RESPECTIVE CONTRIBUTION OF YIELD AND AREA ESTIMATES TO THE ERROR IN CROP PRODUCTION FORECASTING

P. Defourny [a, *], X. Blaes [a], P. Bogart [a]

[a] Dpt of Environmental Sciences and Land Use Planning , Université catholique de Louvain, Croix du Sud, 2 bte 16, 1348 Louvain-la-Neuve, Belgium Pierre.Defourny@uclouvain.be

Commission VIII, WG VIII/10

KEY WORDS: agriculture production forecasting, error, trends analysis, acreage estimate, yield estimate

ABSTRACT:

Improvement of crop production forecasting is a major research subject since many years. A systematic analysis of long time series made available through 2 extended databases for 4 crops and 15 European countries allowed investigating the basic question of the main sources of the year-to-year production variability. This statistical study could not conclude by a simple answer to this simple question. It was shown that most of the inter-annual production variability is actually due to the main producer countries. The year-to-year production variability was mainly related to the yield variation for wheat, potato and sunflower, but to the cultivated area variation for maize. Forecasting the production based on exponential smoothing technique highlighted that, in some countries, it is better to forecast the crop production knowing the cultivated surface while deriving the yield from the previous years' trend. However, the contrary was also found according to the crops and the country. Finally, the accuracy required to improve a forecasting system only based on statistical trends analysis was computed in order to define the performance target for any alternative source of information such satellite remote sensing.

1. INTRODUCTION

Forecasting agriculture production is of critical importance for the policy makers as for the market stakeholders. In a context of globalization, it is indeed very valuable to get as soon as possible accurate overall estimates of crop production at international scale. Since early days of satellite remote sensing till now, optical imagery has been experimented either to better estimate the crop acreage or to assess the canopy development as a proxy variables of the expected yield. However, a very simple but critical question remained apparently quite open: should we focus first on the crop acreage or on the yield estimation to improve the production forecasting? Even in a context of intensive cropping systems like the European Union, such basic question has been hardly addressed. Therefore, the overall objective of this study was to investigate the various sources of errors in the production estimation and to compare the respective relative errors associated to yield or acreage estimates when derived from trend analysis.

2. METHODS AND DATA SET

2.1 Overall approach

The research questions were designed from a user perspective addressing the most relevant issues to set a research agenda for the remote sensing community. The first questions concerned the inter-annual variability of productions, yields and cultivated areas for different levels of aggregation as this variability is the main justification of any forecasting system. These questions were twofold:

1. How large is the inter-annual variability of production for a given crop? How it is spatially distributed?

2. What are the respective contributions of the area and the yield variability on the total production fluctuation for a given crop?

In a second step, a forecasting perspective relying only on the existing statistics was adopted to formulate the following questions:

3. What is better to know - yield or area – to estimate the crop production for the current year?

Based on the previous results, a quite relevant question was expressed to support the remote sensing research agenda definition:

4. What is the required accuracy of any alternative source of information, e.g. optical imagery, to be able to improve a forecasting system based only on trend analysis?

2.2 Available time series of agriculture statistics

This research was carried out for 15 countries of the European Union (EU) because of the availability of long time series. The study actually relied on 2 existing data bases: the national statistics from 1950 to 2003 for 15 European countries (DB Cronos) and time series detailed by region, i.e. at NUTS 2 level in the EU administrative typology, from 1975 to 2002 (DB Regio). Both include the yields and surfaces statistics for 4 crops: wheat, maize, potato and sunflower. They were made

* Corresponding author.

available for the study by the AGRIFISH unit of the Joint Research Center (Ispra, Italy).

These statistics are routinely collected by various national statistical services and conveyed to the statistical office of the European Commission, EUROSTAT. The surface per crop is currently estimated from various sources, including field survey and remote sensing interpretation. In the EU the farmers' declarations and the related control with remote sensing of area based subsidies is a main source of information provided by each country.

3. RESULTS

3.1 Overall trends and inter-annual variability

The overall temporal evolution of the production was observed using the DB Cronos and plotted here for the 3 main producers' country of two crops (figure 1). To estimate the temporal trend, statistical test (f tests) were successively performed to select the best polynomial regression order. The polynomial of degree i (starting with i=0) was compared to the polynomial of degree i+1 thanks to the root mean square error (RMSE) between polynomial trend and real data. The selected polynomial regression order corresponded to the lowest RMSE value.

From these curves, we can not conclude to a single overall trend of the production evolution. The trends vary a lot from crop to crop: wheat and maize show a high production increase but potato shows rather stable production levels. The sunflower production has increased before the 90's and decreased during the last decade. However, some countries show different trends. For example, the production's level of wheat in Italy is stable whereas the three main producers increase their production with different rates.

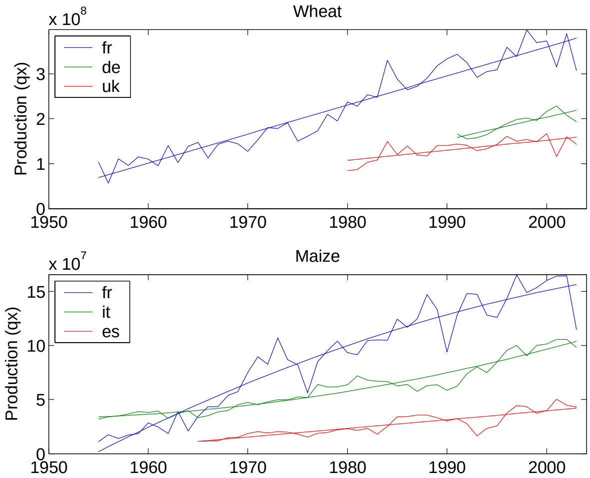


Figure 1: Evolution of the production level in the 3 main producers' country in Europe for wheat and maize.

In order assess the inter-annual variability of each producers' country, the deviation between the recorded production and the production estimated by the trend (root mean square distance) was computed for each crop. These deviations were then plotted as a function of the production averaged over the same period of time (figure 2). Most of the inter-annual variability of the productions in EU comes from the main producers' countries for all 4 crops, i.e. higher the productions level higher the variability. At country level, the inter-annual variability of the production seems to be proportional to the production volume.

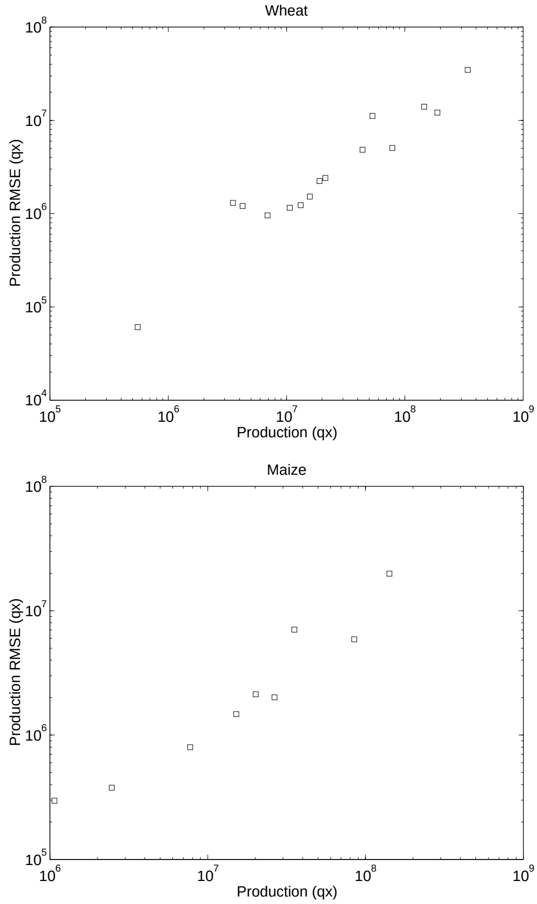


Figure 2: Relationship between the average production levels for the 15 EU member states and the respective inter-annual production variability for wheat (upper) and maize (lower).

3.2 Respective contributions to the inter-annual variability

The total crop production of a region depends on the acreage per crop and the mean yield. Year-to-year variability of the production level observed in the previous section could come from both, the yields' variability and from the areas' variability. In order to improve forecasting system, it is important to know which of these two variables was driving the production fluctuations.

The effect on the production estimate of the missing of either, the yield or the surface information, was assessed for each crop type. The error of the respective production estimate was then compared.

As a preliminary step, the yield and the surface trends of long statistical series (DB Cronos) were respectively modelled for each crop and each country using polynomial functions as described above.

In a second step, the production of each year was estimated for each crop along 2 ways: on one hand by multiplying the observed yield by the surface estimated from the trend and, on the other hand, by multiplying the observed surface by the yield estimated from the trend. The comparison of both production

estimates to the observed production was summarized by the RMSEa and RMSEy for the temporal series (table 1).
These gave an indication of the respective contribution of yields and surfaces variability on the total production for each year.

Member state	Production (T)	RMSEa (qx)	RMSEy (qx)	Main var.
Wheat				
France	3.4E+08	1.5E+07	1.9E+07	Y
Germany	1.9E+08	6.8E+06	1.3E+07	Y
West Germ.	*7.4E+07*	*2.5E+06*	*4.6E+06*	*Y*
U. Kingdom	1.5E+08	1.2E+07	8.2E+06	A
Italy	7.8E+07	4.4E+06	7.2E+06	Y
Spain	5.3E+07	3.1E+06	8.1E+06	Y
Denmark	4.4E+07	2.1E+06	1.7E+06	A
Greece	2.1E+07	1.4E+06	3.7E+06	Y
Sweden	1.9E+07	3.1E+06	1.5E+06	A
Belgium	1.6E+07	7.3E+05	1.0E+06	Y
Austria	1.3E+07	1.1E+06	1.0E+06	A
Netherlands	1.1E+07	8.8E+05	7.1E+05	A
Ireland	6.9E+06	6.9E+05	5.4E+05	A
Finland	4.3E+06	1.2E+06	5.9E+05	A
Portugal	3.5E+06	6.1E+05	1.0E+06	Y
Luxembourg	5.5E+05	3.3E+04	5.4E+04	Y
EU15	**9.4E+08**	**3.7E+07**	**4.9E+07**	**Y**
Maize				
France	1.4E+08	9.6E+06	9.3E+06	A
Italy	8.5E+07	4.0E+06	4.1E+06	Y
Spain	3.5E+07	3.9E+06	2.0E+06	A
Germany	2.7E+07	2.0E+06	2.5E+06	Y
West Germ.	*7.1E+06*	*9.5E+05*	*7.0E+05*	*A*
Greece	2.0E+07	2.3E+06	1.3E+06	A
Austria	1.5E+07	1.2E+06	1.0E+06	A
Portugal	7.7E+06	5.5E+05	4.3E+05	A
Belgium	2.5E+06	1.7E+05	2.2E+05	Y
EU15	**3.4E+08**	**1.9E+07**	**2.1E+07**	**Y**
Potato				
Germany	1.2E+08	1.0E+07	1.2E+07	Y
West Germ.	*1.4E+08*	*5.5E+06*	*1.2E+07*	*Y*
Netherlands	7.3E+07	4.3E+06	4.1E+06	A
France	5.9E+07	1.1E+07	8.8E+06	A
Spain	4.0E+07	2.2E+06	2.2E+06	A
Belgium	2.3E+07	1.2E+06	2.0E+06	Y
Italy	2.2E+07	3.1E+06	1.8E+06	A
Denmark	1.5E+07	1.5E+06	1.4E+06	A
Portugal	1.2E+07	8.3E+05	1.5E+06	Y
Sweden	1.1E+07	4.1E+05	1.4E+06	Y
Finland	7.7E+06	4.4E+05	9.6E+05	Y
Austria	7.5E+06	5.3E+05	7.8E+05	Y
Ireland	5.7E+06	8.7E+05	1.6E+06	Y
Luxembourg	2.3E+05	3.0E+04	1.1E+05	Y
EU15	**4.4E+08**	**1.9E+07**	**2.2E+07**	**Y**
Sunflower				
France	1.9E+07	4.7E+06	1.1E+06	A
Spain	9.6E+06	1.3E+06	1.8E+06	Y
Italy	4.8E+06	3.8E+05	2.4E+05	A
Germany	1.1E+06	8.8E+05	3.6E+05	A
Austria	5.7E+05	1.8E+05	4.8E+04	A
Greece	3.1E+05	3.6E+05	1.1E+05	A
Portugal	3.0E+05	7.3E+04	1.2E+05	Y
EU15	**3.6E+07**	**1.7E+06**	**3.9E+06**	**Y**

Table 1: Root mean square error of the production estimated without knowing the cultivated area (RMSEa) and without knowing yield (RMSEy). Series of West Germany stop in 1994.

Furthermore, for each country and each crop type, the comparison between both estimated production curves with regards to the real production highlighted which one of the yield or area, is more important to know in a crop production estimation perspective (figure 3).

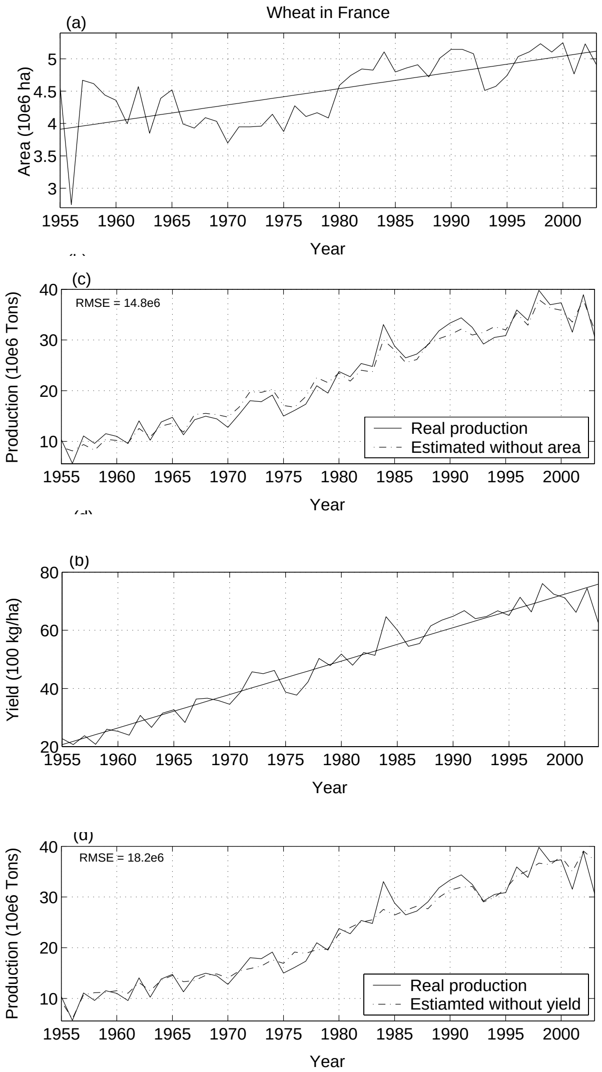


Figure 3: Temporal evolution of the wheat areas (a) and yield (b) in France with their respective trend. Temporal production and production estimated without area (c) and without yield information (d).

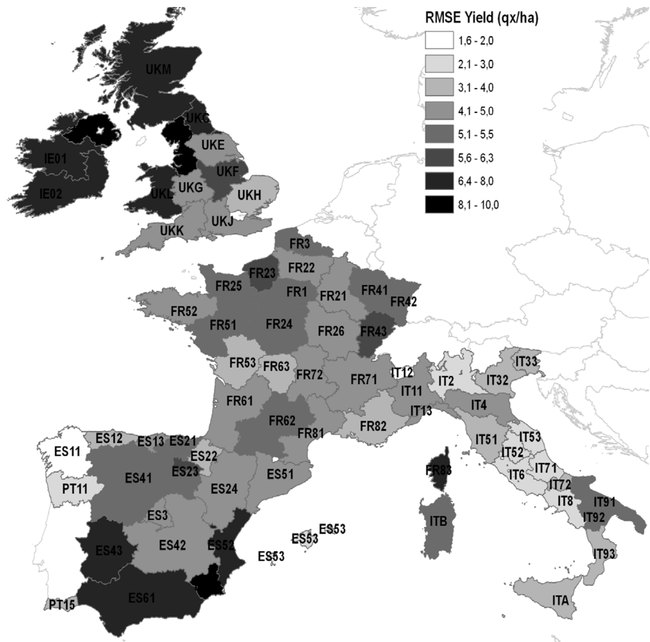
The trends of areas and yields were best estimated with long series. When the series length was shorter than 20 years, the temporal evolution was generally considered as constant but it would have been different if a longer series would have been available. Only the estimation made using long series are discussed. Moreover, as some countries have almost the same RMSEa and RMSEy, only some of the most contrasted situations are discussed here below.

For the wheat, 72% of the European production comes from France, Germany and United Kingdoms (91 % when adding Italy, Spain and Denmark) and the main source of variability in the production at European level is the yield variability.

Six countries are responsible of 96% of the European maize production. While the difference between RMSEa and RMSEy is rather small, the variability of the production at country level is mainly driven by the area variability. The opposite was

found at European level but from a short time series (1988-2003).
In general the potato's yields are highly variable in time whereas the surfaces cultivated in potato show a progressive decrease in time (except in the Netherlands). Whereas the contrary is observed in France and Italy, the main source of variability in production is the yields variability.

The variability of sunflower production is apparently related to the surface variability in most of the countries. However, a detailed analysis of the trends allowed adjusting more appropriate models leading to the opposite conclusion. The yield variability was finally found higher than the area variability.



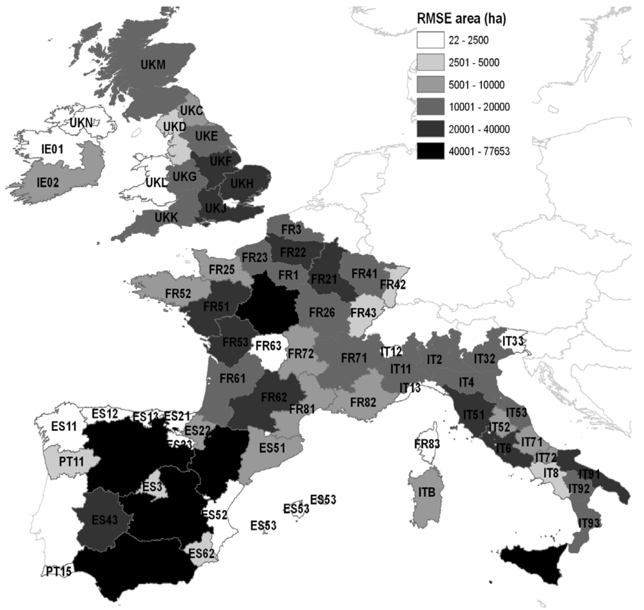


Figure 4: Deviation from the temporal evolution (from 1975 to 2002) of the yield (upper) and the cultivated area (lower) for the wheat in each region.

The crop productions are also quite variable across regions and other scales of variability should be considered, in particular those related to agricultural regions or at NUTS 2 level. For instance, the RMSE of the surface allocated to wheat was calculated for NUTS 2 regions over the period (figure 3).

Regional differences may be of great importance. For instance, the France shows that the inter-regional variability of yield (6-19 qx/ha) is of the same order of magnitude than the inter-annual variability (4-7 qx/ha) observed in the French regions.

3.3 Forecasting production using time series

A crop production is calculated by multiplying the yield by the area. While observed production information is only available after harvest, production forecast is very valuable information during the growing season. Time series of cultivated areas and yields recorded in the previous years can serve to predict areas and yields of the next year. Of course, the year-to-year variability of both, areas and yields, makes inaccurate the prediction based only on the past years and additional information about the current year is very much needed. Remote sensing timely delivers information over large areas and this study could indicate on what variable the satellite observation should focus on. In other words, is the production better forecasted knowing the surface or the yield for the year of estimation?

In order to answer this forecasting question, the total production of the year n was predicted at country level in 2 alternative ways using the Cronos DB. First, the production forecasting was obtained by multiplying the area observed for the year n by the yield predicted for this year n using the yields recorded in the previous years. Alternatively, the production forecasting was obtained by multiplying the area predicted from the areas recorded in the previous years by the yield measured for the year n.

To do so, the area and yield estimate were first predicted using an exponential smoothing (Chatfield, 2003) to weight more the recent past observations, once the overall trends was removed by a linear model. The set of weights decreases by a constant ratio in such a way that the weights lie on an exponential curve defined by the α coefficient value. The α values minimising the RMSE between the predicted values and the recorded ones was selected for each time series. Then productions were computed for each year of the time series and their deviation (RMSE) with regard to the observed production was computed.

	Pred. years	Prod. RMSE (qx) forecast without:		Best to know
		Area	Yield	
Wheat				
France	39	1.4E+07	2.1E+07	Y
Germany	3	4.4E+06	2.8E+07	Y
W. Germ.	*30*	*3.1E+06*	*5.9E+06*	*Y*
UK	14	1.4E+07	7.4E+06	A
Italy	39	4.7E+06	7.0E+06	Y
Spain	29	4.7E+06	9.8E+06	Y
Denmark	39	3.1E+06	2.0E+06	A
Greece	25	1.2E+06	3.9E+06	Y
Sweden	8	4.1E+06	1.2E+06	A
Belgium	39	7.8E+05	1.2E+06	Y
Austria	27	6.9E+05	1.3E+06	Y
Netherlands	39	9.6E+05	7.3E+05	A
Ireland	39	8.0E+05	5.3E+05	A

Finland	4	9.2E+05	4.7E+05	A
Portugal	19	6.1E+05	9.4E+05	Y
Lux.	39	5.9E+04	5.2E+04	A
Maize				
France	39	1.0E+07	1.1E+07	Y
Italy	39	3.9E+06	3.7E+06	A
Spain	29	4.1E+06	2.0E+06	A
Germany	6	3.0E+06	4.6E+06	Y
W. Germ.	*30*	*7.6E+05*	*8.9E+05*	*Y*
Greece	25	2.2E+06	1.6E+06	A
Austria	34	8.8E+05	1.2E+06	Y
Portugal	19	6.0E+05	4.9E+05	A
Belgium	39	2.8E+05	2.5E+05	A
Potato				
Germany	6	1.3E+07	1.2E+07	A
W. Germ.	*30*	*1.0E+07*	*1.1E+07*	*Y*
Netherlands	38	5.2E+06	4.7E+06	A
France	38	9.1E+06	7.2E+06	A
Spain	29	3.0E+06	2.9E+06	A
Belgium	39	1.9E+06	2.1E+06	Y
Italy	39	2.1E+06	1.4E+06	A
Denmark	38	1.5E+06	1.2E+06	A
Portugal	19	1.3E+06	1.1E+06	A
Sweden	11	5.1E+05	8.9E+05	Y
Finland	10	1.8E+05	7.1E+05	Y
Austria	21	6.5E+05	8.8E+05	Y
Ireland	38	1.1E+06	9.6E+05	A
Lux.	39	5.8E+04	7.8E+04	Y
Sunflower				
France	39	2.2E+06	1.3E+06	A
Spain	29	1.7E+06	2.4E+06	Y
Italy	39	5.3E+05	2.5E+05	A
Austria	21	1.4E+05	6.9E+04	A
Greece	25	2.9E+05	1.3E+05	A
Portugal	22	1.1E+05	1.2E+05	Y

Table 2: RMSE of productions forecasting using exponential smoothing and missing one of the two variables for the current year. Column 1 gives the number of predicted years, column 2 and 3 the RMSE of the production forecasted using either the predicted areas or the predicted yields.

As shown at table 2, it is sometimes better to predict the production knowing the surface and predicting the yield for the year of estimation, but the contrary sometimes has to be preferred, i.e. better to know the yield and estimate the surface for the year of interest. In most of the countries, the wheat production is better forecasted when areas are estimated. For the other crop types, it seems to be the contrary. For the maize cultivated in the south of Europe (Italy, Spain, Greece and Portugal), it is better to forecast the production estimating the yields. Potato and sunflower productions are also, in most of the countries, better forecasted when yields are estimated.

3.4 Required accuracy for the variables

Other sources of information, such as remote sensing, can be used to support the production forecasting. To improve the performance of a forecast system based only on trends analysis, the accuracy of the area or yield estimated from any other sources of information must be higher than the deviations obtained from trends analysis only.

For wheat, maize and potato, some countries show significant errors (figure 5). The worst seems to be the estimation of wheat yield in Portugal, with an error of 25%. Such an error could probably be improved using remote sensing information for instance. Other high errors on wheat yield predictions are also observed for Spain and Greece. Sweden and Finland show high errors on the area prediction but short series were used for these countries. The low errors (<+/-10%) obtained for the other countries using the exponential smoothing of statistical series can hardly be improved by alternative source of information.

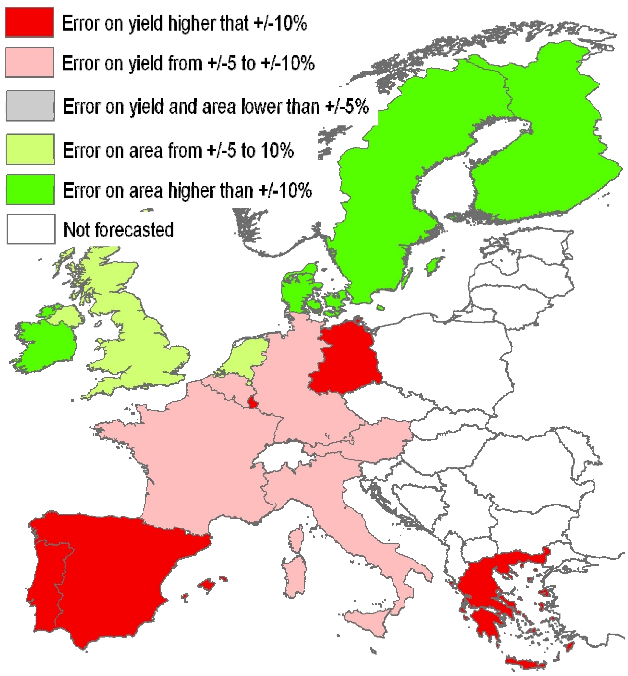


Figure 5: Map representing the maximum error on Area or Yield forecasted for wheat using the exponential smoothing.

4. CONCLUSIONS

This time series analysis demonstrated that a forecasting system for crop production in the EU can not only be based on statistical trends. Furthermore, the use of production trend is generally less efficient than area or yield trend, one of them being known. Research agenda for the improvement of a forecasting system could be defined according to 2 criteria:

- the focus on the main error source (area or yield) to improve the estimate from the alternative source to trend analysis,
- the feasibility to reach the required accuracy to reduce the prediction error obtained for the trend analysis.

The overall study approach successfully analysed the existing time series in order to document in an objective and quantitative way the need and the feasibility to improve current forecasting system simply based on trends analysis. Such an approach could be repeated at various scales and in various part of the world to define the research priorities.

Acknowledgements

This study was supported by the Belgian Science Policy Office under the contract CE/10/03. The different databases were kindly provided by the MARS-STAT project of the *AGRIFISH* ***Unit of the Joint Research Center (Ispra, Italy).***

References

Chatfield, C., 2003. *The analysis of time series: an introduction*, 5th edition, pp 66-76.