Weather induced uncertainty in the European Crop Growth Monitoring System

A.J.W. de Wit, H.L. Boogaard and C.A. van Diepen

Alterra Green World Research
P.O.Box 47
6700 AA Wageningen
The Netherlands
allard.dewit@wur.nl

ABSTRACT

Agrometeorological models that simulate crop growth are used operationally in many parts of the world for monitoring the effect of weather conditions on crop growth and for predicting crop yields on regional to continental scales. These crop monitoring systems typically operate at resolutions of 10 to 100 km. Finding meteorological input data and carrying out model calibrations at these scales becomes a challeging task. Particularly meteorological data is subject to large uncertainty because often interpolations are necessary from scattered weather stations or grid points of low resolution numerical weather prediction models. This paper describes the result of a study on the effect of uncertainty in the meteorological data that are used as input in the European Crop Growth Monitoring System (CGMS). The interpolated radiation and precipitation values of the CGMS 50x50 km grid were replaced by estimates of the highly accurate and high resolution ELDAS radiation and precipitation databases. Next, CGMS was applied to a test site consisting of France and Germany and the differences in the CGMS crop model output were evaluated.

The results demonstrate that there is large uncertainty in the CGMS radiation and precipitation estimates and that this translates into a large uncertainty in the CGMS crop simulations at the level of 50x50 km grids, particularly in the water-limited simulations. However, when the CGMS output is aggregated to larger spatial regions (NUTS 1) the uncertainty decreases. Since CGMS results are currently used only at national or subnational level, the influence of the uncertainty on the yield forecast is probably limited. However, it demonstrates that CGMS output should not be used for making decisions on small areas (i.e. a single grid cell.).

Also the aspect of subgrid variability and spatial scaling was investigated in this study by exploiting the high spatial resolution of the ELDAS databases. CGMS results at the 50x50 km grid were compared with results obtained with an adapted CGMS that used a 10x10 km grid. Our results demonstrate that the variability in the crop simulation results as a result of variability in precipiation and radiation within a 50x50 km grid is large. However, the average of the 10x10 km grids shows an almost perfectly linear relationship with the results at the 50x50 km grid. We therefore conclude that the 50x50 km grid size is a proper spatial resolution for CGMS. The main challenge is to find representative ‘average’ values for radiation and precipitation at this scale level, rather then to increase the spatial resolution of the system.

1 Introduction

Agrometeorological models that simulate crop growth are used operationally in many parts of the world for monitoring the effect of weather conditions on crop growth and for predicting crop yields on regional to continental scales. The models used in these systems vary in complexity, but they have in common that they provide a description of the carbon assimilation process, a description of the phenological development of the crop through its different growth stages and a description of growth limiting factors, usually water availability. The differences between the various models mainly consist of the amount of physical details that they use to describe these processes and the crop management factors that can be applied.

When assuming that a crop model is a proper representation of the processes that govern plant growth, the accuracy of the model output will be a function of the accuracy of the calibration of the model for a particular crop species and variety, the boundary conditions that are being applied (e.g. soil, management) and the accuracy of the model input data (meteorology). For example, a poorly calibrated model may cause modelled phenological development to deviate from phenological development in the field leading to a wrong timing of phenological phases (shooting, flowering, ripening). Poorly defined soil parameters (porosity, rooting depth) may lead to under- or overestimation of the available soil water, while poor meteorological input can have effect on various components (soil moisture, phenology, assimilation).

Many operational crop monitoring systems operate on continental scales which means they typically have a spatial resolution of 10 to 100 km. For this type of systems it is a challenging task to properly calibrate crop varieties over large areas, define representative soil properties from often small-scale soil maps (1:1.000.000 or 1:5.000.000) and to derive meteorological input data from scattered weather stations or numerical weather prediction models with limited spatial resolution (often 1 degree). Also scaling issues play a role, because it is often unclear how representative a crop model output at low spatial resolution is, given the small scale variability in crop varieties, soil types and meteorological conditions. All these factors introduce uncertainty in the crop model outputs and consequently on the yield forecast of the crop monitoring system.

Although the above-mentioned calibration aspects can strongly influence crop model outputs, in practice the need for stringent calibration can be somewhat relaxed when it comes to crop yield predictions. This is related to the methodology that many crop monitoring systems use in order to produce a crop yield forecast. In most cases the systems do not produce a crop yield forecast directly from the crop model outputs. Instead, crop model outputs are aggregated to larger regions (provinces, districts) and regression analyses is applied to relate the aggregated crop model outputs to the yield statistics available for these regions over a period of 10 to 15 years. Often the regression analyses results in a separation of a time trend in the statistical crop yields due to improvements in crop varieties and management on the on hand and the yearly variations on this trend on the other hand. Next, the crop model outputs are used to predict the deviation from this trend rather then the crop yield itself.

The advantage of this type of methodology for crop yield forecasting is that systematic deviations in crop model output due to poor calibration or soil definition can be compensated in the regression analysis. The soil definition and crop calibration parameters are stable factors through time and will therefore influence the relationship between crop model outputs and the trend in the yield statistics. The meteorological inputs on the other hand do vary year by year and are therefore important for predicting the between-year variability in crop yields. This analysis of the different components of a crop yield forecasting system implies that uncertainty in the meteorological input data is the most important factor influencing the uncertainty in the crop yield forecast.

In Europe, the leading crop yield forecasting system is the MARS crop yield forecasting system developed by the MARS unit of the Joint Research Centre in Ispra, Italy. The MARS system was developed in the early nineties and consists of several components including the use of low resolution satellite data and an agrometeorological crop monitoring system. This system is known as the Crop Growth Monitoring System (CGMS). Since 1994, CGMS monitors crop growth in Europe, Anatolia and the Maghreb with a spatial resolution of 50 x 50 km and a temporal resolution of one day.

Currently, CGMS receives near real-time daily weather variables from about 1700 stations over Europe. The total number of 50x50 km CGMS grid cells in Europe is 5625. This means that, in the ideal case, only 30% of the grid cells have a station located in them. In practice, the situation will be worse because weather stations are often clustered and not equally spaced over the area. With regard to the weather input we thus have a severe undersampling of the true weather patterns and therefore uncertainty in the CGMS weather input.

This uncertainty is particularly severe for the parameters radiation and precipitation. Radiation is critical because it is measured directly at only few locations. Often crude correlations using either cloud cover fraction, sunshine duration or temperature need to be applied in order to estimate radiation at locations where no measurements are available. On the other side, precipitation is measured at every station, but precipitation often has a very small spatial correlation length (localized rainstorms) and the limited number of weather stations therefore cannot capture the true variability in precipitation. Interpolation of precipitation from relatively few workstations may therefore lead to highly distorted rainfall patterns.

Currently, it is unknown what the influence is of uncertainty in the CGMS weather input data on the CGMS simulation results and on the crop yield forecasts. The ELDAS project has resulted in precipitation and radiation databases which are highly accurate and with a high spatial resolution. These databases cannot be generated in near real-time and are therefore of no use to the operational CGMS. However, these databases are ideal candidates for a retrospective analyses on uncertainty in CGMS weather input. Therefore, the main objectives of this ELDAS case study were the following:

1. to quantify the uncertainty in the CGMS interpolated weather
2. to evaluate the influence of this uncertainty on the crop simulations
3. to evaluate scaling effects of the meteorological input on CGMS output

2 Models and data

2.1 The Crop Growth Monitoring System

The Crop Growth Monitoring System consists of a weather, soil and crop database, the WOFOST crop simulation model and a GIS (Diepen, 1992; Vossen, 1995) http://home.concepts-ict.nl/ iwan-supit/contents/). Within CGMS three operational levels can be distinguished:

1. interpolation of weather variables to a 50x50 km grid
2. simulation of crop growth
3. forecasting of crop yield

Within the first level, weather variables are processed and interpolated to a 50x50 km grid. At the station level, the system estimates radiation from either cloud cover fraction, sunshine duration or temperature and it calculates evapotranspiration using a modified Penman approach. Next, each grid cell receives values for temperature, radiation, air humidity, evapotranspiration and wind speed as the average from suitable surrounding weather stations. Determination of the most suitable weather stations takes place on the basis of the so-called “meteorological distance” (Voet et al., 1994). This meteorological distance is a virtual distance which is not only based on the true distance between the grid cell and the weather station, but also on factors like altitude, distance to coast and the existence of climate barriers (e.g. mountain ridges, water bodies) between the grid cell and the weather station. In case of rainfall a grid cell receives the value of the weather station with the smallest meteorological distance from the grid cell.

At the second level, the actual crop modelling is carried out. For this purpose, data from the European soil map (Anonymous, 1985; King et al., 1995) are used as input in the calculation of the water balance. Soil moisture contents at different pressure heads (saturation, field capacity, wilting point) determine the water retention of the soil. Besides the hydrologic soil characteristics, the soil map is used to estimate the suitability of soils for different crops. The climatic grid and the soil map are combined in an overlay procedure that results in a number of unique simulation units: combinations of climatic grid cell and soil type. For these simulation units the potential and water-limited crop growth are simulated with the WOFOST crop growth model.

At the third level, CGMS aggregates the daily biomass values per simulation unit into decadal values per grid (so-called grid yield) and per administrative unit (so-called NUTS yield). The CGMS model outputs are gathered for time-series of 15 to 20 years and are regressed against the historic known values of crop yields for each administrative unit.

2.2 The WOFOST crop simulation model

The WOFOST (WOrld FOod STudies) crop simulation model (Diepen et al., 1989, Supit et al., 1994) is a mechanistic crop growth model that describes plant growth by using light interception and CO_2 assimilation as growth driving processes and by using crop phenological development as growth controlling process. The WOFOST model can be applied in two different ways: (1) a potential mode, where crop growth is purely driven by temperature and solar radiation and no growth limiting factors are taken into account. (2) A water-limited mode, where crop growth is limited by the availability of water. The difference in yield between the potential and water-limited mode can be interpreted as the effect of drought. Currently, no other yield-limiting factors (nutrients, pests, weeds, farm management) are taken into account.

WOFOST uses a simple soil water balance model to keep track of moisture in the soil. The purpose of the soil water balance model is to provide information on the relative reduction in transpiration due to drought and its time course over the season. In this soil water balance, the most relevant soil layer is the actually rooted zone, which grows in thickness from an initial value of 10 cm to its maximum value, limited by soil depth or maximum crop rooting depth. The storage of water and all inflow and outflow terms are quantified in the model. The available water capacity is a function of soil texture mainly, possibly corrected for high organic matter content. The soil layer is recharged by rainfall, and when the soil moisture exceeds field capacity (equilibrium status), the excess water is drained away as percolation to the subsoil. Crop water stress is modelled as a crop transpiration reduction factor which equals to ‘1’ if the soil moisture is above a critical soil moisture level (crop dependent) and linearly varies between ‘0’ and ‘1’ when the soil moisture content is between the wilting point and the critical soil moisture.

2.3 ELDAS radiation database

The ELDAS radiation database was provided by the Royal Dutch Meteorological Institute and is based on the output of a numerical weather forecasting model which was improved by assimilating MeteoSat-derived shortwave radiation and cloud patterns into the model. This system is currently known as the ELDORADO assimilation scheme. The system provides three-hourly estimates of average downwelling longwave and shortwave radation at approximately 16-km spatial resolution over Europe for the period 1 October 1999 until 31 December 2000. Validation of the radiation estimates provided by ELDORADO demonstrated that the system provides accurate unbiased estimates of longwave and shortwave radiation with root mean square error (RMSE) values below 2600 KJ/day for shortwave radiation and RMSE values below 1300 KJ/day for longwave radiation. For our purpose, the three-hourly values were aggregated and converted to total daily radiation in KJ/day.

2.4 ELDAS precipitation database

The ELDAS precipitation database was provided by the Institute of Medical Physics and Biostatistics University of Veterinary Medicine in Vienna and consists of daily precipitation values on a 0.2° grid over Europe for the period 1 October 1999 until 31 December 2000. The precipitation values were interpolated using block kriging on more then 20,000 bias-corrected rain gauge measurements. Validation has demonstrated that systematic measurement errors for over 90% of the number of stations are within 1 mm/day. Given the sheer volume of rain gauge measurements that were used to generate this database, it will give a far better estimate of the true rainfall patterns compared to the estimates in the CGMS meteorological database.

3 Test area & experiment setup

3.1 Test area

The test site for our case study consists of Germany and France. The motivations for this choice were that the test site was small enough to allow reasonable processing times. Moreover, from north to south a gradient exists from temperate maritime conditions in Northern Germany to Mediterranean conditions in the south of France. Finally, reasonable quality yield statistics exist for Germany and France which could be used for validation purposes, although this was not yet carried out in this study.

3.2 Setup of the experiment

In order to determine the uncertainty on winter crops as well as summer crops we choose winter-wheat and grain maize as typical examples of a winter and a summer crop. Next, we defined three actions for carrying out our research:

3.2.1 Action 1

Action 1 consisted of the running of the standard CGMS system which uses interpolated weather from about 1700 stations. No modifications were made to this system.

3.2.2 Action 2

In action 2 we replaced the CGMS precipitation and radiation with the average radiation and precipitation of all ELDAS grid points within a 50x50 km CGMS grid box. All other meteorological input variables and other data were kept to the values that were defined in action 1.

3.2.3 Action 3

In action 3 we introduced a major change in the system by increasing the grid density from 50x50 km to 10x10 km. This allows to infer the within-grid variability of the CGMS crop simulation results and to test if the system scales linearly when averaging weather inputs. We defined a 10x10 km grid that coincides with the 50x50 km grid, each 50x50 grid box contained 25 10x10 km grids. Next, the ELDAS grid points were added to the system as pseudo stations, each pseudo station received the associated ELDAS radiation and precipitation values and received the values for the other variables from the underlying 50x50 km grid.

The standard CGMS interpolation procedure was applied to the pseudo stations in order to assign values to the 10x10 km grid. The grid with pseudo stations is so dense that the “meteorological distance” that is calculated is basically irrelevant and each grid receives the values from the surrounding pseudo stations. A more logical approach would be to interpolate from weather stations to the 10x10 km grid first and in the second step replace the value of radiation and precipitation in the 10x10 km grid with the value of the nearest ELDAS grid point. Our approach is basically a shortcut in order to avoid redefining and recalibrating the CGMS interpolation procedure on the 10x10 km grid. We assume that this shortcut has little influence on the simulation results. The final adaption we made was to create an overlay between the soil map and the 10x10 km grid. This procedure results in a new set of simulation units (unique combinations of climatic grid cell and soil type). Finally, the WOFOST model was applied for each simulation unit.

Table 1: Statistics derived from the residuals between CGMS and ELDAS databases for radiation and precipitation.

	Radiation [kJ/m^2]		Precipitation [mm]	
Dekad	Mean	StDev	Mean	StDev
1	564.6	1137.0	-5.4	17.7
10	-231.0	1946.6	-4.8	21.1
19	1182.6	3649.7	-1.6	14.5
28	539.0	1495.8	-3.2	16.9

4 Results

4.1 Comparison of weather variables in CGMS and ELDAS weather variables

We determined the uncertainty in the CGMS radiation fields by calculating the average radiation over a period of 10 days for dekads[1] 1, 10, 19 and 28 for both the CGMS and ELDAS database. The average value of all ELDAS points within a CGMS grid cell was used. Next, we subtracted the ELDAS values from the CGMS values, calculated the statistics of the residuals (Table 1) and created maps of the residuals (not shown). The results demonstrate that large uncertainty exist in the CGMS radiation values with a tendency to overestimate radiation (positive residuals). In general CGMS overestimates radiation values in the southern Mediterranean areas in all four dekads. For all other areas large positive and negative residuals occur, for example in dekad 10 radiation is underestimated with values of 3000 to 4000 kJ/m^2 in northern Europe, while in dekad 19 radiation is overestimated with 3000 to 5000 kJ/m^2 in large parts of central and eastern Europe.

We determined the uncertainty in the CGMS precipitation fields by calculating the total precipitation over a period of 10 days for dekads 1, 10, 19 and 28 for both the CGMS and ELDAS database. The average value of all ELDAS points within a CGMS grid cell was used. Next, we subtracted the ELDAS values from the CGMS values, calculated the statistics of the residuals (Table 1) and created maps of the residuals (not shown). The results demonstrate that large uncertainty exist in the CGMS precipitation values with a tendency to underestimate precipitation (negative residuals). In general no clear trends can be derived from the spatial distribution of the residuals and under- or overestimation of precipitation tends to occur rather randomly.

4.2 Uncertainty of simulated crop yield (action 1 & 2)

The uncertainty in the input weather fields as demonstrated in the previous section will have influence on the crop simulation results. We ran the CGMS system with both ELDAS and CGMS weather input and compared the simulation results at the 50x50 km grid level. We used the simulated biomass of the storage organs to calculate statistical properties of the residuals (Table 2) and to create scatter plots (figure1).

For winter-wheat under potential production, the mean of the residuals (Table 2) is slightly positive (21.3 kg/ha) which can be explained by the overestimation of radiation by CGMS leading to higher CGMS yields compared to ELDAS yields. For water-limited production on the other hand, the mean of the residuals becomes strongly negative (-668.8 kg/ha) due to the underestimation of precipitation by CGMS, but also the standard deviation and RMSE become twice as large. This is confirmed by the scatter plot of CGMS versus ELDAS yield (Figure 1) which shows a considerable deviation from the 1:1 line and a majority of pixels above the diagonal.

For grain maize under potential production, the mean of the residuals is also positive (130.5 kg/ha) which confirms the overestimation of radiation by CGMS. The effect of uncertainty in the precipitation on grain maize simulation is even larger compared to winter-wheat because the standard deviation and RMSE are four

[1]The term *dekad* refers to a 10-day period, this is a convention defined by the Food and Agriculture Organisation of the United Nations in order to avoid confusion with *decade* (10-year period).

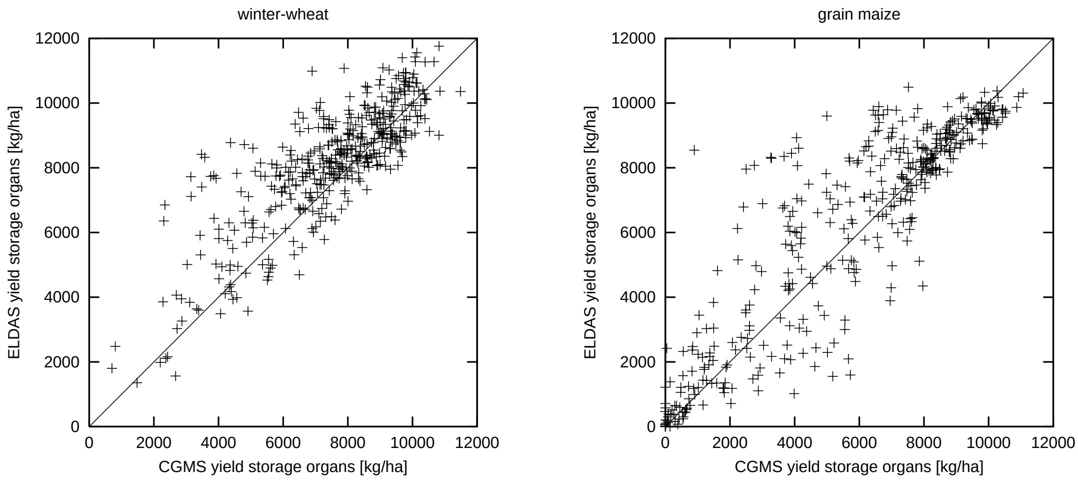

Figure 1: Scatterplot between water-limited grid yield obtained using weather variables from the CGMS database and from the ELDAS database for winter-wheat and grain maize.

Table 2: Mean and standard deviation of the residuals between CGMS and ELDAS yields as well as the root mean square error (RMSE) between CGMS and ELDAS yields. Results for both potential and water-limited production (storage organs) for winter-wheat and grain maize

	Winter-wheat [kg/ha]		Grain maize [kg/ha]	
	Potential production	Water-limited production	Potential production	Water-limited production
Mean	21.3	-668.8	130.5	-420.6
StDev	644.4	1164.9	351.5	1479.4
RMSE	644.0	1342.1	374.6	1536.3

times larger compared to the potential simulations. These results demonstrate that the effect of uncertainty in precipitation is particularly strong for summer crops which are more influenced by high transpiration rates during the summer. The scatter plot (Figure 1) confirms the large uncertainty in the simulation results. The plot also shows a clustering of pixels near the origin and the upper right part of the chart which indicates that yield is not influenced under very wet (potential yield) or very dry (no yield) conditions, but the transition zones are sensitive.

For yield forecasting it is necessary to aggregate the yield of individual grid cells to larger spatial regions. When the CGMS and ELDAS simulation results at the 50x50 km grid level are spatially aggregated to NUTS 1 regions the variability between ELDAS and CGMS yield becomes much smaller (Figure 2). For winter-wheat an upward bias remains which is due to the underestimation of precipitation by CGMS, while the results for grain maize are more evenly distributed along the 1:1 axis. These results demonstrate the necessity to aggregate CGMS results to larger regions in order to decrease uncertainty on the simulation results.

4.3 Scaling of crop yield simulation results (action 2 & 3)

CGMS interpolates meteorological variables to the 50x50 km grid and assumes these estimates representative for the entire 50x50 km grid cell. In practice, subgrid variability in these meteorological parameters exists and this variability is probably large for certain areas and seasons. A related question is how the output of CGMS scales when the weather input variables are spatially averaged. If CGMS output scales highly non-linear when the input variables are averaged, then the application of CGMS on the level of 50x50 km grid cells

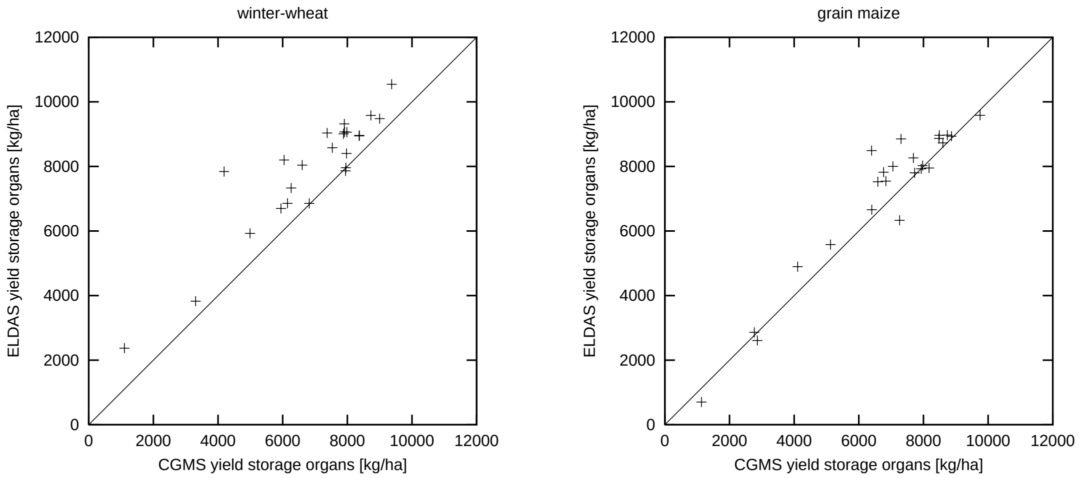

Figure 2: Scatterplot between water-limited NUTS yield obtained using weather variables from the CGMS database and from the ELDAS database for winter-wheat and grain maize.

is inappropriate and a smaller grid size should be chosen.

We investigated the scaling of CGMS by comparing the grid yield at the 50x50 km level, with the mean and standard deviation of the grid yield of the underlying 10x10 km grid (action 3). The crop simulations at the 10x10 km grid used the high resolution ELDAS precipitation and radiation, while the crop simulations at the 50x50 km level used averaged ELDAS precipitation and radiation. The results (figure 3) demonstrate that for water-limited production considerable subgrid variability does exist (Y-error bars) but that CGMS scales almost linearly between the 10x10 km and 50x50 km level. A linear regression on the relationship between the water-limited yields (Table 3) shows that for both winter-wheat and grain maize the regression coefficient is nearly one and R-squared reaches values of 0.96 to 0.97. The intercept was set to zero in the regression.

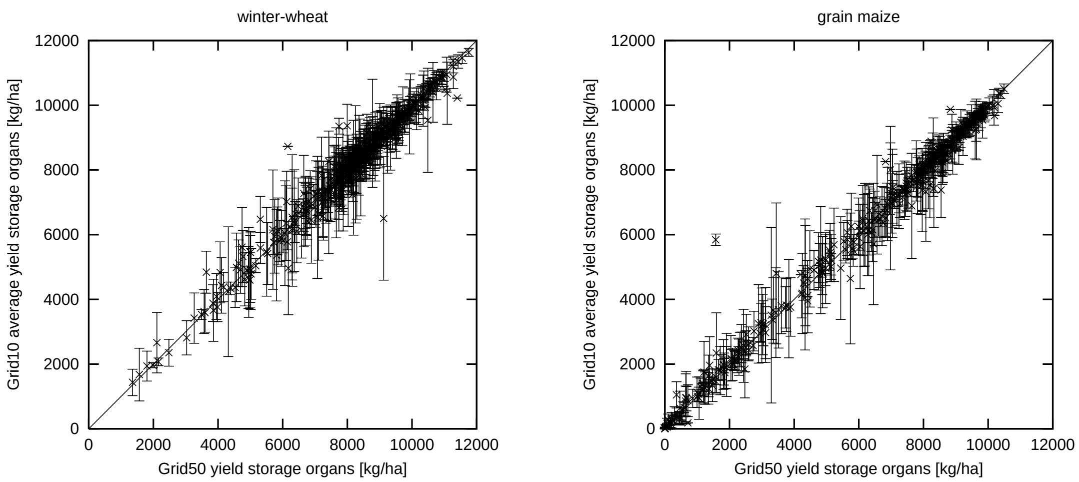

Figure 3: Scatterplot between grid yield obtained using average ELDAS weather for the 50x50 km grid and average grid yield obtained from the 10x10 km grids. Results for water-limited yield storage organs for winter-wheat and grain maize. Error bars are plus and minus one standard deviation.

Table 3: Results from a linear regression between grid50 yield and the average of the underlying grid10 yield, the intercept of the regression is assumed to be zero.

	Regression coefficient	R^2	RMSE
Winter-wheat	0.989	0.967	327.8
Grain maize	0.997	0.989	335.0

5 Conclusions

The results of our study demonstrate that the interpolated radiation and precipitation values used by CGMS at the 50x50 km grid are subject to large uncertainty. In general there is a tendency to overestimate radiation, while precipitation is underestimated. This uncertainty in meteorological input causes uncertainty in the CGMS crop simulations. For summer crops in particular (grain maize), the influence of uncertainty in precipitation is large, but also winter crops are influenced. Much of this uncertainty can be decreased by spatially aggregating simulation results to larger regions. Since CGMS results are currently used only at national or subnational level, the influence of the uncertainty on the yield forecast is probably limited. However, it demonstrates that CGMS output should not be used for making decisions on small areas (i.e. a single grid cell.).

Finally, the aspect of subgrid variability and spatial scaling was investigated in this study. We compared the simulated yield at the 50x50 km grid that was obtained using average precipitation and radiation, with the average and standard deviation of the simulation results that were obtained using a 10x10 km grid that received distributed radiation and precipitation. Our results demonstrate that the variability in the crop simulation results as a result of variability in precipiation and radiation within a 50x50 km grid is large. However, the average of the 10x10 km grids shows an almost perfectly linear relationship with the results at the 50x50 km grid. We therefore conclude that the 50x50 km grid size is a proper spatial resolution for CGMS. The main challenge is to find representative ‘average’ values for radiation and precipitation at this scale level, rather then to increase the spatial resolution of the system.

Acknowledgements

Bart van den Hurk and Franz Rubel are kindly acknowledged for providing the ELDAS radiation and precipitation databases. Paul Skomorowski is kindly acknowledged for providing his software to convert the GRIB format into a database readable format.

References

Anonymous, 1985. Soil map of the European Communities 1 : 1 000 000. Commision of the European Communities, Directorate General for Agriculture, Brussels-Luxembourg, pp. 124.

Diepen, C.A., van, Wolf, J. and Keulen van, H., 1989. WOFOST: a simulation model of crop production. Soil Use and Management, 5: 16-24.

Diepen, C.A., van, 1992. An agrometeorological model to monitor the crop state on a regional scale in the European Community: concept, Implementation and first operational outputs. In: F. Toselli and J. Meyer-Roux (Editors), Conference on application of remote sensing to agricultural statistics. Office for Official publications of the EU, Belgirate, Italy, pp. 269-277.

King, D., Burril, A., Daroussin, J., Le Bas, C., Tavernier, R., Van Ranst, E, 1995. The EU soil geographic database. In: D. King, R.J.A. Jones and A.J. Thomasson (Editors), European land information systems for agro-environmental monitoring. EUR 16232 EN, Office for Official Publications of the EU, Luxembourg, pp. 43-60.

Supit, I., Hooijer, A.A. and Diepen van, C.A., 1994. System Description of the WOFOST 6.0 Crop Simulation Model Implemented in CGMS, Volume 1: Theory and Algorithms. EUR 15956 EN, Joint Research Center, Commission of the European Communities, Luxembourg.
Vossen, P., 1995. Early crop production assessment of the European Union, the systems implemented by the MARS-STAT project. In: J.F. Dallemand and P. Vossen (Editors), Workshop for Central and Eastern Europe on agrometeorological models: theory and applications in the MARS project. Office for Official Publications of the EU, Ispra, Italy, pp. 21-51.