Global Earth Observation – Benefit Assessment: Now, Next, and Emerging

GEO-BENE global database for bio-physical modeling v. 1.0

(Concepts, methodologies and data)

Rastislav Skalský[1], Zuzana Tarasovičová[1], Juraj Balkovič[1] Erwin Schmid[2], Michael Fuchs[3], Elena Moltchanova[4], Georg Kindermann[5] & Peter Scholtz[1]

[1] Soil Science and Conservation Research institute, Bratislava, Slovakia
[2] University of Natural Resources and Applied Life Sciences, Vienna, Austria
[3] Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany
[4] National Public Health Institute, Helsinki, Finland
[5] International Institute for Applied System Analysis, Laxenburg, Austria

Summary:

Digital database presented here was compiled as a purposeful dataset which is to support GEO-BENE project activities targeted towards societal benefit assessment of improved information coming from data-model fusion approach applied in agriculture and ecosystem modeling. Version 1.0 of the global database is the first approximation to bio-physical model EPIC and agricultural and forestry sector optimization model FASOM requirements and a potential of available global observation data and data coming from other sources addressing climate, topography, soil, and crop management. Several hierarchically organized spatial reference objects were defined to geographically display the global database data. The global scale 5' spatial resolution grid covering land surface was created as primary spatial reference for geographical representation of all other spatial objects. Homogenous response units, 30' spatial resolution grid and country-level administrative units serve the secondary spatial reference as well as the basis for delineation of elemental spatial units (tertiary reference) for ultimate spatial geographical representation and organization of data in the global database. Global scale and publically available data sources used for global database compilation were described and classified into four groups (global observation data, digital maps, statistical and census data, and results of complex modeling) for the better understanding of the present situation in sources of the data available for global scale agro-ecosystem modeling. Data treatment and interpretation methodologies and the way of data harmonization are described so that solid metadata basis is available for the data publication and exchange. Global database comprises four thematic datasets addressing all global modeling aspects being under consideration: (i) land cover/land use statistics dataset, (ii) topography and soil data dataset, (iii) cropland management dataset, and (iv) climate dataset. Self-standing (v) spatial reference dataset provides a tool for geographical representing and visualization of the data. Data structure of all the datasets – tables and attributes – is also identified and described in the technical report.

Table of content

1. Introduction

Global database for bio-physical modeling (further referred as global database) described here has been created within 7th EC framework program project GEO-BENE (*Global Earth Observation – Benefit Assessment: Now, Next, and Emerging*, http://www.geo-bene.eu/).

Global database is primarily supposed to support the data-model fusion approach in agriculture sector modeling based on bio-physical model EPIC - *Environmental Policy Integrated Climate* (Williams et al., 1995) and the data on soil, topography, climate, land cover and land use available at global scale. Global database should serve also as a source of data on land cover/land use for land use optimization modeling with FASOM model - *Forest and Agricultural Sector Optimization Model* (Adams et al. 1996). Along with its main purpose, the global database is also a tool for identification of the gaps in availability of necessary global data for successful data-model fusion based interpretations in societal benefit area (SBA) agriculture (Justice et Becker-Reshef, 2007) and this way to support the GEO-BENE participation on a Group on Earth Observation (GEO) activities.

Data-model fusion based geographical modeling of landscape is not a straightforward task. Depending on particular circumstances, it requires more or less complex approach to gather all the necessary input information on landscape and natural or human-driven landscape processes which is consequently used to carry out appropriate basis for (bio-physical) model application and interpretation of the modeling results (c.f. Rossiter 2003). For global-scale modeling, however, it is not possible to gather directly measured input data in such a spatial, temporal and attribute detail which can fully satisfy the requirements of the (bio-physical) model.

This implies that current version of the global database (version 1.0) can be considered only a broad approximation to an optimal landscape model necessary for the fully successful application of (bio-physical) model and rather than real requirements of (bio-physical) model or complex landscape modeling it reflects a compromise between the needs and limits:

- **general modeling requirements** – (needs) - *a)* global-scale geographical landscape modeling, *b)* implementation of base-run and alternative scenarios for climate, land cover and (arable) land management into the modeling, *c)* (optional) clustering of EPIC and FASOM model within a complex landscape model (input/output data communication between individual models);
- **input data requirements of the models** – (needs) - *a)* quantitative data organized within global coverage of landscape units which are homogenous as for topography, soil, climate and management is essential to run the EPIC model (for more information on EPIC model data requirements see the APPENDIX 1), *b)* statistical data for selected statistical units (country-level administrative units further stratified by topography and soil) on area portion of land cover/land use categories and/or average values of environmental indicators coming from EPIC modeling are required by the FASOM model;
- **global data availability and quality** – (limits) - *a)* thematic data is of various origin: geographic data directly based on earth observations (interpreted/non-interpreted), digital thematic maps, census data and other geographical data coming from complex interpretations or modeling, *b)* data quality varies significantly across the available data: spatial, attribute and temporal resolution of the data, its thematic relevance and general accuracy and reliability, *c)* data accuracy and reliability may decrease after necessary

estimations of missing data based on existing data and/or expert knowledge is done, *d)* simplifications of general modeling concept due to the missing data without the possibility of its estimation from existing data sources is necessary;

- **data harmonization** – (limits) - common spatial reference and spatial resolution have to be set for the data of different source and quality and different way of the data spatial referencing (geo-referencing or geo-coding) to the common spatial reference may lead to loose of information.

2. Primary data identification – inputs for the global database

The list of the data sources given below is a result of internet and literature search conducted under the first stages of the GEO-BENE project. The aim of this was to gather the information on the best data available globally which could satisfy the model (and modeling) requirements (for more particular information on EPIC model data requirements see the APPENDIX 1). All of the data sources listed below were directly used for compiling the global database; identification and description of other data sources without such relation to global database would go beyond the scope of this report.

The final input data selection for models (and modeling) followed an assumption of the data sources used for global database should represent a sample of data available from public domains and any special contracts or other restrictions except the copyright or easy-to-get licenses are not needed to get and use the data. Most of data can be directly downloaded in digital form using on-line mapping and downloading services; some of the data have to be requested from the data authorities, whereas some others represent only a hardcopy publications.

Categorization of the data sources we introduce here should emphasize the differences in the way of the data acquisition (data gathering and data processing methods, mapping methods) for this could have significant influence on the quality and accuracy of the information coming from the data interpretations as well as it can better indicate what global earth observation data are up to hand for the SBA agriculture data-model fusion based modeling.

2.1. Global earth observations based digital data

This group represents the data coming directly from earth observation systems and earth observation data only slightly processed or interpreted. Data is originally in raster format with various spatial resolutions (e.g. space or air-born observations) or in form of to single or multi- point related measurements (e.g. weather station data).

2.1.1. Digital elevation data (SRTM, GTOPO30)

The high-resolution global Shuttle Radar Topography Mission digital elevation model (further referred as SRTM) derived by NASA (http://www2.jpl.nasa.gov/srtm/) was used as a source of global elevation data. SRTM digital elevation model is available in 3'' horizontal resolution (approximately 90 m at the equator) for areas between the latitudes from 60 N to 60 S, the altitude measure units are meters above a sea level.

Global 30 Arc Second Elevation Data (further referred as GTOPO30) (http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html) was used as a source of global elevation data. GTOPO30 is a global digital elevation model available in 30'' horizontal resolution (approximately 1 km at the equator); the altitude measure units are meters above a sea level. It was derived from several raster and vector sources of measured and pre-processed topographic information. GTOPO30, completed in late 1996, was developed over a three year period through a collaborative effort led by staff at the U.S. Geological Survey's EROS Data Center.

The Global Land Cover for year 2000 (http://www.gvm.jrc.it/glc2000/defaultGLC 2000.htm) produced as a common activity of several national and international institutions coordinated by JRC (further referred as GLC2000) was used as a basic source of land cover information at global scale. GLC2000 global raster is available in spatial resolution of approximately 32'' (approximately 1 km at the equator). Information on spatial distribution of 21 land cover classes interpreted from SPOT 4 VEGETATION 1 program satellite imagery (http://www.cnes.fr/web/1468-vegetation.php) using Land Cover Classification System of FAO (Di Gregorio et Jansen 2000) is available from GLC2000 legend (Table 2.1).

Table 2.1: *Legend to Global Land Cover 2000 (GLC2000)*

LAND COVER CLASS	CLASS DESCRIPTION
1	Tree Cover, broadleaved, evergreen
2	Tree Cover, broadleaved, deciduous, closed
3	Tree Cover, broadleaved, deciduous, open
4	Tree Cover, needle-leaved, evergreen
5	Tree Cover, needle-leaved, deciduous
6	Tree Cover, mixed leaf type
7	Tree Cover, regularly flooded, fresh water
8	Tree Cover, regularly flooded, saline water
9	Mosaic: Tree Cover / Other natural vegetation
10	Tree Cover, burnt
11	Shrub Cover, closed-open, evergreen
12	Shrub Cover, closed-open, deciduous
13	Herbaceous Cover, closed-open
14	Sparse herbaceous or sparse shrub cover
15	Regularly flooded shrub and/or herbaceous cover
16	Cultivated and managed areas
17	Mosaic: Cropland / Tree Cover / Other natural vegetation
18	Mosaic: Cropland / Shrub and/or grass cover
19	Bare Areas
20	Water Bodies
21	Snow and Ice
22	Artificial surfaces and associated areas
23	No data

2.1.3. Weather data (ECWMF)

The European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom provides an integrated weather forecasting system based on of processing manifold earth observation data (http://www.ecmwf.int/products/). Grid of 2.5°spatial resolution data on daily weather (further referred as ECWMF) was used to calculate monthly statistics required by EPIC weather generator.

2.2. Digital thematic maps

This group represents digital versions of hardcopy maps created by classical mapping methodologies or other digital thematic maps which has not been directly processed on the basis of global earth observations data. Data is originally available in vector format (polygon or polyline data).

2.2.1. Administrative regions (GAUL)

Global Administrative Regions Layer version 2007 () (http://www.fao.org/geonetwork /srv/en/metadata.show?id=12691&currTab=simple, further referred as GAUL) processed under the authority of the FAO and European Commission was used as a source of the country-level administrative regions data. Countries are in GAUL vector layer identified both by country name and country code which can be easily compared with official United Nations coding list of countries and world regions (http://unstats.un.org/unsd/methods/m49/ m49.htm).

2.2.2. Soil data (DSMW)

The digital version of the 1:5 000 000 scale Soil map of the world (further referred as DSMW) version 3.6 (http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116&curr Tab=simple) was used as a source of data on distribution of major soil units across the world. DSMW soil mapping units delineations (available both in vector or 5 arc minutes resolution raster) are attributed with information on soil mapping unit soil components (soil typological units and soil phases) and information on their area portion (%) of the soil mapping unit delineation. Totally, information on 106 soil typological units classified according to map legend (FAO-UNESCO 1974) and 5 miscellaneous non-soil units (glaciers, inland waters, dune and shifting sands, rock debris and outcrops, salt flats) can be retrieved from the map.

2.3. Census and other non-spatial data

This data group represents mostly the statistical data related (geo-coded) to statistical administrative units (country and first sub-country). Some of the data is a result of a slight interpretation or aggregation of original statistics. Data is originally digital attribute data organized in the tables; some data is in form of hardcopy publications (plain text, tables).

2.3.1. Agricultural statistics (FAOSTAT, FAOAQUASTAT, IFA)

FAO on-line data server (further referred as FAOSTAT) provides vide range of country specific agricultural statistics. FAOSTAT served a source of 1961 to 2006 years country-level statistics on crop harvested areas (http://faostat.fao.org/site/_567/default.aspx) and 2002 to 2005 years statistics on total consumption of nutrients (N, P_2O_5, K_2O) for selected countries (http://faostat.fao.org/site/575/default.aspx).

Website AQUASTAT maintained by FAO (further referred as FAOAQUASTAT) provides specific information on water management in agriculture, including irrigation areas statistics and water withdrawal by agriculture statistics. Country and crop specific irrigation calendar for the 90 countries of the world was used for global database as the source information on start and end days of crop planting and harvesting (http://www.fao.org /nr/water/aquastat/main/index.stm).

International Fertilizer Industry Association provides statistics on country and crop specific fertilizer consumptions (further referred as IFA). IFA dataset (http://www.fertilizer.org /ifa/statistics.asp) was used as a source of data on crop and country specific fertilizer application rates and total country-level average nutrients consumption (N, P_2O_5, K_2O).

2.3.2. Crop calendar data and documents (USDA, MARS, Crop and Country calendars)

Climate and crop calendar for the key producing regions and countries is available from U.S. Department of Agriculture document (USDA 1994, further referred as USDA). In the publication concentration zones of major crops for each country are identified and completed with information on historical averages of crop area (ha), yields (t/ha) and production (t). Coarse grain, winter and spring wheat, barley, rice, major oilseeds, sugar crops and cotton are included in USDA publication.

Crop calendars, agriculture practices calendars, crop rotations and other information on selected crops for the part of Europe (Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary, Slovenia, Romania, Bulgaria and Turkey) are available from publications of Kučera et Genovese 2004a, 2004b, and 2004c (further referred as MARS dataset). Wheat, barley, maize, rice, sugar beet, sunflower, soya bean, rape, potato, cotton and olive for Italy, Spain and Greece are given by Narciso et al. (1992). Some other sources of national crop calendar data used for global database are listed in APPENDIX 2.

Crop harvest calendars for sugar beet (FAO, 1959a), sugarcane (FAO, 1959b) and coffee (FAO, 1959c) include selected countries of the world. Rice crop calendar (http://www.irri.org/science/ricestat/) contains the data on planting and harvest dates (months) of rice for selected countries of the world. The crop calendar (planting and harvesting dates) of winter and spring wheat for Albania, Austria, Bulgaria, Czech Republic, Finland, Hungary, Norway, Poland, Romania, Slovak Republic, Sweden, Switzerland, Ex-Yugoslavia is available from publication of Russell et Wilson (1994). Autumn or winter and spring barley crop calendar (planting and harvesting dates) for Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain and United Kingdom is available in publication of Russell (1990)

The potato crop calendar (MacKerron 1992) includes dates of planting and harvesting (decade and month) for Germany, France, Italy, Netherlands, Belgium, United Kingdom, Ireland, Denmark, Greece, Spain and Portugal. Information about potato and sweet potato cultivation with emphasis on developing countries collects International Potato Centre. On their web-site are presented the potato crop calendar for countries of Africa, South and Middle America and Eurasia (http://research.cip.cgiar.org/confluence/ display/wpa/Home). The crop calendar of sweet potato is available for the countries of Africa and Asia (http://research.cip. cgiar.org/confluence/display/WSA/Home).

2.4. Interpreted data or modeling outputs

This group represents the data coming from various interpretations based on existing data and expert knowledge or data coming from modeling (interpolations results, statistical down-scaling results, etc.). Such a data represents secondary data not directly measured or collected. The interpretation methodologies used for data creation influence significantly the data accuracy and reliability. Various digital geo-referenced or geo-coded data available in this group have various spatial resolutions and is available in raster or vector representation depending on the data origin.

2.4.1. Soil data (WISE)

International Soil Reference and Information Centre (ISRIC) 5 by 5' grid of soil properties estimation based on global soil distribution (DSMW) and soil profile data (WISE soil profile database, http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/Global+soil+profile+data.htm) interpretation (further referred as WISE, http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/WISE5by5minutes.htm) was used as a source of DSMW soil typological unit specific data on soil analytical properties for 5 depth intervals of soil profile (20 cm intervals for total depth of 1m). Detailed interpretation methodology for WISE compilation is described in publication of Batjes (2006).

2.4.2. Land use and land management data (GLU, Manure data)

Global crop distribution data processed by International Food Policy Research Institute was used as a source of basic information on land cover and land use. The data (further referred as GLU) represents a global coverage of a regular-shaped statistical units (5' spatial resolution grid) attributed with estimated crop cultivation and harvest areas (physical area in ha) and crop production (tons/ha) for 20 globally most important crops or crop groups (wheat, rice, maize, barley, millet, sorghum, potatoes, sweat potatoes and yams, cassava, bananas and plantains, soybean, other pulses, sugar cane, sugar beet, coffee, cotton, other fiber crops, groundnuts, other oil crops). Estimations are done separately for the four agricultural production systems (high input - irrigated, high input - rainfed, low input - rainfed and subsistence management systems). Final dataset resulted from downscaling of many national and sub-national agricultural census data using additional spatial information on land cover (GLC2000), crop suitability and management system potential yields (agro-ecological zones assessment data, Fischer et al. 2000), irrigation areas (Global irrigation map, Siebert et al, 2007) and population density data (Gridded Population of the World, http://sedac.ciesin.columbia.edu/gpw/index.jsp). All the methodological details are given in the publication of You et Wood (2006). Actually, GLU dataset is not available for download by any public on-line service. For global database GLU data set was provided by IFPRI after personal communication with Dr. Liangzhi You.

Global scale 5' spatial resolution on nitrogen rates coming from manure application (further referred as manure data) compiled by Liu et al. (in press) were used as a source dataset on nitrogen fertilization and total manure application. Manure data was interpreted from global data on livestock density (Gridded Livestock of the World, http://www.fao.org/ag/AGAinfo/resources/en/glw/default.html) developed by FAO in collaboration with the Environmental Research Group Oxford and FAOSTAT data on country and livestock specific data on slaughter weights using published knowledge on nitrogen rates in livestock excretion and amounts of manure recycled to cropland. Actually, manure dataset is not available for download by any public on-line service. For global database the dataset was provided after personal communication with Dr. Junguo Liu.

2.4.3. Weather data (Tyndall)

The Tyndall Centre for Climate Change Research of University of East Anglia, Norwich, UK provides data on historical time series of global weather for the period from 1901 – 2000 and 16 climate change scenarios (further referred as Tyndall, http://www.cru.uea.ac.uk/~timm/grid/TYN_SC_2_0.html). Global scale land surface coverage 0.5°spatial resolution grids of modeled climate data on cloud cover, diurnal temperature range,

precipitation, temperature, vapor pressure were compiled from several interpolated global and regional climate datasets. Methodological details of Tyndall data are described in publication of Mitchel et al. (2004).

3. Reference for geographical representation of the data

Three-level hierarchical system of geographical objects was designed for spatial referencing the global database data on natural and management landscape characteristics. It comprises primary geographical reference grid (chapter 3.1), secondary geographical reference objects (HRU, 30' spatial resolution grid and country-level administrative unit, chapter 3.2) and tertiary reference objects (landscape units, chapter 3.3). Database reference objects (simulation units, chapter 3.3) provide the least non-spatial data reference in the global database.

3.1. Primary geographical reference grid

Primary geographic reference is the basic spatial frame for geographical representation and visualization of all the data stored in global database. It is a basis for spatial referencing of secondary and tertiary reference objects. Moreover, primary geographic reference is the basic harmonization tool for the geographical data coming from various sources.

3.1.1. 5' spatial resolution geographical grid (Global grid)

Global extent (-90° S to 90° N and -180° W to 180° E) grid designed in geographical coordinate system WGS 84 with pixel resolution of 5' (about 10 X 10 km on equator) was created to serve the primary geographical reference for the global database (further referred as global grid). Land surface mask was used to keep only those pixels of global grid which represent land surface (any other pixels representing oceans are not relevant for global modeling). Land surface mask resulted from intersection of the land surface subsets of the most significant geographical data inputs (GTOPO30, SRTM, DSMW, GLC2000, GAUL, IFPRI-GLU) harmonized in global grid (chapter 3.2.). Decision was done also to exclude Antarctica from global grid due to the most of its area is covered by ice or bare rocks. Totally, global grid comprises 2.186.775 pixels.

Each grid cell of the global grid is indexed by column and row number counted from upper left corner of the grid (point coordinates 90° N, -180° W). Column numbers ranges from 1 to 4320, row numbers ranges from 1 to 2160. Independently from column-row indexing each pixel is indexed also by x and y coordinate (decimal degree) of pixel centroid. Point lattice can be then displayed via centroid coordinates to visualize global grid in a map. This makes global grid independent of a particular GIS platform used for compilation and storage of the data as well as it supports the exchange of geospatial data between various GIS or database systems via simple and highly interoperable ASCII text files (e.g. *txt*, *csv*). Classical database tools can be used for storage, maintenance and analyzing the data geo-referenced to point lattice and GIS support is needed only for displaying the data geographically in a map.

Modeling requires that all the data on area of spatial units is reported as real area values (e.g. ha or km^2). Real area (ha) for each pixel of global grid was calculated as a 1/36 of real area of corresponding 30' resolution grid pixel (chapter 3.2.2). Real area of 30' resolution grid resulted from transformation of the geographically not projected data (WGS84 coordinate system) into the equal area geographic projection system (Goode-Homolosine projection) following the routine described by Lethcoe and Klaver (1998).

3.2. Secondary geographic reference

Secondary geographic reference objects provide spatial reference for all the data direct referencing of which to the global grid is not advantageous or possible (chapter 3.2.2, 3.2.3) Secondary reference object also serve as data stratification tool in bio-physical and optimization models communication (chapter 3.2.1). Along with referencing function secondary spatial reference objects play important role in the delineation of landscape units (chapter 3.3., Fig. 3.2).

Particular secondary geographic reference object is spatially referenced to global grid as an attribute value assigned to relevant pixel or group of pixels. It can be geographically displayed on a map as a spatial zone of global grid (Fig. 3.1).

Fig. 3.1: *Secondary geographical reference object zone (e.g. HRU) visualized via global grid.*
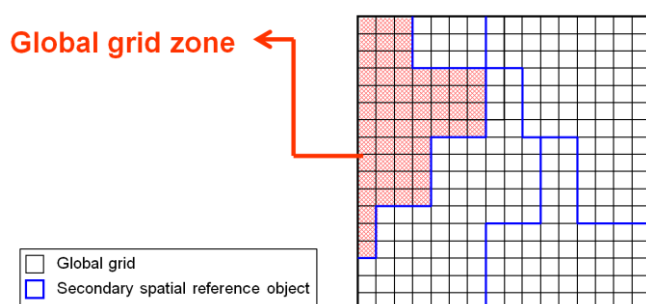
3.2.1. Homogenous response units (HRU)

Concept of homogenous response units (HRU) used here was adopted after slight modification from earlier works (Schmid et al. 2006, Balkovič et al. 2006, Stolbovoy et al. 2007) as a general concept for delineation of basic spatial units. Only those characteristics of landscape, which are relatively stable over time (even under climate change) and hardly adjustable by farmers, were selected. HRU is a basic spatial frame for implementation of climate-change and land management alternative scenarios into global modeling and therefore it is one of basic inputs for delineation of landscape units (chapter 3.4, Fig. 3.4). Moreover, HRU provides a possible interface for communication of bio-physical and optimization models (EPIC model derived and consecutively HRU level aggregated information on environmental indicators can input FASOM optimization modeling).

Tab. 3.2: *Altitude, slope and soil class criteria for HRU delineation*

LAND CHAR.	UNIT	CLASS (CLASS INTERVAL)
altitude	meters	**1** *(0 – 300)*, **2** *(300 – 600)*, **3** *(600 – 1100)*, **4** *(1100 – 2500)*, **5** *(> 2500)*,
slope inclination	degree	**1** *(0 – 3)*, **2** *(3 – 6)*, **3** *(6 – 10)*, **4** *(10 – 15)*, **5** *(15 – 30)*, **6** *(30 – 50)*, **7** *(> 50)*
soil	-	**1** *(sandy)*, **2** *(loamy)*, **3** *(clay)*, **4** *(stony)*, **5** *(peat)*, **88** *(no-soil)*

HRU is spatially delineated as a zone of global grid having same class of altitude, slope and soil (HRU class definitions are listed in Tab. 3.2.):

- Dominant altitude class was calculated by raster algebra as a zonal majority value of pre-classified GTOPO30 raster altitude class over a one global grid pixel area;

- Dominant slope class was calculated by raster algebra as a zonal majority value of pre-classified 30'' spatial resolution temporary raster slope class over a one global grid pixel area. Temporary raster used for calculations was interpreted from original SRTM and GTOPO30 data as follows. The SRTM data calculated slopes at 3'' spatial resolution were grouped into the classes 0°- 3°, 3°-6°, 6°-10°, 10°-15°, 15°-30°, 30°-50° and >50°. For the 30'' resolution raster zonal majority procedure was done to get the 60 N to 60 S extent raster of slope classes. To fill up the missing regions from 60°N to 90°N and 60°S to 90°S a slope raster with the GTOPO30 was calculated. The region 60°N to 60°S was covered by both SRTM and GTOPO30 derived slope data. This overlapping region was used to create a look-up table which allowed transforming the slope from the GTOPO30 to the slope class shares of the SRTM and fill up the missing regions.

- Dominant soil class represents most frequent soil class of DSMW soil mapping unit (as for its relative area) assigned to global grid pixel by intersection (spatial join) of global grid centroid lattice and original DSMW layer. Soil typological units of the particular DSMW soil mapping unit were classified into five pre-defined soil classes. Based on WISE soil profile data on aggregated soil texture classes (coarse, medium and heavy texture) sandy, loamy and clay soil classes were interpreted; soil typological units classification was applied for stony and peat soil classes interpretation. An arbitrary value of 88 was assigned to all non-soil bodies. Sum of the areas of all soil typological units classified to the same soil class or sum of areas of all non-soil bodies having dominant area portion of the total DSMW mapping unit area was then applied as a criterion for global grid pixel dominant soil class.

Fig. 3.2: *Global HRU coverage.*
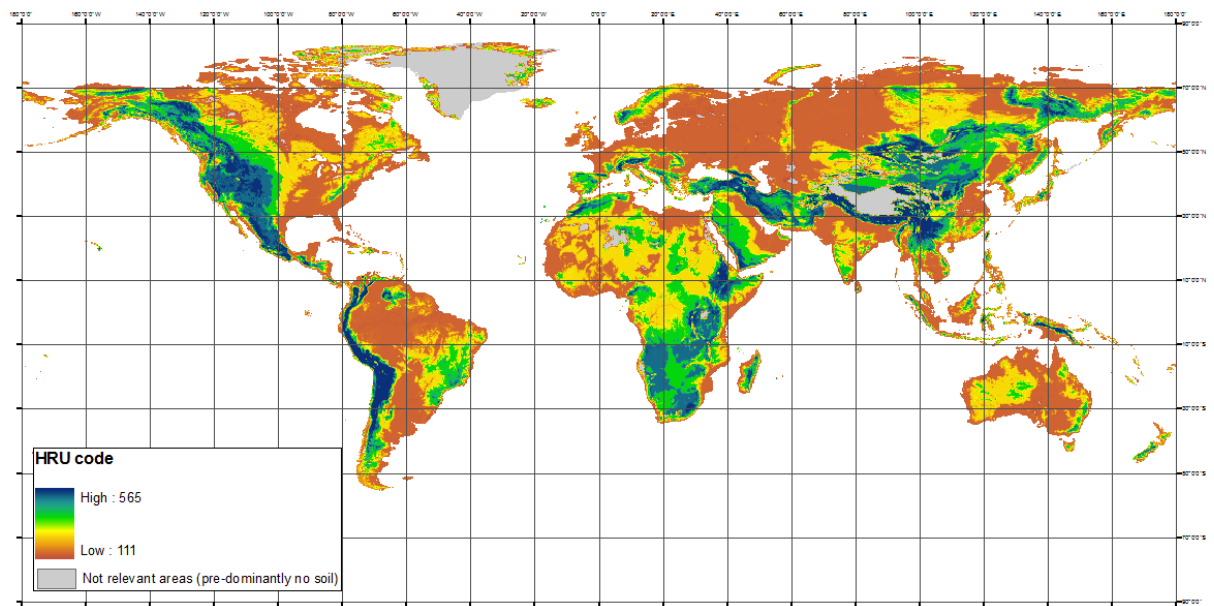
Altitude, slope and soil class value assigned to 5' spatial resolution pixel represents spatially most frequent class value (not average!) taken from input data of higher spatial (GTOPO30, SRTM) or attribute (DSMW) resolution than target dataset (i.e. idea of "the most likely" natural conditions is adopted here). This implies that not absolute information on landscape quality and variability over the 5' spatial resolution pixel area is transferred to global grid in data harmonization process and resulting harmonized information used for HRU delineation is just broad approximation to real variability of the global landscapes.

Totally, 150 unique combinations of altitude, slope and soil class resulted from HRU delineation process (Fig. 3.2). Each delineated HRU zone is indexed by numerical code assembled from code of altitude, slope and soil on first, second and third position, respectively.

3.2.2. 30' spatial resolution geographical grid (PX30)

Overlay of global grid and global extent (-90° S to 90° N and -180° W to 180° E) 30' spatial resolution grid (about 50 X 50 km on equator) designed in geographical coordinate system WGS 84 was done to get regularly-shaped zones of global grid indexed by column and row index of underlying 30' spatial resolution grid (column and row indexes counted from upper left corner, coordinates 90° N, -180° W). Each zone was also assigned with *x* and *y* coordinate values of corresponding 30' resolution pixel centroid.

Fig. 3.3: *Real area change with increasing latitude*
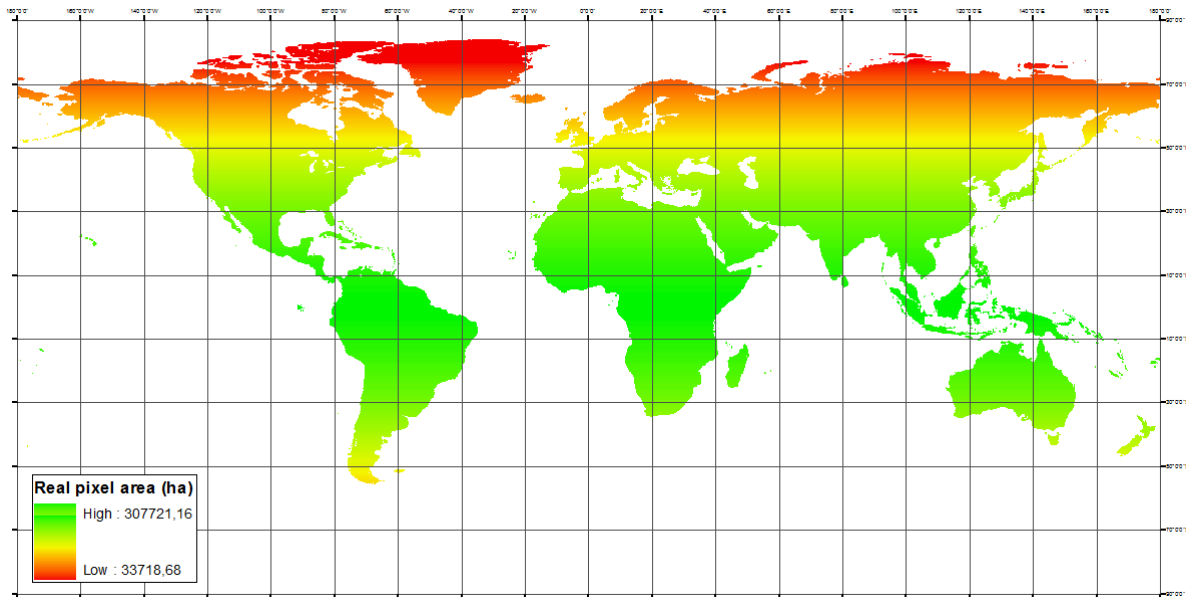
Real area (ha) of 30' resolution grid was calculated after transformation of the geographically not projected data (WGS84 coordinate system) into the equal area geographic projection system (Goode-Homolosine projection) following the routine described by Lethcoe and Klaver (1998). Real area (ha) change with latitude is shown on a map (Fig. 3.3).

30' resolution grid zone (further referred as PX30) provide a direct geographical reference for interpolated weather data (Tyndall) as well as it is an arbitrary spatial frame which is to secure local detail in geographical analyses and interpretation of the input data for bio-physical modeling (topography and soil data, land cover and land use statistics).

3.2.3. Administrative units (COUNTRY)

Simple overlay (spatial join) of GAUL data and the global grid centroid lattice was done to get country and first sub-country level administrative unit identification (country code) for all pixels of global grid. Consequently, original country-level administrative unit codes taken from original GAUL dataset were replaced by a country code from official UN country list. UN country codes from ad-hoc constructed look-up table (Tab. 3.1) were

14

assigned to all administrative regions having no country information in GAUL dataset (geopolitically controversial administrative regions). Original GAUL coding was kept only for sub-country-level administrative regions.

Country level administrative regions (further referred as COUNTRY) provide a spatial reference for geo-coding the most of national-level agricultural census data and national or national level data on crop management calendars. Sub-country-level administrative regions provide an additional spatial reference for geo-coding the sub-national crop management calendar data for selected countries (chapter 4.3.3).

Tab. 3.1. *UN country codes for geopolitically controversial administrative regions.*

GAUL REGION (COUNTRY	UN COUNTRY CODE (COUNTRY NAME)
Aksai Chin	356 (India)
Arunashal Pradesh	356 (India)
Dhekelia and Akrotiri SBA	196 (Cyprus)
Gaza Strip	376 (Israel)
Hala'ib triangle	818 (Egypt)
China/India	156 (China)
Ilemi triangle	736 (Sudan)
Jammu Kashmir	356 (India)
Kuril islands	643 (Russian Federation)
Madeira Islands	620 (Portugal)
Ma'tan al-Sarra	736 (Sudan)
West Bank	376 (Israel)

3.3. Landscape units and simulation units

Landscape units and simulation units provide immediate reference for the global database data displaying in geographical or classical database. Simple scheme illustrating the definition and position of these units in global database is given in Fig. 3.4.

Fig. 3.4: *Landscape zones and simulation units – definition, functional position and referenced data*

15
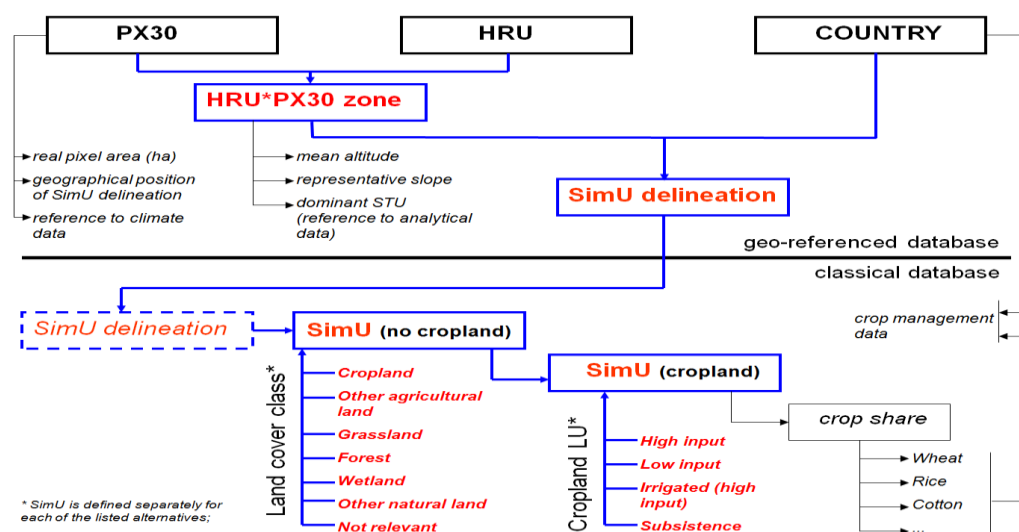
3.3.1. Landscape units (HRU*PX30 zone, SimU delinetion)

Landscape units were implemented to satisfy the bio-physical model requirements for spatial units being homogenous as for its natural conditions and landscape management. Landscape units are the smallest spatial units coming from intersect of HRU, PX30 and COUNTRY spatial delineations. The function of landscape units in global modeling is to provide both the immediate reference to the data required as inputs for bio-physical model (Fig. 3.4) and to provide the least possible spatial delineation (which can be further specified only on non-spatial level, chapter 3.3.2). Two hierarchical levels of landscape units are assumed in global modeling:

- **HRU*PX30** zone (Fig. 3.4.) is delineated as the smallest landscape body homogenous as for its natural conditions (topography, soil and weather). HRU*PX30 zone provides direct spatial reference for EPIC input data on soil and topography (average altitude, representative slope and soil typological unit specific analytical values);
- **Simulation unit delineation** (further referred as SimU delineation, Fig. 3.4, Fig. 3.5) is the HRU*PX30*COUNTRY zone delineated as the smallest landscape body homogenous as for its natural conditions (the same as above) but with further reference to a particular country-level administrative region. SimU delineation provides direct spatial reference for non spatial information on land cover and land use. Maximal area of SimU delineation is equal to an area of one 30' spatial resolution grid pixel and its area decrease from about 300 000 ha on equator to about 30 000 ha in high latitudes; minimal area of SimU delineation is equal to an area of one 5' spatial resolution grid pixel and its value decrease from about 8 500 ha on equator to about 950 ha in high latitudes.

Fig. 3.5: *SimU delineation visualized as a global grid zone.*
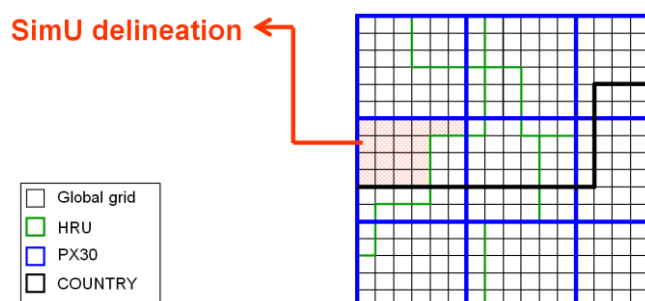
3.3.2. Simulation units (SimU)

Simulation unit (further referred as SimU) is non-spatial semantic unit which represents one of all possible land cover, land use (for cropland) and via this two also land management alternatives which can take place in the area defined by SimU delineation (Fig. 3.4.). In the global database SimU represent a unit which bears definite information on unique natural and management condition existing at particular place. This information is utilized in bio-physical modeling to assemble the set of management alternative specific input files for model runs. SimU provides also a reference for interpretation and displaying the bio-physical modeling outputs.

4. Input data interpretation for global database

4.1. Land cover and land use data

4.1.1. Land cover statistics

Combined GLC2000 and GLU datasets interpretation was done to get global grid and SimU delineation specific data on cropland area necessary for setting up the base-run scenarios for agro-ecosystem modeling (information absent in GLC2000) and data on other land cover classes area (grasslands, forests, wetlands, other natural vegetation) necessary for land cover/land use optimization modeling (information absent in GLU).

Table 4.1: *GLC2000 classes derived global database land cover classes*

GLOBAL DATABASE LAND COVER CLASS	ORIGINAL GLC2000 CLASSESS (Tab. 2.1)
total agricultural land	16,17,18
grassland	13,
forest	1, 2, 3, 4, 5, 6, 9, 10
wetlands	7, 8, 15
other natural vegetation	11, 12, 14
not relevant land covers	19, 20, 21, 22

Zonal analyze of pre-classified GLC2000 layer (Tab. 4.1.) over the global grid was done to get pixel specific statistics on total cropland, grassland, forest, wetland, other natural vegetation and not relevant land cover classes real areas (ha). Pixel specific cropland area (ha) was calculated from original GLU data by summing up all particular crop areas over all assumed management systems. Consistence of GLU and GLC2000 data was checked by subtracting GLU derived cropland area from GLC2000 derived total cropland area. Result of the consistence check is shown on a map (Fig. 4.1).

Fig. 4.1: *GLU and GLC2000 land cover data inconsistence.*

cropland where GLU cropland area is more than GLC total agricultural land area

cropland where GLU cropland area is more than global grid pixel area

other cropland where GLU area is equal or less than GLC2000 area

17
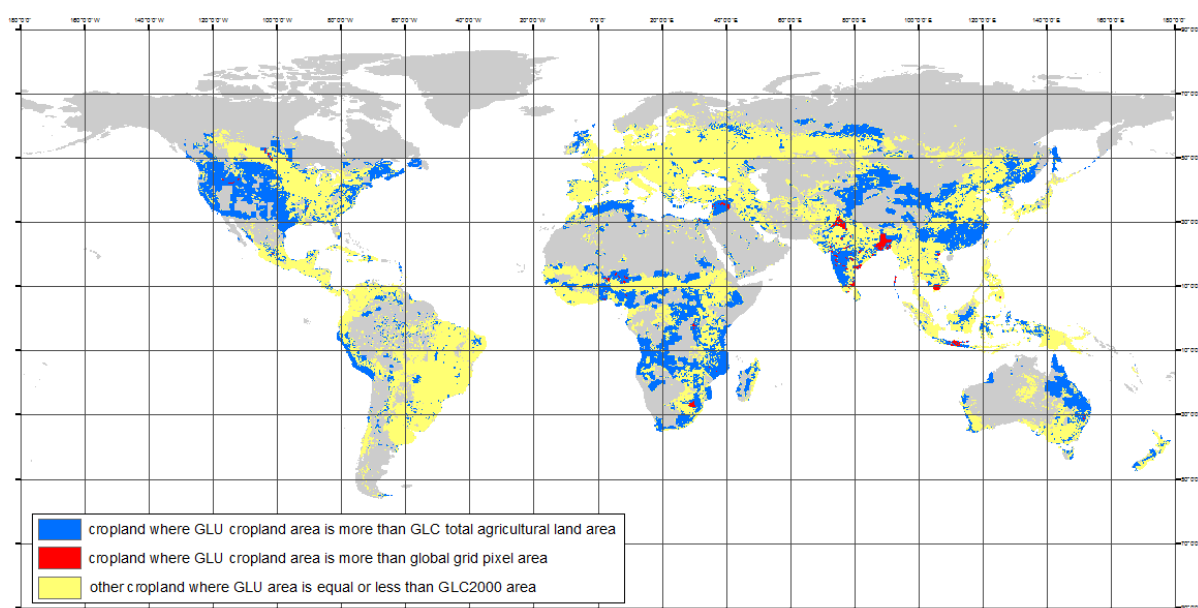
There is lot of pixels having GLU cropland area higher than GLC2000 total agriculture area; even pixels having higher GLU cropland area than total pixel area were identified in dataset. Reason for that resides in an inconsistence of GLC2000 and national agricultural census data (more agricultural land is reported in census than it is available in GLC2000) as well as in GLU model total area of all pixels was arbitrary set to 9000 ha which have could caused some of the inconsistencies (personal communication with Dr. Liangzhi You).

It was decided that GLU cropland area should be secured in final harmonized land cover dataset (cropland land cover class) with only exception in the case of GLU cropland area is higher than total pixel area; and data on all other land cover classes were adjusted following several simple rules:

- if the GLU derived cropland area was less than GLC2000 derived total agricultural land area an additional land cover class (other agricultural land) was introduced and assigned with area value calculated as a difference of GLC2000 total agricultural land area and GLU cropland area, GLC2000 derived areas of all other land cover classes were kept;
- if the GLU derived cropland area was more than GLC2000 derived total agricultural land area additional land cover class (other agricultural land) was given by zero value and areas of all other land cover classes were recalculated based on their area portion of global grid pixel so that resulting sum of their areas was equal to difference of total pixel area and GLU reported cropland area;
- if the GLU derived cropland area was more than total pixel area; cropland area value was arbitrary set to total pixel area and areas of all other land cover classes were set to zero.

For each pixel of global grid a metadata field was created to indicate the method used for land cover data harmonization. Finally, harmonized land cover dataset was spatially aggregated by zonal sum of all land cover classes over the SimU delineation zone to get SimU delineation specific data on land cover.

4.1.2. Land use statistics

Information on cropland land use classes (chapter 2.4.2) is available from GLU dataset. Pixel specific cropland management system area (ha) was calculated from original GLU data by summing up all crop areas over particular management system. Additional area adjustments based on management system area portion of total cropland were done in case of GLU dataset derived specific cropland area was different from cropland area of harmonized land cover dataset. Finally, spatial aggregation of land use dataset was done by zonal sum of all land use classes over the SimU delineation zones to get SimU delineation specific data on land use.

4.1.3. Land cover/land use relevance and visulisation mask

An arbitrary 5% threshold of land cover area portion of total SimU delineation area was set to tell whether SimU delineation is relevant or not for simulation of the defined land cover alternative (i.e. which SimU are to be defined across the SimU delineation area). Similarly, a relevance of cropland land use alternative for simulation was set based on 5%

threshold of cropland management system area portion of total cropland area. Relevance masks were set for to limit bio-physical model runs only to most representative management alternatives across the SimU delineation area. Relevance masks are not mutually exclusive and in a case it meets the threshold more than one land cover/land use alternative can be assumed for the SimU delineation. Absolute SimU delineation relevance mask based on spatial dominance of particular cropland management system across the cropland was in addition set only for cropland land use alternatives (chapter 4.3.2).

Fig. 4.1: V*isualization mask for land cover (land use) data*
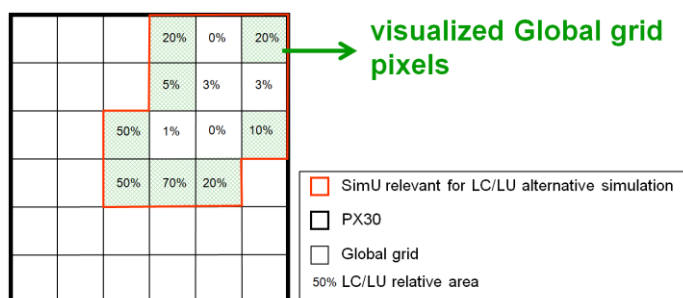
SimU delineation is the least spatial object which can be visualized on the map and all related information on land cover/land use is organized only as a non-spatial statistics (chapter 3.4.). To overcome this problem, for visualization of the SimU specific data (input/output data for bio-physical model, management alternatives spatial pattern, etc.) a visualization mask was defined for each of the global grid pixels.

If a particular land cover/land use alternative is considered for visualization, only those pixels of global grid are visualized which represents relevant SimU delineation zone and which meet the 5% threshold of minimum land cover area portion of total pixel area or land use area portion of total cropland area in particular pixel (Fig. 4.1, Fig. 4.2).

Fig. 4.2: *Land cover/land use* v*isualization mask – example of irrigated cropland management system.*
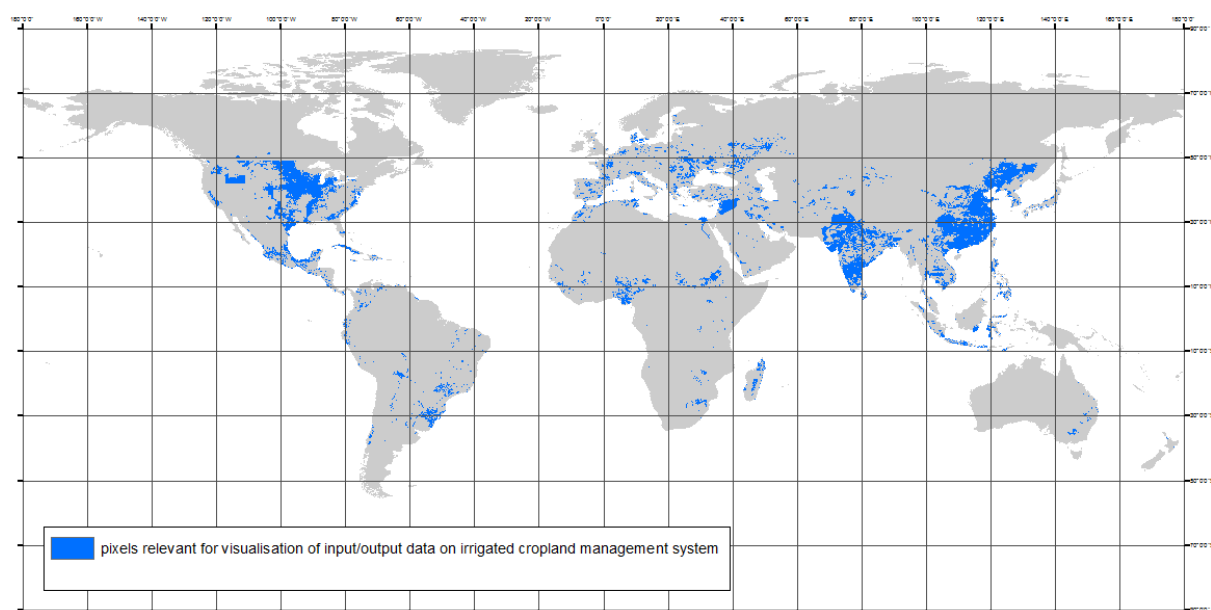
4.2. Topography and soil data

4.2.1. Topography data

Zonal mean of altitude value (m) coming from GTOPO30 was calculated over the HRU*PX30 zone to get average altitude as an input for bio-physical model (Fig. 4.3.).

Fig. 4.3: *Global pattern of HRU*PX30 zone average altitude.*
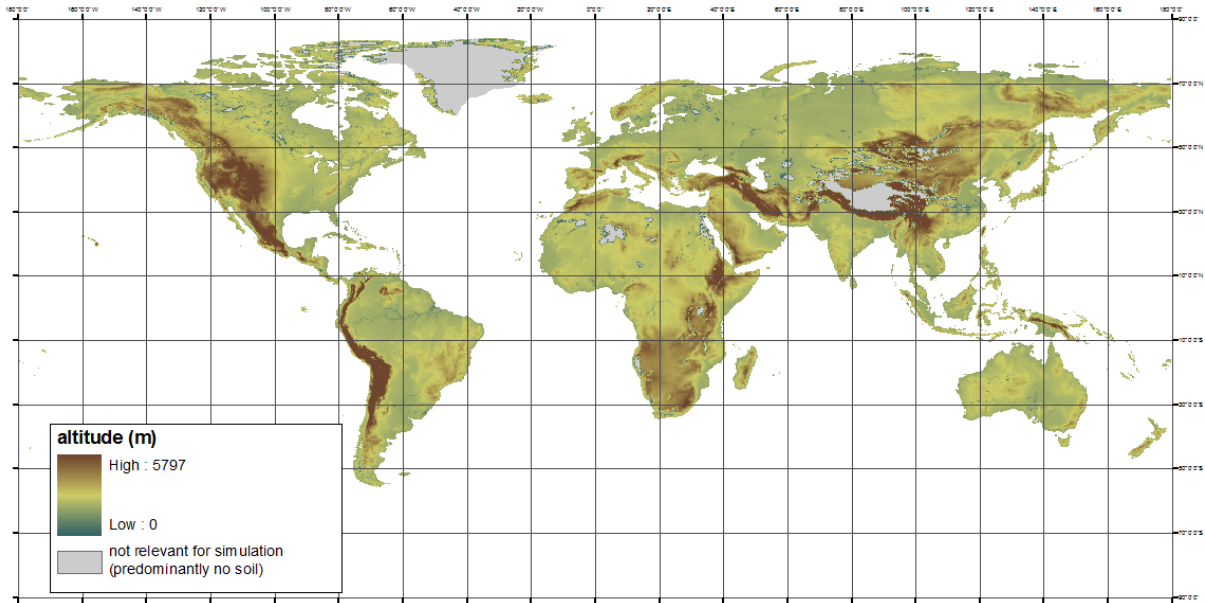
Fig. 4.4: *Global pattern of HRU*PX30 zone representative slope.*
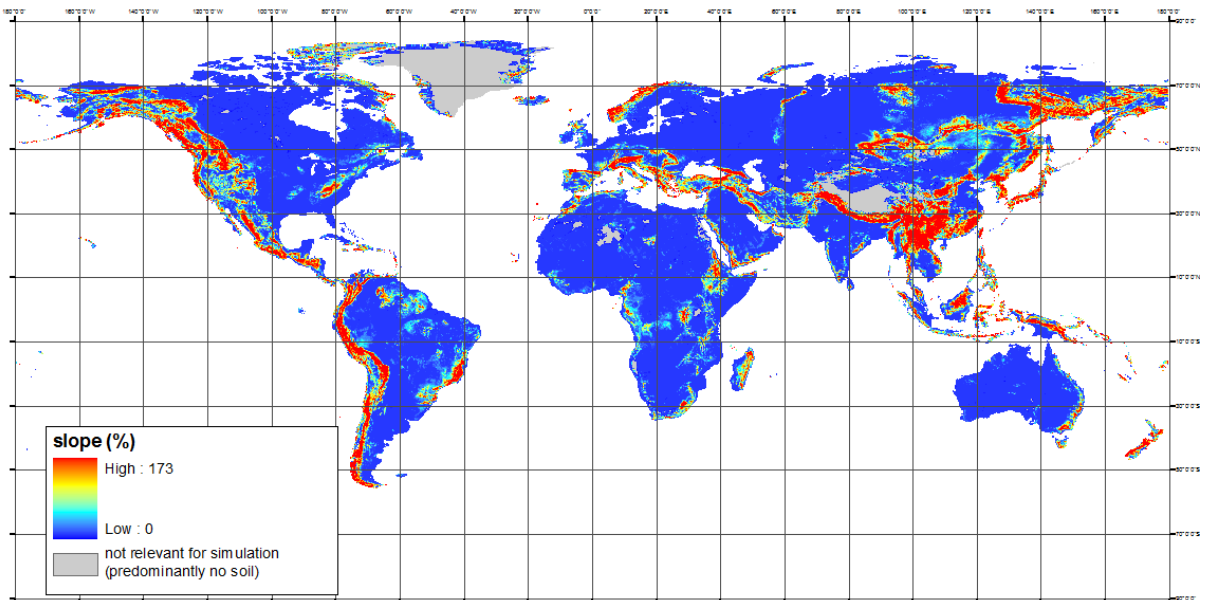
Representative slope value for slope class coming from HRU (Tab. 4.2) was assigned to HRU*PX30 zone after recalculation to percent by simple equation (*slope in % = tg(slope in deg) * 100*) to get slope value required by bio-physical model (Fig. 4.4). Dominant slope

class interval method derived representative value seems to be more appropriate as an average value for the averaging of slope over the HRU*PX30 zone could result in non-realistic (artificial) values in a case of wide range of slopes is presented within a zone.

Table4.2: *Representative slope values assigned to the slope classes*

HRU SLOPE CLASS INTERVAL	REPRESENTATIVE SLOPE VALUE (DEG)
0 - 3	1
3 - 6	4
6 - 10	8
10 - 15	12
15 - 30	23
30 - 50	40
> 50	60

PX30 pixel centroid *x* and *y* coordinate values were assigned to all corresponding HRU*PX30 zones as a representative value of landscape unit geographical position required by bio-physical model. This simplification should not affect negatively the simulations.

4.2.2. Analytical characteristics of soil

Data on analytical characteristics of soil required by bio-physical model are to HRU*PX30 zone related via information on dominant soil typological unit. A slightly modified methodology firstly introduced by Batjes at al. (1995) for DSMW data interpretation was applied to get HRU*PX30 dominant soil typological unit from DSMW data.

Fig. 4.5: *Global pattern of HRU*PX30 zone dominant soil typological units (soils according to FAO-UNESCO 1974), RK1, RK2, DS, and ST stands for no soil bodies, NA for non-identified soil bodies.*

Area percentage of each DSMW soil mapping unit identified within the particular HRU*PX30 zone extent was calculated in the first step using the soil mapping unit data obtained from from intersect of global grid centroid lattice and original DSMW layer. In the next step, soil typological unit specific relative areas (%) calculated from data on DSMW soil mapping unit area percentage and data on DSMW soil mapping unit composition (chapter 2.2.2) were summed over the HRU*PX30 zone to get total area portion (%) for each of the soil typological units. Soil typological unit of the highest area portion was selected as dominant one for HRU*PX30 zone (Fig. 4.5).

Dominant soil typological unit relative area varies from 14 to 100% across the global coverage of HRU*PX30 zones (Fig. 4.6.). This implies that in areas with high diversity of soils (the case of dominant soil typological unit area is less than 60%) dominant soil typological unit based concept cannot be considered more than just a broad approximation to the appropriate model of soil variability.

Fig. 4.6: *Area percentage of dominant soil typological unit of HRU*PX30 zone.*
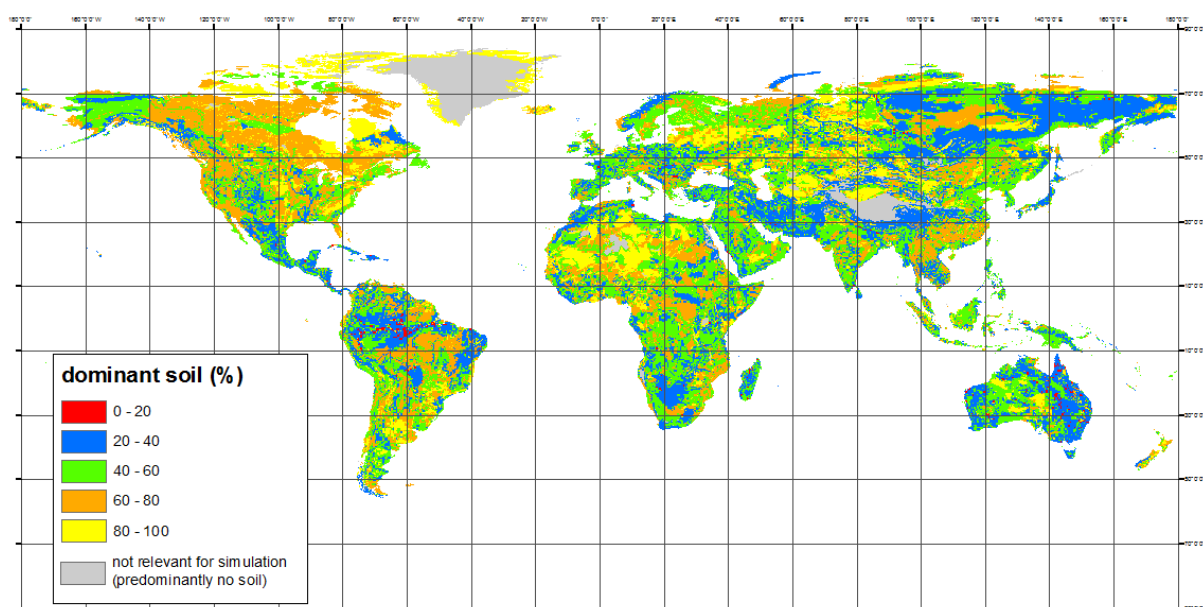
The most of the soil analytical data required by bio-physical model was adopted directly from WISE dataset without or with only slight modifications such as transformation of measurement units or re-calculation of relative to absolute values. Some other soil analytical parameters (hydrological soil group and soil hydrological parameters, albedo of moist soil) missing in WISE dataset were estimated from WISE data by available methodologies or published knowledge.

Hydrological soil group was interpreted for each soil typological unit from WISE analytical data on soil texture by simplified rules from official USDA-NRCS engineering manual (USDA-NRCS 2007). Soil typological unit specific hydro-physical characteristics were estimated from WISE analytical data on soil texture and bulk density by combined methodology. Neural-network algorithm based pedo-transfer model Rosetta (Shaap et Bouten 1996) produced saturated hydraulic conductivity value and soil hydrological parameters necessary for calculation of soil water content at field water capacity and wilting point by Van Genuchten equation (Wösten et Van Genuchten 1988). An arbitrary value of 0,15 was assigned to all soil typological units as albedo of moist soil representing median of the 0.1 – 0.2 albedo interval for dark soil surfaces given by Dobos (2006).

All salt flats and shifting sands no-soil bodies (ST, DS) and all non identified soil bodies (NA) which resulted from HRU*PX30 zone dominant soil typological unit analyze were assigned by analytical values of other available soil typological units after the WISE dataset rules (Batjes 2006). In a case of rock non-soil bodies (RK1, RK2) were identified as dominant soil typological unit, analytical values of rankers or cambisols were given to stony (RK2) or all other (RK1) HRU soil classes, respectively.

Analytical data on organic soils (histosols (O), dystric histosols (O), eutric histosols (O, and gelic histosols (Ox)) are not available in WISE. Soil analytical parameters for organic soils were estimated from the soil horizon data on three histosol soil profiles (Od) available in global soil profile database compiled by ISRIC (Batjes 1995), WOFOST model default soil input files for Europe (Boogaart et al. 1998) and published reference information on soils of the world (FAO-UNESCO 1974, Dreissen at al. 2001) as follows.

Organic soils were assumed as being homogenous over the soil profile and therefore the same analytical values were given to all soil profile depth intervals considered in WISE dataset (five intervals to depth of 1 m). Median of soil albedo interval for dark soil surfaces (0.15) given by Dobos (2006) was set for soil albedo parameter. Arbitrary value (0) was set to volume of stones and calcium carbonate content soil parameters. For the presence of ground water table close to the soil surface in most of organic soils (Driessen et al. 2001) and relatively low value of saturated hydraulic conductivity taken from WOFOST input files soil hydrological group was interpreted as D. Data on representative textural class of organic soils was not available and it was arbitrary set for all organic soils to 30, 40 and 30 for sand, silt and clay fraction, respectively, so that any soil textural extreme is eliminated from simulations. Upper limit of bulk density interval for fibric histosols (0.15 t/m^3) published by Driessen et al. (2001) was given to all organic soils. Cation exchange capacity was calculated as an average across the global soil profile database (Batjes 1995) and set for all organic soils (95 cmol/kg). Base saturation was calculated as an average from global soil profile database (25%) and together with cation exchange capacity value it was used for calculation of base saturation for O, Od and Ox (23.75 cmol/kg). Representative value (75%) of eutric qualifier base saturation interval (50-100%) was taken from DSMW legend (FAO-UNESCO 1974) and used in calculation of base saturation for Oe (71.25 cmol/kg). Average soil pH (4.5) calculated from global soil database (Batjes 1995) was given to O, Od and Ox organic soils. Following both the representative soil pH for eutric histosols given by Driessen at al. (2001) and FAO-UNESCO (1974) eutric histosols class definition soil pH parameter was set to 6.5 for Oe. Soil organic carbon was set to 35 % for all organic soils following the average calculated from global soil profile database (Batjes 1995) and characteristics of histosols given by Driessen et al. (2001). From WOFOST model default soil input files for Europe – EC6 (Boogaart et al. 1998) representative values of 0.52, 0.13, and 5.62 were taken for field water capacity, wilting point and saturated hydraulic conductivity, respectively.

4.3. Cropland management data

Country specific statistics on cultivated crop areas, crop rotation rules, crop management calendar and data on fertilization and other supporting data (various crop classifications and crop specific data) were compiled to satisfy the minimum set of EPIC model requirements for crop management data (APPENDIX 1).

4.3.1. Selection of crops for global database

GLU dataset crop list (chapter 2.4.2) was selected as the primary list of crops to be included in global database. In the final list of crops (Tab. 4.1.) all crop groups included in original GLU dataset (banana and plantains, sweet potato and yams, other pulses, other oil crops, and other fiber crops) were replaced by country specific representative of the crop group. The most frequent crop (as for its harvested area) in the country which was not included in original GLU crop list was taken from FAOSTAT as the crop group representative. If no crop was available for the country in FAOSTAT, the crop group representative of the nearest country was taken.

Final list of crops covers only those crops which are considered globally most important (You et Wood 2006) and many other crops with only regional significance is not included (e.g. temporary pastures or silage corn for most of central and western Europe).

Tab. 4.1: *List of crops assumed in global database.*

CROP ACRONYM IN GLOBAL DATABASE	CROP NAME
BARL	barley
BEAN	beans
CASS	cassava
COFF	coffe
COTT	cotton
GROU	ground nuts
MAIZ	maize
POTA	potatoes
RICE	rice
SORG	sorghum
SOYB	soybean
SUGC	sugarcane
WHEA	wheat
JUTE	jute
BHBE	broad and horse bean
SUNF	sunflower
SWTP	sweet potatoes
BANA	bananas
COCO	coconuts
CPEA	chick peas
FLAX	flax fibre and tow
LENT	lentils
LINS	linseed
MELO	melonseed
MILL	millet
MUST	mustard seed
OLIV	olives
OPAL	oil palm fruits
PEAS	peas, dry
PLAN	plantains

CROP ACRONYM IN GLOBAL DATABASE	CROP NAME
PULS	pulses, nec
RAPE	rapeseed
SAFF	safflower seed
SESA	sesame seed
SISA	sisal
SUGB	sugar beet
YAMS	yams

4.3.2. Crop share and crop rotation rules data

Primary aim of crop share data is to provide a basis for compilation of the hypothetic crop rotations or other crop cultivation strategies (such as plantations) which are to represent the SimU delineation specific land use base-run scenario in bio-physical modeling. Because of lack of relevant data on global scale, multiplied cropping during the year or multi-crop cultivation (more than one crop at the same place and time) was not considered in standing version of global database.

A SimU delineation and management system (high input, irrigated high input, low input and subsistence) specific list of physical areas of all crops from final crop list (further referred as crop share data) was constructed from original GLU data by zonal summing the crop areas over the SimU delineation and consecutively refined to fit the SimU delineation specific land use statistics data (chapter 4.1.2).

Tab. 4.2: *Total duration of cultivation for selected perennial crops.*

CROP	TOTAL LENGTH OF CULTIVATION (years)
sugar cane	4
coffee	25
banana plant	25
plantain plant	25
sisal	15
coconuts palm	70
olive tree	100
oil palm	25

Relevance masks were interpreted and together with various crop classifications included in crop share data table to make possible selection of various crop rotation options which could optionally take a part in base-run scenarios set for SimU delineation:

- Duration of crop cultivation less or more than 365 days was used as a criterion for classification of all crops as annual or perennial. For selected perennial crops the most likely total length of cultivation was interpreted from many available sources as given in table (Tab. 4.2.);
- Based on assumed crop cultivation strategies all crops were distinguished as crops cultivated in rotation or as crops cultivated in monoculture (most of perennial crops and typical plantain crops). For subsistence management system all crops (both annual and perennial) were assumed as they are grown in rotation;

- SimU delineation relevance mask was taken from land cover/land use statistics dataset (chapter 4.1.3) to indicate whether the management system is significant enough to be considered in crop rotation or crop calendar base-run scenario for the particular SimU delineation. Absolute relevance mask was interpreted based on SimU delineation specific management systems areas data coming from land use statistics dataset (chapter 4.1.3) so that only spatially dominant management system can be optionally selected as a representative of the particular SimU delineation in bio-physical simulations;
- Minimum crop area portion of total management system area in particular SimU delineation was set to 5% and used as a criterion for setting up the crop relevance mask. Crop relevance mask is to tell if crop is significant enough to be considered in base-run scenario for SimU delineation;
- In particular management system of some SimU delineations both crops grown in rotation and monoculture crops can be relevant for simulation. Therefore an alternative mask was interpreted to tell which crop cultivation strategy is more significant for particular management system in SimU delineation. Alternative mask was defined based on total area of monoculture crop or sum of all crops grown in rotation in particular SimU delineation and management system.

Tab. 4.3: *Crop rotation rules for global database, main crops in rows, preceding crops in columns (0 – not possible, 1 – not suitable but possible, 2 – suitable and very suitable).*

CROP	BARL	BEAN	CASS	GROU	MAIZ	JUTE	SUNF	BHBE	POTA	RICE	SORG	SOYB	SWTP	WHEA	SUGC	COTT	SUGB	MILL	MUST	SESA	SAFF	RAPE	PEAS	MELO	LINS	LENT	FLAX	CPEA	YAMS	BANA	PLAN
BARL	2	2	2	2	2	1	2	2	2	1	2	2	2	2	1	1	2	1	2	1	1	2	2	1	2	2	2	2	1	1	1
BEAN	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	2	2	1	2	1	1	1	1	2	1	2	1	2	2	2
CASS	1	2	1	2	1	0	0	2	1	0	1	2	1	1	0	0	1	1	1	1	1	1	2	1	0	2	0	2	1	0	0
GROU	2	1	2	1	2	2	2	1	2	2	2	1	2	2	0	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2	2
MAIZ	1	2	2	2	2	0	1	2	2	2	1	2	2	2	1	1	2	1	1	1	1	2	1	2	1	2	1	2	1	1	1
JUTE	1	2	1	2	1	1	1	2	2	2	1	2	0	0	0	0	1	1	1	1	1	1	2	1	1	2	1	2	0	1	1
SUNF	2	1	2	1	2	0	1	1	1	1	2	1	1	2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BHBE	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	2	1	2	1	2	1	1	1	1	1	1	2	2	2	2
POTA	2	2	1	2	2	1	0	2	1	1	2	2	1	2	1	1	1	2	2	1	2	2	2	1	2	2	2	2	1	1	1
RICE	2	2	1	2	1	0	1	2	1	2	1	2	1	2	0	0	0	1	1	1	1	1	2	1	1	2	1	2	0	1	1
SORG	2	2	2	2	2	1	1	2	2	1	2	2	2	2	0	2	2	1	1	1	1	2	1	2	2	2	2	2	2	2	2
SOYB	2	1	2	1	2	1	1	1	2	1	2	2	2	2	1	1	2	2	1	1	1	1	1	1	1	1	1	1	2	2	2
SWTP	0	0	1	2	2	1	2	0	1	2	2	1	2	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	2	2
WHEA	2	2	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	1	2	2	1	2	2	1	2	2	2	2	1	1	1
SUGC	1	2	0	2	2	1	2	2	1	2	2	2	0	1	2	0	1	2	2	2	2	2	2	2	2	2	2	2	0	0	0
COTT	2	2	0	2	2	1	1	2	0	0	2	2	0	2	0	2	0	2	1	1	1	1	2	1	1	2	1	2	0	0	0
SUGB	2	2	0	2	1	1	1	2	1	0	1	2	1	2	0	0	1	1	2	1	1	2	1	2	2	2	2	1	1	1	1
MILL	1	2	1	2	1	1	2	2	2	1	1	2	2	1	1	1	2	1	1	1	1	1	2	1	1	2	1	2	2	2	2
MUST	2	2	1	2	1	1	1	2	2	0	1	2	1	2	1	1	1	2	2	1	1	2	2	2	2	2	1	2	1	1	1
SESA	2	2	1	2	2	1	1	2	2	2	2	2	1	2	1	2	1	2	1	1	1	2	1	1	2	1	2	1	1	1	1
SAFF	2	1	1	2	2	1	1	2	2	1	2	2	1	1	1	1	1	2	1	2	1	1	2	1	1	2	1	2	1	1	1
RAPE	2	2	1	2	2	0	1	1	1	1	0	2	2	2	1	1	0	2	1	1	1	2	1	1	2	1	2	2	1	1	1
PEAS	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	2	2	2	2	1	2	1	1	2	1	2	1	2	1	1
MELO	2	2	1	2	2	1	1	2	1	1	2	2	1	2	1	1	1	1	1	1	2	1	1	2	1	2	1	2	1	2	2
LINS	2	1	1	1	2	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
LENT	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	2	2	2	1	2	1	1	2	1	2	1	2	1	1	1
FLAX	2	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2
CPEA	2	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	2	2	2	2	1	2	1	1	2	1	2	1	2	1	1
YAMS	0	0	1	2	2	1	2	0	1	1	2	1	0	0	0	0	1	2	1	1	1	1	0	1	1	0	1	0	2	2	2
BANA	1	2	0	2	2	1	1	2	2	2	2	2	2	1	0	0	1	2	1	1	1	1	2	1	1	2	1	2	0	1	1
PLAN	1	2	0	2	2	1	1	2	2	2	2	2	1	0	0	1	2	1	1	1	1	2	1	1	2	1	1	2	0	2	2

A crop rotation rules table compiled for Europe (Schmid et al. 2006) was after slight modifications and completing of missing crops from final list of crop (Tab. 4.1) taken as the

basis for global crop rotation rules table. Based on general knowledge on crop-after-crop suitability in rotation and crop calendar data (chapter 4.3.3) a crop combination suitability in rotation was classified into three classes as follows: i) very suitable to suitable (the crops are usually grown in mutual combination), ii) not suitable but possible (crops are not usually grown in mutual combinations or such combination is not recommended but there are not physical restrictions such as overlapping planting/harvesting dates to cultivate crops in combination), and iii) absolutely not suitable (physical restrictions such as overlapping dates of planting/harvesting make crop combination in rotation impossible). Crop rotation rules as it was set for global database are given in the table (Tab. 4.3).

4.3.3. Crop calendar data

A country and crop specific crop calendar from FAOAQUASTAT covering developing countries in tropical and subtropical regions was used as a basic source of data on starting and end decades for crop planting and harvesting. FAOAQUASTAT dataset was completed with other available sources such as USDA document published global scale data on crop management calendar, European regional data (MARS), and national (chapter 2.3.2, APPENDIX 2) or crop specific (chapter 2.3.2) crop management calendar data and table of country and crop specific start and end period of planting and harvesting was created. The resulting table was compared against the country specific crop share data (chapter 4.3.2) and all identified gaps in crop data were filled up by country specific crop calendar of crop with similar cultivation requirements as missing crop (Tab 4.4) or by crop calendar of the same crop from the nearest country.

Tab. 4.4: *Rules for filling up the gaps in missing crops crop calendar data, intepreted from IFA 1992 and Jurášek 1997, 1998 published data.*

ALTERNATIVE CROP	CROPS
maize	sorghum, millet, sunflower
bean	pea abd other pulses
soybean	groundnut
jute	groundnut
sunflower	safflower, maize

Mostly for large countries but also for all other countries with heterogonous natural conditions it is useful if crop calendar data can be interpreted in the way it respects the best the regional particularities. At global scale, however, such data is missing for the most of the countries. Therefore in the global database sub-country national level was used only for small group of large countries respecting the actually available published information. Countries and assumed sub-country regions are listed in table (Tab. 4.5).

The table of starting and end planting/harvesting periods compiled from available sources was then used for calculation of planting and harvesting dates so that they represented middle of reported planting or harvesting period. Length of vegetation period was calculated as difference (in days) of harvesting and planting day and used for final refinements of planting or harvesting dates in case calculated vegetation length value seriously exceeded published data on typical lengths of vegetation periods given in table (Tab. 4.6).

Tab. 4.5: *Countries and assumed sub-country regions for crop calendar data specification (regions according to crop calendar data source used for the country).*

COUNTRY	SUB-COUNTRY REGION
Australia	New South Wales
Australia	Northern Territory
Australia	Queensland
Australia	South Australia
Australia	Victoria
Australia	Western Australia
Brazil	centre-south
Brazil	north
Brazil	north-east
Brazil	south
Canada	Canadian Prairies
Canada	Ontario and Quebec
China	centre
China	north
China	south
Russian Federation	Central
Russian Federation	Siberia
Russian Federation	South
Russian Federation	Ural

Because of lack of relevant global scale data on agricultural management operations it was set arbitrary based on planting and harvesting dates as follows:

- One tillage operation (ploughing) has been set for all crops cultivated in all management systems. The plughing date has been set ten days before planting;
- Manure application has been set for all crops cultivated in high input, irrigated, and low input systems one day before ploughing;
- Nitrogen application has been assumed as it is done on two separate dates for crops cultivated in high input, irrigated and low input management systems. The first application is done before planting and the date is set three days before ploughing. One half of calculated nitrogen rate (chapter 4.3.4) is applied during the first application. The second application is done during the vegetation. The date of second nitrogen application has been for all annual crops set to half of vegetation period, for perennial crops the date has been set to the first year of the cultivation so that the nitrogen is applied in the middle of the period defined by planting date and end of the first year of cultivation. One half of calculated nitrogen rate (chapter 4.3.4) is applied during the second application. Total calculated nitrogen rate (chapter 4.3.4) is applied for selected perennial crops (cassava, sugar cane, banana, and plantains) in the second or third year of cultivation on the middle of the period between start of the second or third (cassava) cultivation year and date of harvest (cassava) or first harvest (sugar cane, banana, and plantain);
- Slightly modified nitrogen application scheme from that above has been set for perennial crops from table (Tab. 4.2) with exception of the sugar cane, banana, and plantains. First year nitrogen application is done the same way as for all other crops. In the second and

each other years of cultivation the nitrogen applied in two separate dates so that the first nitrogen application date is 121 days before the harvest date and the second 121 days after the harvest date. The dates of nitrogen application are refined in that manner that two nitrogen applications are to be done during one cultivation year;

- Phosphorous and potassium application has been set on only one date before planting or during vegetation period (in second and each other years for perennial crops from Tab 4.2. and cassava) together with nitrogen application;
- Due to lack of consistent data on irrigation water amounts on global scale the irrigation dates for crops cultivated in irrigated management system were not set in the standing version of global database.

Tab. 4.6: *Typical lengths of vegetation period (in days) for most of global database relevant crops (minimum, maximum and average lengths) as published by IFA 1992 and Jurášek 1997, 1998.*

| CROP	IFA, 1992		Jurášek 1997, 1998	
	MIN-MAX	AVERAGE	MIN-MAX	AVERAGE
Rice	70 - 160	115	90 - 210	150
Maize	60 - 280	170	90 - 150	120
Wheat	180 - 270	225	120 - 210	165
Barley	0	0	90 - 120	105
Sorghum	105 - 270	188	90 - 210	150
Millet	75 - 300	188	60 - 90	75
Potato	0	0	90 - 210	150
Sweet potato	120 - 180	150	120 - 180	150
Cassava	300 - 365	332,5	365 - 730	548
Yams	210 - 405	308	0	0
Bean	180 - 285	233	0	0
Pea	165 - 255	210	0	0
Chickpea	75 - 180	128	0	0
Lentil	90 - 165	128	0	0
Soybean	90 - 150	120	80 - 160	120
Groundnut	120 - 145	133	90 - 150	120
Sunflower	120 - 160	140	90 - 150	120
Oilseed Rape	320	320	270	270
Safflower	120	120	0	0
Olive_firt harvest	3 y.	3 y.	4 - 12 y.	8
Coconut_first harvest	4,5 y.	4 1/2 y.	6 - 9 y.	7 1/2
Oilpalm	2 1/2 - 3 1/2	3	3 - 4 y.	3 1/2
Sugarcane	300 - 720	510	300 - 690	495
Sugar beet	0	0	180 - 240	210
Cotton	140 - 210	175	160 - 220	190
Flax	90 - 120	105	0	0
Jute	90 - 150	120	90 - 150	120
Sisal_first harvest	900	900	0	0
Coffee_first harvest	720 - 1080	900	4*	4*
Banana	210 - 450	220	270 - 365	318

4.4.4. Fertlization rates

      Fertilization rates data were treated separately for manure application (and nitrogen rate coming from manure application) and N,P, and K fertilizers application.

      High resolution data on nitrogen rates coming from manure application were averaged over the SimU delineation by zonal statistics algorithm and yielded value was treated so that it did not exceed arbitrary set maximum value of 300 kg/ha of nitrogen. Consecutively, a coefficient 0.002 reported by (Williams et al. 1995) as default value of nitrogen amount coming from manure was used for calculation of total manure application from manure data.

      For countries covered by IFA statistics the country and crop specific information on fertilization rates (kg/ha) of nitrogen (N), phosphorous (P_2O_5), and potassium (K_2O) were taken directly for high input and irrigated management systems, for low input system the application rate was decreased to one half of total application rate reported in IFA. No fertilization is aasumed for subsistence management system.

      For other countries not covered by IFA statistics (Fig. 4.1) as well as for all the crops of IFA countries which are not covered by the statistics FAOSTAT country-level data on total fertilizer consumption (1000t) was recalculated to crop not-specific application rates (kg/ha) by simple algorithm taking into account particular management system.

Fig. 4.1: *Countries covered and not covered by IFA fertilizer application rates statistics.*
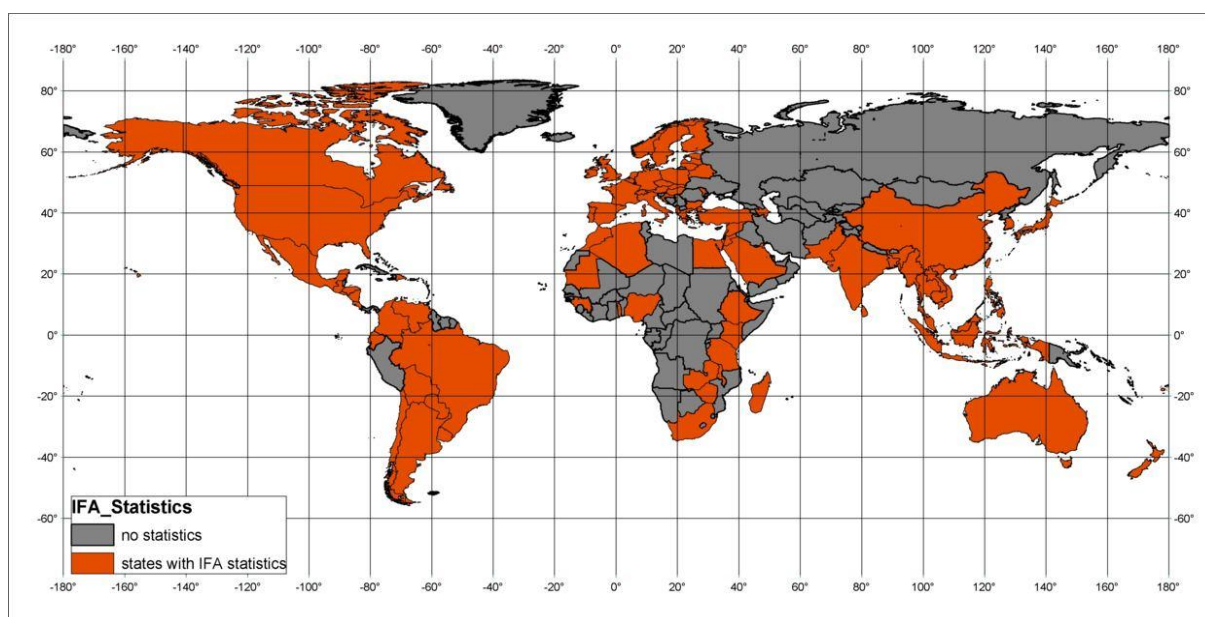
      Country specific high input and irrigated management systems areas were summed and their area portion calculated from total cropland area. The same was done for low input management system. Yielded area portions were used for allocation of total fertilizer application amounts to management systems. Low input management system area portion decreased to one half was used as coefficient for allocation total fertilizers amount (1000t) applied in low input system in particular country, the rest of total fertilizer application amounts was allocated to high input and irrigated management systems. Final value of fertilizers application rate for a particular country (kg/ha) was then calculated as total

allocated amount divided by the corresponding management systems areas. The origin of fertilization rate data has been flagged in the metadata fields of fertilization rate table.

Consecutively, calculated rates were refined so that it fit arbitrary set maximum values of 350 kg/ha or 175 kg/ha of N for high input and irrigated or low input management systems, respectively. Similarly a maximum rates of P_2O_5 and K_2O application rates were set to 300 kg/ha or 150 kg/ha for high input and irrigated or low input management systems, respectively.

Additional rules were applied for all pulses including soybean and groundnuts and flagged in the table so that nitrogen is applied only before harvest and maximum nitrogen application rate has been set to 90 kg/ha or 45 kg/ha for high input and irrigated management system or low input management system, respectively. For all management systems in the country having no corresponding area relevant for simulation, fertilizer application rate was arbitrary set to not-relevant (-999).

4.5. Weather data

Those EPIC model required weather parameters available from Tyndall dataset were directly taken and included in global database. Some other weather parameters were interpreted based on available data. Number of wet days was calculated by simple quadratic equation *Wet_days = α*sqrt(total_precipitation)*, where α was estimated for each grid cell with the precipitation and the number of wet days in the past (Schmid et al. 2006). The radiation was estimated from cloudiness data by the method described by Rivington et al. (2002) where the radiation is treated as a function of day length, latitude, atmospheric transmissivity, solar declination and solar elevation. Original as well as calculated parameters for the past were calculated as averages of different time intervals (10, 25, 50 and 100 years) for each climatic variable.

The variation data and probability values (standard deviation of minimal, maximal temperature and precipitation, skewness of precipitation, number of rainy days, probabilities of wet day after wet day and wet day after dry day) were calculated from daily sequences of ECWMF weather data. The probabilities of wet after wet and wet after dry day were calculated as follows: a number of wet-wet or dry-wet sequences together with dry and wet days was calculated for each month. Consecutively, the ratio of wet-wet vs. wet and dry-wet vs. dry was calculated. If the number of wet (or dry) days was effectively zero, a missing value was returned. Alternatively a non-informative Bayesian correction may be used so that for example PW|W = (wet-wet + 1)/(wet + 2).

5. Global database data structure

GEO-BENE global database for bio-physical modeling v. 1.0 is organized in four individual thematic datasets and one spatial reference dataset for geographical representation and visualization of thematic data:

- Spatial reference dataset (chapter 5.1);
- Land cover and Land use dataset (chapter 5.2);
- Topography and Soil dataset (chapter 5.3);
- Cropland management dataset (chapter 5.4);
- Climate dataset (chapter 5.5)

5.1. Spatial reference database (SpatialReference_v10)

Purpose of spatial reference dataset and domain tables is to reference and visualize thematic data of global database. Database is implemented in MS Office Access 2003 database. All tables and attribute fields contains metadata descriptions. One data table, four domain tables and one select query are included in the *SpatialReference_v10* database:

- *SpatialReference* data table, 5' spatial resolution grid centroid lattice, spatial reference for display global database content, includes all attributes necessary for SimU and HRU*PX30 zone delineation and relevant land cover/land use class visualization masks; (Tab. 5.1);
- *D_HRUalti* domain table, list and description of altitude classes used for HRU delineation (Tab. 5.2);
- *D_HRUslp* domain table, list and description of slope classes used for HRU delineation (Tab. 5.3);
- *D_HRUsoilgrp* domain table, list and description of soil classes used for HRU delineation (Tab. 5.4);
- *D_COUNTRY* domain table, list of countries used for global database, country identification numbers, country names and acronyms according to United Nations Statistics Division (Tab. 5.5);
- *D_ColRow30* domain table, split col_row30 identification into two separate fields (col30, row30), only for export data to EPIC (Tab. 5.6);
- *SpatialReference(DataPreview&EXPORT)* select query set on *SpatialReference* data table, helps with the spatial reference data preview and export the data, use of select query is optional (SQL code of query is given in APPENDIX 3);

Tab. 5.1: *SpatialReference data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
Col_Row	*Text(50)*	identification of 5' spatial resolution pixel (column and row number), counted from upper left corner;
XCOORD	*Number(Double)*	5' spatial resolution pixel centroid x coordinate (decdeg), can be used for data visualisation;
YCOORD	*Number(Double)*	5' spatial resolution pixel centroid y coordinate (decdeg), can be used for data visualisation;

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
PxArea	*Number(Double)*	real area (ha) of 5' spatial resolution pixel;
HRU	*Number(Integer)*	homogenous response unit (HRU) code, assembled of the code of AltiClass (first character), SlpClass (second character) and SoilClass (third character), 88 - soil class = 88 (no-soil prevails);
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
ZoneID	*Number(LongInteger)*	HRU*PX30 zone number, can be used for display HRU*PX30 zone related data geographicaly via spatial reference (topography and soil data);
SimUID	*Number(LongInteger)*	SimU delineation (HRU*PX30*COUNTRY) zone number, can be used for display SimU delineation related data via spatial reference (land cover/land use data, crop management data);
mCrpLnd	*Number(Byte)*	cropland relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mCrpLnd_H	*Number(Byte)*	cropland (high input management system) relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mCrpLnd_L	*Number(Byte)*	cropland (low input management system) relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mCrpLnd_I	*Number(Byte)*	cropland (irrigated high input management system) relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mCrpLnd_S	*Number(Byte)*	cropland (subsistence management system) relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mOthAgri	*Number(Byte)*	other agricultural land relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mGrass	*Number(Byte)*	grassland relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mForest	*Number(Byte)*	forest land relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mWetLnd	*Number(Byte)*	wettland relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mOthNatLnd	*Number(Byte)*	other natural land relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;
mNotRel	*Number(Byte)*	not-relevant land cover class relevant data visualisation mask, (1 = data is visualized, 0 = data is not visualized), based on arbitrary set rules;

Tab. 5.2: *D_HRUalti domain table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
AltiClass	*Number(Byte)*	altitude class for HRU delineation;
Altitude	*Text(50)*	altitude (m) interval;

Tab. 5.3: *D_HRUslp domain table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
SlpClass	*Number(Byte)*	slope class for HRU delineation;
Slope	*Text(50)*	slope (deg) interval;
RepSlp	*Number(Byte)*	slope class representative value (deg);
RepSlpPerc	*Number(Byte)*	slope class representative value (%) calculated by equation RepSlpPerc = tan(RepSlp*(3,14159/180)*100;

Tab. 5.4: *D_HRUsoilgrp domain table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
SoilClass	*Number(Byte)*	slope class for HRU delineation;
Soil	*Text(50)*	soil texture, stoniness and classification by FAO/UNESCO 1974 (Soil map of the world - legend);

Tab. 5.5: *D_COUNTRY domain table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
UN_Name	*Text(250)*	country name by United Nations Statistics Division;
UN_Acro	*Text(4)*	acronym of country by United Nations Statistics Division;

Tab. 5.6: *D_ColRow30 domain table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column, row), counted from upper left corner;
Col30	*Number(Integer)*	30' spatial resolution grid column number;
Row30	*Number(Integer)*	30' spatial resolution grid row number;

5.2. Land cover and Land use database (LandCover&LandUse_v10)

Purpose of the dataset is to equip the bio-physical and optimization modeling with spatial and statistical data on relevant land covers and particular land uses of arable land. Database is implemented in MS Office Access 2003 database. All tables and attribute fields contains metadata descriptions. Two data table, five domain tables and one select query are included in the *LandCover&LandUse_v10* database:

- *PixelStatistics* data table, global grid pixel specific statistics on land cover/land use (Tab. 5.7);
- *SimUstatistics* data table, SimU delineation (HRU*PX30*COUNTRY zone) aggregated statistics on land cover/land use, SimU delineations having HRU = 88 are excluded (Tab. 5.8);

34

- *D_HRUalti* domain table, list and description of altitude classes used for HRU delineation (Tab. 5.2, chapter 5.1);
- *D_HRUslp* domain table, list and description of slope classes used for HRU delineation (Tab. 5.3, chapter 5.1);
- *D_HRUsoilgrp* domain table, list and description of soil classes used for HRU delineation (Tab. 5.4, chapter 5.1);
- *D_COUNTRY* domain table, list of countries used for global database, country identification numbers, country names and acronyms according to United Nations Statistics Division (Tab. 5.5, chapter 5.1);
- *D_ColRow30* domain table, split col_row30 identification into two separate fields (col30, row30), only for export data to EPIC (Tab. 5.6, chapter 5.1);
- *SimUstatistics(DataPreview&EXPORT)* select query set on *SimUstatistics* and *D_ColRow30* tables, helps with the SimU statistics preview and export the data, use of select query is optional (SQL code of query is given in APPENDIX 3);

Tab. 5.7: *PixelStatistics data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
Col_Row	*Text(50)*	identification of 5' spatial resolution pixel (column and row number), counted from upper left corner;
HRU	*Number(Integer)*	homogenous response unit (HRU) code, assembled of the code of AltiClass (first character), SlpClass (second character) and SoilClass (third character), 88 - soil class = 88 (no-soil prevails);
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
SimUID	*Number(LongInteger)*	SimU delineation (HRU*PX30*COUNTRY) zone number, can be used for display SimU delineation related data via spatial reference (land cover/land use data, crop management data);
PxArea	*Number(Double)*	real area (ha) of 5' spatial resolution pixel;
CrpLnd	*Number(Double)*	cropland, total, physical area (ha);
CrpLnd_H	*Number(Double)*	cropland, high input management system, physical area (ha);
CrpLnd_L	*Number(Double)*	cropland, low input management system, physical area (ha);
CrpLnd_I	*Number(Double)*	cropland, irrigated high input management system, physical area (ha);
CrpLnd_S	*Number(Double)*	cropland, subsistence management system, physical area (ha);
OthAgri	*Number(Double)*	other agricultural land (other than cropland), physical area (ha);
Grass	*Number(Double)*	grassland, physical area (ha);
Forest	*Number(Double)*	forest, physical area (ha);
WetLnd	*Number(Double)*	wetland, physical area (ha);
OthNatLnd	*Number(Double)*	other natural land, physical area (ha);
NotRel	*Number(Double)*	not relevant land, physical area (ha);
Flag	*Number(Byte)*	metadata field, identification of method used for land cover data narmonization in particular case: 1 (CrpLnd > PxArea), 2 (CrpLnd > total agricultural area), 3 (CrpLnd < total agricultural area), 4 (CrpLn = 0);

Tab. 5.8: *SimUstatistics data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
SimUID	*Number(LongInteger)*	SimU delineation (HRU*PX30*COUNTRY) zone number, can be used for display SimU delineation related data via spatial reference (land cover/land use data, crop management data);
HRU	*Number(Integer)*	homogenous response unit (HRU) code, assembled of the code of AltiClass (first character), SlpClass (second character) and SoilClass (third character), 88 - soil class = 88 (no-soil prevails);
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
SimUArea	*Number(Double)*	real area (ha) of SimU delineation;
CrpLnd	*Number(Double)*	cropland, total, physical area (ha);
CrpLnd_H	*Number(Double)*	cropland, high input management system, physical area (ha);
CrpLnd_L	*Number(Double)*	cropland, low input management system, physical area (ha);
CrpLnd_I	*Number(Double)*	cropland, irrigated high input management system, physical area (ha);
CrpLnd_S	*Number(Double)*	cropland, subsistence management system, physical area (ha);
OthAgri	*Number(Double)*	other agricultural land (other than cropland), physical area (ha);
Grass	*Number(Double)*	grassland, physical area (ha);
Forest	*Number(Double)*	forest, physical area (ha);
WetLnd	*Number(Double)*	wetland, physical area (ha);
OthNatLnd	*Number(Double)*	other natural land, physical area (ha);
NotRel	*Number(Double)*	not relevant land, physical area (ha);
mCrpLnd	*Number(Byte)*	cropland relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if CrpLnd/SimUArea > 0.05;
mCrpLnd_H	*Number(Byte)*	cropland (high input management system) relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if CrpLnd_H/SimUArea > 0.05;
mCrpLnd_L	*Number(Byte)*	cropland (lowinput management system) relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if CrpLnd_L/SimUArea > 0.05;
mCrpLnd_I	*Number(Byte)*	cropland (irrigated high input management system) relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if CrpLnd_I/SimUArea > 0.05;
mCrpLnd_S	*Number(Byte)*	cropland (subsistence management system) relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if CrpLnd_S/SimUArea > 0.05;
mOthAgri	*Number(Byte)*	other agricultural land relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if OthAgri/SimUArea > 0.05;
mGrass	*Number(Byte)*	grassland relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if Grass/SimUArea > 0.05;
mForest	*Number(Byte)*	forest land relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if Forest/SimUArea > 0.05;
mWetLnd	*Number(Byte)*	wetland relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if WetLnd/SimUArea > 0.05;
mOthNatLnd	*Number(Byte)*	other natural land relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if OthNatLnd/SimUArea > 0.05;
mNotRel	*Number(Byte)*	not-relevant land cover classes relevance for SimU delineation zone (1 = relevant, 0 = not relevant), relevant if NotRel/SimUArea > 0.05;

5.3. Topography and soil database (Topography&Soil_v10)

Purpose of the dataset is to provide EPIC model with mandatory data on topography (geographical position, altitude and slope) and soil (soil profile and soil layer related data on soil properties). Database is implemented in MS Office Access 2003 database. All tables and attribute fields contains metadata descriptions. Four data table, five domain tables and two select queries are included in the *Topography&Soil_v10* database:

- *ZoneData* data table, HRU*PX30 zone specific data on geographic position, altitude, slope and the digital map of the world dominat soil typological unit (Tab. 5.9);
- *SoilAnal_MineralSoil* data table, the digital soil map of the world soil typological unit specific soil analytical data for 5 soil depth intervals (0.0 - 0.2m, 0.2 - 0.4m, 0.4 - 0.6m, 0.8 - 1.0m), mineral soils (Tab. 5.10);
- *SoilAnal_OrganicSoil* data table, the digital soil map of the world soil typological unit specific soil analytical data for 5 soil depth intervals (0.0 - 0.2m, 0.2 - 0.4m, 0.4 - 0.6m, 0.8 - 1.0m), organic soils (Tab. 5.11);
- *ClimateReference* data table, assignes historical climate data pixel (5' spatial resolution) to *HRU*PX30* zone by *Col_Row30* identification (Tab. 5.12);
- *D_HRUalti* domain table, list and description of altitude classes used for HRU delineation (Tab. 5.2, chapter 5.1);
- *D_HRUslp* domain table, list and description of slope classes used for HRU delineation (Tab. 5.3, chapter 5.1);
- *D_HRUsoilgrp* domain table, list and description of soil classes used for HRU delineation (Tab. 5.4, chapter 5.1);
- *D_COUNTRY* domain table, list of countries used for global database, country identification numbers, country names and acronyms according to United Nations Statistics Division (Tab. 5.5, chapter 5.1);
- *D_ColRow30* domain table, split col_row30 identification into two separate fields (col30, row30), only for export data to EPIC (Tab. 5.6, chapter 5.1);
- *Topo&Soil_MineralSoil(DataPreview&EXPORT)* select query set on *ZoneData*, *SoilAnal_MineralSoil*, *ClimateReference* and *D_ColRow30* tables, helps with the topo&soil data preview and export, use of select query is optional (SQL code of query is given in APPENDIX 3);
- *Topo&Soil_OrganicSoil(DataPreview&EXPORT)* select query set on *ZoneData*, *SoilAnal_OrganicSoil*, *ClimateReference* and *D_ColRow30* tables, helps with the topo&soil data preview and export, use of select query is optional (SQL code of query is given in APPENDIX 3);

Tab. 5.9: *ZoneData data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
ZoneID	*Number(LongInteger)*	HRU*PX30 zone number, can be used for display HRU*PX30 zone related data geographicaly via spatial reference (topography and soil data);
HRU	*Number(Integer)*	homogenous response unit (HRU) code, assembled of the code of AltiClass (first character), SlpClass (second character) and SoilClass (third character), 88 - soil class = 88 (no-soil prevails);
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
ZoneLon	Number(Double)	HRU*PX30 zone specific longitude (decimal degrees), x coordinate of 30' spatial resolution grid pixel centroid;
ZoneLat	Number(Double)	HRU*PX30 zone specific latitude (decimal degrees), y coordinate of 30' spatial resolution grid pixel centroid;
ZoneArea	Number(Double)	HRU*PX30 zone real area (ha);
ZoneAlti	Number(LongInteger)	HRU*PX30 zone mean altitude(m);
ZoneSlp	Number(Byte)	HRU*PX30 zone representative slope (%), derived as the mean value of the HRU slope class interval;
ZoneSTU	Text(3)	HRU*PX30 zone dominant digital map of the world soil typological unit, soil unit code according to FAO map of soils of the world legened (FAO 1974), no-soils (RK1, RK2, ST, NA, DS) according to WISE (Batjes 2006);

Tab. 5.10: *SoilAnal_MineralSoil data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
ZoneSTU	Text(3)	HRU*PX30 zone dominant digital map of the world soil typological unit, soil unit code according to FAO map of soils of the world legened (FAO 1974), no-soils (RK1, RK2, ST, NA, DS) according to WISE (Batjes 2006);
HG	Number(Byte)	soil hydrological group by USDA NRCS technical manual, 1 = A, 2 = B, 3 = C, 4 = D;
ALB	Number(Single)	albedo of moist soil surface; arbitrary set constant (albedo = 0,15);
DEPTH1	Number(Single)	depth of lower boundary of 1. soil layer (m), 0.2;
DEPTH2	Number(Single)	depth of lower boundary of 2. soil layer (m), 0.4;
DEPTH3	Number(Single)	depth of lower boundary of 3. soil layer (m), 0.6;
DEPTH4	Number(Single)	depth of lower boundary of 4. soil layer (m), 0.8;
DEPTH5	Number(Single)	depth of lower boundary of 5. soil layer (m), 1.0;
VS1	Number(Integer)	volume of stones (%), content of soil fragments > 2mm; WISE (Batjes 2006) value;
VS2	Number(Integer)	as above;
VS3	Number(Integer)	as above;
VS4	Number(Integer)	as above;
VS5	Number(Integer)	as above;
SAND1	Number(Integer)	sand content (%), WISE (Batjes 2006) value;
SAND2	Number(Integer)	as above;
SAND3	Number(Integer)	as above;
SAND4	Number(Integer)	as above;
SAND5	Number(Integer)	as above;
SILT1	Number(Integer)	silt content (%), WISE (Batjes 2006) value;
SILT2	Number(Integer)	as above;
SILT3	Number(Integer)	as above;
SILT4	Number(Integer)	as above;
SILT5	Number(Integer)	as above;
CLAY1	Number(Integer)	clay content (%), WISE (Batjes 2006) value;
CLAY2	Number(Integer)	as above;
CLAY3	Number(Integer)	as above;
CLAY4	Number(Integer)	as above;

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
CLAY5	Number(Integer)	as above;
BD1	Number(Single)	bulk density (t/m3), WISE (Batjes 2006) value;
BD2	Number(Single)	as above;
BD3	Number(Single)	as above;
BD4	Number(Single)	as above;
BD5	Number(Single)	as above;
CEC1	Number(Single)	cation exchange capacity (cmol/kg), WISE (Batjes 2006) value;
CEC2	Number(Single)	as above;
CEC3	Number(Single)	as above;
CEC4	Number(Single)	as above;
CEC5	Number(Single)	as above;
SB1	Number(Single)	sum of bases (cmol/kg), WISE (Batjes 2006) derived value by equation SB = WISE base saturation/100)* WISE CEC;
SB2	Number(Single)	as above;
SB3	Number(Single)	as above;
SB4	Number(Single)	as above;
SB5	Number(Single)	as above;
PH1	Number(Single)	pH in H2O, WISE (Batjes 2006) value;
PH2	Number(Single)	as above;
PH3	Number(Single)	as above;
PH4	Number(Single)	as above;
PH5	Number(Single)	as above;
CARB1	Number(Single)	(calcium) carbonate content (%), WISE (Batjes 2006) derived value by equation CARB = WISE total carbonate equivalent/1000g)*100;
CARB2	Number(Single)	as above;
CARB3	Number(Single)	as above;
CARB4	Number(Single)	as above;
CARB5	Number(Single)	as above;
CORG1	Number(Single)	soil organic carbon content, WISE (Batjes 2006) derived value by equation CORG = WISE organic carbon content/1000g)*100;
CORG2	Number(Single)	as above;
CORG3	Number(Single)	as above;
CORG4	Number(Single)	as above;
CORG5	Number(Single)	as above;
FWC1	Number(Double)	water content in soil (mm/mm) at field water capacity, WISE (Batjes 2006) sand,silt,clay and BD derived value (Rosetta model and Van Genuchten equation);
FWC2	Number(Double)	as above;
FWC3	Number(Double)	as above;
FWC4	Number(Double)	as above;
FWC5	Number(Double)	as above;
WP1	Number(Double)	water content in soil (mm/mm) at wilting point, WISE (Batjes 2006) sand,silt,clay and BD derived value (Rosetta model and Van Genuchten equation);
WP2	Number(Double)	as above;
WP3	Number(Double)	as above;
WP4	Number(Double)	as above;

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
WP5	*Number(Double)*	as above;
KS1	*Number(Double)*	saturated hydraulic conductivity (mm/hour) of saturated soil, WISE (Batjes 2006) sand,silt,clay and BD derived value (Rosetta model);
KS2	*Number(Double)*	as above;
KS3	*Number(Double)*	as above;
KS4	*Number(Double)*	as above;
KS5	*Number(Double)*	as above;

Tab. 5.11: *SoilAnal_OrganicSoil data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
ZoneSTU	*Text(3)*	HRU*PX30 zone dominant digital map of the world soil typological unit, soil unit code according to FAO map of soils of the world legened (FAO 1974), no-soils (RK1, RK2, ST, NA, DS) according to WISE (Batjes 2006);
HG	*Number(Byte)*	soil hydrological group by USDA NRCS technical manual, 1 = A, 2 = B, 3 = C, 4 = D;
ALB	*Number(Single)*	albedo of moist soil surface; arbitrary set constant (albedo = 0,15);
DEPTH1	*Number(Single)*	depth of lower boundary of 1. soil layer (m), 0.2;
DEPTH2	*Number(Single)*	depth of lower boundary of 2. soil layer (m), 0.4;
DEPTH3	*Number(Single)*	depth of lower boundary of 3. soil layer (m), 0.6;
DEPTH4	*Number(Single)*	depth of lower boundary of 4. soil layer (m), 0.8;
DEPTH5	*Number(Single)*	depth of lower boundary of 5. soil layer (m), 1.0;
VS1	*Number(Integer)*	volume of stones (%), arbitrary set to zero;
VS2	*Number(Integer)*	as above;
VS3	*Number(Integer)*	as above;
VS4	*Number(Integer)*	as above;
VS5	*Number(Integer)*	as above;
SAND1	*Number(Integer)*	sand content (%), arbitrary set to 30, soil texture assumed to be ballanced as for sand, silt and clay fractions portion;
SAND2	*Number(Integer)*	as above;
SAND3	*Number(Integer)*	as above;
SAND4	*Number(Integer)*	as above;
SAND5	*Number(Integer)*	as above;
SILT1	*Number(Integer)*	silt content (%), arbitrary set to 40, soil texture assumed to be ballanced as for sand, silt and clay fractions portion;
SILT2	*Number(Integer)*	as above;
SILT3	*Number(Integer)*	as above;
SILT4	*Number(Integer)*	as above;
SILT5	*Number(Integer)*	as above;
CLAY1	*Number(Integer)*	clay content (%), arbitrary set to 30, soil texture assumed to be ballanced as for sand, silt and clay fractions portion;
CLAY2	*Number(Integer)*	as above;
CLAY3	*Number(Integer)*	as above;
CLAY4	*Number(Integer)*	as above;
CLAY5	*Number(Integer)*	as above;
BD1	*Number(Single)*	bulk density (t/m3), set to upper limit of bulk density interval for fibric histosols given by Dreissen et al. 2001;

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
BD2	Number(Single)	as above;
BD3	Number(Single)	as above;
BD4	Number(Single)	as above;
BD5	Number(Single)	as above;
CEC1	Number(Single)	cation exchange capacity (cmol/kg); set to 95, average value calculated from WISE database (Batjes 1995) data on soil layers of three Od profiles from Guayana, Phillipines and New Zealand;
CEC2	Number(Single)	as above;
CEC3	Number(Single)	as above;
CEC4	Number(Single)	as above;
CEC5	Number(Single)	as above;
SB1	Number(Single)	sum of bases (cmol/kg); calculated as CEC *0,25 (O, Od, Ox) or CEC*0,75 (Oe), 0,25 is average base saturation calculated from WISE profiles (Batjes 1995), 0.75 is median of 0,5-1 base saturation interval defined for eutric qualifier (FAO-UNESCO 1974);
SB2	Number(Single)	as above;
SB3	Number(Single)	as above;
SB4	Number(Single)	as above;
SB5	Number(Single)	as above;
PH1	Number(Single)	pH in H2O; set to 4,5 for O, Od, Ox (avg calculated from WISE profiles, Batjes 1995) or to 6,5 for Oe (arbitrary value based on FAO-UNESCO 1974 and Dreissen et al. 2001 characteristics of eutric histosols);
PH2	Number(Single)	as above;
PH3	Number(Single)	as above;
PH4	Number(Single)	as above;
PH5	Number(Single)	as above;
CARB1	Number(Single)	(calcium) carbonate content (%), arbitrary set to zero;
CARB2	Number(Single)	as above;
CARB3	Number(Single)	as above;
CARB4	Number(Single)	as above;
CARB5	Number(Single)	as above;
CORG1	Number(Single)	organic carbon content, set to 35 based on WISE profile (Batjes 1995) average and FAO-UNESCO 1974 class definition of histosols;
CORG2	Number(Single)	as above;
CORG3	Number(Single)	as above;
CORG4	Number(Single)	as above;
CORG5	Number(Single)	as above;
FWC1	Number(Double)	water content in soil (mm/mm) at field water capacity, set to 0.52, adopted from soil input file for european peat soil (EC6) available in WOFOST model (version 7.1, Boogaard et al. 1998);
FWC2	Number(Double)	as above;
FWC3	Number(Double)	as above;
FWC4	Number(Double)	as above;
FWC5	Number(Double)	as above;
WP1	Number(Double)	water content in soil (mm/mm) at wilting point, set to 0.13, adopted from soil input file for european peat soil (EC6) available in WOFOST model (version 7.1, Boogaard et al. 1998);

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
WP2	*Number(Double)*	as above;
WP3	*Number(Double)*	as above;
WP4	*Number(Double)*	as above;
WP5	*Number(Double)*	as above;
KS1	*Number(Double)*	saturated hydraulic conductivity (mm/hour), set to 5.62, adopted from soil input file for european peat soil (EC6) available in WOFOST model (version 7.1, Boogaard et al. 1998);
KS2	*Number(Double)*	as above;
KS3	*Number(Double)*	as above;
KS4	*Number(Double)*	as above;
KS5	*Number(Double)*	as above;

Tab. 5.12: *ClimateReference data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;
Tyndall_X	*Number(Double)*	HRU*PX30 zone relevant historical climate data (Tyndall dataset) x-coordinate or longitude (decimal degrees) of the pixel;
Tyndall_Y	*Number(Double)*	HRU*PX30 zone relevant historical climate data (Tyndall dataset) y-coordinate or latitude (decimal degrees) of the pixel;
Flag	*Number(Byte)*	metadata field, flag1 = 1 if no counterpart pixel to 30' spatial resolution pixel was found in climate dataset and Tyndall_X(Y) of the nearest pixel was used instead;

5.4. Cropland management database (CroplandManagement_v10)

Purpose of the dataset is to provide EPIC model with data on crop management alternatives and crop shares within the particular land use, crop and crop management calendar (dates of planting, harvesting, tillage, fertilization and irrigation) and inputs to agriculture (fertilizer application rates). Database is implemented in MS Office Access 2003 database. All tables and attribute fields contains metadata descriptions. Seven data table, five domain tables and three select queries are included in the *CroplandManagement_v10* database:

- *CropCalendar* data table, country and crop specific crop management operation calendar; (Tab. 5.13);
- *CropShare* data table, SimU and management system specific list of crops (crop share), includes data on physical crops area and relevance masks for selection the alternatives (Tab. 5.14);
- *FertilizerApplication* data table, country and crop specific N, P, K fertilizer application rates (Tab. 5.15);
- *ManureApplication* data table, SimU delineation specific data on N comming from manure (application rates) and total manure application (Tab. 5.16);
- *CropIdentification* data table, list of original IFPRI dataset crops and crop groups and all crop country alternatives for IFPRI data crop groups (Tab. 5.17);

- *CropRotationRules* data table, main crop suitability for cultivation after the preceeding crop, rules; (Tab. 5.18);
- *CropDuration* data table, duration of crop cultivation for perennial crops (Tab. 5.19);
- *D_HRUalti* domain table, list and description of altitude classes used for HRU delineation (Tab. 5.2, chapter 5.1);
- *D_HRUslp* domain table, list and description of slope classes used for HRU delineation (Tab. 5.3, chapter 5.1);
- *D_HRUsoilgrp* domain table, list and description of soil classes used for HRU delineation (Tab. 5.4, chapter 5.1);
- *D_COUNTRY* domain table, list of countries used for global database, country identification numbers, country names and acronyms according to United Nations Statistics Division (Tab. 5.5, chapter 5.1);
- *D_ColRow30* domain table, split col_row30 identification into two separate fields (col30, row30), only for export data to EPIC (Tab. 5.6, chapter 5.1);
- *D_SubCountry* list of first sub-country region used for further specification of country-level crop calendar data, original sub-country identification numbers and names according to GAUL dataset (Tab. 5.20);
- *CropCalendar(DataPreview&EXPORT)* select query set on *CropCalendar* table, helps with the crop management calendar data preview and export, metadata descrition of attribude field is given in design mode of table, use of select query is optiona (SQL code of query is given in APPENDIX 3);
- *CropShare(DataPreview&EXPORT)* select query set on *CropShare* and *D_ColRow30* tables, helps with the crop share data preview and export, use of select query is optiona (SQL code of query is given in APPENDIX 3);
- *ManureApplication(DataPreview&EXPORT)* select query set on *ManureApplication* and *D_ColRow30* tables, helps with the manure application rates data preview and export, use of select query is optiona (SQL code of query is given in APPENDIX 3);

Tab. 5.13: *CropCalendar data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
CrpCalSpec	*Number(Byte)*	identification number of sub-region for further specification of country level crop calendar;
CROP	*Text(4)*	code of original IFPRI crop and where appropriate IFPRI crop group (SWPY, BANP, OOIL, OPUL) replaced by country most frequent representative of the crop group;
OPER	*Text(50)*	identification of agro-technic operation;
YEA	*Number(Byte)*	year of crop cultivation;
DAY	*Number(Integer)*	number of day of agrotechnical operation (counted from 1.1. to 31.12., 29.2. not included);
DAT	*Date/Time*	date of agro technical operation (1.1. 2001 - 31.12.2005);
mHI	*Number(Byte)*	operation relevance for crop cultivated in high input management system (1 = relevant, 0 = not-relevant);
mIR	*Number(Byte)*	operation relevance for crop cultivated in irrigated high input management system (1 = relevant, 0 = not-relevant);
mLI	*Number(Byte)*	operation relevance for crop cultivated in low input management system (1 = relevant, 0 = not-relevant);
mSS	*Number(Byte)*	operation relevance for crop cultivated in subsistence management

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
		system (1 = relevant, 0 = not-relevant);

Tab. 5.14: *CropShare data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
SimUID	*Number(LongInteger)*	SimU delineation (HRU*PX30*COUNTRY) zone number, can be used for display SimU delineation related data via spatial reference (land cover/land use data, crop management data);
MngSystem	*Text(2)*	identification of management system, HI - high input, LI - low input, IR - irrigated high input, SS - subsistence system, management systems according to You et Wood 2006;
HRU	*Number(Integer)*	homogenous response unit (HRU) code, assembled of the code of AltiClass (first character), SlpClass (second character) and SoilClass (third character), 88 - soil class = 88 (no-soil prevails);
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
CrpCalSpec	*Number(Byte)*	identification number of sub-region for further specification of country level crop calendar;
CROP	*Text(4)*	code of original IFPRI crop and where appropriate IFPRI crop group (SWPY, BANP, OOIL, OPUL) replaced by country most frequent representative of the crop group;
CultType	*Number(Byte)*	crop cultivation type, crop classification according to duration of crop cultivation on the field, 1 (crops cultivated for < 365 days), 2 (crops cultivated for > 365 days);
RotType	*Number(Byte)*	crop rotation type, 1 (crops cultivated in monoculture), 2 (crops cultivated in rotation);
CrpArea	*Number(Double)*	SimU delineation specific physical area (ha) of particular crop;
SimUmask	*Number(Byte)*	SimU relevance mask (1 = relevant, 0 = not relevant), relevant if particular management system fulfills the area portion threshold (taken from land cover/land use statistic dataset);
absSimUmask	*Number(Byte)*	absolute SimU relevance mask (1 = relevant, 0 = not relevant), relevant only for one management system having the spatial dominance in the SimU delineation;
CropMask	*Number(Byte)*	crop relevance mask (1 = relevant, 0 = not relevant), relevant is crop having more than 5% area portion of the particular management system total area;
AltMask	*Number(Byte)*	alternative mask (1 = relevant, 0 = not-relevant), mark SimU relevant management alternative (rotation or monoculture1, monoculture2, monoculture n) which is spatially dominant in SimU delineation;

Tab. 5.15: *FertilizerApplication data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
CROP	*Text(4)*	code of original IFPRI crop and where appropriate IFPRI crop group (SWPY, BANP, OOIL, OPUL) replaced by country most frequent representative of the crop group;
N1_HiIr	*Number(Double)*	nitrogen (N) application 1 (kg/ha), before planting or during

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
		vegetation (perennial crops) together with P, K application, rainfed and irrigated high input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
N2_HiIr	*Number(Double)*	nitrogen (N) application 2 (kg/ha), during vegetation, rainfed and irrigated high input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
N3_HiIr	*Number(Double)*	nitrogen (N) application 3 (kg/ha), during vegetation (third year for selected crops), rain fed and irrigated high input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
P_HiIr	*Number(Double)*	phosphorous (P2O5) application (kg/ha), before planting or during vegetation (perennial crops), rain fed and irrigated high input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
K_HiIr	*Number(Double)*	potassium (K2O) application (kg/ha), before planting or during vegetation (perennial crops), rain fed and irrigated high input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
N1_Li	*Number(Double)*	nitrogen (N) application 1 (kg/ha), before planting or during vegetation (perennial crops) together with P, K application, low input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
N2_Li	*Number(Double)*	nitrogen (N) application 2 (kg/ha), during vegetation, low input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
N3_Li	*Number(Double)*	nitrogen (N) application 3 (kg/ha), during vegetation (third year for selected crops), low input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
P_Li	*Number(Double)*	phosphorous (P2O5) application (kg/ha), before planting or during vegetation (perennial crops), low input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
K_Li	*Number(Double)*	potassium (K2O) application (kg/ha), before planting or during vegetation (perennial crops), low input management system, -999 no relevant (no area of corresponding management system in COUNTRY);
Flag1	*Number(Byte)*	metadata field, if = 1 then data was taken directly from original IFA country and crop specific dataset;
Flag2	*Number(Byte)*	metadata field, if = 1 then data was calculated from FAOSTAT dataset on country specific N, P, K consumption;
Flag3	*Number(Byte)*	metadata field, if = 1 then max. 90 kg/ha (rain fed and irrigated high input) or 45 kg/ha (low input management systems) application rates were arbitrary set for pulses, soya and groundnuts in case calculated value exceeded the threshold;

Tab. 5.16: *ManureApplication data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
SimUID	*Number(LongInteger)*	SimU delineation (HRU*PX30*COUNTRY) zone number, can be used for display SimU delineation related data via spatial reference (land cover/land use data, crop management data);
HRU	*Number(Integer)*	homogenous response unit (HRU) code, assembled of the code of AltiClass (first character), SlpClass (second character) and SoilClass (third character), 88 - soil class = 88 (no-soil prevails);

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
Col_Row30	*Text(50)*	identification of 30' spatial resolution pixel (column and row number), counted from upper left corner;
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
Nrate	*Number(Integer)*	nitrogen (kg/ha) coming from manure application, taken from high resolution global scale data (Liu et al. 2008);
Manure	*Number(LongInteger)*	total manure application (kg/ha), arbitrary set as Nrate/0,002;

Tab. 5.17: *CropIdentification data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
CROP	*Text(4)*	code of original IFPRI crop and where appropriate IFPRI crop group (SWPY, BANP, OOIL, OPUL) replaced by country most frequent representative of the crop group;
CrpName	*Text(50)*	crop name;
Flag	*Number(Byte)*	metadata field, identification of crop source, 1 = original IFPRI data list of crops or crop groups, 2 = country specific representative for crop group;

Tab. 5.18: *CropRotationRules data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
CROP	*Text(4)*	main crop, code of original IFPRI crop and where apropriate IFPRI crop group (SWPY, BANP, OOIL, OPUL) replaced by country most frequent representative of the crop group;
BARL	*Number(Byte)*	preceding crop, suitability for main crop: 0 = not suitable, 1 = not suitable but possible, 2 = suitable or very suitable;
BEAN	*Number(Byte)*	as above
CASS	*Number(Byte)*	as above
GROU	*Number(Byte)*	as above
MAIZ	*Number(Byte)*	as above
JUTE	*Number(Byte)*	as above
SUNF	*Number(Byte)*	as above
BHBE	*Number(Byte)*	as above
POTA	*Number(Byte)*	as above
RICE	*Number(Byte)*	as above
SORG	*Number(Byte)*	as above
SOYB	*Number(Byte)*	as above
SWTP	*Number(Byte)*	as above
WHEA	*Number(Byte)*	as above
SUGC	*Number(Byte)*	as above
COTT	*Number(Byte)*	as above
SUGB	*Number(Byte)*	as above
MILL	*Number(Byte)*	as above
MUST	*Number(Byte)*	as above
SESA	*Number(Byte)*	as above
SAFF	*Number(Byte)*	as above

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
RAPE	*Number(Byte)*	as above
PEAS	*Number(Byte)*	as above
MELO	*Number(Byte)*	as above
LINS	*Number(Byte)*	as above
LENT	*Number(Byte)*	as above
FLAX	*Number(Byte)*	as above
CPEA	*Number(Byte)*	as above
YAMS	*Number(Byte)*	as above
BANA	*Number(Byte)*	as above
PLAN	*Number(Byte)*	as above

Tab. 5.19: *CropDuration data table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
CROP	*Text(4)*	code of original IFPRI crop and where appropriate IFPRI crop group (SWPY, BANP, OOIL, OPUL) replaced by country most frequent representative of the crop group;
DURATION	*Number(Byte)*	duration (years) of crop cultivation;
LastYearRepeat	*Number(Byte)*	number of cultivation years after the last year indiced in crop calendar, the same calendar data as for the last year;

Tab. 5.20: *D_SubCountry domain table attribute fields descriptions.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
COUNTRY	*Number(Integer)*	country code, numerical code by United Nations Statistics Division;
GAUL1_code	*Number(LongInteger)*	identification of first sub-country region, GAUL administrative region code;
CrpCalSpec	*Number(Byte)*	identification number of sub-region for further specification of country level crop calendar;
NameCrpCalSpec	*Text(50)*	name of subregion for further specification of country level crop calendar;

5.5. Climate database (Climate_v10)

The data set provides the necessary climate input data (climate in the past and future climate scenarios). The climate database consists of two separate data sets:

- *HistoricalClimate_v10* is a MS Office Access 2003 database where the Tyndall weather data from 1901 to 2002 together with the monthly statistics derived from ECMWF database are stored. Six data tables hold the data (name specification is *ts_cru_ts_2_10_1901_2002_<interval>_<variable-name>*) _cld (cloud cover in %), _pre (precipitation in mm), _rad (radiation in MJ/m²), _tmn (near-surface temperature minimum in °C), _tmx (near-surface temperature maximum in °C), and _wet (wet day frequency in days) separately for 10, 25, 50, and 100 year intervals. Seven tables contain the statistical values *sd_tmin* (standard deviation of minimum temperature), *sd_tmax* (standard deviation of maximum temperature), *sd_rain* (standard deviation of

precipitation), *sk_rain* (skewness of precipitation), *ww_rain* (probabilities of wet day after wet day) and *dw_rain* (probabilities of wet day after dry day). All tables have the same structure, described in Tab. 5.21;

- Eight Tyndall climate change scenarios are available as postgreSQL databases. The database backups comprise the PCM and HadCM3 scenarios for the 4 different global warming values A1FI, A2, B1 and B2. Each backup file contains seven tables (name specification is: *scenario_<global-climate-model>_<variable-name>*): *_cld* (cloud cover in %), *_pre* (precipitation in mm), *_rad* (radiation in MJ/m²), *_tmn* (near-surface temperature minimum in °C), *_tmx* (near-surface temperature maximum in °C), *_tmp* (near-surface temperature in °C), and *_wet* (wet day frequency in days). All tables have the same structure, described in Tab. 5.22.

Tab. 5.21: *HistoticalClimate_v10 structure of tables and identification of attribute fields.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
period	*Number(Integer)*	time interval (years) for the average value of the Tyndall dataset;
lon	*Number(single)*	longitude of 0.5° (Tyndall) or 2.5° (ECWMF) spatial resolution pixel, pixel identification;
lat	*Number(single)*	latitude of 0.5° (Tyndall) or 2.5° (ECWMF) spatial resolution pixel, pixel identification;
jan	*Number(single)*	mean value for January, depends on the table theme;
feb	*Number(single)*	mean value for February, depends on the table theme;
mar	*Number(single)*	mean value for March, depends on the table theme;
april	*Number(Single)*	mean value for April, depends on the table theme;
may	*Number(single)*	mean value for May, depends on the table theme;
jun	*Number(single)*	mean value for June, depends on the table theme;
jul	*Number(single)*	mean value for July, depends on the table theme;
aug	*Number(single)*	mean value for August, depends on the table theme;
sep	*Number(single)*	mean value for September, depends on the table theme;
oct	*Number(single)*	mean value for October, depends on the table theme;
nov	*Number(single)*	mean value for November, depends on the table theme;
dec	*Number(single)*	mean value for December, depends on the table theme;

Tab. 5.22: *ClimateScenarios backups table structure and identification of attribute fields.*

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
year	*Number(LongInteger)*	year, range 1 (2001) – 100 (2100);
lon	*Number(single)*	longitude of 0.5° spatial resolution pixel, pixel identification;
lat	*Number(single)*	latitude of 0.5° spatial resolution pixel, pixel identification;
jan	*Number(single)*	mean value for January, depends on the table theme;
feb	*Number(single)*	mean value for February, depends on the table theme;
mar	*Number(single)*	mean value for March, depends on the table theme;
april	*Number(Single)*	mean value for April, depends on the table theme;
may	*Number(single)*	mean value for May, depends on the table theme;
jun	*Number(single)*	mean value for June, depends on the table theme;
jul	*Number(single)*	mean value for July, depends on the table theme;
aug	*Number(single)*	mean value for August, depends on the table theme;

ATTRIBUTE FIELD	DATA TYPE	DESCRIPTION
sep	*Number(single)*	mean value for September, depends on the table theme;
oct	*Number(single)*	mean value for October, depends on the table theme;
nov	*Number(single)*	mean value for November, depends on the table theme;
dec	*Number(single)*	mean value for December, depends on the table theme;

References

Adams, D.M.; Alig, R.J.; Callaway, J.M.; McCarl, B.A.; Winnett, S.M., 1996. The forest and agricultural sector optimization model (FASOM): model structure and policy applications. Res. Pap. PNW-RP-495. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 60 p.

Balkovič, J., Schmid, E., Bujnovský, R., Skalský, R., Poltárska, K., 2006. Biophysical modelling for evaluating soil carbon sequestration potentials on arable land in the pilot area Baden-Württemberg (Germany). Agriculture, Vol. 52, No. 4, pp. 169 – 176

Batjes, N.H. (Ed.), 1995. A homogenized soil data file for globalenvironmental research: a subset of FAO. ISRIC and NRCS profiles(Version 1.0). Working Paper and Preprint 95/10b. ISRIC - World Soil Information, Wageningen, 43 p. Available on internet: <http://www.isric.org/ISRIC/WebDocs/Docs/isric_wp95_10.pdf>, Accessed: February 27 2008

Batjes, N., H., 2006. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.1). Report 2006/2, ISRIC - World Soil Information, Wageningen, 46 pp., Available on internet: http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2006_02.pdf, Accessed: February 8, 2008

Batjes, N. H., Bridges, E.M., Nachtergaele, F.O., 1995. World Inventory of Soil Emission Potentials: Development of a Global Soil Data Base of Process-Controlling Factors. In Peng, S., Ingram, K.T., Neue, H.U., Ziska, L.H. (Eds.) 1995. Climate change and rice. Springer-Verlar, Berlin, Heidelberg, p. 102 – 115

Boogaart, H.L., Van Diepen, C.A., Rotter, R.P., Cabrera, J.M., Van Laar, H.H., 1998: User´s guide for the WOFOST 7.1 crop growth simulation model and WOFOST Control Center 1.5. Technical document. DLO Winand Staring Centre, Wageningen, 141 p.

Di Gregorio, A., Jansen, L., 2000. Land cover classification system, classification concepts and user manual. FAO, Rome, Available on internet: <http://www.fao.org/docrep/003/X0596E/X0596E00.HTM>, Accessed: February 14 2008

Dobos, E., 2006. Albedo. In Lal, R. (Ed.). Encyclopedia of Soil Science – Second edition, Volume 1. Taylor&Franciss CRC Press, New York, (2006), pp. 63 – 65

Driessen, P., Deckers, J., Spaargaren, O., Nachtergaele, F. 2001. Lecture notes on the major soils of the world. World Soil Resources Reports 94, FAO, Rome, 334 p.

FAO, 1959a. World crop harvest calendar for sugar beets. Meeting of Experts of Time Reference (7 – 18 Dec.). FAO, Roma, 6 p.

FAO, 1959b. World crop harvest calendar for sugar cane. Meeting of Experts of Time Reference (7 – 18 Dec.). FAO, Roma, 7 p.

FAO, 1959c. World crop harvest calendar for coffee. Meeting of Experts of Time Reference (7 – 18 Dec.). FAO, Roma, 7 p.

FAO-UNESCO, 1974. Soil map of the world, Volume 1 - Legend. Unesco, Paris, 59 p.

Fischer, G., van Velthuizen, H., Nachtergele, F., 2000. Global Agro-Ecological Zones Assessment: Methodology and Results. Interim Report IR-00-064, IIASA, Laxenburg, 48 p., Available on internet: <http://www.iiasa.ac.at/Admin/PUB/Documents/RR-02-002.pdf>, Acessed: February 14 2008

IFA, 1992. World Fertilizer Use Manual. Germany, 1992. 632 p. ISBN 2-9506299-0-3.

Jurášek, P., 1997. World Agriculture. Part I. (in Slovak). AT Publishing, Bratislava, 263 p. ISBN 80-967812-0-0.

Jurášek, P., 1998. World Agriculture. Part II. (in Slovak). AT Publishing, Bratislava, 387 s. ISBN 80-967812-9-4.

Justice C.O. and Becker-Reshef I.(Eds.) (2007) Report from the Workshop on Developing a Strategy for Global Agricultural Monitoring in the framework of Group on Earth Observations (GEO), UN FAO, July 2007, Geography Dept., University of Maryland, 66, pp, Available on internet: <ftp://ftp.iluci.org/GEO_Ag> , Accessed: February 6, 2008

Kučera, L., Genovese, G., (Eds.), 2004a. Crop Monographies on Central European Countries. Vol. 2 Estonia, Latvia, Lithuania, Poland. Office for Official Publications of the European Communities, Luxembourg, 122 pp. ISBN 92-894-8177-3. Available on internet: (http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/MOCA/00000000.HTM) Accessed: March 11, 2008.

Kučera, L., Genovese, G., (Eds.), 2004b. Crop Monographies on Central European Countries. Vol. 3 Czech Republic, Slovakia, Hungary, Slovenia. Office for Official Publications of the European Communities, Luxembourg, 160 pp. ISBN 92-894-8178-1. Available on internet: (http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/MOCA/00000000.HTM) Accessed: March 11, 2008.

Kučera, L., Genovese, G., (Eds.), 2004c. Crop Monographies on Central European Countries. Vol. 4 Romania, Bulgaria, Turkey. Office for Official Publications of the European Communities, Luxembourg, 122 pp. ISBN 92-894-8179-X. Available on internet: (http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/MOCA/00000000.HTM   ) Accessed: March 11, 2008.

Lethcoe, K.J., Klaver, R.W., 1998. Simulating the Interrupted Goode Homolosine Projection With ArcInfo. 18th Annual ESRI International User Conference in San Diego Proceedings, California, Available on internet: <http://gis.esri.com/library/userconf/proc98/PROCEED.HTM>, Accessed: February 12, 2008

Liu J., You L., Obersteiner M., Zehnder A.J.B. and Yang, H., 2008. A high-resolution assessment on global nitrogen flows in cropland. In preparation.

MacKerron, D. K. L., 1992. Agrometeorological Aspects of Forecasting Yields of Potato within the E. C. Office for Official Publications of the European Communities, Luxembourg, 250 pp. ISBN 92-826-3538-4.

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., New, M., 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100). Tyndall Centre for Climate Change Research Working Paper 55. 25 p., Available on Internet: http://www.tyndall.ac.uk/publications/working_papers/wp55.pdf, Accessed: March 14, 2008

Narciso, G., Ragni, P., Venturi, A., 1992. Agro-meteorological aspects of crops in Italy, Spain and Greece: A summary review for common and durum wheat, barley, maize, rice, sugar beet, sunflower, soya bean, rape, potato, tobacco, cotton, olive and grape crops. Office for Official Publications of the European Communities, Luxembourg, 440 pp. ISBN 92-826-3995-6.

Rivington, M., Matthews, K.B., Buchan, K., 2002. A Comparison of Methods for Providing Solar Radiation Data to Crop Models and Decision Support Systems. In Rizzoli, A.E., Jakeman, J.J., (Eds.) Integrated Assessment and Decision Support, Proceedings of the 1st Biennial Meeting of the iEMSs, Volume 3, The International Environmental Modelling and Software Society, Lugano, Switzerland, pp. 193 – 198, Available on Internet: http://www.wcai-infonet.org/servlet/BinaryDownloaderServlet?filename=1065620840070_wdds25.pdf&refID=113918, Accessed: March 14, 2008

Rossiter, D. G. 2003. Biophysical models in land evaluation. In Encyclopedia of Life Support Systems (EOLSS), 2003, 16 p. Available on internet: <www.itc.nl/~rossiter/Docs/EOLSS_1527_Preprint.pdf >, Accessed: February 7, 2008

Russell, G., 1990. Barley Knowledge Base. Office for Official Publications of the European Communities, Luxembourg, 142 pp. ISBN 92-826-1647-9.

Russell, G., Wilson, G. W., 1994. An Agro-Pedo-Climatological Knowledge-Base of Wheat in Europe. Office for Official Publications of the European Communities, Luxembourg, 160 pp. CL-NA-15789-EN-C.

Schmid, E., Balkovič, J., Moltchanova E., Skalský, R., Poltárska, K., Müller, B., Bujnovský, R., 2006: Biophysical process modeling for EU25: concept, data, methods, and results. Deliverable D3 (T30), Final Report, Appendix II., EU FP 6 Project INSEA − Integrated Sink Enhancement Assessment (SSPI-CT-2003/503614 with DG RTD), 2006, 76 p.

Schaap, M.G., Bouten, W., 1996. Modelling Water Retention Curves of Sandy Soils Using Neural Networks. Water Resour. Res., Vol. 32, s. 3033 – 3040.

Siebert, S., Döll, P., Feick, S., Hoogeveen, J., Frenken, K., 2007. Global Map of Irrigation Areas version 4.0.1. Johann Wolfgang Goethe University, Frankfurt an Main, Germany/ FAO, Rome. Available on internet: (http://www.fao.org/nr/water/aquastat/ irrigationmap/index10.stm). Accessed: March 11 2008.

Stolbovoy, V., Montanarella, L., Panagos, P., (Eds.) 2007. Carbon Sink Enhancement in Soils of Europe: Data, Modeling, Verification. JRC technical and scientific reports, EUR 23037 EN, Office for Official Publications of the European Communities, Luxembourg, 183 p. Available on internet: <http://eusoils.jrc.it/ESDB_Archive/eusoils_docs/other/ EUR23037.pdf>, Acessed: February 27, 2008

USDA, 1994. Major World Crop Areas and Climatic Profiles. World Agricultural Outlook Board, U. S. Department of Agriculture. Agriculture Handbook No. 664., 279 p.

USDA-NRCS, 2007. Hydrologic Soil Groups. National Engineering Handbook : Part 630 Hydrology : Chapter 7. 210-VI-NEH-630.7, USDA, Washington DC. Available on internet: <http://directives.sc.egov.usda.gov/media/pdf/H_210_630_7.pdf>, Accessed February 21 2008

Williams, J.R. 1995. The EPIC Model. In Computer Models of Watershed Hydrology (Ed.: V.P. Singh). Water Resources Publications, Highlands Ranch, Colorado, 1995, pp 909-1000.

Wösten, J. H. M., Van Genuchten, M. T., 1988. Using Texture and Other Soil Properties to Predict the Unsaturated Soil Hydraulic Functions. Soil. Sci. Soc. Am. J., Vol. 52, p. 1762 - 1770

You, L. and S. Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems Vol.90, Issues1-3 p.329-347.

Appendix 1: Input data requirements of the EPIC model

Data on four major input data components: weather, soil, topography and management practices is essential to run EPIC.

The **weather variables** necessary for running EPIC are precipitation (in mm), minimum and maximum air temperature (in degree Celsius), and solar radiation (in MJ/m²). If the Penman methods are used to estimate potential evapo-transpiration, wind speed (in m/sec measured at 10 m height), and relative humidity (in %) are also required. If measured daily weather data is available it can be directly input into EPIC. In addition, monthly statistics of this daily weather (mean, standard deviation, skew coefficient, probabilities of wet-dry and wet-wet days, etc.) need to be computed and input in the model. EPIC provides a statistical support program to compute the statistics of relevant variables based on daily weather records. Consequently, long historical daily weather records (20-30 years) for all weather variables are desirable for statistical parameter calculations. Based on the weather variable statistics, EPIC can generate weather patterns for long-run analyses (100+ years), or as indicated above, daily weather records (e.g., from world climate models with downscaling procedures) can be input directly. There is also the option of reading a sequence of actual daily weather and use generated weather afterwards or before within a simulation run.

The **topography** of a field or HRU is described by average field size (ha), slope length (m), and slope (%). In addition, elevation, longitude and latitude are needed for each site or HRU. Information on elevation, longitude and latitude is usually provided by a climate station of a digital elevation map.

A large number of physical and chemical **soil parameters** to describe each soil layer and subsequently an entire soil profile can be input in EPIC. These soil parameters (Tab.1) are separated between general and layer-specific as well as between essential and useful soil information requirement. The essential soil parameters are mandatory input while the remaining ones would help to further describe a soil specific profile situation. In EPIC, a soil profile can be split in up to 10 soil layers of which each is described by a specific set of chemical and physical soil parameters. If, for instance, the description of only two soil layers is available (top-soil and sub-soil), EPIC still allows (optional) to split the soil profile into ten soil layers. It assures that e.g., soil temperature and soil moisture can be appropriately estimated through soil layers and time.

Wide range of **management scheduling** in EPIC allows flexibility in modelling different cropping, forestation, and tillage systems (including crop rotations and crop mixes). However, it requires reliable information of the actual management practice for a given region or site. Generally, information on:

- Date of planting (including potential heat units the crop needs to reach maturity);
- Date, type (commercial; dairy, swine, etc. manure), and amount of fertilizer (elemental NPK) in kg/ha; if manure is applied, information on application rate (in case of grazing the stocking rate) and nutrient composition (orgN, minN (NO3-N + NH3-N), orgP, minP, minK, orgC, and fraction of NH3-N on minN) is needed;
- Date and amount of irrigation (including NO_3 and salt concentration in irrigation water) in mm;
- Date and type of tillage operation (plough, harrow spike, field cultivator, thinning, etc.); and
- Date and type of harvesting (combine, hay cutting, grazing, etc.) is needed.

Table 1: *List of physical and chemical soil parameters needed by EPIC*

ESSENTIAL SOIL INFORMATION	USEFUL SOIL INFORMATION
general soil and hydrologic information	
soil albedo (moist)	initial soil water content (fraction of field capacity)
hydrologic soil group (A, B, C, or D)	minimum depth to water table in m
	maximum depth to water table in m
	initial depth to water table in m
	initial ground water storage in mm
	maximum ground water storage in mm
	ground water residence time in days
	return flow fraction of water percolating through root
	soil weathering (CaCO3 soils; non-CaCO3 soils that
	number of years of cultivation
	soil group (kaolinitic, mixed, or smetitic)
	fraction of org C in biomass pool
	fraction of humus in passive pool
	soil weathering code
by soil layer	
depth from surface to bottom of soil layer in m	bulk density of the soil layer (oven dry) in t/m3
bulk density of the soil layer (moist) in t/m3	wilting point (1500 kPa for many soils) in m/m
sand content in %	field capacity (33 kPA for many soils) in m/m
silt content in %	Initial organic N concentration in g/t
soil pH	sum of bases in cmol/kg
organic carbon in %	cation exchange capacity in cmol/kg
calcium carbonate content in %	coarse fragment content in %vol.
	initial soluble N concentration in g/t
	initial soluble P concentration in g/t
	initial organic P concentration in g/t
	exchangeable K concentration in g/t
	crop residue in t/ha
	saturated conductivity in mm/h
	fraction of storage interacting with NO3 leaching
	phosphorous sorption ratio
	lateral hydraulic conductivity in mm/h
	electrical conductivity in mm/cm
	structural litter kg/ha
	metabolic litter kg/ha
	lignin content of structural litter in kg/ha
	carbon content of structural litter in kg/ha
	C content of metabolic litter in kg/ha
	C content of lignin of structural litter in kg/ha
	N content of lignin of structural litter in kg/ha
	C content of biomass in kg/ha
	C content of slow humus in kg/ha
	C content of passive humus kg/ha
	N content of structural litter in kg/ha
	N content of metabolic litter in kg/ha
	N content of biomass in kg/ha
	N content of slow humus in kg/ha
	N content of passive humus in kg/ha
	observed C content at the end of simulation

Appendix 2: Selected sources of national crop calendar data

COUNTRY	SOURCE
Afghanistan	*http://www.icarda.org/afghanistan/NA/Full/Primary_F.htm*
Angola	*http://www.fao.org/docrep/009/j8081e/j8081e00.htm*
Australia	*http://151.121.3.140/remote/aus_sas/crop_information/calendars/index_of_clndrs.htm*
Bangladesh	*http://151.121.3.140/remote/aus_sas/crop_information/calendars/index_of_clndrs.htm*
Burkina Faso	*http://v4.fews.net/docs/Publications/**Burkina**_200704en.pdf*
Canada	*http://www.fas.usda.gov/remote/canada/index.htm*
Côte d´Ivoire	*http://www.fas.usda.gov/pecad2/highlights/2002/09/West_Africa/pictures/ic_cropcalendar.htm*
Ethiopia	*http://www.fas.usda.gov/pecad2/highlights/2002/10/ethiopia/baseline/Eth_Crop_Calendar.htm*
India	*http://151.121.3.140/remote/aus_sas/crop_information/calendars/index_of_clndrs.htm*
Iran	*http://www.fas.usda.gov/remote/mideast_pecad/iran/iran.htm*
Iraq	*www.nationalaglawcenter.org/assets/crs/RL32093.pdf*
Kazakhstan	*http://www.fas.usda.gov/pecad2/highlights/2005/03/Kazakh_Ag/crop_cal.htm*
Kenya	*http://www.fas.usda.gov/pecad/highlights/2004/12/Kenya/images/crop_calendar.htm*
Madagascar	*http://www.wildmadagascar.org/maps/crops.html*
Malawi	*http://v4.fews.net/docs/Publications/South_200710en.pdf*
Mali	*http://v4.fews.net/docs/Publications/**Mali_2006**12en.pdf*
Mozambique	*http://v4.fews.net/docs/Publications/**Mozambique_2006**07en.pdf*
Nicaragua	*http://usda.gov/pecad2/articles/99-03/nic0399.pdf*
Nigeria	*www.fas.usda.gov/pecad2/highlights/2002/03/**nigeria**/pictures/**crop_calendar_nigeria**.pdf*
Russian Federation	*http://www.fas.usda.gov/remote/soviet/Crop_Calendars_2004/rs_districts.htm*
Rwanda	*www.fews.net/centers/files/**Rwanda_2000**09en.pdf*
Sudan	*http://www.sudan.net/statistic/sudancal.html*
Tanzania	*http://www.fas.usda.gov/pecad2/highlights/2003/03/tanzania/images/crop_calendar.htm*
Turkey	*http://www.fas.usda.gov/remote/mideast_pecad/turkey/turkey.htm*
Zambia	*http://v4.fews.net/docs/Publications/Zambia_200705en.pdf*
Zimbabwe	*http://www.fas.usda.gov/pecad2/highlights/2004/06/zimbabwe/images/zimbabwe_cropcalendar.htm*
Viet Nam	*www.cimmyt.org/english/docs/**maize**_producsys/**vietnam**.pdf*
Uganda	*www.foodnet.cgiar.org/inform/Fews/**fews00may**.pdf*

Appendix 3: SQL codes of select queries for preview and export the data

1. SQL code of *SpatialReference(DataPreview&EXPORT)* select query:

SELECT SpatialReference.Col_Row, SpatialReference.XCOORD, SpatialReference.YCOORD,
SpatialReference.ZoneID, SpatialReference.SimUID, SpatialReference.mCrpLnd, SpatialReference.mCrpLnd_H,
SpatialReference.mCrpLnd_L, SpatialReference.mCrpLnd_I, SpatialReference.mCrpLnd_S,
SpatialReference.mOthAgri, SpatialReference.mGrass, SpatialReference.mForest, SpatialReference.mWetLnd,
SpatialReference.mOthNatLnd, SpatialReference.mNotRel
FROM SpatialReference;

2. SQL code of *SimUstatistics(DataPreview&EXPORT)* select query:

SELECT SimUstatistics.SimUID, SimUstatistics.HRU, D_ColRow30.Col30, D_ColRow30.Row30,
SimUstatistics.COUNTRY, SimUstatistics.SimUArea, SimUstatistics.CrpLnd, SimUstatistics.CrpLnd_H,
SimUstatistics.CrpLnd_L, SimUstatistics.CrpLnd_I, SimUstatistics.CrpLnd_S, SimUstatistics.OthAgri,
SimUstatistics.Grass, SimUstatistics.Forest, SimUstatistics.WetLnd, SimUstatistics.OthNatLnd,
SimUstatistics.NotRel, SimUstatistics.mCrpLnd, SimUstatistics.mCrpLnd_H, SimUstatistics.mCrpLnd_L,
SimUstatistics.mCrpLnd_I, SimUstatistics.mCrpLnd_S, SimUstatistics.mOthAgri, SimUstatistics.mGrass,
SimUstatistics.mForest, SimUstatistics.mWetLnd, SimUstatistics.mOthNatLnd, SimUstatistics.mNotRel
FROM SimUstatistics, D_ColRow30
WHERE (((SimUstatistics.Col_Row30)=[d_colrow30].[col_row30]))
ORDER BY SimUstatistics.SimUID;

3. SQL code of *Topo&Soil_MineralSoil(DataPreview&EXPORT)* select query:

SELECT ZoneData.Zone_ID, ZoneData.HRU, D_ColRow30.Col30, D_ColRow30.Row30, ZoneData.ZoneLon,
ZoneData.ZoneLat, ZoneData.ZoneAlti, ZoneData.ZoneSlp, SoilAnal_MineralSoil.HG,
SoilAnal_MineralSoil.ALB, SoilAnal_MineralSoil.DEPTH1, SoilAnal_MineralSoil.DEPTH2,
SoilAnal_MineralSoil.DEPTH3, SoilAnal_MineralSoil.DEPTH4, SoilAnal_MineralSoil.DEPTH5,
SoilAnal_MineralSoil.VS1, SoilAnal_MineralSoil.VS2, SoilAnal_MineralSoil.VS3, SoilAnal_MineralSoil.VS4,
SoilAnal_MineralSoil.VS5, SoilAnal_MineralSoil.SAND1, SoilAnal_MineralSoil.SAND2,
SoilAnal_MineralSoil.SAND3, SoilAnal_MineralSoil.SAND4, SoilAnal_MineralSoil.SAND5,
SoilAnal_MineralSoil.SILT1, SoilAnal_MineralSoil.SILT2, SoilAnal_MineralSoil.SILT3,
SoilAnal_MineralSoil.SILT4, SoilAnal_MineralSoil.SILT5, SoilAnal_MineralSoil.CLAY1,
SoilAnal_MineralSoil.CLAY2, SoilAnal_MineralSoil.CLAY3, SoilAnal_MineralSoil.CLAY4,
SoilAnal_MineralSoil.CLAY5, SoilAnal_MineralSoil.BD1, SoilAnal_MineralSoil.BD2,
SoilAnal_MineralSoil.BD3, SoilAnal_MineralSoil.BD4, SoilAnal_MineralSoil.BD5,
SoilAnal_MineralSoil.CEC1, SoilAnal_MineralSoil.CEC2, SoilAnal_MineralSoil.CEC3,
SoilAnal_MineralSoil.CEC4, SoilAnal_MineralSoil.CEC5, SoilAnal_MineralSoil.SB1,
SoilAnal_MineralSoil.SB2, SoilAnal_MineralSoil.SB3, SoilAnal_MineralSoil.SB4, SoilAnal_MineralSoil.SB5,
SoilAnal_MineralSoil.PH1, SoilAnal_MineralSoil.PH2, SoilAnal_MineralSoil.PH3, SoilAnal_MineralSoil.PH4,
SoilAnal_MineralSoil.PH5, SoilAnal_MineralSoil.CARB1, SoilAnal_MineralSoil.CARB2,
SoilAnal_MineralSoil.CARB3, SoilAnal_MineralSoil.CARB4, SoilAnal_MineralSoil.CARB5,
SoilAnal_MineralSoil.CORG1, SoilAnal_MineralSoil.CORG2, SoilAnal_MineralSoil.CORG3,
SoilAnal_MineralSoil.CORG4, SoilAnal_MineralSoil.CORG5, SoilAnal_MineralSoil.FWC1,
SoilAnal_MineralSoil.FWC2, SoilAnal_MineralSoil.FWC3, SoilAnal_MineralSoil.FWC4,
SoilAnal_MineralSoil.FWC5, SoilAnal_MineralSoil.WP1, SoilAnal_MineralSoil.WP2,
SoilAnal_MineralSoil.WP3, SoilAnal_MineralSoil.WP4, SoilAnal_MineralSoil.WP5, SoilAnal_MineralSoil.KS1,
SoilAnal_MineralSoil.KS2, SoilAnal_MineralSoil.KS3, SoilAnal_MineralSoil.KS4, SoilAnal_MineralSoil.KS5,
ClimateReference.Tyndall_X, ClimateReference.Tyndall_Y
FROM ZoneData, SoilAnal_MineralSoil, ClimateReference, D_ColRow30
WHERE (((SoilAnal_MineralSoil.ZoneSTU)=[zonedata].[zonestu]) AND
((ClimateReference.Col_Row30)=[zonedata].[col_row30]) AND
((D_ColRow30.Col_Row30)=[zonedata].[col_row30]))
ORDER BY ZoneData.Zone_ID;

4. SQL code of *Topo&Soil_OrganicSoil(DataPreview&EXPORT)* select query:

SELECT ZoneData.Zone_ID, ZoneData.HRU, D_ColRow30.Col30, D_ColRow30.Row30, ZoneData.ZoneLon,
ZoneData.ZoneLat, ZoneData.ZoneAlti, ZoneData.ZoneSlp, SoilAnal_OrganicSoil.HG,
SoilAnal_OrganicSoil.ALB, SoilAnal_OrganicSoil.DEPTH1, SoilAnal_OrganicSoil.DEPTH2,
SoilAnal_OrganicSoil.DEPTH3, SoilAnal_OrganicSoil.DEPTH4, SoilAnal_OrganicSoil.DEPTH5,
SoilAnal_OrganicSoil.VS1, SoilAnal_OrganicSoil.VS2, SoilAnal_OrganicSoil.VS3, SoilAnal_OrganicSoil.VS4,
SoilAnal_OrganicSoil.VS5, SoilAnal_OrganicSoil.SAND1, SoilAnal_OrganicSoil.SAND2,
SoilAnal_OrganicSoil.SAND3, SoilAnal_OrganicSoil.SAND4, SoilAnal_OrganicSoil.SAND5,
SoilAnal_OrganicSoil.SILT1, SoilAnal_OrganicSoil.SILT2, SoilAnal_OrganicSoil.SILT3,
SoilAnal_OrganicSoil.SILT4, SoilAnal_OrganicSoil.SILT5, SoilAnal_OrganicSoil.CLAY1,
SoilAnal_OrganicSoil.CLAY2, SoilAnal_OrganicSoil.CLAY3, SoilAnal_OrganicSoil.CLAY4,
SoilAnal_OrganicSoil.CLAY5, SoilAnal_OrganicSoil.BD1, SoilAnal_OrganicSoil.BD2,
SoilAnal_OrganicSoil.BD3, SoilAnal_OrganicSoil.BD4, SoilAnal_OrganicSoil.BD5,
SoilAnal_OrganicSoil.CEC1, SoilAnal_OrganicSoil.CEC2, SoilAnal_OrganicSoil.CEC3,
SoilAnal_OrganicSoil.CEC4, SoilAnal_OrganicSoil.CEC5, SoilAnal_OrganicSoil.SB1,
SoilAnal_OrganicSoil.SB2, SoilAnal_OrganicSoil.SB3, SoilAnal_OrganicSoil.SB4, SoilAnal_OrganicSoil.SB5,
SoilAnal_OrganicSoil.PH1, SoilAnal_OrganicSoil.PH2, SoilAnal_OrganicSoil.PH3,
SoilAnal_OrganicSoil.PH4, SoilAnal_OrganicSoil.PH5, SoilAnal_OrganicSoil.CARB1,
SoilAnal_OrganicSoil.CARB2, SoilAnal_OrganicSoil.CARB3, SoilAnal_OrganicSoil.CARB4,
SoilAnal_OrganicSoil.CARB5, SoilAnal_OrganicSoil.CORG1, SoilAnal_OrganicSoil.CORG2,
SoilAnal_OrganicSoil.CORG3, SoilAnal_OrganicSoil.CORG4, SoilAnal_OrganicSoil.CORG5,
SoilAnal_OrganicSoil.FWC1, SoilAnal_OrganicSoil.FWC2, SoilAnal_OrganicSoil.FWC3,
SoilAnal_OrganicSoil.FWC4, SoilAnal_OrganicSoil.FWC5, SoilAnal_OrganicSoil.WP1,
SoilAnal_OrganicSoil.WP2, SoilAnal_OrganicSoil.WP3, SoilAnal_OrganicSoil.WP4,
SoilAnal_OrganicSoil.WP5, SoilAnal_OrganicSoil.KS1, SoilAnal_OrganicSoil.KS2, SoilAnal_OrganicSoil.KS3,
SoilAnal_OrganicSoil.KS4, SoilAnal_OrganicSoil.KS5, ClimateReference.Tyndall_X,
ClimateReference.Tyndall_Y
FROM ZoneData, SoilAnal_OrganicSoil, ClimateReference, D_ColRow30
WHERE (((SoilAnal_OrganicSoil.ZoneSTU)=[zonedata].[zonestu]) AND
((ClimateReference.Col_Row30)=[zonedata].[col_row30]) AND
((D_ColRow30.Col_Row30)=[ZoneData].[Col_Row30]))
ORDER BY ZoneData.Zone_ID;

5. SQL code of *CropCalendar(DataPreview&EXPORT)* select query:

SELECT CropCalendar.COUNTRY, CropCalendar.CROP, CropCalendar.CrpCalSpec, CropCalendar.OPER,
CropCalendar.YEA, CropCalendar.DAY, CropCalendar.DAT, CropCalendar.mHI, CropCalendar.mIR,
CropCalendar.mLI, CropCalendar.mSS
FROM CropCalendar
ORDER BY CropCalendar.COUNTRY, CropCalendar.CROP, CropCalendar.CrpCalSpec, CropCalendar.DAT;

6. SQL code of *CropShare(DataPreview&EXPORT)* select query:

SELECT CropShare.SimUID, D_ColRow30.Col30, D_ColRow30.Row30, CropShare.CrpCalSpec,
CropShare.CROP, CropShare.CultType, CropShare.RotType, CropShare.CrpArea, CropShare.SimUmask,
CropShare.absSimUmask, CropShare.CropMask, CropShare.AltMask, CropShare.MngSystem
FROM CropShare, D_ColRow30
WHERE (((CropShare.MngSystem)="HI") AND ((CropShare.Col_Row30)=[d_colrow30].[col_row30]))
ORDER BY CropShare.SimUID;

7. SQL code of *ManureApplication(DataPreview&EXPORT)* select query:

SELECT ManureApplication.SimUID, ManureApplication.HRU, D_ColRow30.Col30, D_ColRow30.Row30,
ManureApplication.COUNTRY, ManureApplication.Nrate, ManureApplication.Manure
FROM ManureApplication, D_ColRow30

WHERE (((D_ColRow30.Col_Row30)=[manureapplication].[col_row30]))
ORDER BY ManureApplication.HRU;