CROP MASKING – NEEDS FOR THE MARS CROP YIELD FORECASTING SYSTEM

B. Baruth [a, *], L. Kucera[b]

[a] EC - Joint Research Centre, IPSC, Agrifish –Unit, 21027 Ispra (Italy)
[b] GISAT, Charkovska 7, 101 00 Praha 10, Czech Republic

Commission VIII, WG VIII/10

KEY WORDS: Agriculture, Land Cover , GIS, Remote Sensing

ABSTRACT:

The objective of the Mars Crop Yield Forecasting Systems is to provide precise, scientific, traceable independent and timely forecasts for the main crops yields at EU level. The forecasts and analysis are used since 2001 as a benchmark by analysts from DG – Agriculture and Rural Development in charge of food balance estimates. Core of the Crop Yield Forecasting System is a spatialized crop growth models that works on simulation units with assumed homogenous growth conditions. These results are aggregated at regional level to be used as yield predictors. Within this process the usage of crop masks for an accurate aggregation is envisaged. Therefore specific crop masks needs have been identified and crop masks tested.

1. INTRODUCTION

The objective of the Mars Crop Yield Forecasting System (MCYFS) is to provide precise, scientific, traceable independent and timely forecasts for the main crops yields at EU level. The forecasts and analysis are used since 2001 as a benchmark by analysts from DG – Agriculture and Rural Development in charge of food balance estimates. The MARS Bulletin with the forecast results and additional analysis is published, nearly monthly, on paper and on an internet site: http://mars.jrc.it/stats/bulletin with public access.
Core of the Crop Yield Forecasting System is a spatialized crop growth model that requires input concerning the area cropped to ensure a proper aggregation of the forecasted results.
Currently this is assessed for the crop growth model via the soil suitability for the simulation unit. Aggregation to higher administrative levels is directly based on area statistics. As the currently used soil suitability doesn't reflect properly the crop pattern, efforts were made to investigate the use of different crop masks.
In addition the system is supported by the use of Remote Sensing data, namely SPOT-VEGETATION, NOAA-AVHRR, MSG-SEVIRI and MODIS TERRA. In this context crop masks are needed for a better exploitation of the mixed signal due to the coarse resolution. Moreover crop masks are used by the experts of the team to focus the analysis.
These different applications request different crop masks in terms of stratification, spatial resolution and update frequency.

2. CROP MASKS

The actual crop pattern results from natural factors, technology and market conditions. Whereas technology and market conditions are difficult to catch, natural factors can be characterized and thereof a maximum cropping area defined. The interaction of these factors and the resulting crop acreage is reflected in the statistics which are in general based on administrative units. As a result crop mask creation can be a process of disaggregating crop area statistics to spatial units suitable for specific crop production derived from biophysical conditions. This process is often supported by land use / land cover information derived from Remote Sensing (e.g. Corine Land Cover 2000, EEA, 2000) or area frame surveys, like LUCAS (EEA, 2006). The finally obtained crop mask can indicate the presence or the acreage of a crop within a spatial unit (e.g. pixel of 1 km * 1 km or landscape units).

2.1 Crop masks for remote sensing

In the frame of the MCYFS mainly sensors with a coarse spatial resolution (1 km) like NOAA-AVHRR and SPOT VGT are used. As a consequence mixed signals are derived that need to be unmixed to better reflect the crop behaviour in the signal. As work is done on profile analysis of time series the crop masks have to ensure enough spatial and temporal aggregation to capture the changing land cover pattern over time, e.g. crop rotation scheme. Typically a crop mask used within the remote sensing work of MARS STAT should indicate the presence or non presence of non-irrigated arable land and crop present, ideally expressed as a percentage of a 1 km resolution map.

2.2 Crop masks in the frame of CGMS

For the crop growth model which is used in the frame of the MCYFS crop masks can be used for the aggregation process of simulated crop growth parameters to derive yields and production estimates for administrative regions. For a better understanding of the crop masks requirements the aggregation method is explained in the next section.

* Corresponding author.

2.3 CGMS aggregation approach. The crop growth model WOFOST (Diepen, 1989) which is used within CGMS is a point model. In order to apply this model throughout Europe regions have to be identified where homogenous crop growth conditions can be assumed. This is done by taking into account climatic conditions and soil properties.

Therefore the soil mapping units (SMU) of the Soil Geographical Database of Europe (SGDBE) are intersected with the MARS climatic grid cells, which have a resolution of 50 km by 50 km. This intersection called Elementary Mapping Unit (EMU) is the smallest simulation unit for which results are produced by the CGM – system. Based on these simulation units the obtained results are aggregated to administrative regions, using acreage statistics, so that they can be used as regional yield predictors (JRC, MARS STAT, 2004).

This aggregation is performed in several steps. Firstly the soil mapping unit which is delineated in the soil map, consists of one or more soil types. Qualitative and quantitative attributes are only given for these soil types. The composition of these soil types (called soil typological unit, STU) that form the soil mapping unit are given in percentage. For these STUs an assessment is performed whether they are suitable or not for crop growth to determine the suitable crop growth area within the SMU. The following parameters are used to assess the crop specific soil suitability: slope, texture, agriculture limiting phase, rooting depth, drainage, salinity and alkalinity. But as a result the selection based on these parameters led to a general overestimation of cultivated areas for the majority of the crops and to a general underestimation of the cultivated area for maize and root crops (European Commission, 2006). As a consequence only very basic soil suitability rules are used in order to exclude non-suitable soils like rocks or sealed areas.

If once the suitable area for possible crop growth is established within the EMU, the simulation result of each EMU is multiplied by the EMU area suitable divided by the sum of all the area suitable in the selected statistical region, in this case NUTS 2 regions (NUTS – Nomenclature of territorial units for statistics).
For the next aggregation level to NUTS 1 a weighted sum is calculated based on NUTS 2 area statistics. This implies that crop masks to be used in the aggregation process from the EMU to the selected administrative unit (NUTS 2) have to be based at least on NUTS 3 statistics or making use of a finer spatial disaggregation by means of enhanced statistical procedures like it was performed for the CAPRI-Dynaspat land cover masks (Koeble 2006).

2.4 Crop masks for the analysis and derivation of C-indicators Within the analysis and interpretation of results from the crop growth models experts concentrate their analysis on the main production areas. In this context actual crop masks are of great importance as there are shifts in production areas from year to year. Furthermore a big part of the analysis is based on indicators weighted by the occurrence of the relevant crop (respectively non irrigated arable land) within a predefined spatial unit (in general NUTS 2), the so called C-indicators (Genovese, 2001).

3. MARS CROP MASK APPROACH

Due to the different crop masks purposes within the MCYFS it is clear that there is not a unique solution in terms of thematic stratification and spatial resolution that can solve all the requirements.
For the aggregation process, ideally a European wide land-use map with yearly updates showing the crop pattern based on at least NUTS 3 statistics should be used. As this is not realistic for the time being a simple solution which is easy to update has been chosen to produce the crop masks in order to facilitate the analysts work.

3.1 Methodology

Corine Land Cover 2000 and Global Land Cover 2000 (Bartholome & Belward, 2005) are used to spatially distribute official statistical data about crop acreages into CGMS grids with 50 km and 25 km resolution.

The processing includes the following steps:

- compilation of arable land mask using selected CLC/GLC data
- GIS overlay n°1: NUTS regions vs. arable land mask
- calculation of the total area of the arable land in each NUTS region
- calculation of crop ratios for each NUTS region: $R_i=C_i/NA_i$ where C is crop area and NA is total area of arable land in each NUTS region
- GIS overlay n°2: CGMS 25km/50km grid vs. GIS overlay n°1
- calculation of crop areas in each grid: $GA_j=\sum(R_i*GNA_i)$ where R_i is crop ratio for each NUTS region and GNA_i is the area of arable land from NUTS region falling into selected grid
- based on these crop areas per grid re-distribution of statistical crop acreage data.

Homogeneous distribution of each crop over the arable land in each NUTS region is assumed in this approach. The smaller the NUTS region areas (i.e. more detailed crop statistical data) are, the more precise results may be expected.

Based on the availability of input data the following area was covered: EU25 countries plus Bulgaria, Norway, Switzerland, Romania and Turkey. The following crops were selected: soft wheat, durum wheat, wheat total, barley, rye, grain maize, rape, sunflower, potato, sugar beet and rice.

3.2 Data sources

3.2.1 Land cover data. An arable land mask is needed to describe the spatial distribution of arable land inside the NUTS regions and CGMS grids. CLC2000 database represents the most comprehensive and homogeneous data layer to be used for this purpose. Nevertheless CLC2000 data do not cover the entire CGMS area where the individual crop masks are targeted. Therefore, the Global Land Cover 2000 data were used for Norway, Switzerland and Turkey to fill the gaps in the CLC2000 database.
Based on comparisons between EUROSTAT and FAO acreage statistics it was decided to choose the following combination of CLC classes to generate arable land mask:

Arable (except rice): CLC 2.1.1 (non-irrigated arable land) + 2.1.2 (permanently irrigated land) + 2.4.1(Annual crops associated with permanent crops) + 2.4.4 (Agro-forestry areas).

The only exception from this rule is the NUTS2 region ES11 in Spain, because none of the classes where mapped, but the

EUROSTAT statistics show more than 300 000 ha of arable land. It was decided to use CLC 2.4.2 class (complex cultivation pattern) as arable land mask in this region.
For the countries covered by GLC the class ‘cultivated and managed areas’ has been selected.
This combination of classes ensures the best match between the statistics and the derived area from the CLC, but nevertheless they are partly large discrepancies between the estimated areas in the different data sources (Table 1).
Table 1: Differences in percentage between arable land in ha derived from EUROSTAT statistics and the area derived from Corine Land Cover class 2.1.*.

Differences (%) between EUROSTAT arable land and CLC 200 arable land (2.1.*)

AT	20,48	FI	26,55	MT	97,81
BE	22,36	FR	15,89	NL	23,34
BG	-15,21	GR	1,20	PL	0,66
CY	-195,53	HU	-10,23	PT	20,79
CZ	-5,43	IE	50,87	RO	13,09
DE	-15,89	IT	3,22	SE	-11,06
DK	-0,65	LT	23,51	SI	36,24
EE	21,39	LU	67,66	SK	-14,47
ES	9,03	LV	50,60	UK	4,96

source: EUROSTAT CRONOS data 2000

3.2.2 Statistical data about crop acreages. Considering NUTS 2 data, the most complete dataset is available through Eurostat Regions Database (http://epp.eurostat.ec.europa.eu /portal/page?_pageid=1335,47078146&_dad=portal&_schema= PORTAL). It gives consistent and reliable data of all main crops for EU 25. However some less important crops are not included (oats, beans, peas). To reduce the influence of yearly variations in crop acreages the time period 2000 – 2003 was used for most countries, as no actual data was available and this time period is close to the GLC 2000 data set.

Table 2: Total number and average area of NUTS2 and NUTS3 regions by country

COUNTRY		NUTS2		NUTS3	
		No. of regions	Average area [ha]	No. of regions	Average area [ha]
AT	AUSTRIA	9	929 540	35	239 703
BE	BELGIUM	12	254 627	43	71 309
BG	BULGARIA	6	1 843 474	28	395 703
CY	CYPRUS	1	924 821	1	924 769
CZ	CZECH REP.	8	982 537	14	563 325
DE	GERMANY	40	889 600	439	81 352
DK	DENMARK	1	4 266 920	15	285 811
EE	ESTONIA	1	4 320 851	5	868 775
ES	SPAIN	18	2 807 922	52	973 046
FI	FINLAND	6	5 434 450	20	1 640 793
FR	FRANCE	26	2 449 652	100	637 932
GR	GREECE	13	1 011 353	51	258 012
HU	HUNGARY	7	1 317 014	20	462 217
IE	IRELAND	2	3 494 358	8	877 193
IT	ITALY	20	1 499 666	103	291 689
LT	LITHUANIA	1	6 459 428	10	648 920
LU	LUXEMBOURG	1	258 722	1	259 588
LV	LATVIA	1	6 428 561	6	1 076 723
MT	MALTA	1	31 583	2	15 794
NL	NETHERLANDS	12	292 465	40	88 083
PL	POLAND	16	1 941 805	45	693 098
PT	PORTUGAL	7	1 315 689	30	307 290
RO	ROMANIA	8	2 967 082	42	566 528
SE	SWEDEN	8	5 472 586	21	2 097 427
SI	SLOVENIA	1	2 022 405	12	168 954
SK	SLOVAKIA	4	1 221 864	8	612 826
UK	UNITED KINGDOM	37	657 515	133	183 707
TR	TURKEY	26	2 964 070	81	952 233
CH	SWITZERLAND	7	570 394	26	153 975
NO	NORWAY	7	4 591 403	19	1 702 265

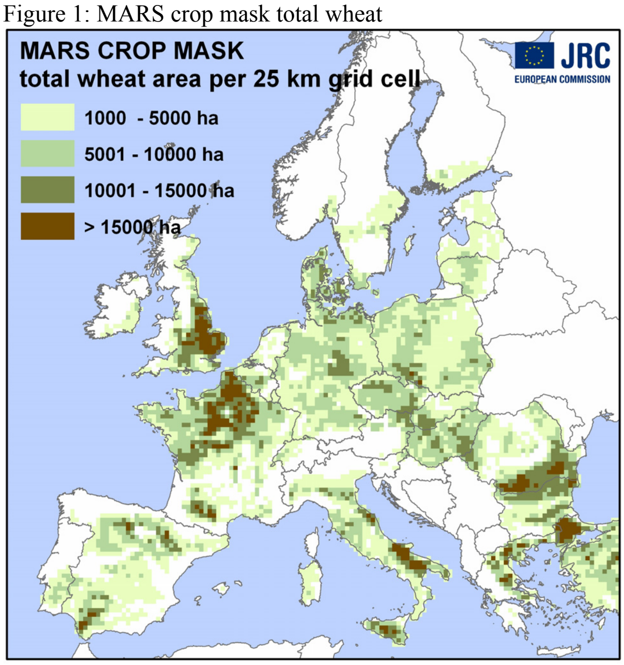
The average size of NUTS2 regions vary from country to country (see Table 2). There are countries where NUTS2 regions correspond to single CGMS 50 km x 50 km grid cells (Netherlands, Belgium) but it is not an exception to have NUTS2 regions covering more than 10 CGMS grid cells (France, Spain, Romania, Sweden, Finland, Norway, Ireland, Turkey). This ratio is even much bigger for countries not using NUTS2 division at all (Denmark, Estonia, Lithuania and Latvia).

For the NUTS 3 data, which is especially important for the MCYFS for the crop mask production the situation is quite diverse not only from country to country but in some cases also within the country (e.g. Germany). Collection and administration of crop statistical data related to NUTS3+ regions are not always the responsibility of the central statistical office (or a similar institution) but this activity can be regionally distributed. Nevertheless a huge amount of crop acreage data at NUTS 3 level or even higher is available on-line but complete spatial and thematic coverage can hardly be achieved without a formal procedure of data collection and harmonization throughout the Member States.
In general the creation based on NUTS3 statistics should be favoured due to the large variation in size of the NUTS2 regions and the decreasing grid size of the MARS system (25 km x 25 km) in the future.

4. RESULTING MARS CROP MASKS

For the crops soft wheat, durum wheat, wheat total, barley, rye, grain maize, rape, sunflower, potato, sugar beet and rice crop masks based on NUTS 2 statistics (average area 2000 – 2003) for 50 km and 25 km grid cells have been created. An example is given in Figure 1. These crop masks are mainly used by the experts to focus the analysis of the crops. So far only for two selected countries crop masks using NUTS 3 statistics have been created.

Figure 1: MARS crop mask total wheat



5. SUMMARY

The selected approach is a basic approach relying only on two information layers – a land cover map and regional acreage statistics. Disaggregation is performed under the basic assumption that the crops are spatial equally distributed within an administrative region. Errors introduced by this assumption are partly compensated if the administrative regions are small enough to reflect also particular situations. An advantage of the approach is the fact that the masks are easy to update if new input data is available and quickly inserted into our system.

As long as a general coverage with NUTS 3 crop acreage statistics can not be achieved the crop masks data will not be used in the CGMS aggregation approach, but used by the experts for a focused analysis and for extraction purposes in generating agricultural specific remote sensing data products.

References

Bartholomé, E., Belward, A.S., 2005. GLC2000: A New Approach to Global Land Cover Mapping From Earth Observation Data *International Journal of Remote Sensing,* 26(9), May 2005, 1959-1977.

Diepen, K., Wolf, J. , Keulen, H. 1989. WOFOST: a simulation model of crop production. In: *Soil use and management, Vol 5. page 16-24.*

European Commission 2006. New soil Information for the MARS Crop Yield Forecasting System. *European Commission EU-report 224999*. Available at: http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/SINFO/index_en.htm.

European Environment Agency 2000. Corine Land Cover Technical Guide – Addendum 2000. *EEA Technical Report No. 40, 105 pp Copenhagen*. Available at: http://reports.eea.europa.eu/.

European Environment Agency 2006. The thematic accuracy of Corine Land Cover 2000 – assessment using LUCAS (land use / cover area frame statistical survey). *EEA Technical Report 07/2006. 85 pp.* Available at: http://reports.eea.europa.eu/.

Genovese, G. et al, 2001. A methodology for a combined use of normalized difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. *Agronomic 21 (2001) 91-111.*

JRC, MARS STAT 2004. Methodology of the Mars Crop Yield Forecasting System. Vol 1- Vol 4. *European Commission EU-report 21291 EN/1-4*. Available at: http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/METAMP/

Koeble, R., Leip, A. , Kempen, M. 2005. CAPRI-GIS link. *CAPRI working paper 04/05. 55 pp*. Available at: http://www.ilr1.uni-bonn.de/agpo/rsrch/dynaspat/publicat.htm