Dynamic Land Use and the Field-level Disaggregation of Crop Production.

Richard Howitt[⊥] and Arnaud Reynaud[τ]

9/23/2002~~11/12/01~~~~11/12/2001~~

Abstract

The disaggregation problem arises each time the question of interest requires knowledge of microeconomic behavior that must be based on aggregate sample data or model results. A related question is, what is the optimal level of disaggregation when facing different types of data at different scale levels? Using the most disaggregate level conserves as much information as possible, but may not be justified given the additional model complexity. In this paper we develop a data-consistent approach to the estimation of cropping choices by farmers at a disaggregate level (field-level) using more aggregate (regional-level) data or the results of an aggregate production policy model. Our data disaggregation procedure requires two steps. The first step consists of specifying a dynamic model of crop allocation and estimating it using aggregate data. In the second step, we disaggregate the outcomes of aggregate model using maximum entropy (ME). Two points should be noticed. First, we explicitly model aggregate cropping pattern choices as a dynamic process. Second, we disaggregate farmer's behavior by ME at the most disaggregate level possible, namely the field-level. Our data disaggregation procedure is applied to a sample of 190 fields located in California and observed from 1986 to 1990. The sample includes six annual crops, namely: Alfalfa, Cotton, Field, Grain, Melons and Tomatoes. Aggregate crop choices are modeled by a second-order Markov process. Field-level crop predictions are therefore limited to years 1988, 89 and 90. We show that the quality of predictions at field-level is relatively good. In 1988, 64% of fields are predicted to produce the crop that is observed. The percent of correct predictions is 72% in 1989 and 60% in 1990. These results show that the micro behavior, inferred from aggregate data with our data disaggregation approach, seems to be consistent with observed behavior~~.~~

Keywords: Data disaggregation, Bayesian method, Maximum entropy, Markov process, Land use.

Code JEL: C11, C44, Q12.

⊥ Department of Agricultural and Resource Economics, University of California at Davis.

τ Department of Agricultural and Resource Economics, University of California at Davis and LEERNA-INRA, University of Social Sciences Toulouse.

I. Introduction

Economic theory generally focuses on behavior of individuals and their resulting decision rules. However, in many cases, researchers only have access to aggregate data to study micro behavior. Aggregation may be over individuals, time, space, and products (among others) or on any combination of these dimensions. Given that only aggregate data are available, two approaches are usually possible. The first one is to postulate some behavioral rules for individual choices, and use them as a basis for specifying aggregate rules. This approach has the drawback of postulating the individual behavior that has yet to be estimated. The second approach is to try to infer from aggregate data the individual behavior that is most consistent with the observed outcomes. We follow the latter approach in this paper by developing a method for disaggregating aggregate agricultural production data. More specifically, the problem we address is the following: observing annual aggregate data on land use amongst crops and an initial micro-level crop allocation, is it possible to estimate yearly micro field-level behavior of farmers in a statistically consistent way?

We argue that a valid data disaggregation method is of interest in agricultural production economics for three main reasons. The first reason deals with availability of data. The difficulty of obtaining adequate micro data is mentioned by many authors (Just 1993, Just and Pope (1999-a), as a major limit on the ability to empirically estimate micro level agricultural production response. Aggregate data for crop and livestock are now relatively abundant, but many authors have emphasized the difficulties of using it to learn about micro-level behavior. See for example Mundlak (1999), or Nerlove and Bessler (1999) among others.

The main criticism of using aggregate data is that the failure to consider heterogeneity across agricultural producers may result in misrepresenting technology, in addition, aggregate data may fail to support the regularity conditions needed to recover technology from estimated structures, Just and Pope (1999-a). Therefore, Just and Pope (1999-b) recognize the need to create 'a broad and complete public panel of agricultural production data' at the farm level. Despite repeated calls for such a database, it does not yet exist[1]. In its absence, we must resort to using the aggregate data actually available. A valid data

[1] It is true that some very detailed databases at field or farm-level have already been used for estimating agricultural production functions. For example, Antle and Capalbo (2001) analyze at field-level the dryland grain production system of the Northern Plains, Montana, using panel data on 425 commercial grain farms. They introduce site-specific biophysical characteristics (soil,

disaggregation method could partially bypass the lack of disaggregate data in agricultural economics by disaggregating the information from aggregate models to the micro level, and thus enabling tests of the micro behavior inferred by the aggregate observations.

The second argument for disaggregation is that consistency among the explanatory variables requires that most microeconomic models are estimated at the level of the least disaggregated variable. However there is probably additional micro-level information in the remaining data set that could be recovered by disaggregation of the limiting variable. In principle, working with disaggregate data allows one to use all the information and to develop more powerful tests, Barker and Pesaran (1990). As a consequence, model estimates based on disaggregated data that preserve the information in the aggregate data, and use additional structural information based on measures of heterogeneity may, in general, result in better predictions.

A third reason to study disaggregation is that agricultural production models are being increasingly used in conjunction with biophysical process models. For example, assessing environmental impacts of agriculture increasingly requires the use of linked disciplinary simulation models, Antle and Capalbo (2001). These latter models are often calibrated at a micro level. Disaggregating the results from more aggregate economic models will enable them to interact more effectively with physical process models.

Miller and Plantinga (1999) have proposed a maximum entropy approach (ME) for estimating land use shares using aggregate data. They use ME to disaggregate land use shares from multi-county scale to county scale. They show that the ME specification encompasses the traditional pooled logistic regression as a particular case. They apply the ME approach to the estimation of land use in three Iowa counties and its impact on soil erosion. Our paper differs from Miller and Plantinga (1999) in two main ways. First we explicitly model cropping pattern choice as a dynamic process. Second, we disaggregate farmer behavior to the most disaggregate level possible, namely the field-level. Four main reasons favor a field-level analysis of cropping choices. First, land use patterns are determined by relative returns that depend on prices and costs (among others), and by land characteristics. As these characteristics may differ from parcel to parcel, the field-level is the only level that fully takes this heterogeneity into account. Second, many agricultural policies have an influence on a farmer's crop choice at the field-level, which we postulate is the basic decision unit for a farmer. For example, conservation programs often target lower-

---

topography and climate) into the production function and analyze crop rotation choices. However, these very extensive types of dataset still remain the exception and we often have to deal with much more aggregated information.

quality lands susceptible to soil erosion. Assessing the impact of such policies requires a better understanding of land allocation between crops and requires a field-level analysis. Third, some environmental problems, such nonpoint source pollution or wetland depletion, often have to be solved on a local basis where the proximity to riparian areas is important. Finally, working at the field-level dynamic crop rotations can be modeled explicitly rather than focusing on a series of 'static' realizations of an aggregated steady state. For these reasons, we believe that a valid approach for disaggregating aggregate land use data to farmer's micro behavior may be of interest for agricultural economists.

Interest in data disaggregation procedures is not limited to agricultural economics. Researchers face data disaggregation problems in many other fields. A few of them are climate science, geography, political science, and marketing. In Climate Science, disaggregation (downscaling) techniques have been used in the past decades to derive finer-scale weather information from numerical weather prediction models and for classifying weather regimes, Von Storch et al. (1994). A growing interest in solving data disaggregation problems may also be found in Geography. Geography deals basically with spatial data. An important problem in working with spatial data is combining information at a variety of scales. The quantity of available geospatial data at multiple spatial and temporal scales is now growing exponentially with development of remote sensing coupled with the widespread use of geographic information systems (GIS).

In Political Science, Ogburn and Goltra (1919) were first to address the issue of finding a valid method for passing from aggregate data to disaggregate data. Another important political science example where data disaggregation is required, is Federal Law implementation, King (1997). The 1965 U.S. Voting Right Act prohibits voting discrimination on the basis of race, color or language. Legally, discrimination only exists if plaintiffs can demonstrate that members of a minority group vote both cohesively and differently from other voters. Since individual voting surveys are rarely available, implementation of the Voting Right Act requires a valid data disaggregation from electoral data and U.S. Census Data. In Marketing it is often not possible to obtain individual demographic characteristics of consumers. Researchers must often rely on average information characterizing people of given area (from U.S. Census Data for example). A valid data disaggregation approach allows researchers to link observed individual consuming to these aggregate characteristics.

In Economics, the data disaggregation problem is, in some sense, the inverse of the aggregation problem. Nearly every study must aggregate over individuals, products, techniques, or time and space, Barker and Pesaran (1990). But aggregation of micro relations leads in general to an aggregation bias.

Aggregation problems have been a question of interest for economists since the seminal paper of Theil (1954). In the case of linear aggregation, Theil (1954) has shown that macro relation performs, at best, as well as the estimated micro relations if micro relations are correctly specified. Nevertheless, a correct specification requires assumptions that are often not verified in practice, namely the absence of heterogeneity, stability of parameters over time, etc. This explains why the definition of a 'good' level of data aggregation is still a topic of discussion Stoker (1993). Working at the most disaggregate level is a way to bypass aggregation problems.

The remainder of the paper is organized as follows. Section 2 presents a dynamic aggregate model of land use and the ME data disaggregation approach. The data disaggregation procedure requires two steps. The first step consists in specifying a model of crop allocation and estimating it using aggregate data. In the second step, we disaggregate the outcome of aggregate model using maximum entropy (ME). Two points should be noticed. First we explicitly model aggregate cropping pattern (or rotational) choices as a dynamic process. We model aggregate land use shares using a stationary r-order Markov process. Second we disaggregate farmers behavior by ME at the most disaggregate level possible, namely the field-level. We show that ME framework provides an efficient method for generating field-level estimates that are consistent with observed aggregate data. In section three we apply our model to a sample of fields located in California. We consider a 190-field sample observed from 1986 to 1990. The sample includes six annual crops, namely: Alfalfa, Cotton, Field, Grain, Melons and Tomatoes. A second-order Markov process is specified as representing aggregate crop choices. Therefore field-level crop predictions are limited to years 1988, 89 and 90. In 1988, 64% of fields are predicted to produce the crop that is observed. This percentage is 72% in 1989 and 60% in 1990. These results show that the micro behavior, inferred from aggregate data with our data disaggregation approach, seems to be consistent with observed behavior.

II. The disaggregation model

In this section we first present the aggregate model of land use. Then, we turn to the data disaggregation problem. We show that the aggregate model used to infer farmer's crop choices at the field-level requires solving an integer constrained cross entropy problem.

2.1 Disaggregating aggregate data: background

Consider a region producing K different crops. We assume that each year we observe *at the regional-level* (or aggregate level) the land allocated to each crop $S_k(t)$, where $k = 1,\ldots,K$ and $t = 1,\ldots,T$ index crops and years respectively. Let $Y_k(t)$ be the probability of producing crop k at date t. By definition:

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

A region is composed of I fields indexed by $i = 1,\ldots,I$ of respective and time-invariant size s_i. We assume that the available information at the *field-level* (or disaggregate level) is limited to:

$$c_i(t) \in \{1,\ldots,K\} \quad \forall i, \forall t = 1,\ldots,r \tag{2}$$

that is the crop produced on field i during the first r periods[2], $r < T$. We assume that we only have information at the district-level for the *first r periods*[3]. This data availability assumption is based on the common situation where there is comprehensive data at aggregate-level, but only partial information at disaggregate-level. This partial information may come from detailed field-level surveys that are not conducted every year due to time and expense.

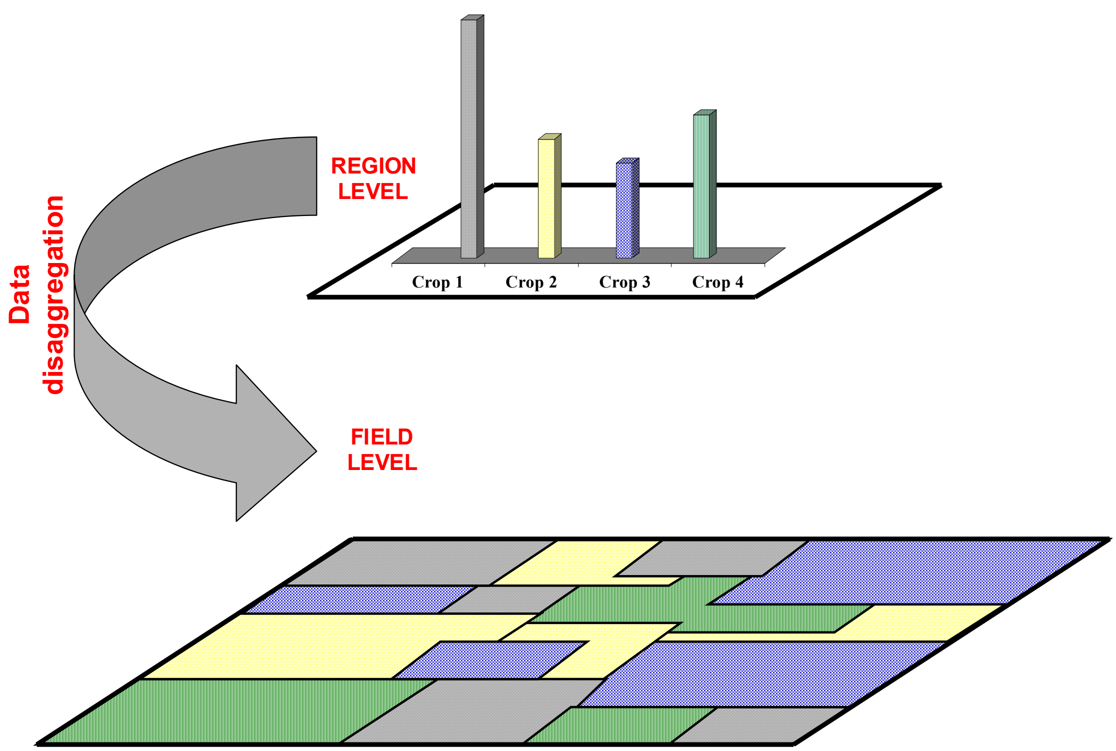
---

[2] We implicitly assume that farmers can only produce one crop per year on each field. We do not take into account possibility of intra-annual rotations.

[3] This informational requirement necessitates some comments. In what follows, we are going to specify the aggregate land use model as a r-order Markov process. In the disaggregation procedure, we need to initialize the Markov process at the field-level. As a consequence we need to observe individual crop choices at field-level for the r first periods. It is clear that the choice of the Markov process order crucially depends on data availability at field-level. If we were able to observe field-level crop choices during a long period we could specify a high-order markov process.

We want to combine the complete information at the regional-level for $t = 1,\ldots,T$ with the partial information at the field-level for $t = 1,\ldots,r < T$, with the aim of recovering crop choices at the field-level for periods $r+1,\ldots,T$: $c_i(t) \quad \forall i$ and $t = r+1,\ldots,T$. Moreover, we want these estimates to be data-consistent with aggregate information. In this paper we define data-consistency as meaning that the disaggregated outcomes at the field-level are compatible with data at the aggregate level. The data disaggregation procedure is conceptually described in Figure 1. for the case where there are four possible crops.

Figure 1: Disaggregating data from regional-level to field-level



Two main difficulties have to be overcome by a valid disaggregation method. First, the disaggregation problem is in most of cases ill-posed and there is a multiplicity of solutions. A valid selection criterion must be defined in order to select the 'most likely field-level allocation of land' compatible with the aggregate data. Second, we have to go from continuous land use distributions at the region-level to discrete choices at the field-level.

The dynamic disaggregation method we propose in this paper requires two steps. In the first step, we estimate, at the aggregate-level, a dynamic model of land use[4]. Then, in the second step, we disaggregate land use to the field-level for specific time periods, using a Generalized Maximum Entropy (GME) framework. This procedure is described in the following two sections.

2.2 A Dynamic model of land use at the aggregate level

2.2.1 Aggregate land use as a Markov process

We want to predict individual farmer's crop choices in a dynamic framework using aggregate data. As a consequence, we need to first consider a dynamic model for aggregate land use. Models of optimal land use allocation have been widely analyzed in agricultural economics. Three main lines of research have been developed. The first one uses representative farm programming models to estimate constrained resource allocation. This class of model includes linear and non-linear programming and positive mathematical programming[5]. This approach is very useful because it allows discrete choices of land use. However, since most mathematical programming models rely on representative farmer behavior, an important limit on its use is that it cannot take into account heterogeneity between agricultural producers. The second approach is based on econometric models that explain land use based on a reduced-form of economic variables (input and output prices considered as proxies for land rents from alternative uses),

---

[4] Farmer's choices are dynamic. Four main types of intertemporal relationships between crop can be mentionned to justify the use of a dynamic process. First, crop rotation may be viewed as a way to reduce the loss of soil productivity due to erosion, El-Nazer and Mac Carl (1986) or Goetz (1997). Second, it may stabilize over time profits of risk-averse agricultural producers. Third, crop rotations may be used for breaking weed and disease cycles. Finally by reducing dependence on external inputs, crop rotation system offer the possibility of atenuating agriculture environmental impacts while maintaining. Therefore the crop choice must be considered as a dynamic process: agricultural producer's decision about crop today has a significative impact on future profitability of cultures.

[5] To our knowledge, King (1953) and Heady (1954) are the first studies using mathematical programming in agricultural economics. These techniques are now widely used. For example, in a recent paper Prato et al. (1996) use mathematical programming in order to assess the impact of climate change on agricultural economics. The interested lector can refer to Howitt (1995-a) and Howitt (1995-b) for a formal presentation of positive mathematical programming.

characteristics of land and policy variables[6]. Usually, land shares are estimated using a logistic specification, despite the ad hoc nature of this specification. The third approach consists of direct estimates of production, cost or profit functions. However, if traditional econometric estimates are used, this approach requires prodigious amounts of data.

In this paper, we use another convenient formulation to specify the aggregate land use model. Land shares are dynamic, stochastic and simultaneously determined. Hence, they may be viewed as a dynamic stochastic process. Predicting the aggregate land allocated to a given crop can be done by estimating transition probabilities from any given crop to the one considered. These simplifying assumptions imply that an individual farmer's crop choices follow a finite Markov process. A Markov process is an appropriate probabilistic model for time series data when the state variable at any point depends only on the previous lagged state values. Lee et al. (1970) show that it is possible to estimate a Markov transition probability matrix, which defines the behavior of the micro units, using only aggregate data in the form of proportional sample data.

Assumption: Land use at the regional-level follows a finite stationary r-order Markov process.

We assume that aggregate land use can be modeled using a finite stationary *r*-order Markov process with $r < T$. Using a *r*-order Markov chain, we implicitly assume that crop choices at a given period *t*, $t \in \{1,\ldots,T\}$, only depend on the *r* previous periods. This assumption corresponds to the view that land use is a dynamic problem and that an agricultural producer's crop rotation horizon is finite. The stationarity assumption means that the time dependence of crop choices is the same at all dates. We implicitly assume here that there is no exogenous shock during the period considered[7]. In the next paragraphs, we estimate by ME the transition probability matrix characterizing the Markov process.

We have made the assumption that aggregate land use follows a *r*-order Markov process. Any *r*-order Markov process may be rewritten as a first-order process by enlarging the space of possible states,

[6] Some recent econometrics estimates of land use may be found in Hardie and Parks (1997), Plantinga (1996) or in Wu and Segerson (1995).

[7] The stationary of aggregate land shares can be easily tested in practice. We should mention that our estimation procedure could be extended to non-stationary aggregate data. It would require however the use in the model of exogenous variables in order to capture the non-stationary trend of land shares.

Kijima (1997). Let us consider the sequence corresponding to crop choices from period *t-r+1* to period *t*. It is straightforward to show that, given that aggregate land use follows a *r*-order Markov process, this *r-observations* sequence follows a first-order Markov process. A Markov state is a sequence of *r*-observed crops. At each period, farmers choose among *K* possible crops hence Markov states are defined on the Markov space $,^{\mathrm{r}} = \left\{1,\ldots,K^r\right\}$. There exist K^r states corresponding to the K^r possible *r*-tuplets. States are indexed by $j \in \{1,\ldots,J\}$ with $J = K^r$. The probability associated to any state j is denoted $Q_j(t)$ and is equal the product of probabilities $Y_k(t)$ corresponding to the crop sequence indexed by *j*. For example, assuming a *second-order* Markov process, the probability of producing k_1 in *t*-1 and k_2 in *t* is given by $Y_{k_1}(t-1) \times Y_{k_2}(t)$. Computed in this way, $Q_j(t)$ satisfies the two properties required for a probability distribution, namely: $Q_j(t) \in [0,1]$ and $\sum_j Q_j(t) = 1$.

Now, let T be the $(J \times J)$ Markov transition matrix associated with the first-order Markov process. $T_{jj'}$ gives the probability of passing from any state $j \in \{1\ldots,J\}$ at date *t* to any state $j' \in \{1\ldots,J\}$ at date *t*+1. The transition probabilities satisfy the two usual properties ~~that~~: $T_{jj'} \geq 0\ \forall j,j'$ and $\sum_{j'} T_{jj'} = 1\ \forall j$. It follows that the probability of being in state *j'* in *t*+1 is given by:

$$Q_{j'}(t+1) = \sum_{j=1}^{J} Q_j(t) \cdot T_{jj'}\ ,\ \forall j \in \{1,\ldots,J\} \text{ and } \forall t \in \{r,\ldots,T-1\}. \tag{3}$$

The first step consists of estimating the Markov transition probabilities from the aggregate data. We use a maximum entropy, (ME) method.

2.2.2 The ME formalism for the Markov transition metrics

Notice that only $J \times (K-1)$ parameters of the Markov transition matrix have to be estimated. First, for a given Markov state the number of possible outcomes is *K*. For example, assuming a second-order Markov process and observing the Markov state $\left\{\bar{k}(t), \bar{k}(t-1)\right\}$ at date *t*, the resulting Markov state in *t+1* belongs to the set $\left\{k(t+1), \bar{k}(t)\right\}$. At most $J \times K$ transition probabilities are strictly positive. Moreover, as the sum of transition probabilities must be equal to 1, $J \times (K-1)$ transition probabilities have to be considered. Remember that we have $T-r>0$ periods for which $Q_j(t)$ can be computed. It follows that $J\times(T-r)$ observations can be used.

If $T-r>K-1$ which results in cases where there are a large number of periods, a limited number of crops and / or a small Markov order, the usual econometric methods can be implemented. Burton (1997) specifies, for example, a stationary and a non-stationary first-order Markov process for describing employment movements in the forestry sector. Burton (1997) uses Zellner's Seemingly Unrelated Regression to estimate the transition probabilities.

However, in many cases the number of periods is small and the $T-r>K-1$ condition does not hold. This is the situation that we consider in the remainder of the paper. In this case, there are more parameters to be estimated than available moment conditions and the problem is ill-posed. The matrix of exogenous variables is non-invertible and usual traditional mathematical procedures would result in arbitrary choosing $J\times(K-1-T+r)$ parameters. Given there exist multiple solutions, we must use a way of selecting a particular transition matrix. In this example, we estimate the transition probabilities associated to the J states by maximum entropy (ME). This approach yields a unique solution.

We assume that we observe the aggregate Markov states with noise. This unobservable disturbance may result from sample and non-sample errors, specification and modeling errors or imperfect measurement of economic agent behavior. Let $e(t)$ be the J dimensional vector of disturbances. The specification of the Markov process with noise is:

$$Q_{j'}(t+1)=\sum_{j=1}^{J}Q_j(t)\cdot T_{jj'}+e_{j'}(t)\,,\ \forall j'\in\{1,\ldots,J\}\text{ and }\forall t\in\{r,\ldots,T-1\}. \tag{4}$$

Following the ME formalism, we reparameterize this equation in terms of currently unknown probability distributions.

- By definition, $T_{jj'}$ is between zero and one. We can define a set $\{u_1,\ldots,u_M\}$ of $M\geq 2$ points with $u_1=0$, $u_M=1$ and a probability distribution $\{T_{jj'1},\ldots,T_{jj'M}\}$ such as $T_{jj'}=\sum_{m=1}^{M}u_m\cdot T_{jj'm}$.
- The unknown disturbances $e_{j'}(t)$ may be treated in a similar way. By defining a set of N points, $\{v_1,\ldots,v_N\}$, with $N\geq 2$ and $\{e_{j'1}(t),\ldots,e_{j'N}(t)\}$ as the associated probabilities, we have:

$e_{j'}(t)=\sum_{n=1}^{N} v_n \cdot e_{j'n}(t)$. Note that $e_{j'}(t)$ is the estimated expected error in equation (4), while $e_{j'n}(t)$ are the n estimated probabilities over which the expected error is taken.

Given this reparametrization, we have $\forall j' \in \{1,\ldots,J\}$ and $\forall t \in \{r,\ldots,T-1\}$:

$$Q_{j'}(t+1)=\sum_{j=1}^{J}\left\{Q_j(t)\cdot\sum_{m=1}^{M}u_m \cdot T_{jj'm}\right\}+\sum_{n=1}^{N}v_n \cdot e_{j'n}(t). \tag{5}$$

The problem of recovering transition probabilities can be formulated in a standard generalized cross-entropy framework (GCE). Given the GCE formalism, we seek to solve the following nonlinear optimization program:

$$\underset{T,e}{Min}\ H(T,e)=-\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{m=1}^{M}T_{jj'm}\cdot\log\left(\frac{T_{jj'm}}{T^{p}_{jj'm}}\right)-\sum_{j'=1}^{J}\sum_{t=r}^{T-1}\sum_{n=1}^{N}e_{j'n}(t)\cdot\log\left(\frac{e_{j'n}(t)}{e^{p}_{j'n}(t)}\right) \tag{6}$$

subject to :

$$Q_{j'}(t+1)=\sum_{j=1}^{J}\left\{Q_j(t)\cdot\sum_{m=1}^{M}u_m \cdot T_{jj'm}\right\}+\sum_{n=1}^{N}v_n \cdot e_{j'n}(t)\quad \forall j'\ \forall t \tag{7}$$

$$\sum_{j'=1}^{J}\sum_{m=1}^{M}u_m \cdot T_{jj'm}=1\ \ \forall j \tag{8}$$

$$\sum_{m=1}^{M}T_{jj'm}=1 \text{ and } T_{jj'm}\in[0,1]\ \forall j,j',m \tag{9}$$

$$\sum_{n=1}^{N}e_{j'n}(t)=1 \text{ and } e_{j'n}(t)\in[0,1]\ \forall j',t,n \tag{10}$$

In the objective function, terms $T^{p}_{jj'm}$ and $e^{p}_{j'n}(t)$ represent prior probabilities associated respectively with the parameters and errors. Following the definition of the entropy, $T_{jj'm}\cdot\log\left(T_{jj'm}/T^{p}_{jj'm}\right)$ is taken to be 0 when $T_{jj'm}=0$. Equation (7) of this program are the data-constraints of the model. Condition (8) corresponds to the second property of transition probabilities. Conditions (9)-(10) are the adding-up constraints. In terms of the GME formalism, the maximum entropy is achieved when the moment relation (7) is not enforced and the distributions of probabilities for the parameters and errors are equal to the priors. Adding the data-constraint (7) causes the posterior distributions to differ from the prior distributions.

2.2.3 The estimated Markov transition matrix

The optimization program (6)-(10) constitutes a standard GCE problem. The interested reader may consult Golan, Judge and Miller (1996) for a complete and detailed derivation of this program's solutions, ($\hat{T}_{jj'm}, \hat{e}_{j'n}(t)$). However, we would like to emphasize two points. First, this GCE program has a unique solution in *(T,e)*. Second, there is no closed-form solution and we must resort to numerical optimization to find this unique solution.

From the GCE estimates of the probabilities $\hat{T}_{jj'm}$ and $\hat{e}_{j'n}(t)$, it is easy to recover a point estimate, both for transition probabilities and error terms defined in equation (4). More formally:

$$\hat{T}_{jj'} = \sum_{m=1}^{M} u_m \cdot \hat{T}_{jj'm} = 1 \ \forall j, j' \text{ and } \hat{e}_{j'}(t) = \sum_{n=1}^{N} v_n \cdot \hat{e}_{j'n}(t) \ \forall j'. \tag{11}$$

At this point of the analysis, we have estimated the transition matrix of a Markov process using aggregate data. We are now able to associate to any state, the set of probabilities of going to any other state. We turn now to the problem of recovering of individual crop choices, that is to the data disaggregation problem.

2.3 Data disaggregation to the field-level

The second step of our analysis consists of disaggregating, for specific time periods, the aggregate observations of land use to field-level crop predictions. The field-level estimates must be consistent with the observed aggregate outcomes. Moreover they must 'fit' the behavior defined by the Markov process. More precisely, we want to estimate cropping choices at ~~field level~~ field-level $\hat{c}_i(t) \in \{1, \ldots, K\}$ for $t \geq r+1$. We want these estimates to be consistent with:

- the aggregate data, defined by $S_k(t)$ for $t \geq r+1$,
- the Markov metrics defined by $\hat{T}$.

Both $\hat{T}$ and $S_k(t)$ are used in the estimation process by specifying a Maximum Entropy (ME) problem that uses the information given by the Markov metrics $\hat{T}$ as priors and the information on the aggregate crop allocation $S_k(t)$ as additional data through the data-compatibility constraint.

2.3.1 Writing the data-compatibility constraint of the disaggregation problem

At each period, field-level cropping choices must be compatible with the observed allocation of land at the regional-level. Let us assume that at some date $t \in \{r+1,\ldots,T\}$, the Markov state associated with field $i \in \{1,\ldots,I\}$ is $j(i)$. Observing the aggregate crop area at date t, $S_k(t)\ \ \forall k = 1,\ldots,K$, results in a set of K constraints:

$$\sum_{i=1}^{I} \sum_{j' \in \Psi(k)} \delta_{j(i)j'}(t) \cdot s_i = S_k(t) \quad \forall k = 1,\ldots,K \qquad (12)$$

where:

- $\Psi(k)$ is the set of Markov states for which crop *k* is produced at the last period;
- $\delta_{j(i)j'}(t)$ is a dummy variable equal to 1, if transition from state $j(i)$ in period $t-1$ to state j' in period t is observed, and to 0 otherwise. By definition, we have: $\delta_{j(i)j'}(t) \in \{0,1\}$ and $\sum_{j'} \delta_{j(i)j'}(t) = 1$ for all $i \in \{1,\ldots,I\}$.

This data-compatibility constraint (12) simply states that the sum of field acreage predicted to produce crop *k – that is all fields falling in a Markov state that belongs to* $\Psi(k)$ *in* t*–* must be equal to the observed aggregated area allocated to crop *k* in t, $S_k(t)$.

Remark 1:

Given we observe individual crop choices at field-level $c_i(t)$ for $t = 1,\ldots,r$ we can associate to each field *i,* the corresponding Markov state $j(i)$ for year $t = r$. For later periods under the open-loop assumption, the crop choice at the field-level is no longer observable. As the data disaggregation program is solved from year to year, we use the predicted crop choice $\hat{c}_i(t)$ for $t = r+1,\ldots,T$, obtained from the previous period in an open-loop, to associate a Markov state to each field.

2.3.2 ME formalism for the data disaggregation problem

First, we add an unobservable noise term $\varepsilon_k(t)$ to the data-constraint (12). This unknown disturbance ε_k may be treated like the error term $e_{j'}$ in the Markov problem, equation (4). By denoting the error support values $\{w_1,\ldots w_N\}$, where $N \geq 2$, and $\{\varepsilon_{k1}(t),\ldots \varepsilon_{kN}(t)\}$ as the associated probabilities, we have: $\varepsilon_k(t) = \sum_{n=1}^{N} w_n \cdot \varepsilon_{kn}(t)$. Using the ME formalism, the data-compatibility constraint may be rewritten for a given date t, $t \in \{r+1,\ldots,T\}$, as:

$$\sum_{i=1}^{I} \sum_{j' \in \Psi(k)} \delta_{j(i)j'}(t) \cdot s_i + \sum_{n=1}^{N} w_n \cdot \varepsilon_{kn}(t) = S_k(t) \quad \forall k = 1,\ldots,K \,. \tag{13}$$

Given this constraint, the problem of data disaggregation can be formulated as an integer generalized cross-entropy framework (GCE). For a given period[8], we have to solve the following nonlinear optimization program:

$$\underset{\delta,\varepsilon}{Min}\, H(\delta,\xi) = -\sum_{i=1}^{I} \sum_{j'=1}^{J} \delta_{j(i)j'} \cdot \log\left(\frac{\delta_{j(i)j'}}{\delta^{p}_{j(i)j'}}\right) - \sum_{k=1}^{K} \sum_{n=1}^{N} \varepsilon_{kn} \cdot \log\left(\frac{\varepsilon_{kn}}{\varepsilon^{p}_{kn}}\right) \tag{14}$$

subject to :

$$\sum_{i=1}^{I} \sum_{j' \in \Psi(k)} \delta_{j(i)j'} \cdot s_i + \sum_{n=1}^{N} w_n \cdot \varepsilon_{kn} = S_k \quad \forall k = 1,\ldots,K \tag{15}$$

$$\sum_{j'=1}^{J} \delta_{j(i)j'} = 1 \quad \forall i = 1,\ldots,I \tag{16}$$

$$\delta_{j(i)j'} \in \{0,1\} \tag{17}$$

$$\sum_{n=1}^{N} \varepsilon_{kn} = 1 \ \forall k = 1,\ldots,K \tag{18}$$

$$\varepsilon_{kn}(t) \in [0,1]. \tag{19}$$

where $\delta^{p}_{j(i)j'}$ and ε^{p}_{kn} represent priors respectively associated with $\delta_{j(i)j'}$ and ε_{kn}. Let us first give an intuitive description of the problem. We want to find first, a degenerate distribution that attributes to each

[8] For reasons of simplicity we omit the time index in the following program. Since we want to predict field-level crop choice for periods $r+1$ to T, $T-1$ nonlinear optimization programs must be solved.

field at each date a Markov state with a probability equal to one and second, a distribution for the usual error term. These two distributions must maximize a GCE objective and have to be consistent with the data-constraints. Returning to the definition of priors $\delta^{p}_{j(i)j'}$, a natural way of choosing these priors is to select transition probabilities from the Markov matrix for aggregate land use: $\delta^{p}_{j(i)j'} = \hat{T}_{j(i)j'}$. Finally, the recovery of the field-level crop choices, $\hat{c}_i$, can be performed using the estimates $\hat{\delta}_{j(i)j'}$. The predicted crop for field *i* for a given date is given by the unique Markov state *j'* such that $\hat{\delta}_{j(i)j'} = 1$.

Remark 2:

An unusual feature of this optimization program is shown in constraint (17). This constraint means that we want $\hat{\delta}_{j(i)j'}$ to be a degenerate distribution. We want to associate a Markov state with a probability equal to one to each field. It follows that we must solve a mixed-integer non-linear problem (MINLP). (See section 3 for computational details.) If we do not impose $\hat{\delta}_{j(i)j'}$ to be a degenerate distribution of probabilities, then constraint (17) becomes $\delta_{j(i)j'} \in [0,1]$. For each field, the solution to this *relaxed data disaggregation* program is a distribution of probabilities associated to the different types of crops. In this case the data-compatibility constraint (15) holds in an expected sense. This relaxed version of the data disaggregation problem is similar to that done by Miller and Plantinga (1997) for a different model of crop use in a static framework. Howitt and ~~Arnaud~~ Reynaud (2001) develop a similar approach as the relaxed model, but in a dynamic context.

Remark 3:

If we solve this program without taking into account the data-constraint expressed by (15), the solution consists in attributing to each field the Markov state with the highest prior. Adding the acreage constraint results in a change of the optimal allocation from this most likely state according to the priors, in a data-consistent way.

Remark 4:

The ME approach provides an easy way to formally incorporate out-of-sample information. Out-of-sample information may either consist in additional constraints in the optimization program or in particular priors for the Markov transition metrics. For example, some specific physical constraints (quality of soil, water availability, agronomic constraints) may prevent farmers from producing particular crops. This information may be added to the data disaggregation program with additional constraints. Any out-of-the sample information on transition probabilities may be added to the model through the prior specification. In fact, we are not obliged to use an estimated Markov process for priors. We could have out-sample priors from other sources of information. For example, such priors may be based on data from previous surveys or from detailed surveys such as the National Resources Inventory (NRI) survey.

Remark 5:

If we have no a priori information on parameters to be estimated, uninformative uniform distributions may be considered as priors.

III. Data disaggregation in practice

In this section we apply our data disaggregation approach to a sample of fields for which we observe crop production in each year. This detailed data set enables us to test the data disaggregation model by evaluating the quality of field-level predictions. When used to disaggregate the outcomes of economic models, the disaggregation approach only requires r time periods (–two in this example) of field observations to initialize the projections, an aggregate time series to estimate the Markov matrix, and the aggregate values from the model.

3.1 The data

We use a sample of 190 fields observed from 1986 to 1990. A field is assumed to produce one crop each year from among the following six: Alfalfa, Cotton, Field, Grain, Melons and Tomatoes[9]. According to the previous section notations, we have $T=5$ and $K=6$. In what follows, the first letter indexes each crop. Therefore, we have $k \in \{A, C, F, G, M, T\}$. Table 1 gives the aggregate crop area by year.

Table 1: Allocation of land per crop

	1986		1987		1988		1989		1990	
Crop	**Acreage**	**%**	**Acreage**	**%**	**Acreage**	**%**	**Acreage**	**%**	**Acreage**	**%**
Alfalfa	2103	12.6	2072	12.4	1833	11.0	1969	11.8	1986	11.9
Cotton	8893	53.2	8902	53.2	10052	60.1	8920	53.3	9160	54.8
Field	2386	14.3	2358	14.1	1508	9.0	2119	12.7	1144	6.8
Grain	1062	6.3	1062	6.3	1102	6.6	1891	11.3	1312	7.8
Melons	729	4.4	765	4.6	426	2.5	418	2.5	1120	6.7
Tomatoes	1556	9.3	1570	9.4	1808	10.8	1411	8.4	2007	12.0
Total	16728	100	16728	100	16728	100	16728	100	16728	100

The data in table 1 are used to estimate the r-order Markov process. Crop choice is represented by a *second*-order Markov process, $r=2$. Hence, a Markov state is a pair of crops observed during two consecutive years. As a consequence, there are 36 possible states each year. States are indexed in the

[9] The initial sample was made up of 364 fields producing 14 possible types of land use: alfalfa, cotton, fallow, field, grain, melons, other, pasture, perennial, rental, tomatoes, unknown, unused, vegetables. We have kept the six most important crops and we have ruled out of the final sample any field not producing one of these crops. The resulting number of observations is 190.

following way: state *kk'* observed in *t* means that crop *k'* was produced in *t-1* and crop *k* in *t,* $k,k' \in \{A,C,F,G,M,T\}$. Notice that each state has six possible outcomes: having observed $\overline{kk}'$ in *t* necessarily results in observing $k\bar{k}$ with $k \in \{A,C,F,G,M,T\}$ in *t+1*. At most, we have to estimate 216 transition probabilities. Finally, as we have $T-r \leq K-1$, the problem is ill-posed and use of the ME approach to estimate the Markov metrics is required.

In order to test our model, we first estimate the Markov transition matrix using aggregate data, then we disaggregate crop choices at the field-level and finally we compare the ~~field level~~ field-level crop predictions, ~~with those actually observed~~ under the open-~~-~~loop and closed-~~-~~loop specifications, with those actually observed. In approximately a third of the fields in the sample the cross entropy integer crop prediction differed from the crop with the highest expected prior probability, showing the additional information added by the integer disaggregation procedure.

3.2 The Markov metrics

3.2.1 A formal specification of the Markov estimation

We use a weighted version of the CGE program (6)-(10) to estimate the Markov metrics. In the objective function of the program defined in equations (6)-(10), an equal weight is implicitly placed on the parameter *p* and the noise *e*. This means that the objectives of minimizing the deviations from the prior rotation predictions, and the precision with which the model predicts the aggregate crop production, are equally weighted. By choosing different weights on the γ parameter, we can tune the estimation to favor the prior rotational predictions or the precision with which the model fits the current aggregate crop data.

The transition matrix is the solution of the following weighted GCE optimization program:

$$\underset{T,e}{Min}\, H(w,p) = -(1-\gamma)\cdot \sum_{j=1}^{36}\sum_{j'=1}^{36}\sum_{m=1}^{M} T_{jj'm}\log\left(T_{jj'm} \big/ T^{p}_{jj'm}\right) - \gamma\cdot \sum_{j'=1}^{36}\sum_{n=1}^{N}\sum_{t=r}^{T-1} e_{j'n}(t)\cdot \log\left(e_{j'n}(t) \big/ e^{p}_{j'n}(t)\right) \quad (20)$$

subject to the data-constraint :

$$Q_{j'}(t+1) = \sum_{j=1}^{J}\left\{Q_j(t)\cdot \sum_{m=1}^{M} u_m \cdot T_{jj'm}\right\} + \sum_{n=1}^{N} v_n \cdot e_{j'n}(t) \quad \forall j' \ \forall t \geq 88 \quad (21)$$

and the adding-up constraints given by (8)-(10). In the objective function, the parameter $\gamma \in [0,1]$ is the weight placed on the error entropy.

Two questions have to be answered before progressing to the estimation phase. First, we have to choose support values for the errors and parameters and second we have to specify the priors. The natural bounds for the Markov support values (u_m) are zero and one. In addition, we have to choose the number M of support values. Since previous studies have shown that increasing the number support points from 3 to 5 points has little effect on the estimates[10], as opposed to the range of support values, we fix M equal to 3. It follows that the vector of support values is $\{0,\ 0.5, 1\}$. The choice of error support values clearly depends on the properties of the errors e. By reference to the Chebyshev's inequality, some authors determine the bounds using a 3σ rule, Golan, Judge and Miller (1996). This is the rule followed here. The number N of values for the error support is 3. Finally, as we do not have any a priori for parameter and error distributions, we use uniform priors: $T^{p}_{jj'm} = 1/M\ \forall m$ and $e^{p}_{j'n} = 1/N\ \forall n$.

3.2.2 Results of Markov estimates

Table 2: Estimation of the Markov metrics

		AA	**CA**	**FA**	**GA**	**MA**	**TA**			**AC**	**CC**	**FC**	**GC**	**MC**	**TC**
ALFALFA	**AA**	0.150	0.347	0.154	0.134	0.078	0.137	**COTTON**	**CA**	0.048	0.672	0.180	0.092	0.000	0.008
	AC	0.086	0.742	0.042	0.042	0.010	0.078		**CC**	0.135	0.635	0.010	0.061	0.028	0.131
	AF	0.130	0.417	0.167	0.124	0.053	0.109		**CF**	0.006	0.705	0.250	0.039	0.000	0.000
	AG	0.165	0.238	0.158	0.154	0.124	0.161		**CG**	0.168	0.296	0.181	0.186	0.054	0.115
	AM	0.160	0.235	0.177	0.160	0.119	0.149		**CM**	0.090	0.200	0.428	0.267	0.001	0.014
	AT	0.167	0.260	0.146	0.147	0.115	0.165		**CT**	0.181	0.157	0.105	0.162	0.200	0.195
		AF	**CF**	**FF**	**GF**	**MF**	**TF**			**AG**	**CG**	**FG**	**GG**	**MG**	**TG**
FIELD	**FA**	0.147	0.362	0.149	0.127	0.077	0.138	**GRAIN**	**GA**	0.156	0.305	0.148	0.140	0.104	0.147
	FC	0.087	0.740	0.039	0.039	0.011	0.084		**GC**	0.086	0.733	0.038	0.042	0.018	0.083
	FF	0.124	0.451	0.156	0.110	0.051	0.108		**GF**	0.148	0.333	0.161	0.140	0.088	0.130
	FG	0.165	0.240	0.156	0.152	0.126	0.161		**GG**	0.165	0.229	0.154	0.153	0.136	0.163
	FM	0.158	0.245	0.174	0.155	0.118	0.150		**GM**	0.163	0.212	0.170	0.162	0.138	0.155
	FT	0.167	0.258	0.144	0.146	0.118	0.167		**GT**	0.164	0.259	0.142	0.145	0.126	0.164
		AM	**CM**	**FM**	**GM**	**MM**	**TM**			**AT**	**CT**	**FT**	**GT**	**MT**	**TT**
MELON	**MA**	0.156	0.281	0.149	0.141	0.120	0.153	**TOMATOES**	**TA**	0.155	0.324	0.157	0.142	0.081	0.141
	MC	0.083	0.697	0.064	0.053	0.027	0.076		**TC**	0.081	0.748	0.051	0.045	0.007	0.068
	MF	0.153	0.297	0.148	0.138	0.114	0.150		**TF**	0.143	0.366	0.168	0.138	0.064	0.121
	MG	0.163	0.221	0.158	0.155	0.142	0.161		**TG**	0.166	0.235	0.160	0.156	0.122	0.161
	MM	0.164	0.202	0.162	0.159	0.150	0.163		**TM**	0.163	0.221	0.175	0.163	0.125	0.153
	MT	0.160	0.250	0.153	0.148	0.131	0.158		**TT**	0.166	0.264	0.151	0.149	0.109	0.161

[10] See for example Golan, Judge and Miller (1996). These authors have shown that passing from 2 points to 3 substantialy decreases the mean-square-error of estimates. Futher increase in M is shown to only result in a smaller improvement.

Note: Table 2 gives transition probabilities of passing from one Markov state to another. For example, the transition probability of producing Cotton in *t+1* after having produced Alfalfa in *t* and Grain in *t-1* is 0.238. It corresponds to the probability of passing from state AG to state CA.

Table 2 gives estimation of the Markov metrics for an error-weight γ in the entropy criterion equal to 0.9. The initial states composed of this and the previous year's cropping history are tabulated in the left hand column, while the state in the next time period is tabulated in the rows. These transition probabilities are estimated using aggregate data, but reflect individual behavior. One way of evaluating the Markov metrics is to compute the normalized entropy for the parameters and the error terms, Golan, Judge and Miller (1996). The normalized entropy is a measure of the relative information content of GME estimates. The normalized entropy is defined for a signal T as:

$$S(\hat{T}) = \frac{-\sum_{j}\sum_{j'}\sum_{m}\hat{T}_{jj'm}\cdot\ln(\hat{T}_{jj'm})}{J\cdot J\cdot\ln(M)}. \qquad (22\text{-a})$$

The normalized entropy varies from zero to one. At one extreme, $S(\hat{T})=0$ means that there is no uncertainty in the parameter distribution whereas $S(\hat{T})=1$ corresponds to an uninformative uniform distribution. The estimate normalized entropy for T is equal to 0.588. A similar measure of normalized entropy for e is given by:

$$S(\hat{e}) = \frac{-\sum_{j'}\sum_{t}\sum_{n}\hat{e}_{j'n}\cdot\ln(\hat{e}_{j'n})}{J\cdot T\cdot\ln(N)}. \qquad (22\text{-b})$$

The normalized entropy for the error terms e is equal to 0.959.

Another way of evaluating the Markov metrics is to compare the predicted crop shares from 1988 to 1990 with the observed aggregate shares. Table 3, presents open-loop simulations of the Markov metrics[11].

Table 3: Observed versus predicted land crop shares, $Y_k(t)$

	1988		1989		1990	
Crop	**Observed**	**Predicted**	**Observed**	**Predicted**	**Observed**	**Predicted**
Alfalfa	0.110	0.113	0.118	0.114	0.119	0.118

[11] Given aggregate land use for 1986 and 1987, the Markov metric gives a prediction of land use shares for 1988. Then given observed land use share in 1987 and predicted in 1988, we estimate land use for 1989. Finally, land allocation in 1990 is based on predicted land use shares in 1988 and 1989.

Cotton	0.601	0.553	0.533	0.550	0.548	0.544
Field	0.090	0.100	0.127	0.101	0.068	0.094
Grain	0.066	0.088	0.113	0.088	0.078	0.090
Melons	0.025	0.044	0.025	0.045	0.067	0.047
Tomatoes	0.108	0.102	0.084	0.103	0.120	0.107

The Markov metrics performs quite well, both in terms of the level prediction and their variation over time. The average Percentage Absolute Predicted Error (PAPE) for year 1988 to 1990 is respectively equal to 12.1%, 14.4% and 13.2%[12]. This shows that the Markov metrics recovers the aggregated crop acreage with an acceptable precision. It is also interesting to notice that the direction of changing crop shares between two dates is correctly predicted eight times out of twelve.

3.3 Data disaggregation to field-level

To test the disaggregation method over a significant, but manageable number of fields, we randomly select a *25*-observation sub-sample[13] from the initial *190*-observation sample. We want to estimate the discrete individual field crop choices for 1988 to 1990, conditional on the estimated Markov metrics and the field-level crop choices for 1986 and 1987.

Table 4: Allocation of land per crop for the sub-sample

	1986		1987		1988		1989		1990	
Crop	**Acreage**	**%**	**Acreage**	**%**	**Acreage**	**%**	**Acreage**	**%**	**Acreage**	**%**
Alfalfa	500.27	21.5	500.27	21.5	364.25	15.6	338.03	14.5	330.37	14.2
Cotton	1385.27	59.5	1403.95	60.3	1334.22	57.3	867.50	37.3	1487.59	63.9
Field	199.92	8.6	181.24	7.8	217.56	9.3	727.46	31.2	0.00	0.0
Grain	66.15	2.8	66.15	2.8	0.00	0.0	203.40	8.7	33.88	1.5

[12] The average Percentage Absolute Predicted Error is $\frac{1}{K \cdot Y_k^o} \sum_k \left| Y_k^o - \hat{Y}_k \right| * 100$ where Y_k^o and $\hat{Y}_k$ represent respectively the observed and estimated probability of producing crop *k*. The PAPE figures presented above do not take into account melons. For this crop, the PAPE varies from 29.9% in 1990 to 80.0% in 1989. However, these results should be related to the fact that melons has the smallest land use share in each year.

[13] We use a sub-sample of fields in order to simplify the empirical application presentation. However, it is clear that for a given number of crops, the higher is the sample size, the more difficult is the data disaggregation. The number of parameters to be estimated increases with the size of the sample *I,* whereas the number of data-constraints only depends on the number of produced crops*, K*. In fact, the complexity of the data disaggregation problem increases with *I-K*

Melons	177.19	7.6	177.19	7.6	20.62	0.9	0.00	0.0	0.00	0.0
Tomatoes	0.00	0.0	0.00	0.0	392.15	16.8	192.40	8.3	476.96	20.5
Total	2329	100	2329	100	2329	100	2329	100	2329	100

Table 4 describes the land allocation in the subsample. See also Table 6 for the observed sequence of crops for each field.

3.3.1 The specification of the data disaggregation problem

The data disaggregation problem for $t \geq 1988$ can be reformulated in the following weighted CGE framework. For each period, we have to:

$$\underset{\delta,\varepsilon}{Min}\, H(\delta,\varepsilon) = -(1-\lambda)\cdot\sum_{i=1}^{25}\sum_{j'=1}^{36}\delta_{j(i)j'}(t)\cdot\log\left(\delta_{j(i)j'}(t)\big/\hat{T}_{j(i)j'}\right) - \lambda\cdot\sum_{k=1}^{6}\sum_{n=1}^{N}\varepsilon_{kn}(t)\cdot\log\left(\varepsilon_{kn}(t)\big/\varepsilon_{kn}^{p}(t)\right) \qquad (23)$$

subject to the data-constraint:

$$\sum_{i=1}^{25}\sum_{j'\in\Psi(k)}\delta_{j(i)j'}(t)\cdot s_i + \sum_{n=1}^{N} w_n\cdot\varepsilon_{kn} = S_k(t) \quad \forall k = 1,\ldots,K \qquad (24)$$

and subject to the integer constraint $\delta_{j(i)j'}(t) \in \{0,1\}$ (17), and the usual adding-up constraints (16), (18) and (19). The λ weighting parameter is similar to γ in the Markov problem and has to be chosen by the user. For the numerical application, an error weight $\lambda = 0.8$ was selected. Three supports are used for the error term.

In the following sections we first consider the *relaxed data disaggregation problem* (CGE programs without the integer constraint) and then the *complete data disaggregation problem* (CGE programs with the integer constraint).

3.3.2 The relaxed data disaggregation problem

The key difference is that we do not impose $\delta_{j(i)j'}$ to be a degenerate distribution of probabilities. The solution of this relaxed program is, for any given field, a distribution of probabilities over the six possible crops. As previously mentioned, this relaxed version is similar to the approach used by Miller and Plantinga (1997) for a different model of crop use. Miller and Plantinga (1997) calculate ME estimates of the multi-county aggregate optimal allocation of land, using out-sample information as priors.

'For land use applications, we may interpret the presample shares as estimates of the expected land use allocations made without reference to additional sample information. The presample shares may be based on data from a previous survey, and, for example, high-quality regional shares may be assembled from detailed plot-level studies such as the periodic National Resources Inventory survey, (NRI.)'.

In our data disaggregation approach, the priors are endogenous. A natural way of choosing priors for 1988 is to select the transition probabilities from the Markov metric associated with the observed state, $T_{j(i)j'}$.

In Table 5, we show the results of the relaxed data disaggregation problem for 1988, $\hat{\delta}_{j(i)j'}(1988)$. Estimates for years 1989 and 1990 could be derived in the same way. The priors shown in Table 5 come directly from the estimated Markov metrics, Table 2. For example, the first line of Table 2 gives the priors for field 14 (corresponding to Alfalfa in 1986 and 1987).

Remark 6:

Table 5 allows a direct comparison of prior and posterior probabilities. There are significant differences between priors and posteriors. For example, the prior for field 9 corresponding to *Alfalfa* in 1988 after having produced *Grain* in 1986 and 1987 is 0.165 whereas the posterior probability is 0.204. We would like to emphasize that these differences do not invalidate the Markov metric estimate. These differences mean that the relaxed data-compatibility constraint contains a significant information signal for field-level allocation. In other words, the relaxed data-constraint modifies the optimal field-level crop pattern away from that yielded by the most likely allocation based on the expected value of the priors.

Remark 7:

It is interesting to note that posteriors for fields with identical initial Markov states differ slightly. For example, the probability of producing *Tomatoes* in 1988 given *Cotton* in 1986 and in 1987 varies from 0.145 (for field 11) to 0.170 (for field 4), in contrast, the prior value from table 2 for the CC initial state is

0.131. This variation is the consequence of field size heterogeneity, which has to be reflected in the relaxed data-compatibility constraint in the optimization program.

Table 5: Priors versus posteriors for 1988 with relaxed data disaggregation approach.

Field	Initial State	ALFALFA		COTTON		FIELD		GRAIN		MELON		TOMATOES	
		Prior	Post.	Prior	Post.	Prior	Post.	Prior	Post.	Prior	Post.	Prior	Post.
1	AA	0.150	0.179	0.347	0.416	0.154	0.189	0.134	0	0.078	0.048	0.137	0.168
2	GG	0.165	0.204	0.229	0.284	0.154	0.194	0.153	0	0.136	0.112	0.163	0.205
3	AA	0.150	0.178	0.347	0.413	0.154	0.187	0.134	0	0.078	0.055	0.137	0.166
4	CC	0.135	0.138	0.635	0.678	0.010	0.013	0.061	0	0.028	0.000	0.131	0.170
5	GG	0.165	0.206	0.229	0.286	0.154	0.196	0.153	0	0.136	0.104	0.163	0.207
6	CC	0.135	0.145	0.635	0.687	0.010	0.012	0.061	0	0.028	0.008	0.131	0.149
7	CC	0.135	0.145	0.635	0.687	0.010	0.012	0.061	0	0.028	0.008	0.131	0.149
8	FF	0.124	0.142	0.451	0.521	0.156	0.189	0.110	0	0.051	0.018	0.108	0.130
9	GG	0.165	0.204	0.229	0.284	0.154	0.194	0.153	0	0.136	0.112	0.163	0.205
10	CC	0.135	0.145	0.635	0.687	0.010	0.012	0.061	0	0.028	0.008	0.131	0.149
11	CC	0.135	0.145	0.635	0.684	0.010	0.011	0.061	0	0.028	0.015	0.131	0.145
12	CC	0.135	0.143	0.635	0.686	0.010	0.012	0.061	0	0.028	0.001	0.131	0.158
13	MM	0.164	0.217	0.202	0.274	0.162	0.252	0.159	0	0.150	0.006	0.163	0.250
14	AA	0.150	0.179	0.347	0.422	0.154	0.209	0.134	0	0.078	0.006	0.137	0.184
15	AA	0.150	0.179	0.347	0.422	0.154	0.209	0.134	0	0.078	0.006	0.137	0.184
16	CC	0.135	0.143	0.635	0.686	0.010	0.012	0.061	0	0.028	0.002	0.131	0.157
17	CC	0.135	0.142	0.635	0.685	0.010	0.012	0.061	0	0.028	0.001	0.131	0.159
18	CF	0.006	0	0.705	0.737	0.250	0.263	0.039	0	0	0	0	0
19	CF	0.006	0	0.705	0.737	0.250	0.263	0.039	0	0	0	0	0
20	CC	0.135	0.143	0.635	0.686	0.010	0.012	0.061	0	0.028	0.001	0.131	0.158
21	CC	0.135	0.143	0.635	0.686	0.010	0.012	0.061	0	0.028	0.001	0.131	0.158
22	CC	0.135	0.145	0.635	0.687	0.010	0.012	0.061	0	0.028	0.007	0.131	0.149
23	FF	0.124	0.142	0.451	0.521	0.156	0.189	0.110	0	0.051	0.018	0.108	0.130
24	AA	0.150	0.178	0.347	0.421	0.154	0.211	0.134	0	0.078	0.004	0.137	0.185
25	FF	0.124	0.142	0.451	0.521	0.156	0.189	0.110	0	0.051	0.017	0.108	0.130

Note: Table 5 gives, for each field of the sub-sample, the initial Markov state in 1988 and distributions of priors (given by the Markov transition matrix) and posteriors (resulting from the relaxed data disaggregation program).

3.3.2 Solution of the complete data disaggregation problem

We turn now to the complete data disaggregation problem, which consists of choosing, for each field, the optimal crop to be produced with probability one.

a- Solving the optimization program with GAMS

The complete data disaggregation problem defined in equations (23) and (24) requires solving a Mixed-Integer NonLinear Programming problem (MINLP). The problem cannot be solved by GAMS as specified, because the binary variables $\delta_{j(i)j'}$ appear in the objective function in a non-linear way. This difficulty is overcome by introducing a continuous variable $\delta^{c}_{j(i)j'}$ into the objective function, and by imposing the constraint $\delta^{c}_{j(i)j'} = \delta_{j(i)j'}$ to hold. By rewriting the optimization program in this way, the continuous variables $\delta^{c}_{j(i)j'}$ and ε_{kn} appear now non-linearly while binary variables $\delta_{j(i)j'}$ appear linearly in the constraints. This program is solved using the DICOPT solver.

However, we face a second computational difficulty. The algorithm underlying DICOPT starts by solving the NonLinear Programming problem (NLP) where the binary constraint imposed to $\delta_{j(i)j'}$ is relaxed. If the solution to this problem does not lead to a 0-1 solution, DICOPT continues with an alternative sequence of NLP program and Mixed-Integer Linear Program (MIP). The NLP programs are solved for the fixed 0-1 variable, that is first predicted by the MIP program. The sequence of MIP and NLP programs stop when having reach a criterion to be specified by the user[14]. Notice that, although this algorithm has provisions to handle non-convexities, it does not necessarily converge to the global optimum. In this case, the starting point of the algorithm may be of some importance. A natural starting point for $\delta_{j(i)j'}$ is 1 for the most likely state according to the Markov metrics and 0 otherwise. But, as the first step of the algorithm consists in solving a relaxed NLP, the starting point for the first MIP will be far away from this natural prior. In order to skip this difficulty, we artificially constraint $\delta^{c}_{j(i)j'}$ to be a binary variable using a penalty function associate to $\delta^{c}_{j(i)j'} \cdot (1-\delta^{c}_{j(i)j'})$ in the objective function. By using a high enough weight, the continuous variable $\hat{\delta}^{c}_{j(i)j'}$ estimated by the first NLP is close to a binary value. This

[14] See the GAMS/DICOPT Solver Manual available on http://www.gams.com/docs/document.htm.

algorithm is more likely to converge to the global optimum. Denoting the positive weight associated with the penalty as W, the objective is now to find the set $\{\delta^c, \delta, \varepsilon\}$ that minimizes:

$$-(1-\lambda)\cdot\sum_{i=1}^{25}\sum_{j'=1}^{36}\delta^c_{j(i)j}(t)\cdot\log\left(\delta^c_{j(i)j}(t)\big/\hat{T}_{j(i)j'}\right)-\lambda\cdot\sum_{k=1}^{6}\sum_{n=1}^{N}\varepsilon_{kn}(t)\cdot\log\left(\varepsilon_{kn}(t)\big/\varepsilon^p_{kn}(t)\right)+W\cdot\delta^c_{j(i)j}(1-\delta^c_{j(i)j})\quad(25)$$

subject to equations (15)-(19).

b- Defining the priors

In the following section there are two different assumptions on how the priors are formed, namely we specify priors according to an open-loop (OL) or a closed-loop approach (CL). The open-loop (OL) specification has no feed back and does not update the state variable values at each stage, whereas the closed-loop (CL) specification uses the most recent information on the crop state. Accordingly, for the CL predictions, we assume that we are able to observe crop choices of previous years. Then a natural way of choosing priors is to use Markov transitions corresponding to the realized, and observed, states. For the OL approach, we assume that we are only able to observe the crop choices in the initial years , 1986 and 1987. As in the CL approach, a natural prior for year 1988 is given by $T_{j(i)j'}$. For year 1989 and 1990, it also seems to be natural to use $T_{\hat{j}(i)j'}$ as prior where $\hat{j}(i)$ is the predicted state from the previous year. Hence in the CL approach, priors are based on observed previous states whereas in the OL approach we only use predicted states. It follows that prediction errors in the OL approach accumulate from one period to another.

c- Results

Table 6 presents estimates[15] of Markov states for the field sub-sample, using the open-loop and closed-loop approaches. The results in Table 6 suggest several comments. Note that the superscript ** indicates that the model correctly predicts the state variable (a sequence of two crops), while a superscript of * indicates that the model correctly predicts the current crop, but not the preceding crop. The underline shows the cases when the predicted state differs from the most likely state based on the priors.

[15] Notice that with our notations the first number of the state index gives the crop produced on the considered field.

Let us initially focus on the closed-loop approach, (CL). First, in several cases the aggregate data-constraint results in changing the optimal crop allocation ~~changing~~ from the state indicated by the expected prior values. In 1988 five of the eighteen correctly predicted fields are predicted to produce a crop that is not the one with the highest prior. In 1989 the number of correct predictions that differed from the priors was eight out of eighteen, and in 1990 the equivalent number of fields was two out of fifteen correct predictions. The implication of these results is that the aggregate data-constraint adds substantial information to the estimates, but the final field-level predictions use information from both the prior predictions and the aggregate data-constraint. By examining all the cases where the prediction differed from the priors we see that the changes from the priors seem to be consistent with observed data. In 1988, five of the eight changes resulted in a correct prediction of the observed state. In 1989 and 1990, the equivalent ratios of correct predictions are respectively 8 out of 9 and 2 out of 6.

The precision of the field-level prediction is relatively good. In 1988, 64% of fields are correctly predicted. This percentage goes up to 72% in 1989 and down to 60% in 1990. These results show that the micro behavior inferred from the aggregate data seems consistent with observed data. Of course, since we are working on a 25-observation sub-sample, these results cannot be considered a full validation of our approach. Nevertheless, they offer some interesting examples of how it could work in practice.

Now consider the results using an open-loop approach (OL). As expected, the goodness of fit deteriorates after the first period using OL approach. In 1989, six predicted states correspond to the observed states compared with 18 out of 25 with the ~~closed loop~~closed-loop estimator. In 1990, the predicted state corresponds to the observed state for only five fields. However, the model behaves better in terms of crop predictions at field-level. In 1989, 48% of fields are predicted to produce the observed crop. The proportion goes up in 1990 and is equal to 52%.

Table 6: Results of the complete data disaggregation method.

Field	**Observed crop sequence**	**Predicted state 1988**		**Predicted state 1989**		**Predicted state 1990**	
	86-87-88-89-90	**CL**	**OL**	**CL**	**OL**	**CL**	**OL**
1	A A A A G	AA**	AA**	AA**	CA	CA	CC
2	G G A A A	CG	CG	CA	CC	CA	CC
3	A A A G A	AA**	AA**	AA	CA	CG	CC
4	C C C F C	TC	TC	FC**	FT*	TF	TF
5	G G C C C	CG**	CG**	GC	CC**	CC**	CC**
6	C C C C C	CC**	CC**	CC**	TC	CC**	CT*
7	C C C C C	CC**	CC**	CC**	CC**	CC**	CC**
8	F F C F C	CF**	CF**	CC	FC**	CF**	CC*
9	G G M C T	CG	CG	CM**	CC*	GC	CC*

10	C C T C T	CC	CC	CT**	CC*	AC	CC
11	C C C T C	CC**	CC**	CC	CC	CT**	CC*
12	C C F C C	CC	CC	FF	AC	CC**	CA*
13	M M C G C	TM	TM	GC**	GT*	CG**	TG
14	A A A A A	AA**	AA**	AA**	CA	AA**	CC
15	A A A A A	AA**	AA**	AA**	CA	AA**	CC
16	C C C C C	CC**	CC**	CC**	TC	CC**	CT*
17	C C T F T	CC	CC	FT**	FC*	CF	AF
18	F C C C C	CC**	CC**	CC**	CC**	CC**	CC**
19	F C C C C	CC**	CC**	CC**	CC**	CC**	CC**
20	C C T C C	CC	CC	CT**	AC	TC	CA*
21	C C C T C	CC**	CC**	TC**	FC	CT**	CF*
22	C C C C C	CC**	CC**	CC**	CC**	CC**	CC**
23	F F C F C	CF**	CF**	CC	CC	CF**	CC*
24	A A C F T	FA	FA	FC**	FF*	CF	AF
25	F F F C T	FF**	FF**	CF**	CF**	CC	CC

Notes: OL for open-loop predicted Markov state. CL for closed-loop predicted Markov state. ** for a correctly predicted Markov state. * for a correctly predicted crop. Underlined for predicted state that does not correspond to the state that is most likely based on the expected priors.

There are two ways to improve the quality of the OL estimates, both require more precision in the Markov process since the OL estimates place greater reliance on the predicted values from this process. The first is to add additional years to the data set used to estimate the Markov metrics, which would allow the order of the Markov process to be increased. The second improvement would be to estimate a non-stationary Markov process. Such a model would allow us to take into account temporal shocks. Data availability prevents us from doing either of these improvements in this paper.

Table 7: Observed aggregate crop areas versus predicted (ac)

	1988			1989			1990		
	Obs.	**CL**	**OL**	**Obs.**	**CL**	**OL**	**Obs.**	**CL**	**OL**
Alfalfa	364.3	343.5	343.5	338	343.5	310.2	330.4	353.1	324.6
Cotton	1334.2	1305.6	1305.6	867.6	888	858.5	0	0	0
Field	217.6	218.9	218.9	727.5	763.8	764.7	1487.6	1516.7	1543.4
Grain	0	0	0	203.4	177.2	177.2	0	0	0
Melons	20.6	0	0	0	0	0	33.9	20.6	0
Tomatoes	392.2	460.9	460.9	192.4	156.4	218.6	477	438.4	460.9

Notes: OL for open-loop predicted Markov state. CL for closed-loop predicted Markov state.

Finally, Table 7 shows the aggregate land use predictions using the CL and OL approaches. One interesting point to be noticed is that despite the fact that these are within-sample land use predictions, the model is able to correctly predict states that are not present in the initial states. For example, despite the absence of any field producing *Grains* in 1988 sub-sample, the model allocates grain to two fields in 1989 (in one case correctly predicting the exact field) and reverts, correctly, to no grain fields in 1990. Again, this illustrates that the final field predictions are based on a combination of prior and aggregate sample information. In this example, the aggregate sample information dominates the prior probabilities for two of the twenty five fields.

IV. Conclusion

In this paper, we have addressed the issue of dynamic data disaggregation in agricultural production. The problem of data disaggregation arises each time economists are restricted to using aggregate data but are interested in studying microeconomic behavior. Disaggregation is also important when facing different types of data at different scales. Disaggregating some data sources and using the rest of the data at a more disaggregate level may preserve more information than an aggregated data set.

Data disaggregation allows one of the most significant obstacles to progress in agricultural production, namely the lack of *better and more detailed* data to be partially bypassed, Just and Pope (1999-b). Aggregate agricultural production data are now available in most countries[16]. Such data can be disaggregated using our procedure. This is especially interesting as periodic site-specific data (soil-surveys, GIS data, satellite images) are becoming increasingly available. Disaggregating economic data allows production economists to work at the most disaggregated level. This facilitates the interaction of biophysical models, defined at this scale, with economic models. In addition, the ME approach is flexible enough to take into account out-of-sample information. Specific out-of-sample information may be added to the data disaggregation program via additional constraints. Out-of-the sample information on transition probabilities may be added to the model via specification of priors.

We have developed a data-consistent way of estimating cropping choices by farmers at a disaggregate level (field-level) using more aggregate data (regional-level). Our data disaggregation procedure requires two steps. The first step consists in specifying a model of crop allocation and estimating it using aggregate

---

[16] In the U.S., aggregate data may be found in the annual publication of the U.S. Department of Agriculture, *Agricultural Statistics*. There also are available in the *Census of Agriculture* published every 5 years. Most of country-level aggregate data are compiled by the Food and Administration Organization (FAO) and are easily available.

data. In the second step, we disaggregate outcomes of the aggregate model using maximum of entropy (ME). Three points should be noticed:

- First, we explicitly model aggregate cropping pattern choices as a dynamic process by using a Markov process. We believe that farmer's crop choices are dynamic *per se*.
- Second, we disaggregate farmer's behavior to the field-level which is, we believe, the most interesting and precise level to represent farmer's crop choices.
- Third, we use a ME approach for disaggregating the data. It selects the optimal solution based on the Kullback-Leibler cross-entropy criterion in cases where traditional estimation methods would result in an unidentified model.

We have applied our data disaggregation procedure to a sample of 190 fields located in California and observed from 1986 to 1990. The sample includes six annual crops, namely: Alfalfa, Cotton, Field, Grain, Melons and Tomatoes. A second-order Markov process is specified to represent aggregate crop choices. Therefore field-level crop predictions are limited to years 1988, 89 and 90. We have shown that the quality of predictions at field-level is relatively good. In 1988, 64% of fields are predicted to produce the crop that is observed. This percentage is 72% in 1989 and 60% in 1990. These results mean that the micro behavior, inferred from aggregate data with our data disaggregation approach, seems to be consistent with observed behavior. In addition, we show that the resulting disaggregated data are consistent with the priors given by the Markov metrics, and with the data represented by the aggregate land use shares.

V. References

Antle, J., and S. Capalbo. 'Econometric-Process Models for Integrated Assessment of Agricultural Production Systems', *American Journal of Agricultural Economics*, 83(2), May, 2001: 389-401.

Antle, J., S. Capalbo and S. Mooney. 'Optimal Spatial Scale for Evaluating Economic and Environmental Tradeoffs', Paper presented at AAEA Annual Meeting, Nashville, August 8-11, 1999.

Barker, T. and M.H. Pesaran, 'Disaggregation in Econometric Modeling – An Introduction.' in Disaggregation in Econometric Modeling. T. Barker and M.H. Pesaran editors. Chapman and Hall, Inc. London. (1990).

Burton, D. M. 'An Astructural Analysis of National Forest Policy and Employment', *American Journal of Agricultural Economics*, 79(3), August, 1997: 964-974.

El-Nazer, T. and B. Mac Carl. 'Choice of Crop Rotation: A Modeling Approach and Case Study', *American Journal of Agricultural Economics,* 68(1), February, 1986: 127-136.

Goetz, R. 'Diversification in Agriculture Production: A Dynamic Model of Optimal Cropping to Manage Soil Erosion' *American Journal of Agricultural Economics*, 79(2), May, 1997: 341-356.

Golan, A., G. G. Judge, and D. Miller. Maximum Entropy Econometrics: Robust Estimation with Limited Data, New York: John Wiley & Sons, 307 pages. (1996)

Hardie, I.W., and P.J. Parks. 'Land Use with Heterogeneous Land Quality: An Application of an Area Base Model.' *American Journal of Agricultural Economics*, 79(May 1997): 153-161.

Heady, E.O. 'Simplified Presentation and Logical Aspects of Linear Programming Technique'. *Journal of Farm Economics*, **24**, 1954: 1035-1048.

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

Howitt, R.E. 'Calibration methods for Agricultural Economic Production Models.' *Journal of Agricultural Economics*, 46(2), 1995-b: 147-159.

Howitt, R.E., and A. Reynaud. **'**Spatial Disaggregation of Agricultural Production Data by Maximum of Entropy'. Mimeo University of California at Davis. (2001).

Just, R., and R. Pope., 'Implications of Heterogeneity for Theory and Practice in Production Economics', *American Journal of Agricultural Economics*, 81(1), February, 1999: 711-718.

Just, R., and R. Pope. 'The Agricultural Producer: Theory and Statistical Measurement' in *Handbook of Agricultural Economics*. B. Gardner and C. Rausser Eds. (1999-b). *Forthcoming.*

Kijima , M. *Markov processes for stochastic modeling.* London : New York : Chapman & Hall, 1997.

King, G. ‘*A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*’, Princeton: Princeton University Press, 1997.

King, R.A. ‘Some Applications of Activity Analysis in Agricultural Economics’, *Journal of Farm Economics*, 25, 1953: 823-833.

Lee, T.C., G. G. Judge, and A. Zellner (1970). *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*. North-Holland Publishing Co., 254 pages. Amsterdam.

Mundlak, Y.,. ‘Production and Supply.’ in *Handbook of Agricultural Economics*. B. Gardner and C. Rausser Eds. (1999-b). *Forthcoming.*

Miller, D., and A. Plantinga. ‘Modeling Land Use Decision with Aggregate Data’, *American Journal of Agricultural Economics*, 81(1), February, 1999: 180-194.

Nerlove, M., and D. Bessler. ‘Expectations, Information and Dynamics.’ in *Handbook of Agricultural Economics*. B. Gardner and C. Rausser Eds. (1999-b). *Forthcoming.*

Plantinga, A. ‘The Effect of Agricultural Policies on Land Use and Environmental Quality.’ *American Journal of Agricultural Economics*, 78(November 1996): 1033-1047.

Prato, T., C. Fulcher, S. Wu, and J. Ma. ‘Multiple-Objective Decision Making for Agroecosystem Management.’ *Agricultural and Resource Economics Review*, **25**, October, 1996: 200-212.

Stocker, T.M., ‘Empirical Approaches to the Problem of Aggregation over Individuals’, *Journal of Economic Literature*, 31(4), December, 1993: 1827-1874.

Von Storch, H., E. Zorita, and U. Cubash, ‘Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime’. *Journal of Climate*, **6**, 1993, 1161-1171.

Von Storch, H., B. Hewitson and L. Mearns, ‘Review of Empirical Downscaling Techniques’**,** Paper presented at REGCLIM workshop, 8 May - 9 May 2000 at Jevnaker, Norway, GKSS Research Center, Institute for Hydrophysics, Germany

Wu, J., and K. Segerson. ‘The Impact of policies and land Characteristics on Potential Groundwater Pollution in Wisconsin.’ *American Journal of Agricultural Economics*, 77(November 1995): 1033-1047.

Wu, J., and R. Adams. ‘Micro vs. Macro Acreage Response Models: Does Site-Specific Information Matter?’ mimeo. Oregon Sate University.