, etc) economic and financial data (the value of production of the different crops, stocks, sales and purchases, production costs, assets, liabilities, production quotas and subsidies)

FADN (Farm Accountancy Data Network)

- The FADN data is collected every year and concerns all EU member states. The annual sample covers approximately 90 % of the total utilized agricultural area (UAA) and accounts for more than 90 % of the total agricultural production of the Union.
- The collected data concern physical and structural data (crop areas, labor force, etc) economic and financial data (the value of production of the different crops, stocks, sales and purchases, production costs, assets, liabilities, production quotas and subsidies)
- Remarks about the FADN:

FADN (Farm Accountancy Data Network)

- The FADN data is collected every year and concerns all EU member states. The annual sample covers approximately 90 % of the total utilized agricultural area (UAA) and accounts for more than 90 % of the total agricultural production of the Union.
- The collected data concern physical and structural data (crop areas, labor force, etc) economic and financial data (the value of production of the different crops, stocks, sales and purchases, production costs, assets, liabilities, production quotas and subsidies)
- Remarks about the FADN:
  - The data is available at the level of administrative units (NUTS II). This level is not appropriate to carry out precise environmental studies;

FADN (Farm Accountancy Data Network)

- The FADN data is collected every year and concerns all EU member states. The annual sample covers approximately 90 % of the total utilized agricultural area (UAA) and accounts for more than 90 % of the total agricultural production of the Union.
- The collected data concern physical and structural data (crop areas, labor force, etc) economic and financial data (the value of production of the different crops, stocks, sales and purchases, production costs, assets, liabilities, production quotas and subsidies)
- Remarks about the FADN:
  - The data is available at the level of administrative units (NUTS II). This level is not appropriate to carry out precise environmental studies;
  - We need to localize the information more precisely to be able to study the effects of public policy;

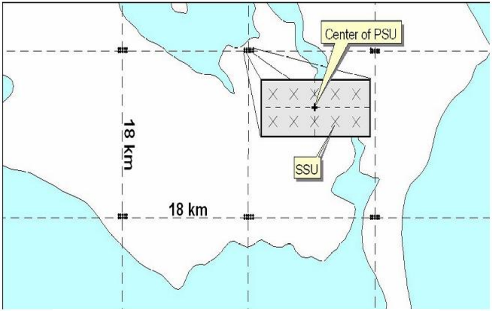
FADN (Farm Accountancy Data Network)

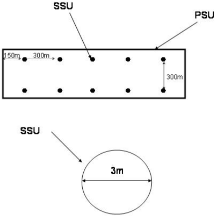
- The FADN data is collected every year and concerns all EU member states. The annual sample covers approximately 90 % of the total utilized agricultural area (UAA) and accounts for more than 90 % of the total agricultural production of the Union.
- The collected data concern physical and structural data (crop areas, labor force, etc) economic and financial data (the value of production of the different crops, stocks, sales and purchases, production costs, assets, liabilities, production quotas and subsidies)
- Remarks about the FADN:
  - The data is available at the level of administrative units (NUTS II). This level is not appropriate to carry out precise environmental studies;
  - We need to localize the information more precisely to be able to study the effects of public policy;
  - The data contains no information on topography, pedology and climate. This kind of information is necessary to study the effects of agricultural practices on the environment.

LUCAS: Land Use/Cover Area frame statistical Survey

- LUCAS: Conducted by Eurostat to obtain at EU level, harmonized data of the main Land Use/Cover areas.
- LUCAS is a two-step survey :
  - Step 1: a field survey in spring to collect data on land cover/use as well as the environment
  - Step 2 : a farmer interview conducted in autumn to gather information on yields and agricultural practices and techniques
- The LUCAS covers the entire EU and uses a systematic area frame sampling with a two-stage sampling design. The total land area of Europe is divided into a $18km \times 18km$ grid. Primary Sampling Units (PSU) are defined as cells of this grid while Secondary Sampling Units (SSU) are 10 observation points evenly distributed across the centre of each PSU. Within each PSU the SSU make up two rows of 5 points each. All points are 300 m apart from each other. The SSU, which are represented as circles of $3m$ in diameter, are units under investigation in the step I of the LUCAS survey.

LUCAS two-stage sampling design





Land use Statistics

Table: Land Use in the Picardie Region (Sources: FADN and LUCAS

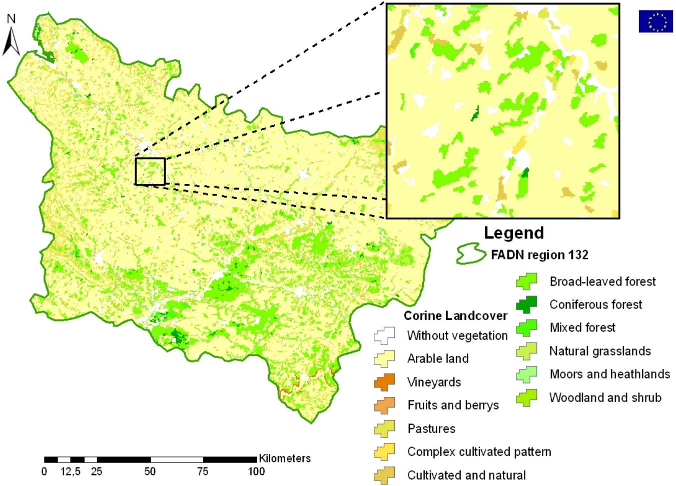
Land use	Abbreviation	% FADN	% LUCAS
Wheat-Barley-Rape	WBR	40.15	37.91
Root Crops	RC	10.21	8.77
Grassland	GL	5.72	14.24
Fallow Land	FL	3.69	2.65
Maize	MA	4.71	4.64
Vegetables and Flowers	VF	1.25	0.83
Temporary Grassland	TG	0.53	1.32
Dry Pulses	DP	0.02	2.81
Permanent Crops	PC	0.01	0.50
Other Crops	OC	6.12	-
Non-agricultural use	NAU	27.60	26.32

CORINE Land Cover

Introduction
Context
The Disaggregation model
Land use model estimation
Land use Disaggregation
Application to the Picardie region
Data
Results
Conclusion and research perspectives

- Corine (COoRdination of INformation on the Environment)land Cover is a geographical database that provides EU wide geo-referenced data.
- Corine Land Cover (CLC) provide a map of the European environmental landscape based on an interpretation of satellite images. CLC provides comparable digital maps of land cover for each country in Europe. This is useful for environmental analysis and comparisons as well as for policy making and assessment.

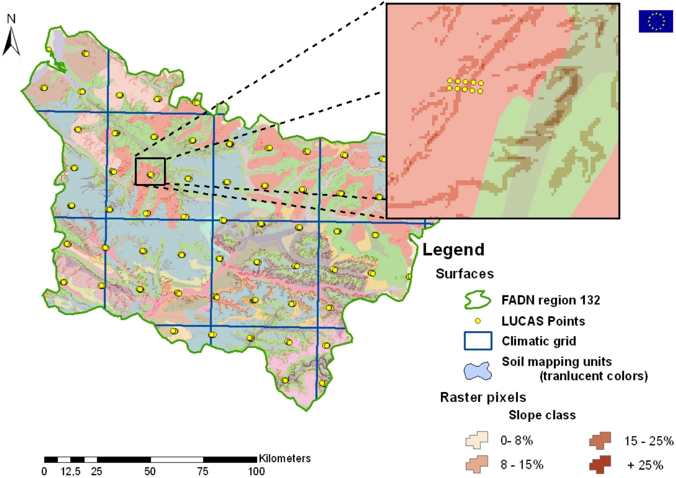
Corine Land cover data



Soil, Climate, DEM data

- Soil: type, texture, carbon content, etc
- Climate : temperature (min, max), pluviometry, etc
- DEM : slope and elevation.

LUCAS, Soil, Climate and DEM data



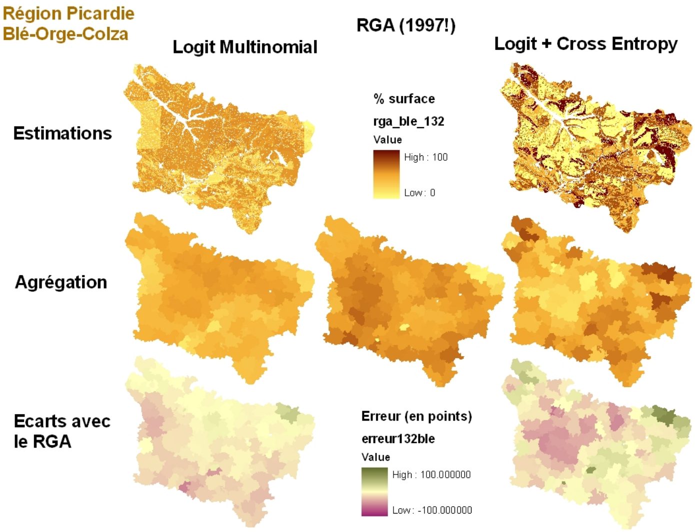
Land use Model Estimation results

Disaggregation results

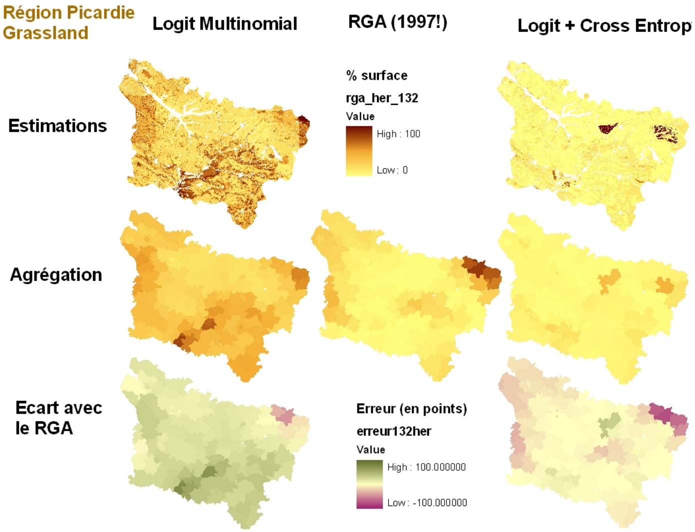
Spatial Disaggregation of Agricultural Production Data: An Econometric Approach using Minimum Cross Entropy

Raja Chakir

Introduction
Context
The Disaggregation model
Land use model estimation
Land use Disaggregation
Application to the Picardie region
Data
Results
Conclusion and research perspectives



Disaggregation results



Spatial Disaggregation of Agricultural Production Data: An Econometric Approach using Minimum Cross Entropy

Raja Chakir

Introduction
Context
The Disaggregation model
Land use model estimation
Land use Disaggregation
Application to the Picardie region
Data
Results
Conclusion and research perspectives

Conclusion and research perspectives

- Very preliminary results to be improved

Conclusion and research perspectives

- Very preliminary results to be improved
- The approach is to be generalized to the EU territory

Conclusion and research perspectives

- Very preliminary results to be improved
- The approach is to be generalized to the EU territory
- The proposed model can be enhanced in several ways.
    - Add the spatial autocorrelation in the MNL land use model
    - Consider the land use choice as a dynamic process (rotations, risk aversion, etc.)

Comments welcome

Thank you for your attention. Any comments, remarks or suggestions are welcome!!