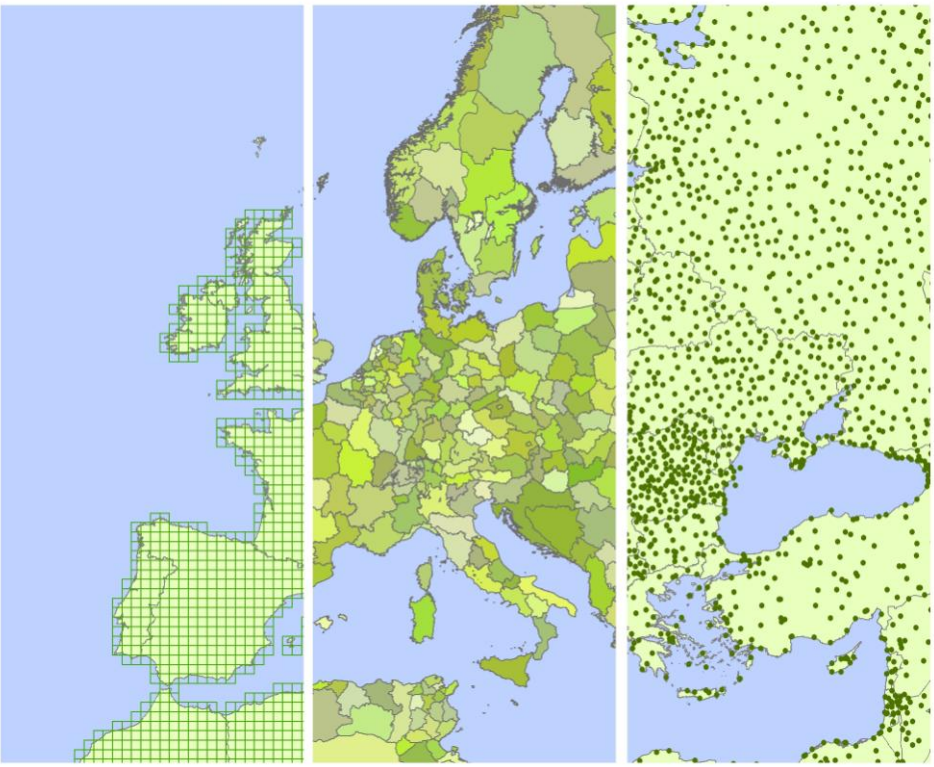
JRC Scientific and Technical Reports

CGMS version 9.2
User manual and technical documentation

Editors: B. Baruth, G. Genovese, O. Leo
Contributors: H. Boogard, J. A. te Roller, K. van Diepen



EUR 22936 EN - 2007

JRC
EUROPEAN COMMISSION

ipsc
Institute for the Protection and Security of the Citizen

The Institute for the Protection and Security of the Citizen provides research based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: I-21020 Ispra (VA)
E-mail: mars-stat@jrc.it
Tel.: 0039-0332-785471
Fax: 0039-0332-789029

http://ipsc.jrc.ec.europa.eu
http://www.jrc.ec.europa.eu

Legal Notice
Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server
http://europa.eu/

JRC 4605

EUR 22936 EN
ISBN 978-92-79-06995-6
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

© European Communities, 2007

Reproduction is authorised provided the source is acknowledged

Printed in Italy

Content

1 Introduction 4

2 Thematic description of CGMS level 1 6

2.1 Goals and assumptions 6

2.2 Acquisition, checks and processing daily station data 7

2.2.1 Data Acquisition and Checks 7
2.2.2 Calculation of Global Radiation 8
2.2.3 Calculation of Evapotranspiration 9

2.3 Spatial Interpolation to regular climatic grid 10

2.3.1 Reasons for Interpolation 10
2.3.2 Interpolation Method 10

2.4 Alternative entry for grid weather data 14

3 Thematic description of CGMS level 2 16

3.1 Goal and assumptions 16

3.2 Collection and processing of input data 17

3.2.1 Grid Weather 17
3.2.2 Crop Data 17
3.2.3 Soil Data 19
3.2.4 Initial Available Soil Water 20
3.2.5 Administrative Data / Area Statistics 21

3.3 Spatial schematization for simulation 22

3.4 Regional Crop simulation 23

3.4.1 WOFOST 23
3.4.2 LINGRA 25

3.5 Spatial schematization for aggregation 26

3.6 Spatial Aggregation 27

3.7 Simulation winter crops starting in autumn 28

3.8 Restart crop simulation 31

3.9 Information on CGMS run 32

4 Additional functionality in the CGMS software for yield forecast 34

4.1 Spatial aggregation of other indicators besides CGMS level 2 34

4.2 Data preparation for yield forecast 35

4.2.1 Linear regression: other indicators 35
4.2.2 Scenario analysis 36
4.2.3 User specified equation 36

4.3 Manage external data 37

5 Overview of different CGMS versions 38

6 CGMS 9.2 database 40

6.1 Table by table 40

6.2 Relation between tables 57

6.3 Database compilation 70

6.3.1 Data collecting 70
6.3.2 Recommendations for building tables 72

7 CGMS User Interface 79

7.1 Installation CGMS 79
7.2 Data base login and execution steps 79
7.3 Weather data calculation 81
7.4 Crop simulation 83
7.4.1 Crop and administrative regions 83
7.4.2 Campaign 85
7.4.3 Weather data 85
7.4.4 Start/end mode 86
7.4.5 Initialisation water balance 86
7.4.6 Simulate with groundwater influence 87
7.4.7 Winterkill 87
7.4.8 Save output for one or more decades 87
7.5 Aggregation 88
7.5.1 Schematization 89
7.5.2 Spatial level 89
7.5.3 Period 89
7.5.4 Select indicator 89
7.6 Preparation of data yield forecast calculation 90
7.6.1 Analyses 91
7.6.2 Administrative regions of interest 92
7.6.3 Period 92
7.6.4 Statistical crop 92
7.6.5 Decade of yield forecast 92
7.6.6 Definition of composed indicators 92
7.6.7 Selection of statistical parameters 94
7.6.8 Storage of prepared data 96
7.7 Import external data to database 97
7.7.1 Data type 97
7.7.2 Filename 99
7.8 Export data 99
References 109
Annex A Using CGMS in batch mode 113

1 Introduction

The following report gives detailed information on the CGMS version 9.2 which is used in an operational context for the MARS Crop Yield Forecasting System. This report is the follow-up of the CGMS version 8.0 User manual and technical documentation (Savin et. al, 2004, EUR 21379 EN) and follows the same structure.

Historically, the MARS Crop Yield Forecasting System (MCYFS) has been developed in Europe around the Crop Growth Monitoring System (CGMS). The CGMS is the combination of the WOFOST crop growth model, a relational database and a statistical yield prediction module.

Nowadays the MARS Crop Yield Forecasting System (MCYFS) is fully operational for the European countries. The decisions n^0 1445/2000/EC and 2066/2003/EC on the application of area frame survey and remote sensing techniques to the agricultural statistics for 1999 to 2007 of the European Parliament moved the MCYFS into the operational phase. The main customers of the system are DG-AGRI and EUROSTAT. Part of the operational service to run MCYFS is outsourced through the MARSOP project (MARS-OPerational), which started in the middle of 2000. The AGRI4CAST action (ex MARS –STAT) (Crop forecasts/Estimates and Climate Change impact on agriculture) of the AGRICULTURE unit of the Joint Research Centre supervises this project and concentrates on yield forecast analysis and synthesis of all information in a bulletin.

In the context of the Global Monitoring for Environment and Security initiative (GMES), JRC formed in 2001 the MARS-FOOD action (Crop monitoring for food security) This action aims at supporting the European Union Food Security and Food Aid Policy through an improved assessment of the crop status in regions/countries stricken by food shortage problems. Accurate and timely information on crop status is needed to properly calibrate and direct European Food Aid, in order to prevent food shortages and consequent human suffering, and to avoid possible market disruptions due to unnecessary food aid distributions. DG AIDCO and RELEX, are the main European Commission customers for this work.

For agro-meteorological monitoring in Russia, Central Asia and the Mediterranean Basin CGMS was extended with different functions as described in the CGMS version 8.0. User Manual and Technical Documentation. This extension was made possible by the contract N 20268-2002-12-FIED ISP NL. As a main result CGMS version 8.0 was developed.

During the years 2004-2007 the development of CGMS continued in the framework of the contracts MARSOP-2 (contract no 21508-2003-12 F1SC ISP NL) and ASEMARS (contract no 22550-2004-12 F1SC ISP NL) leading to a new version CGMS 9.2. The underlying document describes this latest version. New functionality was added or existing functions were changed:

- Weather interpolation for the current year is changed using long term average data from reliable stations in case data for the current day is missing
- The possibility to simulate frost damage (or winterkill) has been implemented
- Several bugs have been solved with regard to the LINGRA model. The LINGRA model is now ready to be used
- CGMS is adapted so that it can be used by the CALPLAT tool:
    - Adjustment of CROP_YIELD (extra column crop, date in stead of decade/year, extra output columns)
    - Adjustment of GRID_YIELD and NUTS_YIELD (date in stead of decade/year, extra output columns)

- Potential and water limited can be run separately
- Introduced CALPLAT mode (store data at daily level and write additional output)

In the near future it is expected that part of the current functionality of CGMS, the spatial aggregation and the data preparation for yield forecast, will be replaced by another tool. Then the CGMS software will be the tool responsible for weather processing and interpolation and crop growth simulation. Next, the data engine of the CGMS viewer, which has been built to generate on-the-fly maps of CGMS level 1 and 2 for the web site, will be responsible for retrieving and aggregating CGMS weather and crop indicators in a fast way, and additionally creating compound meteorological indicators like rainfall around a crop development stage. The tool will read region, crop and decade specific settings from a database for instance the group of indicators that have to be aggregated to regional level so that they can be used in the third tool which is responsible for the regional crop yield forecast: CGMS Statistical Tool (CST). On top of these tools a manage console will exists responsible for running the tools in an operational production line. In addition there is the CGMS viewer (using the same data engine to generate data) to view the data at EMU or grid level.

We hope this information will be useful for the CGMS users, as well as for the specialists in the field of crop growth simulation, agronomists, agro-meteorologists and will lead to fruitful discussion on further CGMS development, and crop growth monitoring methods improvement.

2 Thematic description of CGMS level 1

The weather monitoring component is one of the cornerstones in CGMS. It consists of the two following activities (see Figure 2-1):

- Acquisition, checks and processing of daily meteorological station data
- Spatial interpolation to a regular climatic grid

Resulting weather indicators can be viewed through the tool CGMS viewer.

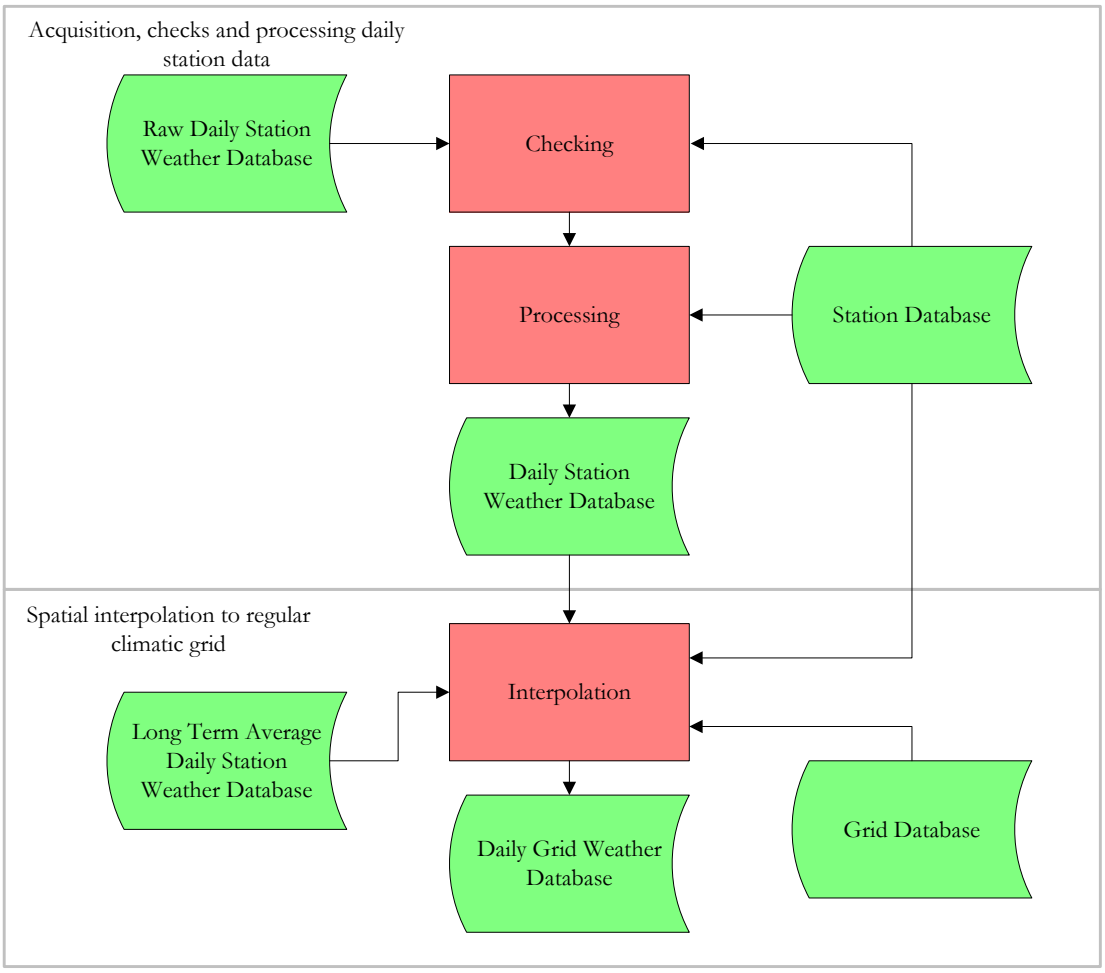


Figure 2-1: Overview of the weather monitoring components of CGMS.

For a detailed background the reader is referred to Micale and Genovese (2004) and in Supit et al. (1994). In the following paragraphs CGMS tables are mentioned. A detailed description of the tables can be found in Chapter 6.

2.1 Goals and assumptions

Daily meteorological station data are used in two ways for crop yield evaluations. First as weather indicators for a direct evaluation of alarming situations such as drought, extreme rainfall during sowing, flowering or harvest etc. Second, as input for the crop growth model WOFOST (see § 3.4).

Weather is the general condition of the atmosphere and the processes occurring in it. These processes vary in time and space from a few seconds for small scale eddies to a few weeks

for large depressions. For the CGMS only the atmospheric conditions near the earth surface are relevant. The CGMS focuses on a time scale of one day and a spatial scale of 50 by 50 km which should be the optimum scales to evaluate effects of weather on crops yields at European level regarding also data availability and resources.

2.2 Acquisition, checks and processing daily station data

2.2.1 Data Acquisition and Checks

The meteorological station data consists of:

- Station information.
- Raw daily meteorological data.
- Processed daily meteorological data.

The stations are limited to those for which data not only are regularly collected but which can also be received and processed in semi-real time (Burrill and Vossen, 1992). Relevant information of stations includes station number, station name, latitude, longitude and altitude. This data are available in the table **WEATHER_STATION**.

The following text describes how (historic) data are collected for the European application. This gives some insight in the problems that can be expected when working with station data. Some of the historic meteorological data are purchased directly from various national meteorological services, others are acquired via the Global Telecommunication System (GTS). As the data are obtained from a variety of different sources, considerable pre-processing is necessary to convert them to a standard format. Two different procedures are applied for distinct subsets of the data set. The historic data were ordered directly from national meteorological services. Around 1992 they represented approximately 380 stations in the EU, Switzerland, Poland and Slovenia with data from 1949 to 1991 (Burrill and Vossen, 1992). Later the historic set have been extended with stations in eastern Europe, western Russia, Maghreb and Turkey. The historic data were converted into consistent units and were checked on realistic values. The database was also scanned for inconsistencies, such as successive days with the same value for a variable, or minimum temperatures higher than maximum temperatures (Burrill and Vossen, 1992).

From 1991 to present, meteorological data are received in near real time from the GTS network for different hours within one day. The data are pre-processed and quality checked using the AMDAC software package (MeteoConsult, 1991) which extracts, decodes and processes the GTS data. After decoding, the following data are checked for consistency and errors: air temperature, dew-point temperature (humidity), pressure at sea level, wind speed, amounts of precipitation, clouds, and sunshine duration. This error checking compares each observation with the corresponding values of the surrounding stations and compares that particular observation with observations at other times in the same day at the same station. Obvious errors in the observations are corrected automatically and a message is written to a log file; other errors are flagged for possible correction by an operator (Burrill and Vossen, 1992). Finally, the data are converted into daily values. This comprises the selection of minimum and maximum temperature, the aggregation of the rainfall, cloud cover and sunshine duration, the calculation of mean vapor pressure etc.

The processed daily meteorological data consists of 30 meteorological parameters including various cloud cover indicators, air temperature, vapour pressure, wind speed and rainfall. Because European stations follow different measurement schemes many records contain blank fields for parameters which are never registered. Stations often also include blank

fields for parameters which where not available for limited periods. However, the stations selected for inclusion in the database are those which normally report at least the minimum and maximum daily air temperature, rainfall, wind speed, vapour pressure (or humidity) as well as either global radiation, sunshine hours or cloud cover (Burrill and Vossen, 1992). Each day the processed daily meteorological data are inserted into the CGMS database in table **METDATA**. However CGMS only needs a sub set of the 30 variables described above and thus only 9 variables are described in underlying user manual and table **METDATA**.

2.2.2 Calculation of Global Radiation

Global radiation is the daily sum of incoming solar radiation that reaches the earth surface. It is mainly composed of wavelengths between 0.3 µm and 3 µm. Approximately half of the incoming radiation with wavelengths between 0.4 and 0.7 µm is Photosynthetically Active Radiation (PAR). Global radiation is the driving variable in the growth-determining CO2 assimilation process and thus crop growth models are sensitive to radiation data (van Diepen, 1992).

A major problem is the scarcity of measured global radiation. In cases where no direct observations are available it must be derived from sunshine duration, cloud cover and/or temperature, on the basis of relatively weak relationships. The global radiation calculation uses one of three formulae (Ångström, Supit, Hargreaves), depending on the availability of meteorological parameters. An important component in these formulae is the amount of Angot radiation which is the extra-terrestrial radiation integrated over the day at a certain latitude on a certain day. In fact, all of the three formulae estimate the fraction of Angot radiation actually received at the earth surface. The calculation of the Angot radiation and the three different formulae are described by Supit *et al.* (1994) and van der Goot (1998a).

The following hierarchical method is used to calculate global radiation (Supit and van Kappel, 1998). If observed global radiation is available it will be used. In the case sunshine duration is available, global radiation is calculated using the equation postulated by Ångström (1924) and modified by Prescott (1940). The two constants in this equation depend on the geographic location.

$$R_g = R_a * (A_a + B_a * (n / L)) \qquad (2\text{-}1)$$

where:

R_g	: global radiation	[J m^{-2} d^{-1}]
R_a	: Angot radiation	[J m^{-2} d^{-1}]
n	: bright sunshine hours per day	[h]
L	: astronomical day length	[h]
A_a, B_a	: regression coefficients (Ångström)	[-]

When sunshine duration is not available but minimum and maximum temperature and cloud cover are known, the Supit formula is used, which is an extension of the Hargreaves formula (Supit, 1994). Again, the regression coefficients depend on the geographic location.

$$R_g = R_a * A_s * (\sqrt{(T_{max} - T_{min})} + B_s * \sqrt{(1 - CC/8)}) + C_s \qquad (2\text{-}2)$$

where:

T_{min}, T_{max}	: minimum and maximum daily temperature	[°C]
CC	: cloud cover in octets	[-]
A_s, B_s	: regression coefficients (Supit)	[-]

C_s : regression coefficient (Supit) [J m^{-2} d^{-1}]

Finally, when only the minimum and maximum temperatures are known the equation of Hargreaves *et al.* (1985) is used. Again, the regression coefficients depend on the geographic location.

$$R_g = R_a * A_h * \sqrt{(T_{max} - T_{min})} + B_h \qquad (2\text{-}3)$$

where:

A_h, : regression coefficient (Hargreaves) [-]

B_h : regression coefficient (Hargreaves) [J m^{-2} d^{-1}]

The main problem with the application of these formulae is the quality of the regression constants. Studies by Supit (1994), Supit and van Kappel (1998) and van Kappel and Supit (1998) showed no relationship between latitude and the coefficients, although such a relation is frequently used to estimate these regression constants.

Supit and van Kappel (1998) and van Kappel and Supit (1998) have obtained sets of regression constants for the above mentioned formulae for as many weather stations as possible, with a geographic distribution that corresponds to the area of interest for the CGMS in Europe. As a result, a set of 256 reference stations has been identified for which a relevant set of measured radiation data and other parameters in the formulae exist. For these stations regression constants have been calculated based on measured radiation data for the three formulae mentioned above. They are stored in table **SUPIT_REFERENCE_STATIONS**.

The program SupitConstants (an additional program in addition to CGMS) uses this set of data, consisting of latitude, longitude, altitude and calculated regression constants, to derive the regression constants for all stations in the CGMS for Europe. Interpolation of the regression constants of the reference stations to other stations is based on a simple distance weighted average of the three nearest stations. More information is given by Kappel and Supit (1998). This process is carried out once, unless the set of reference stations changes or when new stations are added. Interpolated regression constants are written in the table **SUPIT_CONSTANTS**. After the regression constants have been established for all stations, global radiation can be calculated by the CGMS using any one of the above formulae. Finally, the CGMS writes the derived daily global radiation of every station in the table **CALCULATED_WEATHER**.

2.2.3 Calculation of Evapotranspiration

Daily meteorological station data received from GTS does not contain potential evapotranspiration. This parameter is calculated by the CGMS with the well-known Penman formula (Penman, 1948).

In general, the evapotranspiration from a water surface can be described by:

$$E0 = \frac{(\Delta R_{na} + \gamma EA)}{\Delta + \gamma} \qquad (2\text{-}4)$$

where:

$E0$: evapotranspiration from a water surface [mm d^{-1}]

R_{na} : net absorbed radiation [mm d^{-1}]

EA	: Evaporative demand	[mm d^{-1}]
Δ	: Slope of the saturation vapour pressure curve	[mbar°C^{-1}]
γ	: Psychrometric constant (0.67)	[mbar °C^{-1}]

Evapotranspiration from a wet bare soil surface (ES0) and from a crop canopy (ET0) can also be calculated with the above formula. Only the albedo and surface roughness differs for these three types of evapotranspiration as explained below. The net absorbed radiation depends on incoming global radiation, net outgoing long-wave radiation, the latent heat and the reflection coefficient of the considered surface (albedo). For E0, ES0, and ET0 albedo values of 0.05, 0.15 and 0.20 are used respectively. The evaporative demand is determined by humidity, wind speed and surface roughness. For crop canopies (ET0) a surface roughness value of 1.0 is used and for a free water surface and for the wet bare soil (E0, ES0) value 0.5 is taken. For a more detailed description of the underlying formulae we refer to Supit *et al.* (1994) and van der Goot (1997). The calculated E0, ES0, and ET0 are stored in table **CALCULATED_WEATHER**.

2.3 Spatial Interpolation to regular climatic grid

2.3.1 Reasons for Interpolation

To simulate crop growth the CGMS needs to interpolate daily meteorological station data to a regular climatic grid. In theory it is possible to apply the CGMS at station level. However, the problem is that meteorological stations are very far apart and their geographical distribution is irregular. Furthermore, the CGMS not only needs daily meteorological data but also data on soils, crops and land use which are spatially variable too. In order to combine these data layers and simulate their interaction as well as their effects on crop growth it is necessary to know meteorological conditions for any point in between the stations. After interpolation of meteorological data, a representative spatial schematisation of meteorology, soils, crops and land use can be derived.

An additional reason for interpolation is the function of the CGMS as a monitoring and forecasting system in which current results of certain regions are compared to a reference year. Comparison of current weather with the long term average on a station basis is not recommended because the continuity of time series cannot be guaranteed at station level. Stations are sometimes moved to other locations, or they are replaced by others, or new stations do not have historic data. Further, station data may be incomplete because of interruptions in observations or instrument failures. The spatial interpolation procedure of the CGMS secures continuous and complete time series of meteorological data on a fixed geographical grid covering the complete land surface.

2.3.2 Interpolation Method

The daily meteorological data is interpolated towards the centres of a regular climatic grid. The size of the grid cell is flexible and can be different for each specific application. The data of the climatic grid are stored in table **GRID**. The grid is used for two purposes in CGMS. Firstly it is used to describe the area for which the meteorological data is assumed to be homogeneous. Secondly, a crop-calendar and a crop variety is associated with each grid cell (see Chapter 3). The density of the network of meteorological stations is the main factor that determines the size of the grid cells.

The methodology for the spatial interpolation of the data of the existing network of meteorological stations towards the climatic grid cell centres, is based on the studies of Beek *et al.* (1991) and van der Voet *et al.* (1994). It is described by van der Goot (1998a). This

method was chosen because its simple approach made it easy to automate while the accuracy was sufficient to serve as input to the crop growth model. The interpolation is executed in two steps: first the selection of suitable meteorological stations to determine representative meteorological conditions for a specific climatic grid cell. Second, a simple average is calculated for most of the meteorological parameters, with a correction for the altitude difference between the station and grid cell centre in case of temperature and vapour pressure. As an exception rainfall data are taken directly from the most suitable station.

2.3.2.1 General selection of weather stations

Not all meteorological stations broadcast a complete set of data via the GTS, and not all stations broadcast continuously. To increase interpolation reliability, and to reduce the computational requirements, the CGMS performs checks on data availability of weather stations. The first check is based on a classification with respect to the data type that stations can deliver. Three meteo groups are distinguished: rainfall, temperature and all other variables (radiation and evapotranspiration). The second check is based on the temporal availability of the data in these classes. In the selection procedure the CGMS determines for each weather station and for each historic year (year is not equal to record 'CURRENT_YEAR' in the **SYSCON** table) if availability for a data class is above a certain threshold. If so, the station is marked as valid for the concerned class. The threshold value can be selected per station, but is applied to all three categories. The CGMS in Europe applies a threshold value of 80%, i.e. if the station data in a particular class is for more than 80% complete, the station will be used for the interpolation of the data in the concerned class. The timeframe taken into account for the check is the total number of days in the year for the historic years. After determination of the availability the CGMS writes the results into the table **WEATHER_DATA_AVAILABILITY**. This procedure secures a reliable fixed set of stations used for the interpolation of weather data to a climatic grid cell.

For the current year (record 'CURRENT_YEAR' in the **SYSCON** table) the determination of available stations is done in a different way because it is not possible to determine on forehand a fixed set of stations based on this temporal availability criteria. Therefore it was decided to use all available stations per each, individual day.

2.3.2.2 Qualification of weather stations

Only weather stations within a radius of 250 kilometre around the grid cell centre can possibly used for interpolation. The radius is defined in the table **SYSCON**. To define the suitability of a weather station for interpolation, the CGMS applies a selection procedure that relies on the similarity of the station and the climatic grid cell centre. This similarity is expressed as the result of a scoring algorithm that takes the following criteria into account: distance between the station and the grid cell centre, similarity in altitude and distance to the coast between the station and the grid cell centre and relative position of the station and the grid cell centre with regard to climatic barriers (i.e. mountain ranges).

The final score is expressed in kilometres and is derived from these geographic characteristics by empirically converting them into kilometres. The higher the score, the lower the similarity between the station and the grid cell centre. The score is calculated as follows:

$$SimilarityScore = dist + Dalt * Walt + DdCstcorr + ClbInc \qquad (2\text{-}5)$$

SimilarityScore: similarity score of the weather station with relation to the

	concerned grid cell centre	[km]
dist	: distance between the weather station and the grid cell centre	[km]
Dalt	: absolute difference in altitude	[m]
Walt	: weighting factor for *Dalt*	[km.m^{-1}]
DdCstcorr	: absolute difference in corrected distance to coast	[km]
ClbInc	: climate barrier increment	[km]

The weighting factor for altitude differences has been set at 0.5 km per m which is based on the assumption that 100 meters difference in altitude is equivalent to 50 kilometres of distance. The climate barrier increment is set to 1000 when the station and the grid cell centre are separated by a climate barrier such as the Alps and the Pyrenees, otherwise it is set to 0. The difference in the distance to the coast is expected to be more important when the absolute distance to coast is small, and of no importance when the actual distance to the coast is large (more than 200 kilometres). The empirical correction shown in Figure 2-2 maps the true distance to coast to a range between 0 and 100 km.

For each climatic grid cell centre and weather station, information is needed about geographic location, altitude, distance to coast and climate barrier. The altitude of the grid cell should represent the agricultural area within the grid cell.

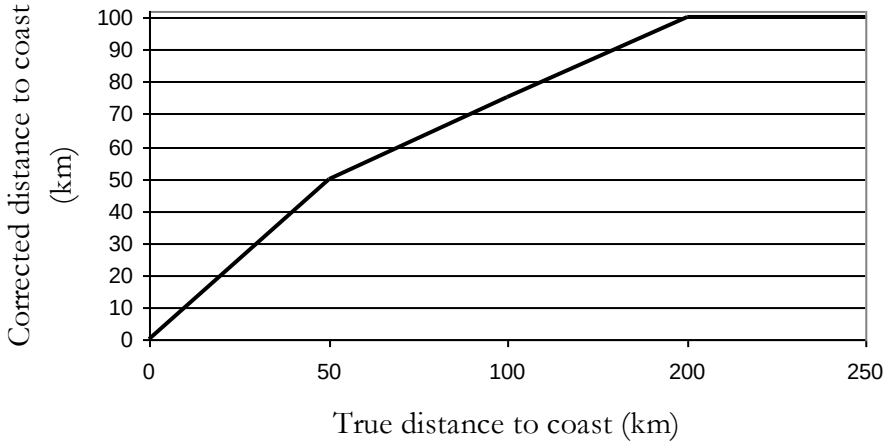


Figure 2-2: Empirical correction for the true distance to coast.

2.3.2.3 Interpolation of rainfall data

Interpolated rainfall should be realistic in terms of number of rainy days, amount of rainfall and temporal distribution. The effects on the soil water balance and consequently on crop growth simulation of various small daily showers is different than for example one large rainfall event per week. When rainfall data from several surrounding stations are averaged, the rainfall peaks are levelled off and the number of rainy days increases. Therefore the CGMS takes for each climatic grid cell the rainfall of the most similar station i.e. the weather station with the lowest score for the concerned grid cell. The rainfall data as well as the weather station, which has been used to determine the rainfall, and the similarity score of this station are stored in the tables **GRID_WEATHER** and **STATIONS_PER_GRID**. In case of the current year the weather station, which has been used to determine the rainfall, and the similarity score of this station are stored in table **STATIONS_PER_GRID_ CURRENTYEAR**.

2.3.2.4 Interpolation of other weather data

All other data are interpolated using data from one up to four stations. This analysis is done per meteo group. To determine the most suitable set of stations the CGMS calculates a combination score or set score in a similar way as the single station score (see equation 2-5). It is based on the mean scores of a set of stations. The set score is calculated as follows:

$$SetScore = distavg + Daltavg * Walt + DdCstcorravg + DCG + FnS * Scoremin \qquad (2\text{-}6)$$

SetScore	: suitability score of a set of weather stations in relation to the concerned grid cell centre	[km]
distavg	: average distance between the weather stations and the grid centre	[km]
Daltavg	: average of the absolute difference in altitude	[m]
Walt	: weighting factor for *Dalt* (= 0.5)	[km.m^{-1}]
DdCstcorr	: average of the absolute difference in corrected distance to coast	[km]
DCG	: distance between grid centre and centre of gravity of set of stations	[km]
FnS	: factor based on the number of stations in the set	[-]
Scoremin	: minimum single station score from the complete set of scores	[km]

The criteria DCG guarantees an optimisation of the distribution of the selected stations around the concerned climatic grid cell centre. The factor *FnS* decreases as the number of weather stations in the set increases. It is 0 for three or four weather stations in the set, 0.2 when two weather stations are used and 0.5 for a set consisting of only one single weather station. The term *FnS*Scoremin* is used to balance the importance of the number of weather stations in relation to the other components.

Theoretically the set score could be calculated for all possible combinations of one up to four weather stations, taken from all available weather stations. However obviously this would lead to many unnecessary calculations. Therefore the CGMS determines the set score only for all combinations of one up to 4 weather stations, taken from the seven weather stations most similar to the climatic grid cell centre. This results in 98 sets for which the set score has to be calculated.

Once the CGMS has determined the best set of weather stations for each climatic grid cell, as expressed by the minimum set score, data are simply averaged. Temperature and vapour pressure are corrected for the difference in altitude between the selected weather station and the climatic grid cell centre. The correction factors used for the temperature and vapour pressure are respectively -0.006 (°C.m^{-1}) and -0.00025 (hPa.m^{-1}). The interpolated data, the set of weather stations used for the interpolation and the set score are stored in the tables **GRID_WEATHER** and **STATIONS_PER_GRID**. In case of the current year the weather stations, which has been used for the interpolation and the set score of this station are stored in table **STATIONS_PER_GRID_CURRENTYEAR.**

The averaging is carried out without weighting for distance, because in a comparative test it appeared that weighting did not improve the accuracy. This is not surprising, because the procedure for selecting the optimum set of stations contains already a weighting element (van Diepen, 1998). It must be stressed that the consequence of the interpolation is that values obtained for each climatic grid cell represents an ‘average’ daily condition. They do not necessarily represent meteorological conditions that could be measured at the climatic

grid cell centre. For instance, the altitude used is not the altitude that can be measured at the climatic grid cell centre, but rather a value that represents the mean altitude of the agricultural activity in the concerned cell (Genovese, 2001).

2.3.2.5 Missing station data

As mentioned before for historic years, an availability threshold of 80% is applied to weather stations used for the interpolation of meteorological data. This implies that a weather station can have missing data for a number of days. In such cases missing values are substituted with the long-term average of the concerned weather station and day. A separate procedure, not included in CGMS, is used to computes these long-term averages every year using the complete set of historic data after which the data are stored in table **REFERENCE_WEATHER**. The calculation of the average for each Julian day and station is based on a 15-days time-window (+/-7 days around the selected day) and averaging all the data that are present in that window. Currently no threshold is applied which means the calculation of an average can be based on a small number of observations. The determination of the threshold is difficult. If the threshold is set too high, some days of some stations could be excluded from table **REFERENCE_WEATHER**. This could have consequences for the spatial coverage of the interpolated station weather. If for a specific station and day the value is missing the long term average value is taken. If for this combination of station and day not have enough data exist to calculate a reference (number of observations lower than the threshold), this combination will be excluded from table **REFERENCE_WEATHER**. If for a certain combination of day and station the long term average value is not available, the missing value cannot be replaced by the long term average. In that case the climatic grid cells which use this station for the interpolation will have a missing value too. Because the crop growth model WOFOST needs a complete set of weather data each year, the CGMS skips the climatic grid cells with missing values. As a consequence these grid cells are not written in the tables **GRID_WEATHER**.

When interpolating weather data in the current year one could also face situations where a large area has no station data for one specific day. In such case climatic grid cells will not have any weather data and the crop simulation will stop as soon as a day without weather data is found. Therefore the interpolation algorithm for the current year uses long term average weather of reliable stations from table **REFERENCE_WEATHER** in case there are no observations for a certain grid cell on a specific day within a range of 250 km. In the current year the available observations within a range of 250 km for a specific day are analyzed per meteo group. In case one meteo group has no observations long term average data are taken from reliable stations within the same range of 250 km. For each meteo group this analysis is done separately. This can lead to situations where a grid cell could get a kind of “mixed” weather: for one meteo group the long term average values are used while for the other the actual values are used. It has been concluded that the “mixing” effect does not harm the crop simulation or the analysis while in this case optimum use is being made of all available weather data. In case there are also no long term average data of reliable stations for that specific grid cell on a specific day, the grid cell will be excluded for that day. Reliable stations are stations that have at least 50% of observations for the last ten years for each meteo group. The reliability stamp is stored in table **WEATHER_STATION**.

2.4 Alternative entry for grid weather data

CGMS has the possibility to enter grid weather through other tables than table GRID_WEATHER. The table **GRID_WEATHER_MODEL_1**, **GRID_WEATHER_MODEL_10** and **GRID_WEATHER_MODEL_30** store respectively model weather with a time step of a

day, ten days or a month. Only mean temperature is extra and the 10-daily and monthly tables have a different date field. Because the crop growth model WOFOST in CGMS does need daily weather data, it is necessary to convert 10-daily and monthly data into daily values. Variables like temperature, given on a monthly or dekad basis, are simply interpolated to obtain daily values. A table **RAINY_DAYS** has been introduced which stores long term average rainy days per dekad and climatic grid cell. The rainy days are necessary to calculate daily rainfall by distributing the given monthly or dekadal rainfall over these rainy days.

3 Thematic description of CGMS level 2

The crop monitoring component produces simulated crop indicators like biomass and yields to show the effect of recent weather on crop growth. The work is divided into four activities (see Figure 3-1). Only the last two are part of the operational services, while the first two are pre-processing tasks:

- Collection and processing of input data.
- Spatial schematisation.
- Regional crop simulation.
- Spatial aggregation.

Resulting crop indicators can be viewed through the tool CGMS viewer.

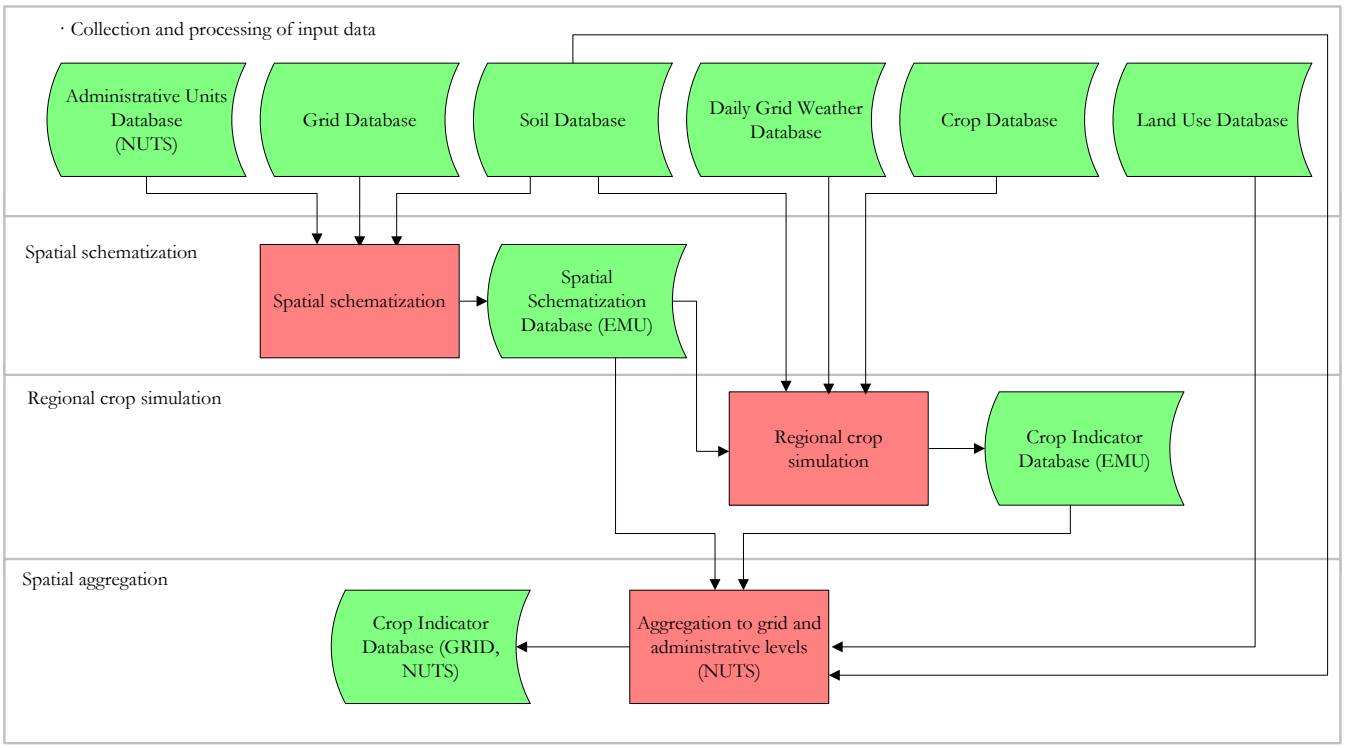


Figure 3-1: Overview of the crop monitoring components of CGMS

For a detailed background the reader is referred to Lazar and Genovese (2004) and in Supit et al. (1994). In the following paragraphs CGMS tables are mentioned. A detailed description of the tables can be found in Chapter 6.

3.1 Goal and assumptions

The CGMS uses the daily interpolated grid weather to simulate biomass accumulation and crop development. Besides regional monitoring of the crop condition, this component issues alarm warnings in the case of abnormal conditions. The outcome of the crop monitoring part is also one of the inputs for the yield prediction.

Van Diepen and van der Wal (1995) described crop growth as a complex process which takes place on farms at field level. Crop yields vary among regions, farms, fields, and years.

Many different factors influence the process. The cause of variation in crop yield may be sought in factors such as:

- A-biotic: weather, soil type.
- Farm management: soil tillage, planting density, sowing date, weeding intensity, fertiliser rates, crop protection against pests and diseases, harvest techniques, post harvest losses, degree of mechanisation.
- Land development: field size, terracing, drainage, irrigation.
- Socio-economic: distance to markets, population pressure, investments, costs of inputs, prices of outputs, education level, skills, infrastructure.

The crop monitoring component of the CGMS studies the influence of one factor, weather, assuming implicitly that the influence of all other omitted factors is constant. However, the results of the analysis cannot be conclusive, when yield is co-determined by factors kept outside the analysis. Furthermore, the influence of these factors may be completely overruled when the overall economic and political situation is not stable, or when crop-damaging catastrophes occur, such as warfare, flooding, earthquakes etc. Therefore in the final synthesis not only results of the CGMS should be included but also other sources.

Many of the omitted factors are important at local scale and may lead to variations in yields. The CGMS assumes that at regional level the influence of these factors is compensated by each other (van Diepen and van der Wal, 1995). Additionally, the CGMS assumes that effects of weather on yield is regardless the degree of fertilisation (Supit, 1999).

The majority of the relations between plant growth and agro-meteorological growing conditions are non-linear. This non-linearity does not allow to first aggregate the input parameters to regional level and next simulate the crop growth at this regional level (Vossen, 1995). Therefore crop growth simulation takes place at the detailed spatial level where distinct areas are more or less homogeneous regarding weather, soil, and crop management.

The soil suitability mode of CGMS assumes that the cultivated area of each crop is evenly distributed over all suitable soils, which implies that the production volume increases linearly with planted area. In the soil suitability mode the crop yield is estimated for all land that is considered suitable for this specific crop. In reality a given crop could be grown on marginal soils, for instance if the best soils are occupied with more profitable crops. And if the crop is grown on the most suitable soils the inter-annual fluctuation of crop acreage could involve only marginal soils. These assumptions should be kept in mind when calculating and studying production volumes which are the straight multiplication of crop yield and planted area.

3.2 Collection and processing of input data

3.2.1 Grid Weather

The daily weather data, interpolated to a regular grid, is the main input for the crop growth model WOFOST. This subject is described in Chapter 2.

3.2.2 Crop Data

In the CGMS (the table **CROP**) a selection of crops is included. The crop parameters of the CGMS fall broadly into two categories. The first category of data contains the data that describes the various crops for the WOFOST model. These so-called crop parameters describe characteristics of the crop like for example the threshold temperature sum from emergence to anthesis, or the leaf area index at emergence. The information is stored in the

table **CROP_PARAMETER_VALUE**. The table **PARAMETER_DESCRIPTION** explains the parameters.

The second category describes the spatial and temporal variation in crop use, for instance which plant is used in a particular location and what the average sowing and harvest dates are for this plant. For crops which are harvested before maturity (green crops) the average harvest date is specified as well. The data are linked to climatic grid cells, and are stored in the table **CROP_CALENDAR**. The start of the crop in table **CROP_CALENDAR** can have three different start types:

- VARIABLE_SOWING: The sowing date is determined by the program. The program starts evaluating sowing conditions 10 days before the earliest sowing date (given by START_MONTH1 and START_MONTHDAY1) and will return the earliest sowing date within the period between the earliest and the latest sowing date (defined by START_MONTH2 and START_MONTHDAY2). Emergence is calculated using the effective daily temperature (defined by the crop parameters TBASEM, TEFFMX and TSUMEM).
- FIXED_SOWING: The sowing date is given by START_MONTH1 and START_MONTHDAY1. Emergence is calculated using the effective daily temperature (defined by the crop parameters TBASEM, TEFFMX and TSUMEM).
- FIXED_EMERGENCE: Emergence takes place on the day given by START_MONTH1 and START_MONTHDAY1.

The end of the simulation is determined by the end type, which can take the following values:

- HARVEST: The simulation stops at the date given by END_MONTHDAY and END_MONTH. If maturity is reached before this date, the model stops at maturity. This option is useful for crops that are harvested in vegetative state, e.g. sugar beet. (MAX_DURATION is not used)
- MATURITY: The simulation stops at maturity, but the simulation will not exceed MAX_DURATION days after emergence. The use of MAX_DURATION will prevent anomalies when a crop never reaches maturity due to low temperatures. (END_MONTHDAY and END_MONTH are not used)
- EARLIEST: The simulation stops at the earliest of maturity, end date or maximum duration.

Different climatic grid cells, which have the same crop, can have different varieties. This is also indicated in table CROP_CALENDAR. These varieties can be different for instance for the required temperature sum to reach a certain crop development stage. Such crop parameters that differ from the original crop (also called basic variety) are stored in the table **VARIETY_PARAMETER_VALUE**. So when a crop for a certain climatic grid cell is basically the same as an existing crop (in terms of the WOFOST crop parameters) it is convenient to express the differences by introducing only those parameters that are different in the variety parameter table. If however the crop is significantly different, it is more convenient (and more logical) to introduce a completely new crop in the crop parameter table. It is important to realise that for the system this is no different, since theoretically all of the crop parameter values can be overwritten by a variety parameter value.

The crop parameters fall into two categories. One category contains the parameters that can be expressed as a single number, e.g. TSUM1, the temperature sum from emergence to anthesis. The other category contains the parameters that can be expressed as a function of another variable, e.g. SLATB, the specific leaf area as a function of DVS. These parameters are expressed as a set of value pairs (x,y) that describe the shape of the function. A special interpolation function (AFGEN) within CGMS will perform an interpolation to obtain the

function values for arbitrary inputs. The naming convention for the multiple parameters is the name of the parameter, post fixed by '_XX'. The values of XX range from 01 to 10.

The rules for interpolation by the AFGEN function are as follows:

- If the argument (x-value) is less than the first x-value in the descriptive array return the first y-value.
- If the argument is between two x-values, return the linear interpolation of the corresponding y-values.
- If the argument is larger than the last x-value, return the last y-value.

When building the crop database for CGMS Europe information was often available at the wrong scale. For example, the sowing dates for a region are often known only for a small sample of fields. Information how representative these samples are, is usually unavailable while this information is needed for accurate scaling up from the site-specific information to a NUTS level. Similarly, values of crop modelling parameters obtained from individual trials will differ from those that would have been obtained if a complete enumeration had been achieved. This is of particular importance wherever the relations are not linear (Vossen and Rijks, 1995). Another problem is that, although the inventories were carefully compiled, information frequently is not available for certain parameters and in certain regions or countries. In such cases a 'best guess' was made.

Each crop is assigned to one of the following groups: cereals, root crops and maize (see the table CROP_GROUP). These groups describe the requirements of a crop with respect to soil characteristics such as agricultural limiting phase, texture, slope, drainage, rooting depth, alkalinity and salinity. The requirements determine whether a soil is suitable to grow a specific plant which is relevant if data are aggregated according the soil suitability mode.

In case of LINGRA the table CROP_**CALENDAR** is filled as follows. The start type is "FIXED_EMERGENCE", the growing starts at the first of January and ends at 31st of December with end type is "HARVEST". The maximum duration can be set on 365 days. From thematic point of view the crop calendar is not needed to run LINGRA as grass is growing continuously on the field. However CGMS needs the information otherwise the crop cannot be simulated.

3.2.3 Soil Data

The digital soil map is one of the most complicated sources of information for CGMS. However, as far as the GIS aspect is concerned, its use is primarily to indicate the soil mapping units. These are the smallest cartographic units on the soil map. The scale of the soil map has a direct impact on the number of elementary mapping units, and therefore a direct impact on the size of the system and the processing time required.

The information associated with the soil mapping units, or simply the soil data, is used in two ways in CGMS. Firstly it is used to determine the physical soil characteristics taken into account by the simulation. These characteristics determine fundamental parameters used by the crop simulation, e.g. soil moisture content, seepage parameters, rooting depth etc. Secondly, the soil data can be used to determine whether or not to perform a simulation for a soil mapping unit for a particular crop. The latter goal depends on the availability of actual land-use data. If such data are missing the decision to simulate is simply based on the suitability of a particular soil for a particular crop. If at least part of the soil mapping unit is deemed suitable then the simulation will be performed.

Soil mapping units (table SOIL_MAPPING_UNIT) are considered to consist of one or more soil types (table SOIL_TYPOLOGIC_UNIT) that, at the scale of the soil map in use, are not being mapped as individual cartographic units. These sub-units are called 'soil typologic units'. The soil mapping unit is described by the percentages that the different STUs occupy in the SMU. This information is stored in the soil_association_compostion table.

The soil map is used to determine the rooting depth and the soil physical group which are needed to assess whether the crop will suffer from droughts. Each STU has one soil physical group defining the available water capacity (AWC) and the infiltration capacity. The AWC is a static soil characteristic and gives the amount of water between field capacity (wet soil) and wilting point (no water available for plants anymore) per unit length rooting depth. Multiplication of AWC and rooting depth gives the maximum available water which a soil can supply to a plant during a period of drought. It should be noted that the rain fed crop yields of the CGMS are more sensitive to the rooting depth than to the soil physical group (van der Goot, 1998b). The CGMS stores these data in the tables ROOTING_DEPTH and SOIL_PHYSICAL_GROUP. Be aware that the number of soil physical groups or rooting depth classes is not fixed. If the soil characteristics to be mapped cannot be described by any of the available soil physical groups or rooting depth classes, a new soil physical group or rooting depth class can be constructed

In the original CGMS version (unix operating system) a subroutine ("watgw.for") was included to simulate influence of groundwater. This subroutine was first implemented in the windows CGMS version 8.0 (Savin et al., 2004). The table **SOIL_PHYSICAL_GROUP** has been extended with data on additional soil physical data like water retention and hydraulic conductivity curves. The table **INITIAL_SOIL_WATER** is used to supply data on initial ground water level and drainage depth.

In case of ground water influence the initial soil moisture profile is not based on the WAV parameter (available water in potential rooting zone) given in table **INITIAL_SOIL_WATER** but this profile is calculated assuming a equilibrium situation in relation to the ground water level. If the ground water level is deep the initial soil moisture profile will be very dry. To avoid such a dry start the ground water module will not be used if the initial ground water table (ZTI) is already deeper than 600 cm below soil surface.

The parameters that describe the infiltration are stored in table **SITE**. They describe the redistribution or loss of rainfall due to run off and surface storage. These parameters are system wide and have no linkage with soil mapping units.

In case the aggregation is done through the soil suitability mode suitable soils are determined per crop group on the basis of crop growth limiting properties of these soils. Examples of limiting soil properties are slope, texture, agriculture limiting phase, rooting depth, drainage, salinity and alkalinity. Next, pedotransfer rules can be used to determine whether a soil is suitable for root crops, cereals and grain maize based on these limiting properties.

The suitable STU's and the percentage of the suitable area of SMU's are available in the tables **SUITABILITY** and **SMU_SUITABILITY**. Because a SMU can consist of more than one STU the percentage suitable area must be calculated.

3.2.4 Initial Available Soil Water

The spatial and temporal variability of initial soil moisture can be entered in the table **INITIAL_SOIL_WATER**. Two different issues can be set:

- The amount of available soil water (field WAV) used to initialize the potential rooting depth. Note that WAV is the amount of water added to the potential rooting depth in addition to the amount of water that is already stored at pressure head equals wilting point. When the available water is more than the potential root zone can contain, the surplus of the initial water is supposed to percolate to greater depth. For all crops except rice the maximum soil moisture content is the soil moisture content at field capacity while for rice it is the soil moisture content at saturation.
- The start of the soil water balance. The possible choices offered through the user interface are:
  - "No initialisation" = no initialization and thus the soil water balance start at the emergence date
  - "Automatic, use all weather prior to emergence" = this means automated initialization using the available grid weather from the first day of the campaign. Note that the campaign year is defined in the user interface.
  - "Fixed number of days prior to emergence" = specifies a fixed number of days of days prior to emergence. Note that the number is limited for the available grid weather from the first day of the campaign
  - "Fixed date available in table INITIAL_SOIL_WATER". In this case the CGMS reads the fixed date from the field GIVEN_STARTDATE_WATBAL in table **INITIAL_SOIL_WATER**.

So the WAV determines the initial available water (between wilting point and field capacity) in the potential rooting depth. Besides it is possible to initialize the soil moisture prior to emergence. This will lead to a better estimate of the soil moisture in the initial rooting depth. In case the potential rooting depth, below the initial rooting depth, has a lower soil moisture content than the content related to field capacity this part of the potential rooting depth will change and become wetter in case of a climatic rainfall excess that cannot be stored in the initial rooting depth.

3.2.5 Administrative Data / Area Statistics

The administrative units are used to relate the calculated results from CGMS to real world data. It is therefore important to choose the administrative units for this part of the system in such a way that relevant data is actually available for these units. It is always possible to 'aggregate' the information for smaller units to larger units, e.g. region to province, or province to country. At this stage the smallest useful unit should be considered. As with the soil map, the size and, therefore the number, of these units have a direct impact on the number of elementary mapping units.

The simulated crop indicators are aggregated from simulation units to administrative regions so that they can be used as regional yield predictors. The administrative regions in EU are called Nomenclature des Unités Territoriales Statistiques (NUTS). The NUTS system is organised as follows: the highest level, the whole country, is called NUTS-0, which is divided in macro regions: NUTS-1. These macro regions are subdivided in NUTS-2 sub-regions. The CGMS keeps these data in the table **NUTS**.

Preferably acreage statistics of administrative regions (in stead of soil suitability data) are used to aggregate simulated crop indicators from the smallest administrative level 2 to administrative level 1 and 0. These data can be found in the table **AGGREGATION_AREAS**. Missing records for an administrative region are added through aggregation of information if the lower administrative level regions (the 'children' of the administrative region) all have acreage data. In case a group of administrative regions, which belong to the same higher

level administrative region ('parent'), still has (a) missing record(s), the long term average values or soil suitability information can be used. If all available, the acreage values of the whole group are taken from their long term average values. Otherwise these values are based on soil suitability information. Note that always the whole group is changed so that the crop acreage of the individual administrative regions of one 'parent' is based on either actual values or long term average values or soil suitability information. This constraint is needed to avoid inconsistencies in the spatial aggregation of simulated crop indicators.

3.3 Spatial schematization for simulation

The following text is mainly based on the document written by van der Goot (1998b). The crop growth model WOFOST of the CGMS is a point model. To apply this model at a larger scale, areas where meteorological data, soil characteristics and crop parameters can be assumed homogeneous have to be identified. It is assumed that the simulated crop growth is representative for those areas. Furthermore, the model output (simulated yield indicators etc.) should be representative for an administrative region for which statistical yield data are available and for which the yield has to be forecasted. The regional yield statistics are normally available at national (level 0) or at provincial level (level 1 or 2). The areas for which soil characteristics are constant raise more a problem. Most soil maps have a smallest cartographic unit, the 'soil mapping unit' (SMU), although often it consists of various Soil Typologic Units. Still, the smallest unit that can be defined for the soil data is a SMU. Finally, we need a definition of the area for which the meteorological data can be assumed to be homogeneous. For this purpose, the area of interest is divided into regular climatic grid cells. The crop data are linked to the climatic grid cells, assuming these crop data (varieties, planting dates, crop parameters etc.) do not change within a climatic grid cell.

Two data layers (climatic grid cell and SMU) are intersected which results in so-called Elementary Mapping Unit's (EMU) (see **Figure 3-2**). This is the smallest unit for which the CGMS produces results. The elementary mapping unit is fundamental to the system, and any change in any of these two components has major implications for the system. The EMU is not given a unique identifier in the database, but is stored as a (unique) combination of its properties, i.e. soil mapping unit number and grid identifier plus area in the table **ELEMENTARY_MAPPING_UNIT.**

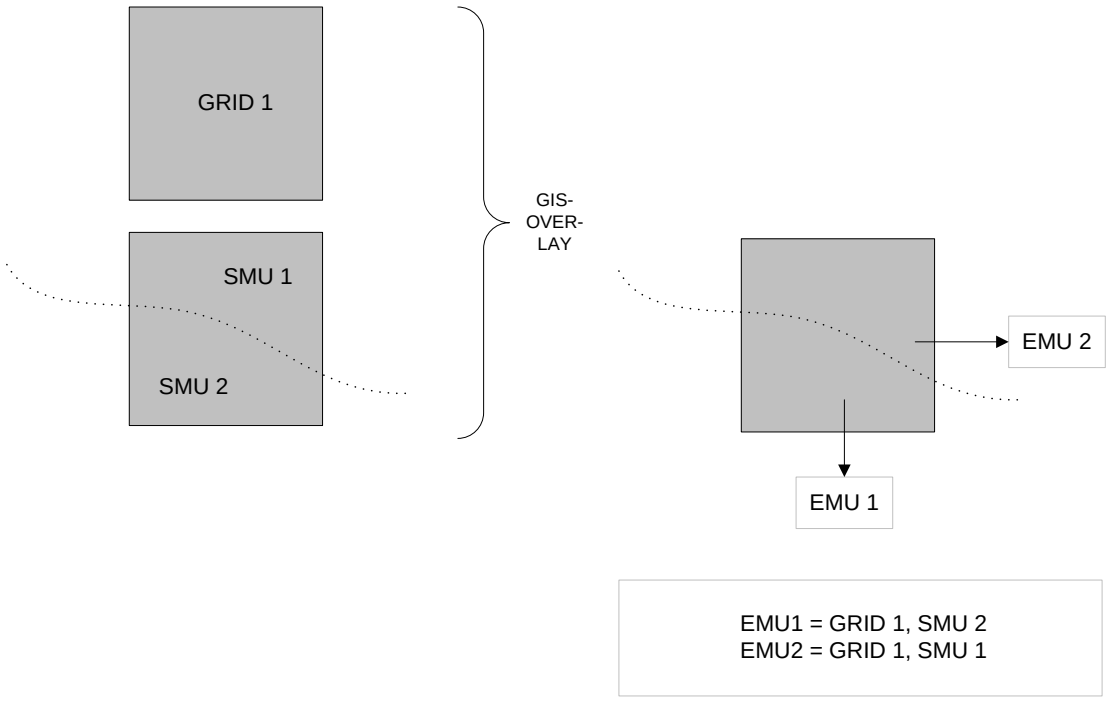


***Figure 3-2**: In the CGMS the Elementary Mapping Unit (EMU) is defined as the intersection of climatic grid cells and Soil Mapping Units (SMU's).*

The crop growth simulations are performed for all unique suitable STUs in a climatic grid cells. These simulations units, stored in the table **SIMULATION_UNIT**. Note that the user has the possibility to exclude STUs based on soil suitability rules. This information is stored in table **SUITABILITY.** After the crop growth simulation, results of simulation units are converted to results at EMU level, again.

The CGMS does not directly need GIS software to produce its results. However, a GIS is necessary for a meaningful presentation of the results, and is also indispensable for the initial creation of the soil, grid and EMU database. The interpolated meteorological data are stored at climatic grid cell level, the simulated yields are stored at EMU level and aggregated to climatic grid cells or various levels of administrative regions. EMU's, climatic grid cells and administrative regions can be linked to a GIS. CGMS does not impose the use of any particular geographic or co-ordinate system. However, all geographic data used in the same CGMS installation should be compatible.

When producing the EMU coverage, care must be taken to carefully clip the individual coverages. A straightforward intersection of the two underlying datasets could produce many artefacts, especially around the country borders and shoreline, and result in EMUs that are not properly defined.

3.4 Regional Crop simulation

The heart of the CGMS two crop growth models do exist: WOFOST and LINGRA.

3.4.1 WOFOST

The crop growth simulation model WOFOST is dynamic, explanatory point model. In the CGMS this point model is applied for each individual simulation unit. The core of the WOFOST's crop growth sub model has been taken from the SUCROS model (Spitters *et al.*, 1989; Van Laar *et al.*, 1992). The principles of WOFOST's water module are described in Van Keulen and Wolf (1986). The initial version of this model was developed by the Centre for World Food Studies and AB-DLO (van Diepen *et al.*, 1988; 1989). In the CGMS, WOFOST version 6.0 has been used (Hijmans *et al.*, 1994).

In WOFOST, crop growth is simulated on the basis of eco-physiological processes. The major processes are phenological development, CO_2-assimilation, transpiration, respiration, partitioning of assimilates to the various organs, and dry matter formation. This is illustrated in **Figure 3-3**. Potential and water-limited growth is simulated dynamically, with a time step of one day.

In the CGMS WOFOST simulates two production levels: potential and water-limited. The potential situation is only defined by temperature, day length, solar radiation and crop parameters (e.g. leaf area dynamics, assimilation characteristics, dry matter partitioning, etc.). For this situation the effect of soil moisture on crop growth is not considered and a continuously moist soil is assumed. The crop water requirement, which in this case is equal to the water consumption, is quantified as the sum of crop transpiration and evaporation from

the shaded soil under a canopy. To calculate the potential crop growth, the soil parameters rooting depth and soil physical group are not needed. Therefore in a climatic grid cell all EMU's have the same simulation results for the potential situation.

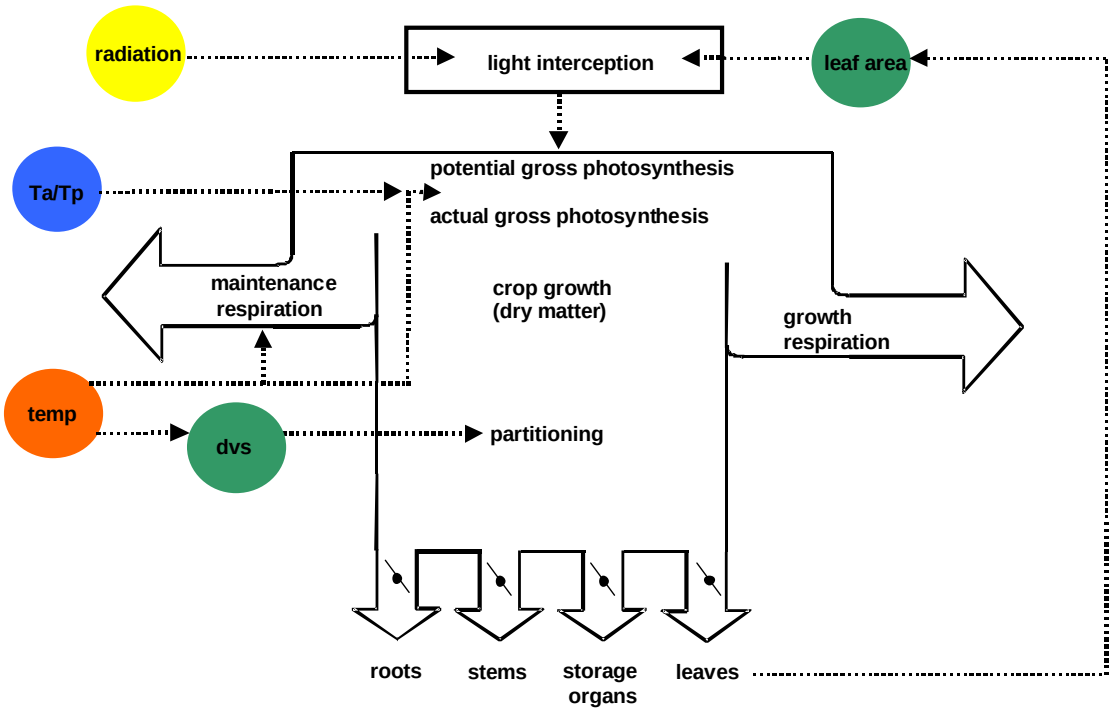


Figure 3-3: Crop growth processes ('Ta' and 'Tp' are actual and potential evapotranspiration rate; 'temp' is temperature and 'dvs' is development stage) (Kropff and van Laar, 1993).

In the water-limited situation soil moisture determines whether the crop growth is limited by drought stress. In both, the potential and water limited, situations optimal supply of nutrients is assumed. For each situation, dry matter per hectare of above-ground biomass and storage organs such as grains and roots (potatoes and sugar beets) are simulated from sowing to maturity or harvest on the basis of physiological processes as determined by the crop's response to daily weather, soil moisture status and management practices (i.e. sowing density, planting date, etc.). For this simulation the CGMS Windows program is used. The required inputs for WOFOST per simulation unit are daily weather data, soil characteristics, crop parameters and management practices. The output data are stored in table **CROP_YIELD**. To save disk space the results on EMU level can be saved for the last day of a dekad in stead of each day. The results are stored for each simulation period, defined by a start and end decade, requested by the user. Note that the simulation itself is always carried out from the beginning of the growing season.

In case the groundwater influence is simulated the fields FSMUR, RUNOFF, SOIL_EVAPORATION and LOSS_TO_SUBSOIL are not calculated. However CGMS write value "0" into the **CROP_YIELD** table. The reason why the fields are not calculated is simple. The fields were introduced because of a soil water balance study assuming free drainage conditions. At that time there was no need and time to implement this structure into the ground water module as well. Thus when simulating crop yield with the ground water influence option these fields should be ignored.

Implementation of WOFOST in the CGMS and its structure is described by Supit *et al.* (1994) and can be found on internet :

- http://www.treemail.nl/download/treebook7/index.htm
- http://agrifish.jrc.it/marsstat/Crop%5FYield%5FForecasting/METAMP/

Technical descriptions and user manuals have been prepared by van Raaij and van der Wal (1994), van der Wal, (1994), Hooijer *et al.* (1993). Further development of the individual WOFOST model took place in the framework of other projects. This resulted in WOFOST version 7.1 which has never been implemented into the CGMS. Main changes in relation to version 6.0 were the development of a user friendly graphical user interface (WOFOST Control Centre version 1.5), the use of FSEOPT for calibrating crop parameters and the removal of minor bugs. Only minor thematic changes took place between WOFOST 6.0 and 7.1 (Boogaard *et al.* (1998)).

3.4.2 LINGRA

The GRASSLAND GROWTH MODEL - LINGRA (LINTUL GRAssland) was developed to predict growth and development of perennial rye grass across the member states of the EC at the level of potential production and water-limited production. The model is based on the LINTUL (Light INTerception and UtiLisation simulator) concepts as proposed by Spitters (Schapendonk et al., 1998; Bouman et al., 1996). The main principle of this concept is that crop growth is proportional to the amount of light intercepted by canopy.

The integration level is kept high and the number of processes has been restricted to key parameters, and only a small number of processes involving these parameters are dynamically simulated. On the other hand, parameters that have relatively little impact on crop growth, or which knowledge is scarce, have been treated using a static approach.

Common modules with CGMS are: soil water balance and weather parameters (ETP).

In contrast to arable crops, the grassland plants are frequently defoliated due to grazing or management activities. The consequence of defoliation is reduction of photosynthesis rate. After defoliation, new leaves must be formed in order to assure continuation of production. The formation of the new leaves is based on amounts of carbohydrates stored in the stubble of the plant before defoliation. This induces an alternation of the periods with assimilate shortage with periods when the surplus of assimilate is stored. This process is strongly influenced by environmental conditions and cultural practices.

Assimilate demand (the sink) is associated with leaves elongation, leaf appearance and tillering rate, where assimilate supply (the source) is controlled by photosynthesis which is depending on the amount of light that is intercepted by canopy. In LINGRA, the dynamic fluctuation of assimilate demand (ΔW_d) and the assimilate supply (ΔW_s) are simulated semi-independently. The term “semi-independent” is used because each day, crop-growth rate is estimated from the most limiting process, either ΔW_d or ΔW_s as driving rate variable. All other state variables are derived from the growth rate at that particular day and are not integrated independently for source or sink limitations.

Both hypothetical levels of potential production (depending only on intercepted solar radiation and temperature) and water-limited production are simulated. Soil nutrients are considered to be at optimal level and there is no simulation of mineral nutrition. Also, the effects of pests, diseases and weeds are not taken into consideration.

A description of the model LINGRA model as implemented in CGMS, may be found in Bouman et al., 1996.

The required inputs for LINGRA per simulation unit are daily weather data, soil characteristics, crop parameters and management practices. The output data are stored in table **CROP_YIELD**. To save disk space the results on EMU level are only saved for the last day of a decade. The results are stored for each simulation period, defined by a start and end decade, requested by the user. Note that the simulation itself is always carried out from the beginning of the growing season.

3.5 Spatial schematization for aggregation

By making a clear separation between simulation and aggregation, unnecessary reprocessing of crop growth is avoided when land cover maps and/or administrative regions changes. Two different weighing schemes can be used in the aggregation from EMU to climatic grid cell and/or administrative region:

- Weights based on soil suitability
- Weights based on crop areas

These two different weighing schemes need also two different spatial schematizations. When using soil suitability the following three data layers have to be intersected: climatic grid cell, SMU and administrative region at level 2 (see **Figure 3-4**). The distinct unique combination of climatic grid cell, SMU and administrative region at level 2 plus related area are stored in table **EMU_PLUS_NUTS.** In case the user only aggregates to climatic grid cell the data in table **ELEMENTARY_MAPPING_UNIT** is sufficient.

When applying crop areas as weights in the aggregation the spatial schematization given in table **EMU_PLUS_NUTS** has to be corrected for crop areas. For each Elementary Aggregation Unit (combination of climatic grid cell, SMU and administrative region at level 2) in table **EMU_PLUS_NUTS** that area has be to corrected for the part covered by the specific crop. This should be done by intersecting the EMU_PLUS_NUTS coverage with the crop specific masks. The distinct unique combination of climatic grid cell, SMU, administrative region at level 2 and crop plus related area are stored in table **EMU_PLUS_NUTS_LANDCOVER.**

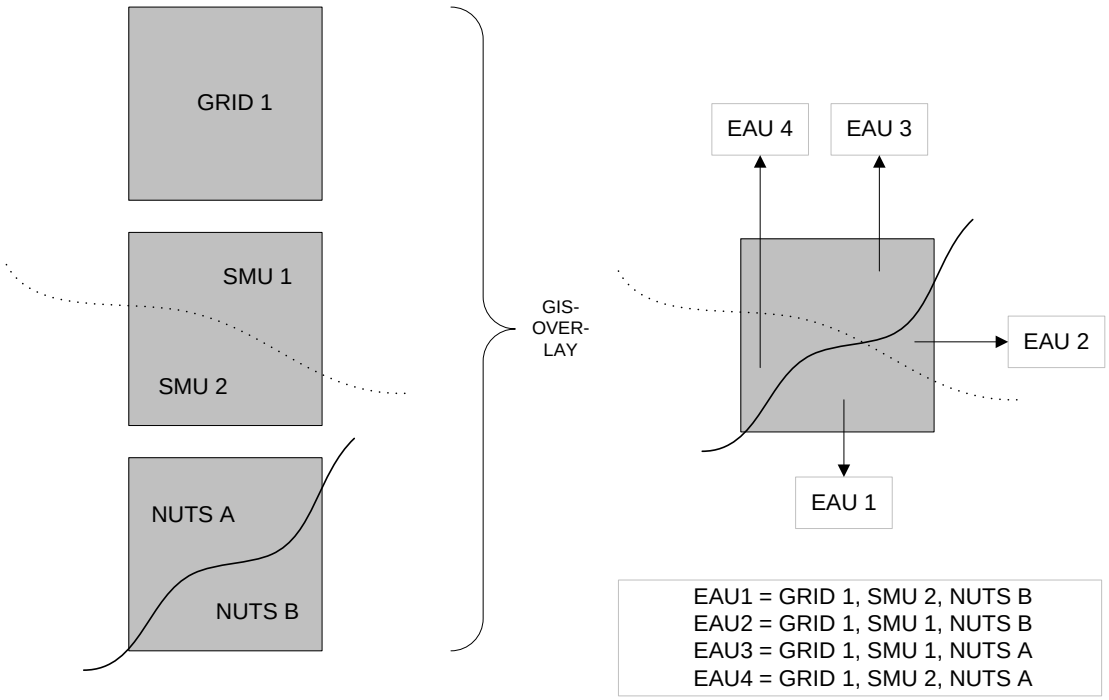


***Figure 3-4**: In the CGMS a spatial schematization exists for aggregation from EMU to the administrative regions at level 2 which is defined as the intersection of climatic grid cells, Soil Mapping Units (SMU's), and administrative regions (Nomenclature des Unités Territoriales Statistiques, NUTS). Note EAU stands for Elementary Aggregation Unit*

3.6 Spatial Aggregation

Simulated crop indicators of the EMU's can be spatially aggregated to the climatic grid cells for the production of crop indicator maps (table **GRID_YIELD**) and to the administrative level 2, 1 and 0 as a basis for yield forecasts (table **NUTS_YIELD**).

The aggregation from EMU to climatic grid cell or administrative region at level 2 is based on the weight of each EMU within the climatic grid cell or administrative region at level 2 as shown by equation (3-9 and 3-10).

The selected spatial schematization determines the weight:

- Weight is the area fraction of the EMU with suitable soils in relation to the total suitable area of all EMU's within the climatic grid cell or administrative region at level 2
- Weight is the area fraction of the EMU which is covered by the selected crop in relation to the total area of all EMU's within the climatic grid cell or administrative region at level 2 covered by the selected crop

$$Y_G = \frac{\sum_{i=1}^{n} C_{E,i} A_{E,i} Y_{E,i}}{\sum_{i=1}^{n} C_{E,i} A_{E,i}} \qquad (3\text{-}9)$$

where:

Y_G	: simulated yield at climatic grid cell	[kg ha^{-1}]
Y_E	: simulated yield at EMU level	[kg ha^{-1}]
A_E	: EMU area	[ha]

C_E : fraction of the EMU area suitable for the crop group
or fraction of the EMU area covered by the selected crop [-]
n : number of EMU's in climatic grid cell [-]

$$Y_{N2} = \frac{\sum_{i=1}^{n} C_{E,i} A_{E,i} Y_{E,i}}{\sum_{i=1}^{n} C_{E,i} A_{E,i}} \qquad (3\text{-}10)$$

where:
Y_{N2} : simulated yield for the administrative region at level 2 [kg ha^{-1}]
n : number of EMU's in the administrative region at level 2 [-]

Finally, simulated crop indicators at administrative level 2 are aggregated to level 1 via:

$$Y_{N1} = \frac{\sum_{i=1}^{n} C_{N2,i} A_{N2,i} Y_{N2,i}}{\sum_{i=1}^{n} C_{N2,i} A_{N2,i}} \qquad (3\text{-}11)$$

where:
Y_{N1} : simulated yield of a region at administrative region level 1 [kg ha^{-1}]
Y_{N2} : simulated yield of a region at administrative region level 2 [kg ha^{-1}]
A_{N2} : crop area of a region at administrative region level 2 [ha]
n : number of sub regions at level 2 for a region at level 1 [-]

Simulated yields at level 0 are obtained in a similar way. Area data can be found in the table **AGGREGATION_AREAS**.

3.7 Simulation winter crops starting in autumn

Usually winter crops start their simulation on 1 January as vernalisation is not described by WOFOST. However from technical point of view it is possible to start the simulation of winter crops in autumn. The following situations after winter are identified:

- Crop failure
- Crop was affected by frost (in the most cases part of leafs or all leafs died)
- No frost effect on crop (crop biomass doesn't changed during the winter)

In the second case the crop will have a delay in the phenological development and a smaller leaf area index. First the idea was to stop the simulation when winter starts and re-start the simulation when winter ends and correct state variables externally. Note that this option has been implemented in CGMS. However some important drawbacks exist with regard to this approach:

- All the current states variables describing the leaf characteristics have to be stored in the database The storage of the arrays m_ASLA[0..365], m_ALVAge[0..365] and

ALV[0..365] leads to enormous storage of data. However these parameters are needed for correction and the re-start after winter.

- The state variables have to be corrected for the amount of winterkill. This correction is complicated. The biomass of leaves is schematized in different leaf classes with their specific age, biomass etc. A correction of the leaf characteristics means that these arrays have to be changed starting with the oldest leaf classes because oldest leaves die first.

Therefore an alternative approach has been implemented to simulate winter crops from autumn without stopping during the winter. This has been organized by inserting one extra parameter (fDSLv3) describing the fraction of the leaves that have died because of severe winter conditions. The value of this parameter is based on the following approach taken from the CERES model (Ritchie, 1991).

First the hardening of a crop is estimated, distinguishing two hardening stages. The first hardening stage requires an average crown temperature between -1 and 8. Ten days at this temperature are required to complete this phase. If the second hardening stage is reached an increase of hardening is established by an average crown temperature below zero. Twelve days of the second phase are assumed to result in a fully hardened plant. Dehardening occurs when the maximum temperature exceeds 10 degrees. The daily dehardening rate is 0.04*(TempCrownMax -10) in stage 2 and 0.02*(TempCrownMax -10) in stage 1.

The killing of a crop by cold and the reduction of the LAI is given by the following code: killing temperature = - 6.0 * (1 + HI). This threshold value is compared with the minimum temperature at crown level. The hardening index (HI) of the crop determines the threshold at which below the minimum crown temperature leads to a killing of the crop
A crop could be damaged partly in case the minimum crown temperature is above this threshold but still significantly below zero because of the low temperatures. In such case the LAI is reduced. The cK is determined as follows:

```
IF (TempCrownMin > -6) Then
        cK = 1
Else
        HelpVar = (0.02 * HI - 0.1) * (TempCrownMin * 0.85 + TempCrownMax * 0.15 + 10 +
                    0.25 * SnowDepth)
        cK = MIN(0.96,MAX(0,HelpVar))
End If
```

The minimum crown temperature must be significantly below zero to get a reduction of LAI. A maximum crown temperature above zero and presence of snow means an even lower minimum temperature (below zero) before reduction of LAI takes place because of the contrasting effect of a maximum crown temperature above zero and snow depth. The minimum crown temperature, maximum crown temperature and snow depth not only determine if there is a reduction but also determine the size of reduction. Besides, the magnitude of the effect is determined by the hardening index (HI): how larger the HI how smaller the reduction of LAI.

The determination of the HI and the damage functions have been implemented in CGMS as follows:

- If (one day) the minimum temperature at crown level is lower than the killing temperature the crop is killed and simulation stops

- If the crop is not killed but a reduction is calculated the biomass of leaves is reduced. The cumulated value of the leaves died by cold is written to the database.

Possible problems have been studied when CGMS continues to simulate crops during the winter. In regions with a temperate climate strange results do occur. Within these regions the mean day temperature during winter is above zero degrees but can drop below zero for a limited number of days. In such climate the phenological status of the crops keeps developing during winter. **Figure 3-5** shows the impact of this development during winter on the LAI, biomass and yield which become too low when starting to simulate the crop in the autumn.

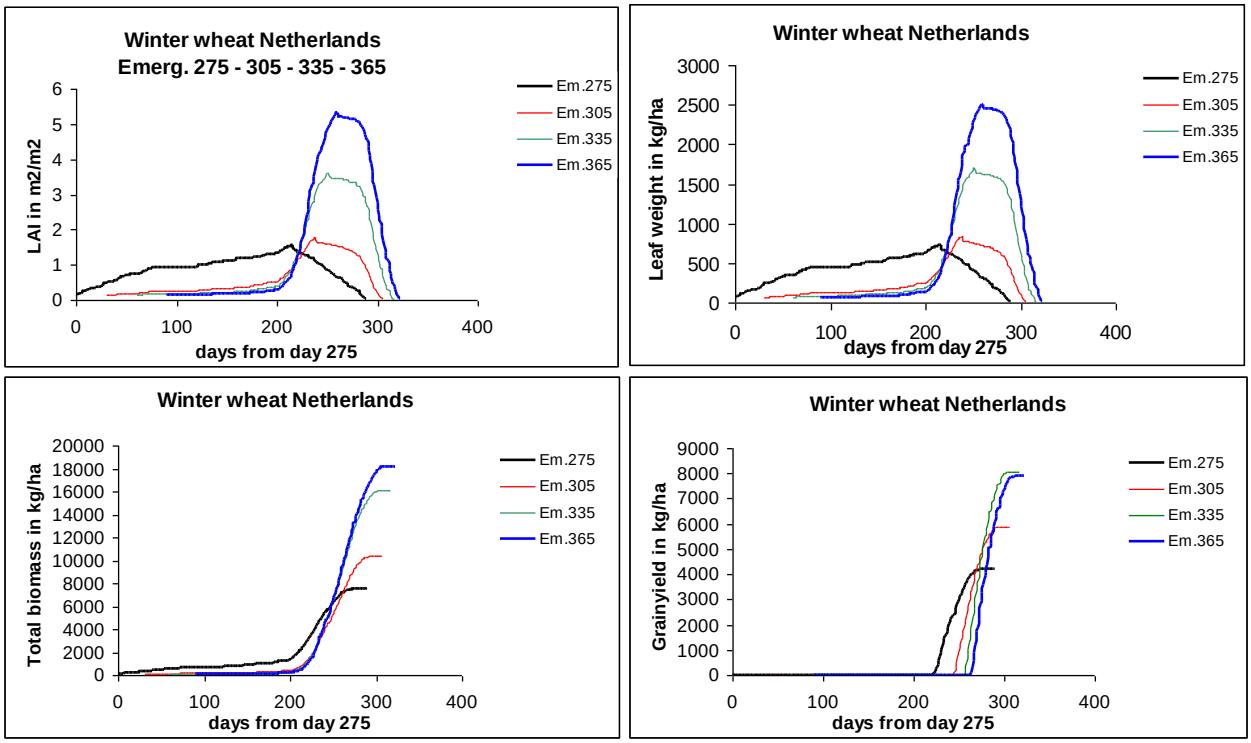


Figure 3-5: Effects on LAI, biomass and yield in the Netherlands when starting to simulate winter wheat at varying days in autumn

But when doing the same simulation runs for a location with a continental climate results are different (see **Figure 3-6**). During the winter the mean daily temperature is below zero and CGMS will not accumulate temperature. In other words: the phenological development is stopped and will continue when spring announces. This proofs that it could be possible to run CGMS during winter time for climatic grid cells with a continental climate.

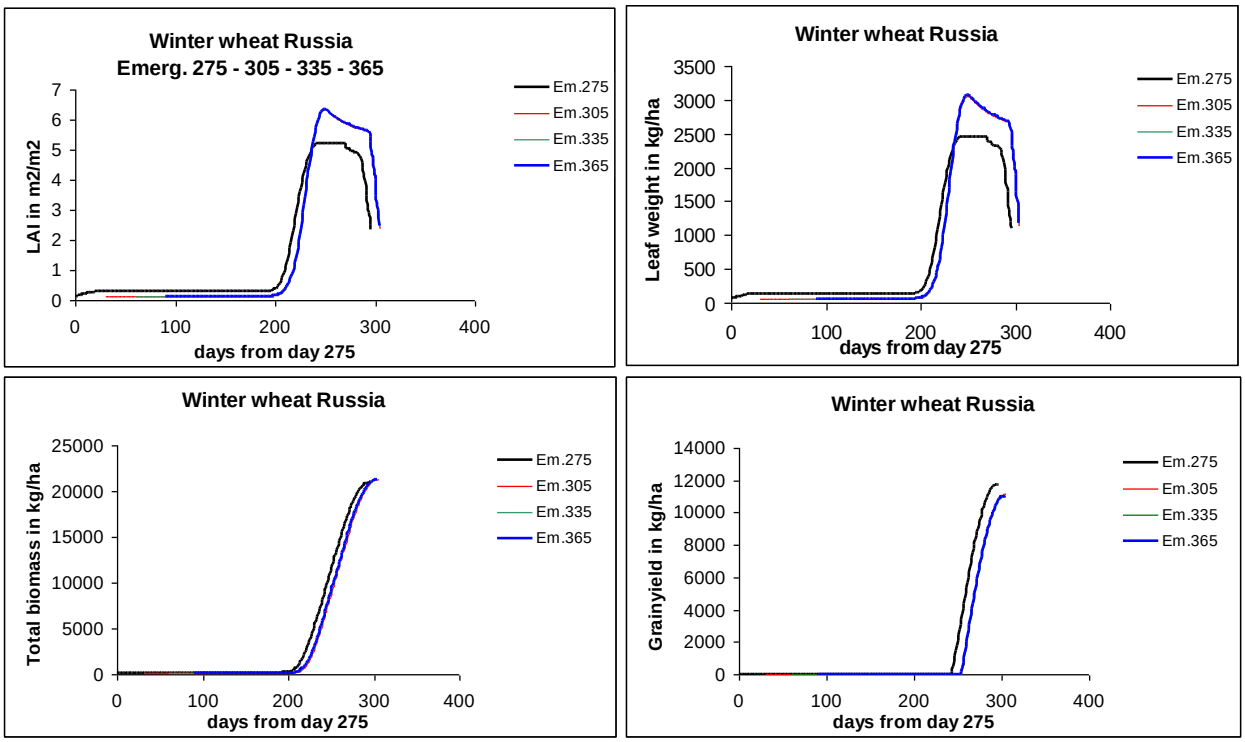


Figure 3-6: Effects on LAI, biomass and yield in Russia when starting to simulate winter wheat at varying days in autumn

.

3.8 Restart crop simulation

In CGMS a so called “restart” functionality is available to correct the biomass of winter crops because of frost damage. It enables the user to stop the simulation at a certain day, correct the biomass and continue the simulation at the same day. However some important drawbacks exist with regard to this approach:

- All the current states variables describing the leaf characteristics have to be stored in the database The storage of the arrays m_ASLA[0..365], m_ALVAge[0..365] and ALV[0..365] leads to enormous storage of data. However these parameters are needed for correction and the re-start after winter.
- The state variables have to be corrected for the amount of winterkill. This correction is complicated. The biomass of leaves is schematized in different leaf classes with their specific age, biomass etc. A correction of the leaf characteristics means that these arrays have to be changed starting with the oldest leaf classes because oldest leaves die first.

Therefore an alternative approach has been implemented to simulate winter crops from autumn without stopping during the winter.

But perhaps the user has another reason for using the restart option and therefore this function is described in the following text.

The user can stop the simulation at a specific day or when a specific condition has been met. When the crop growth model is forced to stop (intermediate stop), values of state variables are written to database table. The values are stored per climatic grid cell, STU, crop and year.

Three simulation modes are distinguished:

- **normal simulation** (start day is based on the crop calendar, no intermediate stop).
- **intermediate stop** (start day is based on the crop calendar, an intermediate stop is required and the stop day is read from a table in the database or is determined by applying a temperature rule (see below)).
- **intermediate start** (the start day is read from a table in the database or is determined by applying a temperature rule (see below), no intermediate stop).

The temperature rule is defined as follows. In case of an intermediate stop (in combination with the temperature rule) the simulation will stop when during three consecutive days the average daily temperature is below a certain threshold. The crop parameter TBASEM is used as threshold. CGMS needs to know the day from which this temperature rule must be applied. A new crop parameter is introduced, called STARTDAY_ISTOP, which indicates the Julian day to start applying the temperature rule. In case of an intermediate start the simulation will begin when three consecutive days the average daily temperature is above a certain threshold. Again TBASEM is used as threshold and the crop parameter STARTDAY_ISTART indicates when to start checking the rule.

A table called **VALUE_STATE_VARIABLES** has been defined which can store values of state variables, for all unique combinations of climatic grid cell, STU and crop. And a table called **STOP_AND_START_DAYS** has been created. These Julian days are entered per climatic grid cell, year and crop.

Total number of values of state variables to be stored per simulation unit (climatic grid cell / stu), per crop and per campaign year amounts to 1124. For large study areas this table will be quite large.

3.9 Information on CGMS run

Some general characteristics of the CGMS run are written into table **CGMS_SYSLOG**. Information on the CGMS level 2 run like crop calendar and how the crop ended the simulation are written in table **SIMULATION_LOG**. The following end status values can occur with regard to WOFOST:

- NSU No simulation units for this grid
- NGW No grid weather for this grid
- MAX Exceeded lifemax (as read from database)
- DEC Exceeded end decade, but not yet lifemax
- BND Number of leaf classes exceeds 365
- DVS DVS >= DVSEND (development stage)
- LAI No living leaves (anymore)
- NEQ Number of simulated rows not equal for water limited and potential
- SOW Invalid sowing date
- EME Invalid emergence date

- IED Invalid end date
- EBE End date before emergence
- WLG Crop failure due to water logging

For LINGRA the following status values can occur:

- MAX Exceeded lifemax
- OK Simulation successful
- ERR Partitioning error

In addition a log file is written to the local file system using the windows environment variable called TMP.

4 Additional functionality in the CGMS software for yield forecast

In 2003 some additional functionality has been introduced in the CGMS software which is responsible for the preparation of data for the yield forecast procedure. It covers issues like aggregation of other indicators, determination of special indicators, import of indicators etc. Note that since 2003 new developments took place resulting in software tools (data engine of CGMS viewer, the CGMS Statistical tool, CGMS data management console) that will probably replace the functions written in this chapter.

More information on the thematic background of the yield forecast procedure can be found on the internet: (http://agrifish.jrc.it/marsstat/Crop%5FYield%5FForecasting/METAMP/) and in the literature (Genovese, G. and Bettio, M., 2004).

4.1 Spatial aggregation of other indicators besides CGMS level 2

The spatial aggregation procedure of CGMS can be used to aggregate the following indicators besides the crop indicators generated with CGMS level 2:

- FAO Water Index model results stored in table **FAO_WATER_INDEX_MODEL** can be aggregated to climatic grid cell level and to administrative regions (the tables **FAO_WATER_INDEX_MODEL_SMU**, **GRID_FAO_WATER_INDEX_MODEL** and **NUTS_FAO_WATER_INDEX_MODEL** are introduced to store spatially aggregated indicators)
- NDVI stored in table **NDVI** can be aggregated to climatic grid cell level and to administrative regions (the tables **GRID_NDVI** and **NUTS_NDVI** are introduced to store spatially aggregated indicators). Note that also a table **CNDVI** exits that store aggregated NDVI data of administrative regions but in this aggregated by an external party.
- Ground weather stored in table **GRID_WEATHER** can be aggregated to administrative regions (the table **NUTS_WEATHER** is introduced to store spatially aggregated observed station weather)
- Model weather – daily stored in table **GRID_WEATHER_MODEL_1** can be aggregated to administrative regions (the table **NUTS_WEATHER_MODEL_1** is introduced to store spatially aggregated daily model data)
- Model weather – 10-daily stored in table **GRID_WEATHER_MODEL_1** can be aggregated to administrative regions (the table **NUTS_WEATHER_MODEL_10** is introduced to store spatially aggregated 10-daily model data)

Different indicators can be aggregated to climatic grid level or administrative region level. Because ground and model weather are already entered into the database at climatic grid level aggregating to administrative region is the only logical option for these indicators.

Monthly model weather data has been excluded from the aggregation. In the yield forecast procedure, which needs aggregated indicators, monthly model weather data is not required.

Except for “ground weather” and “model weather – daily”, all indicators available in the aggregation procedure are given per decade. In the spatial aggregation to climatic grid cell and administrative regions this temporal resolution is maintained. The “ground weather” and “model weather – daily” indicators are aggregated from daily values to decade values during the spatial aggregation to administrative regions. Therefore the intermediate tables **GRID_WEATHER_10** and **GRID_WEATHER_MODEL_1_10** are introduced. For mapping calendar days to decades the table **DAY_DECADE** is used.

4.2 Data preparation for yield forecast

In CGMS different statistical analysis can be selected. This selection determines the data needed for this analysis and the statistical parameters. The user can select and compose crop and/or weather indicators to be used in the yield forecast functions. The data offered to the user are data from tables that store indicators at administrative region level as described in 4.1. The preparation of data for the yield forecast leads to new composed indicators. The way they are composed (variable, decade number etc) is indicated in the name of the new indicator. The following name convention is used:

- Indicator code
- Start and end period (in case of meteo) or end period (in case of other indicators)
- Absolute or relative
- Temperature threshold used in formula (in case of meteo)
- Formula to calculate average or sum (in case of meteo)
- Weather data source (in case of meteo)

The selected data are written to table **DATA_FOR_YIELD_FORECAST**.

General choices like which type of analysis are written to the table **PARAM_MAIN**.

In the database the table **STAT_CROP** determines the mapping between the simulated crop numbers and the crop numbers as they appear in the yield statistics. The yield statistics per crop and administrative region are given in table **EUROSTAT**.

4.2.1 Linear regression: other indicators

The previous forecasting module uses four fixed indicators. These indicators are potential biomass, water limited biomass, potential storage organ and water limited storage organ. The main statistical routines for the forecasting module are included as a Fortran library (DLL) and were managed by the S-PLUS program (Kuyper, 2001). The restriction to four fixed indicators is set by the current Fortran library implementation. In this CGMS version all indicators can be applied, but the number of indicators remains fixed to four indicators among which the best performing indicator will be selected.

In 2004 the S-PLUS program was replaced by a dedicated Delphi program including the statistical functions: linear regression (Fortran DLL), scenario analysis and user specified equation. This program is suitable for a production line. The Delphi program has a batch mode. For the statistical functions MathPack Professional Edition (http://www.lohninger.com/mathpack.html) is used. Note that since 2007 the CGMS Statistical Tool (CST) exists which has a richer functionality to carry out the yield forecast. However the CST has not been linked to CGMS! Thus the CGMS version described here works with the older version developed in 2004.

The indicators are selected from the table **DATA_FOR_YIELD_FORECAST** the yield statistics from table **EUROSTAT**. Parameter settings like parameters sliding window, minimum observations and window size can be changed through the user interface. Their final values are written to the table **META_DATA_REGR**. The results of the linear regression are written to the tables **FORECASTED_NUTS_YIELD_REGR** (yield forecast) and **PARAM_REGR** (some intermediate results).

The calculations can be logged in a log file by setting the parameter “show_debug” to value “1” (no logging is value “0”). Intermediate reports and graphs can be saved by setting the parameter “store_results” to value “1” (no reports/graphs is value “0”). These reports and graphs are useful when the user would like to study the results in more detail. The results are written to two directories, “logs” *and “graphs”* under the user specified workspace. The filename is composed by using the type of analysis (REGR = linear regression; USER = user specified equation; SCEN = scenario analysis), a timestamp, code for the administrative region and the (statistical) crop number.

4.2.2 Scenario analysis

Scenario analysis can be used to define year-analogues for simulated, meteorological or remote sensing indicators. This scenario analysis requires multivariate data analysis like Principal Component Analysis (PCA) and Cluster Analysis (CA). The implementation is based on a command file as applied by JRC to perform this analysis in SPSS. After the procedure has defined the most similar years compared to the current year a yield range is calculated. Official yields of the similar years are corrected for the trend and projected on the most recent year in the defined period.

The indicators are selected from the table **DATA_FOR_YIELD_FORECAST** the yield statistics from table **EUROSTAT**. Parameter settings can be changed through the user interface. Their final values are written to the table **META_DATA_SCEN**. The results of the scenario analysis are written to the tables **FORECASTED_NUTS_YIELD_SCEN** (yield forecast range) **PARAM_SCEN** (used parameter settings) and **PARAM_SCEN_SIMYEARS** (selected similar years).

Like the linear regression it is possible to store information on the calculations and reports/graphs.

4.2.3 User specified equation

Different forecast methods can be incorporated in the system by providing the option to apply user-specified equations. In this case, the user specifies a regression equation between yield and separate indicators. The equation has the form:

$y = a*X_1\text{^}d+b*X_2+c$

where a, b, c and d are coefficients entered by the user through the CGMS user interface and, X_1 and X_2 are variables linked to user defined indicators like crop indicators, FAO, NDVI, meteo etc. The description of the equation is stored in table **META_DATA_USER_SPEC_EQUATION**. The values related to X_1 and X_2 are stored in table **DATA_FOR_YIELD_FORECAST**.

The indicators are selected from the table **DATA_FOR_YIELD_FORECAST** the yield statistics from table **EUROSTAT**. The equation can be specified through the user interface. The final equation is written to the table **META_DATA_USER_SPEC_EQUATION** (variables x_1 and x_2 and coefficients a, b, c and d). The results of the user specified equation are written to the tables **FORECASTED_NUTS_YIELD_USER** (yield forecast) and **PARAM_USER** (used equation).

4.3 Manage external data

Besides indicators of CGMS level1 and 2 other indicators can be added. This can be data of the FAO Water Index Model and remote sensing data at different spatial levels. Results of the FAO Water Index Model should be given for combinations of climatic grid cell, STU, year and decade (table **FAO_WATER_INDEX_MODEL**). CGMS offers functions to aggregate these data to climatic grid cell or administrative regions. NDVI data should be entered for combinations of grid cell, SMU, year and decade (table **NDVI**). Likewise, these data can be aggregated to climatic grid cells or administrative regions. Besides, CNDVI data can be stored per combination of administrative region, year and decade (table **CNDVI**).

CGMS has options to import data from separate TXT (CSV) files with a predefined structure. These data refer to:

- Weather model data.
- Remote sensing data (NDVI and CNDVI).
- Output of the FAO Water Index Model.
- Initial hydrology conditions.
- “stop” and “start” days to stop and restart simulation.
- Start day and start type (table **CROP_CALENDAR**).

Further, CGMS has options to export data to separate TXT (CSV) files. These data refer to:

- Selection from the **CROP_YIELD** tables. For combinations of crop, year and decade all records are written to the specified file in comma separated value format.
- Selection from the **FORECASTED_NUTS_YIELD_REGR** table. For combinations of crop and year all records are selected of the latest run code and are written to the specified file in comma separated value format.

5 Overview of different CGMS versions

The first version of the Crop Growth Monitoring System (CGMS) was developed by the DLO Winand Staring Centre (SC-DLO) in the Netherlands on behalf of the Institute of Remote Sensing Applications of the Joint Research Centre of the European Commission at the beginning of nineties. This version was completely developed in FORTRAN, and used ARC/INFO as a main tool for database preparation, and for visualization of the results (Hooijer et al., 1993). Then, due to increasing size and complexity of the system it became necessary to effectuate a major redesign of the CGMS software by introducing ORACLE for database functions. The redesign took place in 1993-1994, which was realized into creation of the version of 2.0 of the CGMS. FORTRAN was still used in this version for interpolation of grid-weather, simulation of crop yield and calculation of the regression parameters. The following versions 3.0 and 3.1 were applications using ARC/INFO, ORACLE and FORTRAN. They were developed on a VAX-VMS computer and then ported to a Sun-Unix work station (Hooijer, van der Wal, 1994; van der Wal, 1994). The Unix CGMS version 3.1 was the first which was used in operational mode at JRC. It was transformed consequently into versions 4.1 and 5.1 because of thematic improvement. The Unix version 5.1 at the end of the nineties had been redesigned and rewritten in C++ by E. van der Goot and G. Cifani at JRC into Windows version 1.5a.. Moreover, the windows CGMS version 1.5a was used during last years as a basis for further thematic improvement which leaded to creation of the versions 2.0a, 2.0b, 2.1, 2.2 and 2.3. During 2003 new functionality was introduced:

- Storage of alternative grid weather
- Simulate groundwater influence
- Restart of crop growth
- Aggregation of crop indicators, weather variables etc
- Data preparation for the yield forecast module

Because of confusion with the version numbering of the Unix CGMS it was decided to raise the version number to 8.x. Some small changes lead to CGMS version 8.1.2 which is now being used in operational mode by JRC

Below a complete overview is given:

Unix (ProFortran, ASCII/ORACLE):

- CGMS 1.1 (Fortran, ArcInfo, ASCII files).
- CGMS 3.1 ((PRO*)Fortran, ArcInfo, ORACLE).
- CGMS 4.1 (thematic improvements).
- CGMS 5.1/5.2 (thematic improvements).

Windows (C++, ORACLE, openlink drivers (1.5a, 2.0a, 2.1, 2.2)):

- CGMS 1.5a: covers additional weather, weather availability, interpolation, crop growth simulation and is based on Unix version 5.1.
- CGMS 2.0a: is based on CGMS 1.5a and covers introduction of campaign season, possible initialization of water balance, logging simulation run data.
- CGMS 2.0b: is based on CGMS 2.0a and has a batch mode. This version is created in September 2003 via another project called “Delivery of meteo data on a daily basis” (the same has been done for the program MeteoInsert).
- CGMS 2.1: is based on CGMS 2.0a and covers spatial and temporal variability of initial soil moisture, fixed date to start water balance in stead of x-number of days prior emergence, temporal variability of sowing dates, simulated soil moisture rooted zone and potential root zone written to database. It has been created in 2000 by Alterra.

- CGMS 2.2: is based on CGMS 2.1 and covers a bug free LINGRA module, has an Oracle ODBC connection instead of Open link driver, is tested on XP platform. It has been created in the autumn of 2003 by JRC.
- CGMS 2.3: is based on CGMS 2.2 and covers adaptation for partitioning tables, a batch mode is created, determination of available stations for the current year is based on a daily analysis and the NUTS_CODE string is removed from the ELEMENTARY_MAPPING_UNIT and CROP_YIELD_x tables.
- CGMS 8.0: is based on CGMS 2.3. Because of confusion with the version numbering of the Unix CGMS it was decided to raise the version number. New functionality is introduced:
    - Storage of alternative grid weather
    - Simulate groundwater influence
    - Restart of crop growth
    - Aggregation of crop indicators, weather variables etc
    - Data preparation for the yield forecast module
- CGMS 8.1.2: is based on CGMS version 8.0 but some bugs were solved and the following items were introduced or adapted:
    - Introduced winter kill
    - Weather interpolation for the current year is changed using long term average data from reliable stations in case data for the current day is missing
    - Aggregation of weather data from day to decade: rainfall and evapotranspiration are now aggregated as sum in stead of average
    - Solved error in calculation of radiation in LINGRA: variable was multiplied with 1000000, should be 1000
    - Additional output of the soil water balance, of runoff, loss to sub soil and soil evaporation, is saved to the database
- CGMS 9.2: is based on CGMS 8.1.2 but now CGMS is adapted so that it can be used in CALPLAT. The following changes:
    - Adjustment of CROP_YIELD (extra column crop, date in stead of decade/year, extra output columns)
    - Adjustment of GRID_YIELD and NUTS_YIELD (date in stead of decade/year, extra output columns)
    - Potential and water limited can be run separately
    - Introduced CALPLAT mode (store data at daily level and write additional output)
    - Found bug in interpolation of monthly weather data to daily data

6 CGMS 9.2 database

The database is the main part of the CGMS system and contains all input, output and system definition data. Most of the tables in the database form a permanent part of the database, but some tables are temporal (created and deleted by the executable program).

6.1 Table by table

Notation:

- Underlined indicates these fields form the primary key for searching and sorting.
- NOT NULL indicates this field is obligated.
- NUMBER(8,2) means the value must be a number and will be stored with two digits.
- VARCHAR2(1) means the value is a character string with one character.

AGGREGATION_AREAS -

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
AREA_CULTIVATED (area cultivated for indicated crop)		NUMBER	10^3 ha
SOURCE (source of information)	NOT NULL	VARCHAR2(20)	-
DAY (calendar day when data is inserted)	NOT NULL	DATE	-
PERCENTAGE (percentage of area in region cultivated for indicated crop)		NUMBER(5,2)	-

CALCULATED_WEATHER

WMO_NO (station number)	NOT NULL	NUMBER(5)	-
DAY (calendar day)	NOT NULL	DATE	-
E0 (daily transpiration of water surface)		NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)		NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)		NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)		NUMBER(9)	$KJ.m^{-2}.d^{-1}$

CGMS_SYSLOG

RUNID (identification number of run)	NOT NULL	NUMBER(5)	-
DBUSER (username)		VARCHAR2(20)	-
RUNDATE (run date)		DATE	-
RUNINFO (codes explaining kind of run executed)		VARCHAR2(20)	-
CAMPAIGN_START (start month of campaign (agricultural season))		NUMBER(2)	-
WEATHER_START (start date of weather calculations)		DATE	-
WEATHER_END (end date of weather calculations)		DATE	-
START_YEAR (start year of crop simulation)		NUMBER(4)	-
CAMPAIGN_YEARS (number of campaing years to be simulated)		NUMBER(2)	-
RUNSTATUS (final status of run)		VARCHAR2(20)	-
START_MODE (normal, intermediate start or stop)		VARCHAR2(20)	-
WEATHER_DATA_SOURCE (ground or model (daily, decade, month))		VARCHAR2(20)	-
INITIALISATION_WATER_BALANCE (way of initialisation)		NUMBER(4)	-
GROUND_WATER_BALANCE (simulation of ground water influence)		NUMBER(1)	-

CNDVI

NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-

CNDVI (result)	NOT NULL	NUMBER(8,3)	-

CROP

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
CROP_NAME (name of crop)	NOT NULL	VARCHAR2(40)	-
CROPGROUP_NO (group where crop belongs to)	NOT NULL	NUMBER(2)	-
CROP_MODEL (WOFOST (0) or LINGRA (1))	NOT NULL	NUMBER(1)	-

CROP_CALENDAR

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
VARIETY_NO (variety number)	NOT NULL	NUMBER(4)	-
YEAR (year campaign started)	NOT NULL	NUMBER(4)	-
START_TYPE (the way crop starts)	NOT NULL	VARCHAR2(20)	-
START_MONTHDAY1 (earliest day in month crop starts)	NOT NULL	NUMBER(2)	-
START_MONTHDAY2 (latest day in month crop starts -optional)		NUMBER(2)	-
START_MONTH1 (earliest month crop starts)	NOT NULL	NUMBER(2)	-
START_MONTH2 (latest month crop starts - optional)		NUMBER(2)	-
END_TYPE (the way a crop ends)	NOT NULL	VARCHAR2(20)	-
END_MONTHDAY (day in month crop ends)	NOT NULL	NUMBER(2)	-
END_MONTH (month crop ends)	NOT NULL	NUMBER(2)	-
MAX_DURATION (maximum duration for crop on field)	NOT NULL	NUMBER(3)	d

CROP_GROUP

CROPGROUP_NO (group where crop belongs to)	NOT NULL	NUMBER(2)	-
CROPGROUP_NAME (name of crop group)	NOT NULL	VARCHAR2(40)	-

CROP_PARAMETER_VALUE

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
PARAMETER_CODE (character code of crop parameter)	NOT NULL	VARCHAR2(20)	-
PARAMETER_XVALUE (crop parameter value)	NOT NULL	NUMBER(38,6)	-
PARAMETER_YVALUE (crop parameter value – multiple)		NUMBER(38,6)	-

(in case of single parameters PARAMETER_XVALUE gives the value; in case of multiple parameters PARAMETER_YVALUE
gives the value for a certain development stage or temperature described by PARAMETER_XVALUE)

CROP_YIELD

CROP_NO (crop number)	NOT NULL	NUMBER(4)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
SMU_NO(soil mapping unit number)	NOT NULL	NUMBER(8)	-
DAY (date)	NOT NULL	DATE	-
POTENTIAL_YIELD_BIOMASS (potential dry weight biomass)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
POTENTIAL_YIELD_STORAGE (potential dry weight storage organs)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
WATER_LIM_YIELD_BIOMASS (water limited dry weight biomass)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
WATER_LIM_YIELD_STORAGE (water limited dry weight storage organs)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
POTENTIAL_LEAF_AREA_INDEX (potential leaf area index: leaf area divided by surface area)	NOT NULL	NUMBER(5,3)	$m^2.m^2$
WATER_LIM_LEAF_AREA_INDEX (idem for water limited)	NOT NULL	NUMBER(5,3)	$m^2.m^2$

(water limited leaf area index: leaf area divided by surface area)
DEVELOPMENT_STAGE NOT NULL NUMBER(4) -
(development stage of crop 0-200)
RELATIVE_SOIL_MOISTURE NOT NULL NUMBER(8,4) -
(percentage of (field capacity minus wilting point))
TOTAL_WATER_CONSUMPTION NOT NULL NUMBER(8,4) cm
(sum of water limited transpiration
TOTAL_WATER_REQUIREMENT NOT NULL NUMBER(8,4) cm
(sum of potential transpiration)
FSM NOT NULL NUMBER(3,2) -
(volumetric soil moisture content in rooted zone)
FSMUR NOT NULL NUMBER(3,2) -
(volumetric soil moisture content in not rooted zone)
LEAVES_DIED_BY_COLD NUMBER(8,3) -
(fraction of leaves died by cold wrt total biomass)
RUNOFF NUMBER(8,3) cm
(run off)
SOIL_EVAPORATION NUMBER(8,3) cm
(soil evaporation)
LOSS_TO_SUBSOIL NUMBER(8,3) cm
(amount of water drained to the sub soil and therefore lost for the crop)

NB In case of CALPLAT the table CROP_YIELD has a slightly different definition (see CALPLAT manual)

DATA_FOR_YIELD_FORECAST

NUTS_CODE (character code administrative unit) NOT NULL VARCHAR2(8) -
YEAR (calendar year) NOT NULL NUMBER(4) -
DECADE (decade) NOT NULL NUMBER(2) -
STAT_CROP_NO (crop number) NOT NULL NUMBER(2) -
INDICATOR_CODE (short character string for identification) NOT NULL VARCHAR2(50) -
INDICATOR_VALUE (result of indicator) NOT NULL NUMBER(8,3) -

DAY_DECADE

DAY (day in month) NOT NULL NUMBER(2) -
MONTH (month) NOT NULL NUMBER(2) -
DECADE (decade) NOT NULL NUMBER(2) -

ELEMENTARY_MAPPING_UNIT

GRID_NO (grid number) NOT NULL NUMBER(8) -
SMU_NO(soil mapping unit number) NOT NULL NUMBER(8) -
AREA (area of elementary mapping unit) NOT NULL NUMBER(10) m^2

EMU_PLUS_NUTS

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
SMU_NO(soil mapping unit number)	NOT NULL	NUMBER(8)	-
AREA (area of elementary mapping unit)	NOT NULL	NUMBER(10)	m^2

EMU_PLUS_NUTS_LANDCOVER

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
SMU_NO(soil mapping unit number)	NOT NULL	NUMBER(8)	-
CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
AREA (area of elementary mapping unit)	NOT NULL	NUMBER(10)	m^2

EUROSTAT

STAT_CROP_NO (‘statistic’ crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
AREA_CULTIVATED (acreage cultivated)		NUMBER(10,3)	10^3 ha
OFFICIAL_YIELD (yield (fresh weight))		NUMBER(6,3)	10^3 $kg.ha^{-1}$

FAO_WATER_INDEX_MODEL

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
STU_NO (soil typologic unit number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
WSI (result)	NOT NULL	NUMBER(8,3)	-

FAO_WATER_INDEX_MODEL_SMU

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
SMU_NO (soil typologic unit number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
WSI (result)	NOT NULL	NUMBER(8,3)	-

FORECASTED_NUTS_YIELD_REGR

NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
STAT_CROP_NO (‘statistic’ crop number)	NOT NULL	NUMBER(2)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
RUN_CODE (code indicating date/time of run)	NOT NULL	VARCHAR(25)	-
FORECASTED_YIELD (forecast of crop yield as fresh weight of storage organs)		NUMBER(6,3)	10^3 $kg.ha^{-1}$

FORECASTED_NUTS_YIELD_SCEN

NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
STAT_CROP_NO (‘statistic’ crop number)	NOT NULL	NUMBER(2)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
RUN_CODE (code indicating date/time of run)	NOT NULL	VARCHAR(25)	-
FORECAST_MIN (minimum of possible yield range)	NOT NULL	NUMBER(6,3)	10^3 $kg.ha^{-1}$

FORECAST_MAX (maximum of possible yield range)	NOT NULL	NUMBER(6,3)	10^3 kg.ha^{-1}
COUNT_SIM_YEARS (number of similar years)	NOT NULL	NUMBER(2)	-

FORECASTED_NUTS_YIELD_USER

NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
STAT_CROP_NO (‘statistic’ crop number)	NOT NULL	NUMBER(2)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
RUN_CODE (code indicating date/time of run)	NOT NULL	VARCHAR(25)	-
FORECAST	NOT NULL	NUMBER(6,3)	10^3 kg.ha^{-1}

GRID

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
LATITUDE (latitude of grid cell centre)	NOT NULL	NUMBER(8,5)	dd
LONGITUDE (longitude of grid cell centre)	NOT NULL	NUMBER(8,5)	dd
ALTITUDE (average altitude of grid cell)	NOT NULL	NUMBER(5)	m
CLIMATE_BARRIER_NO (climate barrier)	NOT NULL	NUMBER(2)	-
DISTANCE_TO_COAST (distance to coast from cell centre)	NOT NULL	NUMBER(5,1)	km

GRID_FAO_WATER_INDEX_MODEL

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
WSI (result)	NOT NULL	NUMBER(8,3)	-

GRID_NDVI

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
NDVI (result)	NOT NULL	NUMBER(8,3)	-

GRID_WEATHER (INTERPOLATED DAILY STATION WEATHER)

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
DAY (calendar day)	NOT NULL	DATE	-
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	m.s^{-1}
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	mm.d^{-1}
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	mm.d^{-1}
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	mm.d^{-1}
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	mm.d^{-1}
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	KJ.m^{-2}.d^{-1}
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

GRID_WEATHER_10 (INTERPOLATED DAILY STATION WEATHER – STORED AT DECADE LEVEL)

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-

DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

GRID_WEATHER_MODEL_1 (MODEL WEATHER DATA PER DAY)

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
DAY (calendar day)	NOT NULL	DATE	-
MEAN_TEMPERATURE (mean air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

GRID_WEATHER_MODEL_1_10 (DAILY MODEL WEATHER DATA – STORED AT DECADE LEVEL)

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
MEAN_TEMPERATURE (mean air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

GRID_WEATHER_MODEL_10 (MODEL WEATHER DATA PER DECADE)

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
MEAN_TEMPERATURE (mean air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$

CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

GRID_WEATHER_MODEL_30 (MODEL WEATHER DATA PER MONTH)

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
MONTH (calendar month)	NOT NULL	NUMBER(2)	-
MEAN_TEMPERATURE (mean air temperature)	NOT NULL	NUMBER(3,1)	^0C
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^0C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^0C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

GRID_YIELD

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
DAY (date)	NOT NULL	DATE	-
POTENTIAL_YIELD_BIOMASS (potential dry weight biomass)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
POTENTIAL_YIELD_STORAGE (potential dry weight storage organs)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
WATER_LIM_YIELD_BIOMASS (water limited dry weight biomass)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
WATER_LIM_YIELD_STORAGE (water limited dry weight storage organs)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
POTENTIAL_LEAF_AREA_INDEX (potential leaf area index: leaf area divided by surface area)	NOT NULL	NUMBER(5,2)	$m^2.m^2$
WATER_LIM_LEAF_AREA_INDEX (idem for water limited) (water limited leaf area index: leaf area divided by surface area)	NOT NULL	NUMBER(5,2)	$m^2.m^2$
DEVELOPMENT_STAGE (development stage of crop 0-200)	NOT NULL	NUMBER(4)	-
RELATIVE_SOIL_MOISTURE (percentage of (field capacity minus wilting point))	NOT NULL	NUMBER(8,4)	-
TOTAL_WATER_CONSUMPTION (sum of water limited transpiration	NOT NULL	NUMBER(8,4)	cm
TOTAL_WATER_REQUIREMENT (sum of potential transpiration)	NOT NULL	NUMBER(8,4)	cm
FSM (volumetric soil moisture content in rooted zone)	NOT NULL	NUMBER(3,2)	-
FSMUR (volumetric soil moisture content in not rooted zone)	NOT NULL	NUMBER(3,2)	-
LEAVES_DIED_BY_COLD (fraction of leaves died by cold wrt total biomass)		NUMBER(8,3)	-
RUNOFF (run off)		NUMBER(8,3)	cm
SOIL_EVAPORATION (soil evaporation)		NUMBER(8,3)	cm
LOSS_TO_SUBSOIL (amount of water drained to the sub soil and therefore lost for the crop)		NUMBER(8,3)	cm

NB In case of CALPLAT the table GRID_YIELD has a slightly different definition (see CALPLAT manual)

INITIAL_SOIL_WATER

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
STU_NO (soil typological unit number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
ZTI (depth of initial ground water table)	NOT NULL	NUMBER(4,1)	cm
DD (depth of drainage)		NUMBER(3,1)	cm
WAV (water available in potential rooting depth)		NUMBER(3,1)	cm
GIVEN_STARTDATE_WATBAL (date for which WAV is valid)		DATE	-

META_DATA_REGR

WINDOW_SIZE (window size (default 12))	NOT NULL	NUMBER(2)	-
MIN_OBS (minimum observations (default 8))	NOT NULL	NUMBER(2)	-
SLIDING_WINDOW (sliding window (1) or not (0))	NOT NULL	NUMBER(2)	-

META_DATA_SCEN

MAX_PCOMPCOUNT	NOT NULL	NUMBER(2)	-
(maximum number of principal components to process (default 2))			
MIN_EXPL_VAR	NOT NULL	NUMBER(3,1)	-
(minimum percentage of variance: between 0 and 1) (default 0.9))			
CUTOFF_D1	NOT NULL	NUMBER(3,1)	-
(Cutoff score dissimilarity #1 - save pairs with dissimilarity < CutOffD1 (default 1))			
CUTOFF_D2	NOT NULL	NUMBER(3,1)	-
(Cutoff score dissimilarity #2 - save pairs with dissimilarity < CutOffD2 (default 2))			
MIN_SIM_YEARS	NOT NULL	NUMBER(2)	-
(minimum number of similar years (default 5))			
MIN_OBS	NOT NULL	NUMBER(2)	-
(minimum observations (default 8))			

META_DATA_ USER_SPEC_EQUATION

INDICATOR_CODE_X1	NOT NULL	VARCHAR2(40)	-
INDICATOR_CODE_X2	NOT NULL	VARCHAR2(40)	-
COEFFICIENT_A	NOT NULL	NUMBER(6,3)	-
COEFFICIENT_B	NOT NULL	NUMBER(6,3)	-
COEFFICIENT_C	NOT NULL	NUMBER(6,3)	-
COEFFICIENT_D	NOT NULL	NUMBER(6,3)	-

METDATA

STATION_NUMBER (station number)	NOT NULL	NUMBER(5)	-
DAY (calendar day)	NOT NULL	DATE	-
SUNSHINE (daily sunshine duration) (2)		NUMBER(6,2)	h
WIND_10 (daily mean wind speed at 10 m height)		NUMBER(6,2)	$m.s^{-1}$
VAP_PRES (daily mean vapour pressure)		NUMBER(6,2)	hPa
TEMP_MIN (minimum air temperature) (3,4)		NUMBER(6,2)	^0C
TEMP_MAX (maximum air temperature) (3,4)		NUMBER(6,2)	^0C
RAIN (daily rainfall)		NUMBER(6,2)	$mm.d^{-1}$
CLOUD_DAYTIME_TOTAL (daily mean of total cloud cover) (3)		NUMBER(6,2)	oktas
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm
RAD_MEA (daily global radiation at surface) (1)		NUMBER(11,2)	$KJ.m^{-2}.d^{-1}$

1 this variable returns directly radiation
2 this variable can be used to derive radiation via the angstrom formula
3 these variables can be used to derive radiation via the Supit formula
4 these variables can be used to derive radiation via the Hargreaves formula

NDVI

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
SMU_NO (soil mapping unit number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
NDVI (result)	NOT NULL	NUMBER(8,3)	-

NUTS

NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
NUTS_NAME (name of administrative unit)		VARCHAR2(60)	-
NUTS_LEVEL (hierarchic level of administrative unit)	NOT NULL	NUMBER(1)	-
BELONGS_TO (parent administrative unit that belongs to indicated unit)		VARCHAR2(8)	-

NUTS_FAO_WATER_INDEX_MODEL

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
WSI (result)	NOT NULL	NUMBER(8,3)	-

NUTS_NDVI

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
NDVI (result)	NOT NULL	NUMBER(8,3)	-

NUTS_WEATHER

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

NUTS_WEATHER_MODEL_1

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
MEAN_TEMPERATURE (mean air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C

MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

NUTS_WEATHER_MODEL_10

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
YEAR (calendar year)	NOT NULL	NUMBER(4)	-
DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
MEAN_TEMPERATURE (mean air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MAXIMUM_TEMPERATURE (maximum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)	NOT NULL	NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (mean daily vapour pressure)	NOT NULL	NUMBER(4,2)	hPa
WINDSPEED (mean daily wind speed at 10 m height)	NOT NULL	NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)	NOT NULL	NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)	NOT NULL	NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)	NOT NULL	NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of crop canopy)	NOT NULL	NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)	NOT NULL	NUMBER(6)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

NUTS_YIELD

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
DAY (date)	NOT NULL	DATE	-
POTENTIAL_YIELD_BIOMASS (potential dry weight biomass)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
POTENTIAL_YIELD_STORAGE (potential dry weight storage organs)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
WATER_LIM_YIELD_BIOMASS (water limited dry weight biomass)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
WATER_LIM_YIELD_STORAGE (water limited dry weight storage organs)	NOT NULL	NUMBER(8,3)	$kg.ha^{-1}$
POTENTIAL_LEAF_AREA_INDEX (potential leaf area index: leaf area divided by surface area)	NOT NULL	NUMBER(5,2)	$m^2.m^2$
WATER_LIM_LEAF_AREA_INDEX (idem for water limited) (water limited leaf area index: leaf area divided by surface area)	NOT NULL	NUMBER(5,2)	$m^2.m^2$
DEVELOPMENT_STAGE (development stage of crop 0-200)	NOT NULL	NUMBER(4)	-
RELATIVE_SOIL_MOISTURE (percentage of (field capacity minus wilting point))	NOT NULL	NUMBER(8,4)	-
TOTAL_WATER_CONSUMPTION (sum of water limited transpiration	NOT NULL	NUMBER(8,4)	cm
TOTAL_WATER_REQUIREMENT (sum of potential transpiration)	NOT NULL	NUMBER(8,4)	cm
FSM (volumetric soil moisture content in rooted zone)	NOT NULL	NUMBER(3,2)	-
FSMUR (volumetric soil moisture content in not rooted zone)	NOT NULL	NUMBER(3,2)	-
LEAVES_DIED_BY_COLD (fraction of leaves died by cold wrt total biomass)		NUMBER(8,3)	-
RUNOFF (run off)		NUMBER(8,3)	cm

SOIL_EVAPORATION (soil evaporation)		NUMBER(8,3)	cm
LOSS_TO_SUBSOIL (amount of water drained to the sub soil and therefore lost for the crop)		NUMBER(8,3)	cm

PARAM_MAIN

SNAME (name of variable)	NOT NULL	VARCHAR2(30)	-
SVALUE (value of variable)	NOT NULL	VARCHAR2(50)	-

Contents:

NAME	VALUE
analysis	selected statistical analysis
start_year	first year of period for which data have been prepared
end_year	last year of period for which data have been prepared
store_results	indicator to store (1) or not store (0) intermediate results of statistical analysis
show_debug	indicator to show all executed statements (1) or not (0)

PARAM_REGR

NUTS_CODE (character code administrative unit)	NOT NULL	VARCHAR2(8)	-
STAT_CROP_NO (‘statistic’ crop number)	NOT NULL	NUMBER(2)	-
DECADE (decade)	NOT NULL	NUMBER(2)	-
RUN_CODE (code indicating date/time of run)	NOT NULL	VARCHAR(25)	-
TYPE_OF_DATA_SET (data set over years used to derive parameters (W=window, T=total)	NOT NULL	VARCHAR2(1)	-
START_YEAR (start of sequence where parameters are based on)	NOT NULL	NUMBER(4)	-
END_YEAR (end of sequence where parameters are based on)	NOT NULL	NUMBER(4)	-
NUMBER_OF_OBS_REGRESSION (number of observation (years) for regression; in case of window: window size)	NOT NULL	NUMBER(2)	-
NUMBER_OF_OBS_TOTAL (number of total observations (years) to select the predictor)	NOT NULL	NUMBER(2)	-
MISSING_YEARS_TOTAL (number of missing years from total data set)	NOT NULL	NUMBER(2)	-
MEAN_YEAR (year to be used in forecasting formula)	NOT NULL	NUMBER(6,1)	-
MEAN_OFFICIAL_YIELD (mean official yield)	NOT NULL	NUMBER(8,3)	10^3 kg.ha^{-1}
YEARLY_INCREASE (yearly increase in yield)	NOT NULL	NUMBER(8,3)	10^3 kg.ha^{-1}.yr^{-1}
USED_REGRESSION_COEFFICIENT (indicates which (simulated) variable is to be used for forecasting)	NOT NULL	VARCHAR2(50)	-
COEFF_INDICATOR_X1 (multiplier for delta X1)	NOT NULL	NUMBER(6,3)	-
COEFF_INDICATOR_X2 (multiplier for delta X2)	NOT NULL	NUMBER(6,3)	-
COEFF_INDICATOR_X3 (multiplier for delta X3)	NOT NULL	NUMBER(6,3)	-
COEFF_INDICATOR_X4 (multiplier for delta X4)	NOT NULL	NUMBER(6,3)	-
MEAN_INDICATOR_X1 (mean X1)	NOT NULL	NUMBER(8,3)	10^3 kg.ha^{-1}
MEAN_INDICATOR_X2 (mean X2)	NOT NULL	NUMBER(8,3)	10^3 kg.ha^{-1}
MEAN_INDICATOR_X3 (mean X3)	NOT NULL	NUMBER(8,3)	10^3 kg.ha^{-1}
MEAN_INDICATOR_X4 (mean X4)	NOT NULL	NUMBER(8,3)	10^3 kg.ha^{-1}

CODE_INDICATOR_X1 NOT NULL VARCHAR(50) -
(description indicator X1)
CODE_INDICATOR_X2 NOT NULL VARCHAR(50) -
(description indicator X2)
CODE_INDICATOR_X3 NOT NULL VARCHAR(50) -
(description indicator X3)
CODE_INDICATOR_X4 NOT NULL VARCHAR(50) -
(description indicator X4)
STUDENT_VALUE_P5 NOT NULL NUMBER(6,3) -
(T-value resulting from regression, for model indicator)
RSQ_P0 NOT NULL NUMBER(6,3) -
(R-square of trend only)
RSQ_P5 NOT NULL NUMBER(6,3) -
(R-square of selected regression)
REL_RMSQ_RESIDUAL_ERR_P5 NUMBER(5,1) %
(relative rmsq residual error of selected regression)
REL_RMSQ_JACKKNIFE_ERR NUMBER(5,1) %
(relative rmsq jackknife error of prediction based on all years)
REL_RMSQ_ONE_YEAR_AHEAD_ERR_P5 NUMBER(5,1) %
(relative rmsq one year ahead error of complete prediction rule)
REL_RMSQ_TWO_YEAR_AHEAD_ERR_P5 NUMBER(5,1) -
(relative rmsq two year ahead error of complete prediction rule)
ONE_YEAR_AHEAD_ERR_P0 NUMBER(8,3) -
(one year ahead error of trend only)
ONE_YEAR_AHEAD_ERR_P5 NUMBER(8,3) -
(one year ahead error of the selected regression)
TWO_YEAR_AHEAD_ERR_P0 NUMBER(8,3) -
(two year ahead error of trend only)
TWO_YEAR_AHEAD_ERR_P5 NUMBER(8,3) -
(two year ahead error of selected regression)
STANDARD_DEV_P5 NOT NULL NUMBER(7,4) -
(standard deviation of selected regression)

PARAM_SCEN

RUN_CODE (code indicating date/time of run) NOT NULL VARCHAR(25) -
MAX_PCOMPCOUNT NOT NULL NUMBER(2) -
(maximum number of principal components to process (default 2))
MIN_EXPL_VAR NOT NULL NUMBER(3,1) -
(minimum percentage of variance: between 0 and 1) (default 0.9))
CUTOFF_D1 NOT NULL NUMBER(3,1) -
(Cutoff score dissimilarity #1 - save pairs with dissimilarity < CutOffD1 (default 1))
CUTOFF_D2 NOT NULL NUMBER(3,1) -
(Cutoff score dissimilarity #2 - save pairs with dissimilarity < CutOffD2 (default 2))
MIN_SIM_YEARS NOT NULL NUMBER(2) -
(minimum number of similar years (default 5))
MIN_OBS NOT NULL NUMBER(2) -
(minimum observations (default 8))

PARAM_SCEN_SIMYEARS

NUTS_CODE (character code administrative unit) NOT NULL VARCHAR2(8) -
STAT_CROP_NO (‘statistic’ crop number) NOT NULL NUMBER(2) -
DECADE (decade) NOT NULL NUMBER(2) -
RUN_CODE (code indicating date/time of run) NOT NULL VARCHAR(25) -
SIM_YEAR (similar year) NOT NULL NUMBER(4) -

PARAM_USER

RUN_CODE (code indicating date/time of run) NOT NULL VARCHAR(25) -

INDICATOR_CODE_X1	NOT NULL	VARCHAR2(50)	-
INDICATOR_CODE_X2	NOT NULL	VARCHAR2(50)	-
COEFFICIENT_A	NOT NULL	NUMBER(6,3)	-
COEFFICIENT_B	NOT NULL	NUMBER(6,3)	-
COEFFICIENT_C	NOT NULL	NUMBER(6,3)	-
COEFFICIENT_D	NOT NULL	NUMBER(6,3)	-

PARAMETER_DESCRIPTION

PARAMETER_CODE (character code of crop parameter)	NOT NULL	VARCHAR2(20)	-
MULTIPLE (Y = single value, N = value pairs)	NOT NULL	VARCHAR2(1)	-
PARAMETER_DESCR (description of crop parameter)	NOT NULL	VARCHAR2(80)	-

RAINY_DAYS

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
DECADE (calendar decade)	NOT NULL	NUMBER(2)	-
RAINY_DAYS (long term average rainy days per decade)	NOT NULL	NUMBER(2)	-

REFERENCE_WEATHER

WMO_NO (station number)	NOT NULL	NUMBER(5)	-
DAY (calendar day for leap year)	NOT NULL	DATE	-
MAXIMUM_TEMPERATURE (maximum air temperature)		NUMBER(3,1)	^{0}C
MINIMUM_TEMPERATURE (minimum air temperature)		NUMBER(3,1)	^{0}C
VAPOUR_PRESSURE (daily mean vapour pressure)		NUMBER(4,2)	hPa
WINDSPEED (daily mean wind speed at 10 m height)		NUMBER(5,1)	$m.s^{-1}$
RAINFALL (daily rainfall)		NUMBER(4,1)	$mm.d^{-1}$
E0 (daily transpiration of water surface)		NUMBER	$mm.d^{-1}$
ES0 (daily transpiration of wet bare soil)		NUMBER	$mm.d^{-1}$
ET0 (daily transpiration of closed crop)		NUMBER	$mm.d^{-1}$
CALCULATED_RADIATION (daily radiation at surface)		NUMBER(9)	$KJ.m^{-2}.d^{-1}$
SNOW_DEPTH (daily mean snow depth)		NUMBER(6,2)	cm

ROOTING_DEPTH

CLASS (rooting depth class)		NUMBER(5)	-
MIN_DEPTH (depth to which crop roots extends)		NUMBER(5)	cm

SIMULATION_LOG

YEAR (year)	NOT NULL	NUMBER(4)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
VARIETY_NO (variety number)	NOT NULL	NUMBER(2)	-
SOWING_DATE (simulated sowing date)		DATE	d
EMERGENCE_DATE (simulated emergence date)		DATE	d
ANTHESIS_DATE (simulated anthesis date)		DATE	d
END_DATE (simulated end date)		DATE	d
STATUS (way crop ends for instance development stage)		VARCHAR2(3)	-

SIMULATION_UNIT

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
STU_NO (soil typological unit number)	NOT NULL	NUMBER(8)	-

SITE

IFUNRN (infiltration rate depends on storm size, 0 = no, 1 = yes)		NUMBER(1)	-
NOTINF (non infiltrating fraction, 1 = no infiltration, 0 = infiltration)		NUMBER(2,1)	-
MAX_SURFACE_STORAGE (maximum surface storage)		NUMBER(5,3)	cm

SMU_SUITABILITY

SMU_NO (soil mapping unit number)	NOT NULL	NUMBER(8)	-
CROPGROUP_NO (group where crop belongs to)	NOT NULL	NUMBER(2)	-
PERCENTAGE (perc. of soil mapping unit suitable for crop group)	NOT NULL	NUMBER(3)	-

SOIL_ASSOCIATION_COMPOSITION

SMU_NO (soil mapping unit number)	NOT NULL	NUMBER(8)	-
STU_NO (soil typological unit number)	NOT NULL	NUMBER(8)	-
PERCENTAGE (percentage of soil typological unit in SMU)	NOT NULL	NUMBER(3)	-

SOIL_MAPPING_UNIT

SMU_NO (soil mapping unit number)	NOT NULL	NUMBER(8)	-
COUNTRY_CODE (country code)	NOT NULL	VARCHAR2(3)	-

SOIL_PHYSICAL_GROUP

SOIL_GROUP_NO (soil physical group number)	NOT NULL	NUMBER(3)	-
PARAMETER_CODE (character code of soil phys. parameter)	NOT NULL	VARCHAR2(30)	-
PARAMETER_XVALUE (soil phys. parameter value)	NOT NULL	NUMBER(38,6)	-
PARAMETER_YVALUE (soil phys. parameter value – multiple)		NUMBER(38,6)	-

SOIL_TYPOLOGIC_UNIT

STU_NO (soil typological unit number)	NOT NULL	NUMBER(8)	-
SOIL (character code to characterise the soil)	NOT NULL	VARCHAR2(5)	-
SLOPE1 (slope)	NOT NULL	NUMBER(1)	-
TEXT1 (texture of the soil)	NOT NULL	NUMBER(2)	-
PHASE1 (agricultural limitation)	NOT NULL	NUMBER(2)	-
CALCULATED_ROOTING_DEPTH (rooting depth class)	NOT NULL	NUMBER(5)	-
SOIL_GROUP_NO (soil physical group number)	NOT NULL	NUMBER(3)	-
DRAINAGE (the drainage of the soil)	NOT NULL	VARCHAR2(2)	-
ALKALINITY (alkalinity of the soil)	NOT NULL	VARCHAR2(10)	-
SALINITY (salinity of the soil)	NOT NULL	VARCHAR2(10)	-

STAT_CROP

STAT_CROP_NO (‘statistic’ crop number)	NOT NULL	NUMBER(2)	-
STAT_CROP_NAME (name of crop)	NOT NULL	VARCHAR2(40)	-
SIM_CROP_NO (simulated crop number)	NOT NULL	NUMBER(2)	-
SIM_CROP_NAME (name of simulated crop)	NOT NULL	VARCHAR2(40)	-

STATIONS_PER_GRID

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-

YEAR (year)	NOT NULL	NUMBER(4)	-
RAIN_STATION (station number used for rainfall)	NOT NULL	NUMBER(5)	-
TEMP_STATION_1 (first station used for interpolation temperature)	NOT NULL	NUMBER(5)	-
TEMP_STATION_2 (second station used for interpolation temperature)		NUMBER(5)	-
TEMP_STATION_3 (third station used for interpolation temperature)		NUMBER(5)	-
TEMP_STATION_4 (fourth station used for interpolation temperature)		NUMBER(5)	-
REST_STATION_1 (first station used for interpolation other variables)	NOT NULL	NUMBER(5)	-
REST_STATION_2 (second station used for interpolation other variables)		NUMBER(5)	-
REST_STATION_3 (third station used for interpolation other variables)		NUMBER(5)	-
REST_STATION_4 (fourth station used for interpolation other variables)		NUMBER(5)	-
RAIN_SCORE (similarity score for station used rainfall)		NUMBER(6,2)	km
TEMP_SCORE (similarity score for first station used interpolation temperature)		NUMBER(6,2)	km
REST_SCORE (similarity score for first station used interpolation other variables)		NUMBER(6,2)	km

STATIONS_PER_GRID_CURRENTYEAR

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
DAY (julian day: 1-366)	NOT NULL	NUMBER(3)	-
RAIN_STATION (station number used for rainfall)	NOT NULL	NUMBER(5)	-
TEMP_STATION_1 (first station used for interpolation temperature)	NOT NULL	NUMBER(5)	-
TEMP_STATION_2 (second station used for interpolation temperature)		NUMBER(5)	-
TEMP_STATION_3 (third station used for interpolation temperature)		NUMBER(5)	-
TEMP_STATION_4 (fourth station used for interpolation temperature)		NUMBER(5)	-
REST_STATION_1 (first station used for interpolation other variables)	NOT NULL	NUMBER(5)	-
REST_STATION_2 (second station used for interpolation other variables)		NUMBER(5)	-
REST_STATION_3 (third station used for interpolation other variables)		NUMBER(5)	-
REST_STATION_4 (fourth station used for interpolation other variables)		NUMBER(5)	-
RAIN_SCORE (similarity score for station used rainfall)		NUMBER(6,2)	km
TEMP_SCORE (similarity score for first station used interpolation temperature)		NUMBER(6,2)	km
REST_SCORE (similarity score for first station used interpolation other variables)		NUMBER(6,2)	km

STOP_AND_START_DAYS

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
YEAR (calendar year number)	NOT NULL	NUMBER(4)	-
CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
STOP_DAY (Julian day to stop crop growth)	NOT NULL	NUMBER(4)	-
START_DAY (Julian day to restart crop growth)	NOT NULL	NUMBER(4)	-

SUITABILITY

CROPGROUP_NO (group where crop belongs to)	NOT NULL	NUMBER(2)	-
STU_NO (soil typological unit number)	NOT NULL	NUMBER(8)	-

SUPIT_CONSTANTS

WMO_NO (station number)	NOT NULL	NUMBER(5)	-
ANGSTROM_A (ångström constant A using sun shine duration)	NOT NULL	NUMBER	-
ANGSTROM_B (ångström constant B using sun shine duration)	NOT NULL	NUMBER	-
SUPIT_A (supit constant A using cloud cover and temperature)	NOT NULL	NUMBER	-
SUPIT_B (supit constant B using cloud cover and temperature)	NOT NULL	NUMBER	-
SUPIT_C (supit constant C using cloud cover and temperature)	NOT NULL	NUMBER	$KJ.m^{-2}.d^{-1}$
HARGREAVES_A (hargreaves constant A using temperature)	NOT NULL	NUMBER	-
HARGREAVES_B (hargreaves constant B using temperature)	NOT NULL	NUMBER	$KJ.m^{-2}.d^{-1}$

SUPIT_REFERENCE_STATIONS

STATION_NAME (station number)	NOT NULL	VARCHAR2(32)	-
LATITUDE (latitude)	NOT NULL	NUMBER(7,5)	dd
LONGITUDE (longitude)	NOT NULL	NUMBER(8,5)	dd
ALTITUDE (altitude)	NOT NULL	NUMBER(5)	m
ANGSTROM_A (angstrom constant A using sun shine duration)		NUMBER	-
ANGSTROM_B (angstrom constant B using sun shine duration)		NUMBER	-
SUPIT_A (supit constant A using cloud cover and temperature)		NUMBER	-
SUPIT_B (supit constant B using cloud cover and temperature)		NUMBER	-
SUPIT_C (supit constant C using cloud cover and temperature)		NUMBER	$KJ.m^{-2}.d^{-1}$
HARGREAVES_A (hargreaves constant A using temperature)		NUMBER	-
HARGREAVES_B (hargreaves constant A using temperature)		NUMBER	$KJ.m^{-2}.d^{-1}$

SYSCON

SNAME (name of variable)	NOT NULL	VARCHAR2(30)	-
SVALUE (value of variable)	NOT NULL	VARCHAR2(30)	-

Contents:

NAME	VALUE
IPOL_MAXDIST	maximum distance in km for using stations in interpolation
CURRENT_YEAR	when year to be simulated is equal to this value the available stations to be used in the interpolation are determined on a daily basis. For other years this is done on the basis of the whole year
SIMLOG	POT or WL (this indicates for which run extra info will be written to the simulation log table; pot = potential and wl = water limited)
CALLBACK_DLL	full pathname of the callback dll. If present, callback will be enabled for the functions present in the dll
MAXCOSPH	guard used in 'sphere distance'. dist = ((float) acos(cosph)) * EarthRadius, otherwise 0

VALUE_STATE_VARIABLES

GRID_NO (grid number)	NOT NULL	NUMBER(8)	-
STU_NO (soil typological unit number)	NOT NULL	NUMBER(8)	-
YEAR (year)	NOT NULL	NUMBER(4)	-
CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
PARAMETER_CODE (character code of crop parameter)	NOT NULL	VARCHAR2(20)	-
PARAMETER_XVALUE (crop parameter value)	NOT NULL	NUMBER(38,6)	-
PARAMETER_YVALUE (crop parameter value – multiple)		NUMBER(38,6)	-

VARIETY_PARAMETER_VALUE

CROP_NO (crop number)	NOT NULL	NUMBER(2)	-
VARIETY_NO (variety number)	NOT NULL	NUMBER(4)	-
PARAMETER_CODE (character code of crop parameter)	NOT NULL	VARCHAR2(20)	-
PARAMETER_XVALUE (crop parameter value)	NOT NULL	NUMBER(38,6)	-
PARAMETER_YVALUE (crop parameter value – multiple)		NUMBER(38,6)	-

(in case of single parameters PARAMETER_XVALUE gives the value; in case of multiple parameters PARAMETER_YVALUE gives the value for a certain development stage or temperature described by PARAMETER_XVALUE)

WEATHER_DATA_AVAILABILITY

YEAR (year)	NOT NULL	NUMBER(4)	-
WMO_NO (station number)	NOT NULL	NUMBER(5)	-
AVAILABILITY_TYPE (group of variable(s))	NOT NULL	VARCHAR2(30)	-

WEATHER_STATION

WMO_NO (station number)	NOT NULL	NUMBER(5)	-
WMO_NAME (station name)	NOT NULL	VARCHAR2(32)	-
LATITUDE (latitude)	NOT NULL	NUMBER(7,5)	dd
LONGITUDE (longitude)	NOT NULL	NUMBER(8,5)	dd
Y_COORDINATE (not used in CGMS)	NOT NULL	NUMBER(4)	-
X_COORDINATE (not used in CGMS)	NOT NULL	NUMBER(4)	-
ALTITUDE (altitude)	NOT NULL	NUMBER(5)	m
CLIMATE_BARRIER_NO (climate barrier)	NOT NULL	NUMBER(2)	-
COUNTRY_CODE (country code)	NOT NULL	VARCHAR2(3)	-
DISTANCE_TO_COAST (distance to coast)	NOT NULL	NUMBER(4)	km
PERCENT_COMPLETE (critical-% for including station for interpolation)		NUMBER(3)	%
RELIABLE_STATION	NOT NULL	NUMBER	-

TEMPORARY "SIMULATION" TABLES

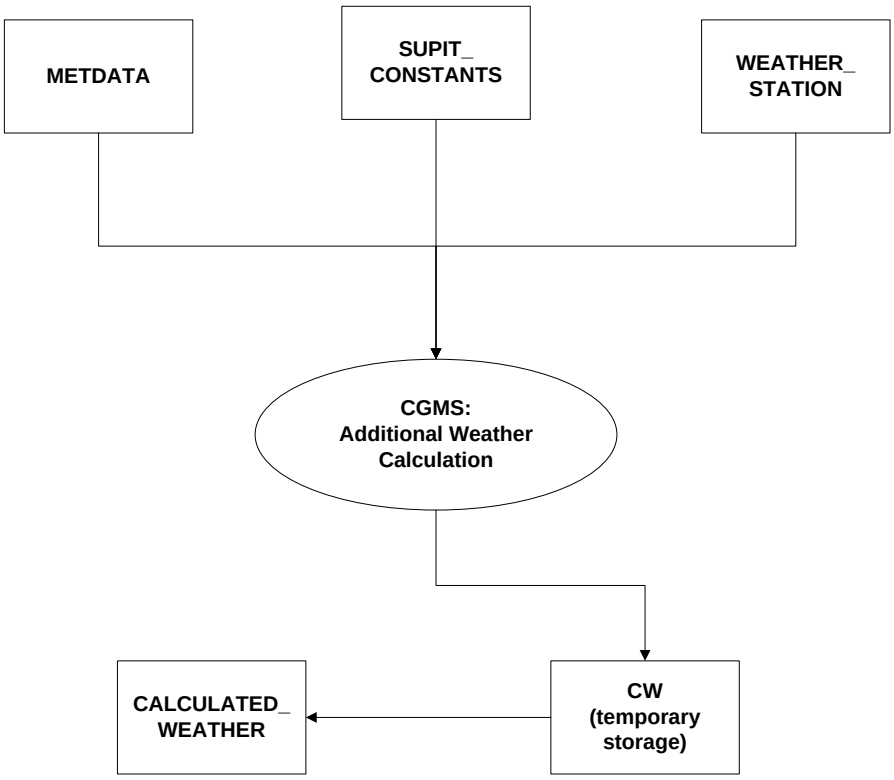
The system creates these tables for each run, and writes the results of simulation to these tables. At the end of the run, the content of the temporary tables is inserted in the main tables (which are indexed). This is done as a single database transaction, and in a single operation. This allows the database to optimise the recreation of the index.

CW: temporary storage for the additional weather calculation
GW: temporary storage for the grid weather calculation
CY: temporary storage for the crop yield calculation

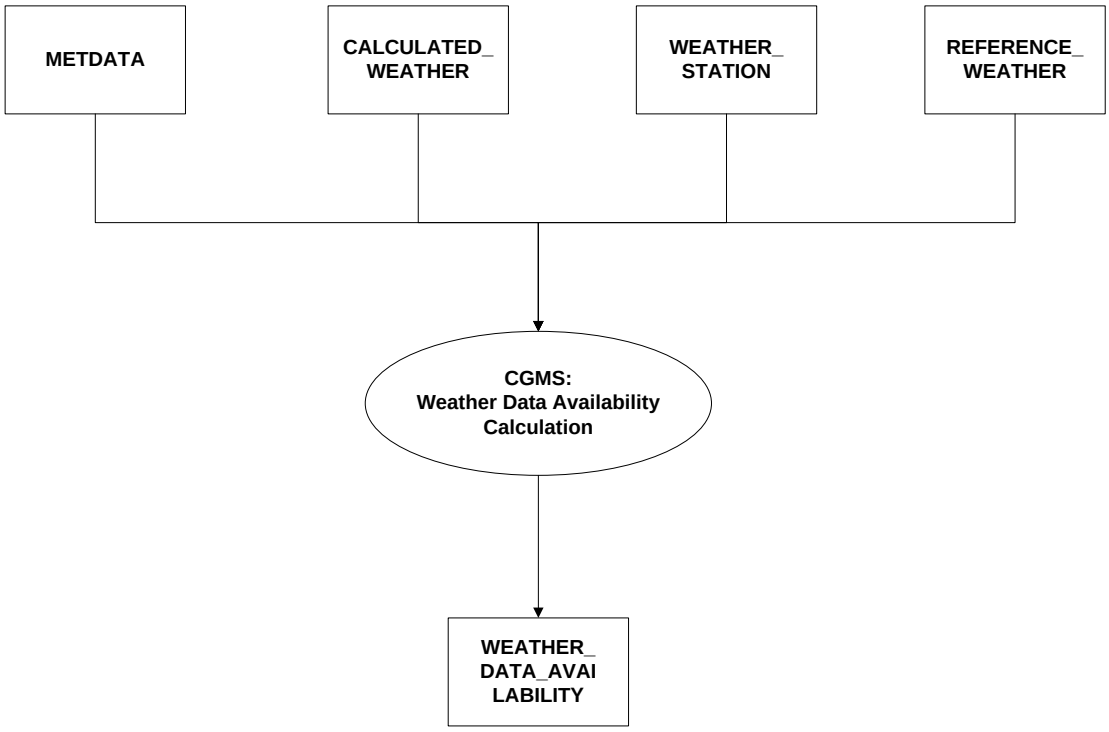
6.2 Relation between tables

The following pages contain:

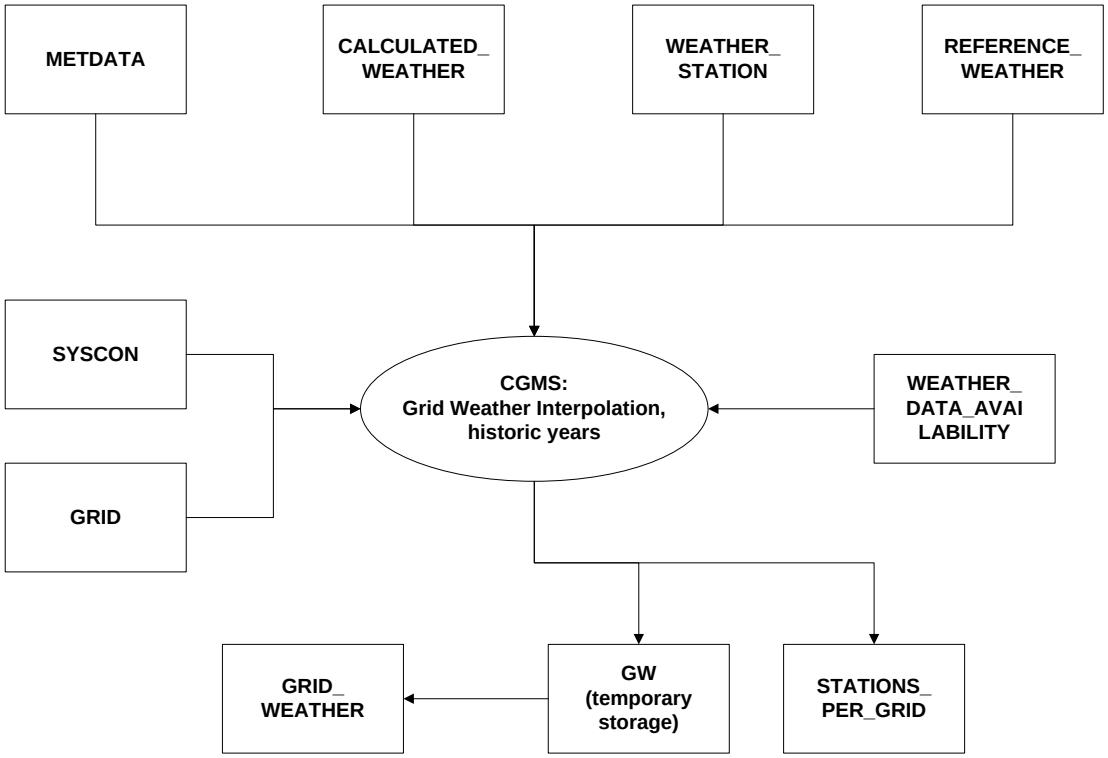
Flow chart 1: All tables used as input and output to calculate global radiation and evapotranspiration at station level
Flow chart 2: All tables used as input and output to determine available weather stations for spatial interpolation (only historic years)
Flow chart 3: All tables used as input and output to spatially interpolate weather from stations to grid cells for historic years
Flow chart 4: All tables used as input and output to spatially interpolate weather from stations to grid cells for the current year
Flow chart 5: All tables used as input and output for the crop growth simulation
Flow chart 6: All tables used as input and output for the aggregation to grid level
Flow chart 7: All tables used as input and output for the aggregation to smallest administrative level
Flow chart 8: All tables used as input and output for the aggregation to administrative level 1 and 0
Flow chart 9: All tables used as input and output for the data preparation for the yield forecast calculation.
Flow chart 10: All tables used as input and output for the statistical analysis: linear regression.
Flow chart 11: All tables used as input and output for the statistical analysis: scenario analysis.
Flow chart 12: All tables used as input and output for the statistical analysis: user specified equation.



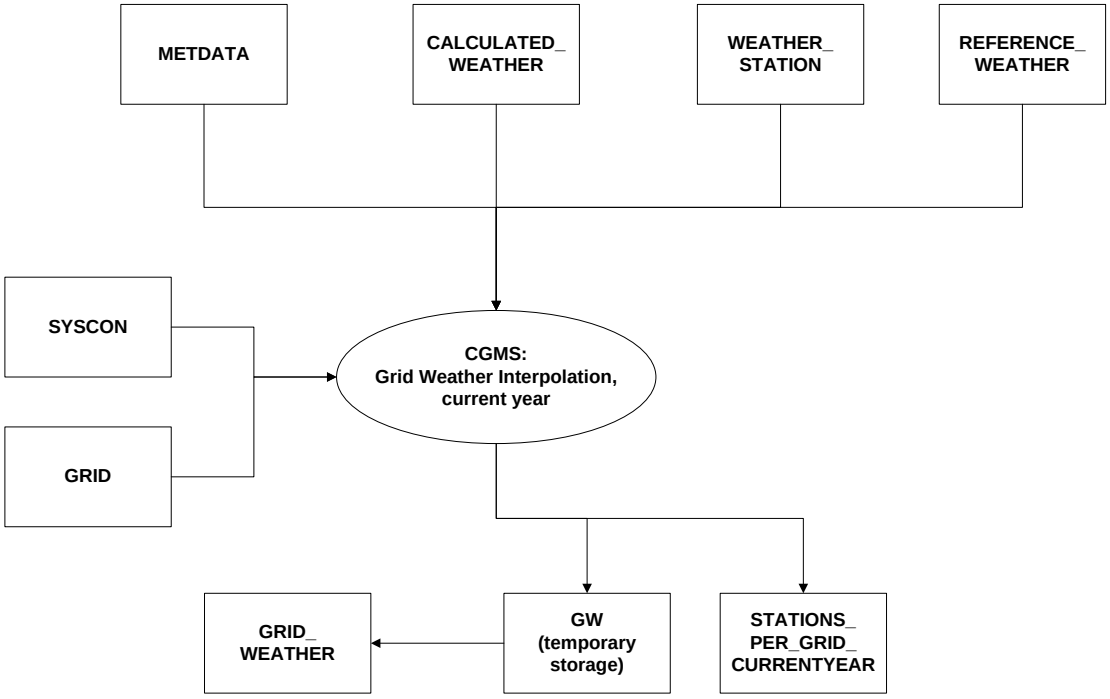
Flow chart 1: All tables used as input and output to calculate global radiation and evapotranspiration at station level



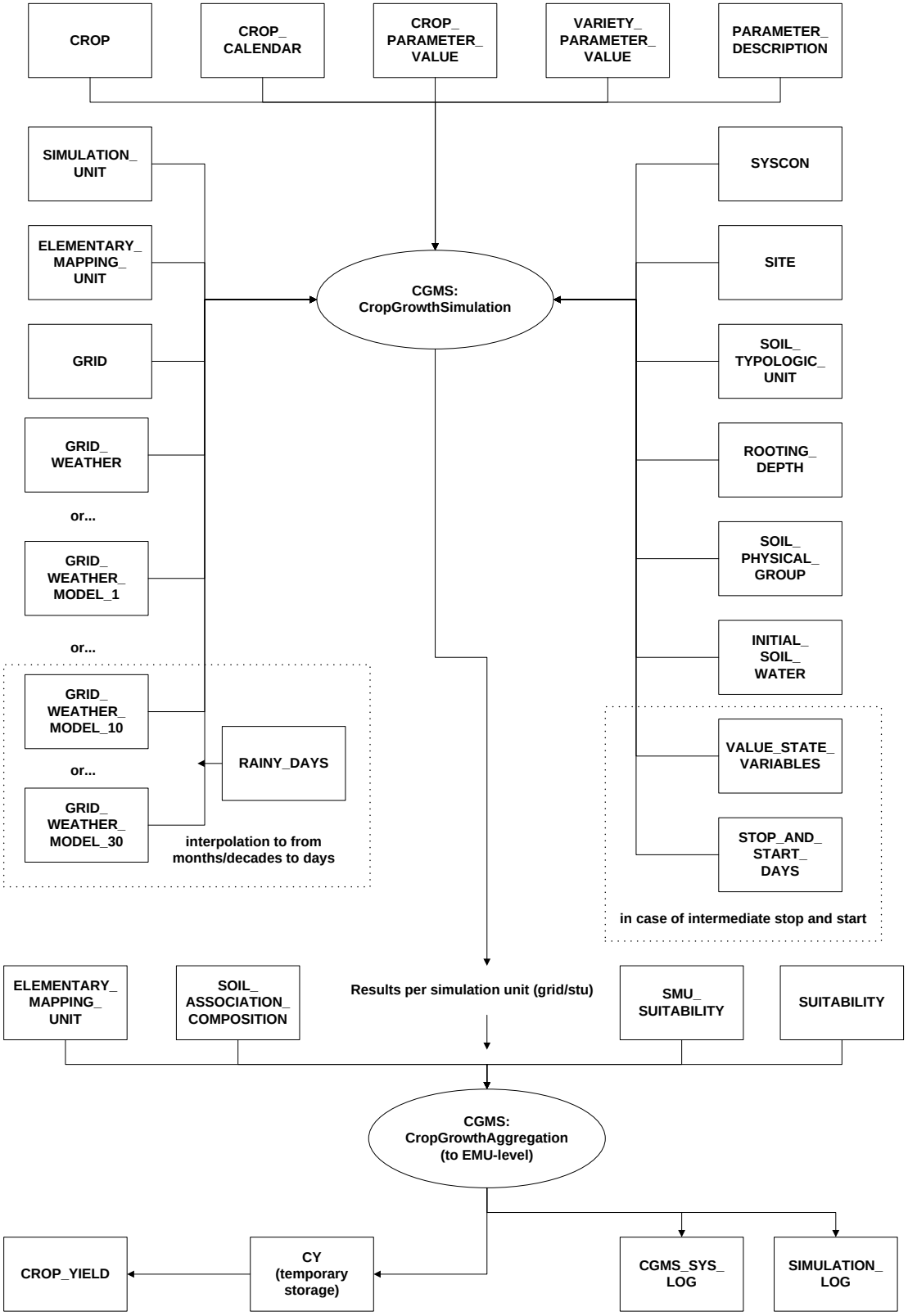
Flow chart 2: All tables used as input and output to determine available weather stations for spatial interpolation. This is only valid for historic years, thus years not equal to value 'CURRENT_YEAR' in table SYSCON



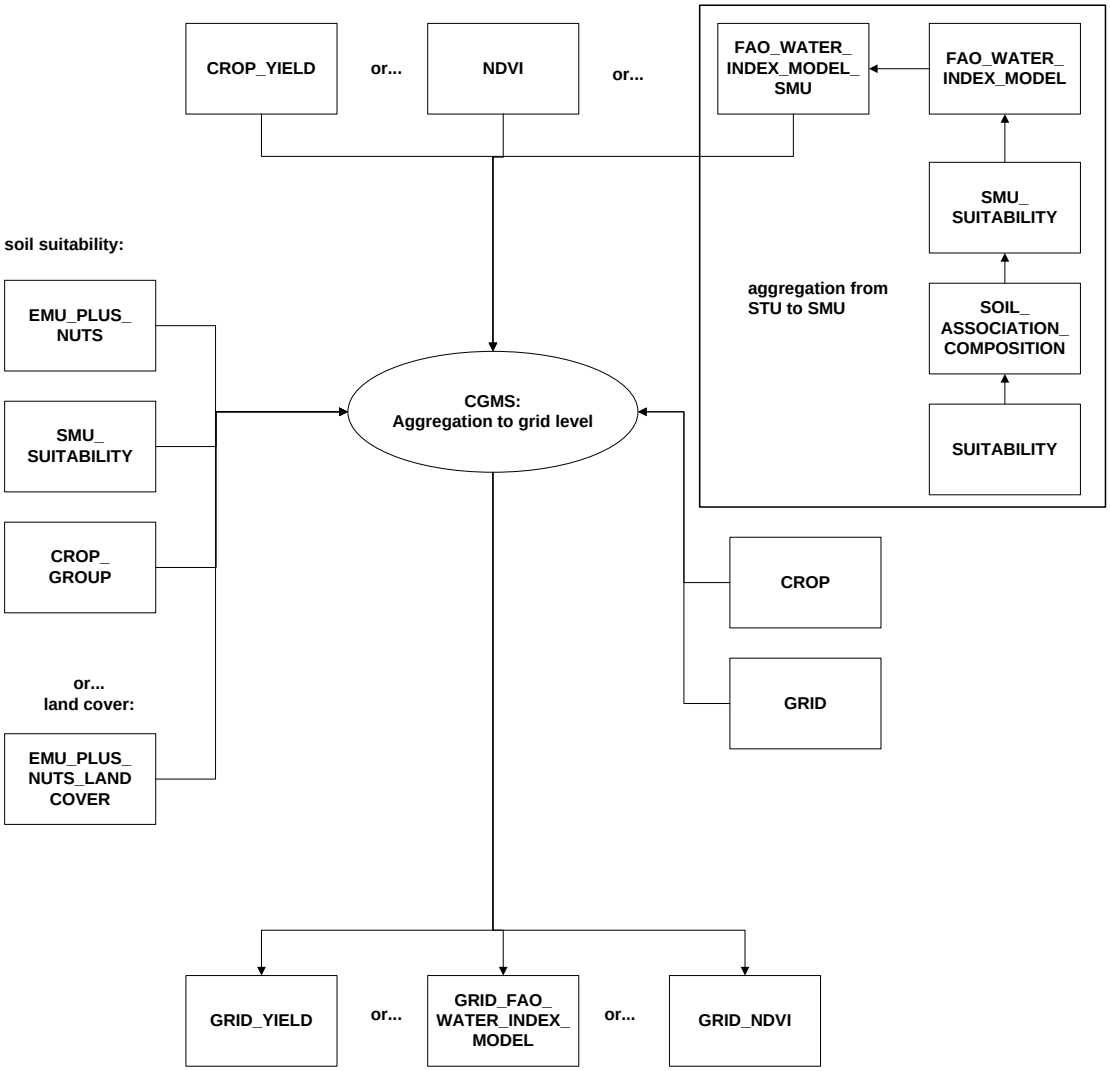
Flow chart 3: All tables used as input and output to spatially interpolate weather from stations to grid cells for historic years, thus years not equal to value 'CURRENT_YEAR' in table SYSCON



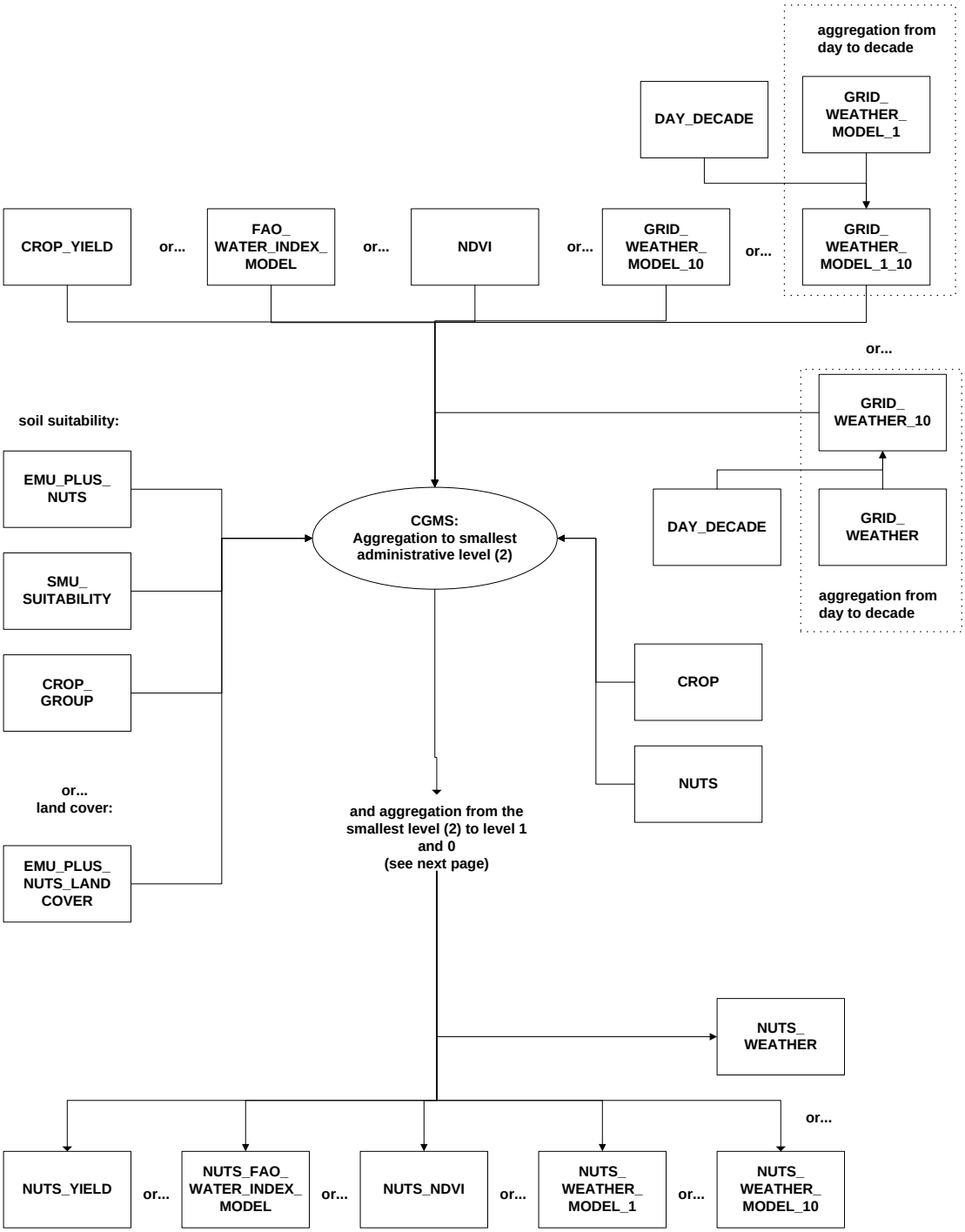
Flow chart 4: All tables used as input and output to spatially interpolate weather from stations to grid cells for the current year, thus the year equal to value 'CURRENT_YEAR' in table SYSCON. The station availability is checked on a daily basis.



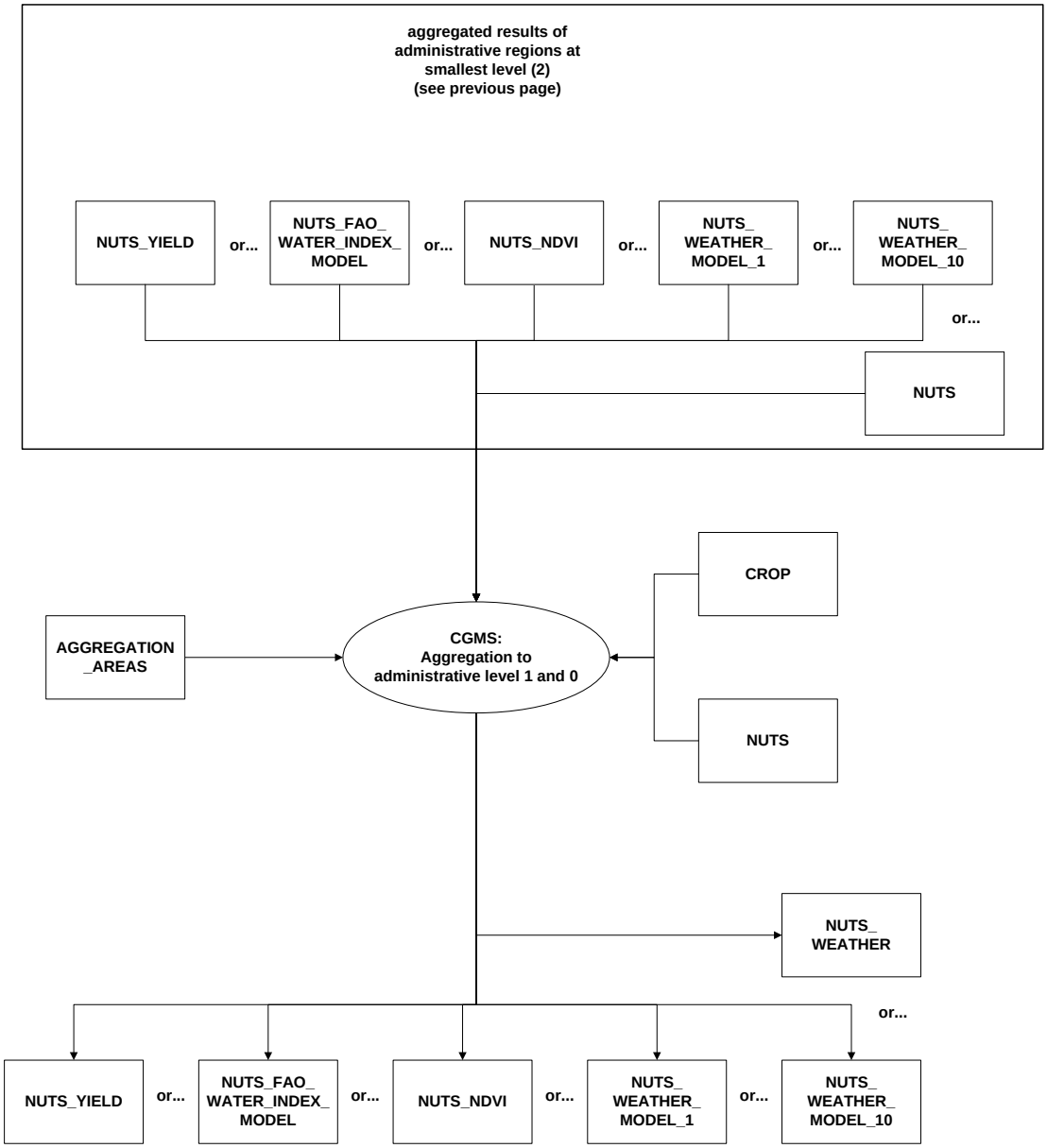
Flow chart 5: All tables used as input and output for the crop growth simulation



Flow chart 6: All tables used as input and output for the aggregation to grid level

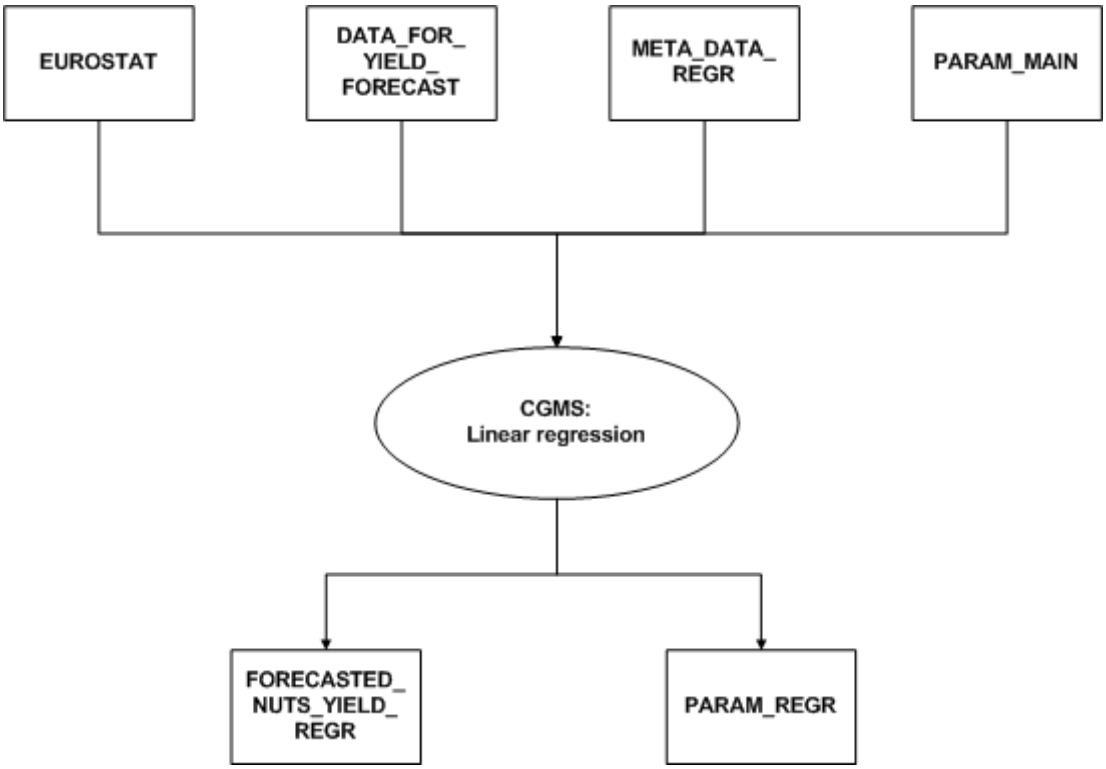


Flow chart 7: All tables used as input and output for the aggregation to smallest administrative level

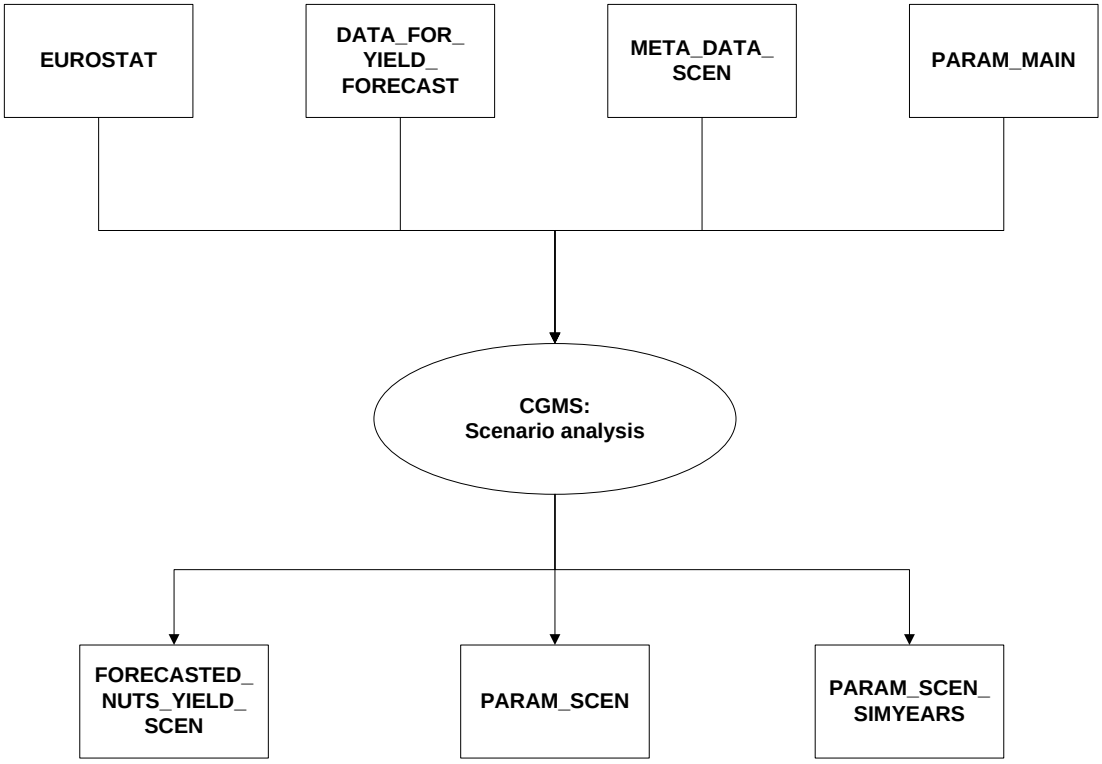


Flow chart 8: All tables used as input and output for the aggregation to administrative level 1 and 0

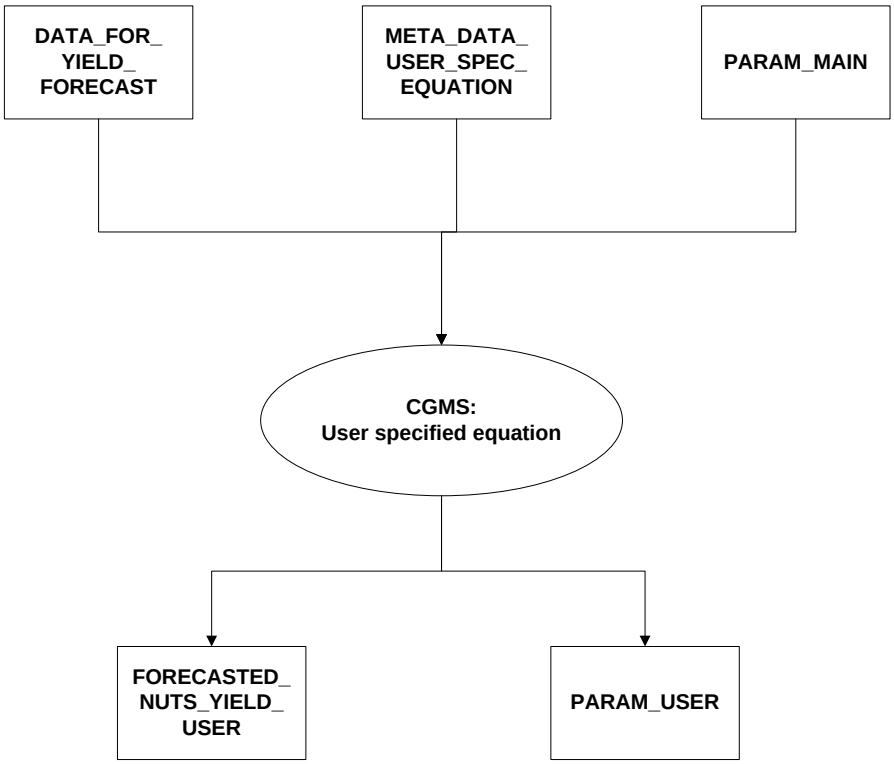
Flow chart 9: All tables used as input and output for the data preparation for the yield forecast calculation



Flow chart 10: All tables used as input and output for the statistical analysis: linear regression



Flow chart 11: All tables used as input and output for the statistical analysis: scenario analysis



Flow chart 12: All tables used as input and output for the statistical analysis: user specified equation

6.3 Database compilation

6.3.1 Data collecting

The first step to the CGMS database compilation is to collect necessary data.

6.3.1.1 Weather

The length of weather time series should be a minimum of 15 years weather data, as a basis for quantifying the annual deviations from long term averages. Technically the minimum length of time series is three years. The following data are required:

- Station number
- Calendar day (dd-mm-yyyy)
- Daily sunshine duration (2) (hours)
- Daily global radiation at surface (1) ($KJ.m^{-2}.d^{-1}$)
- Minimum air temperature (4) degrees Celcius
- Maximum air temperature (4) degrees Celcius
- Daily mean vapor pressure (hPa)
- Daily mean wind speed at 10 m height ($m.s^{-1}$)
- Daily rainfall ($mm.d^{-1}$)
- Daily mean of total cloud cover (3) (oktas)

To derive the global radiation four methods are available, in preferred order (the numbers that occur in the list of data above relate to one of these 4 methods):

1) Measured global radiation
2) Derive global radiation from daily sunshine duration and Angstrom formula
3) Derive global radiation from daily mean of total cloud cover and Supit formula
4) Derive global radiation from temperature and Hargreaves formula

One of the three data sources (measured, sunshine, cloud cover) should be available, otherwise temperature is used. Thus the minimum basic set is temperature, vapor pressure, wind speed and rainfall.

When data are not available, they could be estimated from other data. For example there exist formulas to convert wind speed from one height to another height. Also formulas to derive vapor pressure from dew point temperature, relative humidity or wet bulb temperature (Drift and Van Diepen, 1992).

It is also necessary to collect meta information about the weather stations:

- Latitude (decimal degrees)
- Longitude (decimal degrees)
- Altitude (m)

6.3.1.2 GIS data

The following GIS-data should be available:

- Vector coverage or grid with meteorological grid boundaries and the altitude.
- Vector coverage of soil map.
- Vector coverage with administrative boundaries.
- Not necessary but recommended: a vector coverage of the spatial distribution of the crops.

6.3.1.3 Soil data

The soil map is used to determine the maximum available water which a soil can supply to a plant during a period of drought. Further, if crop masks are missing the soil map can be used as a “land use probability map” to define which crops have to be included in the crop growth simulation for a given soil unit.

6.3.1.4 Crop data

For each crop (and each crop variety) for each agricultural zone for each year (as many years as possible) the following data is needed:

- Sowing date (ddd)
- Emergence date (ddd)
- Leaf area index: maximum value and when reached
- Flowering date (ddd)
- Maturity date (ddd)
- Grain yield ($kg.ha^{-1}.yr^{-1}$)

The minimum set of data is the start of the crop calendar thus sowing or emergence data. But if calibration of the crop parameters is required more data are needed like the list given above.

For the linkage of simulated crop yields to regionally observed crop yields it is necessary to have a time series yield statistics (at least 10 years). It may be helpful to have soil data, irrigation data and fertiliser use data on the regions as well. These data are used for the crop file parameters calibration.

6.3.2 Recommendations for building tables

6.3.2.1 CGMS level 1 – weather monitoring

To work with CGMS you first have to set-up the database. This means you have to fill the basic tables for CGMS level 1 (see also Chapter 2):

- METDATA
- WEATHER_STATION
- SUPIT_REFERENCE_STATIONS
- GRID

Or/and the following tables:

- GRID_WEATHER_MODEL_1
- GRID_WEATHER_MODEL_10
- GRID_WEATHER_MODEL_30

Next, procedures have to be run to fill all other tables that occur in level 1:
- SUPIT_CONSTANTS
- CALCULATED_WEATHER
- WEATHER_DATA_AVAILABILITY
- GRID_WEATHER
- REFERENCE_WEATHER
- STATIONS_PER_GRID
- STATIONS_PER_GRID_CURRENTYEAR

The tables SYSCON and CROP are needed to start up the CGMS executable. All mentioned tables are described in detail before.

After setting up the database, the database should be kept up-to-date. Therefore the METDATA table should be updated at a regularly basis and that leads to the execution of procedures to update all related tables for the level 1. New stations or a different grid need other actions like updating the WEATHER_STATION or GRID and again the execution of procedures to update the whole level 1.

The SYSCON table has to be changed regularly. If you finished previous year you should change the ‘current_year’ field in the table SYSCON to the current year. This is important because of the weather interpolation procedure.

If you update or replace records in the METDATA table for historic years you must execute the actions:
- Import (addition) or update daily station weather (METDATA)
- Calculate additional weather variables
- Calculate long term average daily station weather
- Determine available stations for interpolation
- Interpolation of daily station weather data to grid cells

If you add records in METDATA in the current year you can limit the procedures additional weather, availability and interpolation to grid to the days involved.

In case you delete records from the METDATA table be careful! The best thing to do is to delete the same station/day-combinations from CALCULATED_WEATHER and (if you delete a whole station) the same stations from WEATHER_STATION, SUPIT_CONSTANTS, REFERENCE_WEATHER and WEATHER_DATA_ AVAILABILITY.

Changes in static data like the GRID, SUPIT_REFERENCE_STATIONS, WEATHER_STATION table will have effect on all days in the database. For example, another Supit constant will effects all calculated radiation values for all days. That means the whole period in the database have to be reprocessed.

For updating or adding records in the GRID table, the next procedures have to be done:
- Import (addition) or update grid data (GRID)
- Interpolation of daily station weather data to grid cells (all years, all days)

In case you delete records from the GRID table be careful! The best thing to do is to delete the same grid cells from GRID_WEATHER, STATIONS_PER_GRID, STATIONS_PER_GRID_CURRENTYEAR (and all tables in CGMS level 2 that have GRID_NO as field).

If you would like to update or add records in the SUPIT_REFERENCE_STATIONS table, the next procedures have to be run:
- Import (addition) or update Supit reference constants (SUPIT_REFERENCE_STATIONS)
- Calculate Supit constants
- Calculate additional weather variables (all years, all days)
- Calculate long term average daily station weather
- Determine available stations for interpolation (all years, all days)
- Interpolation of daily station weather data to grid cells (all years, all days)

In this case deleting records from the SUPIT_REFERENCE_STATIONS does not lead to other manual deletes in other tables. You just can follow the same chain procedures as described for the update or addition of records.

If you add records in the WEATHER_STATION table the next procedures have to be run:
- Import (addition) or update station data (WEATHER_STATION)
- Calculate Supit constants

If you update records (location and altitude) in the WEATHER_STATION table you have to run the next procedures:
- Calculate Supit constants
- Calculate additional weather variables (all years, all days)
- Calculate long term average daily station weather
- Determine available stations for interpolation (all years, all days)
- Interpolation of daily station weather data to grid cells (all years, all days)

In case you delete records from the WEATHER_STATION table be careful! The best thing to do is to delete the same station from METDATA, SUPIT_CONSTANTS, CALCULATED_WEATHER, REFERENCE_WEATHER and WEATHER_DATA_AVAILABILITY.

6.3.2.2 CGMS level 2 – crop monitoring

To set-up the database for the level 2, you have to fill the following basic tables (see also Chapter 3):

- CROP (crop related)
- CROP_GROUP (crop related)
- CROP_PARAMETER_VALUE (crop related)
- PARAMETER_DESCRIPTION (crop related)
- VARIETY_PARAMETER_VALUE (crop related)
- CROP_CALENDAR (crop related)
- STOP_AND_START_DAYS (crop related)
- SMU_SUITABILITY (soil related)
- SIMULATION_UNIT (soil related)
- SOIL_TYPOLOGIC_UNIT (soil related)
- SOIL_ASSOCIATION_COMPOSITION (soil related)
- SOIL_PHYSICAL_GROUP (soil related)
- ROOTING_DEPTH (soil related)
- SITE (soil related)
- ELEMENTARY_MAPPING_UNIT (soil related)
- INITIAL_SOIL_WATER (soil related)
- SUITABILITY (soil related)
- RAINY_DAYS (weather related)

Next, procedures have to be run to fill the CROP_YIELD table.

Additional remarks for the mentioned above tables are the following.

A table RAINY_DAYS has been introduced which stores long term average rainy days per decade and climatic grid cell. The rainy days are necessary to calculate daily rainfall by distributing the given monthly or decade rainfall over these rainy days. Daily weather input is requested by the crop growth model.

Crop parameters in CGMS are stored in the tables CROP_PARAMETER_VALUE, VARIETY_PARAMETER_VALUE and PARAMETER_DESCRIPTION. The CROP_PARAMETER_VALUE table contains values for all the parameters as described in the PARAMETER_DESCRIPTION table. These parameters have values used by CGMS to simulate behavior of the crop. To introduce new crops there are two possibilities. When a crop is basically the same as an existing crop (in terms of the WOFOST crop parameters) it is probably convenient to express the differences by introducing only those parameters that are different in the VARIETY_ PARAMETER_VALUE table. If however the crop is significantly different, it is more convenient (and more logical) to introduce a completely new crop in the CROP_PARAMETER_VALUE table. It is important to realize that for the system this is no difference, since theoretically all of the crop parameter values can be overwritten by a variety parameter value.

A crop can have different varieties in a region. Differences in crop parameter values between the basic crop variety and other additional varieties are always very limited and mainly relate

to differences in TSUM1 and TSUM2. Crop parameters of an additional variety that are different compared to the basic crop variety are stored in table VARIETY_PARAMETER_VALUE. If for a certain grid cell / point an additional variety exists than CGMS looks for crop parameters values in VARIETY_PARAMETER_VALUE table, too. If CGMS finds also values in this table it will overrule the value found in the CROP_PARAMETER_VALUE table. The basic crop variety described in CROP_PARAMETER_VALUE has number 1.

Parameters fall into two categories. One category contains the parameters that can be expressed as a single number, e.g. TSUM1, the temperature sum from emergence to anthesis. The other category contains parameters (multiple parameters) that can be expressed as a function of another variable, e.g. SLATB, the specific leaf area as a function of DVS. These parameters are expressed as a set of value pairs (x,y) that describe the shape of the function. A special interpolation function (AFGEN) within CGMS will perform an interpolation to obtain values for arbitrary inputs.

The naming convention for multiple parameters is the name of the parameter, post fixed by '_XX'. The values of XX range from 01 to 10 or 15, depending on the parameter (see the parameter description table). At the moment of writing these limits are fixed.

The rules for interpolation by the AFGEN function are as follows:

- If the argument (x-value) is less than the first x-value in the descriptive array return the first y-value.
- If the argument is between two x-values, return the linear interpolation of the corresponding y-values.
- If the argument is larger than the last x-value, return the last y-value.

The CROP_CALENDAR table describes temporal and spatial distribution of crops and is linked to grid cells / points. When for a specific grid cell / point a crop has to be simulated, the crop calendar contains information about the crop variety to be simulated, and information about sowing day and harvest and simulation strategies. If a crop (variety) occurs at a grid cell / point than you enter a record in the CROP_CALENDAR table, defining the grid cell / point and specific crop number. If you want to simulate the basic crop variety, then enter 1 for the field 'variety_no', otherwise enter an other variety number that corresponds with the variety number described in VARIETY_PARAMETER_VALUE.

For each record you enter you need to define when and how the crop starts growing and when and how the crop ends. The start of the crop simulation is defined by start type, which can have value:

- **variable_sowing**: sowing date is determined by CGMS. CGMS starts evaluating sowing conditions 10 days before the earliest sowing date (given by START_MONTH1 and START_MONTHDAY1) and will return the earliest sowing date within the period between the earliest and the latest sowing date (defined by START_MONTH2 and START_MONTHDAY2). Emergence is calculated using the effective daily temperature (defined by the crop parameters TBASEM, TEFFMX and TSUMEM).
- **fixed_sowing**: the sowing date is given by START_MONTH1 and START_MONTHDAY1. Emergence is calculated using the effective daily temperature (defined by the crop parameters TBASEM, TEFFMX and TSUMEM).
- **fixed_emergence**: emergence takes place on the day given by START_MONTH1 and START_MONTHDAY1.

The end of the simulation is determined by the end type, which can have value:

- **harvest**: simulation stops at the date given by END_MONTHDAY and END_MONTH. If maturity is reached before this date, the model stops at maturity. This option is useful for crops that are harvested in vegetative state, e.g. sugar beet. (MAX_DURATION is not used)
- **maturity**: simulation stops at maturity, but the simulation will not exceed MAX_DURATION days after emergence. The use of MAX_DURATION will prevent anomalies when a crop never reaches maturity due to low temperatures. (END_MONTHDAY and END_MONTH are not used)
- **earliest**: simulation stops at the earliest of maturity, end date or maximum duration.

In the European CGMS database soil mapping units (SMU) are considered to consist of one or more soil types that, at the scale of the soil map in use, are not being mapped as individual cartographic units. These sub-units are called 'soil typologic units'. The soil typologic units (STUs) determine the physical soil parameters, and the soil mapping unit is described by the percentages that the different STUs occupy in the SMU.

The soil physical group parameters are stored in a table called SOIL_PHYSICAL_GROUP. These parameters describe all soil characteristics used in CGMS. In theory you could define as many soil physical groups as STUs. But in reality many STUs have the same soil physical characteristics. Therefore, soil physical groups can be limited to a small set. A small number of soil physical groups will limit the number of simulation units within a grid cell / point and leads to less processing time.

The rooting depth parameter in the SOIL_TYPOLOGIC_UNIT table refers to a so-called rooting depth class. These classes are described by a minimum depth. The depth values for each class are stored in a table called ROOTING_DEPTH, and describe the depth to which the soil allows root growth. It should be noted that the model sensitivity is higher with respect to correct value of rooting depth than towards soil physical group.

The initial soil moisture content, the presence of drainage and the depth of shallow ground water tables are described per climatic grid cell and STU and can be entered in the table INITIAL_SOIL_WATER. Some information closely linked to the soil water balance model is not differentiated with respect to the STUs, but is considered to be 'system wide' and stored in the table SITE. These parameters describe how much water infiltrates into the soil.

The elementary mapping unit (EMU) is formed by the intersection of the soil map and the climatic grid. The elementary mapping unit is fundamental to the system, and any change in any of these two components has major implications for the system. The EMU is not given a unique identifier in the database, but is stored as a (unique) combination of its properties, i.e. soil mapping unit number and grid identifier. The only other information stored for an EMU is its area (table ELEMENTARY _MAPPING_UNIT).

To run CGMS we need to know which soils are suitable to grow cereals, or even better: which soils are currently used to grow cereals. In case no data on land use is available expert rules describing suitability of soils are an alternative. This suitability is not expressed in terms of the actual crops, but in terms of a so-called crop-group. Every crop is labeled as belonging to one of a number of crop groups. The SUITABILITY table contains every valid combination of STU number and crop group number. The table SMU_SUITABILITY is only introduced because of performance reasons.

The user can stop the simulation at a specific day or when a specific condition has been met. In case of an intermediate start the start day is read from the table STOP_AND_START_DAYS in the database or is determined by applying a temperature rule.

The crop growth simulation is carried out for all unique combinations of suitable STUs and climatic grid cells (table SIMULATION_UNITS). The final results produced by the crop growth model are stored in the table CROP_YIELD per climatic grid cell and SMU. To aggregate from STU to SMU the table SOIL_ASSOCIATION_COMPOSITION is used. The name of the CROP_YIELD table is slightly misleading since many results are stored here, not just the crop yield. The results are stored as function of the simulation period (called 'decade'), year and EMU. The results are stored for each simulation period requested by the user. Note that the simulation itself is always carried out from the beginning of the growing season. In case the campaign month in table SYSCON is not equal to '1', the simulation period has a decade numbering which is different compared to calendar decades. For example: if campaign month is '5' than the decade '1' in table CROP_YIELD corresponds to the first decade of May.

In table SIMULATION_LOG table CGMS writes information about the simulations. The table contains data about sowing, emergence, anthesis and end data used and simulated by the model. In the SYSCON table you can decide whether you want data about the potential run or the water limited run.

6.3.2.3 CGMS level 3 – yield forecast

For the yield forecast it is necessary to aggregate crop growth indicators to the administrative units, which are linked with the statistical information about crop yield. Before the aggregation of the indicators and yield forecasting, the following additional tables must be filled:

- AGGREGATION_AREAS
- CNDVI (optional)
- DAY_DECADE
- EMU_PLUS_NUTS
- EMU_PLUS_NUTS_LANDCOVER
- EUROSTAT
- FAO_WATER_INDEX_MODEL (optional)
- META_DATA_REGR
- META_DATA_SCEN
- META_DATA_USER
- NDVI (optional)
- NUTS
- PARAM_MAIN
- STAT_CROP

Some tables from the list are “optional”, which means, that it will be possible to produce yield forecast without filling this tables, and its aggregation. In this way, only simulated crop indicators could be used.

Within the spatial aggregation from EMU to grid or administrative regions the tables DAY_DECADE, AGGREGATION_AREAS, NUTS, EMU_PLUS_NUTS and EMU_PLUS_NUTS_LANDCOVER are necessary. For more information see paragraph 3.6.

The table EUROSTAT is needed to calculate yield forecasts. To specify the parameters settings of the yield forecast functions the tables are META_DATA_REGR, META_DATA_SCEN and META_DATA_USER must be filled. The table STAT_CROP describes the relation between the “simulated” crops and the crops for which yield statistics exist.

7 CGMS User Interface

7.1 Installation CGMS

The installation of the software is quite simple. Run the set-up executable and two files are copied to selected installation directory.

- cgms.exe (CGMS version 9.2 executable)
- cgmsbat.exe (CGMS yield forecast module executable)
- cgmsbulk.dll (A Fortran library for the bulk calculations of CGMS level 3)
- dforrt.dll (this is a Fortran Visual Studio library. The CMGS installation program installs it only in case it does not exist or when it is newer than an existing one. If you do not have Fortran on your PC you may delete this library if you want CGMS removed from your system)

In addition three sub directories are created:

- Database (with a small MS Access database example)
- Logs (used to write log files of yield forecast module)
- Queries (queries used to perform the yield forecast)
- Doc (user manual)
- SourceCode (source code of CGMS version 9.2)

You only need to work with the CGMS version 9.2 executable. To test the new software one MSAccess database has been created. The database describes a simple example. To test and learn CGMS it is adviced to use the MSAccess database called “CGMS_92.mdb”. Before starting CGMS you must first create a new data source for this small database:

- go to control panel/administrative tools/odbc;
- select tab "system dsn";
- select MS Access driver;
- enter a string to define the data source name for instance CGMS_9_2;
- select the right the database (“CGMS_9_2.mdb”).

After the crop growth simulation CGMS writes detailed information to a log file on a temporary directory (directory is defined in your TMP system variable).

CGMS can also be run in batch mode in stead of operated through the user interface (see Appendix A).

7.2 Data base login and execution steps

Two type of databases have been tested:

- MSAccess database (2003)
- ORACLE (9)

A data source name, linked to the database, must be defined before you can start CGMS. In case of ORACLE you need to install ORACLE client and edit the file “tnsnames.ora”. Next, you must define a data source name (DSN) using the normal ORACLE drivers.

In the CGMS graphical user interface (GUI) you can skip the username/password fields if you connect to MS Access database (see Figure 7-1).

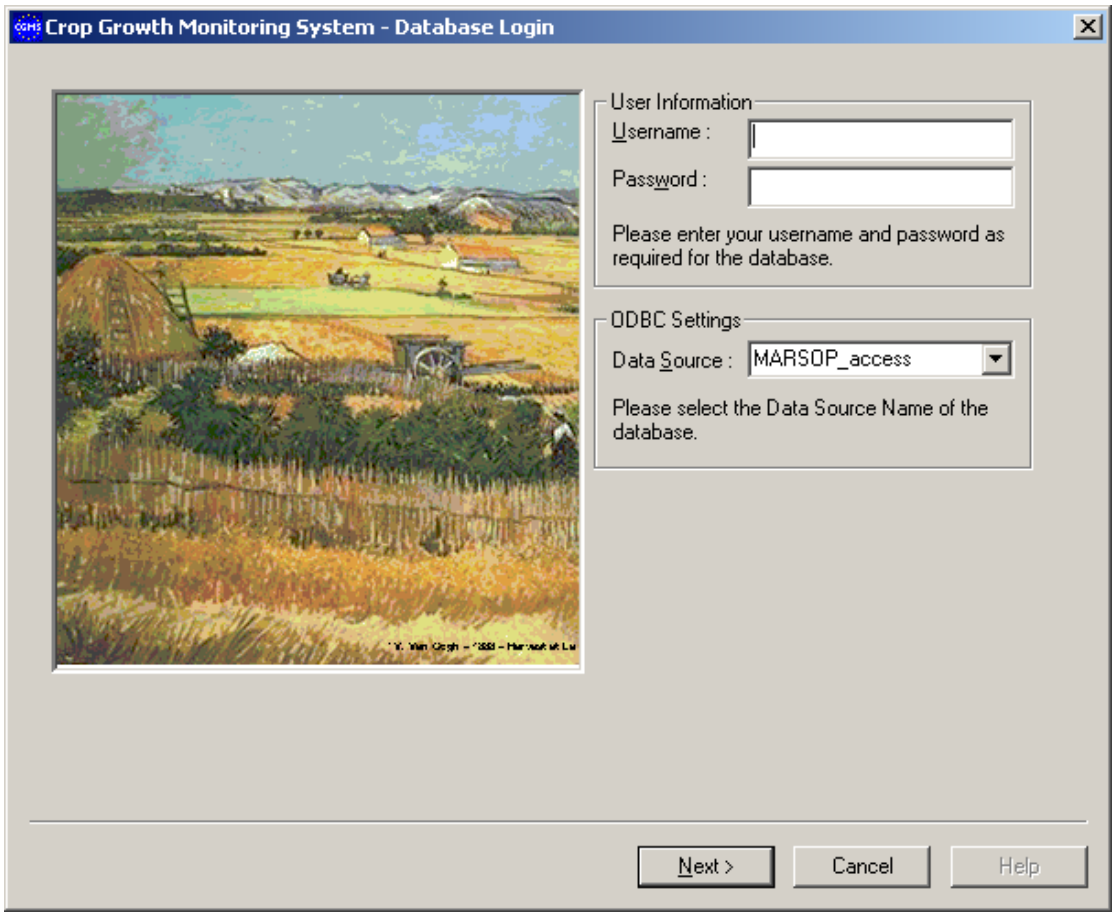


Figure 7-1: Database login screen

When activating the “next”-button the window “execution steps” appears (see Figure 7-2).

Different procedures can be chosen:

- Weather data calculation;
- Crop simulation;
- Aggregation;
- Preparation of data for the yield forecast calculation;
- Import external data to the database;
- Export data from the database.

After clicking on the “next”-button the user confirms the choice and a new window appears.

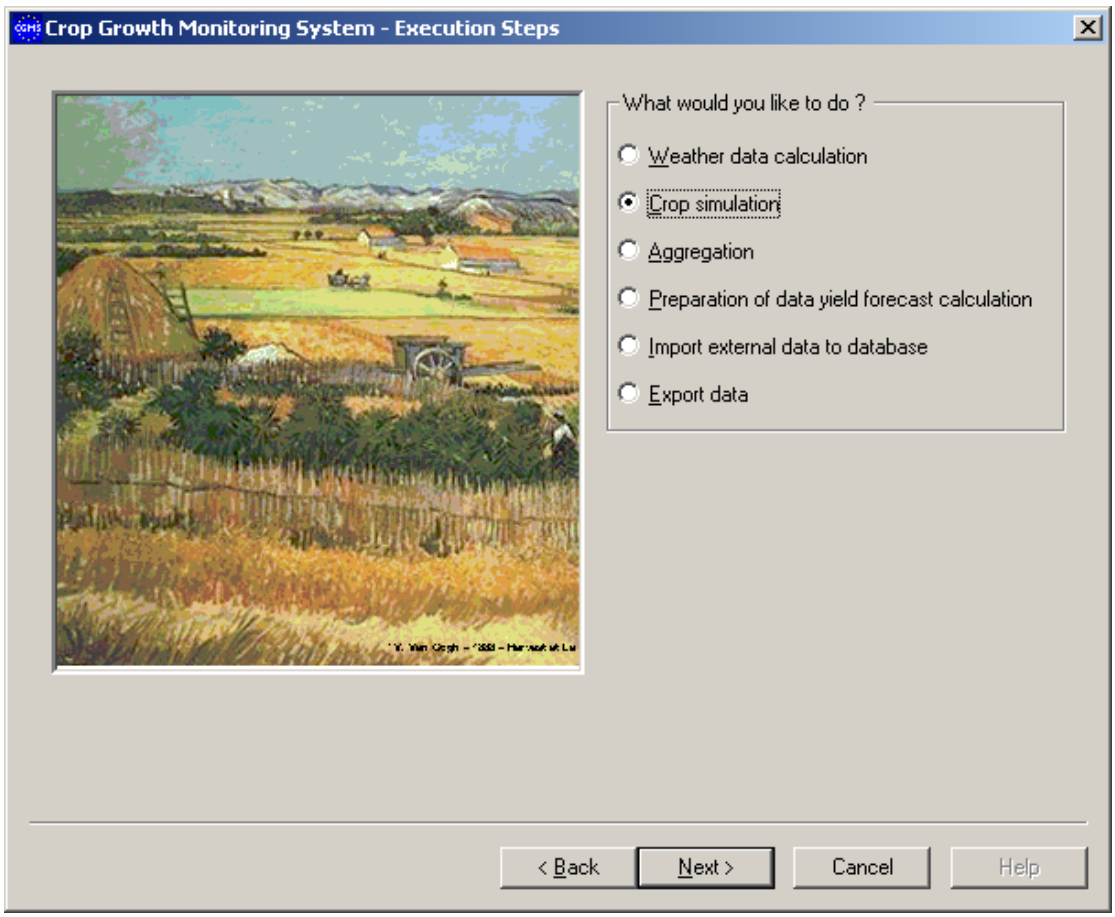


Figure 7-2: Execution steps

7.3 Weather data calculation

When choosing “weather data calculation” in the “execution steps” window, the “weather data calculation window” appears (see Figure 7-3).

First, a period must be defined by entering a start date (day in month, month, year) and an end date (day in month, month, year). For all days within this period, CGMS can:

- Calculate additional weather parameters.
- Determine weather stations available for spatial interpolation.
- Interpolate daily station weather to a regular climatic grid.

In case of “additional weather parameters” you can select two different ways for calculating the global radiation according Supit/Hargreaves or Choisnel.

For more information on the thematic background and the specific database tables see other chapters of this document.

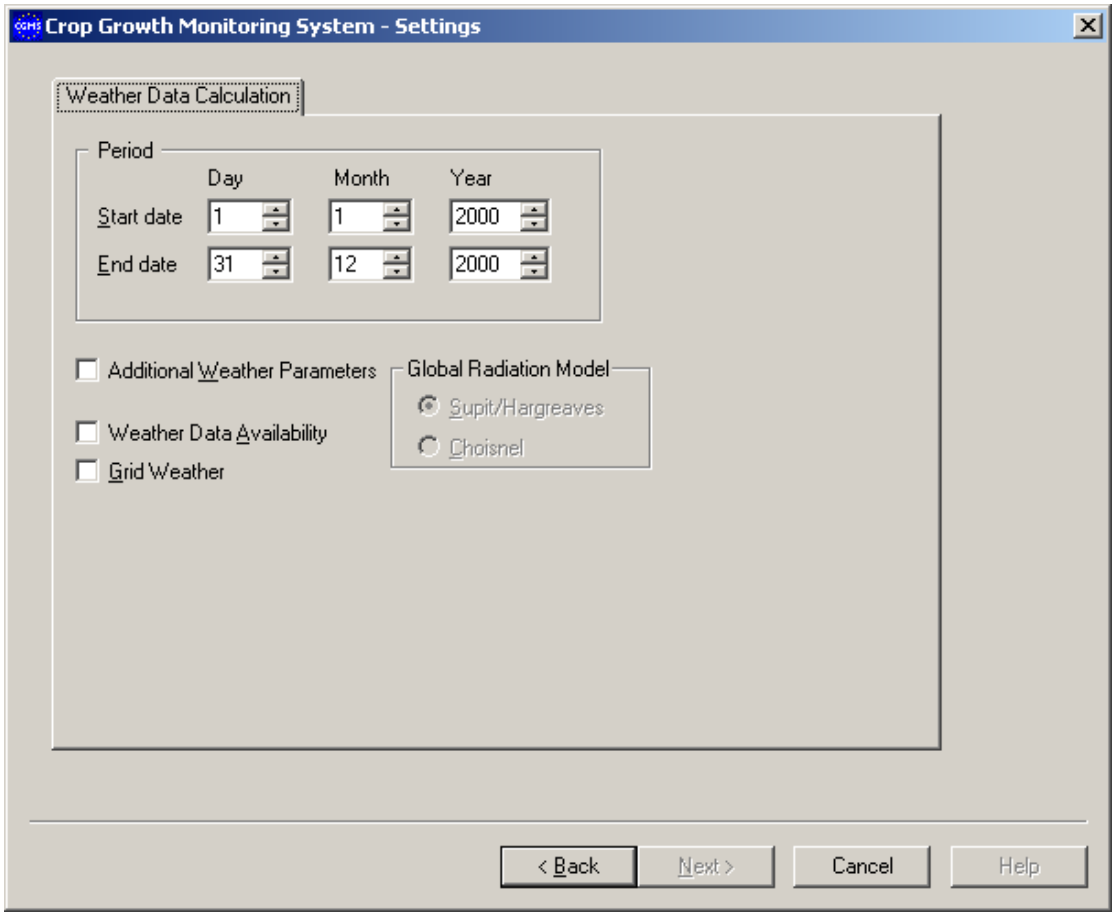


Figure 7-3: Weather data calculation window

After selecting the right options you can press the “next”-button and you will see the summary window. In this window you click on the “finish”-button to effectuate all your choices and start the calculations.

7.4 Crop simulation

When choosing “crop simulation” in the “execution steps” window, the “crop simulation window” appears (see Figure 7-4) containing two tab sheets.

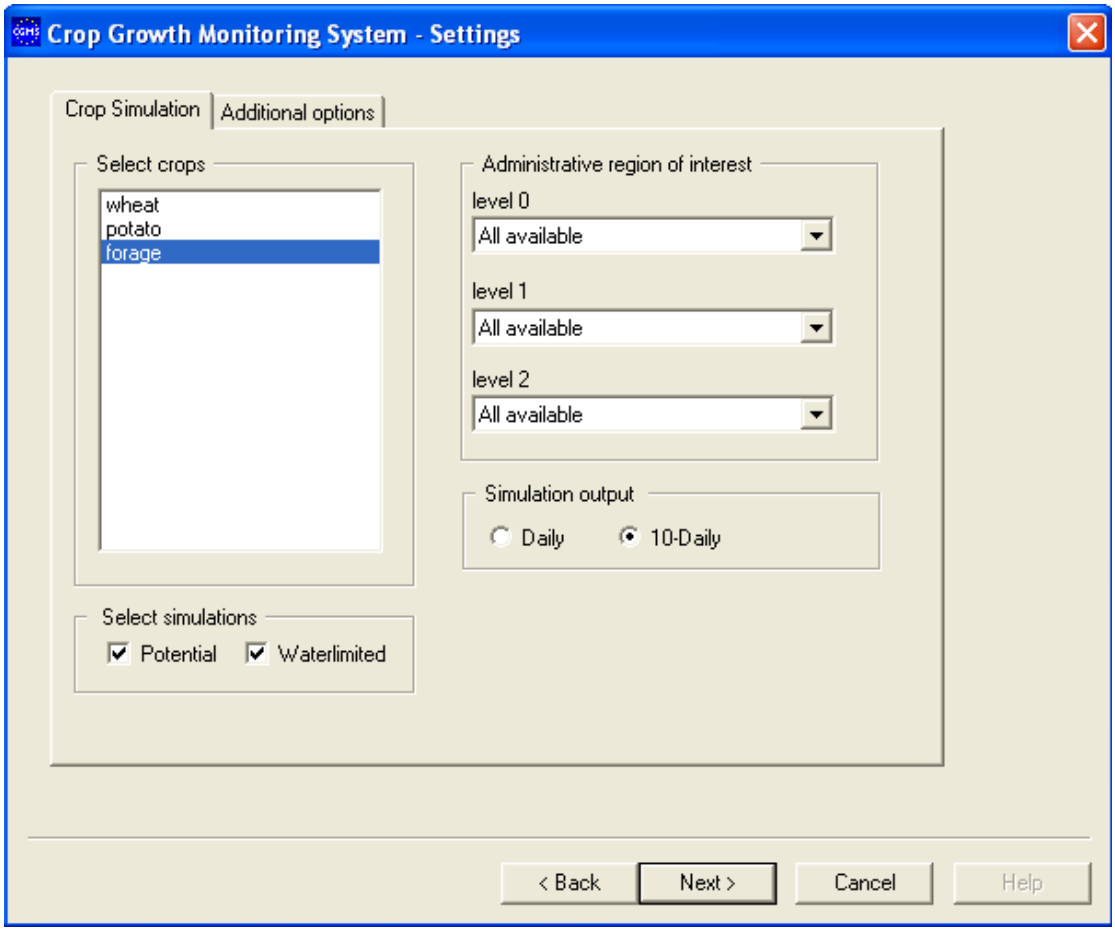


Figure 7-4: Crop simulation window –“crop simulation”-tab sheet

7.4.1 Crop and administrative regions

First, you select one or more crops. Next, the administrative region(s) of interest can be chosen. Default the string “all available” is given indicating all regions will be simulated. The administrative regions are organised as follows: the highest level, for instance the whole country, is called “level 0”, which is divided in smaller regions called “level 1”. These smaller regions are again divided in the smallest regions called “level 2”. Via the pull down menus you can select one region at “level 0” and simulate for this “level 0”-region all “level 2”-regions that belong to this selected region. Or you can continue to limit the “level 2”-regions by selecting only one “level 1”-region within the “level 0”-region. Finally, you can decide to simulate only one “level 2”-region.

Thus the user can choose a limited set of administrative regions in the crop simulation window. Main reason behind this is to limit execution time. If a user chooses a set of regions CGMS will simulate all grid cells that are (partly) covered by these regions. As a

consequence all simulation units within these regions will be simulated and additionally some extra simulations units of grid cells which are only partly covered by these regions.

The ideal solution would have been to simulate only the grid/STU combinations within the selected administrative regions and only aggregate these results to grid / SMU combinations within the administrative regions. But this would have lead to a major change in the concept. In that case data on administrative regions would have been needed during the crop growth simulation to determine simulation units and to control the aggregation from grid cell / STU combinations to grid cell / SMU combinations. This in contradiction to the concept to avoid data on administrative regions in the data processing for crop growth simulation and only use these data in the aggregation from grid cell / SMU combinations to administrative regions.

The solution to translate the selected regions into selected grid cells that are totally or partly within the selected regions has only minor implications:

- Execution time is little bit higher.
- The crop yield table contains not only the grid cell / SMU combinations within the selected regions but also the combinations in other adjacent regions in case the grid cell is partly covered by other regions. The same is true for the nuts yield table which will include results for adjacent regions which are not part of the selected regions.

Finally the user can decide to run only potential or only water limited or both modes together and the required time step to save the results: day or 10-daily (the last day of the dekad).

After you have selected a crop, the “next”-button comes available. Before you click this button you must first check the other simulation options in the “additional options”-tab (see Figure 7-5).

Additional options are:

- Campaign
- Weather data
- Start/end mode
- Initialisation water balance (only relevant for initial rooting depth)
- Simulate with groundwater influence
- Simulate winterkill
- Produce output for user defined period in the current year

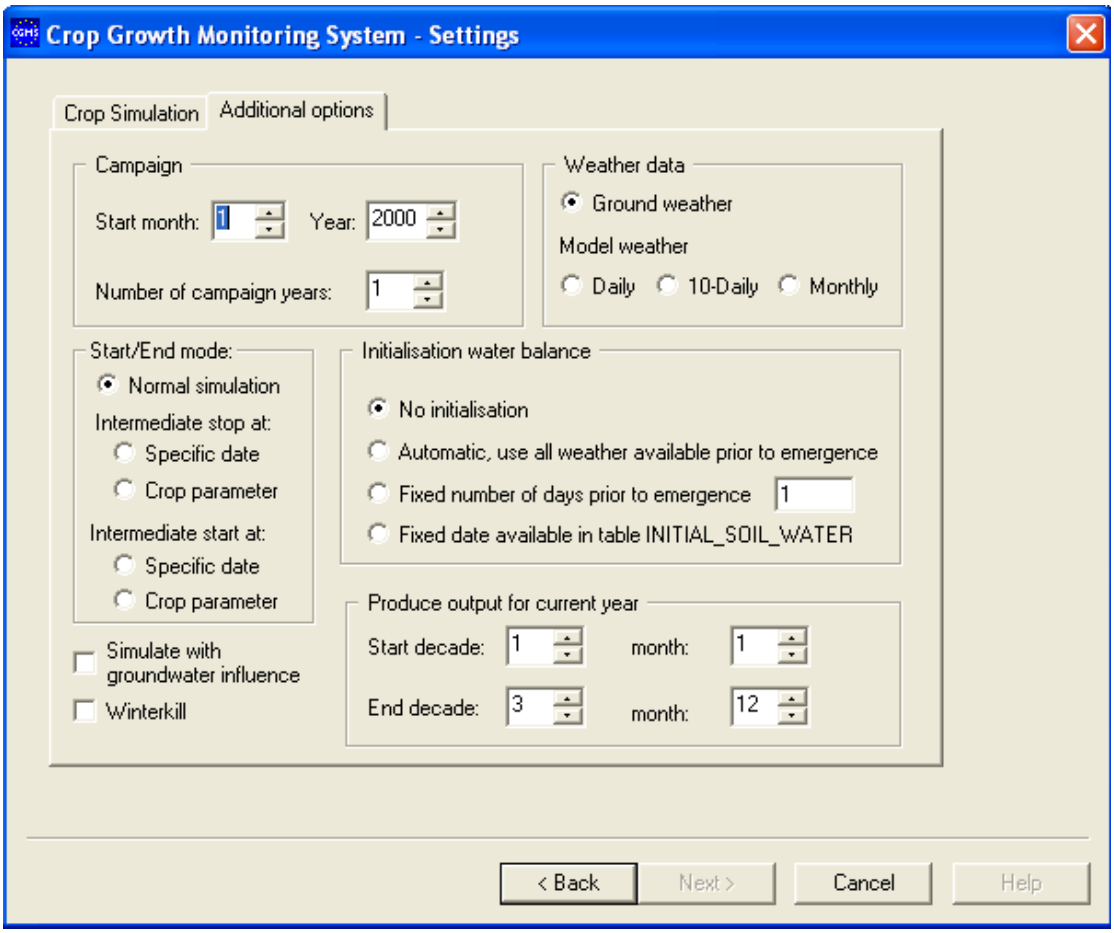


Figure 7-5: Crop simulation window –"additional options"-tab sheet

7.4.2 Campaign

In the "additional options"-tab you can define the campaign year and the number of campaigns which must be simulated. The campaign year has been introduced in CGMS version 2.0a to enable simulation for areas in the Southern Hemisphere. The month indicates the start month of the campaign. The default value is 1, which corresponds to January. In the CROP YIELD table the 36 decades are relative to the campaign start, i.e. decade X, year Y is decade X of the campaign that started in year Y. For instance if the campaign month is set on June (6), then decade 1 of campaign year Y in the CROPYIELD table corresponds to decade 1 of June in calendar year Y and decade 22 of campaign year Y corresponds to decade 1 of January in calendar year Y+1. In previous versions the campaign month was given in table SYSCON, now it is entered via the user interface.

7.4.3 Weather data

Different sources of weather data can be chosen. Interpolated station weather data is available as it was in the previous versions. Starting with CGMS version 8.0 weather model data on daily, 10-daily or monthly basis can be selected for crop growth simulation. Data are stored in the tables:

- GRID_WEATHER_MODEL_1 (day)
- GRID_WEATHER_MODEL_10 (decade)
- GRID_WEATHER_MODEL_30 (month)

A table RAINY_DAYS has been introduced which stores long term average rainy days per decade and climatic grid cell. The rainy days are necessary to calculate daily rainfall by distributing the given monthly or decade rainfall over these rainy days. Other variables like temperature, given on a monthly or decade basis, are simply interpolated to obtain daily values.

7.4.4 Start/end mode

A “restart” functionality is introduced to enable correction of the biomass of winter crop if the crop is damaged by frost. One possible way is to stop the simulation at a specific day or when a specific condition has been met. When the crop growth model is forced to stop (intermediate stop), all values of state variables are written to the table VALUE_STATE_VARIABLES. Three simulation modes are distinguished:

- **normal simulation** (start day is based on the crop calendar, no intermediate stop).
- **intermediate stop** (start day is based on the crop calendar, an intermediate stop is required).
- **intermediate start** (the start day is read from a table in the database or is determined by applying a temperature rule, no intermediate stop).

You can stop and restart the crop growth simulation by using dates given in table STOP_AND_START_DAYS (option “specific date”) or by applying a temperature rule based on a threshold value given by crop parameter TBASEM in the CROP_PARAMETER_VALUE table (option “crop parameter”). The temperature rule is defined as follows. In case of an intermediate stop the simulation will stop when during three consecutive days the mean daily temperature is below a certain threshold. The crop parameter TBASEM is used as threshold. CGMS needs to know the day from which this temperature rule must be applied. A new crop parameter is introduced, called STARTDAY_ISTOP, which indicates the Julian day to start applying the temperature rule. In case of an intermediate start the simulation will begin when three consecutive days the mean daily temperature is above a certain threshold. Again TBASEM is used as threshold and the crop parameter STARTDAY_ISTART indicates when to start checking this rule.

However some important drawbacks exist with regard to this approach:

- All the current states variables describing the leaf characteristics have to be stored in the database The storage of the arrays m_ASLA[0..365], m_ALVAge[0..365] and ALV[0..365] leads to enormous storage of data. However these parameters are needed for correction and the re-start after winter.
- The state variables have to be corrected for the amount of winterkill. This correction is complicated. The biomass of leaves is schematized in different leaf classes with their specific age, biomass etc. A correction of the leaf characteristics means that these arrays have to be changed starting with the oldest leaf classes because oldest leaves die first.

Therefore an alternative approach has been implemented to simulate winter crops from autumn without stopping during the winter and at the same time simulate frost damage or winterkilll.

7.4.5 Initialisation water balance

The initialisation of the soil water balance was introduced in version 2.0a. There are four possibilities to initialise the water balance:

1. no initialisation (the soil water balance start at emergence)
2. automatic, use all weather prior to emergence (starting from first day of the first campaign month)
3. 'fixed' number of days prior to emergence (note that the system will retrieve the same number of days of grid weather prior to the campaign start to assure grid weather is available to run the water balance)
4. 'fixed' date available in table INITIAL_SOIL_WATER.

The last possibility has been introduced in version 2.1.

You can choose one of these four options. In case of option 3 you must indicate how many days prior to emergence the simulation of the water balance must start.

However this only relates to a better estimate of the soil moisture in the upper 10 cm of the soil profile. Obviously this is less important than the estimation of the WAV parameter (table INITIAL_SOIL_WATER) which determines the initial available soil moisture of the total rooting depth (maximum rooting depth including the relatively small initial rooting depth of 10 cm)

7.4.6 Simulate with groundwater influence

Starting from CGMS version 8.0 it is possible to include the influence of groundwater when simulating the soil water balance. The table SOIL_PHYSICAL_GROUP has been extended with data on additional soil physical data such as hydraulic conductivity. The table INITIAL_SOIL_WATER is used to supply data on initial ground water level and drainage depth. The latter data are given per grid cell and soil typologic unit (STU).

In case of ground water influence the initial soil moisture profile is not based on the WAV parameter (available water in potential rooting zone) given in table INITIAL_SOIL_WATER but this profile is calculated assuming a equilibrium situation in relation to the ground water level. If the ground water level is deep the initial soil moisture profile will be very dry. To avoid such a dry start the ground water module will not be used if the initial ground water table (ZTI) is already deeper than 600 cm below soil surface.

7.4.7 Winterkill

Usually winter crops start their simulation on 1 January as vernalisation is not described by WOFOST. However from technical point of view it is possible to start the simulation of winter crops in autumn and simulate frost damage. This has been organized by inserting one extra parameter (fDSLv3) describing the fraction of the leaves that have died because of severe winter conditions. The value of this parameter is based on the approach taken from the CERES model (Ritchie, 1991).

7.4.8 Save output for one or more decades

When simulating the current year (table SYSCON) it is possible to only save results for a certain period defined by decade and months. This is important when using CGMS in an

operational production line where you have to update simulated crop indicators each 10 days. Each 10 days (or decade) you have to rerun the crop simulation but by indicating the period it is possible to only save the last 10 days. This way unnecessary database actions are avoided like deleting and written simulation of the previous 10-day periods. In case of historic years all decades from emergence until decade 36 are written to the database.

After selecting the right options you can press the “next”-button and you will see the summary window. In this window you click on the “finish”-button to effectuate all your choices and start the calculations. The options chosen in the second window (see Figure 7-5) are stored in table CGMS_SYSLOG.

In chapter 6 you can find a flow chart with all tables used as input and output for the crop growth simulation.

7.5 Aggregation

When choosing “aggregation” in the “execution steps” window, the “aggregation” appears (see Figure 7-6). Aggregated indicators on administrative region level are needed as input for the yield forecast procedure. Aggregated indicators on climatic grid cell level are needed to present results in a GIS.

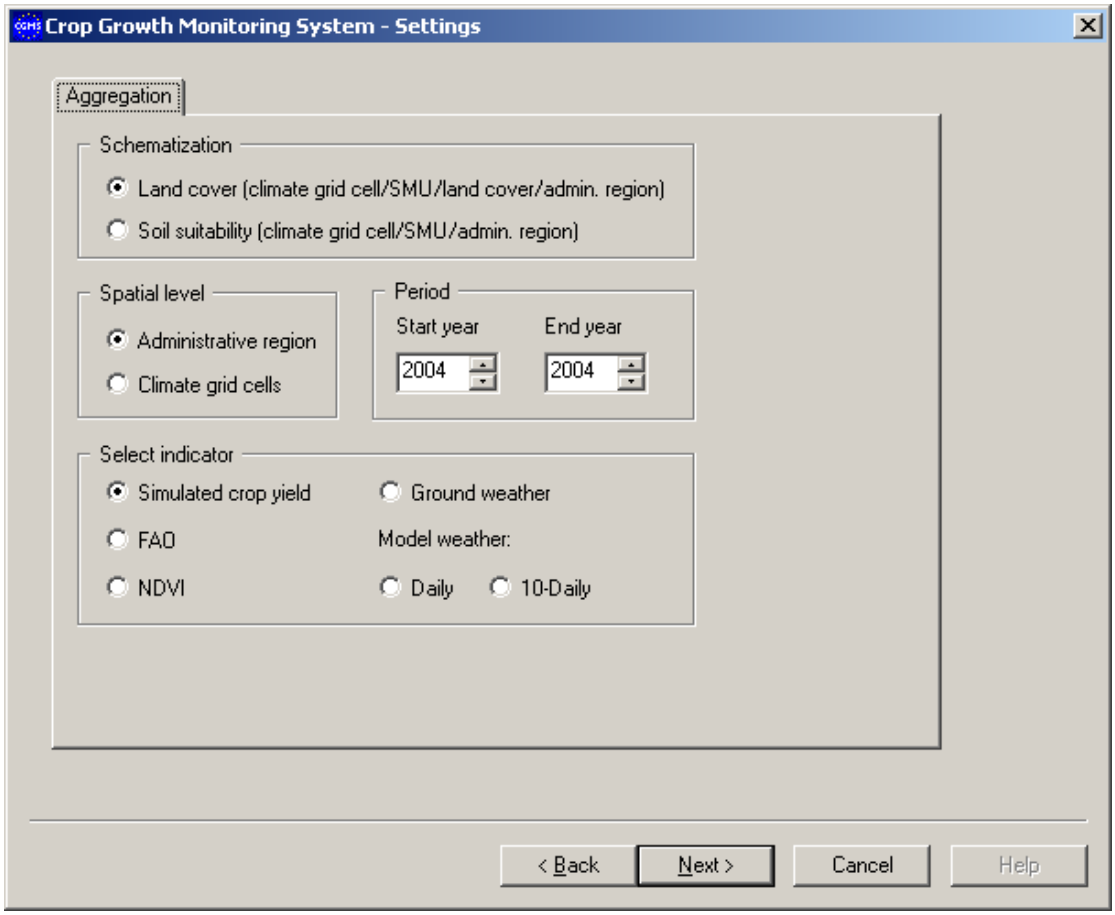


Figure 7-6: Aggregation window

The user can select:

- A type of schematisation (land cover or soil suitability).

- Spatial level to which the indicator will be aggregated.
- Period for which the indicator will be aggregated.
- Type of indicator.

7.5.1 Schematization

You have two possible ways of aggregation: one using the soil suitability rules and one using actual land cover data. Two additional tables are created to support these aggregations. One table called EMU_PLUS_NUTS which is an intersection of climatic grid cells, SMUs and administrative regions. This table will be used to aggregate to grid cells and administrative regions using soil suitability rules. The other table is called EMU_PLUS_NUTS_LANDCOVER and is an intersection of climatic grid cells, SMUs, administrative regions and land cover. This table enables aggregation based on land cover thus only include the areas covered by the specific land cover.

The aggregation from the most detailed administrative region level 2 to more global levels is based on the area statistics given per region (table AGGREGATION_AREAS) and not on the two previously discussed schematisations.

7.5.2 Spatial level

Different indicators can be aggregated to climatic grid level or administrative region level. In case of administrative regions the aggregation will be done for all three levels. Because interpolated station weather and model weather are already entered into the database at climatic grid level aggregating to administrative region is the only logical option for these indicators.

Except for “ground weather” (interpolated station weather) and “model weather – daily”, all indicators available in the aggregation procedure are given per decade. In the spatial aggregation to climatic grid cell and administrative regions this temporal resolution is maintained. The “ground weather” and “model weather – daily” indicators will be aggregated from daily values to decade values within the spatial aggregation to administrative regions.

7.5.3 Period

The period determines the time series for which the aggregation will be executed.

7.5.4 Select indicator

The following indicators are available in the aggregation
- Simulated crop yield (see CROP_YIELD table)
- FAO Water Index model results (see FAO_WATER_INDEX_MODEL table)
- NDVI (see NDVI table)
- Interpolated station weather (see GRID_WEATHER table)
- Model weather – daily (see GRID_WEATHER_MODEL_1 table)
- Model weather – 10-daily (see GRID_WEATHER_MODEL_10 table)

To avoid erroneous aggregated values all fields in the above listed tables must be complete thus no missing values (null values). Monthly model weather data has been excluded from the aggregation. In the yield forecast procedure, which needs aggregated indicators, monthly model weather data is not required.

Aggregated indicators must be stored in the database. Therefore several tables are available:

- GRID_YIELD table to store aggregated simulated crop indicators at grid level.
- GRID_NDVI table to store aggregated NDVI at grid level
- GRID_FAO_WATER_INDEX_MODEL table to store aggregated FAO water indices at grid level
- NUTS_NDVI table to store aggregated NDVI at administrative region level
- NUTS_FAO_WATER_INDEX_MODEL table to store aggregated FAO water indices at administrative region level
- NUTS_WEATHER table to store aggregated interpolated station weather data (originally given per day) at administrative region level
- NUTS_WEATHER_MODEL_1 table to store aggregated model weather (originally given per day) at administrative region level
- NUTS_WEATHER_MODEL_10 table to store aggregated model weather at administrative region level

The table NUTS_YIELD to store simulated crop indicators, already existed in previous versions of CGMS.

Some new tables have been created to store intermediate results or supply data needed in the aggregation:

- DAY_DECADE to store definitions of decades.
- FAO_WATER_INDEX_MODEL_SMU table to store aggregated FAO water indices at GRID/SMU level.
- GRID_WEATHER_10 table to store the result of the aggregation of daily interpolated station weather from day to decade.
- GRID_WEATHER_MODEL_1_10 table to store the result of the aggregation of daily model weather from day to decade.

After selecting the right options you can press the “next”-button and you will see the summary window. In this window you click on the “finish”-button to effectuate all your choices and start the calculations.

In chapter 6 you can find a flow chart with all tables used as input and output for the aggregation.

7.6 Preparation of data yield forecast calculation

When choosing “preparation of data yield forecast calculation” in the “execution steps” window, the “preparation of data yield forecast calculation – step 1” appears (see Figure 7-7). This procedure prepares data for the yield forecast calculation. The first window offers the following options:

- Analysis (type of statistical analysis).
- Administrative region of interest.
- Period.
- Statistical crops.

- Decade of yield forecast.

The next step leads to a window where the user can define the indicators and the statistical parameters.

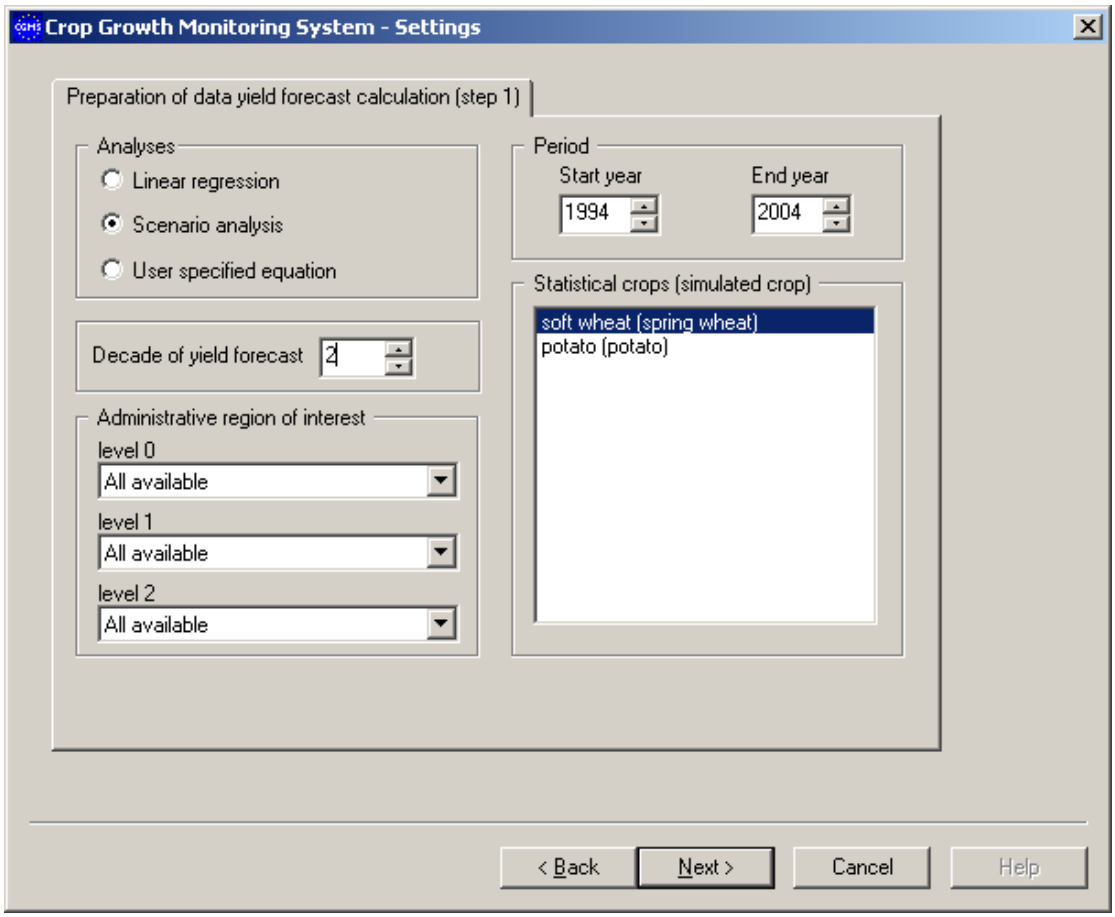


Figure 7-7: Preparation of data yield forecast calculation – "step 1" window

7.6.1 Analyses

Because the type of analysis determines the input data, you must first select an analysis. There are three type of statistical analysis offered:

- **Linear regression.** In the CGMS statistical sub-system all available indicators in the database at administrative region level can be applied, but the number of indicators remains fixed to four indicators among which the best performing indicator will be selected.
- **Scenario-analysis.** Defining year-analogues applying Principal Component Analysis (PCA) and Cluster Analysis (CA) on the basis of all available indicators.
- **User specified equation.** Different forecast methods can be incorporated in the system by providing the option to apply user-specified equations. In this case, you specify a regression equation between yield and separate indicators. The equation will have the form $y = a*X_1\text{^}d+b*X_2+c$, where a, b, c and d are coefficients entered by the user through the CGMS user interface and, X_1 and X_2 are variables linked to two among all available indicators.

7.6.2 Administrative regions of interest

Next, the administrative region(s) of interest can be chosen (see table NUTS). Default the string “all available” is shown, indicating all regions will be simulated. The administrative regions are organised as follows: the highest level, for instance the whole country, is called “level 0”, which is divided in smaller regions called “level 1”. These smaller regions are again divided in the smallest regions called “level 2”. Via the pull down menus you can select one region at “level 0” and simulate for this “level 0”-region all “level 2”-regions that belong to this selected region. Or you can continue to limit the “level 2”-regions by selecting only one “level 1”-region within the “level 0”-region. Finally, you can decide to simulate only one “level 2”-region.

7.6.3 Period

The period determines the time series for which the statistical analysis is applied.

7.6.4 Statistical crop

You can choose one or more so-called “statistical” crops. In the table STAT_CROP these “statistical” crops are linked to the simulated crops as they occur in table CROP.

7.6.5 Decade of yield forecast

The decade for which the yield forecast is performed must be indicated. This is used to store the results for this decade.

7.6.6 Definition of composed indicators

To select the indicators you must click the “next”-button. This leads to a second window (see Figure 7-8) containing definition of indicators.

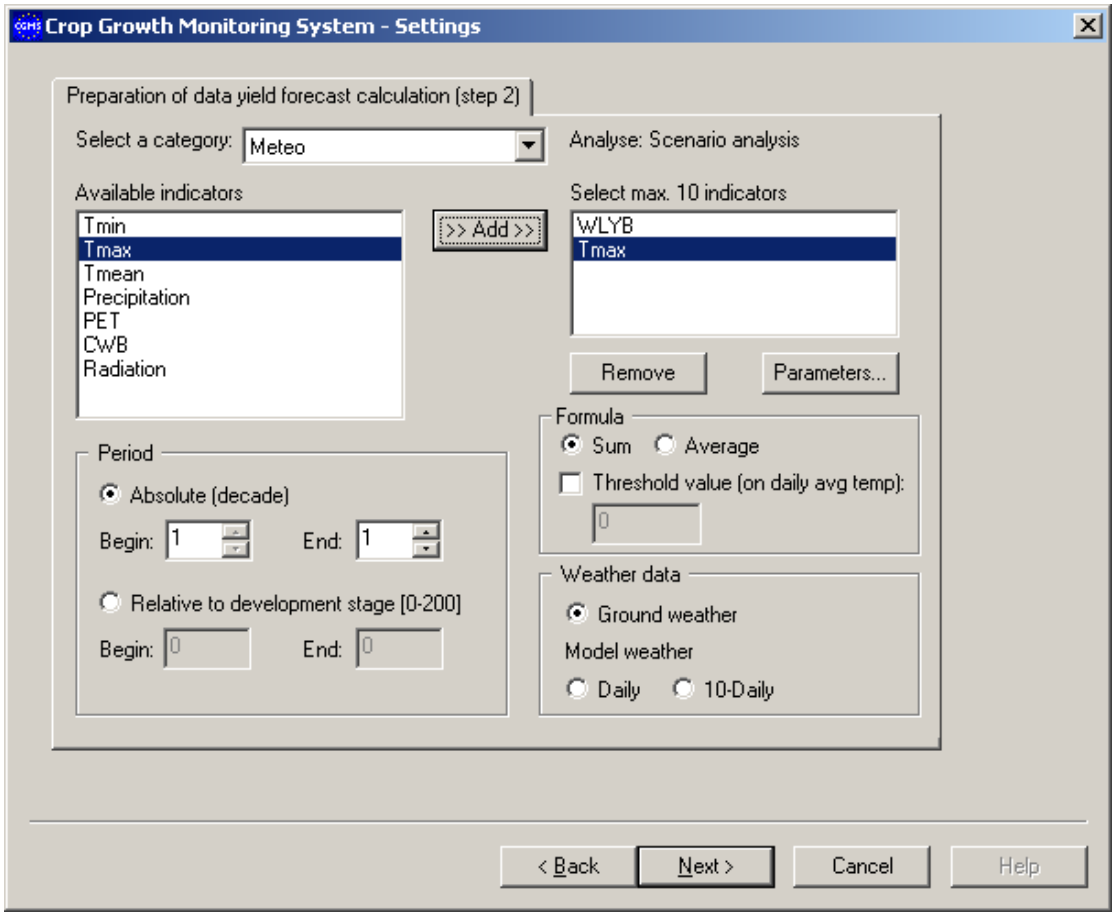


Figure 7-8: Preparation of data yield forecast calculation – "step 2" window

You can select three different categories (see Figure 7-8):

- Meteo:
  - Tmin (minimum temperature)
  - Tmax (maximum temperature)
  - Tmean (mean temperature)
  - Precipitation (rainfall)
  - PET (potential evapotranspiration, ET0)
  - CWB (climatic water balance = rainfall minus potential evapotranspiration)
  - Radiation (radiation)
- Simulated crop (simulated crop indicators):
  - PYB (potential dry weight biomass)
  - WLYB (water limited dry weight biomass)
  - PYS (potential dry weight storage organs)
  - WLYS (water limited dry weight storage organs)
  - PLAI (potential leaf area index: leaf area divided by surface area)
  - WLAI (water limited leaf area index: leaf area divided by surface area)
- Other:
  - WSI (FAO Water Satisfaction Index)
  - NDVI (Normalised Difference Vegetation Index)
  - CNDVI (Corine-NDVI)

Via the "add"-button you can select one or more indicators of one or more different categories. There are limits for each type of statistical analysis. For the "Linear regression"

you must select 4 indicators; for the “scenario-analysis” you can select up to 10 indicators and for “user specified equation” 2 indicators. Via the “remove”-button you can undo the selection.

In the right box with the selected indicators you can define each indicator in more detail by selecting one indicator and make the additional choices. These additional choices depend on the category to which the indicator belongs.

For “meteo”-indicators the following choices must be made:

- Kind of period:
  - absolute: defining a start and end decade over which the indicator will be calculated.
  - relative to development stage: defining a start and end development stage over which the meteo indicator will be calculated.
- Formula: a “sum” or an “average” can be calculated over the period. The “threshold value” is based on the mean daily temperature and is applied as follows. If the mean daily temperature is below the threshold, values of the specific indicator are not included in the calculation of the sum or the average. In the special case of indicator “Tmean”, the value is corrected by subtracting the threshold from the original value. This corresponds to the definition of temperature sums. The usage of the threshold is optionally (see the check box).
- Weather data: the user can choose between different sources of weather data:
  - Ground weather (interpolated station weather)
  - Daily model weather
  - 10-daily model weather

In case of “simulated crop”-indicators or “other”-indicators the choices are simple: only a decade (end decade) or development stage can be selected.

7.6.7 Selection of statistical parameters

When you have planned to do the “user specified equation” and you have selected two composed indicators, an additional button appears through which you can define the equation. After clicking on button “equation” you must enter the coefficients of the equation (see Figure 7-9). The two composed indicators and the coefficients are written in table META_DATA_USER_SPEC_EQUATION.

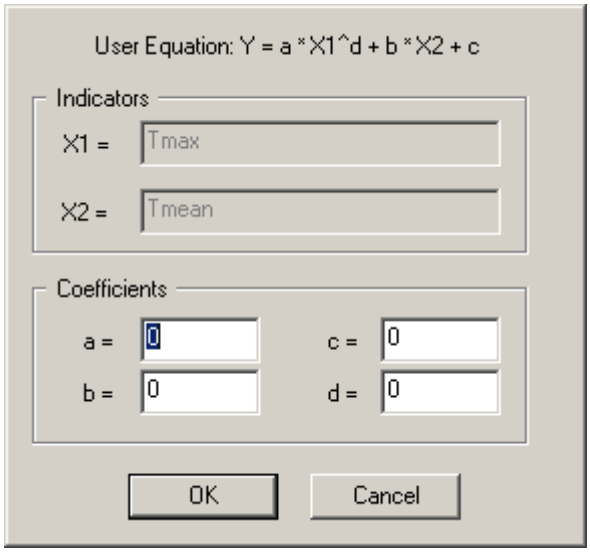


Figure 7-9: Definition of the user specified equation - window

In case you have chosen the “linear regression” and selected the 4 composed indicators you can click an additional button to change the parameters in a separate window (see Figure 7-10):

- WINDOW_SIZE window size (default 12)
- MIN_OBS minimum observations (default 8)
- SLIDING_WINDOW sliding window (1) or not (0) (default 0)

The changed (or default, if not changed) values are written to table META_DATA_REGR.

In case you have chosen the “scenario analysis” and selected the composed indicators you can click an additional button to change the parameters in a separate window (see Figure 7-10):

- MAX_PCOMPCOUNT maximum number of principal components to process (default 2)
- MIN_EXPL_VAR minimum percentage of variance: between 0 and 1 (default 0.9)
- CUTOFF_D1 Cutoff score dissimilarity #1 - save pairs with dissimilarity < Cutoff D1 (default 1)
- CUTOFF_D2 Cutoff score dissimilarity #2 - save pairs with dissimilarity < Cutoff D2 (default 2). This is only used when the number of similar years is smaller than MIN_SIM_YEARS
- MIN_SIM_YEARS minimum number of similar years (default 5)
- MIN_OBS minimum observations (default 8)

The changed (or default, if not changed) values are written to table META_DATA_SCEN.

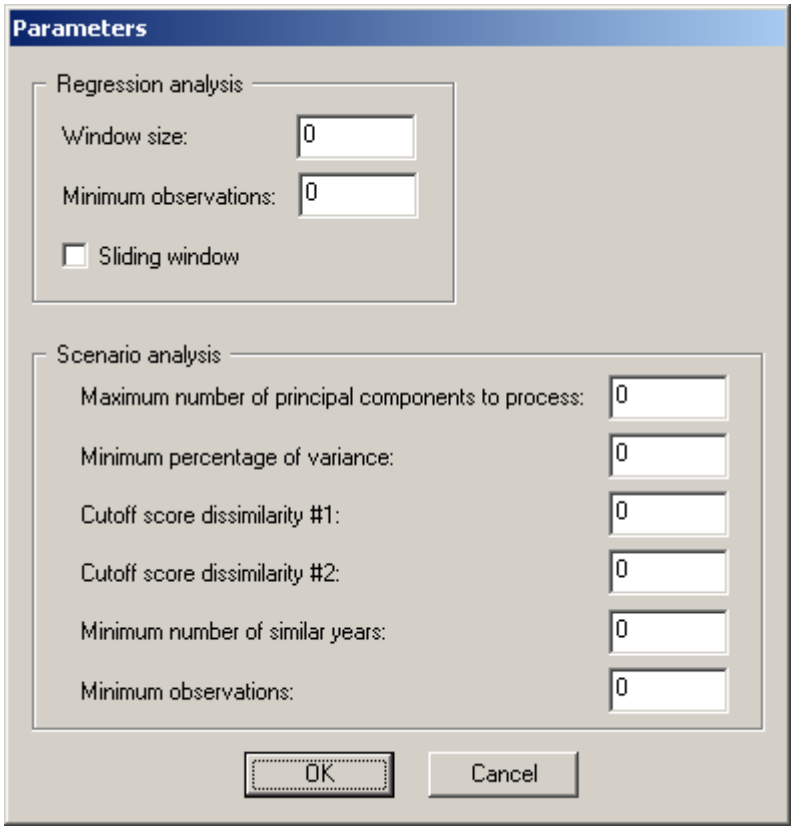


Figure 7-10: Window to edit parameters of 'linear regression' and 'scenario-analysis'

7.6.8 Storage of prepared data

After selecting the right options you can press the "next"-button and you will see the summary window. In this window you click on the "finish"-button to effectuate all your choices and start the calculations. For the selected years, statistical crops, and administrative regions the values of all newly composed indicators are written to the table DATA_FOR_YIELD_FORECAST. In this table the field "INDICATOR_CODE" holds the definition of the new composed indicator using the:

- indicator code
- start and end period (in case of meteo) or end period (in case of other indicators)
- absolute or relative
- threshold (in case of meteo)
- formula (in case of meteo)
- weather data source (in case of meteo)

.

Finally, CGMS writes information to the table PARAM_MAIN. This table contains the fields:

- analysis selected statistical analysis
- start_year first year of period for which data have been prepared
- end_year last year of period for which data have been prepared
- store_results indicator to store (1) or not store (0) intermediate results of statistical analysis
- show_debug indicator to show all executed statements (1) or not (0)

After the data have been written to the database, the button "Launch forecast" is enabled. Clicking on this button will start the CGMS statistical sub-system.

In chapter 6 you can find a flow chart with all tables used as input and output for the data preparation for yield forecast calculation.

Besides there is a button called “Export prepared data” to export the data set just prepared by the user. The data will be written to comma separated value file on a directory indicated by the user.

7.7 Import external data to database

When choosing “import external data to database” in the “execution steps” window, the window appears containing an “import” tab sheet (see Figure 7-11). The import procedure reads external data, formatted as comma separated value files, and writes the data into the database. The window offers the following options:

- Data type
- Filename

7.7.1 Data type

You can select one data source type among the following:

- FAO Water Index Model (see for contents table FAO_WATER_INDEX_ MODEL)
- NDVI remote sensing indicators (see for contents table NDVI)
- CNDVI remote sensing indicators (see for contents table CNDVI)
- Initial hydrology conditions (see for contents table INITIAL_SOIL_WATER)
- “Stop” and “start” days (see for contents table STOP_AND_START_DAYS)
- Start day and type (see for contents table CROP_ CALENDAR)
- Model weather – daily
- Model weather – 10-daily
- Model weather – monthly (see for contents table GRID_WEATHER_ MODEL_30)

The model weather (daily and 10-daily) to be imported differs from the contents and format of the tables GRID_WEATHER_MODEL_1 and GRID_WEATHER_ MODEL_10. Table 7-1 and Table 7-2 explain how daily and 10-daily model weather data are converted into the format defined in the database.

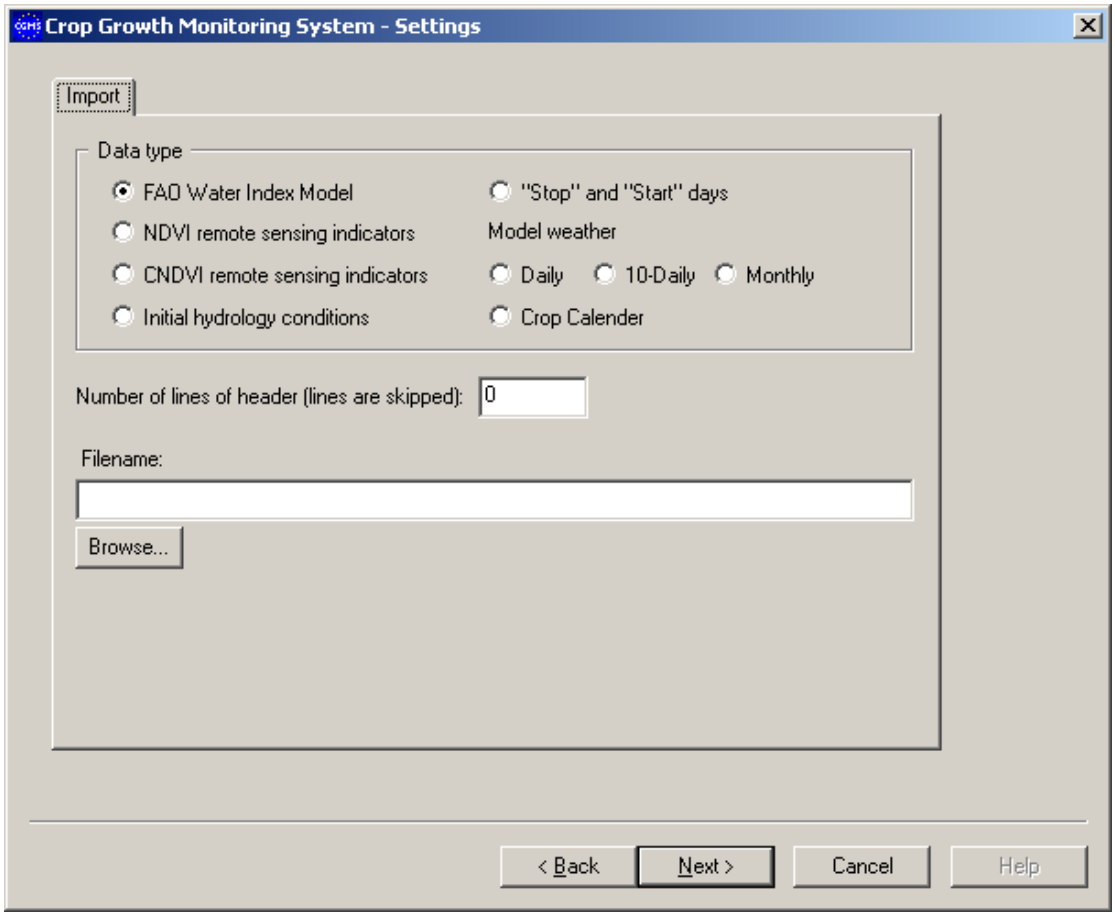


Figure 7-11: Import window

Table 7-1: Conversion of daily model weather data during import into table GRID_WEATHER_MODEL_1

Fields of file to be imported	Fields in table GRID_WEATHER_MODEL_1
Latitude	Converted into unique grid number using formula: (500+latitude)*1000+(500+longitude)
Longitude	
Day in month	Converted into date format
Month	
Year	
Minimum temperature	Minimum temperature
Maximum temperature	Maximum temperature
Mean temperature	Mean temperature
Rainfall	Rainfall
ET0	ET0
Climatic water balance	**Is not imported**
Global radiation	Global radiation
ES0	ES0
E0	E0
Not available	Vapour pressure (given value ‘0’)
Not available	Wind speed (given value ‘0’)
Not available	Snow depth (given value ‘0’)

Table 7-2: Conversion of 10-daily model weather data during import into table GRID_WEATHER_MODEL_10

Fields of file to be imported	Fields in table GRID_WEATHER_MODEL_1
Latitude	Converted into unique grid number using formula: (500+latitude)*1000+(500+longitude)
Longitude	
Decade	Converted into year decade (1-36)
Month	
Year	Year
Minimum temperature	Minimum temperature
Maximum temperature	Maximum temperature
Temperature sum	**Is not imported**
Mean temperature	Mean temperature
Rainfall	Rainfall
ET0	ET0
Climatic water balance	**Is not imported**
Global radiation	Global radiation
ES0	ES0
E0	E0
Not available	Vapour pressure (given value ‘0’)
Not available	Wind speed (given value ‘0’)
Not available	Snow depth (given value ‘0’)

The data must be stored in “comma separated value” files with optionally a header describing the data. The user can indicate how many lines this header contains (0 for no header).

7.7.2 Filename

After the selection of the type you must point the specific data file. Via the “browse” button you navigate to the right directory and click on the right filename.

After selecting the right options you can press the “next”-button and you will see the summary window. In this window you click on the “finish”-button to effectuate all your choices and start the import.

7.8 Export data

The export procedure (see Figure 7-12) can be used to export the following data:

- Selection from the CROP_YIELD tables. The user must select the crop and year. The filename is based on the table and the selections. For this combination of crop, year and decade all records are written to the specified file in comma separated value format on the directory where the executable is located.
- Selection from the FORECASTED_NUTS_YIELD_REGR table. The user can select the crop and year. The filename is based on the table and the selections. For this combination of crop and year all records are selected of the latest run code and are written to the specified file in comma separated value format on the directory where the executable is located.

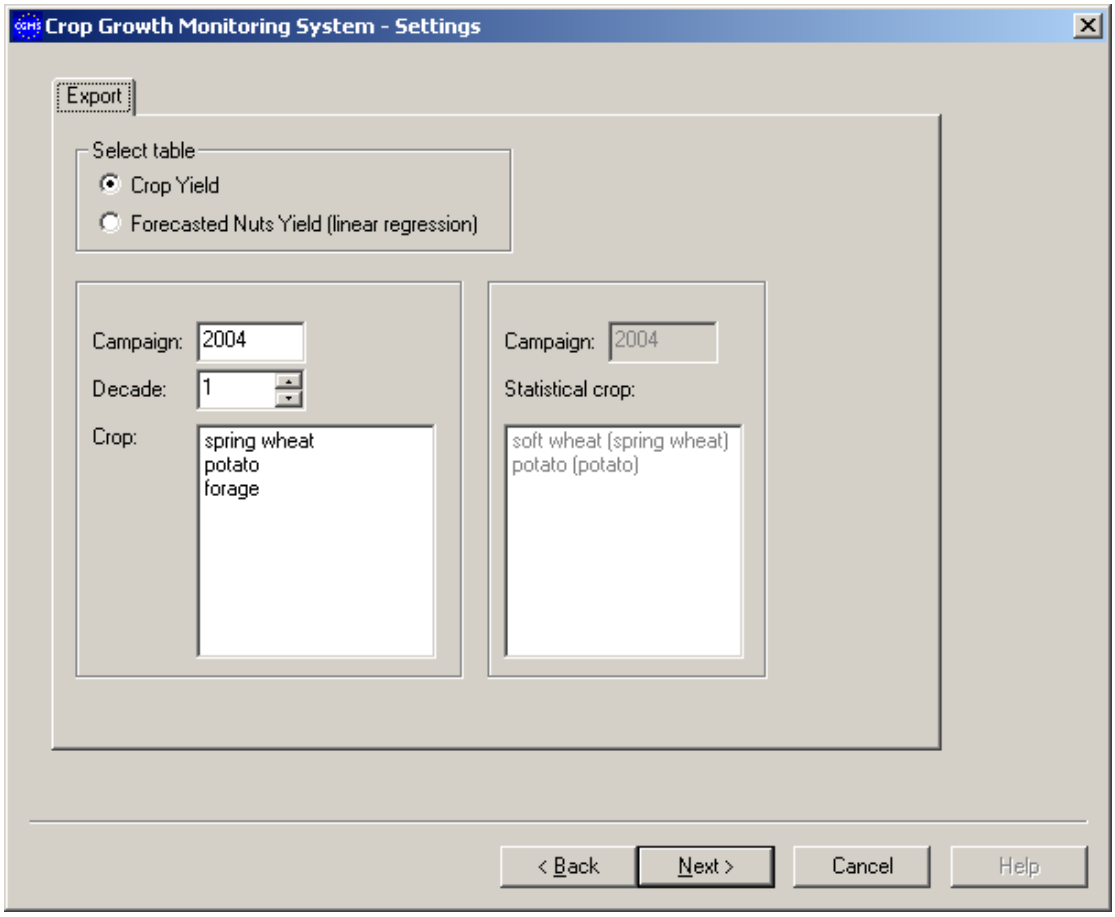


Figure 7-12: Export - window

After selecting the right options you can press the “next”-button and you will see the summary window. In this window you click on the “finish”-button to effectuate all your choices and start the export.

References

Ångström, A, 1924. Solar and terrestrial radiation. Quarterly Journal of the Royal Meteorological Society, 50:121-125.

Beek, E.G., 1991. Spatial interpolation of daily meteorological data. Theoretical evaluation of available techniques. Report 53.1, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 43.

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.

Bouman, B.A.M., Schapendonk, A.H.C.M., Stol, W. and Kraalingen D.W.G van, 1996. Description of the grassland growth model LINGRA as implemented in CGMS. AB-DLO.

Burrill, A., Vossen, P., 1992. Development of an operational agro-meteorological monitoring system. In: F. Toselli, J. Meyer-Roux (eds). Proceedings of conference on the application of remote sensing to agricultural statistics, 26-27 November 1991, Belgirate, Italy. EUR 14262 EN, Office for Official Publications of the EU, Luxembourg, p 357-360.

Diepen, C.A. van, Rappoldt, C., Wolf, J., Keulen, H. van, 1988. Crop growth simulation model WOFOST. Documentation version 4.1. for World Food Studies, Wageningen, The Netherlands.

Diepen, C.A. van, Wolf, J., Keulen, H. van, 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, J. Meyer-Roux (eds). Proceedings of conference on the application of remote sensing to agricultural statistics, 26-27 November 1991, Belgirate, Italy. EUR 14262 EN, Office for Official Publications of the EU, Luxembourg, p 269-277.

Diepen, C.A. van, 1998. Application of simple interpolation methods in agrometeorology. In: B. Gozzini, M. Hims (eds). Proceedings of workshop on dealing on spatialisation, 24-25 September 1996, Toulouse. EUR 18473 EN, Office for Official Publications of the EU, Luxembourg, p 3-17.

Diepen, C.A. van, Wal, T. van der, 1995. Crop growth monitoring and yield forecasting at regional and national scale. In: J.F. Dallemand, P. Vossen (eds). Workshop for Central and Eastern Europe on agrometeorological models: theory and applications in the MARS project, 21-25 November 1994, Ispra, Italy. EUR 16008 EN, Office for Off. Pub. of the EU, Luxembourg, p 143-157.

Drift, van der J.W.M., Diepen, C.A. van, 1992. The DBMETEO data base on the countries of the European Communities. Description of the data and data-treatments
Technical Document 4, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 53.

Genovese, G.P., 2001. Introduction to the MARS Crop Yield Forecasting System (MCYFS). Meeting on 4 and 5 October 2001, Luxembourg. Space Applications Institute, Joint Research Centre of the European Commission, Ispra, Italy, pp 15.

Genovese, G., Bettio, M. (eds). 2004. Methodology of the MARS Crop Yield Forecasting System. Volume 4. Statistical data collection, processing and analysis. In: European Communities (ISBN 92-894-8183-8)

Goot, E. van der, 1997. Technical description of interpolation and processing of meteorological data in CGMS. Joint Research Centre of the European Commission, Ispra, Italy, pp 23.

Goot, E. van der, 1998a. Spatial interpolation of daily meteorological data for the Crop Growth Monitoring System (CGMS). In: M. Bindi, B. Gozzini (eds). Proceedings of seminar on data spatial distribution in meteorology and climatology, 28 September - 3 October 1997, Volterra, Italy. EUR 18472 EN, Office for Official Publications of the EU, Luxembourg, p 141-153.

Goot, E. van der, 1998b. Data requirements and preparation for CGMS. Joint Research Centre of the European Commission, Ispra, Italy, pp 26.

Hargreaves, G.L., Hargreaves G.H., Riley, J.P., 1985. Irrigation water requirement for Senegal River Basin. Journal of Irrigation and Drainage Engineering, ASCE 111, 3:265-275.

Hijmans, R.J., Guiking-Lens, I.M., Diepen, C.A. van, 1994. WOFOST 6.0; User's guide for the WOFOST 6.0 crop growth simulation model. Technical Document 12, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 143.

Hooijer, A.A., Bulens, J.D., Diepen, C.A. van, 1993. CGMS version 1.1; user manual. Technical Document 14.2, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 81.

Kappel, R.R., Supit, I., 1998. Collection of meteorological data for the MARS, MERA and PHARE projects. Joint Research Centre of the European Commission, Ispra, Italy, Joint Research Centre of the European Commission, Ispra, Italy, pp 13.

Keulen, H. van, Wolf, J., (eds.), 1986. Modelling of agricultural production: weather, soils and crops. Simulation Monographs, Pudoc, Wageningen, The Netherlands, pp 479.

Kropff, M.J. and H.H. van Laar (eds.), 1993. Modelling crop-weed interactions. CAB-International, Oxford.

Kuyper, M.C., 2001. CGMS statistical sub-system v4.2. Transformation from Unix to Windows platform. Technical manual. Alterra, Green World Research, Wageningen, The Netherlands, pp 23.

Laar, H.H. van, J. Goudriaan, Keulen, H. van (eds.), 1992. Simulation of crop growth for potential and water-limited production situations (as applied to spring wheat). Simulation Reports CABO-TT 27. CABO-DLO, WAU-TPE, Wageningen.

Lazar, C. and G. Genovese (eds). 2004. Methodology of the MARS Crop Yield Forecasting System.Volume 2. Agrometeorological data collection, processing and analysis. In: European Communities (ISBN 92-894-8181-1).

MeteoConsult, 1991. AMDaC System Manual. MeteoConsult, Wageningen, The Netherlands, pp 40.

Micale, F. and G. Genovese (eds). 2004. Methodology of the MARS Crop Yield Forecasting System. Volume 1. Meteorological data collection, processing and analysis. In: European Communities (ISBN 92-894-8180-3)

Penman, H.L., (1948). Natural evaporation from open water, bare soil and grass. Proceedings Royal Society, Series A, 193:120-146.

Prescott, J.A., 1940. Evaporation from a water surface in relation to solar radiation. Transactions of the Royal Society of South Australia, 64:114-118.

Raaij, B.F.T. van, Wal, T. van der, 1994. CGMS 3.1 Technical description. Technical Document 15.3, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 205.

Ritchie, J.T. 1991. Wheat phasic development. p. 31–54. In R.J. Hanks and J.T. Ritchie (ed.) Modeling plant and soil systems. Agron. Monogr. 31. ASA, CSSA, and SSSA, Madison, WI.

Savin, I., Boogaard, H., Diepen, C. van, Nègre, T. (eds), 2004. CGMS version 8.0. User Manual and Technical Documentation. European Communities (ISBN 92-894-8339-3).

Schapendonk, A.H.C.M., Stol, W., Kraalingen, D.W.G. van, Bouman, B.A.M., 1998. LINGRA, a sink/source model to simulate grassland productivity in Europe. European Journal of Agronomy 9 (1998) 87-100.

Spitters, C.J.T., H. van Keulen and D.W.G. van Kraalingen, 1989. A simple and universal crop growth simulator: SUCROS87. In: Rabbinge et al., 1989. 147-181.

Supit, I., Hooijer, A.A., Van Diepen, C.A., Edts, 1994 (reprint 1996). System description of the WOFOST 6.0 crop simulation model implemented in CGMS. *Volume 1: Theory and Algorithms*. EUR Publication N° 15959 EN of theOffice for Official Publications of the EC. Luxembourg, 146 pp.

Supit, I., Kappel, R.R. van, 1998. A simple method to estimate global radiation. Solar Energy, 63:147-160.

Voet, P. van der., Diepen, C.A. van, Oude Voshaar, J., 1994. Spatial interpolation of daily meteorological data. A knowledge-based procedure for the regions of the European Communities. Report 53.3, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 105.

Vossen, P., 1995. Early crop production assessment of the European Union, the systems implemented by the MARS-STAT project. In: J.F. Dallemand, P. Vossen (eds). Workshop for Central and Eastern Europe on agrometeorological models: theory and applications in the MARS project, 21-25 November 1994, Ispra, Italy. EUR 16008 EN, Office for Off. Publ. of the EU, Luxembourg, p 21-51.

Vossen, P., Rijks, D., 1995. Early crop yield assessment of the EU countries: the system implemented by the Joint Research Centre. EUR 16318 EN, Office for Official Publications of the EU, Luxembourg, pp 180.

Wal, T. van der, 1994. CGMS version 3.1 Data installation guide and operator's manual. Technical Document 15.2, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 60.

Wal, T. van der, 1994. CGMS version 3.1 Data installation guide and operator's manual. Technical Document 15.2, DLO Winand Staring Centre, Wageningen, The Netherlands, pp 60.

Appendix A Using CGMS in batch mode

To start CGMS in batch mode use the parameter /batch. For example:
C:\Data\cgms_food\cgms.exe /batch

Instead of opening the UI an ini-file is read with the choices that are normally made in the UI. This file is called “cgms.ini” and should be located at the same directory as the executable “cgms.exe”. The content of the file looks like this:

```
[Database]
DSN= CGMS_92
User=
Password=

CalplatDSN=CGMS_92_CALPLAT
CalplatUsername=
CalplatPassword=

RunMode= OPERATIONAL
ShowBatchInfo=1

[Calculations]
WeatherCalculation=0
CropSimulation=1
Aggregation=1
PreperationForecast=0
Import=0
Export=0

[Weather]
WeatherAdd=1
WeatherAvail=1
weatherGrid=0
WeatherStartDate=01\01\1994
WeatherEndDate=31\12\1994
GlobalRadiationModel=0

[Simulation]
SimCroOutputYield=1
SimCroRunPot=1
SimCroRunWL=0
SimCroStartMonth=1
SimCroStartYear=1993
SimCroCampaignYears=1
SimCroWeatherData=1
simcroOutputStartDecade=19
simcroOutputEndDecade=19
StartEndMode=1
InitWaterBalance=0
GWInfluence=0
Winterkill=0
```

```
Crop=1,1,spring wheat,0
Crop=7,2,potato,0
SimCroAdminLevel1=All available
SimCroAdminLevel2=All available
SimCroAdminLevel3=All available

[Aggregation]
Schematization=1
SpatialLevel=1
AggIndicator=4
AggStartYear=1994
AggEndYear=1994

[Preparation]
Analyse=1
PrepAdminLevel1=All available
PrepAdminLevel2=All available
PrepAdminLevel3=All available
PrepStartYear=1993
PrepEndYear=1993
StatCrop=soft wheat,2,spring wheat,1
StatCrop=potato,11,potato,7
Indicator=1,TMax,ABS,1,16,SUM,NoThreshold,ground
Indicator=1,TMin,REL,1,150,AVG,0.5,model1
Indicator=1,PET,REL,1,150,AVG,0.5,model1
Indicator=2,PYB,ABS,10
EquationCoefs=1,2,3,4

[Import]
DataType=1
HeaderLines=0
ImpFileName=C:\Data\projecten\MARSCGMS\CGMS_Food\Data\FAO_WIM.txt

[Export]
ExpYear=1993
ExpDecade=0
ExpCropNr=2
```

Empty lines are skipped. Also lines with the words in brackets are skipped or lines starting with semicolon (;). CGMS just looks for the keywords like DSN, User, Password etc. The search is case-insensitive.

If the file is read correct and a connection with a database could be made the selected calculations start. The program opens a DOS-box and reports the progression. When all calculations are finished the box is closed automatically.

[Database]

For test purpose an MS-Access database is used. If User and Password are empty the program assumes that an MS-Access database is used. To insert in an Oracle database all the fields under [Database] should have a value.

In case CGMS is used for CALPLAT a different database can be used to provide weather data (grid_weather). Therefore one should use CalplatDSN, CalplatUsername and CalplatPassword.

Runmode indicates the mode: CALPLAT or the OPERATIONAL mode (default).

ShowBatchInfo:
1 = Show the cgms-window when run in batch mode (default)
0 = Minimize the window

[Calculations]
At this part indicate with “1” the calculations that should be performed.
- WeatherCalculation : Weather data calculation
- CropSimulation : Crop simulation
- Aggregation : Aggregate data
- PreperationForecast : Preparation of data for yield forecast calculation
- Import : Import external data to database
- Export : Export data

[Weather]
This section includes the settings for the weather data calculations.
First 3 items indicate whether the calculation should be performed.
- WeatherAdd : Additional Weather Parameters
- WeatherAvail : Weather Data Availability
- WeatherGrid : Grid Weather

Period of the calculations: WeatherStartDate and WeatherEndDate
For the dates instead of the slash (“\”) it is allowed to use another date-separator. Always use 2 digits for the day and month, and 4 for the year.

Select Global radiation model: 0 = Supit/Hargreaves, 2 = Choisnel

[Simulation]
This section includes the settings for the crop simulation. Select first the period of the simulation. Start month, start year and the number of campaigns (years) to simulate: SimCroStartMonth, SimCroStartYear, SimCroCampaignYears. For the current year select the period to produce output to database: SimcroOutputStartDecade and SimcroOutputEndDecade.

Because of the CALPLAT mode it is possible to save results per day in stead of decade. Therefore use the variable SimCroOutputYield:
0 = Decade
1 = Daily

Independent you can run the Potential and/or Waterlimited run.

SimCroRunPot: Run Potential simulation
SimCroRunWL: Run Water limited simulation

Select weather data to use for simulation: SimCroWeatherData
1 = Ground Weather
2 = Model Weather daily
3 = Model Weather 10-daily
4 = Model Weather monthly

Select Start/stop mode: StartEndMode
1 = Normal simulation
2 = Intermediate stop at specific date
3 = Intermediate stop at Crop parameter
4 = Intermediate start at Specific date
5 = Intermediate start at Crop parameter

Initialisation of water balance: InitWaterBalance
0 = No initialisation
1 = Automatic, use all weather available prior to emergence
-x = Fixed number of days (x) prior to emergence
2 = Fixed date available in table INITIAL_SOIL_WATER

Simulate with or without ground water influence GWInfluence:
0 = without ground water influence
1 = with ground water influence

Simulate with or without winterkill: Winterkill
0 = without winterkill
1 = with winterkill

Select all crops to simulate: Crop
Enter in this order: CropNo, CropgroupNo, CropName, CropModel

Select Administrative region of interest. All available or a specific selected area. In the latter case enter NUTS_NAME from the NUTS table for the specific NUTS level:
SimCroAdminLevel1, SimCroAdminLevel2, SimCroAdminLevel3

[Aggregation]
This section includes the settings for the data aggregation.

First select schematization to use: Schematization
1 = Landcover (climate grid cell/SMU/land cover/admin. region)
2 = soil suitability (climate grid cell/SMU/admin. region)

Select Spatial level: SpatialLevel
1 = Administrative region
2 = Climate grid cells

Select indicator: AggIndicator
1 = Simulated crop yield
2 = FAO

3 = NDVI
4 = Ground weather
5 = Daily
6 = 10-Daily

Select Period in years: AggStartYear and AggEndYear

[Preparation]
This section includes the settings for the preparation of data for yield forecast calculation.

Select yield forecast analysis: Analyse
1 = Linear regression
2 = Scenario analyses
3 = User specified equation

Select Administrative region of interest. All available or a specific selected area. In the latter case enter NUTS_NAME from the NUTS table for the specific NUTS level: PrepAdminLevel1, PrepAdminLevel2, PrepAdminLevel3

Select Period: PrepStartYear and PrepEndYear

Select all crops to prepare: StatCrop
Enter in this order: StatCropName, StatCropNo, SimCropName, SimCropNo

Select the indicators. The number of indicators depends on the analysis:
Linear regression 4 indicators,
Scenario analyses max 10 indicators,
User specified equation 2 indicators.

For the category meteo enter is this order:
category, indicator name, period, begin, end, formula, threshold, weather

Category:
1 = Meteo
2 = Simulated crop
3 = Other

Indicator Name:
Group Meteo:
Tmin (minimum temperature)
Tmax (maximum temperature)
Tmean (mean temperature)
Precipitation (rainfall)
PET (potential evapotranspiration, ET0)
CWB (climatic water balance = rainfall minus potential evapotranspiration) Radiation (radiation)
Group simulated crop:
PYB (potential dry weight biomass)
WLYB (water limited dry weight biomass)
PYS (potential dry weight storage organs)
WLYS (water limited dry weight storage organs)

PLAI (potential leaf area index: leaf area divided by surface area)
WLAI (water limited leaf area index: leaf area divided by surface area)
Group other:
WSI (FAO Water Satisfaction Index)
NDVI (Normalised Difference Vegetation Index)
CNDVI (Corine-NDVI)

Period:
ABS = Absolute (decade) [1-36]
REL = Relative to development stage [0-200]

Begin and End indicate period as indicated before. It means decade or development stage.

Formula:
SUM = Sum
AVG = Average

Threshold: NoThreshold if no threshold value or else enter double value.

Weather:
Ground = Ground Weather
Model1 = Model Weather daily
Model10 = Model Weather 10-daily

For the categories Simulated crop and others just enter
category, indicator name, period, end (DVS value or decade)

In case of user specified analysis also enter the equation coefficients of the equation:

Y = a*X1^d + b*X2 + c

Enter EquationCoefs in this order: a, b, c, d

[Import]
This section includes the settings for the import of external data.

Select data type: DataType
1 = FAO, Water Index Model
2 = NDVI remote sensing indicators
3 = CNDVI remote sensing indicators
4 = Initial hydrologic conditions
5 = “Start” and “Stop” days
6 = Model weather daily
7 = Model weather 10-daily
8 = Model weather montly
9 = Crop Calendar

Enter number of lines of header (lines are skipped): HeaderLines
Enter name (full path) of the data file: ImpFileName

[Export]
This section includes the settings for the export of external data.

Select table to export. To select the table FORECASTED_NUTS_YIELD set ExpDecade to 0. To export table CROP_YIELD ExpDecade is decade to export.

Select year to export: ExpYear

Select crop to export: ExpCropNr. Enter Crop Number for export of CROP_YIELD or StatCropNumbers for the export of FORECASTED_NUTS_YIELD.

European Commission

EUR 22936 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: CGMS version 9.2 - User manual and technical documentation
Author(s): B. Baruth, G. Genovese, O. Leo
Luxembourg: Office for Official Publications of the European Communities
2007 – nnnn pp. – x cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-06995-6

Abstract.
The following report gives detailed information on the CGMS version 9.2 which is used in an operational context for the MARS Crop Yield Forecasting System estimating European yields or the major crops.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB- NA- 22936- EN- C

Name	Value
Publication Year	2007
EUR Number	22936 EN
ISSN	1018-5593
ISBN	978-92-79-06995-6
Catalogue Number	LB-NA-22936-EN-C
DOI	10.2788/37265
URL	
Publisher Name	OPOCE
Publisher City	Luxembourg
Publisher Country	Luxembourg

Publications Office
Publications.eu.int

Bar code