A COMPARISON AND EVALUATION OF PERFORMANCES AMONG CROP YIELD FORECASTING MODELS BASED ON REMOTE SENSING: RESULTS FROM THE GEOLAND OBSERVATORY OF FOOD MONITORING

G. Genovese [a], S. Fritz [a], M. Bettio [a]
[a] EC - Joint Research Centre, IPSC, Agriculture – Unit, 21027 Ispra (Italy)

Commission VIII, WG VIII/10

KEY WORDS: Crop Yield Forecast, Remote Sensing, Modelling

ABSTRACT:

In the context of the GEOLAND EC FP6 project the comparison of different remote sensing based approaches for yield forecasting over large areas in Europe are tested and results inter-compared. In particular the methods tested include the ones in use within the MARS-Crop Yield Forecasting System as the results from the Crop Growth Monitoring System model and vegetation indicators derived from Low Resolution SPOT-VGT and NOAA Images, METEOSAT based yield forecasting and ERS-Scatterometer Crop Performance Index. Performances of the different models were tested in Spain, Belgium and Poland.
The inter-comparisons of the crop yield forecasts were mainly based on the forecasting error obtained from the different approaches based on the Root Mean Square Forecast Error (RMSFE). This error was derived by comparing the predicted yields of the different models with the official yield as from official statistics (EUROSTAT). The comparison of the RMSFE was used to verify the convergence of results from the different models, the reliability of the information, i.e. precision and bias, and its precocity compared to the crop cycle. The results showed that the indicators are able to give reliable information with some differences: remote sensing indicators are more precise and accurate in southern areas (less cloud cover) while in northern areas good results are obtained under the use of better local calibrations of traditional crop yield forecasting systems and/or the use of additional information for instance remote sensing data as inputs into advanced crop modelling systems. Furthermore, in order to take care of the different time series length available, a qualitative indicator called Performance Score (Ps) was introduced. The analysis of the Ps showed that when a long time series of observation is available greater advantages are obtained from RS rather than from more advanced crop models.

1. INTRODUCTION

GEOLAND is an Integrated European Research Project (FP6) of the EC (http://www.gmes-geoland.info/), carried out in the context of GMES to build up a European capacity for Global Monitoring of Environment and Security. Within this project the use of remote sensing is being researched within a semi-operational context for different targets, including studies on agriculture grouped under the Observatory of Food Security and Yield Monitoring (OFM). More specifically, the OFM focuses on developing a global crop yield and crop area forecasting service for food security purposes.
In this context, the analysis described hereafter is an evaluation of the final crop yield forecasts generated during the season by different model-based approaches. These reference models generate indicators using either direct remote sensing observations, mixing them with exogenous data such as meteorological observations, or building more sophisticated crop growth models. The indicators generated in such a way describe different part of the plant biomass or of its cycle during the year, and are then used as predictors in the regression models that produce as outcome the yield forecasts.
The forecasts to be compared were chosen by the GEOLAND partners as the ones generated for the countries of Belgium, Poland and Spain with reference to wheat yield. Wheat has been chosen as the crop of interest. The analysis window for Belgium and Poland was April to August and for Spain March to July; this choice reflects the crop cycle in the mentioned countries and the fact that within the period of the plant cycle the forecasts have a greater value than after the plants have been harvested.
Five project partners provided the wheat yield forecasts (hind cast) heterogeneously in a period ranging from 1998 to 2004; forecasts were produced in a standard approach in which each of them is based on a regression model where the regressors are the indicator/predictors generated by the different models and the true observations are the official EU statistics available for each country (EUROSTAT CRONOS and REGIO DB sources).
The evaluation reported in this paper takes care of different aspects that are expected to be fulfilled by a forecasting approach or system: the global error/uncertainty of the forecasts (precision and accuracy), their precocity coupled with their goodness, the forecasting performance of the systems in extreme years (including bias) and their capacity to keep performances in different areas and provide enough information to feed the statistical/mathematical object used in the forecasting solution. However, the different approaches cause some bias in this evaluation. In fact the different time series of the predictors affect both precision and uncertainty of the prediction. A qualitative approach was then introduced in order to weight for the structural differences of the error indicators: a synthetic Performance Indicator (Ps) was defined weighting the errors indicators according to the "importance", as given by the forecast customers, of the different aspects they refer to. The results are presented and discussed.

2. DATA and METHODOLOGY

The data and models which have been used within the Geoland OFM can be grouped into 4 categories.

- 1st Category: are the forecasts which are not based on remote sensing, they are the benchmark forecasts produced automatically from a Crop Growth Monitoring System

(CGMS). CGMS is a crop growth simulation model based on the WOFOST model (Boogart et *al*., 1998) and in use operationally in the MARS (Monitoring Agriculture with Remote Sensing) System (Lazar et *al*., 2004). The forecasts obtained basing on crop indicators/predictors produced by CGMS were delivered by the project partner Alterra (NL). The CGMS based forecasts can also be simple extrapolation of trend time series according to the an error evaluation based on jack-knife statistical method comparing trend based forecast with crop simulated parameters based forecasts. The approach is described in De Koning et *al.* (1993).

- 2nd Category: includes those forecasts which are derived from indicators/predictors obtained from a modification of the CGMS model by plugging in some way Remote Sensing (RS) data. The modifications and runs of CGMS basing on RS inputs were made by Alterra. In particular, scatterometer-radar data (soil water balance; Wagner et *al.* 1999), and METEOSAT based indicators (radiation and temperatures estimates; Rosema et *al.* 1998) were used.

- 3rd Category: are those forecasts which are based purely on RS indicators (e.g. Scatterometer-NEO; NDVI-VITO; VCI/TCI-IGIK); Dabrowska-Zielinska K et *al*., Wagner et *al*., Eerens et *al*.)

- 4th Category: are those forecasts derived from RS data further transformed according to crop models (assimilation models) such as DMP/fAPAR-VITO Eerens et *al*.), and the EARS-indicators (Rosema et *al.*)

For further information and a detailed description of the indicators the reader can consult the GEOLAND-OFM Methods Compendium Report (GEOLAND 2006)

The indicators/predictors produced by each approach were basically used as in-dependent variables according to a classical regression approach where:
(Observed)Yield= f (RS indicator, METEO data) where in most of the cases the relation was assumed to be linear, for instance:
(Observed)Yield= a+b RSind + error. Where the error is a white noise. A similar approach is described in Genovese et *al*. and De Koning et *al.*
Once the parameters were estimated the models were used to produce the forecasts. Due to the fact that it was necessary to use the major part of the time series for calibration, in some cases the forecast could be computed only for few years. Therefore, often only a short time series of the forecast was available for testing the model in an ‘operational’ condition. This was taken into account in the forecast error calculation.

2.1 Category 1: Forecasts based on the Crop Growth Monitoring System:

In this paper these forecast are referred with the suffix: CGMS, or CGMS-EUR, and CGMS enhanced). The difference among the EUR and the “enhanced” versions of the model is that the latter was improved with a re-calibration using more recent observations.
The standard CGMS indicators provided by ALTERRA (NL) (see Boogart at al, 2002) where:
CGMS-WLY_STORAGE – Uses as predictors the simulations of the total weight of the storage organs (grains) for wheat under water limited conditions (rainfed);
CGMS-WLY_BIOMASS – Uses as predictors the simulations of the total weight of the above ground biomass for wheat under water limited conditions (rainfed);
CGMS-PY_STORAGE - Uses as predictors the simulations of the total weight of the storage organs (grains) for wheat under potential conditions (non-water limited);
CGMS-PY_BIOMASS - Uses as predictors the simulations of the total weight of the above ground biomass for wheat under potential conditions (non-water limited).
An enhanced CGMS was prepared using in the study areas more updated and refined data for calibration (see GISAT, 2003). The resulting forecasts were referred as:
CGMS_ENH-WLY_STORAGE
CGMS_ENH-WLY_BIOMASS
CGMS_ENH-PY_STORAGE
CGMS_ENH-PY_BIOMASS

2.2 Category 2: Forecasts provided based on Crop Growth Model modified with RS data

The related forecasts are referred here with the suffix CGMS SCAT, when scatterometer-radar data where introduced in the model, and CGMS EARS indicators, since they when based on METEOSAT indicators. The Scatterometer based water balance data were supplied by the Univ of Vienna and NEO (NL).
The water balance of the CGMS crop growth model (see Boogart et al, 2000) was substituted with an approach of estimation of water balance based on data from the **ERS-Scatterometer**). The corresponding forecasts in this contribution are:
CGMS_SCAT-WLY_STORAGE
CGMS_SCAT-WLY_BIOMASS

Meteorological variables (radiation, temperature and precipitation) derived from **METEOSAT** were used as an alternative data set into CGMS (see De Wit et al, 2004). The integration in CGMS and the calculation of the new predictors was performed by ALTERRA (NL). The corresponding forecasts are referred in this contribution as

CGMS_EARS-WLY_STORAGE
CGMS_EARS-WLY_BIOMASS
CGMS_EARS-PY_STORAGE
CGMS_EARS-PY_BIOMASS

2.3 Category 3: Forecasts based on remote sensing SPOT-VGT and NOAA-AVHRR data.

The forecasts where developed and provided by IGIK (PL). The assumption of the approach (see Dabrowska et al, 2002) is that the maximum amount of vegetation is developed in years with optimal weather conditions. Conversely, minimum vegetation amount develops in years with extremely unfavourable weather, which suppresses vegetation growth. Therefore, the absolute maximum and minimum of NDVI are calculated from several years. The different stages of crop growth are sensitive to crop development conditions, which are monitored by remote sensing. Developed indices are based on the reflection properties of vegetation in the visible and infrared spectrum and to the radiative temperature of crop. The dekad number in each year has been expressed in the value of accumulated NDVI representing each of the development stage in the given year. For these periods it is essential to deduce soil moisture using developed method of calculating evapotranspiration and

Vegetation Condition Indexes (VCI) and Temperature Condition Indexes (TCI) indices.
The corresponding forecasts are named:
PTVCI_from_SPOT (not available in for Belgium)
PTAVCI_from_SPOT (not available for Spain and Belgium)
MTCI from NOAA (not for available for Belgium and Poland)
VCIAVG_average_of_VCI_from_NOAA_model16km2 (not available for Spain and Belgium)

2.4 Category 3: Forecasts based on Remote Sensing: ERS-SCATTEROMETER.

Crop water limitation is derived from the Soil Water Index (SWI). The methodological approach to retrieve soil moisture information is based on a change detection algorithm exploiting the short revisit capabilities of the ERS scatterometer. The ERS Scatterometer is a single frequency (5.3 GHz) single polarised (VV) low resolution active band. Forecasts of this type are referred as NEO (NL) (see Wagner, 1999).

2.5 Category 4: Forecasts based or Remote Sensing models and/or coupled with RS based on Spot-VGT

The list of indicators evaluated includes: NDVI; DMP; Improved DMP (with improved fAPAR); the Vegetation productivity indicators (see Eerens et *al*, 2000). These indicators were provided by the VITO-TAP Centre (B).

Table 1: Description of Vito indicators

Code	Description
VGT_DMP_No-unmixing	Overall mean DMP of all pixels n for the region of interest
VGT_DMP_Weight-means_50%	Weighted mean DMP of all pixels m with fraction f > 0.5
VGT_DMP-improved_fAPAR_No-unmixing	Overall mean DMP (calculated with the improved fAPAR derived by means of a neural network on S10 images) of all pixels n for the region of interest
VGT_DMP-improved_fAPAR_Weight-means_50%	Weighted mean DMP (calculated with the improved fAPAR derived by means of a neural network on S10 images) of all pixels m with fraction f > 0.5
VGT_Hist_classified_DMP_No-unmixing	Overall mean of all pixels n for the region of interest. (The historical classified DMP is calculated on an analogue way as the VPI. Instead of using NDVI, DMP is used.)

2. 6 Category 4: Forecasts based on model coupled with RS data, based on METEOSAT.

The EARS (NL) Crop Yield Forecasting System (CYFS) is based on satellite derived radiation and actual evapotranspiration data from the Energy and Water Balance Monitoring System (EWBMS). These data enter into a crop growth model, which generates crop yields. Underlying theories are those of radiation and energy transfer at the surface and in the atmosphere. More details on the crop yield modelling EARS approach can be found in Rosema et al, 1998. Three indicators were used; the corresponding forecasts are named:

- EARS_actual yield: water limited, simulated with radiation and relative (actual/potential) evapotranspiration as observed;
- EARS_NON_WATER_LIMITED_YIELD: simulated with radiation as observed and relative evapotranspiration is set to 1 (so for: potential evapotranspiration);
- EARS_act_yield_rel2_non_water_limited_yield: called relative yield, is the ratio of the two previous indicators.

3. FORECAST ERRORS AND PERFORMANCE SCORES INDICATORS

The forecasting power of the OFM models was evaluated with respect to the following criteria:
A – the global error/uncertainty of the forecasts (precision and accuracy);
B – the precocity (how early in the year the forecast is made) of the forecasts coupled with its goodness;
C – the forecasting performance of the systems in extreme years including bias;
D – the capacity of the system to keep performances in different areas and provide enough information to feed the statistical/mathematical object used in the forecasting solution.
The evaluation of the listed criteria leads to a ranking of the performances of the different systems and the evaluation of the significance of the differences between one system and the other. A similar approach was undertaken in previous studies (see Genovese et al., 2004).
Assuming that the final user of the forecasts has as quadratic reference loss function (i.e. the higher the error in both directions, the higher the impact on his decision) the error indicator selected was the root mean square forecast error (RMSFE). Defining the forecast error as the difference between the forecast and the true value of the variable of interest:

$$e_t = \hat{Y}_t - Y_t\,,$$

where Y_t is the true value of the variable at time point t and $\hat{Y}_t$ is its forecast, the RMSFE is an aggregation over a given period of time consisting of T points according to the following formula:

$$\sqrt{\sum_{t=1}^{T} e_t^2 \Big/ T}\,.$$

The RMSFE takes only positive values, and gives an average size of forecast error over the given period.

Three types of comparison have been performed based on on the RMSFE:

- Comparison of overall error, computed by averaging over all months and all years of the interested period separately for each forecasting system;
- Comparison of error across month, averaging for each forecasting system over the figures for the same months in all the years of the interested period. This analysis has been carried out in order to assess the performance of each forecasting system across the whole crop growing season, quantifying for each of them its risk of mistakes and reliability along the forecasting year;

- Comparison of error across years, averaging for each forecasting system over the figures for all months of each year. The purpose of this comparison is to determine whether there are years in which one system performs significantly better than the others, and to test their improvement and refinement in the last years.

Furthermore, the forecast performances have been evaluated by comparing the same error indicators with forecasts obtained with simple linear trends.

Finally, a synthetic Performance score (Ps) has been calculated. The final purpose of this synthetic Performance indicator is to score the results obtained taking into account "uncertainty" given by the heterogeneity of the initial data. A similar approach has been used in soil science to evaluate pedo-tranfer rules models (Donatelli *et al*., 2004). The advantage of such a synthetic performance indicator is to be able to summarize aspects related to forecasting (precision, accuracy, timeliness, consistency and cost) into one single information, taking into account the final user view. This last point is also a drawback as weights are given to the different component of the synthetic indicator subjectively. In this context the weight are calculated according to the number of sub-components indicated by the users as important for the evaluation.

The Ps is a qualitative indicator defined as the count of the results of the following criteria:

- *Criteria 1*: Length of time series: N° of available years >5. Is equal to 1 if the time series of forecasts available includes at least five years, 0 otherwise;
- *Criteria 2*: Overall RMSFE≤0,2 (i.e. less than 2 quintals/ha). Is equal to 1 if the overall (averaged over years and dekades) RMSFE is ≤0,2, 0 otherwise. The same criterion was repeated to analyse forecasts results at the beginning of the forecasting season and at the end (*Criteria 3*);
- *Criteria 4*: Stability of the forecast: April forecast > July forecast. Is equal to one if the RMSFE is higher in April than in July, 0 otherwise. The same criterion was also analysed taking into account inter-forecast variability;
- *Criteria 5*: RMSFE ≤0,2 in the extremes bad years. Is equal to 1 if the RMSFE averaged over dekades of the extreme year, 0 otherwise;
- *Criteria 6*: At least in one year in the period analyzed 50% dekades are better than trend; is equal to 1 if there is at least one year in the period under study in which the forecast are at least 50% of times better than a simple trend extrapolation, 0 otherwise. This criterion was evaluated also at sub-regional level.

The number of criteria to evaluate was discussed and decided together with final customers, which are the operational users of the forecasts. The number of subcriteria associated to the criteria class determined the importance (weight) in the final calculation.

Only exception concerned the length of the time series that counts for nearly 30% in the computation of the final score, being this requirement a fundamental in order to draw conclusion about the performance of each indicator. The weight of each criterion in the calculation of the Ps is reported in the following table:

Table 2: Description of Criteria used for the Ps

CRITERIA	Final weight %
1- Lenght of time series	28%
2 – Overall RMSFE<=0.2	8%
3- Timeliness and accuracy of forecast	17%
4- Stability of forecasts	17%
5- Extreme years	14%
6- Vs trend	17%

Calculation of scores per indicator/predictor
The scores were summed up for each indicator, giving a sum (Sc) ranging from 0 to a maximum of 36 (27 criteria counting one each except for the criterion measuring the length of the time series, counting 10).

Calculation of penalties
A decision had to be taken in case the Criteria were not applicable to a predictor or the subcriteria summed 0. For instance a method gave predictors only for few years and only on one test area. In such cases the method had to be evaluated as operationally limited. Penalties were attributed in the following way:

Pe = Sc * Mc/6
Pe: penalty
Sc: sum of scores compliant with criteria
Mc: number of missing criteria or number of criteria giving 0 (max possible 6).

The Mc ranges from 0, when all of the criteria scores 1, to the maximum of 6 when no criteria are compliant (missing or 0).

Calculation of the final score x indicator/predictor
A final score was then obtained for each predictor according to the following:

Ps = (Sc-Pe)/Ac
Ps: Performance Score
Pe: penalty
Sc: sum of scores compliant with criteria
Mc: number of missing criteria
Ac: Total number of accounted sub criteria (i.e. where either 0 or 1 was available)

The Ps ranges between 0 and 1. The higher the Ps the higher the overall performance of the predictor. In the extreme situation that a predictor has an Ac of 0 the Ps is undetermined. In fact this is obtained when no one of the 6 criteria was available for evaluation.

4. SOME RESULTS

Complete results are available in the GEOLAND technical report on the yield inter-comparison study (G. Genovese et.al, 2006). Before reading them it is recommended to look at them as a guide to take decisions on the approaches not for an eventual absolute conclusions. Table 3 reports the results of the overall RMSFE obtained by country for each model, and their average over countries (mean).

Table 3: RMSFE for BE, PL, ES and overall mean (error is expressed in t/ha).

	BE	PL	ES	Mean
CGMS_EARS-WLY_BIOMASS	0.3	0.3	0.3	0.3
CGMS_EARS-WLY_STORAGE	0.3	0.3	0.2	0.3
CGMS_EARS-PY_BIOMASS	0.3	0.3	0.3	0.3
CGMS_EARS-PY_STORAGE	0.3	0.3	0.3	0.3

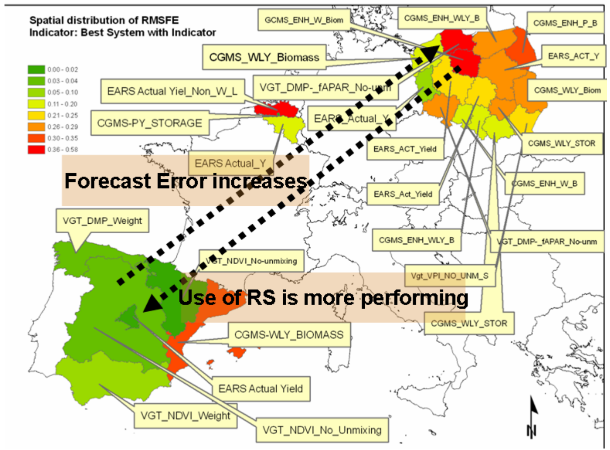
PTVCI_from_SPOT		0.4	0.3	0.4
VGT_Hist_classified_DMP_Weight-means_50%_MAvg_Sum	0.5	0.5	0.1	0.4
PTAVCI_from_SPOT		0.4		0.4
CGMS_ENH-WLY_STORAGE	0.5	0.3	0.4	0.4
CGMS_ENH-WLY_BIOMASS	0.5	0.3	0.3	0.4
VGT_NDVI_Weight-means_50%_MAvg_Sum	0.5	0.4	0.4	0.4
CGMS-WLY_STORAGE	0.5	0.3	0.4	0.4
VGT_NDVI_No-unmixing_MAvg_Max	0.47	0.74	0.1	0.4
CGMS-WLY_BIOMASS	0.56	0.37	0.4	0.4
CGMS_ENH-PY_STORAGE	0.51	0.33	0.5	0.4
CGMS_ENH-PY_BIOMASS	0.55	0.29	0.5	0.5
VGT_DMP-improved_fAPAR_No-unmixing_MAvg_Max	0.6	0.4	0.4	0.5
VGT_DMP-improved_fAPAR_Weight-means_50%_MAvg_Sum	0.5	0.5	0.3	0.5
VCIAVG_average_of_VCI_from_NOAA_model16km2				0.5
CGMS-PY_STORAGE	0.5	0.4	0.5	0.5
EARS_act_yield_rel2_non_water_limited_yield	0.5	0.4	0.5	0.5
VGT_VPI_No-unmixing_Sum	0.5	0.5	0.5	0.5
CGMS-PY_BIOMASS	0.6	0.4	0.5	0.5
VGT_DMP_Weight-means_50%_MAvg_Sum	0.5	0.5	0.4	0.5
VGT_VPI_No-unmixing_MAvg_Max	0.7	0.4	0.3	0.5
EARS_NON_WATER_LIMITED_YIELD	0.6	0.5	0.4	0.5
VGT_Hist_classified_DMP_No-unmixing_MAvg_Max	0.8	0.4	0.3	0.5
EARS_actual_yield	0.5	0.4	0.6	0.5
CGMS_SCAT-WLY_STORAGE	0.5	0.4	0.7	0.5
CGMS_SCAT-WLY_BIOMASS	0.7	0.4	0.6	0.6
VGT_DMP_No-unmixing_MAvg_Max	0.5	0.5	0.3	0.7
NEO	1.5	1.1	0.5	1.0
MTCI_from_NOAA			3.4	

Only in few cases the RMSFE was < 0,2 while in several cases it was <= 0,3. CGMS based results appeared as a whole better than the ones based on RS.

From the regional analysis conducted at NUTS2 level the following results emerged: a general trend (Fig.1) is observable at sub-regional level, i.e. an increasing error patterns (south-north) coupled with the use of the best models.

Figure 1: Spatial distribution of RMSFE of best forecasting indicator (error is expressed in t/ha) per NUTS region, and names of best systems.

A statistical test on the differences among the error distributions of the different forecast has revealed that only in few cases the methods gives significantly different forecasts. Again this analysis is affected by the different populations confronted in terms of number of cases and years covered.
An analysis on the extreme years results in the conclusion that all methods in all regions (with different error results) underestimated in good years and overestimated in bad years. In the hypothesis the reference statistics are of good quality, this should then be a matter of reflection for a choice of future crop forecasting systems.



From the Ps analysis the following results emerged by category:

- 1 In the first best 10 Ps (RMSE Criteria = 0.2) 5 are within the CGMS pure model category. Within these 6 the Potential Biomass is excluded and the CGMS enhanced (re-calibrated) water limited biomass shows the best Ps. The absolute best is an indicator of the RS-pure group: VCIAVG_average_of_VCI_from_NOAA_model16km2, followed by PTVCI_from_SPOT in second position. Within the category of RS-models only the DMP (Hist Class) based on VGT seems of interest. Thus according to the requests of the user (to have a method giving an error on wheat not higher than 2 quintals) the CGMS models seems more attractive basing on the available data in the OFM
- 2 When relaxing the RMSE criteria to 0.3 (3 quintals of error) CGMS pure seems still the best strategy (7 bests in the first 10). On the remote sensing side VCIAVG_average_of_VCI_from_NOAA_model16km2 and the PTVCI-SPOT approach are still the best in absolute. This conclusion can be interesting in the areas where it is not possible to use different approaches than just remote sensing data (lack of other data) and the user can accept an error of 3 quintals per hectare fault of better information.
- 3 Conclusions would change dramatically in case longer time series of forecasts would be simulated on the available methods. All the first 10 positions would be occupied by indicators from RS: 4 of them would be on VGT data (NDVI, VPI), being the best in

absolute the one based on MAX Ndvi from VGT. In the category RS-Model we would find now 4 others best placed among which the second absolute which would be a DMP (Hist Classif.). Although these results come from a simulation we can see the advantages of having longer time series in calibrating stable models. **Continuity in the Remote sensing missions seems fundamental to consolidate the usefulmess of the data.**

From this study there is an indication that for certain regions in some months early in the year a remote sensing indicator might deliver better results than the CGMS model. From our regional analysis it is evident that when RS is performing well, the absolute level of error is lower than what obtainable with crop models (CGMS) (see fig.1). Furthermore, there is an indication that Remote Sensing data can support the CGMS-system, and potentially improve the forecasts when included into the crop model (CGMS). For instance, there is an indication that by incorporating data from METEOSAT into the CGMS an improvement might be achieved. Finally, the data available so far indicates that the enhanced CGMS system with a better crop calendar and certain modifications performs better.

A number of more specific conclusions can be drawn: first of all, the incorporation of the Meteosat data delivered by EARS into CGMS, a Crop Growth Simulation Model existing (and running operationally in MARS), shows promising results and performs very well in the 2-3 year time period considered.
The analysis purely based on the EARS data undertaken for a 3 year period (2002 – 2004) does give reasonable results.
The indicators PTVCI and PTAVCI from IGIK show promising results in particular for Poland: the forecasts for these two indicators show nice convergence, the forecast later on in the year closer to the harvest tends to be better than the ones early in the year.
The forecasts based on the Vito indicators show in particular promising results for Spain, however forecasts tend to get worse over the season for Spain and for Belgium. This implies that the Vito indicators based on NDVI might give better forecasts early than later in the year.
The NUTS 1 (and NUTS 2 for Poland) analysis has revealed that there is a relatively high variation of the RMSFE over the different NUTS regions. Furthermore, the regional analysis has shown that in South-western regions (Spain) the error performance of Remotely Sensed (RS) derived indicators such as SPOT-VGT is very good, while more complex models which are not using RS at all (e.g. CGMS-ENH) or those models which incorporate information from RS (e.g. EARS models) perform better in North-eastern regions.

5. CONCLUSIONS AND RECOMMENDATIONS

- The incorporation of RS data into Crop Growth Simulation Model can be a valid and performing strategy;
- The indicators based on NDVI - PTVCI and PTAVCI (IGIK) - show promising results with forecast at the end of the crop cycle more precise than earlier in the year; in SPAIN simple NDVI based approaches (VITO) are well performing, also early in the season;
- In terms of behaviour in extreme years there is currently no clear pattern visible as data availability is scarce. Nevertheless RS based indicators did perform better in some situations (NDVI, SPAIN);
- the regional analysis has revealed that in South-western regions (Spain) the error of Remotely Sensed (RS) derived indicators such as VGT is the lowest, while more complex models which are not using RS at all (e.g. CGMS-ENH) or models which incorporate information from RS (e.g. EARS models) perform better in North-eastern regions. The complexity of the approach (so possibly the costs) increases with a northern direction .Nevertheless, moving from southern regions northwards the error of the forecast increases as well as the associated variability (uncertainty);
- Based on the Performance score (Ps) analysis and according to the user requests to have a method giving an error on wheat not higher than 2 quintals/ha, the CGMS models seem more attractive among the available data in the OFM;
- Based on the Performance score (Ps) analysis, in case the error is relaxed to 3 quintals/ha, remote sensing based results appear in many situations attractive;
- Based on the Performance score (Ps) analysis following the criteria of 2 quintals error, when long time series of data are available from RS (see indicator PTVCI - based on SPOT) the corresponding forecasts are better performing than CGMS. Therefore continuity in the Remote sensing missions is a MUST to have adequate time series to support a performing crop yield forecasting system.

As a further step a cost/benefit and a sustainability analysis of the forecasting methods should be added. The sustainability of the methods implies as well discussions on the durability of satellite platforms and missions.

6. REFERENCES

Boogaard, H.L., Diepen, C.A. van, Savin, Y. 2000. Monitoring drought affected crop yields based on ERS-scatterometer data; Exploration of possibilities to integrate ERS-scatterometer derived soil moisture into the CGMS crop model for a Russian-Ukrainian study area. Report 009, Alterra, Green World Research, Wageningen, The Netherlands, pp 96.

Boogaard, H.L., Diepen, C.A. van, Rötter, R.P., Cabrera, J.M.C.A., Laar, H.H. van, 1998. WOFOST 7.1; user's guide for the WOFOST 7.1 crop growth simulation model and WOFOST Control Center 1.5. Technical Document 52, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 144.

Dabrowska-Zielinska K., Kogan.F., Ciolkosz A., Gruszczynska M., Kowalik W., 2002; Modelling of crop growth conditions and crop yield in Poland using AVHRR based indices, International Journal of Remote Sensing (Elsevier), Vol 23, No 6, pp. 1109-1123.

De Koning G. H. J., Jansen M.J.W., Boons-Prins E. R., van Diepen C. A., Penning de Vries F. W. T., 1993, Crop growth simulation and statistical validation for regional yield forecasting across the European Community, Simulations Reports CABO-TT, n° 31, Wageningen Agricultural University, CABO-DLO, J.R.C.

De Wit, A.J.W., H.L. Boogaard and C.A. van Diepen, 2004. Using NOAA-AVHRR estimates of land surface temperature for regional agrometeorological modeling. Int. J. of Applied Remote Sensing and Geoinformation, in press.

Donatelli M., Acutis M., Nemes A., Wosten H, 2004, Integrated Indexes for Pedotransfer Function Evaluation, in Development of Pedontranfer Functions in Soil Hydrology, chap.2, editors Pachepsky Ya. & Rawls W.J., Elsevier

Eerens H, Wouters K, Buffet D, Oger R, Dehem D & Tychon B, 2000. Use of 1km²-resolution imagery in the Belgian Crop Growth Monitoring System (B-CGMS), Proc. Conf. VEGETATION-2000, 3-6 April, 2000, Belgirate, Italy, p. 361-366.

GEOLAND, 2006 "Yield Inter-comparison study" technical report. http://www.gmes-geoland.info/index.php

Genovese G., Bettio M., Orlandi S., Petrakos M., Stavropoulos P., Tassoula I., Glossioti M., Boogard H., 2004. Statistical Data Collection Processing and Analysis - Methodology of the MARS Crop Yield Forecasting System VOL. 4, Publication Office of the EC – ISBN 92-894-8176-5 EUR 21291 EN/4, PP. 93.

G. Genovese, M. Bettio, S. Fritz - Report on Yield Inter-comparison Study - EUR 22622 EN, PB/2006/IPSC/4237, 2006 http://www.gmes-geoland.info/index.php

Lazar C., Genovese G., Orlandi S., Van der Goot E., Supit I., Boogard H., Van Diepen K., Schapendonk A.H.C.M, 2004. Agrometeorological Data Collection, Processing and Aanalysis – Methodology of the MARS Crop Yield Forecasting System VOL.2, Publication Office of the EC – ISBN 92-894-8181-1, EUR 21291 EN/2, PP. 98.

Rosema, A., R. Roebeling and D. Kashasha (1998) "Using Meteosat for Water Budget Monitoring and Crop Early Warning", in *Agrometeorological Applications for Regional Crop Monitoring and Production Assessment*, D. Rijks, J.M. Terres and P.Vossen eds, Accounts of the EU Support Group on Agrometeorology (SUGRAM), EUR.17735 EN.

Wagner, W., G. Lemoine, H. Rott (1999) A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data, Remote Sens. Environ., Vol. 70, pp.191-207.

7. ACKNOWLEDGEMENT

The authors would like to acknowledge the GEOLAND OFM partners for the availability of the forecasts.