

## Agricultural and environmental information systems: the integrating role of area samples

L. Ambrosio<sup>1\*</sup>, C. Marín<sup>1</sup>, L. Iglesias<sup>1</sup>, V. Pascual<sup>1</sup>,  
A. Fuertes<sup>2</sup> and M. A. Mena<sup>2</sup>

<sup>1</sup> *Departamento de Economía y Ciencias Sociales Agrarias. Escuela Técnica Superior de Ingenieros Agrónomos. Universidad Politécnica de Madrid. Avda. Complutense, s/n. 28040 Madrid. Spain*

<sup>2</sup> *Ministerio de Medio Ambiente y Medio Rural y Marino. Paseo Infanta Isabel, 1. 28014 Madrid. Spain*

---

### Abstract

This article aims to be a contribution to the improvement of agricultural and environmental official statistics. Methods are applied to integrate information from state agency registers regarding crop area with ground data observed in random area samples. To improve the precision of crop area estimates in small areas (municipalities), methods using ground survey and remote sensing are applied. To improve temporal resolution of crop area estimators, methods based on time series analysis are applied. Agro-meteorological models are applied to improve crop yields statistics. A method is shown whereby crop rotation models may be a useful tool to forecast changes in the dynamics of the use of natural resources (soil, water and air) by agriculture and to foresee their environmental impact. Finally, a method to update and disaggregate information from territorial censuses on land uses is applied. These methods and models are illustrated in the framework of an information system belonging to the Spanish Ministry of the Environment and Rural and Marine Affairs. The relative improvement offered by each method is assessed by evaluating the precision gain of the proposed crop area estimates versus those currently used by the aforementioned information system.

**Additional key words:** agrometeorological models, crop rotation models, sampling in time, small area estimation, spatiotemporal econometric models, updating regional censuses and disaggregation of information.

### Resumen

#### Sistemas de información agraria y medioambiental: el papel integrador de las muestras de áreas

Este artículo pretende ser una contribución a la mejora de las estadísticas oficiales agrarias y medioambientales. Se aplican métodos para mejorar la precisión de las estimaciones anuales de la superficie de cultivos, integrando la información relativa a la superficie de cultivos registrada en los registros administrativos con los datos de campo observados en una muestra aleatoria de áreas. Para mejorar la precisión de las estimaciones de la superficie de cultivos en áreas pequeñas (municipios), se aplican métodos que integran datos de teledetección con datos de campo observados en muestras de áreas. Para mejorar la resolución temporal de los estimadores de la superficie de cultivos se aplican métodos de análisis de series temporales a los datos observados en muestras de áreas. Para mejorar las estadísticas sobre el rendimiento de las cosechas se aplican modelos agrometeorológicos integrados en muestras de áreas. Se muestra cómo los modelos de rotación de cultivos basados en muestras de áreas pueden ser una herramienta útil para anticiparse a cambios en la dinámica del uso de los recursos naturales (suelo, agua y aire) por parte de la agricultura y prever su impacto medioambiental. Finalmente, se aplica un método para actualizar y desagregar la información de censos territoriales sobre los usos del suelo, a partir de modelos de rotación de cultivos y de muestras de áreas. Estos métodos y modelos se ilustran en el marco de un sistema de información perteneciente al Ministerio de Medio Ambiente, Medio Rural y Marino. Los nuevos métodos mejoran la precisión de los estimadores de la superficie de cultivos y la ganancia de precisión ofrecida por cada método se evalúa en el marco del citado sistema.

**Palabras clave adicionales:** actualización de censos territoriales y desagregación de su información, estimación en áreas pequeñas, modelos agrometeorológicos, modelos de rotación de cultivos, modelos econométricos espaciotemporales, muestreo repetido.

---

\* Corresponding author: [luis.ambrosio@upm.es](mailto:luis.ambrosio@upm.es)

Received: 09-12-08; Accepted: 21-10-09.

## Introduction

Information is key for decision-making, as it allows us to transform uncertainty into risks and to quantify and reduce risks. The greater the quantity and quality of information (precision, opportunity and specificity), the lower the uncertainty and the better the risk conditions in which economic and social actors reach their decisions.

Production of information is a complex task that economic and social actors are not in a position to perform individually as this requires not only data on quality, but also knowledge to select the most relevant data in each applicable case and interpret this data. This is a collective task where public authorities are responsible for data production and private bodies then interpret data in the specific context where the decision is reached, with the help of economists and other social scientists. This traditional pattern of division of reporting tasks is not exact (Wolf *et al.*, 2001) because individual economic actors cooperate with the authorities by supplying data and the authorities interpret the data they produce in a context that is increasingly more global and less specific.

In agriculture, the nature of the information required is very diverse, given the heterogeneity of the agents operating in the sector: farmers, input providers, agro-food industries, marketing agents, politicians and state agencies, among others. Farmers associate risk to factors outside their control which have a large impact on their production costs and the economic results of their concerns, including: weather, perceived and paid prices, decisions taken by competitors (Kuhlmann and Brodersen, 2001) and environmental restrictions (Burt, 2003). They require information on these factors that is concrete and precise to allow them to detect deviations from their production plan versus targets and to implement the necessary strategy changes to turn the situation around. The systems known as Decision Supporting Systems, however, have not been widely used by farmers, perhaps because they prefer to use tactics to overcome obstacles caused by factors outside their control, instead of production strategies that may be optimal

but rarely lead to overcoming these obstacles (McCown, 2002). Research and development of new production techniques is also required. Food industries and marketing agents require information on crops and input suppliers require information on consumption of production means.

Society, through its political representatives, demands farming systems that provide food safety and bio-energy, without irreversibly exhausting or polluting natural resources (soil, water and air) and respecting biodiversity. All of the above in a context of climate change. Information required on these systems is basically limited to large macro-economic results (internal offer and demand and the outlook of international markets), while the interaction between the economy and the environment has been relegated to the background. Today, this interaction has moved to the forefront and so has the demand for information to monitor the impact of agricultural systems on the environment, control of waste produced and the design of sustainable agricultural systems which remain productive over time.

This article applies several statistical methods to produce timely, precise, specific and high-resolution spatiotemporal information, in the framework of an information system. The methods applied are illustrated using the Farming and Environmental Information System belonging to the Spanish Ministry of the Environment and Rural and Marine Affairs. This system comprises elements of diverse origins that were designed with a variety of very different specific purposes and this article provides methods to adapt them to the current needs for information. The focus of attention is the bridging role that random area samples can play to transfer the advantages from one element of the system to another, and remedy any deficiencies. This usefulness of random samples is illustrated using the Survey on Surfaces and Crop Yields in Spain (ESYRCE), a key element within the Spanish system.

Surveys, state agency registers and censuses are the basic elements encompassed in an information system. In the Spanish system, we can highlight ESYRCE, a territorial survey originally established to cater for European Union (EU) information demands on member state harvests. In Spain, this information has been

---

Abbreviations used: BLUE (best linear unbiased estimator), BLUP (best linear unbiased predictor), CGMS (crop growth monitoring system), DE (direct estimator), ESYRCE (encuesta sobre superficies y rendimientos de cultivos en España = survey on surfaces and crop yields in Spain), EU (European Union), OLIWIN (agro-meteorological models for the estimation at harvest of olive and vine yield), SAS (statistical analysis software), SAS/IML (interactive matrix language of SAS), SIGA (sistema de información geográfico agrario = geographical information system on farming data), SIGPAC (sistema de información geográfica de parcelas agrícolas = geographic information system to identify agricultural plots), SIMWAT (simulation water balance), UTM (universal transverse mercator), WOFOST (world food studies).

gathered so far by means of subjective processes, on the basis of local expert crop area estimates. Although regulations in the EU allow data to be obtained by subjective procedures, they recommend the preferential use of objective methods. Data obtained by objective methods is directly convertible between EU member states, and this involves a substantial advantage over subjective data.

Given this recommendation from the EU, a random sample of areas was designed (ESYRCE). The sample was selected on the basis of the National Topography Map 1:50,000, using as a sampling unit (segment) an area of  $700 \times 700$  m, inscribed on a  $1 \times 1$  km square of the Universal Transverse Mercator (UTM) projection. This is a systematic sample and it covers the entire national territory, supplying high-resolution spatial data on soil uses and crop yields. Its results are objective and directly comparable and standardised among Spanish Autonomous Communities, EU member states and third countries (Ambrosio and Gallego, 1998).

State agency registers are an important source of data but, however, this data is also subject to unknown measurement errors that must be combined with data from random samples for their validation. This article applies an integration method for data from state agency registers and data from random samples, resulting in integrated data which validates data from this register and is more precise than the data it encompasses. The method is illustrated integrating data from the Spanish system registries with other data from ESYRCE.

In the framework of official information systems, samples are generally designed to obtain precise annual details on a national level and for large regions (Autonomous Communities), meaning that data from the level of small regions, provinces and municipalities do not have the level of precision required for well-informed decisions to be taken. The required level of temporal resolution is also absent: this is the case of ESYRCE, whose results are obtained after the end of the campaign and are frequently not available at the right time to make decisions. This article applies a method to improve the precision of crop area estimates in small areas and another to provide crop area forecasts, by combining field data with auxiliary information from other sources. Both methods are illustrated using ESYRCE.

In order to prevent an undesired environmental impact, it is especially relevant to forecast individual land use decisions made by farmers and to assess their environmental impact. This article applies a method to forecast the alternation and rotation of crops using

field data observed in area samples. Alternation of crops, combined with auxiliary information on doses of water, fertiliser, herbicides, pesticides, and greenhouse gas emissions allows the assessment of the impact of agricultural systems on the environment. In addition, crop rotation models are useful tools for updating territorial censuses. Territorial censuses, such as the Geographic Information System to identify Agricultural Plots (SIGPAC) or the Geographical Information System on Farming Data (SIGA), provide spatial resolution to the information system but keeping them up-to-date is costly. The way in which rotation models can be used to update these censuses is shown in this article.

Agro-meteorological models are a useful tool for carrying out forecasts on the crop yields depending on weather conditions and this article applies this tool. Area frames are useful to select farmer samples to interview, to gather information that is not directly observable on the ground, and this article applies a method for areas with extensive agriculture, analogous to that used for intensive farming areas.

A computer programme written for the IML procedure of the SAS statistical package has been specifically created for each one of the methods applied in the case of ESYRCE.

## Integration of data from different sources

There is frequently data on the same issue from a variety of sources and this article applies an integration method for the data series registered by Spanish agencies and the data series estimated by ESYRCE. The method is illustrated by applying it to the estimated cultivated surface. To achieve data integration, data from each source must have a statistically significant relationship with the real surface. When the source is a random sample, then the data have a statistically significant relationship with the true surface and it is possible to estimate errors on the basis of the sample (Ambrosio *et al.*, 2003). However, when the source is a state agency register, statistical models are required to estimate measurement errors.

## The model

Let  $y_t$  be the (unknown) surface of a certain crop in year  $t$ , in a given region (Autonomous Community).

Two series of crop area estimates for  $\{y_t; t=1,2,\dots,T\}$  are available. One  $\{\hat{y}_{Et}; t=1,2,\dots,T\}$  is the Direct Estimator (DE) used in ESYRCE, which is based solely on field data observed on a random area sample:

$$\hat{y}_{Et} = \frac{100}{49} \frac{M}{m} \sum_{i=1}^m y_{it}$$

where  $M$  is the number of systematic samples in the population,  $m$  is the number of systematic samples selected and  $y_{it}$  is the surface of the crop in the  $i^{th}$  systematic sample in year  $t$ . And the other  $\{\hat{y}_{At}; t=1,2,\dots,T\}$  is based on a state agency register. The statistical properties of the DE estimator,  $\hat{y}_{Et}$ , are known as these are based on a random sample (Ambrosio *et al.*, 2003). However, the properties of  $\hat{y}_{At}$  are unknown and must be assessed using statistical models. To this end, the model:

$$\hat{y}_{At} = \beta_0 + \beta_1 y_t + e_t \tag{1}$$

$$\hat{y}_{Et} = y_t + u_t \tag{2}$$

is considered, where Eq. [1] is a model of the register error structure, specified as the sum of a fixed component,  $\beta_0 + \beta_1 y_t$ , and a random component,  $e_t$ . The random component has a zero mean and its variance is  $\sigma_e^2$ . In addition it is also assumed that  $e_t$  is independent from  $y_t$ . Eq. [2] describes the statistical properties of  $\hat{y}_{Et}$ : it is an unbiased estimator of  $y_t$  and the error  $u_t = \hat{y}_{Et} - y_t$  is random with a zero mean and variance  $\sigma_u^2$ .

**Integration**

In order to integrate  $\{\hat{y}_{Et}; t=1,2,\dots,T\}$  and  $\{\hat{y}_{At}; t=1,2,\dots,T\}$  the best linear unbiased estimator (BLUE) of  $y_t$  is proposed, based on [1] and [2]:

$$\hat{y}_t = \hat{y}_{Et} + \frac{\beta_1 \sigma_u^2}{\sigma_e^2 + \beta_1^2 \sigma_u^2} (\hat{y}_{At} - \beta_0 - \beta_1 \hat{y}_{Et})$$

whose variance is:

$$V(\hat{y}_t|y_t) = \sigma_u^2 \frac{\sigma_e^2}{\sigma_e^2 + \beta_1^2 \sigma_u^2} \text{ (Fuller, 1987; pp. 20-22).}$$

Note that the new integrated data,  $\hat{y}_t$ , involves a  $\hat{y}_{Et}$  correction of an amount that is a function of the deviations between the crop area estimates of the random sample and the register,  $v_t = \hat{y}_{At} - \beta_0 - \beta_1 \hat{y}_{Et}$ . The correction

factor,  $g = \frac{\beta_1 \sigma_u^2}{\sigma_e^2 + \beta_1^2 \sigma_u^2}$ , depends on the degree of precision,  $\sigma_u^2$ , of the random estimator. Integrated estimator  $\hat{y}_t$  improves  $\hat{y}_{Et}$  because it is more precise: both are unbiased and the integrated data has a lower

variance,  $V(\hat{y}_t|y_t) \leq \sigma_u^2$ . Estimate  $\hat{y}_{At}$  contributes to improving the precision of  $\hat{y}_{Et}$  in a  $g = 1 - V(\hat{y}_t|y_t) / \sigma_u^2$  proportion.

**Estimation and validation**

The model defined by Eqs. [1] and [2] has been estimated using data concerning crop surfaces registered by Spanish agencies, using ESYRCE as the random area sample (the length of the series is from 1990 to 2005). Model parameters are estimated using the estimators suggested in Fuller (1987; pp. 13-15), and assuming that  $\sigma_u^2$  is known:

$$\hat{\beta}_1 = \frac{m_{\hat{y}_{Et}\hat{y}_{At}}}{m_{\hat{y}_{Et}\hat{y}_{Et}} - \sigma_u^2}, \quad \hat{\beta}_0 = \bar{Y}_E - \hat{\beta}_1 \bar{Y}_A$$

$$\text{and } \hat{\sigma}_e^2 = m_{\hat{y}_{At}\hat{y}_{At}} - \hat{\beta}_1 m_{\hat{y}_{Et}\hat{y}_{At}}$$

$$\text{where } m_{\hat{y}_{Et}\hat{y}_{Et}} = \frac{1}{T-1} \sum_{t=1}^T (\hat{y}_{Et} - \bar{Y}_E)^2$$

$$m_{\hat{y}_{At}\hat{y}_{At}} = \frac{1}{T-1} \sum_{t=1}^T (\hat{y}_{At} - \bar{Y}_A)^2$$

$$m_{\hat{y}_{Et}\hat{y}_{At}} = \frac{1}{T-1} \sum_{t=1}^T (\hat{y}_{Et} - \bar{Y}_E)(\hat{y}_{At} - \bar{Y}_A),$$

$$\bar{Y}_E = \frac{1}{T} \sum_{t=1}^T \hat{y}_{Et} \quad \text{and} \quad \bar{Y}_A = \frac{1}{T} \sum_{t=1}^T \hat{y}_{At}.$$

Integration requires the relationship between data from the sources not to be spurious, that is, the deviations  $v_t = \hat{y}_{At} - \beta_0 - \beta_1 \hat{y}_{Et}$  have to be stable in the long run: in this case it is said that data series generated by both estimators,  $\hat{y}_{Et}$  and  $\hat{y}_{At}$ , are co-integrated. If the series are stationary (so that the mean and the variance do not change over time), then they are co-integrated. The Augmented Dicky-Fuller test has been used to verify the stationarity of the series. If one or both series are not stationary (and they are a random walk), the Phillips-Ouliaris test is used to verify co-integration (Maddala and Kim, 1998).

**Results**

Co-integration of the data series registered by Spanish agencies (the document known as the «Anuario», based on farmers' statements about the acreage of their crops) and the data series of crop area estimates provided by

ESYRCE (using the DE estimator), concerning large crop groups—cereals, legumes, root crops, industrial crops, fodder crops, citrus crops, fruit trees, vines and olive trees—has been studied for the crops and for the Spanish Autonomous Communities where the extension of the series available was enough to allow these tests to be performed.

If the data series of crop areas registered in the administrative registers and the series of crop areas provided by ESYRCE are co-integrated, then the BLUE estimator outperforms the DE estimator currently used in ESYRCE, as a result of integrating both series. Table 1 shows the estimates of the error of these two estimators, together with the gain in the precision of crop area estimates due to the BLUE estimator. This gain is achieved without cost increase, since the required data are already available.

Concerning cereal crops in Castile and Leon, Andalusia and Aragon, the «Anuario» cereal area data series are co-integrated with the series of cereal area estimates provided by ESYRCE. In these Regions, the «Anuario» cereal area data series are statistically validated (they are coherent with ground data observed in the area sample) and their integration with ESYRCE's cereal area series reduces the error of the BLUE estimator with

respect to that of the DE estimator used in ESYRCE, as Table 1 shows. In Castile-La Mancha and Extremadura over half of the «Anuario» data series concerning cereal area are co-integrated with the series of cereal area estimates provided by ESYRCE and for these estimates, the error of the BLUE estimator is reduced in quantities analogous to the aforementioned data. In the remaining Spanish Autonomous Communities (excluding Galicia, Asturias, Cantabria, the Basque Country and the Canary Islands where cereals are not grown) this is not the case and hence the «Anuario» data series on cereal areas are not statistically validated (they are not coherent with ground data observed in the area sample).

Legume harvesting is centred in the Autonomous Communities of Castile and Leon, Castile-La Mancha and Andalusia, meaning that only these three regions have the necessary information available to analyse co-integration of the series. In these regions, the «Anuario» data series on legume area are co-integrated with the series of legume area estimates provided by ESYRCE, and hence they are statistically validated and the integration of the «Anuario» data series with ESYRCE reduces the error of the BLUE estimator, as Table 1 shows.

**Table 1.** Gain of precision of the BLUE crops area estimator with respect to the direct estimator (DE) crops area estimator currently used in ESYRCE

Crop	Spanish Autonomous Communities	Error (%) of the crops area estimator		Gain of BLUE vs DE (%): [1]-[2]
		DE estimator [1] (currently used in ESYRCE)	BLUE estimator [2] (proposed)	
Cereal	Castile and Leon	6.32	3.64	2.68
	Andalusia	13.47	8.76	4.71
	Aragon	6.34	5.48	0.86
Legumes	Castile and Leon	25.55	22.64	2.91
	Castile-La Mancha	29.42	17.34	2.08
	Andalusia	13.15	9.03	4.12
Potatoes	La Rioja	22.30	14.92	7.38
Industrial crops	Castile and Leon	20.44	18.49	1.95
	Andalusia	6.62	5.33	1.29
	Aragon	15.92	15.60	0.32
Lucernes	Castile and Leon	7.70	4.25	3.45
	Aragon	5.76	3.56	2.20
Citrus	Andalusia	16.66	11.61	5.05
Vineyards	Castile-La Mancha	2.96	1.46	1.50
	Valencia	10.42	9.02	1.40
Olive trees	Andalusia	3.07	2.35	0.72
	Castile-La Mancha	2.93	2.58	0.35

With reference to root crops, the «Anuario» root area data series are not co-integrated with the series of root area estimates provided by ESYRCE in any of the Spanish Autonomous Communities. The potatoes crop in La Rioja is the exception and its integration with ESYRCE reduces the error of the BLUE estimator. The majority of the «Anuario» data series concerning industrial crops area are co-integrated with the series of industrial crops area estimates provided by ESYRCE in Castile and Leon and its integration reduces the error of the BLUE estimator (see Table 1). In Andalusia and Aragon, this happens in less than half the cases and for these cases (cotton and sunflower) the error of the BLUE estimator can be reduced as shown in Table 1. In the remaining Autonomous Communities this does not happen in any of the cases and this is why the «Anuario» data series concerning industrial crops area are not statistically validated in those Regions.

Six fodder crops have been considered – maize, lucernes, vetch, turnip, beetroot and cabbage, and there is only enough data on the first three crops to perform an analysis in all Autonomous Communities: In Aragon and Castile and Leon, the «Anuario» lucernes area data series are co-integrated with the series of lucernes area estimates provide by ESYRCE and its integration reduces the error of the BLUE estimator as shown in Table 1. In the remaining Regions, the «Anuario» data series are not co-integrated with the series provided by ESYRCE, so the «Anuario» data series concerning fodder crops area in those Regions are not statistically validated.

As far as citrus crops are concerned, there is a close relationship between both series in Valencia. However, there were notable differences in 1998 and 1999 and these deviations have brought about the lack of co-integration. In Andalusia, the «Anuario» orange trees area data series are co-integrated with the orange trees area estimates provided by ESYRCE and the error of the BLUE estimator can be reduced as is shown in Table 1. In other Regions the series are not co-integrated. Most of the data series registered by Spanish agencies concerning fruit trees area are not co-integrated with the series of fruit tree area estimates provided by ESYRCE: in Valencia there are two exceptions, apricot trees and plum trees, and the integration of both series reduces the error of the BLUE estimator from 11.74% and 20.63% to 8.68% and 17.85%, respectively.

The «Anuario» data series on vineyard areas registered in Castile-La Mancha and Valencia are co-integrated with the series of vineyard area estimates provided by ESYRCE and its integration reduces the error of the

BLUE estimator as is shown in Table 1. Concerning olive trees, the «Anuario» olive tree area data series are co-integrated with those of olive tree area estimates provided by ESYRCE in Andalusia and Castile-La Mancha and its integration reduces the error of the BLUE estimator as shown in Table 1. More details on these results can be found in Ambrosio *et al.* (2007).

## Estimation in small areas

In official information systems, samples are generally designed to provide *direct estimates* (based solely on field data), that are robust (not dependent on models) and precise, in large regions. However, *direct estimates* in small areas (province and/or municipality) based solely on the sample designed for the large area which they are part of are not sufficiently precise in most cases, given that the part of the field sample in each area is usually very small. A further disadvantage is that it is not possible to obtain *direct estimates* from small areas where no field data are available.

Sometimes, it is possible to obtain the data required in provinces and/or municipalities, on the basis of censuses, government agency data or experts' opinions. However, censuses are costly and they are carried out over a lengthy period and is for this reason that they are not updated; meanwhile, government agency data and experts' opinions, although not subject to sampling errors, are however subject to unknown measurement errors and, as a consequence, their estimates tend to be poor (Jiang and Lahiri, 2006a) and they cannot be validated.

Specific estimation methods are required for small areas, the definition of «small area» being an area where *direct appraisal* techniques do not provide estimates with the required degree of precision (Rao, 2003). The literature suggests a number of *direct appraisal* techniques using *auxiliary* information and using statistical methods to combine auxiliary information with field information (Särndal, 1984). Some of these techniques are *model-dependent*, as their estimates have good statistical properties but only if the model is correct (they are not robust). In order to reduce the risks of an incorrect specification of the model, mixed appraisal techniques have been suggested, providing *indirect estimates based on models and consistent according to design* (robust) (You and Rao, 2002; Jiang and Lahiri, 2006b).

This article assesses the techniques used to improve precision of appraisal in small areas, and ESYRCE is used to illustrate their application. The aim is to estimate

the total  $Y_i = \sum_{j=1}^{N_i} y_{ij}$  of the surface of a certain type of crop in one of the municipalities,  $m$ , in a given region. It is considered that the  $i$ -th municipality  $\{i=1,2,\dots,m\}$ , comprises  $N_i$  sampling units (segments). Associated with the  $j$ -th segment  $\{j=1,2,\dots,N_i\}$  of the  $i$ -th municipality we find a set of values  $\{y_{ij}; \underline{x}_{ij}\}$ , where  $y_{ij}$  is the real crop surface (unknown, but fixed) and  $\underline{x}_{ij}$  is a vector of known values of the auxiliary variables. Remote sensing is the most commonly used auxiliary information source.

In the *indirect appraisal* strategy that we will follow, the appraisal problem is limited to a problem surrounding the prediction of the values of  $y_{ij}$  in the segments not included in the field sample (Royall, 1970; Royall and Herson, 1973). When  $n_i$  is the size of the field sample in the  $i$ -th municipality and  $\{(y_{ij}, \underline{x}_{ij}); j=1,2,\dots,n_i\}$  are the data observed in this sample, an estimator of the total would be  $\hat{Y}_i = n_i \bar{y}_i + (N_i - n_i) \hat{Y}_i^*$ , where:

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad \hat{Y}_i^* = \frac{1}{N_i - n_i} \sum_{j=1}^{N_i - n_i} \hat{y}_{ij}^*,$$

and  $\hat{y}_{ij}^*$  is the prediction of the value of the crop surface in the  $j^*$ -th sampling unit not included in the field sample.

### The model

The  $\hat{y}_{ij}^*$  prediction is obtained from a statistical model linking field data,  $\{y_{ij}; j=1,2,\dots,n; i=1,2,\dots,m\}$ , with auxiliary information,  $\{\underline{x}_{ij}; j=1,2,\dots,N_i; i=1,2,\dots,m\}$ . Two types of (mixed) models have been proposed: a disaggregated model on the level of individual sampling units (Battese *et al.*, 1988) and a further aggregated model on the level of small units (Fay and Herriot, 1979).

In this case, the disaggregated model:

$$y_{ij} = \underline{x}_{ij}^T \underline{\beta} + u_{ij} \tag{3}$$

shall be used.

According to this model it is possible to obtain a measurement,  $Ey_{ij} = \underline{x}_{ij}^T \underline{\beta}$ , of the crop surface in a certain segment,  $\{y_{ij}; j=1,2,\dots,N_i; i=1,2,\dots,m\}$ , depending on the auxiliary information and a parameter vector,  $\underline{\beta}$ . The error of this measurement,  $u_{ij} = y_{ij} - \underline{x}_{ij}^T \underline{\beta}$ , is the effect of numerous causes that are grouped into two large categories,  $u_{ij} = v_i + e_{ij}$ : the effect of some of them is  $v_i$  and they affect all segments of a same municipality equally and, in other cases, there is a variation from one segment to another whose effect is  $e_{ij}$ .

Model [3] is a way of specifying that the measurement errors,  $u_{ij}$ , are correlated. For example, in remote sensing, auxiliary information,  $\underline{x}_{ij}$ , is the refraction of sunlight on the Earth's surface, captured by the sensors. This light is dispersed in such a way that the refraction of a pixel is distributed over several adjacent pixels, with the input of an error,  $v_i$ , that depends on the environmental surroundings —soil and climate— where the measurements are taken and has equal effects on all segments of same surroundings. The specific error component in each segment,  $e_{ij}$ , is attributable to the instrument itself (sensor) and the data processing process from the sensor to the image (Labovitz and Masuoka, 1984; Webster *et al.*, 1989; Haining, 1990). Both error components  $v_i$  and  $e_{ij}$  are regarded as independent random variables with zero means and variances  $\sigma_v^2$  and  $\sigma_e^2$ , respectively. The total error,  $u_{ij} = v_i + e_{ij}$ , with zero mean and variance  $Vu_{ij} = \sigma_v^2 + \sigma_e^2$ , is auto-correlated due to  $v_i$  and the auto-correlation coefficient is  $\rho = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2)$ .

### Estimation

The BLUE of  $y_{ij}^*$  is  $\hat{y}_{ij}^* = \underline{x}_{ij}^{*T} \hat{\underline{\beta}} + g_i (\bar{y}_i - \bar{\underline{x}}_i^T \hat{\underline{\beta}})$ , where  $\hat{\underline{\beta}} = (\underline{X}^T \underline{V}^{-1} \underline{X})^{-1} (\underline{X}^T \underline{V}^{-1} \underline{Y})$ ,  $\underline{Y}$  is a column vector  $(\sum_{i=1}^m n_i \times 1)$  with the  $y_{ij}$  field data,  $\underline{V}^{-1} = \text{diag} (\underline{V}_1^{-1}, \underline{V}_2^{-1}, \dots, \underline{V}_i^{-1}, \dots, \underline{V}_m^{-1})$  is a diagonal matrix by blocks with:

$$\underline{V}_i^{-1} = \frac{1}{\sigma_e^2} \underline{I}_{(n_i)} - \frac{g_i}{n_i \sigma_e^2} \underline{1}_{(n_i)} \underline{1}_{(n_i)}^T,$$

where  $\underline{I}_{(n_i)}$  is the identity matrix of order  $n_i$  and  $\underline{1}_{(n_i)}$  is a column vector  $(n_i \times 1)$  of ones,  $g_i = \sigma_v^2 / [\sigma_v^2 + (\sigma_e^2 / n_i)]$  and

$\bar{\underline{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \underline{x}_{ij}$ . By substituting this predictor in  $\hat{Y}_i = n_i \bar{y}_i +$

$(N_i - n_i) \hat{Y}_i^*$  the BLUE estimator of the total is obtained,  $\hat{Y}_i = N_i \hat{Y}_i(g_i)$ , where  $\hat{Y}_i(g_i) = \bar{\underline{X}}_i^T \hat{\underline{\beta}} + g_i (\bar{y}_i - \bar{\underline{x}}_i^T \hat{\underline{\beta}})$  and

$$\bar{\underline{X}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \underline{x}_{ij}.$$

Several works can be found in the literature (Battese *et al.*, 1988; Ambrosio and Iglesias, 2000) where model [3] is checked using remote sensing data. In particular, it is shown that parameters  $\underline{\beta}$  and  $\sigma_v^2$  are significantly different from zero.

The direct estimator is the mean of field data,

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.$$

## Results

Ambrosio and Iglesias (2000) show that for a case similar to that studied here, the relative efficiency of the *indirect estimator* proposed vis-à-vis the *direct estimator* ranges between 2.08 (when the field sample in the small areas is of 17 segments) and 46.37 (when the field sample in the small area is of solely one segment) from one small area to another. In other terms, the *direct estimator* requires, in the first case, a sample of  $2.08 \times 17 = 35$  segments to reach the same precision as the *indirect estimator* using a sample of 17 segments and the auxiliary remote sensing information. In the second case, the *direct estimator* requires a sample of 46 segments to reach the same precision as the *indirect estimator* using a sample of one segment and the auxiliary remote sensing information.

## Crop area forecasts based on models from the series of annual crop area estimates

Very often, initial official information systems do not reach the level of temporal resolution required, as they do not report the changes occurring over a given campaign and the results are obtained after the termination of the campaign, meaning that they are not available at the right time to take decisions. In order to improve the temporal resolution of these systems, we suggest obtaining forecasts from surveys using models on their evolution and trends and correcting results of these models using field data to be observed on several occasions over the course of the campaign under way, according to the timetable scheduled. The method applied is illustrated with ESYRCE.

The estimation of the total annual surface  $y_t$  of a given crop in a certain region is currently calculated in ESYRCE using solely field data observed in year  $t$ . In this article, to improve ESYRCE precision (though at the expense of its robustness), and produce forecasts at the start of each campaign, before field data become available, we also suggest using an estimator from previous years  $\{\hat{y}_t; t = t-1, t-2, \dots, 1\}$ , with the help of a statistical model to transfer information from one year to another.

### The model

The field data in each sample of the sequence is grouped into «elementary estimates» (Gurney and

Daly, 1965),  $\{\hat{y}_t; t = 1, 2, \dots, T\}$ , using the DE estimator (unbiased) in each period of time. The statistical properties of the series of elementary DE crop area estimates of  $\{y_t; t = 1, 2, \dots, T\}$ , are those specified in model [2] above  $\hat{y}_t = y_t + u_t$ . According to [2], variability of the series of elementary DE estimators is due to two main factors: one is sampling,  $V(\hat{y}_t|y_t) = \sigma_u^2$ , and the other is the natural variability of the series  $\{y_t; t = 1, 2, \dots, T\}$ . The variance (marginal) of  $\{\hat{y}_t; t = 1, 2, \dots, T\}$  is  $V(\hat{y}_t) = V(y_t) + \sigma_u^2$ . Variability due to sampling,  $\sigma_u^2$ , depends on the type of design and what follows is considered known and the same as that of the DE estimator currently used in ESYRCE.

In order to model the natural variability of the series  $\{y_t; t = 1, 2, \dots, T\}$  several models have been proposed in the literature. Patterson (1950) deems that the model

$$y_t - \bar{Y}_t = \rho(y_{t-1} - \bar{Y}_{t-1}) + \varepsilon_t, \text{ where } \bar{Y}_t = y_t/N_t = \frac{1}{N_t} \sum_{i=1}^{N_t} y_{it} \text{ is the}$$

population average,  $y_{it}$  is the crop surface of the crop in question in the  $i$ -th sampling unit and  $\varepsilon_t$  is the perturbation term of the model of zero mean, without correlation with the  $u_t$  or among themselves. In addition, it is supposed that  $\rho$  is known and that  $\sigma_u^2$  is approximately constant over time. Note that no relationship is specified between the  $\bar{Y}_t$  and the  $\bar{Y}_{t'}$  for  $t' = t-1, t-2, \dots, 1$ .

If the  $\bar{Y}_t$  evolve over time, it is possible to improve estimates specifying this relationship with a statistical model (Tam, 1986, 1987). Blight and Scott (1973) suggest adding the equation  $\bar{Y}_t - \mu = \lambda(\bar{Y}_{t-1} - \mu) + \eta_t$  to the Patterson model. The effect of including these models in the estimation process was explored by Scott and Smith (1974), using the results of the theory on signal extraction in the face of stationary noises. A comparison of these approaches can be found in Jones (1979).

Jones (1980), using model [2]  $\hat{y}_t = y_t + u_t$ , shows how different approaches to the work mentioned above can be unified in an approach based on mixed models (with fixed and random coefficients) in the form,  $\hat{y} = X\beta + Zy + u$ , where  $\hat{y}$ ,  $y$  and  $u$  are vectors ( $T \times 1$ ) of components  $\{\hat{y}_t; t = 1, 2, \dots, T\}$ ,  $\{y_t; t = 1, 2, \dots, T\}$  and  $\{u_t; t = 1, 2, \dots, T\}$ , respectively,  $X$  and  $Z$  are known matrices and  $\beta$  is a vector of  $P$  unknown parameters that are fixed (fixed effects).  $u$  and  $y$  (random effects) are random vectors independent from one another,  $Cov(y, u) = Q$ , with averages  $E\bar{u} = \bar{0}$  and  $E\bar{y} = \bar{\mu}$ , and covariances  $Var(\bar{u}) = \bar{R}$  and  $Var(\bar{y}) = \bar{G}$ , respectively. The vector of estimators  $\hat{y}$  is random with expectation  $E\hat{y} = E\bar{y}$  and covariance  $Var \hat{y} = \bar{V} = \bar{G} + \bar{R}$ .

### Estimation

The aim is to estimate  $\underline{y}$  more efficiently than  $\hat{\underline{y}}$  and to predict  $y_{T+1}$ . An optimum estimator for  $\underline{y}$ , which is therefore more efficient than  $\hat{\underline{y}}$ , is the Best Linear Unbiased Predictor (BLUP),  $\hat{\underline{y}}_{(BLUP)} = \underline{G}\underline{Z}^T V^{-1} (\hat{\underline{y}} - \underline{X}\hat{\underline{\beta}})$ , where  $\hat{\underline{\beta}} = (\underline{X}^T V^{-1} \underline{X})^{-1} \underline{X}^T V^{-1} \hat{\underline{y}}$ . Note that  $\hat{\underline{y}}_{(BLUP)}$  makes use of the sample for period  $t$  and the prior estimates  $\{\hat{\underline{y}}_i; i = t-1, t-2, \dots, 1\}$ . Eq. [2], where the vectorial notation is  $\hat{\underline{y}} = \underline{y} + \underline{u}$ , is a particular case of mixed model where it is assumed that series  $\{y_i; i = 1, 2, \dots, T\}$  is stationary as far as the mean (constant) is concerned, that is, there are no fixed effects and  $\underline{\beta} = \underline{0}$ , and  $\underline{Z}$  is the identity matrix. In this specific case,  $\hat{\underline{y}}_{(BLUP)}$  is reduced to  $\hat{\underline{y}}_{(BLUP)} = \hat{\underline{y}} - \underline{R}V^{-1} \hat{\underline{y}}$  and its covariance to  $Var \hat{\underline{y}}_{(BLUP)} = \underline{R} - \underline{R}(\underline{R} + \underline{G})^{-1} \underline{R}$ , where  $\underline{R}V^{-1} \hat{\underline{y}}$  is the BLUP of  $\underline{u}$  given  $\hat{\underline{y}}$ . The estimator  $\hat{\underline{y}}_{(BLUP)}$  requires the specification of the structure of correlations of the series of values to estimate,  $\{y_i; i = 1, 2, \dots, T\}$ , and the series of errors of the ESYRCE estimation,  $\{u_i; i = 1, 2, \dots, T\}$ , that is, the identification of  $Var(\underline{y}) = \underline{G}$  and  $Var(\underline{u}) = \underline{R}$ .

In order to identify the structure of correlations of  $\{y_i; i = 1, 2, \dots, T\}$ , level 1 (stationary) auto-regressive processes and random walks (non-stationary) are considered. The model  $y_t = \rho y_{t-1} + \varepsilon_t$  with  $|\rho| < 1$  and  $E\varepsilon_t = 0$ ;  $Cov(\varepsilon_t, \varepsilon_{t'}) = \sigma_\varepsilon^2$ , if  $t = t'$  and  $Cov(\varepsilon_t, \varepsilon_{t'}) = 0$  in other cases, defines an auto-regressive process of level 1,

$AR(1)$ , with the form  $y_t = \sum_{s=0}^{\infty} \rho^s \varepsilon_{t-s}$ , whose function is

the auto-covariance,  $C(s) = \sigma_y^2 \rho^s$ , allowing calculation of covariances,  $Cov[y_t, y_{t'}]$ , depending on the interval  $t' = t - s$ , for  $s = 0, 1, 2, \dots, \infty$  and of  $\sigma_y^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$ . The

auto-correlation function is  $\rho(s) = \rho^s$ , as  $C(0) = \sigma_y^2$ . The model  $y_t = y_{t-1} + \varepsilon_t$  defines a random walk. The matrix of variances and covariances of a random walk is  $Var \underline{y} = \sigma_\varepsilon^2 \{ \min(t, t') \}$ .

Note that in a random walk, variance,  $Vy_t = \sigma_\varepsilon^2 t$ , grows with  $t$ , while in the level one auto-regressive variance (stationary) it converges to a constant when  $t$  increases,  $Vy_t \rightarrow \frac{\sigma_\varepsilon^2}{1 - \rho^2}$ . However, the random walk may be transformed in a stationary process by differentiating it once,  $\Delta y_t = y_t - y_{t-1} = \varepsilon_t$ , in such a way that a zero mean process  $E\Delta y_t = E\varepsilon_t = 0$  results, with finite variance,  $V\Delta y_t = \sigma_\varepsilon^2$ . Given that series  $\{\Delta y_i; i = 1, 2, \dots, T\}$  is stationary with a structure of correlations  $Var \Delta \underline{y} = \underline{\Delta} Var \underline{y} \Delta^T$ , where  $\underline{\Delta}$  is the matrix  $(T-1) \times T$  of operators with first differences, whose  $t$ -th generic row only

includes null elements except those in positions  $t-1$  and  $t$  with the values  $-1$  and  $+1$ , respectively.

The structure of correlations of the series of estimation errors of the area sample,  $\{u_i; i = 1, 2, \dots, T\}$ , is

$$Cov(u_i, u_{i+\zeta}) = Cov(\hat{y}_i, \hat{y}_{i+\zeta}) = (1 - f_0) \frac{S_{0\zeta}^2}{n_0}; \forall \zeta \neq 0 \text{ where}$$

$$f_0 = \frac{n_0}{N_0} \text{ and } S_{0\zeta}^2 = \sum_{t=1}^T \sum_{i=1}^{N_t} (y_{it} - \bar{Y}_t)(y_{i+\zeta t} - \bar{Y}_{t+\zeta}) / \sum_{t=1}^T (N_t - 1).$$

The matrix  $\underline{R}$  is estimated deeming  $S_{0\zeta}^2$  in the usual way, given the design of the simple. The estimation of  $\underline{G}$ , depends on the correlation structure of  $\{y_i; i = 1, 2, \dots, T\}$ , estimated on the basis of estimation series  $\{\hat{y}_i; i = 1, 2, \dots, T\}$ . If the estimation series is stationary (auto-regressive of level 1), then  $\underline{G}$  is estimated

$$\hat{\rho} = \frac{\sum_{t=2}^T \hat{y}_t \hat{y}_{t-1}}{\sum_{t=2}^T \hat{y}_{t-1}^2} \text{ and } \hat{\sigma}_y^2 = \hat{\sigma}_\varepsilon^2 / (1 - \hat{\rho}^2), \text{ where}$$

$$\hat{\sigma}_\varepsilon^2 = [1/(T-1)] \sum_{t=2}^T (\hat{y}_t - \hat{\rho} \hat{y}_{t-1})^2. \text{ If the series of estimates}$$

is a random walk, then the series  $\{\Delta \hat{y}_i; i = 2, \dots, T\}$  is stationary with  $Var \Delta \hat{y}_i = \Delta Var \hat{\underline{y}} \Delta^T$  where  $Var \hat{\underline{y}} = \hat{\sigma}_\varepsilon^2 \{ \min(t, t') \}$  and  $\underline{G}$  are estimated using  $\hat{\sigma}_\varepsilon^2 =$

$$[1/(T-1)] \sum_{t=2}^T (\hat{y}_t - \hat{y}_{t-1})^2. \text{ In the case of a random walk,}$$

BLUP is obtained from the model  $\Delta \hat{\underline{y}} = \Delta \underline{y} + \Delta \underline{u}$  and it is  $\hat{\underline{y}}_{(BLUP)} = \hat{\underline{y}} - \underline{R} \Delta^T [Var(\Delta \hat{\underline{y}})]^{-1} \Delta \hat{\underline{y}}$ , with variance  $Var(\hat{\underline{y}}_{(BLUP)}) = \underline{R} - \underline{R} \Delta^T (Var(\Delta \hat{\underline{y}}))^{-1} \Delta \underline{R}$ , where  $Var(\Delta \hat{\underline{y}}) = \underline{\Delta} Var(\hat{\underline{y}}) \Delta^T = \underline{\Delta} (\underline{G} + \underline{R}) \Delta^T$ .

### Predictions

In order to predict the series of values  $\{y_{T+h}; h = 1, 2, \dots\}$  on the basis of estimated series  $\{\hat{\underline{y}}_{(BLUP)}; t \leq T\}$ , we suggest using the predictor,  $\hat{y}_{T+h} = \underline{a}^T \hat{\underline{y}}_{(BLUP)}$ , where

$$\underline{a} = \left[ Var(\hat{\underline{y}}_{(BLUP)}) \right]^{-1} \left[ \underline{C} + \underline{1} \left( \underline{1}^T \left[ Var(\hat{\underline{y}}_{(BLUP)}) \right]^{-1} \underline{1} \right)^{-1} \right. \\ \left. (1 - \underline{1}^T \left[ Var(\hat{\underline{y}}_{(BLUP)}) \right]^{-1} \underline{C}) \right],$$

$\underline{C} = [C_{1,T+h} \ C_{2,T+h} \dots \ C_{t,T+h} \dots \ C_{T,T+h}]^T$  and  $C_{t,T+h} = Cov(\hat{y}_t, \hat{y}_{T+h})$  are considered auto-covariance functions with the form  $C_{t,t+\tau} = C(\tau)$ . In the case of a random walk, the auto-covariance function, is  $C_{t,t+\tau} = C(t, \tau) = \sigma_\varepsilon^2 (t - \tau) = \sigma_\varepsilon^2 \min(t - \tau); \{t, s = 1, 2, \dots, T\}$ , meaning that  $\underline{C} = \sigma_\varepsilon^2 [C_{1,T+h} \ C_{2,T+h} \dots \ C_{t,T+h} \dots \ C_{T,T+h}]^T = \sigma_\varepsilon^2 [1 \ 2 \dots \ t \dots \ T]^T$ .

**Table 2.** Gain of cereal area BLUP estimator with respect to the DE estimator currently used in ESYRCE

Spanish Autonomous Communities	Error (%) of the cereal area estimators		Gain of BLUP vs DE (%): [1]-[2]
	DE estimator currently used in ESYRCE [1]	BLUP [2]	
Castile and Leon	6.32	4.89	1.43
Castile-La Mancha	6.45	4.91	1.54
Andalusia	13.47	12.34	1.13
Aragon	6.34	5.24	1.10
Extremadura	13.15	10.65	2.50

**Results**

In addition to producing forecasts at the start of each campaign, before field data become available, the error of the DE estimator currently used in ESYRCE can be reduced using this BLUP. Table 2 shows the reductions achieved by the BLUP concerning cereal area estimates.

Reductions of this same order can be achieved for the error concerning legumes area estimates, industrial crops area estimates, olive tree area estimates, vineyard area estimates and orange tree area estimates. In Ambrosio *et al.* (2007), the results from applying this model to the series of ESYRCE estimates for the main crops can be found.

**Crop rotation models**

It is common to use longitudinal or panel samples, where the same sampling unit (segment) is repeatedly observed year after year. This is the case of ESYRCE, supplying information to estimate crop rotation, that is, the distribution of the surface,  $Y_{i,t-1}$ , occupied by a given crop,  $l$ , in campaign  $t - 1$ , between each and every one of the  $J$  crops of the alternative in the subsequent campaign,  $t$ . Knowing this rotation may be useful, among other reasons, to gain forecasts of the crop alternatives of a campaign depending on the previous campaign, to specify econometric spatial models and as a tool for updating regional censuses and disaggregating this information. We will follow an approach that mixes a *model-dependent* approach with a *direct estimation* approach, to get *indirect estimates assisted by models which are consistent according to their design*, that is, robust. *Direct estimation* is based on field data observed in one of the three systematic samples making up ESYRCE.

**The model**

It is assumed that the distribution of the surface  $Y_{i,t-1}$  containing a certain crop  $l$  in campaign  $t - 1$ , between each and every one of the  $J$  crops of the alternative in the following campaign is the multinomial,  $\{MN(Y_{i,t-1}; p_{i,t}); i_t = 1, 2, \dots, n_i; l = 1, 2, \dots, J; t = 1, 2, \dots, \}$ , dependent on the parameters vector,  $p_{i,t} = [p_{i,t1} p_{i,t2} \dots p_{i,tj} \dots p_{i,tJ}]^T$ , where  $\sum_{j=1}^J p_{i,tj} = 1$  and  $p_{i,tj}$  represents the proportion of the crop surface  $l$  in the  $i^{th}$  sample unit during the campaign  $t - 1$ , that becomes crop  $j$  in the following campaign,  $t$ . Parameters,

$$p_{i,tj} = \frac{e^{\sum_{j'=1}^J Y_{i,t-1j'} \beta_{j'j}}}{\sum_{j'=1}^J e^{\sum_{j'=1}^J Y_{i,t-1j'} \beta_{j'j}}} \tag{4}$$

are specified depending on the crop alternatives of the previous campaign,  $\{Y_{i,t-1l}; l = 1, 2, \dots, J\}$ .

**Estimation and validation**

Regression coefficients,  $\{\beta_{j'j}; j' = 1, 2, \dots, J\}$  are considered homogeneous between sampling units and these are estimated over time following an approach of maximum generalised crossed entropy (Golan *et al.*, 1996). Forecasts are calculated for the immediately subsequent campaign,  $t = T + 1$ . The forecast of crop surface  $j$  in the  $i^{th}$  sampling unit of the sample in campaign  $T + 1$  is  $\hat{Y}_{i,T+1j} = \sum_{l=1}^J \hat{p}_{i,T+1lj} \times Y_{i,Tl}$ , where  $\hat{p}_{i,T+1lj}$  is the estimate of  $p_{i,T+1lj}$  resulting from the substitution in [4] of coefficients  $\beta_{j'j}$  for their estimates  $\hat{\beta}_{j'j}$ . The forecast error is  $V\hat{Y}_{i,T+1j} = V\left(\sum_{l=1}^J \hat{p}_{i,T+1lj} \times Y_{i,Tl}\right) = \sum_{l=1}^J Y_{i,Tl}^2 V\hat{p}_{i,T+1lj}$  assu-

ming independence between  $\{\hat{p}_{i_{T+1}lj}; t = 1, 2, \dots, J\}$  where  $Var\hat{p}_{i_{T+1}lj} = [p_{i_{T+1}lj}(1 - p_{i_{T+1}lj})\underline{x}_{i_{T+1}}] Var(\hat{\beta}_{lj}) [p_{i_{T+1}lj}(1 - p_{i_{T+1}lj})\underline{x}_{i_{T+1}}]^T$ , where  $\underline{x}_{i_{T+1}}$  is the  $(1 \times J)$  vector of values of auxiliary variables  $\{Y_{ij}; j' = 1, 2, \dots, J\}$ , where  $\hat{\beta}_{lj}$  is a vector whose components are  $\{\beta_{j'lj}; j' = 1, 2, \dots, J\}$  and  $Var(\hat{\beta}_{lj}) = (X^T V_{lj}^{-1} X)^{-1}$ , where  $V_{lj} = diag\{p_{i_{T+1}lj}(1 - p_{i_{T+1}lj}); i_t = 1, 2, \dots, n_t; t = 1, 2, \dots, T\}$  and  $X$  is a  $(\sum_{t=1}^T n_t) \times J$  matrix whose generic line is  $\{x_{it}; i_t = 1, 2, \dots, n_t; t = 1, 2, \dots, T\}$ .

### Crops area forecasts

We will illustrate use of model [4] to gain forecasts of the surfaces with crops on the Regional level. Obtaining these forecasts from the area sample is proposed, using forecasts  $\{\hat{Y}_{i_{T+1}j}; j = 1, 2, \dots, J; i_{T+1} = 1, 2, \dots, n_{T+1}\}$  instead of values  $Y_{i_{T+1}j}$  that shall be observed at the end of the campaign  $T + 1$  but are not observable in the current time  $T$ . In ESYRCE, the estimator used is  $\hat{Y}_{j,T+1} = \frac{100}{49} \frac{M}{m} \sum_{k=1}^m \hat{Y}_{k_{T+1}j}$ , where  $M$  is the number of systematic samples in the population and  $m$  is the number of systematic samples selected.  $\hat{Y}_{k_{T+1}j} = \sum_{i_{T+1,k}=1}^{M_k} \hat{Y}_{i_{T+1,k}j}$  is the sum of forecasts  $\hat{Y}_{i_{T+1,k}j}$  of crop surface  $j$  in the segments of the  $k^{th}$  systematic sample, comprising  $M_k$  segments.

Estimator  $\hat{Y}_{j,T+1}$  is (approximately) unbiased and its variance is:

$$V\hat{Y}_{j,T+1} = \frac{100}{49} \frac{M}{m} \sum_{i_{T+1}=1}^M V\left(\sum_{l=1}^J \hat{p}_{i_{T+1}lj} \times Y_{i_{T+1}l}\right) + \left(\frac{100}{49}\right)^2 M^2 (1-f) \frac{S_{j,T+1}^2}{m}, \text{ where}$$

$$S_{j,T+1}^2 = \frac{1}{(M-1)} \sum_{k_{T+1}=1}^M (\hat{Y}_{k_{T+1}j} - \bar{Y}_j)^2 \text{ and } \bar{Y}_j = \frac{1}{M} \sum_{k=1}^M \hat{Y}_{k_{T+1}j}.$$

Forecasts shall be corrected over the campaign, estimating model [4] again each time new field data are available over the campaign and obtaining new forecasts  $\{\hat{Y}_{i_{T+1}j}; j = 1, 2, \dots, J; i_{T+1} = 1, 2, \dots, n_{T+1}\}$ , from which new estimates of  $\{\hat{Y}_{j,T+1}; j = 1, 2, \dots, J\}$  and its variation coefficients can be obtained.

At the end of the campaign, direct observation is made of the value of  $Y_{i_{T+1,k}j}$  that was necessary to predict during the campaign using  $\hat{Y}_{i_{T+1,k}j}$  for all  $k$  except one, resulting in  $Y_{k_{T+1}j} = \sum_{i_{T+1,k}=1}^{M_k} Y_{i_{T+1,k}j}$  and the estimator is reduced

to the usual ESYRCE estimator,  $\hat{Y}_{i,T+1} = \frac{100}{49} \frac{M}{m} \sum_{k=1}^m Y_{k_{T+1}j}$  and also its variance.

### Results

The estimation procedure for parameters defined in [4] has been checked using ESYRCE data and the results were satisfactory. Validation has been performed on two levels: one, more disaggregated, is the segment or sampling unit; the other, more aggregated, is the study zone level. In the first case, a comparison has been made with predictions  $\hat{p}_{i_{T+1}lj}$  of the proportion of a certain crop,  $l = 1, 2, \dots, J$ , becoming another crop  $j = 1, 2, \dots, J$  in the rotation of a year  $t - 1$  to the following year  $t$  ( $t = 1, 2, \dots, T$ ) with the data observed,  $f_{i_{T+1}lj}$ , in each unit of the sample,  $i_t = 1, 2, \dots, n_t$ ; the linear correlation coefficient for prediction  $\hat{p}_{i_{T+1}lj}$  with observed data  $f_{i_{T+1}lj}$  is above 0.99 for cotton, olive groves and soft wheat; 0.88 for sunflower and 0.79 for durum wheat. In the second

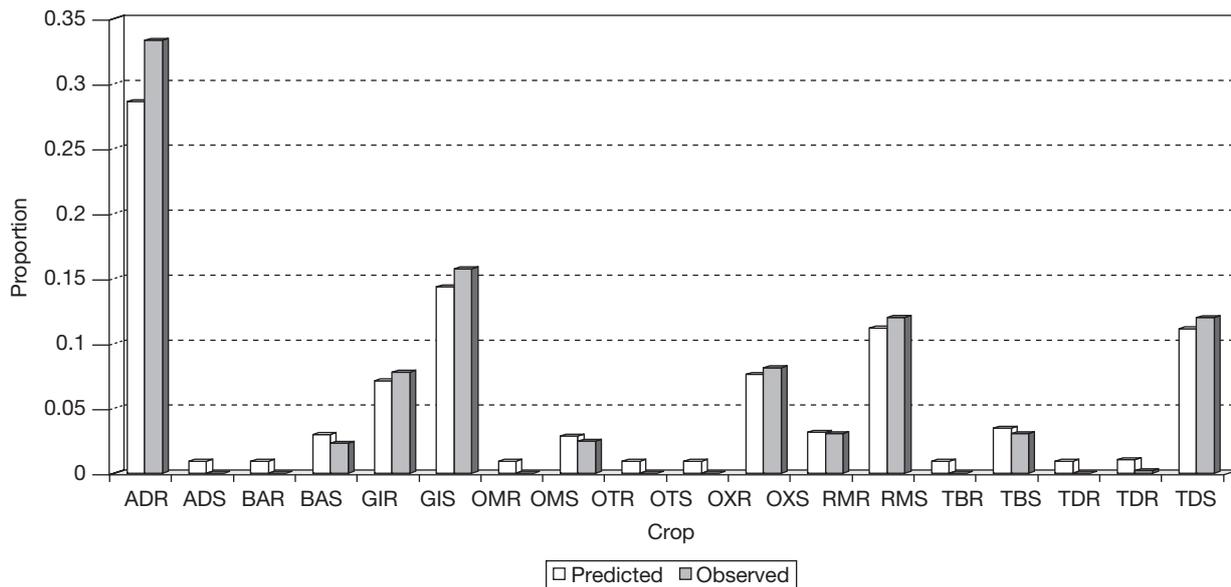
case, the sample averages,  $\bar{p}_{lj} = \frac{1}{T} \sum_{t=1}^T \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{p}_{i_{T+1}lj}$  and  $\bar{f}_{lj} = \frac{1}{T} \sum_{t=1}^T \frac{1}{n_t} \sum_{i=1}^{n_t} Y_{i_{T+1}lj}$  for  $l = 1, 2, \dots, J$  y  $j = 1, 2, \dots, J$  have been compared. Figure 1 shows both averages ( $\bar{f}_{lj}$  in light grey and  $\bar{p}_{lj}$  in grey) when  $l$  is cotton in irrigation and  $j$  is the crop indicated in  $x$ -coordinates. Figure 2 shows both averages when  $j$  is durum wheat and  $l$  is the crop indicated in the  $x$ -coordinates.

In addition, a comparison has been made in each year showing the approximation of forecasts,  $\bar{Y}_{jt} = \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{Y}_{i_{T+1}j}$ , to data observed,  $\bar{Y}_{jt} = \frac{1}{n_t} \sum_{i=1}^{n_t} Y_{i_{T+1}j}$ . In Figure 3 it is shown that the approximation,  $\bar{Y}_{jt} \approx \bar{Y}_{jt}$ , is verified.

However, there is a forecast error increasing the estimation error based solely on the field data. In our area of study, we estimate that the increase of the standard deviation of the estimator based on the forecast, vis-à-vis the deviation based on data observed at the end of the campaign is 17.72% for irrigation cotton (18.64% table olive in irrigation, 27.33% for sunflower).

### Other applications of crop rotation models

The crop rotation model defined by [4] is a useful tool for other applications, including econometric spatial

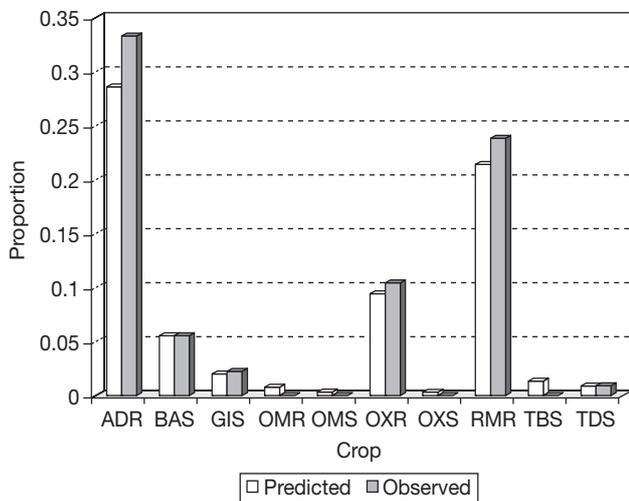


**Figure 1.** Proportion of the irrigation cotton surface in year  $t-1$  becoming the crop rotation in the following year  $t$  (average per segment and year) [irrigated cotton (ADR), unirrigated cotton (ADS), irrigated fallow land (BAR), unirrigated fallow land (BAS), irrigated sunflower (GIR), unirrigated sunflower (GIS), irrigated table olive crop (OMR), unirrigated table olive crop (OMS), irrigated oil olive crop (OTR), unirrigated oil olive crop (OTS), irrigated sugar beet (RMR), unirrigated sugar beet (RMS), irrigated soft wheat (TBR), un irrigated soft wheat (TBS), irrigated durum wheat (TDR), unirrigated durum wheat (TDS), irrigated «other crops» (OXR), unirrigated «other crops» (OXS)].

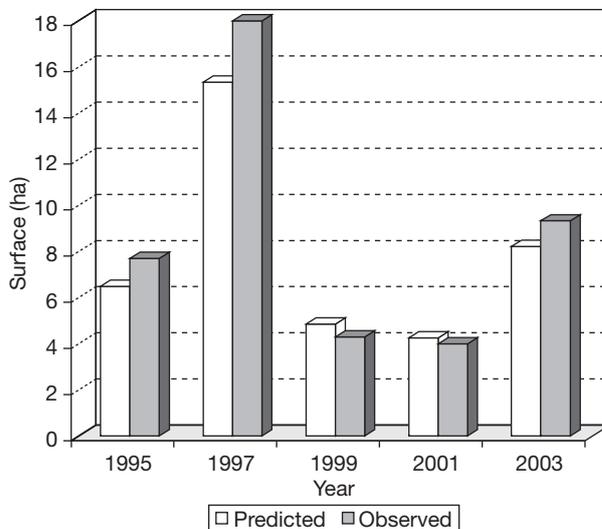
models on soil usage dynamics and for updating territorial censuses. In order to prevent any undesired environmental impact, it is especially relevant to forecast individual decisions made by farmers and to assess their environmental impact. As shown by Ambrosio *et al.* (2008), a crop rotation model is a basis to specify

econometric spatial models as a tool to monitor and control the use of natural resources (soil, water and air) by agriculture as well as synthetic waste products used in agriculture, such as pesticides and chemical fertilisers.

Crop rotation models are also useful for updating territorial censuses and the disaggregation of infor-



**Figure 2.** Proportion of the surface of crop shown in  $x$ -coordinates in year  $t-1$  becoming irrigation cotton in the crop rotation in the following year  $t$  (average per segment and year). Crops: see Fig. 1.



**Figure 3.** Average cotton surface (hectares) per segment predicted in year  $t$  together with the surface effectively observed.

mation. The degree to which territorial censuses are updated depends on the dynamics of soil uses and this dynamic may be reflected by using model [4] on crop rotation. Given a facility  $r$  of the area census  $Y_r$ , and given its disaggregation in alternative crops,  $\{Y_{rj}; j = 1, 2, \dots, J\}$ , where  $Y_{rj}$  is the part of  $Y_r$  occupied by the crop  $j$  at this time,  $t$ , and  $Y_r = \sum_{j=1}^J Y_{rj}$ . An estimate of  $Y_{rj}$

is gained by iterating  $\hat{Y}_{rj} = \sum_{j'=1}^J \hat{Y}_{rj'} \hat{p}_{j'j}$  and using model [4] in each iteration to rotate the crop alternative on the basis of the initial  $\{Y_{rj_0}; j = 1, 2, \dots, J\}$ .

In order to estimate  $\hat{p}_{j'j}$  on the basis of [4], the estimates of the most recent regression coefficients  $\hat{\beta}_{j'j}$  are used in each iteration, together with  $\{\hat{Y}_{rj'-1}; j' = 1, 2, \dots, J\}$ . The updating error is  $V\hat{Y}_{rj} = \sum_{j'=1}^J \hat{Y}_{rj'-1} V\hat{p}_{j'j}$  (assuming independence between  $\{\hat{p}_{j'j}; j' = 1, 2, \dots, J\}$  and conditioned to  $\hat{Y}_{rj'-1}$ ) where  $V\hat{p}_{j'j} = [p_{j'j}(1-p_{j'j}) \underline{x}_{j'j}]^T \text{Var}(\hat{\beta}_{j'j}) [p_{j'j}(1-p_{j'j}) \underline{x}_{j'j}]$ , where  $\underline{x}_{j'j}$  is the vector  $(1 \times J)$  of auxiliary values  $\{Y_{rj'-1}; j' = 1, 2, \dots, J\}$  and  $\hat{\beta}_{j'j}$  is the vector whose components are  $\{\beta_{j'j}; j' = 1, 2, \dots, J\}$  and  $\text{Var}(\hat{\beta}_{j'j}) = (X^T V_{j'j}^{-1} X)^{-1}$ , where  $V_{j'j} = \text{diag} \{p_{j'j}(1-p_{j'j}); t = 1, 2, \dots, T\}$  and  $X$  is a matrix  $T \times J$  whose generic row is  $\{\underline{x}_{j'j}; t = 1, 2, \dots, T\}$ .

Note that  $\sum_{j=1}^J \hat{Y}_{rj} = \sum_{j=1}^J \hat{Y}_{rj'-1} \left( \sum_{j=1}^J \hat{p}_{j'j} \right)$  and  $\sum_{j=1}^J \hat{p}_{j'j} = 1$  so that  $\sum_{j=1}^J \hat{Y}_{rj} = \sum_{j=1}^J \hat{Y}_{rj'-1} = \sum_{j=1}^J \hat{Y}_{rj_0} = Y_r$ . An update of the territorial census is obtained merely by aggregating  $\{\hat{Y}_{rj}; j = 1, 2, \dots, J\}$  in the usage categories considered in the census. In the event that coefficients  $\beta_{j'j}$  were heterogeneous among sampling units and time, homogeneous zone types and/or time intervals may be established.

## Agro-meteorological models

Agro-meteorological models are a useful tool to carry out forecasts on the crop yields depending on weather conditions, and the EU has developed two of such models: one (CGMS) for annual crops and another (OLIWIN) for multi-annual crops. These are computing systems delivering indicators on crop yield with a high spatiotemporal resolution (on the plot level and every 10 days over the campaign) on the basis of environmental data (a ground map and a geo-referenced weather database) and a crop model.

In CGMS (Supit *et al.*, 1994; Van Raaij and van der Wal, 1994; Van der Wal, 1994) the WOFOST crop model is used to calculate yield indicators, such as biomass generated by the crops, dry matter in reserve organs and leaf area indices, as well as the vegetative development level of the crop, including crop water needs (Hijmans *et al.*, 1994). In OLIWIN (European Commission, 1997), the SIMWAT model is used to calculate the water balance. Other elements of the Spanish system also produce vegetative crop development indicators based on vegetation indices calculated on the basis of satellite images (remote sensing). These indicators can be useful for crop yields forecasts and, likewise, indicators calculated on the basis of CGMS and OLIWIN may be useful for the purposes pursued by those other elements.

## The model

Crop yields forecasts are gained from statistical models relating the crop yield observed in the field at the end of the campaign with the indicators calculated on the basis of CGMS and OLIWIN over the campaign. These are *indirect estimates* resulting from the combination of field information with *auxiliary* information using the statistical model [3], with slight scale and notation changes. This model is illustrated using the crop yields included in ESYRCE on the plot level, grouping data into UTM blocks of  $10 \times 10$  km.

It is assumed that at moment  $t - \tau$  of the campaign under way, it is possible to obtain a measurement  $E_{y_{ijt}} = \underline{x}_{ij,t-\tau}^T \underline{\beta}$  of the yield of a certain crop in plot  $j$  of block  $i$  at the end of campaign  $t$ , depending on auxiliary information  $\underline{x}_{ij,t-\tau}$  (indicators of the agro-meteorological model in  $t - \tau$ ) and a parameters vector,  $\underline{\beta}$ . The error of this measurement,  $u_{ijt} = y_{ijt} - \underline{x}_{ij,t-\tau}^T \underline{\beta}$ , is its deviation regarding the true value,  $y_{ijt}$ , and it is the result of numerous causes grouped under two large categories,  $u_{ijt} = v_{it} + e_{ijt}$ : some of which equally affect all plots within a same block  $i$  within a same campaign  $t$ ,  $v_{it}$ , and others varying from one plot to another,  $e_{ijt}$ . This model is a way of specifying that the measurement errors are correlated. For example, in the calculation of the indicators  $\underline{x}_{ij,t-\tau}$  ground maps are used where the same physical and chemical properties are assigned to the ground of a same minimal mappable area, inserting an error that affects all plots to the same degree with all or part of its land on said minimum unit.

The characteristics of error components are considered the same as those of the model [3] and the model parameters are calculated on the basis of observations  $\{(y_{ijt}, x_{ij,t-\tau}) \mid i = 1, 2, \dots, m; j = 1, 2, \dots, n_i; t = 1, 2, \dots, T_j\}$ , where  $y_{ijt}$  is the yield of the crop studied observed in the field and  $x_{ij,t-\tau}$  are yield indicators calculated on the basis of the agro-meteorological model. The forecast of average crop yields  $\bar{Y}_{it} = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ijt}$  within block  $i$  is obtained using the BLUE estimator for  $\bar{Y}_{it}$ , and also the average squared error of the estimator.

## Estimation

The method applied to determine crop yield forecasts has been checked against a variety of crops—common wheat, sunflower, beetroot (irrigated and unirrigated), potato in irrigation, table olive and olive groves for oil production (irrigated and unirrigated)—from two pilot areas in Andalusia: *Campiña de Sevilla* and *Comarca de Antequera*. The field data used are those collected in ESYRCE: Data on the yield of 158 plots of unirrigated wheat (distributed in 26 blocks) has been used, and 270 plots of unirrigated sunflowers (in 44 blocks); in irrigation we have used data on yield in 15 wheat plots (in 5 blocks), in 24 sunflower plots (in 6 blocks), in 28 beetroot plots (in 6 blocks) and in 5 potato plots (in 2 blocks). For table olive, we have used yield data in 17 unirrigated blocks and 3 irrigated blocks and for olive groves for oil production, in 20 unirrigated blocks and 2 irrigated blocks.

Model [3] has been estimated over the campaign for each 10 days from January 1. In each plot, we have calculated the CGMS indicators for each 10-day period and campaign and the OLIWIN data for each month. Model checking is carried out for CGMS results concerning potential biomass in irrigation crops (with no water restrictions) and that obtained with rainwater (without irrigation) and for those of OLIWIN concerning the water balance of June and September. The independent term is not significantly different from zero in any case, and this means that in annual crops the expected yield follows the form  $Ey_{ijt} = \beta x_{ij,t-\tau}$ , where  $x_{ij,t-\tau}$  is the potential biomass of irrigated crops and unirrigated biomass, calculated on the basis of the CGMS data. Coefficient  $\beta$  measures the effect on expected crop yield at the end of the campaign,  $t$ , due to changes in  $x_{ij,t-\tau}$  observed during the campaign via the CGMS in the moment  $t - \tau$ .

## Results

This effect is statistically significant (t-Student > 2) for all crops studied and the value of  $\beta$  is estimated at 0.19 for unirrigated wheat and 0.15 for irrigated wheat, at 0.45 for unirrigated sunflower and 0.31 for irrigated sunflower, at 3.17 for irrigated beetroot and 2.44 for potatoes. As far as table olive and olive for oil production are concerned, the expected crop yield follows the form  $Ey_{ijt} = \beta_1 x_{1ij,t-\tau} + \beta_2 x_{2ij,t-\tau}$  where  $x_{1ij,t-\tau}$  y  $x_{2ij,t-\tau}$  is the water balance in June and September respectively. In table olives, the coefficients  $\beta_1$  and  $\beta_2$  are not significantly different from zero (t-Student < 2), meaning that the effect of indicators of OLIWIN over the expected yield at the end of the campaign is not sufficient to contribute to forecasts of these yields. They are sufficient, however, for olives for oil production (t-Student > 2): The values of  $\beta_1$  and  $\beta_2$  are estimated at 20.3 and -12.8 respectively.

The error component,  $v_{it}$ , is significantly different from zero for annual unirrigated crops considered (p-value of 0.0042 for wheat and 0.0049 for sunflower), although this is not the case for irrigated crops (p-value of 0.087 for wheat, 0.06 for sunflower, 0.43 for beetroot and 0.25 for potato). An interpretation of this result may be that the variability observed in the crop yields between blocks, not explained by the factors that CGMS takes into consideration, may be partly due to factors linked to the soil, that are not observable, such as differences in soil fertility. These differences are significant in the case of unirrigated crops. However, they do not appear to be significant in the case of irrigated crops, where fertility is more uniform. The hypothesis regarding the existence of non-observable factors linked to time has been tested and, in all cases, such hypothesis was rejected, this would lead one to believe that CGMS correctly records the effect of temporal variability of weather variables. However, it does not report the effect of variability in soil fertility as well. Error component,  $v_{it}$ , is also significant in the case of olive groves for oil production (p-value 0.018), but not in the case of table olives.

### *Predictions and their level of precision*

In order to measure the degree of precision of forecasts obtained on the basis of the model we have estimated its standard error in some cases. The standard error of irrigated wheat yield forecasts based on potential

biomass estimated via CGMS in the first 10 days of June was 7.39%. For unirrigated wheat the error of the prediction for that same 10-day period was 13.93%. The standard error of irrigated sunflower yield forecasts based on potential biomass estimated via CGMS in the first 10 days of August was 3.56%. For unirrigated sunflower, the error of the prediction for that same 10-day period was 16.41%. The typical error of irrigated beetroot yield forecasts based on potential biomass estimated via CGMS in the second 10-day period of November was 5.75%. The typical error for yield predictions of unirrigated olive groves for oil production carried out in September was 3.85%.

### Selection of farmer samples on the basis of an area sampling frame

A large part of the information required is not directly observable on the ground and it must be collected by means of direct interviews with the farmers. Generally speaking, samples of farmers to be interviewed are selected on the basis of a list sampling frame (directory of farms). Keeping such a list up-to-date is costly, but this is not the case of an area sampling frame. Thus, the selection of samples on the basis of an area sampling frame offers other advantages among which we can highlight the reduction of coverage biases due to incomplete sampling frames (González-Villalobos *et al.*, 1996) and in this article the selection of the sample of farmers on the basis of an area sampling frame is suggested. The idea is to identify farmers farming lands within the area sample segments and choose a sample of these farmers.

The Department for Agriculture and Fishing of the Andalusian Government has developed an agricultural survey design procedure based on an area sample for the estimation of surfaces, yields and other structural and technical-economic features of vegetable crops and intensive crops as a whole that is analogous to the system applied in this article. Intensive crops, in particular vegetables, are grown successively over the year on a same plot, depending on variable market circumstances, and this means it is necessary to interview farmers. A step-by-step description—construction of the area sampling frame, selection of the segments sample and the farmers sample, collection of information and calculation of estimates and the estimation error—, together with the results of the samples designed with this methodology for the structural and technical-

economics features of intensive crops in Andalusia and for the estimation of production costs and their profitability threshold may be found in Ambrosio *et al.* (1999, 2006).

### Discussion

In order to make decisions, economic actors require information which is timely, precise, specific and which has a high spatiotemporal resolution. Society, via its political representatives, demands information to follow the impact of agricultural systems on the environment, to control waste produced and to design sustainable agricultural systems, which stay productive over time.

Production of this information is a complex task that public authorities are taking care of. This article applies statistical methods to improve spatial and temporal resolution of official agricultural and environmental information, combining field data with auxiliary information from other sources, including administrative registers and remote sensing. The methods applied are checked, validated and illustrated using the agricultural and environmental information system of the Spanish Ministry of the Environment and Rural and Marine Affairs.

A method is applied to integrate data from registries of the authorities with data from random samples, which allows both to the validation of crop area data from registries and the improvement of precision of crop area estimates without cost increase, since the required data are available. This improvement involves a reduction of the crop area estimator error, a reduction which varies from one crop to another, ranging from 0.35 (olive groves for the production of oil) to 12.08 (legumes) percentage points (see Table 1).

A specific method is applied to improve precision of crop area estimates in small areas (municipalities) with the aid of satellite images. The relative precision gain offered by the *indirect estimator* proposed vis-à-vis the *direct estimator* currently used ranges between 2.08 and 46.37 from one small area to another: this means that the typical error of the crop area indirect estimator is between 31% and 85% lower than the typical error of the direct estimator currently used.

In order to improve the temporal resolution of the information system, a method is applied, based on the analysis of the temporal series, allowing result forecasts to be issued at the beginning of each farming campaign. In addition, this last method allows a reduction of the

error of the crop area estimator, without cost increase: for cereal acreage, the reduction varies (see Table 2) from one Spanish Autonomous Community to another, with values ranging from 1.10% (Aragon) to 2.50% (Extremadura). In the case of other crops, the reductions have similar magnitudes.

It is shown that agro-meteorological models are a useful tool to carry out forecasts on the crop yields depending on weather conditions and that the forecast error ranges from 3.85% for unirrigated olive groves for oil production to 16.41% for unirrigated sunflower. Crop rotation models are estimated and validated using panel area samples and the results are satisfactory as can be seen on Figures 1, 2 and 3. Various uses for crop rotation models are provided, among others, applications for the update and disaggregation of information in territorial censuses (SIGPAC) and the development of spatiotemporal econometric models. These last models are especially useful to monitor and control the use of natural resources (soil, water and air) by agriculture and synthetic product waste used in agriculture, such as pesticides and chemical fertilisers.

## Acknowledgements

This study was supported by the Spanish Ministry of the Environment and Rural and Marine Affairs. Thanks are due to José María Fernández del Pozo and Porfirio Sánchez for their commitment to the improvement of the Spanish agricultural information system. We wish to thank José Ramón Sanz Almodóvar, for the creation of the databases used. We would like to thank the referees and the subject editor for helpful comments and suggestions.

## References

- AMBROSIO L., GALLEGO J., 1998. Crop area agricultural survey program based on area frame sampling methods. Spain (1988-1997). In: Multiple frame agricultural surveys. Vol. 2. FAO Stat Dev Ser, 10. Rome.
- AMBROSIO L., IGLESIAS L., MONTAÑÉS J., RUBIO L.A., 1999. Diseño de encuestas agrícolas basadas en un marco de áreas. Metodología y aplicación a los cultivos intensivos en Andalucía. Informaciones Técnicas, 68/99. Junta de Andalucía, Consej Agric Pesc. 98 pp. [In Spanish].
- AMBROSIO L., IGLESIAS L., 2000. Land cover estimation in small areas using ground survey and remote sensing. Remote Sens Environ 74, 240-248. doi:10.1016/S0034-4257(00)00114-0.
- AMBROSIO L., IGLESIAS L., MARÍN C., 2003. Systematic simple design for the estimation of spatial means. Environmetrics 14, 45-61. doi: 10.1002/env.564.
- AMBROSIO L., IGLESIAS L., MARÍN C., MONTAÑÉS J., PECHARROMÁN V., RUBIO L.A., SERRANO A., 2006. Diseño de encuestas agrícolas basadas en un marco de áreas. Características estructurales y costes de producción de la agricultura intensiva en Andalucía. Junta de Andalucía, Consej Agric Pesc. [In Spanish].
- AMBROSIO L., BARRERO C., FUERTES A., IGLESIAS L., MARÍN C., MARTÍN A., MENA M.A., PASCUAL V., 2007. Desarrollos metodológicos en materia de estadísticas oficiales agrarias. Informe técnico. Subdir Gral Estadísticas, MAPA - Dept Economía y Ciencias Sociales Agrarias, ETSIA, UPM. [In Spanish].
- AMBROSIO L., IGLESIAS L., MARÍN C., PASCUAL V., SERRANO A., 2008. A spatial high-resolution model of the dynamics of agricultural land use. Agr Econ 38, 233-245. doi: 10.1111/j.1574-0862.2008.00296.x.
- BATTESE G.E., HARTER R. M., FULLER W.A., 1988. An error-components model for prediction of county crop areas using surveys and satellite data. J Am Stat Assoc 83, 28-36.
- BLIGHT B.J.N., SCOTT A.J., 1973. A stochastic model for repeated surveys. J R Stat Soc B 35, 61-66.
- BURT T.P., 2003. Monitoring change in hydrological systems. Sci Total Environ 310, 9-16. doi:10.1016/S0048-9697(02)00618-6.
- EUROPEAN COMMISSION, 1997. Oliwin Project: agro-meteorological models for the estimation at harvest of olive and vine yield; regional and national level. Final Report, Vol. 1.
- FAY R.E., HERRIOT R.A., 1979. Estimates of income for small places: an application of James-Stein procedures to census data. J Am Stat Assoc 74, 269-277.
- FULLER W., 1987. Measurement error models. Wiley, NY. 464 pp.
- GOLAN A., JUDGE G.G., MILLER D., 1996. Maximum entropy econometrics: robust estimation with limited data. Wiley, NY.
- GONZÁLEZ-VILLALOBOS A., WALLACE M., VOGEL F., COTTER J., KESTLE R., 1996. Multiple frame agricultural surveys volume 1: Current surveys based on area and list sampling methods. FAO Stat Dev Ser, 7. Rome.
- GURNEY M., DALY J.F., 1965. A multivariate approach to estimation in periodic sample surveys. Proc Social Statist Sect Amer Statist Assoc. pp. 242-257.
- HAINING R., 1990. Spatial data analysis in the social and environmental sciences. Cambridge Univ Press.
- HIJMANS R.J., GUIKING-LENS I.M., VAN DIEPEN C.A., 1994. WOFOST 6.0: User's guide for the WOFOST 6.0 crop growth simulation model. Wageningen, DLO Winand Staring Centre, Technical Document 12.
- JIANG J., LAHIRI P., 2006a. Mixed model prediction and small area estimation. Test 15, 1-96. doi: 10.1007/BF02595419.

- JIANG J., LAHIRI P., 2006b. Estimation of finite population domain means- a model assisted empirical best prediction approach. *J Am Stat Assoc* 101, 301-311. doi:10.1198/016214505000000790.
- JONES R.G., 1979. The efficiency of time series estimators for repeated surveys. *Aust NZ J Stat* 21, 1-12. doi: 10.1111/j.1467-842X.1979.tb01118.x.
- JONES R.G., 1980. Best linear unbiased estimators for repeated surveys. *J R Stat Soc B* 42, 221-226.
- KUHLMANN F., BRODERSEN C., 2001. Information technology and farm management developments and perspectives. *Comput Electron Agr* 30, 71-83. doi:10.1016/S0168-1699(00)00157-5.
- LABOVITZ M.L., MASUOKA E.J., 1984. The influence of autocorrelation in signature extraction- an example from a geobotanical investigation of Cotter Basin, Montana. *Int J Remote Sens* 5, 315-332. doi: 10.1080/01431168408948811.
- MADDALA G.S., KIM I.M., 1998. Unit roots, cointegration and structural change. Cambridge Univ Press. 505 pp.
- McCOWN R.L., 2002. Changing systems for supporting farmers' decisions: problems, paradigms and prospects. *Agr Syst* 74, 179-220. doi:10.1016/S0308-521X(02)00026-4.
- PATTERSON H.D., 1950. Sampling on successive occasions with partial replacement of units. *J R Stat Soc B* 12, 241-255.
- RAO J.N.K., 2003. Small area estimation. Wiley, NY. 313 pp.
- ROYALL R.M., 1970. On finite population sampling theory under certain linear regression models. *Biometrika* 57, 377-387. doi:10.1093/biomet/57.2.377.
- ROYALL R.M., HERSON J., 1973. Robust estimation in finite populations I & II. *J Am Stat Assoc* 68, 880-893.
- SÄRNDAL C.E., 1984. Design-consistent versus model-dependent estimators for small domains. *J Am Stat Assoc* 79, 624-631.
- SCOTT A.J., SMITH T.M.F., 1974. Analysis of repeated surveys using time series methods. *J Am Stat Assoc* 69, 674-678.
- SUPIT I., HOOIJER A.A., VAN DER DIEPEN C.A. (eds), 1994. System description of the WOFOST 6.0 crop simulation model implemented in CGMS. Publication EUR 15956 EN, EU Office Official Publ, Luxembourg.
- TAM S.M., 1986. Optimal prediction in stochastic regression models with application to the analysis of repeated surveys. *Aust NZ J Stat* 28, 345-353. doi: 10.1111/j.1467-842X.1986.tb00707.x.
- TAM S.M., 1987. Analysis of repeated surveys using a dynamic linear model. *Int Stat Rev* 55, 63-73.
- VAN RAAIJ B.F.T., VAN DER WAL T., 1994. CGMS version 3.1. Technical description. JRC/SC-DLO/Gray. Technical Document 15.3. Wageningen, The Netherlands.
- VAN DER WAL T., 1994. CGMS version 3.1. Data installation guide and operator's manual. Technical Document 15.2. Wageningen, The Netherlands.
- WOLF S., JUST D., ZILBERMAN D., 2001. Between data and decisions: the organization of agricultural economic information systems. *Research Policy* 30, 121-141. doi: 10.1016/S0048-7333(99)00096-7.
- WEBSTER R., CURRAN P. J., MUNDEN W. J., 1989. Spatial correlation in reflected radiation from the ground and its implications for sampling and mapping by ground-based radiometry. *Remote Sens Environ* 29, 67-78.
- YOU Y., RAO J.N.K., 2002. A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Can J Stat* 30, 431-439.