Improvement of the AROPAj model covering a large range of agricultural activities at wide (UE) and high resolution (mapping of farm types) scales

Edouard Baranger, Melissa Clodic, Elodie Galko,

Pierre-Alain Jayet, Paul Zakharov

UMR INRA210-AgroParisTech, BP01, Centre INRA Versailles-Grignon, F – 78850 Grignon

Contact : jayet@grignon.inra.fr

Abstract

Mathematical programming models suit policy makers' needs for assessing impacts of new policies and modellers' needs for new parameters, constraints and activities to represent the phenomenon under study. The AROPAj model aims at taking into account any geographical extension of the European Union as well as the continuously changing Common Agricultural Policy. As well as the European policies will include more and more environmental problems, AROPAj will be more and more designed toward environmental assessments. Two improvement axes are set up. Biophysical models are used to refine the design of farming systems represented in the economic model and to estimate nitrogen pollutant emissions. Down-scaling methods are coupled with the regional farm group models, aiming at providing results at a fine resolution scale useful for policy evaluation and for consistent analysis of environmental impacts calling for physical models. This helps to deal with new challenges like renewable energy production consistently with agricultural potential and environmental requirements.

Key words: supply model, EU, linear programming, agriculture, environment, coupling, down-scaling

1. Aimed objectives and modelling requirements

1.1. Policy design and impact valuation for the agricultural sector and the environment

Agricultural policies and agricultural economy opened or broadened the field of modelling in different ways. A large range of models were developed, from farm-level model based on mathematical tools related to operational research until to macro-econometrical sector models inspired by the general equilibrium framework. After seminal works initiated by series of precursors, Nerlove or Heady among them, some synthetic compilations are regularly provided on modelling work applied to the agricultural economics (among last ones, see Heckelei et al., 2001, and Arfini, 2005).

Any of the type of model takes advantage of new development in methodology as well as in computer sciences. Focusing on models based on mathematical programming (MP), modellers take advantage of progress in linear and non linear optimization regarding efficient software, allowing for instance the development of positive mathematical programming (PMP). This last way of modelling tends to represent now the main stream regarding models based on optimization methods, beside econometrical models (Howitt, 1995). Nevertheless the choice of the class of model strongly depends on available data, which remains one of the most important "limiting factors" of modelling. Among optimizing models, PMP seems to become the standard procedure particularly when information highly lacks even if the estimates of numerous parameters based on a few data let econometricians doubtful. On the other side, linear programming (LP) extended towards integer variables offers the possibility of solving very large models compatible with accurate representation of complex technical systems. It goes on to remain an efficient modelling tool adapted to large systems belonging to different sectors (oil refinery, transport industry, distribution networks, and multi-scale agricultural systems). At last,

the final user of modelling results would ever be better when results come from quite different approaches.

Let us focus on mathematical programming which allows modellers supplying of results related to "future contexts", in the same time which are not in the trend expanded from the past and which are realistic regarding the characteristics of the modelled systems. Regarding the agricultural sector, the major reason which steers the choice in modelling tools remains the trade off between large databases and accurate "production sets" on one hand and more sophisticated expressions (i.e. non linear introduction of variables in the objective function) included in models using less informative contend on the other hand. Another important reason is the trade off between the ability of PMP to provide calibrated solution very close to the observation (i.e. areas devoted to different crops) and the easiness of LP to deliver large sets of primal and dual variables in quite different contexts.

Numerous successive changes in European agricultural policies as well as important potential change in the global environment will strongly impact on farming systems. The prior assessment of these impacts should be of high interest for economic agents and policy makers when they intend to modify the management of systems that they have in charge. Cost-benefit analysis should be conducted on models themselves, including the high costs of data collection. Advantages will be increased by model competition which benefits users and requires that different models remain steadily operational. On the cost side, models paradoxically take advantage of more and more abundant data requested by the implementation of policy tools (standard CAP and compensated good agricultural practices lead to keep steady the European Farm Accounting Data Network –FADN– and physical databases).

Instead of higher computation costs (software and hardware) induced by mathematical programming models compared to standard econometric sector models, MP will go on supplying added value in term of scientific contend and of results. Obviously the choice in modelling depends on available data and on goals assigned to the models. The key question could not be the selection of a class of models (MP versus econometrics, or LP versus PMP). It could be the capacity of making operational linking and coupling of models dedicated to the representation of economical and physical phenomena running at different scales and leading to difficult problems of global consistency.

1.2. Requested bases and characterisation of the modelling concept

There is no evident hierarchy among the characteristics of a model. There will be a continuous trade-off between upgradeability in a broad sense and working control. Clearly a model should remain something working with respect to user-friendliness and to "good agreeable economical conditions". Regarding upgradeability some rules seem to be useful. First of all something which ran should remain working after upgradeability. Modeller should accept not to use all available information if the related data are not reproducible and not extendable to the entire area covered by the model.

The "short-term agricultural supply" AROPAj model was developed thanks to incentives delivered by different institutions and research programmes (i.e. GENEDEC[1]). It relies on FADN which is considered as a database reproducible along the time and covering a large range of crops and animal activities in all member States (MS) of the European Union (EU). The basic farm model is designed in such a way that it is adaptable to different European farm types. All these characteristics make the AROPAj model generic and flexible as long as it allows us to easily include new crops and new

---

[1] http://www.grignon.inra.fr/economie-publique/genedec/

member States (see section 2). Strong investment in modelling code was initially expended in order to preserve these characteristics. Parameter management is particularly paid by attention in the model structure.

Now we turn towards improvement of modelling around the AROPAj model through the use and/or the coupling with other models.

The AROPAj model is strongly based on technical relations between activities making us sure of the feasible meaning of results regarding the physical properties. But some of the weakest points of the model can still be improved thanks to the improvement of generic biophysical models. Adjustable yields and N-inputs should change their role in the model, when these initial parameters are changed into optimised variables. This significant gap is based on the use of the generic crop model STICS developed by the French Institute of Agronomic Research. It leads us to insert yield functions in the AROPAj model (see section 3).

The lowest scale of statistically representative information delivered by the FADN is the Regional one. The geographical dimension is a key point which strongly characterizes this model. One of the major stakes is to make it able to provide disaggregated results at a scale lower than the Regional scale. This is now possible thanks to reproducible physical databases provided by the Join Research Centre when information includes accurate geographical reference (see section 4).

AROPAj is a supply model. It means that the entire price system is included like parameters. We know that excellent general or partial equilibrium models are continuously improved or kept in steady conditions. Even if we can remain doubtful regarding the difficulty in forecasting of several years increasing world prices of raw materials (i.e. energy, agriculture, metals), we consider that it is more efficient to use the existing models to feed AROPAj with prices. Nevertheless the European agriculture makes unrealistic the "small country assumption" regarding the weight of this agriculture in the world one. Some attempts of coupling between AROPAj and this kind of models were initiated during the GENEDEC programme. This is not presented in this paper. We just insist on the multi-scale consistency of economic models like we face this consistency problem in other fields of modelling (i.e. dynamic climate modelling regarding the different scales and grids of spatial and time analysis).

Finally the AROPAj model should be considered like an element of a more general modelling architecture. Flexible internal structure of AROPAj as well as flexible coupling with different categories of models lead to conceivable improvement of analysis in the agri-environmental field, as shortly presented in section 5.

2. The benchmark model and its recent evolution

2.1. Presentation of the model structure

The benchmark version of the AROPAj model is now the version developed during the European FP6 programme referred by the GENEDEC acronym. Even if the model does not cover all farm types and all agricultural products, it represents a large part of the used agricultural area (UAA) devoted to "grandes cultures" (soft wheat, durum wheat, barley, corn, rice, oats, rye, other cereals, rapeseed, sunflower, soya, potatoes, sugar beet, peas, other proteins), forages (maize, beets) and grasslands, and major animal productions (bovine, goat and sheep herds, poultry and pigs). The bovine activities are particularly disaggregated (27 activities related to milk or meat products, ages, gender).

The basic farming system is a "production set" defined by relations between crops and livestock activities. The basic set of relations is split into three categories: (i) agronomic and crop rotation constraints; (ii) nutriment requirements for animals; (iii) balances between bovine numbers taking account of the gender, the final product (milk or meat) and the age.

An additional specific module is devoted to the insertion of a large range of policy tools, including all pricing tools, all policy tools calling for thresholds on input (i.e. land) or output (i.e. quota).

Technical modules are added according to specific environmental estimations. In this way, the estimate of greenhouse gas emissions is the first operational module included in the model (De Cara and Jayet, 2000, De Cara et al, 2005).

This production set leads us to define a vector of parameters. The programming shell uses Unix scripts and Fortran codes and the core of the model description is written in the MGG language[2]. The MGG language is now only a way of writing of the kernel of the model (i.e. the list of parameters, variables and constraints subject to conditional activation, the objective function, the constraints, the "elements" which cross variables and constraints, and functions of parameters as routines feeding the "element" section). The MGG code is transformed into GAMS code through FORTRAN routines.

The steps of the building of the operational model are (i) the typology process leading to transform the FADN sample into anonymous and virtual farm groups; (ii) the first step estimate of parameters related to any farm group; (iii) the calibration of LPs related to farm groups; (iv) proceeding which manages the runs of the model (including the calibration step). User-friendly "menus" help designers and users to proceed in any step in a Unix environment.

The table 1 delivers the number of farm groups, the related UAA and represented farm numbers per Member State for the three versions of the model developed in response to incentives from the European Commission[3]. The benchmark version is the V2 version related to the 2002 FADN covering the EU-15. The last database provided by the European Commission a few weeks before the end of the GENEDEC programme will lead to the first version covering the EU-25, named V3 in the table.

Some details about the AROPAj farm design are delivered in De Cara and Jayet (2000), De Cara et al (2005), and deliverables related to the GENEDEC programme (D2, D3 available on the public area of the GENEDEC website). A detailed explanation of the clustering of farms into farm groups is provided by Chakir et al (2005), added by elements describing the estimation and calibration steps. It should be noticed that the "limiting step" is the calibration step which is now run-time costly. Let us recall that this step has to ensure us of the existence of a feasible solution of any farm group LP, and that it leads to a solution of the programme ($\mathcal{C}$) :

$$(\mathcal{C}) \qquad \min_{\boldsymbol{q}_k \in \Theta} \sum_{j \in J} (x_{jk}^*(\boldsymbol{q}_k) - \hat{x}_{jk})^2$$

in which $\boldsymbol{q}_k$ is a parameter sub-set related to the farm group k, Θ is a set of "realistic" values, x_{jk}^* and $\hat{x}_{jk}$ are respectively the solution of the k LP (depending on the $\boldsymbol{q}_k$ parameters) and the initial estimates of these variables (from FADN), and $\mathcal{J}$ is a sub-set of the J LP command variables. The

---

[2] see Simons, 1986, cited by H.J. Greenberg:
http://catt.bus.okstate.edu/itorms/volumes/vol1/papers/greenberg/bibs.html

solving of ($\mathcal{C}$) is provided by a combination of random and gradient computational methods. This generally leads to a sub-optimal solution depending of the chosen number of run iterations[4].

Table 1. Number of AROPAj farm groups, number of farms and used agricultural area represented.

MS	FADN regions	V1 (2002): FADN-1997			V2 (2005): FADN-2002			V3 (2008): FADN-2004		
		Farm groups	Farms (1000)	UAA (M ha)	Farm groups	Farms (1000)	UAA (M ha)	Farm groups	Farms (1000)	UAA (M ha)
Belgium	2	15	36.49	1.355	27	33.77	1.432	24	29.95	1.339
Denmark	1	10	48.23	2.349	22	43.59	2.437	21	37.34	2.397
Germany	13	99	255.51	14.736	144	204.94	14.728	138	200.03	15.414
Greece	4	45	273.39	1.418	54	198.10	1.220	59	208.54	1.651
Spain	16	99	302.06	11.816	155	329.96	11.330	150	274.97	11.333
France	22	131	320.19	22.972	157	313.91	24.042	144	299.40	24.321
U.K.	6	51	123.54	11.360	59	118.23	11.647	64	93.57	10.337
Irland	1	8	128.74	4.014	22	76.91	3.198	23	68.99	3.087
Italy	21	161	596.58	8.295	278	373.95	7.553	207	353.71	8.125
Luxemburg	1	5	1.52	0.104	13	1.64	0.130	11	1.51	0.124
Netherlands	1	11	65.35	1.867	19	51.30	1.511	18	41.60	1.234
Austria	1	23	79.28	1.882	38	74.09	1.788	39	72.93	1.853
Portugal	5	43	208.83	2.253	37	78.98	1.990	32	72.18	1.716
Finland	4	18	49.29	1.765	20	42.89	1.889	21	42.18	1.978
Sweden	3	15	38.02	3.228	29	39.00	2.639	27	30.57	2.560
EU-15		*734*	*2527.03*	*89.414*	*1074*	*1981.24*	*87.534*	*978*	*1827.45*	*87.467*
Cyprus	1							9	8.46	0.137
Czech Republic	1							23	12.21	3.430
Estonia	1							11	6.53	0.669
Hungary	7							45	66.78	3.477
Latvia	1							19	19.87	1.000
Lithuania	1							22	29.52	1.470
Poland	4							176	438.29	7.706
Slovakia	1							13	3.42	1.761
Slovenia	1							11	33.44	0.430
EU-24								*1307*	*2445.97*	*107.547*

The AROPAj model is FADN dependent. This means that FADN data lacks impact on the estimation step and could require patches even concerning the kernel of the model. This is for instance the case of sugar-beet quotas not referred for all MS. Globally the kernel extension depends on new objectives assigned to the model, and the covering extension (activities, MS) follows what FADN supplies. The real (and logical) limiting aspects come from legal restriction. Rights of access and use are under control of the European Commission. Access comes with European programmes. The major practical use constraint is the 15 sample farms limit in clustering into farm groups.

Compared to the V2 benchmark version, V3 is characterized by added industrial crops implemented in the model (tobacco, cotton, and flax). The FADN year is 2004. Due to information lack and due to a hopefully better representation of the sugar-beet pricing, the sugar-beet system is re-founded in the model. The first major change comes with the covering of the "new Member States" (NMS) leading to cover the EU-24 (the EU-25 except Malta for which FADN data provided by the EC are provisional) and added policy tools (mainly the top-ups). The second important change is the introduction of

---

[4] The dimension of the $\boldsymbol{q}_k$ vector is around 130 (the total parameter number related to a farm group is greater than 1200), the $\mathcal{J}$ dimension is around 50 (the number of LP primal variables is greater than 1000 when the 2000 CAP applies), and the number of runs is about 1200 for any farm group k.

energy crops and of the two-side related CAP conditions (i.e. the premium devoted to crops aiming at energy production on non set-aside areas, and the authorization of cropping on set-aside areas).

2.2. The Luxembourg agreement viewed through the AROPAj model

From a practical point of view, the introduction of the Luxembourg reform in AROPAj is a two-step process. The first step is devoted to the identification of the parameters of the policy support relative to the Agreement and consistent with the design of the model. That requires some "stylization" of the various national specifications the CAP implementation allows for. The second step is to let the model run with the reference values of the parameters so that the past-dependent parameters of the CAP (land set-aside, subsidies) are estimated. In our approach, the "past" refers to the reference year upon which the model is calibrated thanks to the FADN. The Luxembourg agreement has lead to defining modified ways for delivering subsidies offered to farmers. Therefore, the core of the model needed to be slightly modified. Likewise, Member States keep the right to (partially) re-couple some supports.

The computation of individual or regional decoupled payments is based on the results obtained using the Agenda 2000 policy as input. In the model, the single year of simulation is representative of the 3-year reference period. The prior AROPAj supports have been broken up according to different items related to possible decoupling combinations. The decoupling reform as modelled with AROPAj is subject to the structure of the model, based on FADN regions and farm-groups. Thus, it is possible to compute single farm payment for each farm-group or a unique regional entitlement. Some supports from the Agenda~2000 policy are currently not in the model (i.e. FADN does not distinguish between starch and food potatoes, consequently neither does the model). Nevertheless, the total European budget devoted to agricultural policy is well represented[5]. The regionalisation option is implemented without any difficulty when the regions covered by this option correspond to FADN regions (the case for Germany). Otherwise, some hypotheses are necessary about which part of a FADN region a farm-group belongs to (England).
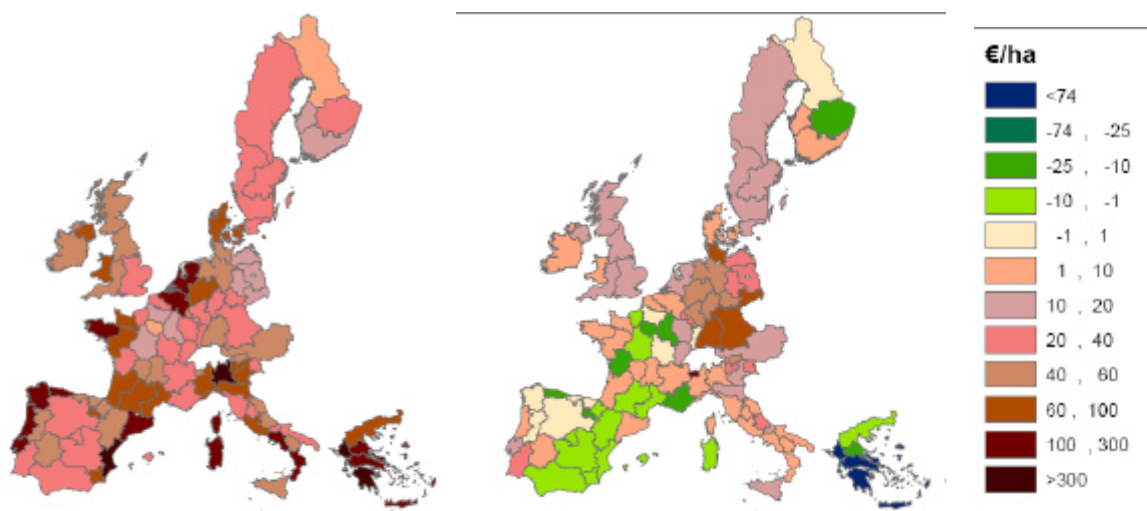

Figure 1: Variation in gross margins by hectare with successively the quasi-fix livestock adjustment (15% max.) applied to Agenda2000 (left) and the implementation of the Luxembourg reform (right).

---

[5] The EAGGF Guarantee Section budget amounted to 43.2 billion €in 2002, and 30.8 when we only consider the activities modelled in AROPAj. In AROPAj, the EAGGF Guarantee Section is estimated to 27.2 billion €

The improvement has allowed us to estimate the change in land allocation of the major agricultural productions (crops, forage), the on-farm consumptions and the marketed productions, as well as the shadow prices of quasi-fix factors (land, livestock, quota of milk and sugar). Moreover, we are able to assess indirect consequences like environmental impacts. Falling into line with De Cara et al (2005), change in emissions of greenhouse gas by farming systems can be estimated. We show that the Luxembourg reform presents a double dividend: emissions decrease with the implementation of the reform at the same time as gross margins increase. By focussing on gross margins, a major impact of the reform is revealed. In accordance with theoretical approaches, AROPAj shows that a decoupling policy induces an increase in gross margin, as shown at the regional level in Figure 1 (decoupled payments are computed from the reference without no livestock adjustment compared to the FADN estimates, and the sum up of the double-effect of livestock adjustment and CAP reform keeps any farm group gross margin greater than it is in the reference year).

Each farm-group has a given amount of area seen like a quasi-fix factor. The "individual" UAA resources are parameters of the model. The dual value related to land availability provides the shadow price of land. An important feature of MP models is that the optimal value of the objective is shared between the different quasi-fix factors. We focus here on land. With the implementation of the Luxembourg reform, the total shadow price of land increases for all Member States (table 2), as expected, as a consequence of a fundamental change in the subsidy. Even if some re-coupled subsidies still do exist for animal production, a large part of the subsidies which were devoted to animal production are devoted to land. Partial re-coupled subsidies devoted to crops make slighter the increase in gross margins and shadow prices particularly in numerous French and Spanish regions.

Table 2: Land shadow price variation with the implementation of the Luxembourg reform

MS		MS		MS		MS		MS	
Belgium	+8%	Greece	+29%	U.K.	+31%	Luxembourg	+27%	Portugal	+11%
Denmark	+11%	Spain	+6%	Irland	+18%	Netherlands	+3%	Finland	+21%
Germany	+21%	France	+8%	Italy	+11%	Austria	+6%	Sweden	+15%
				EU15	+14%				

3. From point yields to N-response functions

3.1. Methodological foundation

Regarding crops, the basic version of the AROPAj model makes yields and variable costs into parameters estimated at the regional scale for any set of crop and farm group. We focus on nitrogen inputs and on the relation between nitrogen fertilizing and the yield because of its key role in the scientific agronomic background and because of its strong environmental implications. The consequence on the model architecture is to replace mathematical "points" (i.e. any "N-input and yield" set) by functions, and the corollary is the transformation of the initial LP model into a NLP model (non-linear). Another consequence is to make central in the model the on-farm use of farming products. Beside the on-farm consumption of cereals for feed, already included in the AROPAj model, we add the potential use of nitrogen brought by manure.

Let us recall how the NLP enters the LP framework. When the optimal solution of the non-linear programme leads to strictly positive marketed crop productions and strictly positive marketed N-fertilizer inputs and when the LP model does not include any additional constraints holding inputs or

outputs (like quotas), the solution of the NLP is exactly the solution of the following transformed LP. Let us respectively denote the crop and N-input prices by $p_{j,k}$ and w_k related to the j crop and the k AROPAj farm group. The transformed LP is the initial LP in which new values of N-inputs ($N_{j,k}$) and yields ($y_{j,k}$) come from the solution of the sub-programme $\max\limits_{N \geq 0} p_{j,k} \cdot y_{j,k}(N) - w_k \cdot N$ where $y_{j,k}(N)$ denote the yield functions. Any strictly optimal solution of this sub-programme is such as the derivative of the yield function is equal to the relative price $w_k / p_{j,k}$. When the optimal solution of the NLP leads to zero marketed output (i.e. the whole production of crop j by the farm group k is devoted to on-farm consumption) or zero marketed N-input (the whole N-input related to crop j and the farm group k is brought by manure), the LP should be fed by N-input and yield values based on implicit prices which are dual values of the LP itself. Now in this case we accept one-step sub-optimal solutions provided by real prices instead of dual prices.
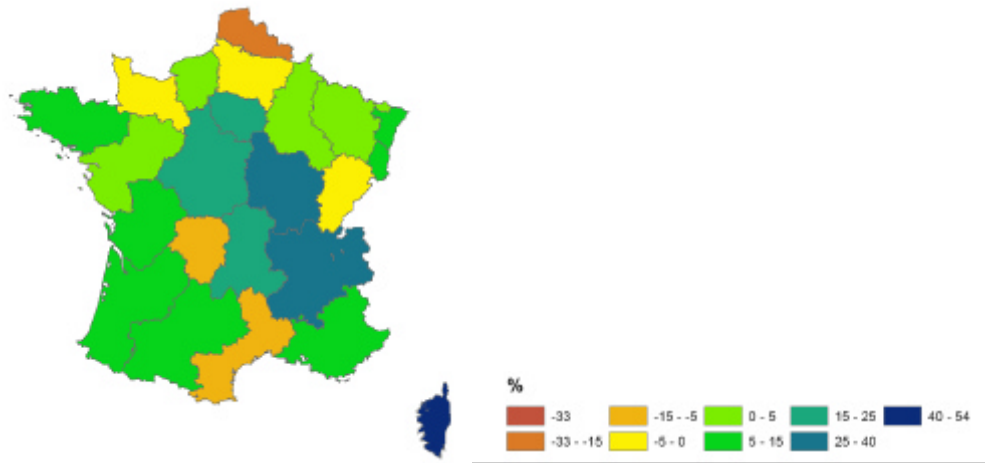

Figure 2. Variation in gross margin related to a climate change scenario.

Data as well as the choice of the formal shape of the yield functions are key preliminary points. Godard et al. (2007) propose a comprehensive method built in order to be applied for any AROPAj set of crop and farm group at the European scale. Let us just recall the guidelines of the method when data are available (i.e. requested information on soils, climate and crop management provided by European databases), when the functional form of yield functions is adopted (i.e. exponential functions $r_{jk}(N) = B_{jk} - (B_{jk} - A_{jk})\exp(-t_{jk}N)$) and when the crop model is selected (i.e. the STICS model). The objective is the selection a set of physical parameters related to a set { j, k }, for any set taken into account by the AROPAj model. The AROPAj model provides estimates of yields and crop prices, marketed N-costs, and manure. The first step is the use of the STICS model to build series of points { $N_{j,k}, y_{j,k}$ } related to different selections of "soils" and "crop management" parameters, the { j, k } and the "climate" given, when $N_{j,k}$ is incremented from 0 toward high values acceptable from an agronomic point of view. These series of points lead to adjusted exponential functions $y_{j,k}(N)$. The second step is the selection of the "best curve" related to { j, k } which (i) makes the initial yield

(provided by the initial AROPAj estimate) reachable, and (ii) makes the slope of the curve when this yield is reached the closest to the price ratio $w_k / p_{j,k}$.

This method was applied to the French farm groups related to the V2 version of the AROPAj model (Galko-Debove, 2007), leading to 584 yield functions. When there is no selected yield function (i.e. when there is no function crossing the initial yield, or when the method is not still applicable like in the case of fodders), the model keeps the initial "points". Galko-Debove used climate change scenarios and the STICS model to modify the previous yield functions. This allows us to assess *ceteris paribus* the impact of climate change on farming systems (see Figure 2 for the impact of the climate change on average regional gross margins).

3.2. Methodological improvement

One of the problems encountered by the method presented above (see section 3.1) is the difference between the initially estimated N-input and the N-input derived from the yield function when the yield and the prices are equal to the initial values. In other words, generally the best function selected by the method does not lead to the ideal function meeting a given point with a given slope.

This difference can be obviously explained by the imperfect selected models, and by the lack of consistency between physical and economical variables used for the selection of yield functions. Another explanation comes from the market price ratio supposed to be equal to the derivative of this function when the gross margin per area unit is optimal. Let us consider the stylized programme ($\mathscr{P}$) of a crop farming system which includes possible quotas:

$$\max_{N_j, s_j, y_j} \sum_j (p_j\, y_j - w\, N_j\, s_j)$$

($\mathscr{P}$)
$$\sum_j s_j \leq S \qquad [\boldsymbol{1}]$$
$$y_j \leq r_j(N_j)\, s_j \qquad [\boldsymbol{r}_j]$$
$$y_j \leq Q_j \qquad [\boldsymbol{a}_j]$$

In this simple programme, N_j, s_j, y_j respectively denote the N-input, the marketed output and the area devoted to the crop j. The total used area is denoted by S and Q_j denotes the quota (which is "infinite" when there is no quota). The yield functions are denoted by $r_j(N)$ (and supposed to be monotonous increasing and concave). The Lagrange multipliers are designed by the Greek letters. Let us consider the optimal solution and a crop $_j$ for which N_j and y_j are strictly positive (and consequently s_j is strictly positive). The first order necessary conditions characterizing the solution leads to the relation $\dfrac{dr_j}{dN} = \dfrac{w}{p_j - \boldsymbol{a}_j}$. When the quota constraint holds, the dual price $\boldsymbol{a}_j$ is strictly positive and the price ratio is greater than the market price ratio $\dfrac{w}{p_j}$.

Regarding the LP AROPAj model, there is a lot of reasons leading to take account of price ratios different (greater or lower) from the market price ratios in the selection step of the yield function building described in section 3.1. Among them we have :

- explicit constraints involving inputs or outputs (i.e. quota)
- existence of strictly positive output accompanying zero marketed output (i.e. when the on-farm consumption for feed makes the cereal worth greater than the market price)
- the real behaviour of farmers accepting sub-optimal choices (i.e. the unique level of N-input per area unit related to a given crop even when the production is supplied at different net prices, like we can observe it in the case of sugar-beet pricing)

Turning back to the selection step of yield function, we consider that it is difficult to compute the virtual prices and to use them in the selection itself. But we find here an argument useful for making lower the difference between the initially estimated N-input and the N-input derived from the yield function mentioned at the beginning of this sub-section. The arbitrary but simple mean to eliminate this difference is to introduce a multiplying parameter distorting the selected yield function.

4. Towards estimates of probability of farm group location

4.1. Objective and method

Agriculture is widely involved in physical and economical processes locating it at the core of cross-interactions between the economic field and the environment. One of the best examples of this situation is supplied by relationship between agriculture and climate, when the one necessarily impacts on the other. When agriculture impacts on climate through greenhouse gas emissions (direct agricultural emissions come from non-CO_2 GHG which are nitrous oxide and methane), this kind of externality is global and the estimate of impacts do not require to be accurately located. But major external effects caused by agriculture would be as better analysed as they are precisely located. And any case environmental regulation concerns farming systems widely diversified in term of individual contribution and marginal cost of abatement. At last policy makers would find interesting information through mapping of linkage between farming systems and the environment.

Mapping of agricultural activities enters a field of research steered towards spatial disaggregation and more often dedicated to compensate the lack of data (Howitt and Reynaud, 2003). Here we point out that a lot of information is available and is geographically informed when it is of physical nature. Difficulties come when economic and private information is at stake. FADN samples are protected in term of private information, and the AROPAj farm clustering means that we just know the regional membership of the AROPAj farm groups. The mapped assessment of farming impacts related to external change (i.e. the CAP or the global environment) will be improved by the estimate of likelihood of farm group mapping.

The three-step procedure is managed at the regional scale and summarised as following. The two first steps are devoted to the computation of land use likelihood (Chakir, 2006). The basic theoretical model relates the land characteristics to the land use, the CORINE Land cover (CLC, see EC-EEA, 1993) map providing at a very fine resolution (100m x 100m) an underlying structure of the FADN region for the desaggregation (but CLC is not detailed enough to distinguish the different kinds of agriculture). Land characteristics are completed by other European databases (soil data base, climate, digital elevation model). Observations and locating at the pixel scale are provided by European databases. The econometric specification refers to the multinomial logit model and leads to the

estimated probability of locating the land activity j (among J) on the pixel i $\boldsymbol{p}_{ij} = \dfrac{\exp(\boldsymbol{b}_j\, x_{ij})}{\sum\limits_{l=1,J}\exp(\boldsymbol{b}_l\, x_{il})}$ where

x_{ij} is the vector of land characteristics used as explanatory variables (this includes the CLC map, altitude criteria, soil properties, climate). Land use observation $\boldsymbol{p}_{ij}$ comes from the LUCAS database[6]. After estimating of the vector of parameters $\boldsymbol{b}_j$ leading to the estimated prior probability $\hat{\boldsymbol{p}}_{ij}$, we use the "cross entropy" approach and the FADN to estimate the posterior probability p_{ij} of locating the land activity j on the pixel i. This is summarised by the programme ($\mathcal{E}$) where R_j is the land dedicated to the crop j in the FADN and u_i is the i pixel area :

$$
\begin{aligned}
&\min_p \sum_{i,\,j} p_{ij} \ln \frac{p_{ij}}{\hat{\boldsymbol{p}}_{ij}} \\[4pt]
(\mathcal{E}) \qquad &\forall j: \sum_i p_{ij}\, u_i = R_j \\[4pt]
&\forall j: \sum_j p_{ij} = 1 \qquad \forall i, j: p_{ij} \geq 0
\end{aligned}
$$

At last, the probability of locating of the AROPAj farm group k on the pixel i is designed by the formula $q_{ik} = \sum\limits_{j=1}^{J} \dfrac{p_{ij} S_{jk}}{\sum\limits_{n=1,K} S_{jn}}$ where S_{jk} denotes the initially estimated area devoted by the AROPAj farm group k to the crop j and K denotes the number of AROPAj farm groups related to the considered region. It is easy to check that q_{ik} is a probability distribution ($\sum\limits_{k=1}^{K} q_{ik} = 1$).

4.2. Mapped impact of the Luxembourg agreement

We applied the methodology described above to the mapping of change in probability of crop location when CAP reforms are implemented in the AROPAj model.
Figures 3 and 4 show for Benelux and France the change related to the Luxembourg agreement for cereals and grasslands. A possible interpretation of the observed variations holds in the cross-effect of the relative decreasing support of animal production compared to crops and of the relative increasing support of grasslands compared to crops. Moreover we have to take care of the difference which should be made between areas and productions. Detailed AROPAj results show that decreasing cereal area is concomitant with decreasing on-farm consumption and stabilized marketed output. This conveys the complexity of substitution effects reflected by the model, between the different uses of the land (crops and grasslands), between the differentiated supported outputs (crops and animals), and between the different feeding systems (on-farm cereals, fodders, and industrial concentrated feed).
Let us too recall that the initial implementation of the Luxembourg agreement implies to maintain set-aside more or less at a level equal to the historical reference level. As long as set-aside commitments remain steady, higher productive and larger farms will not change the part of land out of production.

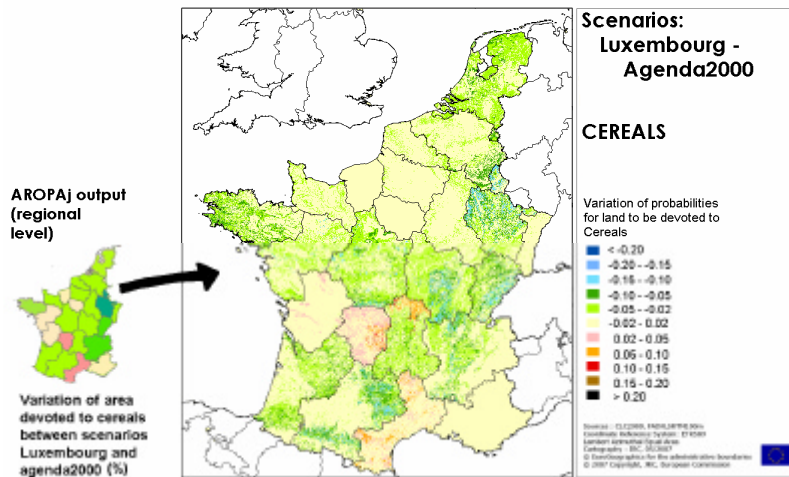[6] Land Use/Cover Area frame statistical Survey, conducted by Eurostat.

Figure 3. Change of mapped probability of cereals location related to the Luxembourg agreement.
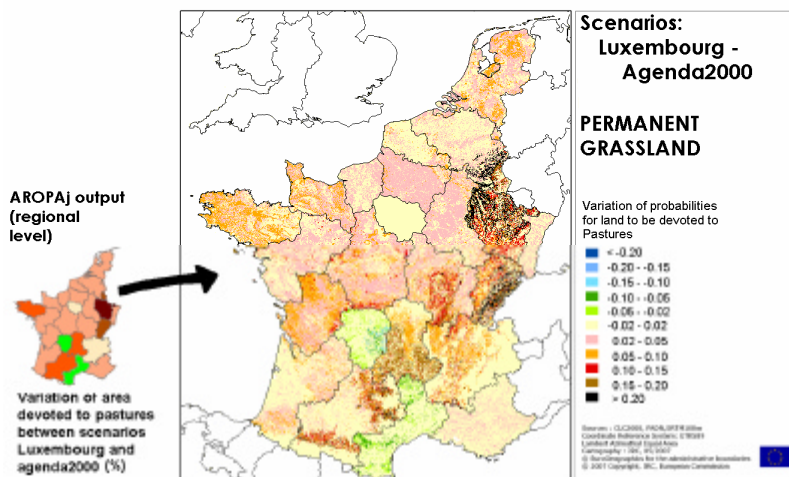

Figure 4. Change of mapped probability of grasslands location related to the Luxembourg agreement.
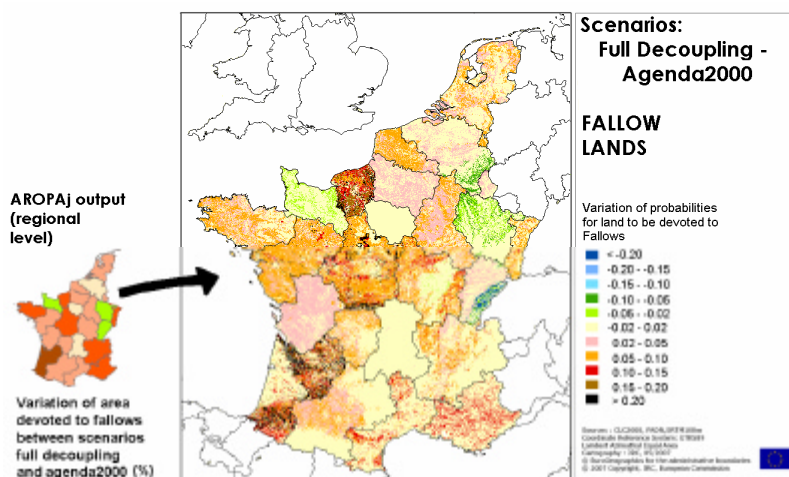

Figure 5. Change of mapped probability of fallow location related to a full decoupling scenario.

The situation quite differs in other farms with possible increasing land devoted to fallow. Figure 5 focuses on the fallow when a "full" decoupling scheme is implemented (no set-aside requirement, no

re-coupled direct subsidy, and a unique farm premium per hectare is implemented). Fallow land could slightly increase in the majority of regions, it could strongly increase in a few of them, but some others could supply decreasing fallow area.

Turning back to the maps themselves, regional limits appear as artefacts that we have to reduce through revisiting of statistical methods of mapping. These limits not surprisingly appear and they are caused by the "Regional" effect playing along all the process of the AROPAj building based on the FADN in which the region is the geographical finest relevant level. Nevertheless even before the use of smoothing methods of mapping, these artefacts are not really significant considering the change of colour related to slight change of probability level. The other interesting point is the revealed existence of sub-regional differences even in homogenous elevation regions.

5. Perspective for estimates of agro-environmental impacts and for policy design

The AROPAj model will be expanded into new modules focusing on phenomena strongly implied in environmental problems. A module dedicated to the nitrogen balance is already partly activated when the yield functions are used (see section 3). Additional part dedicated to animal manure spreading and to organic origin nitrogen is inserted in the model kernel. It will be used when all required data are got and checked. It will allow us to include explicit policy constraints on animal manure spreading. A quite consistent module will be available when the yield functions take account of organic nitrogen inflows. Following the seminal work of Godard (2005) we propose to insert the different sources of nitrogen so that the yield function is $r_{jk}(N_1, N_2,...) = B_{jk} - (B_{jk} - A_{jk}) \exp\left(-\sum_i t_{jki} N_i\right)$ where the parameter t_{jki} refers to the nutritional contribution of the N-input of type i to the crop j in the farm group k.

Figure 6. Overlapping of the Seine river basin and the French regions.

The "response function" concept will be expanded to environmental outputs related to nitrogen. These outputs are nitrates (NO_3 from lixiviation), ammonia (NH_3 from volatilisation) and nitrous oxide (N_2O from denitrification). First work was devoted to the N_2O greenhouse gas in GENEDEC (see the deliverable D4). It is shown how important would be the differences in emission ratios depending on the crop, on the soil, on the climate, and how fragile is the coefficient retained by the International Panel for Climate Change (the last 2007 report retains 1% as the mass ratio of emitted nitrogen on brought nitrogen). A research programme includes the assessment of these response functions at the scale of the Seine river basin (see the map on figure 6). This programme starting on 2007 is mainly devoted to the coupling between AROPAj and the MODCOU hydrological model and focuses on nitrate transfers besides other N-pollutants emissions[7].

The problem of N_2O emission also occurs when we programme an analysis including agronomic and economic modelling aspects to assess the impacts of soil compaction[8]. Soil compaction impacts on private economy of farming systems through change in crop yields, and impacts on environment through change in N_2O emission. The improvement of the agronomic STICS model will benefit to economic assessment based on the use of the AROPAj model.

Besides feed and food, final use of agricultural production can be energy. In the EU besides other countries, biofuels appear as good opportunity leading the EC to propose to cancel set-aside commitments in 2008. When process, prices or products differ, activities related to a generic product (i.e. wheat) can be split and easily introduced in the MP model. One of the stakes now is the way in which the European Directive on the promotion of the use of biofuels will interfere with the European agricultural production, taking account the market sharing between gasoline and diesel. AROPAj will be used and improved in a two-dimension programme9. First this model will be coupled with an industrial MP model of EU refinery dedicated to the energetic raw matter demand, and it will be refined on the taking account of co-products. Secondly we intend to introduce new ligno-cellulosic productions and more globally the so-called "biomass to liquid" outputs aiming at increasing yields in term of energetic contend rather than in proteins. This will allow us to assess environmental impacts related to greenhouse gas emissions as well as water demand and nitrogen balance.

References (Times New Roman 12 bold)

Chakir, R. (2006), Spatial Disaggregation of Agricultural Production Data: An Econometric Approach using Minimum Cross Entropy, 61st Econometric Society European Meeting, Wien, 2006/08, 28 p.

Chakir, R., Debove, E. and Jayet, P.A. (2005), Decoupling of direct payments: An application of the AROPAj model projecting regionally differentiated impacts on the EU-15 farming sector, Proceedings of the expert workshop, Séville, 6-7 October 2005, 15 p.

De Cara, S. and Jayet P.-A. (2000). Emissions of greenhouse gases from agriculture : the heterogeneity of abatement costs in France, *European Review of Agricultural Economics*, vol. 27, n° 3, 2000/09, pp. 281-303

---

[7] This refers to the French PIREN programme financed by Agence de l'Eau Seine-Normandie besides other institutions, and managed by CNRS and Paris VI University.

[8] This programme named DST and financed by the French National Research Agency is lead by the Unit of Agronomy at INRA Orléans.

[9] Such a programme will be realized in collaboration with the IFP (French institute of research on oil).

De Cara, S., Houzé, M. and Jayet, P.A. (2005). Green house gas emissions from agriculture in the EU: a spatial assessment of sources and abatement costs. *Environmental and Resource Economics*, vol. 32, n° 4, 2005. pp 551–583.

EC-EEA, 1993, CORINE Land Cover: technical guide, EUR 12585EN, Office for the publications of the European Communities, Luxembourg, 144p, http://reports.eea.europa.eu/

Galko, E. (2007), Modélisation de l'offre agricole européenne face à de nouveaux enjeux : réformes politiques, effet de serre et changement climatique, PhD-thesis.

Godard, C. (2005), Modélisation de la réponse à l'azote du rendement des grandes cultures et intégration dans un modèle économique d'offre agricole à l'échelle européenne - Application à l'évaluation des impacts du changement climatique, PdD thesis.

Godard, C., Roger-Estrade, J., Jayet, P .A., Brisson, N. and Le Bas, C. (2007), How to use the available information at a European level to construct crop nitrogen response curves for the regions of the EU, accepted in *Agricultural systems*.

Heckelei, T., Witzke, H. P. and Henrichsmeyer, W (2001). Agricultural Sector Modelling and Policy Information Systems. Proceedings of the 65th European Seminar of the European Association of Agricultural Economists, 29–31 March 2000, Bonn (Germany). Vauk, Kiel, Germany, ISBN: 3-8175-0329-6, 331 pp

Howitt, R.E. (1995). Positive Mathematical Programming. *American Journal of Agricultural Economics* 77(2): 329-342.

Howitt, R.E and Reynaud, A. (2003). Spatial Disaggregation of Agricultural Production Data Using Maximum Entropy, *European Review of Agricultural Economics*, 30(3), 1-29.

Simons, R.V. (1986). Mathematical Programming Modeling Using MGG, *IMA Journal of Mathematics in Management* 1, 267-276