



---

# A new Global Agro-Environmental Stratification (GAES)

Sander Mûcher, Lorenzo de Simone, Henk Kramer, Allard de Wit, Laure Roupioz, Gerard Hazeu, Hendrik Boogaard, Rini Schuiling, Steffen Fritz, John Latham, and Anouk Cormont



---

# A new Global Agro-Environmental Stratification (GAES)

Sander Mächer<sup>1</sup>, Lorenzo de Simone<sup>2</sup>, Henk Kramer<sup>1</sup>, Allard de Wit<sup>1</sup>, Laure Roupioz<sup>1</sup>, Gerard Hazeu<sup>1</sup>, Hendrik Boogaard<sup>1</sup>, Rini Schuiling<sup>1</sup>, Steffen Fritz<sup>3</sup>, John Latham<sup>2</sup>, and Anouk Cormont<sup>1</sup>

1 Wageningen Environmental Research (Alterra)

2 Food and Agriculture Organization of the United Nations (FAO)

3 International Institute for Applied Systems Analysis (IIASA)

This research was conducted by Wageningen Environmental Research (Alterra) and funded by the European Commission (FP7, Grant agreement no. 603719) and the Dutch Ministry of Economic Affairs (Topsector Agri&Food, BO-31.03-0007-003).

Wageningen Environmental Research  
Wageningen, December 2016

---

Report 2761  
ISSN 1566-7197

---

Mücher, S., L. De Simone, H. Kramer, A. De Wit, L. Roupioz, G. Hazeu, H. Boogaard, R. Schuiling, S. Fritz, J. Latham, and A. Cormont, 2016. *A new Global Agro-Environmental Stratification (GAES)*. Wageningen, Wageningen Environmental Research, Report 2761. 70 pp.; 46 fig.; 12 tab.; 71 ref.

The GAES database (Version 01a) is a newly developed Global Agro-Environmental Stratification within the EU SIGMA (Stimulating Innovation for Global Monitoring of Agriculture) project. GAES will serve as a new agro-environmental stratification for better global monitoring of the agricultural production on the basis of Earth Observation data and crop growth models. It is anticipated that GAES will be exploited for a wider range of applications, some within SIGMA, towards data gap analysis that identifies agro-environmental strata with limited capacity and monitoring data on agricultural production. GAES was produced by applying segmentation techniques to newly available global agro-environmental data with a high spatial resolution re-sampled to 1 km spatial resolution. The datasets were able to stratify the agricultural production zones according to the region's agro-environmental characteristics, including climatic regimes, soil, terrain, elevation conditions, water availability and land cover proprieties. The GAES strata obtained by segmentation at four different spatial levels (with Level 4 as the most detailed) have been further characterised and described in terms of phenology (e.g. start and peak of the growing season), agricultural (water) management practices, field size, biotic constraints, national and sub-national crop production statistics, GDP, transport infrastructure conditions or market accessibility. The GAES database has four hierarchical layers, with 92 attributes. GAES Level 1 has 194 agro-environmental (AE) types (818 strata); GAES Level 2 has 300 AE types (1,688 strata); GAES Level 3 has 374 AE types (2,087 strata); GAES Level 4 has 516 AE types (3,208 strata). GAES typology is a combination of temperature, altitude, parent material and land cover characteristics. GAES Version 01 has become freely available.

Keywords: global zonation, agricultural production, monitoring, earth observation data.

The pdf file is free of charge and can be downloaded at <http://dx.doi.org/10.18174/400815> or via the website [www.wur.nl/environmental-research](http://www.wur.nl/environmental-research) (scroll down to Publications – Wageningen Environmental Research reports). Wageningen Environmental Research does not deliver printed versions of the Wageningen Environmental Research reports.

© 2016 Wageningen Environmental Research (an institute under the auspices of the Stichting Wageningen Research), P.O. Box 47, 6700 AA Wageningen, The Netherlands, T +31 (0)317 48 07 00, E [info.alterra@wur.nl](mailto:info.alterra@wur.nl), [www.wur.nl/environmental-research](http://www.wur.nl/environmental-research). Wageningen Environmental Research is part of Wageningen University & Research.

- Acquisition, duplication and transmission of this publication is permitted with clear acknowledgement of the source.
- Acquisition, duplication and transmission is not permitted for commercial purposes and/or monetary gain.
- Acquisition, duplication and transmission is not permitted of any parts of this publication for which the copyrights clearly rest with other parties and/or are reserved.

Wageningen Environmental Research assumes no liability for any losses resulting from the use of the research results or recommendations in this report.

Wageningen Environmental Research Report 2761 | ISSN 1566-7197

Photo cover: Wheat fields in the Ukraine

---

# Contents

<b>1</b>	<b>Introduction SIGMA</b>	<b>5</b>
1.1	Objective WP3.1	6
<b>2</b>	<b>Literature review</b>	<b>8</b>
2.1	Introduction	8
2.2	Theoretical background and state-of-the-art	9
2.2.1	Definition	9
2.2.2	Global Agro-Environmental Zone (GAEZ)	9
2.2.3	Global Environmental Stratification (Gens)	10
2.2.4	Sustainable And Global Environment scheme (SAGE)	10
2.2.5	Global Yield Gap Atlas (GYGA-ED)	10
2.2.6	Schemes derived from GAEZ and SAGE: HCAEZ and GLI	10
2.2.7	IIASA gridded EPIC	10
2.3	Objectives current stratifications	11
2.3.1	Global Agro-Environmental Zone (GAEZ)	11
2.3.2	Global Environmental Stratification (GEnS)	11
2.3.3	Sustainable And Global Environment scheme (SAGE)	11
2.3.4	Global Yield Gap Atlas (GYGA-ED)	12
2.3.5	Schemes derived from GAEZ and SAGE: HCAEZ and GLI	12
2.3.6	IIASA gridded EPIC	12
2.4	Stratification methods and data	12
2.4.1	Global Agro-Environmental Zone (GAEZ)	12
2.4.2	Global Environmental Stratification (GEnS)	13
2.4.3	Sustainable And Global Environment Scheme (SAGE)	13
2.4.4	Global Yield Gap Atlas (GYGA-ED)	14
2.4.5	Schemes derived from GAEZ and SAGE: HCAEZ and GLI	14
2.4.6	IIASA gridded EPIC	14
2.4.7	Use of remote sensing in stratification schemes	16
2.5	Stratification applications and evaluations	16
2.6	Conclusions	18
<b>3</b>	<b>Input datasets</b>	<b>19</b>
3.1	Framework for the selection of input data	19
3.2	Climate	21
3.2.1	WorldClim global climate dataset	21
3.2.2	Cloud fraction	23
3.2.3	Results	26
3.2.4	Conclusions	29
3.3	Slope and diversity in altitude	31
3.4	Soil productivity	32
3.4.1	Nutrient availability (SQ1)	33
3.4.2	Nutrient retention capacity (SQ2)	33
3.4.3	Rooting conditions (SQ3)	33
3.4.4	Oxygen availability (SQ4)	34
3.4.5	Excess salts (SQ5)	34
3.4.6	Toxicities (SQ6)	34
3.4.7	Workability (SQ7)	35
3.4.8	Overlay model	35
3.5	Gross Primary Production (GPP)	36
3.6	Crop types	37

---

3.7	Phenology	38
	3.7.1 Phenodata	38
3.8	Number of growing cycles	39
	3.8.1 Global Water Satisfaction Index (GWSI)	40
	3.8.2 Integration Phenodat with GWSI	40
3.9	Field size	42
3.10	Irrigated agriculture	43
	3.10.1 Global Map of Irrigation Areas (GMIA)	43
3.11	Crop land mask	43
	3.11.1 IIASA hybrid crop land layer	43
	3.11.2 FAO GLC-SHARE	45
	3.11.3 Integration of IIASA crop land mask and FAO GLC-SHARE	46
<b>4</b>	<b>Global stratification methodology</b>	<b>47</b>
	4.1 Global input data layers	47
	4.2 Isodata clustering	50
	4.3 Segmentation	51
<b>5</b>	<b>Results: a Global Agro-Environmental Stratification (GAES)</b>	<b>54</b>
	5.1 GAES.gdb: data gathering, input data and preprocessing	55
	5.2 Analysis, models and tools	56
	5.3 GAES statistical outputs	57
	5.3.1 Joining outputs to strata	58
	5.4 GAES final results	59
	5.5 GAES dissemination and outreach	62
	<b>References</b>	<b>63</b>
	<b>Annex 1 Monthly cloud fraction across latitudes</b>	<b>68</b>

---

# 1 Introduction SIGMA

Sustainable agricultural growth is a critical component in efforts to meet the demands and challenges faced by agriculture worldwide and discover new opportunities for poverty reduction in the developing and transitional world. Agriculture's capacity to feed the world is under threat from a combination of existing- and emerging trends and challenges, whilst global hunger and malnutrition remain pervasive. As the global population is anticipated to reach more than nine billion by 2050, food production will need to grow by 70% worldwide, and up to 100% in developing countries, according to the United Nations (UN) (2015). Sustainable and well-balanced planning and management of agricultural resources are essential to achieve this. In this regard, the Earth Observation Community can contribute to some extent by improving the monitoring capacities on agricultural production.

In 2002, the Group on Earth Observations (GEO) was formed at the Summit on Sustainable Development in South Africa. Its main vision is to build a Global Earth Observation System of Systems (GEOSS) through globally coordinated activities based on remote sensing. Sustainable agriculture is one of its core domains. In June 2011, G20 launched its 'Global Agricultural Geo-Monitoring' (GEOGLAM) and the Agricultural Market Information System (AMIS) initiatives. The main objective of GEOGLAM is to improve crop forecasts and, thereby, to increase transparency on agricultural production, through the creation of an operational global agricultural monitoring 'system of systems' based on Earth Observation and *in situ* observations.

SIGMA, an FP7 project, is part of Europe's contribution to GEOGLAM. It involves actively networking expert organisations, united worldwide, in a common effort to enhance current remote sensing-based agricultural monitoring techniques. The project aims to develop innovative methods for agricultural monitoring, by focusing in particular on sustainability, and interaction with the environment.

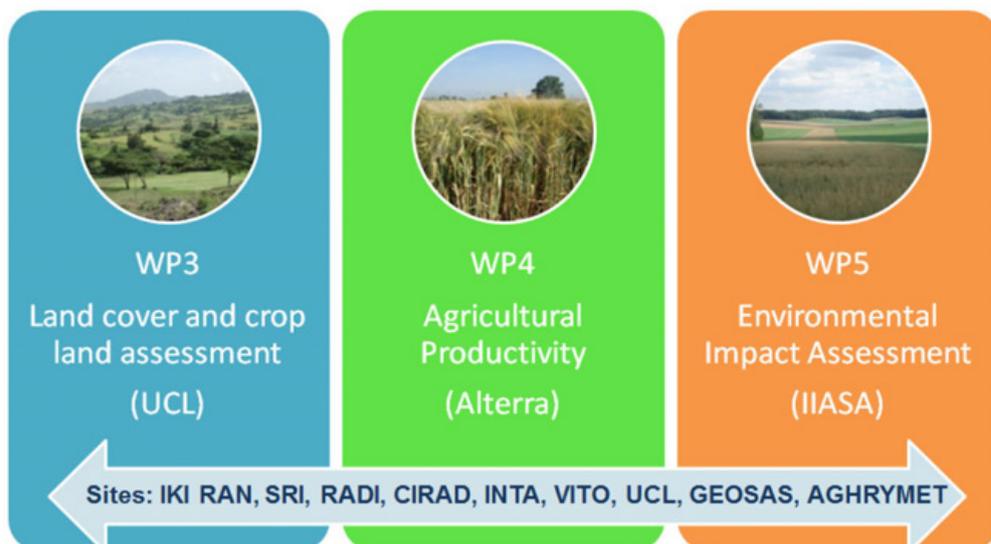
The activities are implemented in the European Union, Russia, Ukraine, China, Vietnam, Argentina Brazil and Sub Saharan African countries.

One aspect that the project addresses is the way that 'global' agricultural production is assessed through relevant indicators, in particular, those describing:

- Agricultural Expansion and crop land dynamics and their impact on other ecosystems, GHG emissions, and biodiversity in general.
- Agricultural Intensification: including shifts in cultivation practises, including cropping patterns, irrigation, seed breeding and their impact on soil fertility, erosion, pollution and water use efficiency.

In summary, SIGMA intends to develop methods and products addressing agricultural sustainability issues, including:

1. How and where do changes in crop land distribution affect other ecosystems?
2. How and where do changes in cropping practices create environmental impacts in the region?
3. How can we ensure integration of developed methods in global monitoring systems?



**Figure 1** Main activities of SIGMA. WP3.1 is part of the overall WP3 'Land cover and crop land assessment'

This deliverable summarises activities covered by work package (WP) 3.1 'Data Gap Analysis and Global Agro-Environmental Stratification' with focus on the production of Global Agro-Environmental Stratification that will help to produce a global cropland map within SIGMA.

## 1.1 Objective WP3.1

The overall objective of this WP is to compile relevant agro-ecological and socio-economic data towards the creation of a global map of agro-environmental strata. This map will be used as:

1. An analytical framework for data gap analysis; and
2. Direct input for WP3.2 on crop land classifications and land cover change.

The Global Agro-Environmental Stratification (GAES) based on environmental data sets and Earth Observation will be used as an analytical framework to assess agricultural resources, dynamics and potentials investigated in the other WPs. The data sets should stratify agricultural production zones based on segmentation of the environmental data sets into relatively homogeneous regions (depending on scale) of climatic regimes, soil, terrain, elevation conditions, water availability and land cover characteristics. In a second stage, the strata of the GAES will be further characterised and described in terms of: crop phenology (e.g. start and peak of growing season); agricultural (water) management practices; field size; biotic constraints; local agricultural- and other economic statistics (e.g. GDP); transport infrastructure conditions and market accessibility. The starting point for this work is the former Global Agro-Ecological Zones (GAEZ) from FAO and IIASA, which provides geo-spatial information on crop production. It includes land suitability and potential attainable yields under different management (irrigation) conditions. GAEZ enables the harmonization and integration of a variety of databases, model scenarios and assessments in both time and space, and is becoming a fundamental tool in land use planning and management, and sustainable development that addresses food security. It also prepares for a more reliable FAO perspective analysis and offers policy options within the context of international conventions and agreements on various areas of interest (e.g. natural resources monitoring and management, climate change, biodiversity and land degradation). GAEZ is used in combination with a selected number of newly available, high resolution data sets used to prepare the GAES at a 1km spatial resolution. Subsequently, a data gap analysis on basis of the GAES is performed to highlight specific uncertainties in the global agricultural monitoring and aims to assist prioritization of research activities to fill the existing gaps in *in situ* monitoring of agricultural production. In close cooperation with other WPs, WP3.1 has three main objectives:

1. To set-up Global Agro-Environmental Stratification (GAES) as a framework to map global cropland and further analysis of agricultural production.

- 
2. To further characterise and describe the strata to enhance activities in the other WPs, such as the regionalisation of yield gaps.
  3. To undertake a data gap analysis to highlight specific, *in situ* uncertainties that affect global agricultural monitoring.

The work is divided into various tasks and activities: Task 1: Global Agro-Environmental Stratification (GAES) coordinated by Wageningen Environmental Research (Alterra); Task 2: Further characterisation and description of the GAES coordinated by FAO; Task 3: Data Gap Analysis coordinated by IIASA. Task 3 is not part of this report and will be described in a subsequent report D31.2 'Report on data gap analysis'.

---

## 2 Literature review

Accurate estimation of agricultural potential and actual production at global scale is essential to cope with the increasing demand for food that will occur in the coming decades. Several studies have been developed for agricultural monitoring, through the estimation of food production capacity and yield gap at local level. Using data acquired from long-term weather, crop rotations, management practices and soil properties, those estimations were also made available at global level. To do so, global agro-environmental zonation schemes were used to up-scale location-specific estimates of relevant parameters, allowing evaluation of global agricultural production potential.

Establishment of a suitable Global Agro-Environmental Stratification (GAES) is an important step towards enabling global monitoring of agriculture, as per the aims within the SIGMA project. The former Global Agro-Ecological Zones (GAEZ) was created by FAO and IIASA, but does not provide global stratification. GAEZ forms a starting point for our work. GAEZ contains a large number of data sets available at global scale that can be used as inputs, directly or following update, to build the GAES. The literature review presented in this report aims to provide information that allows us to identify the most relevant method(s) to construct the GAES by providing a good overview of the major existing global stratification schemes. The background, objective, methodology and evaluation of each stratification scheme are presented. The different schemes are compared and evaluated in relation to the goals of SIGMA. Finally, this review leads to conclusions about the best way to develop the stratification to be performed for SIGMA.

### 2.1 Introduction

The need to increase agricultural production to respond to the growing demand for food is one of the main concerns for the coming decades (Godfray, Beddington et al. 2010). In this context, the accurate estimation at global scale of actual and potential crop yields is fundamental to identify yield gaps and determine where the yields of already cultivated lands could be increased. Food production capacity and yield gap are already used broadly at local level; as derived from research plots or simulation models. More recently, several studies have proposed methods to estimate agricultural production potential at the global level using data derived from long-term weather, crop rotations, management practices and soil properties. The methodology applied in those studies is based on the use of global agro-environmental zonation schemes to up-scale, location-specific estimates of parameters that allow evaluation of agricultural production potential, such as yield potential and water limited yield potential, which are the basis for estimating yield gaps at regional-, national-, and global scales (Van Wart, van Bussel et al. 2013).

SIGMA aims to develop innovative methods, based upon the integration of *in situ* and earth observation data, to enable the prediction of the impact of crop production on ecosystems and natural resources. In SIGMA, the main objective of WP31 is to compile relevant data in order to create a global map of agro-environmental strata, with improved zonations based on mainly agro-ecological factors. This Global Agro-Environmental Stratification (GAES) will serve as the basis for better global monitoring of the agricultural production. The first step of this work comprises of an extensive literature review that gives an overview of the existing global zonation schemes considering their background and objectives, the data used, as well as their current applications and evaluations. The former Global Agro-Ecological Zones (GAEZ) provides a large amount of agro-environmental datasets available at global scale have been designated as the starting point for this work. Furthermore, collecting information on already existing global stratification schemes, along with feedbacks and experiences accumulated after few years of development in that domain, enables us to build upon existing knowledge to create the GAES to be used in SIGMA, and to improve stratification methods for agriculture monitoring purposes. The literature review then provides the information required to refine the specifications of the precise aim of the stratification and to set-up a Global Agro-Environmental Stratification as an analytical framework.

---

Section 2.2 provides a general definition of an Agro-Environmental Zone (AEZ), along with a brief presentation of the main existing global stratification schemes, as a basis for common understanding of the stratification concept. Section 2.3 details the objectives of each stratification, whilst Section 2.4 details the methodology and data used for zoning. A summary of the evaluation and actual applications of stratification schemes found in literature is presented in Section 2.5. Finally, Section 2.6 summarises the findings and points out potential next steps.

## 2.2 Theoretical background and state-of-the-art

### 2.2.1 Definition

An Agro-Ecological Zone is a land resource mapping unit, defined in terms of climate, landform and soils, and/or land cover, and having a specific range of potentials and constraints for land use (FAO 1996). For our purposes, an Agro-Ecological Zone can be considered as a geographical area that exhibits sufficiently homogeneous ranges of key climate and biophysical variables as to make it useful in agricultural production estimation. This notion of an Agro-Ecological Zone has a number of features that complicate its implementation in practice. These include:

- Determining the key variables that might be appropriate and under what circumstances.
- Determining 'sufficiency' of homogeneity.
- Characterising and physically locating the relevant boundaries of agro-ecological zones.

### 2.2.2 Global Agro-Environmental Zone (GAEZ)

The AEZ methodology has been developed by the Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA). Rapid developments in information technology have produced increasingly detailed and manifold global databases, which made the first global AEZ assessment possible in 2000. Since then, global AEZ assessments have been performed every few years. With each update of the system, the issues addressed, the size of the database, and the number of results, have multiplied (Tóth, Kozłowski et al. 2012).

The entire database consists of hundreds of data layers and covers five thematic areas:

- *Land and water resources*, including soil resources, terrain resources, land cover, protected areas and selected socio-economic and demographic data;
- *Agro-climatic resources*, including a variety of climatic indicators;
- *Suitability and potential yields* for up to 280 crops/land utilisation types under alternative input and management levels for historical-, current- and future climatic conditions;
- Downscaled actual yields and production of main crop commodities;
- *Yield and production gaps*, in terms of ratios and differences between actual yield and production and potential for main crops.

The GAEZ database can be used for various applications including the quantification of land productivity. Results are commonly aggregated for current major land use/cover patterns and by administrative units, land protection status, or broad classes reflecting infrastructure availability and market access conditions (Fischer, Nachtergaele et al. 2012). It is important to underline that the GAEZ is not a zonation scheme, as such. It is a database with an individual data layer that can be used to create different zonation scheme, e.g. IGAEZ (Tatsumi, Yamashiki et al. 2011), M-GAEZ (Hasegawa, Fujimori et al. 2013), GAEZ-LGP (Fischer, Nachtergaele et al. 2012).

---

### 2.2.3 Global Environmental Stratification (Gens)

Global Environmental Stratification (GEnS), presented in Metzger et al. 2013b, is the first cluster methodology that aims to establish a global, climate-explicit zonation system. The GEnS is a statistically derived global bioclimatic classification, providing a novel global spatial framework for the integration and analysis of ecological and environmental data. The dataset consists of 125 strata, aggregated into 18 global environmental zones (Metzger, Bunce et al. 2011). The stratification has a 30' resolution (equivalent to 1km<sup>2</sup> at the Equator). GEnS was developed within the Group on Earth Observations Biodiversity Observation Network (GEOBON) and will be available to assist in further research on global ecosystems.

### 2.2.4 Sustainable And Global Environment scheme (SAGE)

This zonation scheme follows the logic of Prentice et al. (Prentice, Cramer et al. 1992), who described relationships between climate and global biomes. The SAGE zonation scheme is using two parameters known to be fundamental drivers of plant growth to describe a region's climate – growing degree days (GDD) and a crop soil moisture index (the ratio of actual evapotranspiration to potential evapotranspiration) (Licker, Johnston et al. 2010). It has generated 100 different climate zone combinations.

### 2.2.5 Global Yield Gap Atlas (GYGA-ED)

Previous yield gap analyses are too coarse or too empirical to serve the targeted objectives, and they lack transparency with respect to methods, sources of data, and underpinning assumptions, which makes it difficult to validate estimates or improve upon them (Van Ittersum, Cassman et al. 2013). Starting from this observation and the fact that for irrigated systems and rainfed systems, respectively, potential yield (Y<sub>p</sub>) and water-limited yield (Y<sub>w</sub>) are the relevant benchmarks for yield gap analysis (<http://www.yieldgap.org>). Unlike other efforts to estimate yield gap that rely on gridded weather data, GYGA seeks to use a 'bottom-up' approach with location-specific observed weather data (Van Wart, van Bussel et al. 2013). To extrapolate results from location-specific observed data, the GYGA approach utilises a hybrid zonation scheme, called the GYGA Extrapolation Domain (GYGA-ED), which combines components of the other zonation schemes. The challenge of using a bottom-up approach is the time, expense and access needed to acquire observed weather data, as well as associated location-specific information about crop rotations, soil properties and farm management, which are required for robust estimates of yield potential and water limited yield potential (Van Ittersum, Cassman et al. 2013). Therefore, the GYGA approach strives for a zonation scheme that balances the need to minimise the number of location-specific sites that require weather-, soil-, and crop management data, with the goal of minimising climatic heterogeneity within the climate zones.

### 2.2.6 Schemes derived from GAEZ and SAGE: HCAEZ and GLI

The Harvest Choice AEZ scheme (HCAEZ), developed for analyses in Sub-Saharan Africa, comprises of modified aspects of both the SAGE and GAEZ (Wood, Sebastian et al. 2010). It is a matrix with 21 zones based on GAEZ-LGP and thermal regime classes for the tropics, sub-tropics, temperate, and boreal zones, distinguished by highland- and lowland regions. The zonation scheme proposed by the Global Landscapes Initiative (GLI) Group at the University of Minnesota in the US, is derived from the SAGE zonation scheme (Mueller, Gerber et al. 2012).

### 2.2.7 IIASA gridded EPIC

IIASA gridded EPIC (Balkovic et al. 2014) is the representative of an approach that is actually implemented within many other systems that are used increasingly for crop growth modelling at continental- or global scales (e.g. Folberth et al., 2012; Liu, 2009; Liu et al., 2007, Stehfest et al., 2007; Tan and Shibasaki, 2003), or integrated within more complex, agricultural-sector models (e.g. Leip et al., 2008). Originally field-scale, process-based crop models (e.g. EPIC, DNDC, DayCent) are coupled with geographical data on key environmental factors, such as climate, topography, soil, and land use. Model and data fusion then yields in spatially explicit estimates of crop biomass (including

effective yield) over a given spatial domain and time period based on pre-defined yield-forming parameters (model inputs) and their interactions (process-based model itself). IIASA gridded EPIC integrates the EPIC model (Williams, 1995), which has been set up at global scale with spatial detail of 5 arc min (10\*10 km at the equator), and calibrated for 16 major crops worldwide. The EU27 regional implementation of IIASA gridded EPIC with 1\*1k resolution has also been established and calibrated for 12 major crops (Balkovic et al., 2013).

**Table 1** Global stratification schemes overview

Stratification scheme	Construction	Resolution	Crop area (CA) or whole terrestrial area (WTA)	Number of zones
GAEZ	Matrix	10*10km	WTA	-
GAEZ-LPG	Matrix	10*10km	WTA	16
GEnS	Cluster	1*1km	WTA	125
HCAEZ	Matrix	10*10km	WTA	21
SAGE	Matrix	10*10km	WTA	100
GLI	Matrix	10*10km	CA	100
GYGA-ED	Matrix	10*10km	CA	265
IIASA EPIC	Matrix	10*10km	CA	131.326

## 2.3 Objectives current stratifications

### 2.3.1 Global Agro-Environmental Zone (GAEZ)

The Agro-Ecological Zones (AEZ) methodology has been developed over the past 30 years for assessing agricultural resources and potential. The GAEZ modelling framework was developed to quantify land productivity, to spatially analyse agricultural systems and to evaluate the impacts of agricultural policies at a global scale (Fischer, Hizznyik et al. 2009). Of the various AEZ schemes used in the GAEZ framework, GAEZ-LGP is based on the Length of the Growth Period (LGP). It utilises the most agronomically relevant categorical variables and has the smallest, and presumably, the most climatically homogenous zones within the GAEZ-family of AEZ schemes (Van Wart, van Bussel et al. 2013).

### 2.3.2 Global Environmental Stratification (GEnS)

The objective of the GEnS is to develop a consistent quantitative stratification of the land surface of the world into relatively homogeneous bioclimatic strata to provide a global spatial framework for the integration and analysis of ecological and environmental data (Metzger, Bunce et al. 2011). It aims at providing a framework for coordination and analysis of global biodiversity observation efforts and research, aggregating observations and for the comparison of trends within similar environments, and will be publicly available to support global ecosystem research and monitoring.

The data set provides a generic classification that can be adapted for a specific objective, as well as providing suitable zonation for environmental reporting (Metzger, Shkaruba et al. 2012).

### 2.3.3 Sustainable And Global Environment scheme (SAGE)

The SAGE zonation scheme was developed to determine within-zone maximum yield for specific crops within each of the 100 zones. If the zonal-maximum yield was exceeded observed yields for a particular region within the zone, the authors considered this a yield gap and identified the region as offering the opportunity for increasing yields (Licker, Johnston et al. 2010).

### 2.3.4 Global Yield Gap Atlas (GYGA-ED)

The goal of the GYGA project is to estimate the yield gap for major food crops in all crop-producing countries based on locally observed data ([www.yieldgap.org](http://www.yieldgap.org)). Methods used to estimate yield gaps in the Global Yield Gap Atlas (GYGA) were developed to improve:

1. Prioritisation of public- and private sector investments in agricultural research and development.
2. National forecasts of food production capacity, potential for food self-sufficiency, and variability of food production capacity due to variation in weather.
3. Interpretation of historical yield trends and projections of future yield trends at regional-, national- and global scales.

### 2.3.5 Schemes derived from GAEZ and SAGE: HCAEZ and GLI

The HCAEZ scheme was developed to improved AEZ schemes for yield gap analysis and the GLI scheme was used for yield gap estimate.

### 2.3.6 IIASA gridded EPIC

Along with the basic objective of the stratification, which is to provide homogenous spatial units (as for their climate, topography, soil, and land use) for the process-based model, Homogenous Response Units (HRU) and Simulation Units (SimU) based spatial stratifications implemented within the IIASA gridded EPIC are considered to provide a direct link between spatially explicit crop production and other environmental parameters coming out of the process-based model and socio-economical land use optimisation models, such as EU FASOM, or GLOBIOM at coarser resolutions (e.g. Schneider et al. 2011, Havlik et al., 2011).

**Table 2** Global stratification schemes objectives summary

Stratification scheme	Agricultural potential/Land productivity	Agricultural system	Impact of agricultural policies	Global ecosystem research and monitoring	Yield gap analysis	Yield potential
GAEZ	x	x	x		x	x
GEnS				x		
HCAEZ					x	x
SAGE					x	x
GLI					x	x
GYGA-ED					x	x
IIASA EPIC	x		x		x	x

## 2.4 Stratification methods and data

### 2.4.1 Global Agro-Environmental Zone (GAEZ)

The GAEZ is a framework that provides data sets that can be used to create a zonation scheme, but it is not a zonation scheme, as such. The database has been used to create different zonation schemes, such as the GAEZ-LGP, in which LGP is derived from temperature, precipitation, and soil water-holding capacity, as categorical variables.

Categorical variables used, or derived, from the data listed in the previous table are:

- a. Accumulated temperature sums for mean daily temperature above a base temperature [Growing Degree Days (GDD)].
- b. Annual temperature profiles, based on mean annual temperature and within-year temperature trends.

- c. Delineation of continuous, discontinuous, sporadic and no permafrost zones.
- d. Quantification of soil water balance and actual evapotranspiration for a reference crop.
- e. Length of Growing Period (LGP), defined as the sum of days when mean daily temperature exceeds 5°C and evapotranspiration for the reference crop exceeds half of the potential evapotranspiration.
- f. Multiple cropping classification, which indicates if annual single, double or triple cropping is possible in a given zone, based on the LGP, and assuming a growth duration per crop of 120 days.
- g. (See user manual for the whole GAEZ database: (Fischer, Nachtergaele et al. 2012)).

#### 2.4.2 Global Environmental Stratification (GEnS)

A broad set of climate-related variables were considered for inclusion within a quantitative model that partitions geographic space into bioclimatic regions. Statistical screening produced a subset of relevant bioclimatic variables, which were further compacted into fewer independent dimensions using Principal Components Analysis (PCA). An ISODATA clustering routine was then used to classify the principal components into relatively homogenous environmental strata. The strata were aggregated into global environmental zones on the basis of attribute distances between strata to provide structure and support a consistent nomenclature (Metzger, Bunce et al. 2011). This cluster zonation uses monthly gridded climate data from the WorldClim database, and annual aridity and potential evapotranspiration seasonality derived from the CGIAR Consortium for Spatial Information, with 30' resolution (approximately 1 km<sup>2</sup> at the equator)(Van Wart, van Bussel et al. 2013). The Global Environmental Stratification was constructed using rigorous statistical procedures. It provides a robust spatial analytical framework for the aggregation of local observations, identification of gaps in current monitoring efforts, and systematic design of complementary and new monitoring and research. The GEnS has the potential to support global environmental assessments, and has been identified as a focal geospatial data resource for tasks of the recently launched Group on Earth Observation Biodiversity Observation Network. GEnS was constructed in two steps (Metzger, Shkaruba et al. 2012):

- In the first stage, 42 categorical variables were screened to remove those that were auto-correlated. Among the variables with high auto-correlation, researchers selected the most sensitive parameters and eliminated the others to prevent over-weighting of the zonation by co-linear variables.
- In the second step, statistical clustering analysis was performed on remaining variables: annual cumulative GDD using base temperature = 0°C, temperature and potential evapotranspiration seasonality's (month-to-month variation), and an annual aridity index (calculated as the ratio of mean annual total precipitation to mean annual total potential evapotranspiration). The statistical clustering was carried out using principle component analysis and iterative, self-organising data analyses, resulting in 125 zones. The classification procedure is further described in Metzger et al. (Metzger, Bunce et al. 2013).

#### 2.4.3 Sustainable And Global Environment Scheme (SAGE)

The SAGE zonation scheme was generated using global, gridded data for two variables known to be important drivers for crop development and crop growth: growing degree-days (GDD) and a crop soil moisture index, with the latter calculated as the ratio of actual to potential evapotranspiration following the approach of Prentice et al. (Prentice, Cramer et al. 1992) and Ramankutty et al. (Ramankutty, Foley et al. 2002). By downscaling the weather data from a 10' to a 5' resolution, calculations were carried out on a 5' grid basis (10\*10 km). The global ranges of the two categorical variables were each divided into ten classes, which were then used to develop a matrix of 100 unique combinations of growing degree-day and soil moisture conditions. Separate zonation schemes were developed for each of 18 crop species using crop-specific base temperatures for calculation of growing degree-days (e.g., 8°C for maize, 5°C for rice).

The global crop data sets from Monfreda et al. (Monfreda, Ramankutty et al. 2008) document the harvested area and yields for 175 crops around the world for the year 2000. The data sets represent agricultural conditions on a 5' \* 5' (approximately 10 km \* 10 km) geographic grid of the Earth. These data sets were created by merging a detailed library of global census data (from the national-, state- and county level for over 20,000 political units); with three different, detailed satellite data sets of global land cover conditions. Thus, whilst the data for yield and area harvested are presented at 5' resolution, they were originally reported at varying, coarser resolutions (Licker, Johnston et al. 2010).

---

#### 2.4.4 Global Yield Gap Atlas (GYGA-ED)

GYGA-ED is constructed from three categorical variables also used by the GEnS:

1. GDD with base temperature of 0°C.
2. Temperature seasonality (as the standard deviation of monthly average temperatures).
3. An aridity index (annual total precipitation divided by annual total potential evapotranspiration).

Grid cell size for the underpinning weather data was based on the SAGE framework (5' grid, or roughly 100 km<sup>2</sup> at the equator). Both GDD and temperature seasonality were calculated using climate data from WorldClim; the aridity index values were taken from CGIAR-CSI (Van Wart, van Bussel et al. 2013). Following Mueller et al. (Mueller, Gerber et al. 2012), only terrestrial surface covered by at least one of the major food crops was considered in this zonation scheme. To avoid inclusion of areas with negligible crop production, only grid cells with sum of the harvested area of major food crops > 0.5% of the grid cell area were accounted for, based on HarvestChoice SPAM crop distribution maps (You, Crespo et al. 2009), which update geospatial crop distribution data of Monfreda et al. (Monfreda, Ramankutty et al. 2008). The resulting range in values for GDD and aridity index were divided into 10 intervals, each with 10% of grid cells with harvested area of the major food crops, and combined in a grid matrix with three ranges of temperature seasonality to give a total of 300 AEZ classes. Of these, only 265 occur in regions where major food crops are grown. (See more on <http://www.yieldgap.org/web/guest/methods-overview>).

#### 2.4.5 Schemes derived from GAEZ and SAGE: HCAEZ and GLI

Concerning the HCGAEZ scheme, mean temperatures, elevation and GAEZ-LGP are used to define thermal regimes and temperature seasonality. Essentially, HCAEZ is a combination, or intersection, of several distinct and independent zonation schemes used in the GAEZ framework. Although it uses data of more recent origin, the HCAEZ resembles an earlier AEZ scheme developed by the Technical Advisory Committee (TAC) of the Consultative Group on International Agricultural Research (CGIAR) (Van Wart, van Bussel et al. 2013).

As the SAGE zonation scheme, the GLI zonation method also used the classification based on crop-specific GDD, but replaced the crop soil moisture index by annual total precipitation. Another modification was that only terrestrial surface covered by harvested area for a specific crop was considered based on geospatial crop distribution maps of Monfreda et al. (Monfreda, Ramankutty et al. 2008). Climate zones were developed for each crop by dividing GDD and precipitation each into ten classes, the intersection of which, formed a matrix of 100 individual climate zones. Instead of using equal ranges for the classes, zones were determined using an algorithm, such that 1% of the global harvested area of that specific crop was in each zone, a methodology known as the 'equal-area approach'.

#### 2.4.6 IIASA gridded EPIC

Concepts of Homogenous Response Units (HRU) and Simulation Units (SimU) have been implemented into IIASA gridded EPIC (Skalsky et al., 2008; Balkovic et al., 2013; Balkovic et al., 2014). HRU group spatially together those individual 5 arc min pixels having the same class of altitude (five classes), slope (seven classes), and soil texture (five classes). HRU represents the basic natural spatial units, from which conditions can only be barely, if at all, changed by climate, land cover or land use and, therefore, some kind of homogenous response to theoretically assumed climate, land cover or land use change can be expected within its borders. SimU adds further precision in HRU by grouping them within additional spatial domains – country borders and 30 arc min resolution pixels, which are supposed to maintain a direct link with other possible spatial data on land use, climate, or crop calendars. SimU consistently provides all the necessary inputs for the EPIC model, including climate, topography, soil, and crop management data. As such, SimU are considered as the smallest homogenous spatial units.

Within the HRU and SimU's stratified structure, the IIASA gridded EPIC integrates many original data sources, including the Princeton Global Dataset of Meteorological Forcings for Land Surface Modelling

(Sheffield et al., 2006), Shuttle Radar Topographic Mission Data (SRTM, Werner, 2001), Global 30 Arc Second Elevation Data (GTOPO, <http://eros.usgs.gov>), Digital Soil Map of the World (DSMW, <http://www.fao.org/geonetwork>), World Inventory of Soil Emission Potential Database (WISE, Batjes, 2009), Global Administrative Unit Layers (GAUL, <http://www.fao.org/geonetwork>), Mineral N, P and K fertilizer application rates (Mueller et al., 2012), crop yields and harvested areas (Monfreda et al., 2008), and Monthly Irrigated and Rainfed Crop Areas around the year 2000 (MIRCA2000, Portmann et al., 2010).

**Table 3** Main variables used in the global stratifications and their respective data sources

Stratification scheme	Variables considered	Data sources
GAEZ-LPG	Temperature, precipitation, potential ET and soil characteristics are used to calculate LGS.	Monthly weather data (20*20 km) from the CRU and the GPCC (More details in (Fischer, Nachtergaele et al. 2012))
GEnS	GDD (base t° of 0°C), aridity index, ET seasonality, t° seasonality (used in Iso-Cluster analysis to 'cluster' grid-cells into zones of similarity)	Monthly gridded climate data from the WorldClim database Annual aridity and potential ET seasonality derived from the CGIAR-CSI (1 km <sup>2</sup> ).
HCAEZ	Mean t°, elevation, and GAEZ-LGP are used to define thermal regimes and t° seasonality.	Monthly weather data (20*20 km) from the CRU and the GPCC (More details in (Fischer, Nachtergaele et al. 2012))
SAGE	GDD ( $\sum$ Tmean-crop-specific base t°) and crop soil moisture index (actual ET divided by potential ET).	33-year monthly averaged weather database from CRU (20*20 km) Soil texture data used to estimate the soil moisture index from the ISRIC (10*10 km)
GLI	Harvested area of target crop, crop-specific GDD and annual total precipitation.	33-year monthly averaged weather database from CRU (20*20 km) Geospatial crop distribution maps of Monfreda et al. (Monfreda, Ramankutty et al. 2008).
GYGA-ED	GDD (base t° of 0 °C), t° seasonality (quantified as the stdev of monthly average t°), aridity index (annual total precipitation divided by annual total potential ET).	Climate data from WorldClim (10*10 km) Aridity index values from CGIAR-CSI HarvestChoice SPAM crop distribution maps Location-specific observed data
IIASA EPIC	Many, coming derived from EPIC model	Princeton Global Dataset of Meteorological Forcings for Land Surface Modelling (Sheffield et al., 2006), topography data (GTOPO30, NASA SRTM), soil data (DSMW, WISE), land cover data (GLC2000, SPAM), land use data (Monfreda et al., 2008; Mueller et al. 2012, Portman et al., 2010) (More details in Balkovic et al. 2014, Skalsky et al. 2008)

**Table 4** Main data used in the global stratification

Stratification scheme	Variable calculated	Variable								
		Precipitation	Temperature	Potential ET	Actual ET	Soil characteristics	DEM	Crop characteristics	Crop Mask	
GAEZ	Length of growing season	x	x	x		x				
GENS	GDD (base t° of 0°C), aridity index, ET seasonality, t° seasonality	x	x	x						
HCAEZ	Thermal regimes and t° seasonality	x	x	x		x	x			
SAGE	GDD and crop soil moisture index		x	x	x	x		x		
GLI	GDD and annual total precipitation	x								x
GYGA-ED	GDD, t° seasonality, aridity index	x	x	x						x
IIASA EPIC	Many, derived from EPIC model					x	x			

#### 2.4.7 Use of remote sensing in stratification schemes

Only a few examples exist, in which, alongside land cover, remote sensing is directly used within stratification schemes. These are often very local:

- Discriminating cropping systems and agro-environmental measures by remote sensing (Pena-Barragan, 2008).
- Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery (Castillejo-Gonzales, 2009).

## 2.5 Stratification applications and evaluations

Agro-Ecological Zoning is a method that uses biophysical attributes of the land to cluster land-use types into more homogeneous areas. This exercise facilitates planning for the sustainable use of natural resources and has shown promising results for application in agriculture monitoring. However, some studies pointed out limitations concerning the global presentation schemes, as presented in this report. In general, remarks about the need to account for topography and the necessity to add soil information were the main remark. In Quiroz et al. (Quiroz, Zorogastúa et al. 2000), the application of AEZ is shown as limited by the lack of geospatial data, particularly in mountainous areas. Thus, they used remote sensing and process-based models for both climate interpolation and crop and livestock production in a watershed above 3,800m, and with the incorporation of these new tools, AEZ becomes a dynamic and more robust method. Mueller et al. (Mueller, Schindler et al. 2010), also highlighted that at an international level or global scale, methods, such as agro-ecological zoning or ecosystem and crop modelling, provide assessments of land productivity, but contain little information on soil. In his report, White et al. (White, Lubulwa et al. 2001), also presented that the definition of agro-climatic and agro-ecological zones will improve through applying digital elevation models, climate surfaces, plant growth models, field and remote sensing data, and geographic information systems. In his study, the use of the GROWEST model as the basis of the agro-climatic definition appears to be significantly superior, to approaches based simply on length of growing period, particularly in temperate environments.

Upon examining the global schemes that have been presented in more detail, some more targeted remarks can be made. According to van Wart et al. (Van Wart, van Bussel et al. 2013), the GAEZ-LGP

---

and HCAEZ schemes are too 'coarse' to use in estimating and extrapolating yield gap analyses, because climate heterogeneity within zones is too large. In Hazeu et al. (Hazeu, Metzger et al. 2011), the GEnS developed at the European level called EnZ is presented. The EnZ have been used to provide broad European environmental and as units for summary reporting (e.g., Thuiller et al., 2005; Metzger et al., 2008a; Smit et al., 2008). The European Commission has used the EnZs as the basis for assessing High Nature Value farmland (Paracchini et al., 2008) and the identification of potential areas for cultivation of bio-energy crops (EEA, 2007). According to Licker et al. (Licker, Johnston et al. 2010), the GLI is still, by far, the most complete description of agricultural landscapes – including the area and yield of a large number of crops – for the planet today.

Data model fusion approaches, such as IIASA gridded EPIC, require the establishment of a reliable crop production baseline, therefore, many evaluation exercises are necessary (e.g. Elliot et al. 2014, Rosenzweig et al., 2014, van Oijen et al. 2014).

Examples of applications are listed below.

GAEZ is used in:

- Alcamo, J., Dronin, N., Endejan, M., Golubev, G., Kirilenko, A. A new assessment of climate change impacts on food production shortfalls and water availability in Russia (2007) *Global Environmental Change*, 17 (3-4), pp. 429-444.
- Tatsumi, K., Yamashiki, Y., Valmir da Silva, R., Takara, K., Matsuoka, Y., Takahashi, K., Maruyama, K., Kawahara, N. Estimation of potential changes in cereals production under climate change scenarios (2011) *Hydrological Processes*, 25 (17), pp. 2715-2725. => Propose an improved version of the GAEZ called IGAEZ.
- Teixeira, E.I., Fischer, G., Van Velthuis, H., Walter, C., Ewert, F. Global hot-spots of heat stress on agricultural crops due to climate change (2013) *Agricultural and Forest Meteorology*, 170, pp. 206-215.
- Hasegawa, T., Fujimori, S., Shin, Y., Takahashi, K., Masui, T., Tanaka, A. Climate change impact and adaptation assessment on food consumption utilizing a new scenario framework (2014) *Environmental Science and Technology*, 48 (1), pp. 438-445 => Use a modified version of the GAEZ called M-GAEZ.

GEnS/EnS is used in:

- Zomer, R.J., Trabucco, A., Wang, M., Lang, R., Chen, H., Metzger, M.J., Smajgl, A., Beckschäfer, P., Xu, J. Environmental stratification to model climate change impacts on biodiversity and rubber production in Xishuangbanna, Yunnan, China (2014) *Biological Conservation*, 170, pp. 264-273.
- The EnS strata can be linked to climate change scenarios, providing insights into broad environmental shifts (Metzger et al., 2008b), as well as providing a basis for the prediction of future crop yields (Ewert et al., 2005) and changes in biodiversity (Verboom et al., 2007).

Finally, the EnS has been used as a core data layer in a number of other European classifications, including the four described below.

- Bunce et al. (2008) have illustrated how the EnS can be used as a sampling framework for assessing stock and trends in European habitats. The EnS have been developed further under the European Union (EU) European Bio-diversity Network (EBONE) project, which aims to create a framework for surveillance and monitoring of species and habitats in Europe.

GLI, as a revision of the SAGE zonation scheme formed the basis of the yield gap estimates in Foley et al. (2011) and Mueller et al. (2012). The SAGE zonation was employed by Johnston et al. (2011) to examine opportunities to expand global biofuel production through agricultural intensification in regions with similar growing conditions.

Integration of biophysical crop model outputs (IIASA gridded EPIC) with economic models describe the land-based production sector in the very detailed spatial heterogeneity:

- Very high resolution leads to diagnosis of the many means of farmers to adjust to climate change – importance of soil x climate x crop interaction for yield (Leclere et al., 2014).

- 
- Very high resolution leads to diagnosing potential for development of the livestock or bioenergy sectors (Havlik et al., 2014, Frank et al., 2013).
  - Such information overlays with high resolution data of socioeconomic nature, e.g. transport and processing costs form infrastructure (Mosnier et al., 2012, Cohn et al., 2014).

Agro-Ecological Zonation schemes have been used to:

- Identify yield variability and limiting factors for crop growth (Caldiz et al., 2002; Williams et al., 2008).
- Regionalise optimal crop management recommendations (Seppelt, 2000).
- compare yield trends (Gallup and Sachs, 2000).
- Determine suitable locations for new crop production technologies (Geerts et al., 2006; Araya et al., 2010).
- Analyse impacts of climate change on agriculture (Fischer et al., 2005).

## 2.6 Conclusions

The literature review emphasised the fact that robust global stratification schemes already exist and that they could be used as reliable basis to build the GAES. Considering the review and the objectives of the SIGMA project, the most relevant stratifications are the GEnS and the GYGA-ED. Both stratifications are performed on complete coverage, and the climatic heterogeneity within zones is the lowest among all the considered stratifications. The GYGA-ED scheme has been designed for agricultural production and yield gap estimation, whilst the GEnS scheme, created using the cluster method, has broader objectives, covering environmental monitoring in general. Even if the latter can be adapted and refined for different purposes, the GYGA-ED suits better the goals of SIGMA. From this, ways to improve the selected methodology were identified. In his article, van Wart (Van Wart, van Bussel et al. 2013) concluded that the GYGA-ED scheme should be extended using data on soils and management. Several studies have also pointed out that using information about the slope is essential to improve the yield gaps estimates (Quiroz, Zorogastúa et al. 2000).

---

## 3 Input datasets

### 3.1 Framework for the selection of input data

Data sources were collected and processed before use in global segmentation and characterisation of the resulting strata. Harmonisation of all data sets into one geo-database, with the same extent, geographic projection and origin, and spatial resolution was an important step. The following criteria were used as a starting point (Mücher et al., 2003):

1. The various components must be distinguished: abiotic-, biotic- and cultural aspects that are meaningful for agricultural production.
2. Clear distinctions should be made between primary data (e.g. rainfall data), interpreted data (e.g. land cover data derived from remotely sensed imagery) and attached values (e.g. ecological value, economic value, priority for conservation of natural or cultural aspects).
3. The components or characteristics mentioned under Point 1 are often interrelated in history, in actual functioning and in spatial distribution (correlative complex). This enables us to construct units (classes, legend-units) with a specific combination of characteristics that are frequently met in reality.
4. The type of relationships, in terms of dependencies between phenomena in a correlative matrix, can be envisaged as asymmetric. Some components are relatively stable and independent, whilst other data are dependent. In general, it is accepted that relatively independent abiotic phenomena determine the presence and nature of relatively dependent biotic phenomena. Changes in abiotic characteristics, therefore, lead to change in biotic components. It is possible to rank and order the various phenomena in the following way (see the functional hierarchy below). This hierarchy shows an increasing dependency at lower levels. An environmental or landscape classification should incorporate these insights by ordering and ranking phenomena in a comparable way (e.g. Klijn, 1995). Human influences have grown in impact and importance during history and one can observe the various serious impacts even on relatively independent natural components. They can be ordered and ranked according to their specific impacts or degree of interference with the ecosystem (e.g. Mücher, 1992; Stomph et al. 1997) and effects on components on the various hierarchical levels (e.g. geomorphology, soils and vegetation).
5. Agronomically relevant spatial data sets. Although independent phenomena (e.g. abiotic conditions) determine dependent phenomena (e.g. vegetation), classifications of the former are only relevant when they really match ecologically relevant distinctions; for example climate typologies should be based upon ecologically relevant thresholds or they should be determined by statistical analysis.
6. Agricultural production is strongly linked to environmental, as well as the cultural conditions. Generally, it is found that cultural aspects are only partly determined by physical phenomena e.g. climate, geology, geomorphology, soil conditions. Demography, cultural history and political history also act as independent factors that explain a certain (former) land use type or occupation pattern. A second aspect, namely 'cultural heritage' is the social dimension of a landscape, e.g. the social embeddedness of a rural population and the socio-economic interests of stakeholders. Some cultural phenomena, however, reflect physical conditions, for example the distribution of vineyards and related phenomena in buildings are historically conditioned by climate zones. Although management activities determine agricultural production (e.g. farmers harvesting crop or cutting willows), the most persistent and dominant aspects result from management, namely the appearance of the agricultural landscape, either seen with the eyes of the local observer, or from photos and digital snapshots from aerial photography and remote sensing.
7. From previous observations, it can be concluded that agricultural landscape character is based on a regionally distinctive correlation between various sets of bio-physical and cultural features. Since cultural aspects, such as land use type and land use intensity are not always or are only partly directly correlated with the agricultural landscape character. This means that any attempt to design a global classification, legend and map that delineate homogeneous areas with distinct combinations of successively abiotic, biotic and cultural features are often not in agreement with the reality.

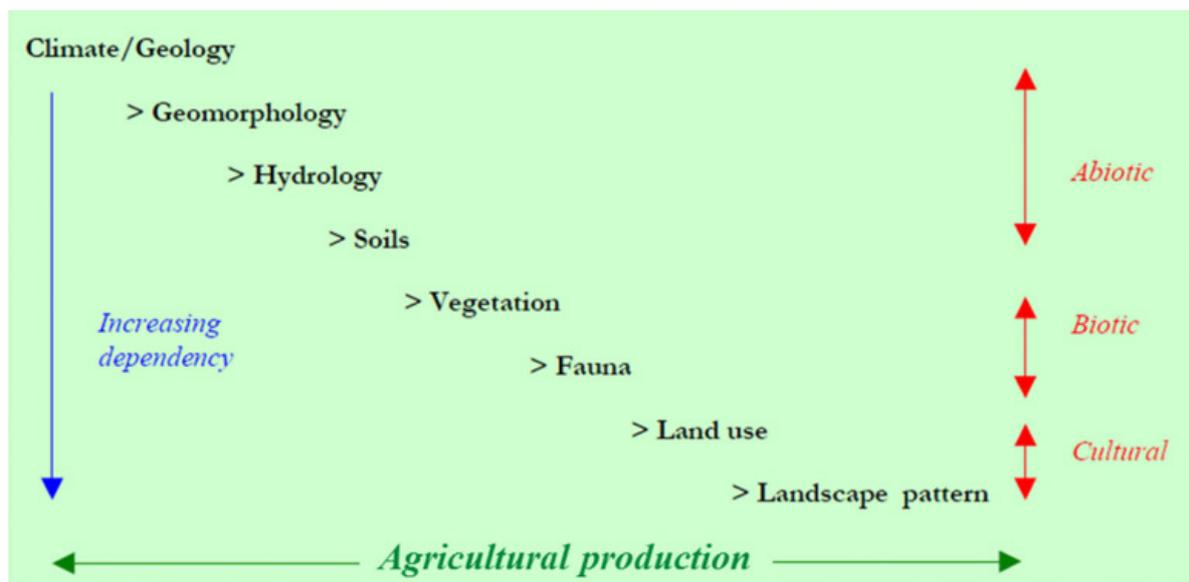
- 
8. Delineation of homogeneous agricultural landscape types that are based on successive combinations of abiotic-, biotic- and land use features is likely to ignore a variety of other regionally relevant cultural features.
  9. A basic requirement for multi-purpose use is, hence, flexibility. The technical facilities for data collection, storage and retrieval sustain this. It leads to the conclusion that flexible storage in a GIS and the acceptance of various classification procedures and evaluation systems should be key factors in designing information systems. However, it can still be worthwhile to produce one or more maps with a more definitive classification, because these maps could present the state-of-the-art of e.g. a landscape typology, providing that comparisons with other maps and underlying data can be made. In policy-making a map can receive a certain status, when used to express political agreements about the values or status of protection of certain areas. For this reason, a more rigid classification can be required.
  10. The role of classification (i.e. determining a number of thematic classes related to the nature or number of phenomena) and the description of each classification are crucial to deliver meaningful information.
  11. The choice for a scale should be related to both the applications and the availability of data. Data analysis should be an important module in research plans, whereas the lack of available or easily accessible data forms an important handicap.
  12. Information with a high resolution should be easily aggregated and generalised to more general and coarser-grained information by up-scaling techniques.

During the FAO Workshop in April 2014, it was decided to target especially those agro-environmental data layers that determine the work of WP3, namely, strategies for EO classification of a global cropland mask. As an example, climate data, such as cloud coverage, determines if optical or radar satellite imagery is required (in cases of almost continuous cloud cover, optical EO data cannot be used and research is largely dependent upon RADAR data). Crop type diversity and cropping pattern determine if temporal EO imagery is required. For example, if only one crop is being cultivated (e.g. monoculture) in a specific region, it means that one satellite image is sufficient to cover the whole region. When more than one crop is cultivated, it means that in most cases multi-temporal imagery is necessary to identify all crops. The same reasoning applies for the number of growing cycles over the year. More growing seasons means that more satellite imagery is required to identify all cropland. Phenology of the crops determines the EO acquisition time. Field size is an additional data layer that was selected, as it determines the spatial resolution of the satellite imagery. For very large fields, medium resolution satellite imagery, such as MODIS, can be used, whilst in the case of smaller fields, high resolution satellite imagery, such as Landsat or SENTINEL, must be used. For fields smaller than one hectare, very high spatial resolution satellite imagery, such as Pleiades or Worldview, is required. In areas where no crops are cultivated, no satellite imagery might be acquired to monitor crop production. This creates a preliminary list of agro-environmental data layers.

Preliminary selection of important agro-environmental data layers selected during the FAO workshop included:

1. Cloud coverage.
2. Crop type diversity.
3. Number of growing seasons per year.
4. Phenology of the growing cycles.
5. Field size.
6. Rainfed *versus* irrigated agriculture.
7. Cropland mask.
8. Slope and diversity in altitude.

At a later stage, additional climate variables, such as temperature and precipitation (where already important variables from the original framework, see Figure 2), or vegetation variables, such as gross primary production (GPP), will be selected to improve a global stratification of the environment for our specific purposes.



**Figure 2** Agricultural production as a functional hierarchy of abiotic-, biotic- and cultural phenomena (Mücher et al., 2003)

## 3.2 Climate

Rather than using an existing climate stratification, such as GEns (Metzger et al, 2013a) or GYGA-ED (van Wart et al., 2013), it was decided to use the original WORLDCLIM data for the delineation of our strata, whilst using the Gens or GYGA-ED as an additional descriptor for the strata created.

### 3.2.1 WorldClim global climate dataset

Hijmans et al. (2005) compiled the monthly averages of climate variables, as measured at weather stations from a large number of global-, regional-, national-, and local sources, mostly from the period 1950–2000. They interpolated these data using the thin-plate smoothing spline algorithm implemented in ANUSPLIN (Hutchinson, 2014) and created global (land areas only, excluding Antarctica) climate surfaces for monthly precipitation and minimum-, mean-, and maximum temperature. The spatial resolution of the database is 30 arc second spatial resolution; this is equivalent to approximately 0.86 km<sup>2</sup> at the Equator, and less elsewhere, and is commonly referred to as 1-km resolution. The data are referred to as the 'WorldClim' database and are available for download from <http://www.worldclim.org>.

Bioclimatic variables (BIOCLIM) form part of the WORLDCLIM database and are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These are often used in ecological niche modelling (e.g., BIOCLIM, GARP). The bioclimatic variables represent annual trends (e.g., annual mean temperature, annual precipitation) seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest- and warmest month, and precipitation of the wet- and dry quarters). These are coded as follows:

- BIO1 = Annual Mean Temperature.
- BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)).
- BIO3 = Isothermality (BIO2/BIO7) (\* 100).
- BIO4 = Temperature Seasonality (standard deviation \*100).
- BIO5 = Max Temperature of Warmest Month.
- BIO6 = Min Temperature of Coldest Month.
- BIO7 = Temperature Annual Range (BIO5-BIO6).
- BIO8 = Mean Temperature of Wettest Quarter.

BIO9 = Mean Temperature of Driest Quarter.  
BIO10 = Mean Temperature of Warmest Quarter.  
BIO11 = Mean Temperature of Coldest Quarter.  
BIO12 = Annual Precipitation.  
BIO13 = Precipitation of Wettest Month.  
BIO14 = Precipitation of Driest Month.  
BIO15 = Precipitation Seasonality (Coefficient of Variation).  
BIO16 = Precipitation of Wettest Quarter.  
BIO17 = Precipitation of Driest Quarter.  
BIO18 = Precipitation of Warmest Quarter.  
BIO19 = Precipitation of Coldest Quarter.

Annual Mean Temperature (BIO1) and Annual Precipitation (BIO12) were selected as the most important climate variables. Cloud fraction, as derived from data sources from the Climate Change Initiative (CCI), was next, as discussed in the next section.

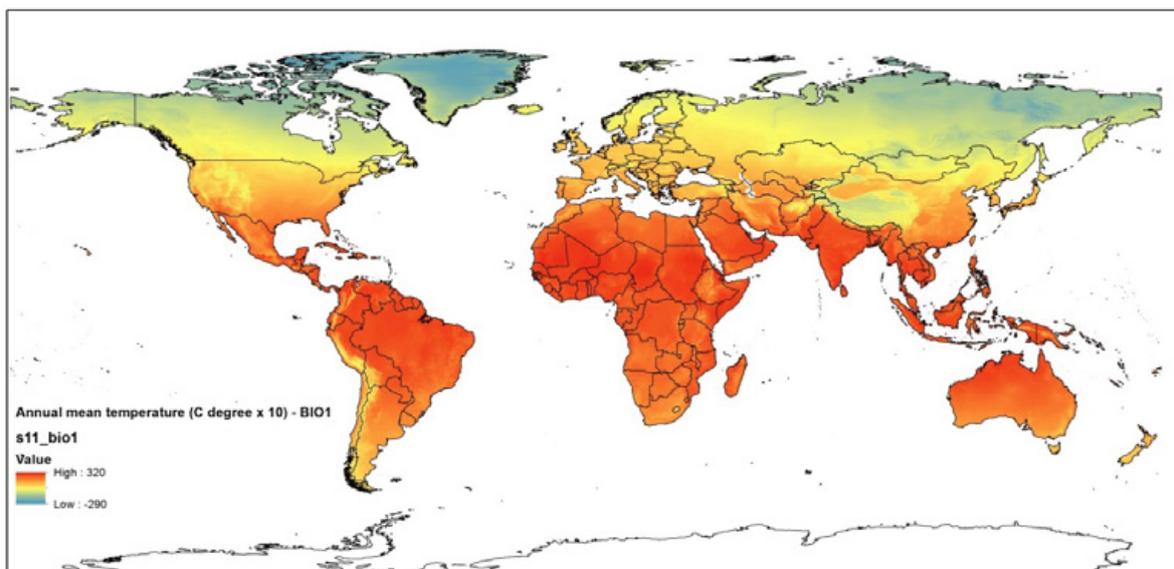
More information can be found at <http://www.worldclim.org/>

Selected data layers as diagnostic criteria:

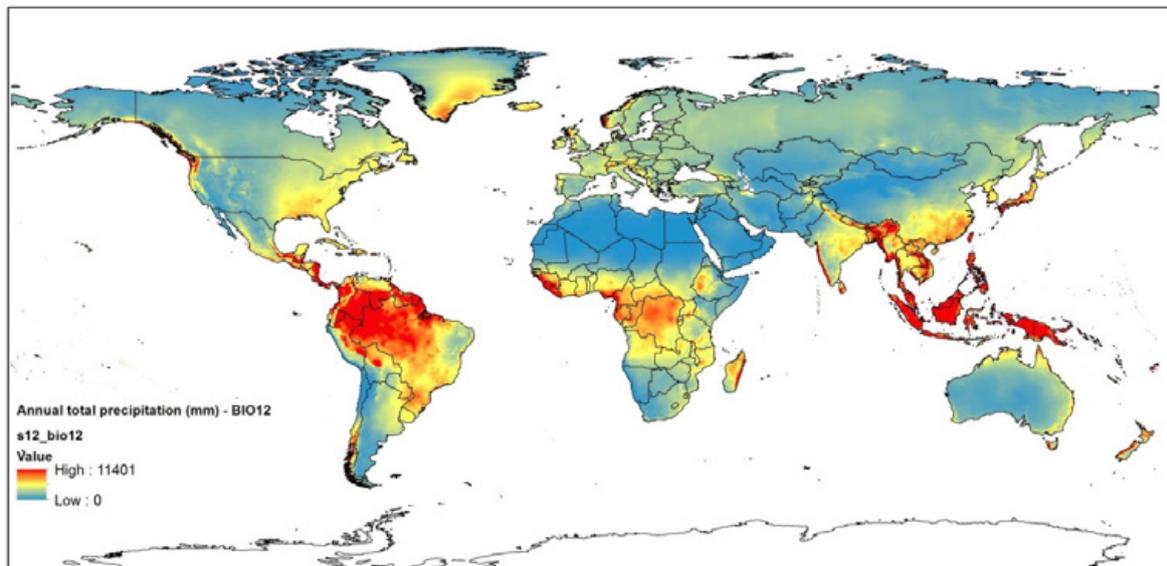
1. Annual Mean Temperature, BIO1 (unit is degree C \* 10).
2. Annual Precipitation, BIO12 (mm).

Selected data layers as descriptors:

1. Mean temperature by months (WorldClim min and max monthly temperature).



**Figure 3** Annual Mean Temperature (BIO1) from the BIOCLIM database - part of the WORLDCLIM database



**Figure 4** Annual Total Precipitation (BIO12) from the BIOCLIM database - part of the WORLDCLIM database

### 3.2.2 Cloud fraction

Within the SIGMA work, a new global cloud fraction data layer with a spatial resolution of 0.1x0.1 degree was produced. The climatology was derived from products provided by the ESA Cloud CCI, in which multiple sensors could be used for estimating the cloud fraction over the period 2007-2009. Two different climatologies were generated, one including both ascending- (afternoon) and descending (morning) passes, and a second, containing only descending (morning) passes. This section describes the climatology calculated from the descending passes only. An analysis of the results demonstrates that the product is temporally and spatially consistent. Spatial-temporal patterns in the cloud fraction climatology can be linked to climatological differences, whilst relatively small details caused by local elevation differences are clearly recognisable in the spatial structure of the cloud fraction. The products from the ESA Cloud CCI have been validated and their accuracy is described. Nevertheless, the accuracy of this global cloud fraction climatology has not yet been quantified, and it should preferably be compared with other cloud fraction products that have (very recently) become available (Whitcraft et al. 2014). An inter-comparison against MODIS derived cloud fraction would give insight into the effects active during the limited time-period (2007-2009) that was used to build this product, whilst a comparison against MeteoSAT second generation could be of interest to estimate the differences between this product and the diurnal effect on the estimated cloud fraction.

#### 3.2.2.1 Introduction to cloud cover data

Observations by optical polar-orbiting satellites provide a major contribution to the development and improvement of global agricultural monitoring, as is currently being carried out within the framework of the Group on Earth Observations Global Agricultural Monitoring (GEOGLAM) Initiative. Nevertheless, the capacity for agricultural monitoring by optical satellites is strongly hampered by cloud cover, which is persistent in many parts of the world during the agricultural season. To cope with persistent cloud cover, agricultural monitoring has mainly utilised data of operational polar-orbiting satellites that have global coverage with daily- or two-day revisit cycles (VGT, AVHRR, MODIS) albeit at a relatively low spatial resolution of 250m to 1km. The large drawback of this type of satellite observations is that spatial resolution of the sensor does not match the typical spatial scale of the agricultural landscape in many parts of the World. Therefore, the observed signal is often a mixture of the response of different fields with different crop types and natural vegetation. This mixing effect severely limits our ability to target the monitoring of individual crops and, therefore, most operational monitoring systems either monitor the level of 'cropland' or can only monitor on a crop-specific level, if the response of a single crop is strongly dominant in the satellite signal. Currently available and upcoming optical satellite missions (DMC, RapidEye, Sentinel2) with higher spatial resolution have a lower temporal resolution, in general, compared to the VGT/MODIS/AVHRR type missions that are currently being used for

vegetation monitoring. Information on the expected cloudiness during different sections of the growing season is, therefore, of high importance in estimating the probability of getting cloud free high-resolution satellite data. Together with the information of the field size, knowledge on cloudiness helps in the selection of suitable satellite data for crop mapping and monitoring.

### 3.2.2.2 Data and methods

#### *Available data for estimating cloudiness*

A short review was carried out to assess available data input for generation cloud fraction climatology. We focused on existing products, such as global gridded cloud masks or cloud fraction.

Table 5 gives an overview of the available products. Most are swath- or data granule products (1, 3, 4) that require considerable download and processing time, in coarse monthly intervals or at a very low spatial resolution (5). Therefore, the most suitable product to derive cloud fraction climatology is the ESA gridded global Cloud-CCI (Climate Change Initiative) data set, which is daily available at a spatial resolution of 0.1 degrees (~10x10km). Moreover, this data set is generated from different satellite sensors, although the historical availability is somewhat limited (2007-2009) compared to the other products. A potential drawback of the Cloud CCI data set could be that it focuses on optimal estimates of cloud properties, rather than on cloud masks, to derive land surface properties.

**Table 5** Overview of available products for cloud fraction estimation

Source	Product	Period	Type	Temporal resolution	Spatial resolution
NASA	Clouds and Earth's Radiant Energy System (CERES) <sup>1</sup>	2000-2013	Gridded	Monthly	1x1 degree
			Swath	Daily	20km
ESA	ESA Cloud CCI project overview <sup>2</sup>	2007-2009	Gridded	Monthly	0.5x0.5 degree
			Gridded	Daily	0.1x0.1 degree
			Swath	Daily	5 km
NASA	MODIS MOD/MyD035 L2 products <sup>3</sup>	2000-2014	Swath	Daily	250 m or 1 km
NASA	MODIS MOD/MYD08 L3 products <sup>4</sup>	2000-2014	Gridded	Daily	1x1 degree
			Gridded	8 Daily	
			Gridded	Monthly	
ESA	TEMIS Cloud Fraction (GOME, OMI, SCHIAMACHY) <sup>5</sup>	2002-2013	Swath	Daily	50 km footprint

1. [http://ceres.larc.nasa.gov/order\\_data.php](http://ceres.larc.nasa.gov/order_data.php)

2. <http://www.esa-cloud-cci.org/?q=information>

3. [http://modis-atmos.gsfc.nasa.gov/MOD35\\_L2/](http://modis-atmos.gsfc.nasa.gov/MOD35_L2/)

4. [http://modis-atmos.gsfc.nasa.gov/\\_docs/L3\\_ATBD\\_2008\\_12\\_04.pdf](http://modis-atmos.gsfc.nasa.gov/_docs/L3_ATBD_2008_12_04.pdf)

5. <http://www.temis.nl/fresco/>

### 3.2.2.3 The ESA Cloud CCI products

The ESA Cloud CCI Initiative derived products from a constellation of satellite sensors, including ENVISAT-AATSR/MERIS, NOAA-AVHRR and TERRA/AQUA-MODIS (Table 6). These data have been processed at their native resolution to derive a spectrum of cloud related products, including cloud cover, cloud phase and cloud optical properties. The Cloud CCI provides products at different processing levels, varying from instantaneous swath products at native resolution, to daily gridded products at 0.1 degree, and finally to monthly aggregated products at 0.5 degree resolution (Table 7).

**Table 6** Equator crossing times and spatial resolution of the used satellites carrying the instruments AATSR, MODIS and AVHRR (after ESA Cloud CCI Product User Guide, Table 2-1)

Sensor	Satellite	Equator crossing time	Swath width	Spatial resolution
AATSR/MERIS	ENVISAT	10:00 am	512 km	1 km
AVHRR	NOAA-15	10:00 am/pm (drifting)	3000 km	4 km (GAC resolution)
	NOAA-16	02:00 am/pm (drifting)		
	NOAA-17	10:00 am/pm (drifting)		
	NOAA-18	01:30 am/pm (drifting)		
MODIS	AQUA	01:30 pm	2330 km	1 km
	TERRA	10:30 am		

**Table 7** Processing levels of Cloud CCI data products

Data level products	Abbreviation	Spatial resolution	Description
Level 2	L2	1 to 4 km	Retrieved cloud variables at the same resolution and location as the level 1 source
Level 3U (Uncollated)	L3U	0.1° lat/lon	Cloud properties of L2 data granules remapped to a global space grid without combining any observations from overlapping orbits. Only sampling is done. Common notation for this processing level are also L2b and L2G. Temporal coverage of this product is one day.
Level 3C (Collated)	L3C	0.5° lat/lon	Cloud properties combined (averaged) from a single instrument into a global space-time grid. Temporal coverage of this product is one month.
Level 3S	L3S	0.5° lat/lon	Cloud properties combined (averaged) from multiple instruments/platforms into a global space-time grid. This product only exists for the AVHRR-MODISAATSR product family. Temporal coverage of this product is one month.

For the generation of the SIGMA cloud fraction climatology, the L3U product was used. This provides a cloud mask in the form of a global gridded product at 0.1 degree spatial resolution. The product is generated by combining the individual L2 swaths of each platform/sensor for a single day and a single orbit (asc/desc) into one global product. For interpolating the L2 products towards the 0.1 degree resolution, a simplistic approach was chosen, through which a particular L2 pixel was chosen to represent the L3 grid cell at 0.1 degrees. The L3U product is then split up into ascending- and descending satellite nodes. The viewing zenith angle sampling and the separation into the two nodes effectively leads to a larger temporal and spatial coherence of atmospheric patterns compared to the previously considered, combined product, in which a random sampling was applied. The drawback from this sampling approach is that it leads to a strong reduction of the number of pixels used for estimating the cloud coverage, although no bias is introduced regarding the clouded- or unclouded state of the pixel.

#### 3.2.2.4 Retrieving and processing of products from ESA Cloud CCI

The L3U NetCDF v3 dataset was obtained from the Cloud CCI website. Monthly zip archives per sensor/platform contain the daily datasets for global gridded cloud properties. However, only the binary cloud mask (cloud-free/clouded) was retrieved from the NetCDF archive and stored in a separate file. Annex 1 provides an overview of the available data files per month and sensor/platform. Two separate cloud fraction products were produced, one including all daytime overpass (full) and a second only including only the descending daytime overpasses (desc\_only), which have an equator passing time in the morning (see Table 6). Further processing of the daily cloud masks into a decadal cloud fraction product consisted of a fairly straightforward aggregation process:

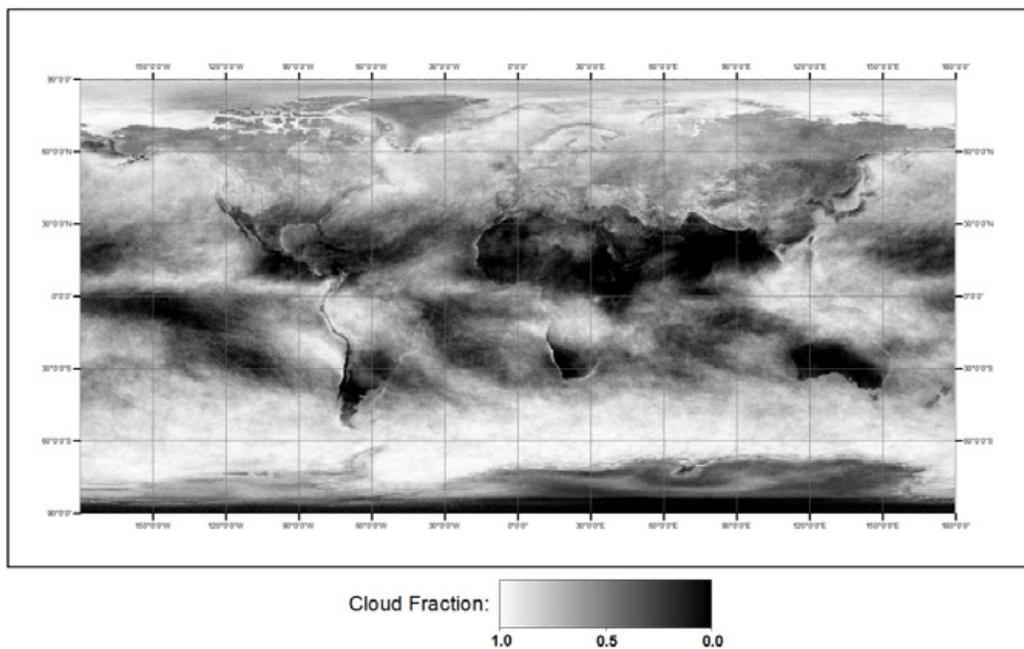
1. Cloud data were collected for all dates in 2007, 2008, 2009 from all sensor/platforms.

2. The number of clouded observations was computed for each pixel and divided by the total number of observations (excluding missing values) to estimate the fraction of clouded observations. The estimate of the cloud fraction obtained from this approach should be regarded as an estimate of the expected **cloudiness at satellite overpass time**. In general, it is not representative of the cloud fraction over the entire day. As the overpass times of the different satellites do not differ widely, the cloud masks of the different sensor/platforms are not strictly independent.

### 3.2.3 Results

#### 3.2.3.1 Overall results

The final output from this WP3 sub work package is a 0.1 degree global cloud fraction product providing an estimate of the average cloudiness at dekadal steps (Figure 5). The figure clearly shows the difference in average cloudiness over the globe for the first dekad of the year, with cloud-free area in large parts of Northern Africa, India and Bangladesh, South America and Southern Australia.

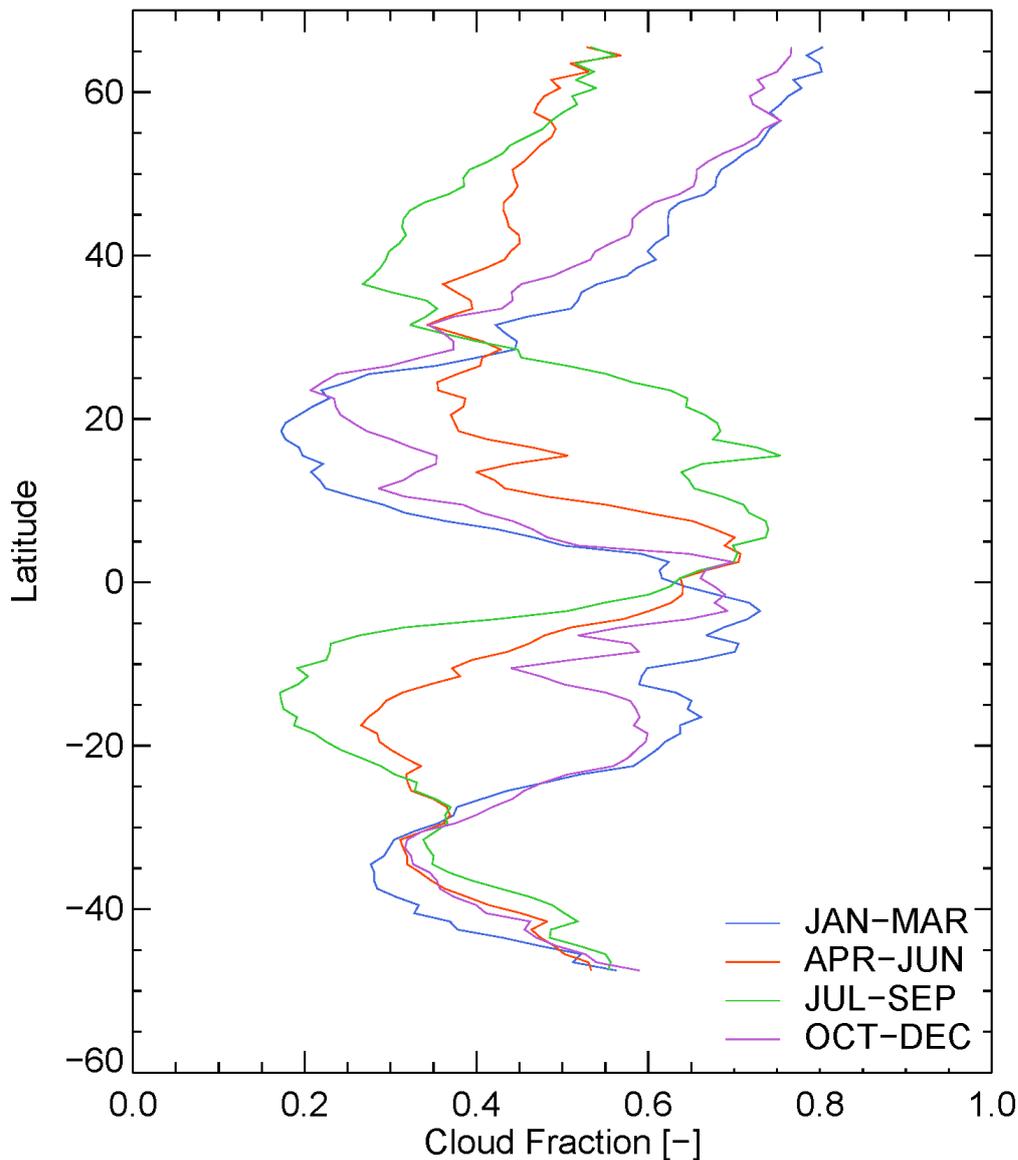


**Figure 5** Cloud fraction in the first dekad of January in the Global Cloud Fraction Climatology produced from products derived from the ESA Cloud CCI

#### 3.2.3.2 Cloud fraction by latitude and season

The seasonal- and spatial variability in cloud fraction can be visualized by calculating the average cloud fraction in latitudinal bands. To focus on areas with dominant agriculture, only pixels with a major share of agricultural land use were selected using the crop mask developed by FAO and IIASA (2007). Only pixels with more than 20% agricultural land use were taken into consideration to calculate the average cloud fraction.

The results from this analysis (Figure 6) clearly demonstrate the variability in cloudiness across the latitudes. High cloudiness along the equator that shifts with the season, moves northwards during the period from April to September, and southward from October to March. Large differences in cloudiness are evident in the Northern Hemisphere, between the Summer period from April to September and the Winter period from October to March. Finally, in the extreme South (below -30 degrees), differences between seasons are remarkably small. A similar figure, but aggregated by month, can be found in Figure 46.



**Figure 6** Average cloud fraction per season and per 1x1 degree latitude box

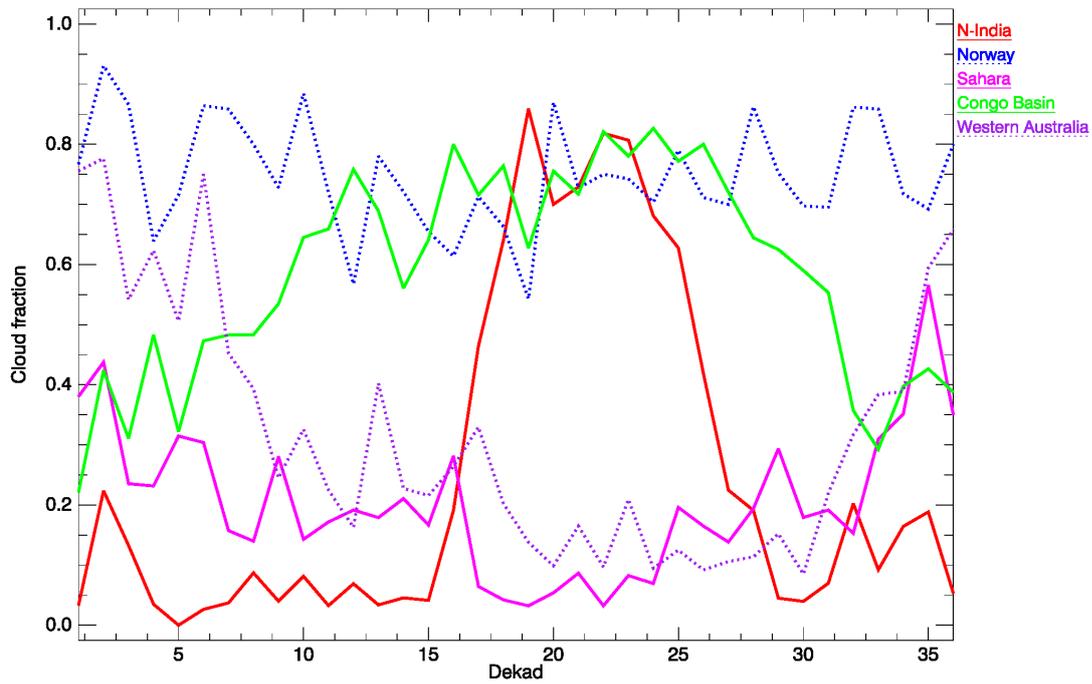
### 3.2.3.3 Detailed results for selected locations

The seasonality of the cloud fraction can also be visualised for selected points (Figure 7) showing the extreme difference in cloudiness for Northern India, with nearly complete cloudiness during the monsoon period and low cloudiness in the rest of the year. The coastal areas of Norway experience consistently high cloudiness throughout the year, whilst the Sahara Desert shows a persistently low cloudiness, but with some seasonal effects. Furthermore, the tropical rainforest in the Congo Basin experiences high cloudiness during most of the year, with a clear, lower cloudiness during the period from November through February.

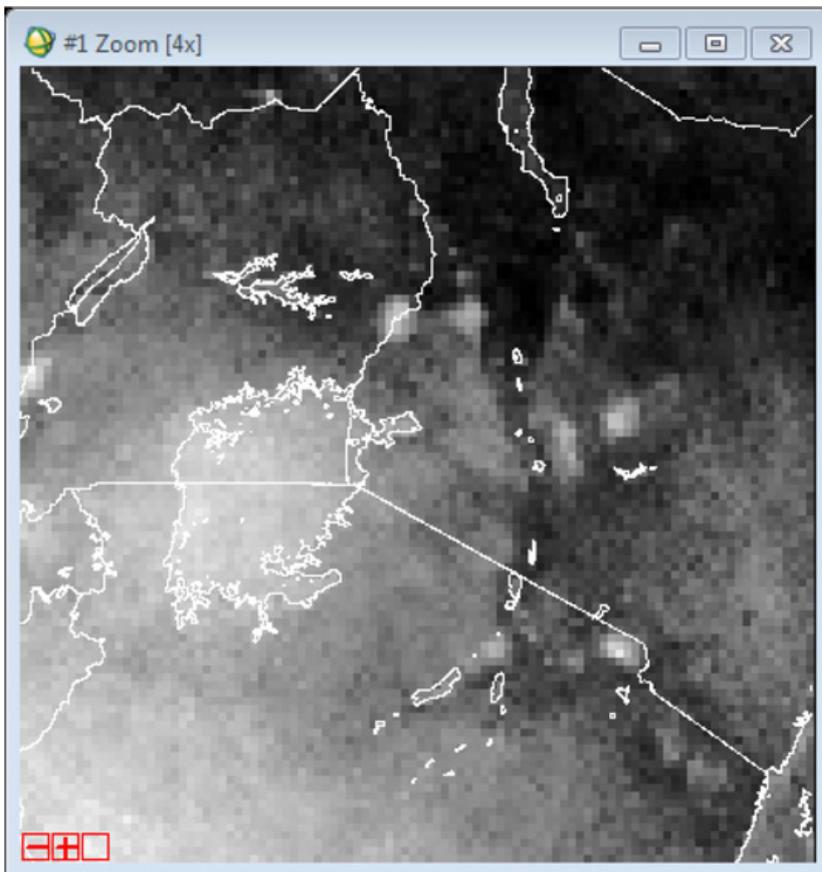
Two further details from the cloud fraction climatology are presented in Figure 8 and Figure 9. The first shows Lake Victoria and the Africa Rift area, where several large- and small volcanoes are located. These volcanoes cause local cloudiness due to the difference in elevation compared with the surrounding area, and they can be easily identified in the cloud fraction climatology. This demonstrates that even at a relatively coarse resolution of 0.1x0.1 degrees, such local features can be identified.

The second example shows the cloud fraction for Central Asia (including Iran, Afghanistan, Pakistan, Turkmenistan and Uzbekistan). In this area also, there is generally a low cloud fraction and the impact of local elevation differences on cloud fraction is evident. The morphology of mountain ranges in Central Iran is clearly visible and matches the structure that can be found in digital elevation models.

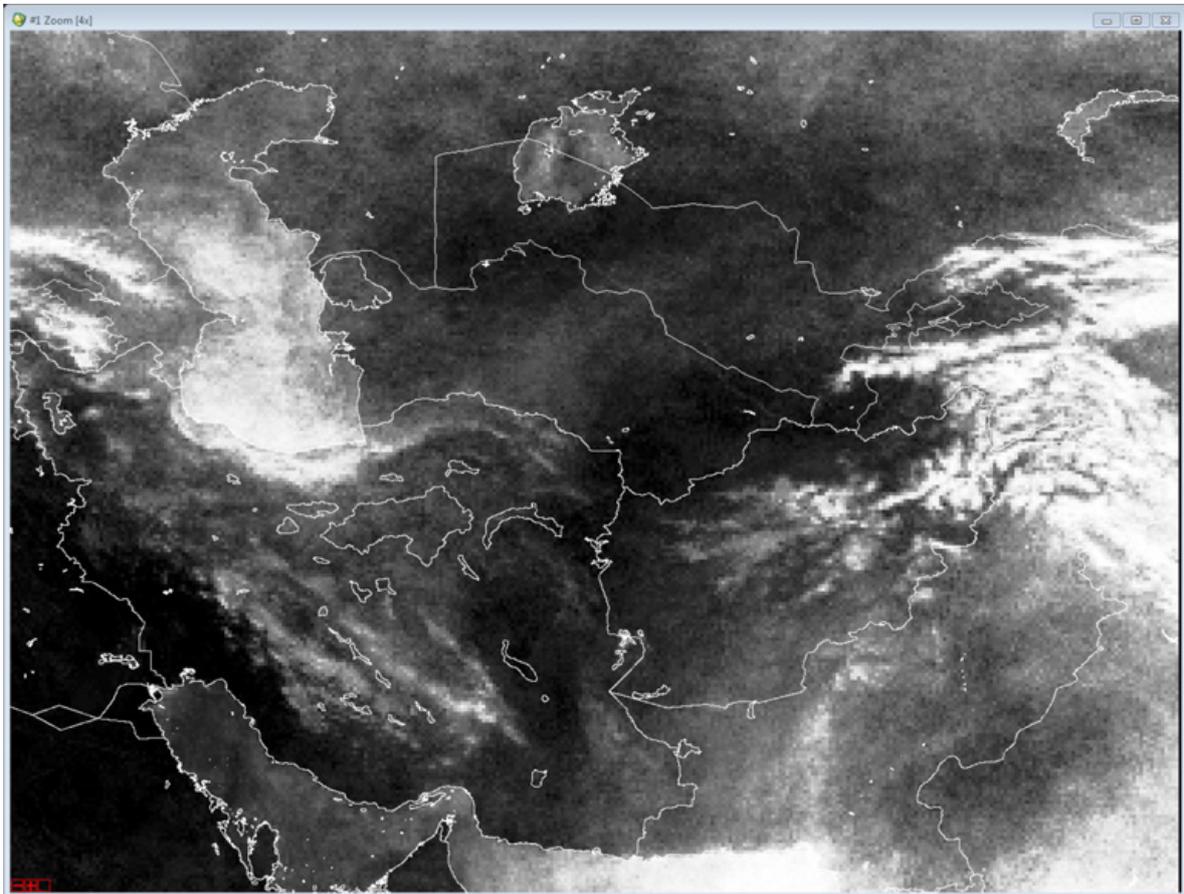
In addition, the morphology of mountain ranges in Eastern Afghanistan, Northern Pakistan and Turkmenistan can be easily identified.



**Figure 7** Examples of seasonality in cloud fraction for selected locations



**Figure 8** Detail of the cloud fraction climatology for the Lake Victoria area for the 1<sup>st</sup> dekad of January. Yellow circles show the location of major volcanoes in this area



**Figure 9** Detail of the cloud fraction climatology for Central Asia (Caspian Sea in the upper left corner) for the 1<sup>st</sup> dekad of January

### 3.2.4 Conclusions

For the SIGMA WP3, a global cloud fraction climatology was produced with a spatial resolution of 0.1x0.1 degrees. The climatology was derived from the ESA Cloud CCI dataset and, therefore, multiple sensors were involved for estimating the cloud fraction over the period 2007-2009. Two different climatologies were generated, one including both ascending- (afternoon) and descending (morning) passes, and a second one, containing only descending (morning) passes. This document describes the climatology calculated from the descending passes only. An analysis of the results demonstrates that the product is temporally- and spatially consistent. Spatial-temporal patterns in the cloud fraction climatology can be linked to climatological differences, whilst relatively small details caused by local elevation differences are clearly recognisable in the spatial structure of the cloud fraction. The accuracy of this global cloud fraction climatology has not yet been quantified and it should preferably be compared with other cloud fraction products that have (very recently) become available (Whitcraft et al. 2014). An inter-comparison against MODIS derived cloud fraction would give insight into the effects of the limited time-series (2007-2009) that was used to build this product, whilst a comparison against those derived from the MSG could be of interest to estimate the differences between this product and the diurnal effect on the estimated cloud fraction.

#### **Additional processing**

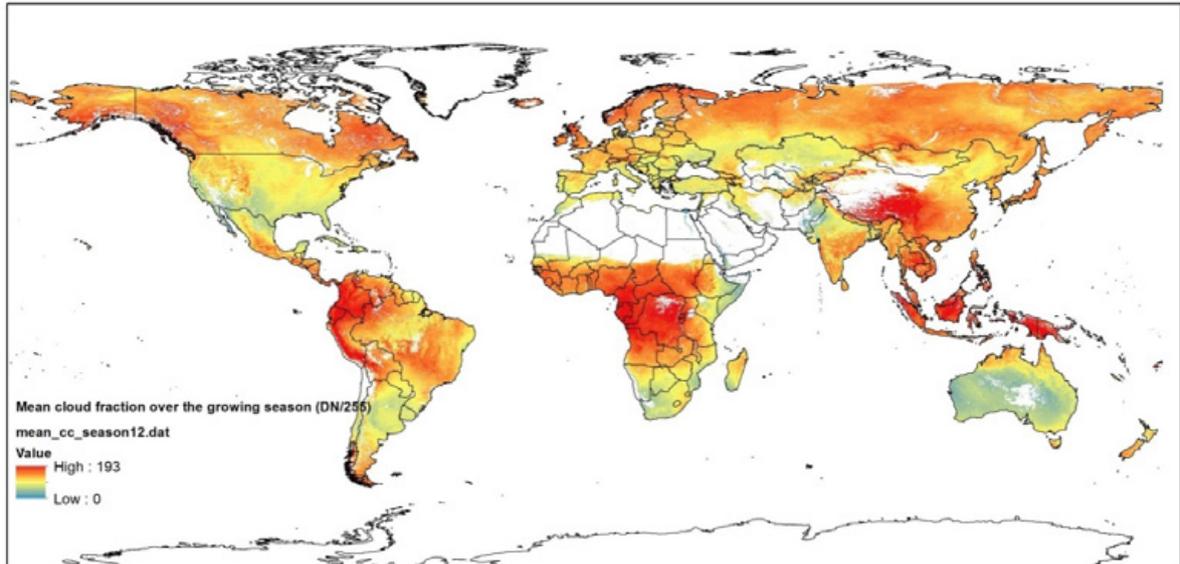
For the production of a global crop land mask, it is only important to know what the cloud fraction is over the growing season or around the maximum peak of the vegetation. Therefore, phenological data (PhenoDat) are used to calculate cloud climatology variables only for the specified periods. Selected cloud fraction products for the stratification/segmentation process were:

1. Mean cloud fraction over the growing season -> Mean\_cc\_season12: average cloud cover for season 1&2, based on ns1 and ne1, (cloud fraction: DN/255), scaled between 0 (0% cc) and 255 (100% cc).

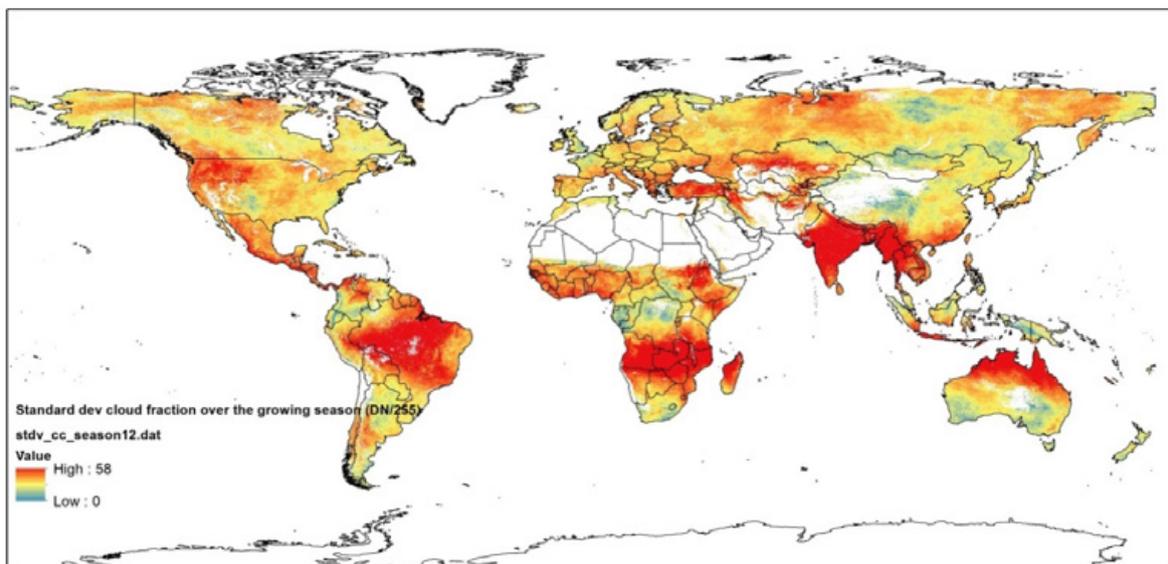
- Standard deviation of the cloud fraction over the growing season -> Stdv\_cc\_season12: standard deviation of cloud cover for season 1&2, based on ns1 and ne1, (cloud fraction: DN/255), scaled between 0 (0% cc) and 255 (100% cc).

Selected data layers as descriptors for the obtained strata were:

- Monthly cloud fraction (dekadal aggregation).



**Figure 10** Mean cloud fraction (2007-2009) over the growing season, scaled between 0 (0% cc) and 255 (100% cc) (source CCI/Wageningen Environmental Research (Alterra))



**Figure 11** Standard deviation of the cloud fraction (2007-2009) over the average growing season, scaled between 0 (0% cc) and 255 (100% cc) (source CCI/Wageningen Environmental Research (Alterra))

---

### 3.3 Slope and diversity in altitude

Global altitude values were derived from three data sets: the SRTM DEM v4 with a 90 m spatial resolution, the ASTER DEM with a 30 m spatial resolution and GTOPO30 with resolution of 30 arc seconds. ASTER was used only for those areas at high latitudes, where crops are cultivated:

- SRTM v4: complete data coverage from 54S to 60N.
- GTopo30: to extend the global coverage from 60N to 75N.
- ASTER-GDEM: for European cultivated areas between 5E to 35E and 60N to 75N.

The slope values were computed on a 90m grid by projecting altitude data to appropriate UTM zones (Carter and Stuiver, 2014). A global slope map was obtained by merging and re-sampling the whole 90m dataset into a 100m grid in a geographic lat/lon projection. The following statistics were generated:

Altitude variation (100m pixels) for each square kilometre

- Minimum Altitude.
- Mean DEM.
- Maximum DEM.
- Delta DEM (Max. DEM minus Min. DEM).
- DEM standard deviation.

Slope variation (100m pixels) for each square kilometre

- Minimum and Maximum Slope Values<sup>1</sup>
- Average Slope Value (calculated using;)
- Slope Standard Deviation.
- FAO Slope Categories Distribution.

FAO slope categories are:

1. Terrain slope 0-0.5%.
2. Terrain slope 0.5-2%.
3. Terrain slope 2-5%.
4. Terrain slope 5-8%.
5. Terrain slope 8-16%.
6. Terrain slope 16-30%.
7. Terrain slope 30-45%.
8. Terrain slope >45%.

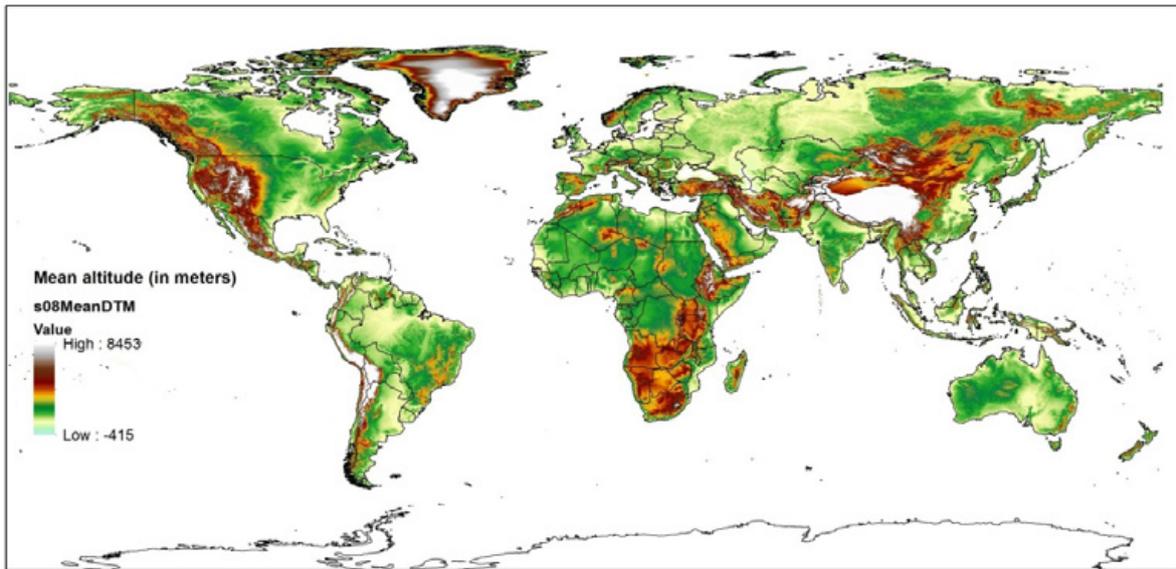
The following tools from ArcGIS 10.2.1 were used in the process:

- Mosaic Dataset, used to create a single dataset from the original tiles (per data source).
- Project Raster, used to project between WGS84 and Mollweide Spatial reference.
- Slope, used to calculate the Slope Value, which is defined as the change in the y-coordinates divided by the change in the x-coordinates. The measurement unit is percentage rise. The range is from 0 to nearly infinity.
- Raster Math Minus function, used to calculate the Delta DEM data layer.
- Export data, used to create TIF files from the mosaic datasets to be used in the R processes.

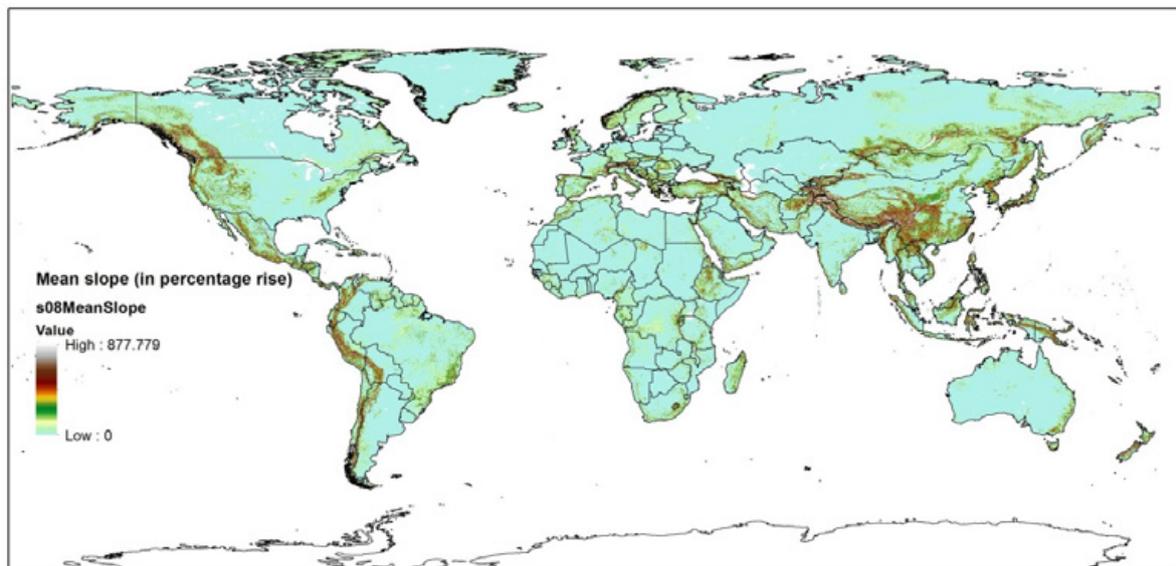
The R software provides an easy tool for tile-based statistical computing. The 100m mosaic DEM were input in a 5 by 5 degree tile structure. The values of the operators Min, Mean, Max and Standard deviation were derived at one kilometre grid cell. The FAO slope categories distribution by 1km pixel dataset was calculated using the zonal count function from the R raster package.

---

<sup>1</sup> The measurement unit for the slope value is percentage rise. The range is from 0 to nearly infinity. A flat surface is 0 percent rise, a 45 degree surface is 100 % rise and an almost 90 degree surface will have a percent rise approaching infinity. So very high slope values do exist.



**Figure 12** Mean altitude at 1km resolution, as derived from SRTM and complementary data sets ASTER Dem and GTOPO30



**Figure 13** Mean slope in percentage rise aggregated from 90m to 1 km resolution as derived from SRTM and complementary data sets ASTER Dem and GTOPO30

### 3.4 Soil productivity

Soil productivity was calculated on the basis of soil parameters provided by the Harmonized World Soil Database v 1.2 (HWSD, <http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>). The Harmonized World Soil Database is a 30 arc-second raster database with over 15,000 different soil mapping units that combines existing regional- and national updates of soil information worldwide (SOTER, ESD, Soil Map of China, WISE), with the information contained within the 1:5 000 000 scale FAO-UNESCO Soil Map of the World (FAO, 1971-1981). A soil map unit is: (i) A conceptual group of one to many delineations identified by the same name in a soil survey that represent similar landscape areas comprised of either: (1) the same kind of component soil, plus inclusions, or (2) two or more kinds of component soils, plus inclusions, or (3) component soils and miscellaneous area, plus inclusions, or (4) two or more kinds of component

---

soils that may or may not occur together in various delineations, but all have similar, special use and management, plus inclusions, or (5) a miscellaneous area and included soils. (ii) A loose synonym for a delineation. The resulting raster database consists of 21,600 rows and 43,200 columns, which are linked to harmonised soil property data. The use of a standardised structure allows for the linkage of the attribute data with the raster map to display or query the composition in terms of soil units and the characterisation of selected soil parameters (organic carbon, pH, water storage capacity, soil depth, cation exchange capacity of the soil and the clay fraction, total exchangeable nutrients, lime- and gypsum contents, sodium exchange percentage, salinity, textural class and granulometry). For more information, see <http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>

Soil productivity was calculated on the basis of the following soil parameters:

1. nutrient availability.
2. nutrient retention capacity.
3. rooting conditions.
4. oxygen availability to roots.
5. excess salts.
6. Toxicities.
7. Workability.

Such soil qualities are related to the agricultural use of the soil and more specifically to specific crop requirements and tolerances. For the comparison of soil qualities, maize was selected as the reference crop, because of its global importance and wide geographical distribution.

#### 3.4.1 Nutrient availability (SQ1)

Soil quality is decisive for successful low input level farming, and to some extent, also for intermediate input levels. Diagnostics related to nutrient availability are manifold. Important soil characteristics of the topsoil (0-30 cm) are: Texture/Structure, Organic Carbon (OC), pH and Total Exchangeable Bases (TEB). For the subsoil (30-100 cm), the most important characteristics considered are: Texture/Structure, pH and TEB. Soil characteristics relevant to soil nutrient availability are also to some extent correlated. For this reason, the most limiting soil characteristic is combined with the average of the remaining less limiting soil characteristics in the evaluation to represent soil quality SQ1.

#### 3.4.2 Nutrient retention capacity (SQ2)

Nutrient retention capacity is of particular importance for the effectiveness of fertilizer applications and is, therefore, of special relevance for intermediate- and high input level cropping conditions. Nutrient retention capacity refers to the capacity of the soil to retain added nutrients against losses caused by leaching. Plant nutrients are held in the soil on the exchange sites provided by the clay fraction, organic matter and the clay-humus complex. Losses vary with the intensity of leaching, which is determined by the rate of drainage of soil moisture through the soil profile. Soil texture affects nutrient retention capacity in two ways; through its effects on available exchange sites on the clay minerals, and by soil permeability.

The soil characteristics used for topsoil are, respectively: Organic Carbon (OC), Soil Texture (Text), Base Saturation (BS), Cation Exchange Capacity of soil (CECsoil), pH, and Cation Exchange Capacity of clay fraction (CECclay). Soil pH serves as indicator for aluminium toxicity and for micro-nutrient deficiencies. The most limiting of these soil characteristic is combined with the average of the remaining less limiting soil characteristics to estimate nutrient retention capacity SQ2.

#### 3.4.3 Rooting conditions (SQ3)

Rooting conditions include effective soil depth (cm) and effective soil volume (volume %) related to the presence of gravel and stoniness. Rooting conditions may be affected by the presence of a soil phase by either limiting the effective rooting depth or decreasing the effective volume accessible for root penetration. Rooting conditions address various relations between soil conditions of the rooting zone and crop growth. The following factors are considered in the evaluation:

- Adequacy of foothold, i.e., sufficient soil depth for the crop for anchoring.
- Available soil volume and penetrability of the soil for roots to extract nutrients.
- Space for root and tuber crops for expansion and economic yield in the soil.
- Absence of shrinking and swelling properties (vertic properties) that affect root- and tuber crops.

Soil depth/volume limitations affect root penetration and may constrain yield formation (roots and tubers). Relevant soil properties considered are: soil depth, soil texture/structure, vertic properties, gelic properties, petric properties, and the presence of coarse fragments. This soil quality is estimated by multiplying the soil depth limitation with the most limiting soil or soil phase property. Soil phases that relevant for rooting conditions vary somewhat with source of soil map and soil classification used. In the HWSO these are:

- FAO 74 soil phases: stony, lithic, petric, petrocalcic, petrogypsic, petroferric, fragipan and duripan.
- FAO 90 soil phases: rudic, lithic, petroferric, placic, skeletal, fragipan and duripan.
- ESB (European Soil Bureau) soil phases and other soil depth/volume related characteristics: stony, lithic, petrocalcic, petroferric, fragipan and duripan, and the presence of gravel or concretions, obstacles to roots (six classes), and impermeable layers (four classes).

#### 3.4.4 Oxygen availability (SQ4)

Oxygen availability in soils is largely defined by drainage characteristics of soils. The determination of soil drainage classes is based on procedures developed at FAO (FAO 1995). These procedures take soil type, soil texture, soil phases and terrain slope into account. Apart from drainage characteristics, the soil quality of oxygen availability may be influenced by soil and terrain characteristics that are defined through the occurrence of specific soil phases. For the FAO '74, these include classification soil phases that indicate phreatic conditions, and for the FAO '90 classification, soil phases that indicate phreatic, anthraquic, inundic, or placic conditions, respectively.

#### 3.4.5 Excess salts (SQ5)

Accumulation of salts can cause salinity. Excess of free salts referred to as soil salinity is measured as Electric Conductivity (EC in dS/m) or as saturation of the exchange complex with sodium ions, which is referred to as sodicity or sodium alkalinity and is measured as Exchangeable Sodium Percentage (ESP). Salinity affects crops by inhibiting the uptake of water. Moderate salinity affects growth and reduces yields; high salinity levels can kill the crop. Sodicity causes sodium toxicity and affects soil structure leading to massive- or coarse columnar structure with low permeability. Apart from soil salinity and sodicity, conditions indicated by saline (salic) and sodic soil phases may affect crop growth and yields. In case of simultaneous occurrence of saline (salic) and sodic soils, the limitations are combined. The most limiting of the combined soil salinity and/or sodicity conditions and occurrence of saline (salic) and/or sodic soil phase were selected.

#### 3.4.6 Toxicities (SQ6)

Low pH leads to acidity related toxicities, e.g., aluminium-, iron-, manganese toxicities, and to various deficiencies, e.g., of phosphorus and molybdenum. Calcareous soils exhibit generally micronutrient deficiencies, for instance, of iron, manganese, and zinc, and in some cases, molybdenum toxicity. Gypsum strongly limits available soil moisture. Tolerance of crops to calcium carbonate and gypsum varies widely (FAO, 1990; Sys, 1993). Low pH and high calcium carbonate and gypsum are mutually exclusive. Acidity related toxicities, such as aluminium toxicities and micro-nutrient deficiencies are accounted for respectively in SQ1, nutrient availability, and in SQ2, nutrient retention capacity. This soil quality, SQ6, therefore, only includes calcium carbonate and gypsum related toxicities. The most limiting of the combination of excess calcium carbonate and gypsum in the soil, and occurrence of petrocalcic and petrogypsic soil phases is selected for the quantification of SQ6.

### 3.4.7 Workability (SQ7)

Diagnostic characteristics to indicate soil workability vary by type of management applied. Workability or ease of tillage depends upon interrelated soil characteristics, such as texture, structure, organic matter content, soil consistence/bulk density, the occurrence of gravel or stones in the profile or at the soil surface, and the presence of continuous hard rock at shallow depth, as well as rock outcrops. Some soils are easy to work, independent of moisture conditions, other soils are only manageable at an adequate moisture status, in particular, for manual cultivation or light machinery. Irregular soil depth, gravel and stones in the profile and rock outcrops, might prevent the use of heavy farm machinery. The soil constraints related to soil texture and soil structure are particularly affecting low and intermediate input farming land use types, whilst the constraints related to irregular soil depth and stony and rocky soil conditions primarily affect mechanized land preparation and harvesting operations, of high input level mechanised farming land use types. Workability constraints are, therefore, handled differently for low/intermediate and high inputs. The workability soil quality SQ7 includes physical hindrance to cultivation, and limitations to cultivation imposed by texture/clay mineralogy. The soil quality SQ7 is derived by combining the most limiting soil/soil phase attribute with the average of the remaining attribute conditions. Soil phases considered in the quantification of SQ7 are stony, lithic, petric, petrocalcic, petroferric, fragipan and duripan (FAO '74), and lithic, petroferric, rudic, skeletal, duripan and fragipan (FAO'90).

### 3.4.8 Overlay model

Seven soil quality layers were downloaded from the HWSD page, named SQ1, SQ2 through to SQ7. Each layer represents in raster format a specific soil quality, as shown in Table 8.

**Table 8** Soil qualities and associated soil characteristics

Code	Soil Qualities	Soil characteristics
SQ1	Nutrient availability	Soil texture, soil organic carbon, soil pH, total exchangeable bases
SQ2	Nutrient retention capacity	Soil Organic carbon, Soil texture, base saturation, cation exchange capacity of soil and of clay fraction
SQ3	Rooting conditions	Soil textures, bulk density, coarse fragments, vertic soil properties and soil phases affecting root penetration and soil depth and soil volume
SQ4	Oxygen availability to roots	Soil drainage and soil phases affecting soil drainage
SQ5	Excess salts	Soil salinity, soil sodicity and soil phases influencing salt conditions
SQ6	Toxicity	Calcium carbonate and gypsum
SQ7	Workability (constraining field management)	Soil texture, effective soil depth/volume, and soil phases constraining soil management (soil depth, rock outcrop, stoniness, gravel/concretions and hardpans)

Each soil layer a specific soil quality has been estimated for the sequence 1 soils of the HWSD in each grid cell, with maize as the reference crop. The derived maps for the individual soil qualities, therefore, represent the qualities of 'main soils' only within the HWSD. Each layer codes quality with values from 1 to 4, where 1 is very low and 4 is very high. Values 5 is rock and 6 is water. Firstly, a reclassification was applied to filter out classes 5 and 6 and leave only 1 to 4. Secondly, an additive weighted overlay model was run, such that:

$$WSQ_{n_{1to7}} = (SQ1+SQ2+SQ3+SQ4+SQ5+SQ6+SQ7)/7$$

Where WSQ is the weighted soil quality layer for maize, and SQs are the individual soil quality layers. Finally, average values of WSQ were calculated for each stratum and added in the attribute table in the field 'Soil productivity'.

### 3.5 Gross Primary Production (GPP)

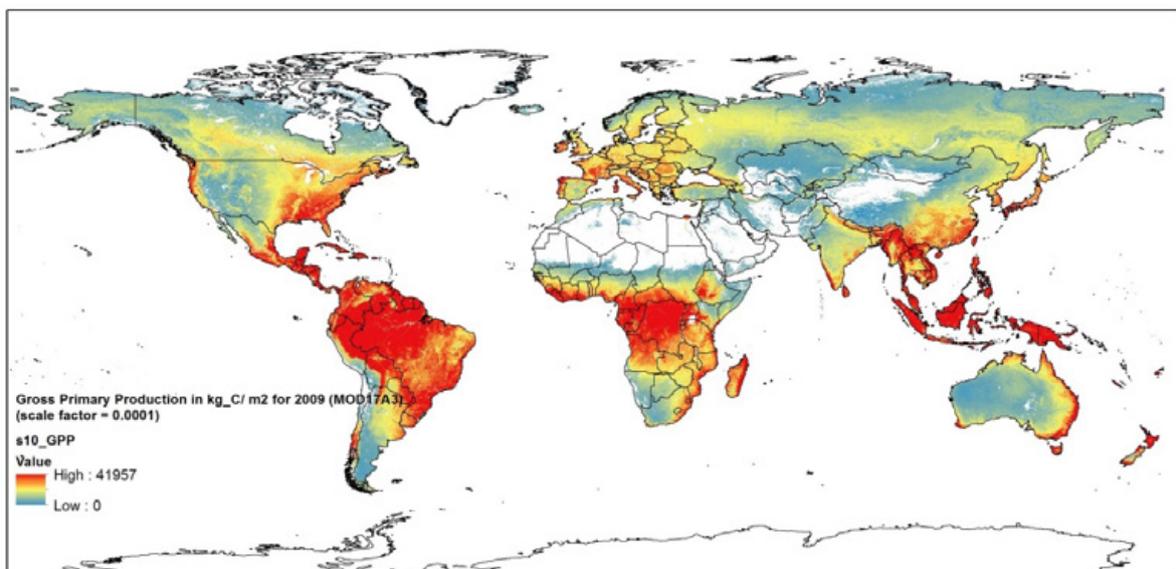
Primary production is the synthesis of organic compounds from atmospheric or aqueous carbon dioxide. Primary production is distinguished as either net (NPP) or gross (GPP), the former accounting for losses to processes, such as cellular respiration, the latter not. The Gross Primary Production (GPP) was derived from the MODIS product MOD17A3. The Net Primary Production (NPP), as well as GPP are both expressed in kg C/m<sup>2</sup> (data range is from 0 – 65,500 and scale factor 0.0001). For the calculation of global GPP, two main input parameters were used: FPAR/APAR, as well as land cover. MOD17A3 data are globally available from 2000 – 2009 at a resolution of 1 km. Version-55 Terra/MODIS NPP products are validated to Stage-3; this means that its accuracy was assessed and uncertainties in the product were well-established via independent measurements made in a systematic and statistically robust way that represents global conditions ([https://lpdaac.usgs.gov/products/modis\\_products\\_table/mod17a3](https://lpdaac.usgs.gov/products/modis_products_table/mod17a3)).

**Table 9** Primary production and plant biomass for the earth

Ecosystem type	Area (10 <sup>6</sup> km <sup>2</sup> )	Mean NPP (g/m <sup>2</sup> /yr)	World NPP (10 <sup>9</sup> ton/yr)	Mean biomass (kg/m <sup>2</sup> )	World biomass (10 <sup>9</sup> ton)
Cultivated land	14.0	650	9.1	1.0	14

(Source: From R.H. Whittaker, quoted in Peter Stiling (1996), "Ecology: Theories and Applications", Prentice Hall).

We experimented for Africa with the most recent GPP (2009), as well as with the mean value per grid cell over the period 2000 – 2009. GPP 2009 gave the best result for the segmentation procedure, as described in Chapter 4. Whilst forests were better visible in the mean over the period 2000-2009, agriculture was more visible with one single year, and GPP 2009 was, therefore, selected.



**Figure 14** Gross Primary Production (GPP) at a 1km resolution for the year 2009 (source: MOD17A3)

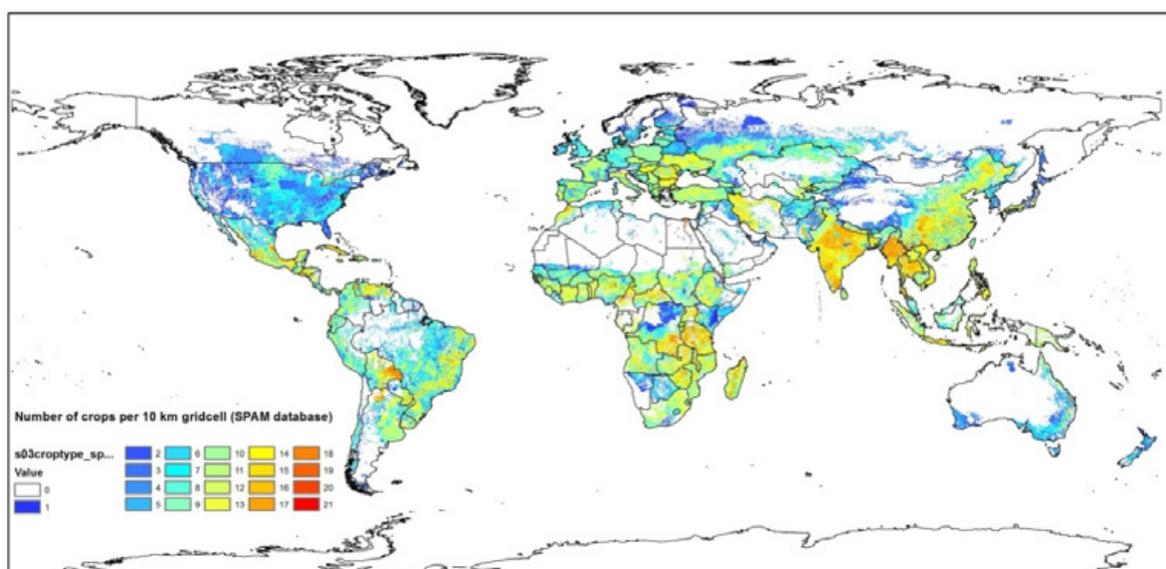
## 3.6 Crop types

Information on crop type diversity was derived from the state-of-the-art SPAM database of IFPRI (International Food Policy Research Institute in Washington, USA). SPAM stands for Spatial Production Allocation Model and reallocates agricultural census to a global grid with a 10 by 10 km spatial resolution (5 arc minute), in terms of major crop types, crop area, production and crop yield (You et al., 2006; You et al., 2009a; You et al., 2009b; You et al., 2014). The SPAM method is achieved through a downscaling approach that accounts for spatial variation in the biophysical conditions that influence the productivity of individual crops within the crop land extent, and that uses crop prices to weigh the gross revenue potential of alternate crops when considering how to prioritize the allocation of specific crops to individual grid cells. The proposed methodology also allows for the inclusion of partial, existing sources of evidence and feedback on local crop distribution patterns through the use of spatial allocation priors that are then subjected to an entropy-based optimisation procedure that imposes a range of consistency and aggregation constraints (You et al., 2014). The database contains information about the following 21 crops (rainfed and irrigated): wheat, rice, maize, barley, millet, sorghum, potato, sweet potato, cassava and yams, plantain and banana, soybean, dry beans, other pulse, sugar cane, sugar beets, coffee, cotton, other fibres, groundnuts, and other oil crops. Each crop is divided into three sub-systems: I – irrigated (“I”), H – rainfed high-input/commercial (“H”), and L – rainfed low-input/ subsistence (“L”). In fact, the database contains several databases, including: harvested area, physical area, production and yield. We used physical area mainly, as we are in the first instance interested in the total crop land area. We used Version 3.0.2 of the database (Version 3 Release 2; April 5, 2010), however, a newer version has just become available for the African continent.

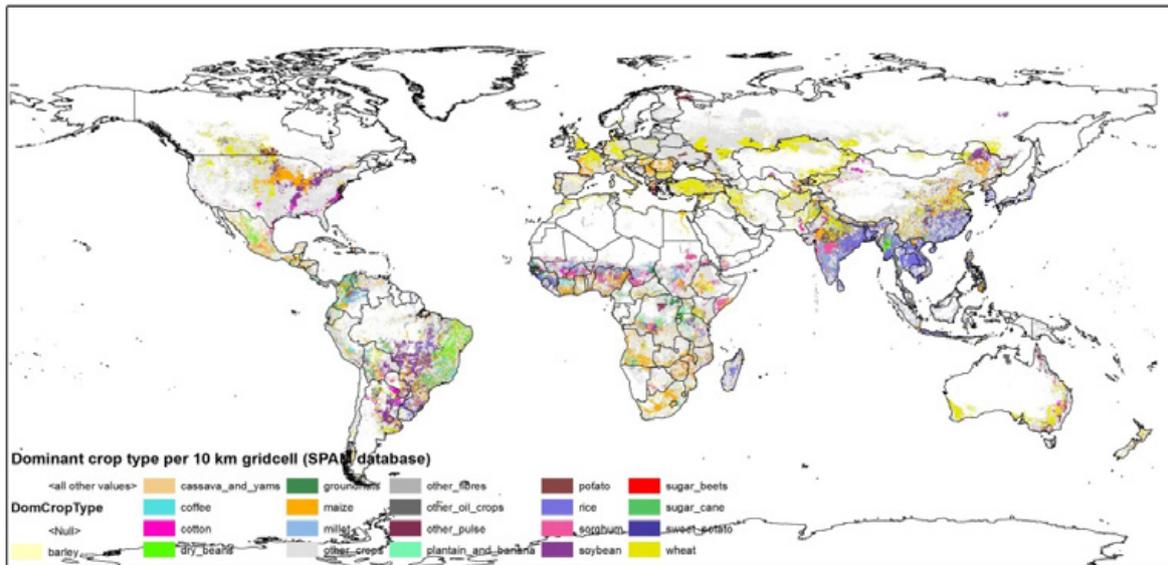
Additional processing was done to calculate:

- Number of crops per grid cell.
- Crop type combinations per grid cell.
- Indication of irrigation per grid cell.
- Percentage of crops per grid cell.
- Dominant crop per grid cell.

All output was exported to raster grid files with a 1km resolution and geo-referenced within the SIGMA database.



**Figure 15** Number of crops per grid cell, as derived from SPAM database



**Figure 16** Dominant crop type for each grid cell, as derived from SPAM database

## 3.7 Phenology

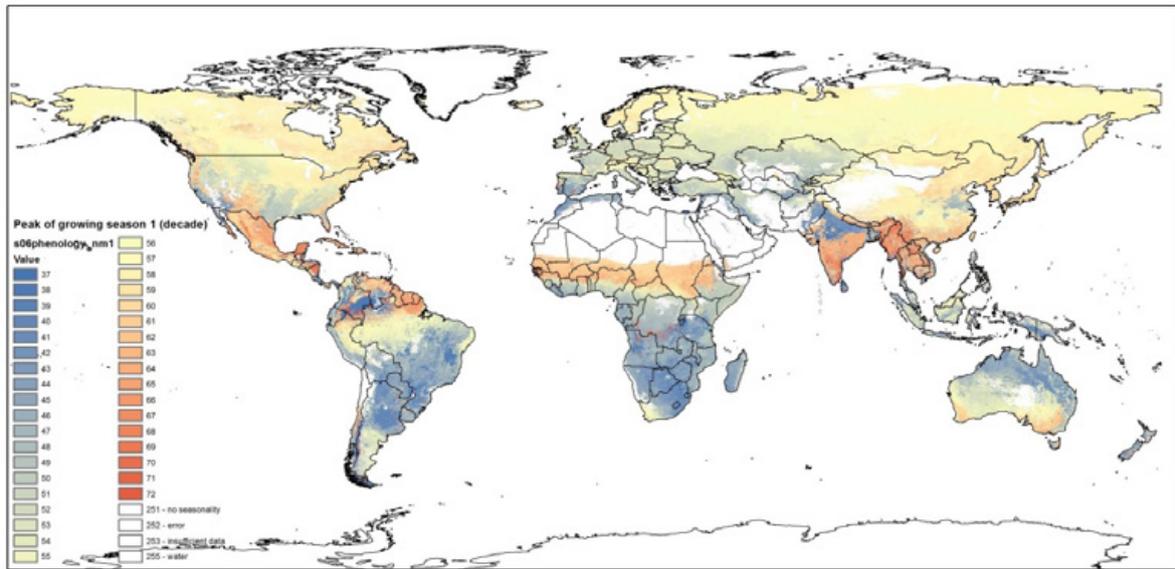
The main data source used to identify vegetation phenology on a global scale - the Phenodata database of VITO and is also directly related to SPIRITS software of JRC. SPIRITS (Software for the Processing and Interpretation of Remotely Sensed Image Time Series) - was developed by VITO for the Monitoring Agricultural Resources unit (MARS) of the Joint Research Centre (JRC) of the European Commission (EC). The software aims to facilitate the analysis of time series of low and medium resolution remote sensing images. SPIRITS was programmed by Dominique Haesen and Herman Eerens from VITO (<http://spirits.jrc.ec.europa.eu>).

### 3.7.1 Phenodata

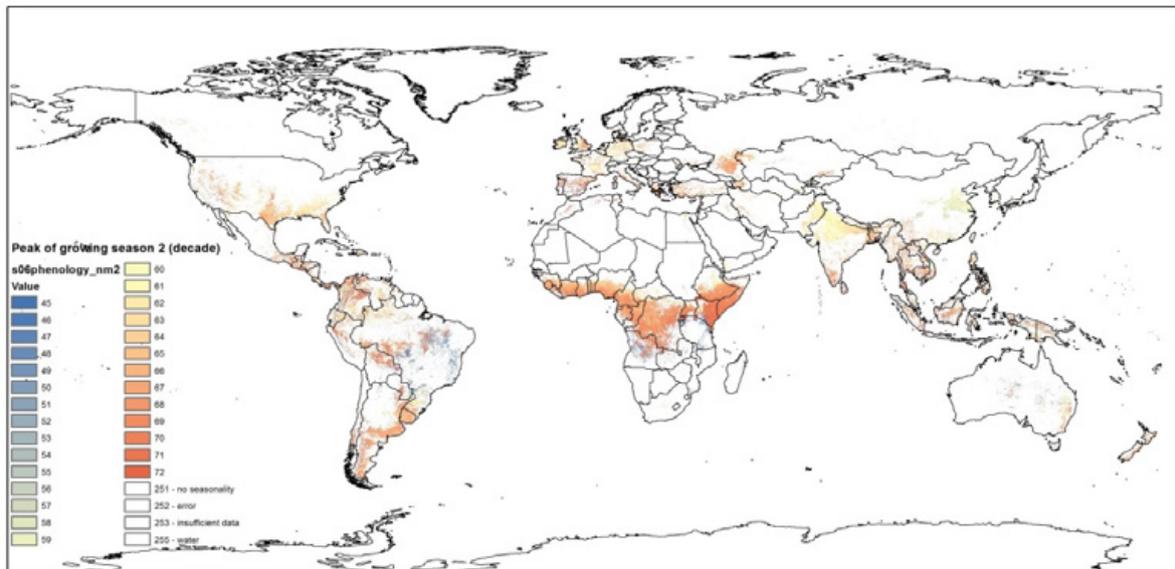
Phenological information, such as start of a growing season ( $ns$ ), maximum development of the biomass ( $nm$ ) and the end of the growing season ( $ne$ ) were derived from time series analysis of SPOT-VEGETATION satellite data with a 1 km spatial resolution over the period 1999-2011. The time series analysis using vegetation Indices ( $VI=NDVI, fAPAR, SAVI, \dots$ ) detects, the number of green seasons (0, 1 or at most 2) for a given civil year (January - December), and the start and end of the season for each such cycle (SOS and EOS). The input is a series of dekadal VI-images, each of which always cover exactly three civil years ( $3 \times 36=108$  dekads), with the 'target year' in the centre. Output raster files are created with SOS/EOS (Start of growing season, End of growing Season) for two seasons/cycles per year at most. These dates are expressed in dekads since the start of the three-yearly IN-series (1-36=previous year, 37-72=target year, 73-108=following year).

The most important derived phenological parameters are:

1.  $Ns1$ : Start of first growing season (dekad).
2.  $Ne1$ : End of first growing season (dekad).
3.  $Nm1$ : Peak of first growing season (dekad).
4.  $Ns2$ : Start of second growing season (dekad).
5.  $Ne2$ : End of second growing season (dekad).
6.  $Nm2$ : Peak of second growing season (dekad).



**Figure 17** Nm1image with dekad value when peak of Growing Season 1 has been reached (source VITO)



**Figure 18** Nm2image with dekad value when peak of Growing Season 2 has been reached (source VITO)

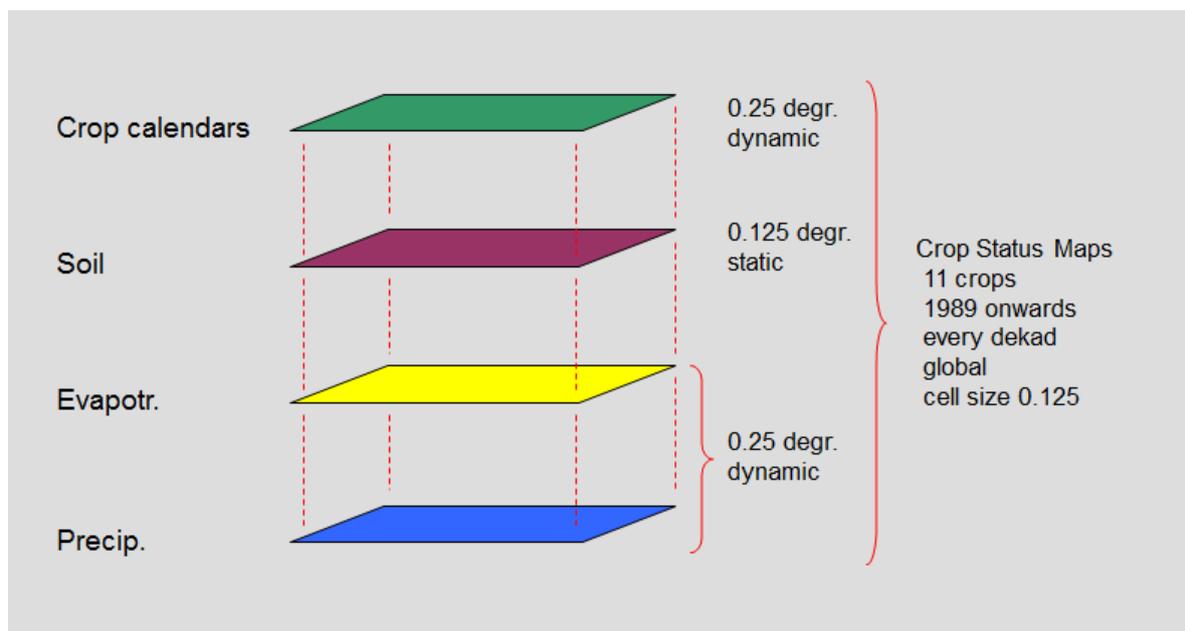
### 3.8 Number of growing cycles

The main data source used was the Phenodata database of VITO. However, the Phenodata database does not identify more than two growing seasons, all potential locations with more than two growing seasons in a year were identified on basis of GWSI (Global Water Satisfaction Index) data set of Wageningen Environmental Research (Alterra). Only those areas within the current maximised cropland mask (combination of GLCshare and IIASA cropland mask, see section 3.10) were analysed to identify the number of growing cycles.

### 3.8.1 Global Water Satisfaction Index (GWSI)

The MARS GWSI (Global Water Satisfaction Index) is an application developed by the JRC MARS Unit in collaboration with Wageningen Environmental Research (Alterra), and applied since 2008 in monitoring rainfed agriculture, with the main goal to detect near real time hot-spots due to water stress at a global level (Baruth B. et al, 2008; Massart M. et al, 2010). The approach of MARS GWSI is based on the agro-meteorological FAO water balance model using the meteorological information of the European Centre for Medium-Range Weather Forecast (ECMWF) and Tropical Applications of Meteorology using **SAT**ellite data and ground based observations (TAMSAT) (Frère and Popov, 1979; Gommès R.,1993; Rojas et al, 2005). The model compares water requirements during the crop growing season with calculated available water, which includes rainfall and easily available water stored in the soil at the start of the growing season. The approach determines a cumulative water balance for each period of 10 days ("dekad") from sowing to harvest. The output is an index expressing the percentage of crop water requirements met. The MARS GWSI is based on the FAO Crop Specific Soil Water Balance (CSSWB) (Frere and Popov, 1986), a simple and physically sound soil water balance developed for operational use, but it follows a completely different spatial approach by using grid cells (0.125 degree) and specific *ad hoc* models for the identification of the crop specific optimal calendar (climatic optimal growing season model), the detection of sowing dates (sowing date detection rule) and the estimation of the initial soil moisture content.

GWSI is calculated for each 10-day period for a fine grid cell of 0.125 degree, and has a 0-100% value range as an indication of water stress, next to number, length, start and climatic seasons favourable for specific crops. In general, a crop requires a 90 days growing season with no water deficit. So, at a location with more than 30 dekad (300 days) with no water deficit, three crops could be cultivated theoretically in one year.



**Figure 19** Schematic illustration of the spatial schematisation in MARS GWSI (Pini et al., 2015)

### 3.8.2 Integration Phenodat with GWSI

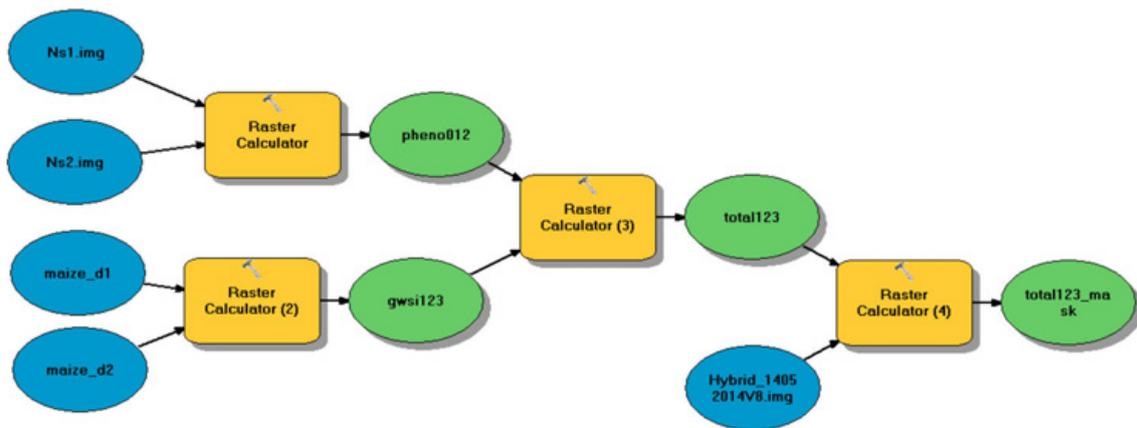
The values of the number of growing seasons from 0, 1, or 2 are derived from Phenodat database of VITO (see also section 3.5 on Phenology). The differentiation towards > 2 growing seasons is made on basis of GWSI. To identify the potential of having more than two growing seasons, the following criteria were applied on the GWSI database for maize (maize\_d1 and maize\_d2):

$GWSI_{d1} \geq 30$  OR  $(GWSI_{d2} \geq 18$  AND  $GWSI_{d1} \geq 9)$  OR  $(GWSI_{d1} \geq 18$  AND  $GWSI_{d2} \geq 9)$

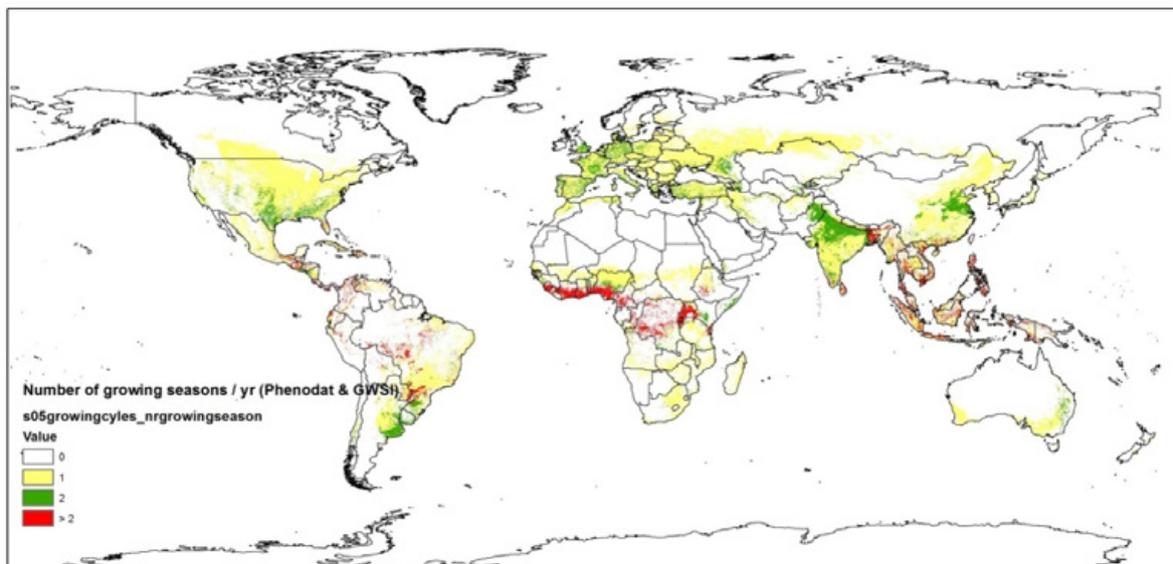
d1 = length of the first climate season favourable for specific crops.  
d2 = length of the second season favourable for specific crops.

So, if first climate season is favourable for a specific crop for more than 300 days it enables three crop seasons. With first- and second climatic seasons within one year favourable for a specific crop, the second climate season should have at least 180 days to enable three crop rotations in a year. Third option is when first climate season is more than 180 days favourable for a specific crop and there is also a second climate season (at least 90 days), it is possible to have three crops rotations in a year. Note that the GWSI analysis was completed for maize, but it is assumed that if climate conditions are favourable to grow maize (with no water stress), in principle, other crops could be cultivated as well.

One, two, or more growing seasons are identified only for those areas where crops are cultivated (use of the IIASA cropland mask, see section 3.10).



**Figure 20** Flowchart of the methodology to identify one, two, or more than two growing seasons based on Phenological database of VITO (Phenodat) and Global Water Stress Index (GWSI) database of Wageningen Environmental Research (Alterra)



**Figure 21** Number of growing seasons over a year based on Phenodat (VITO) and GWSI (Wageningen Environmental Research (Alterra))

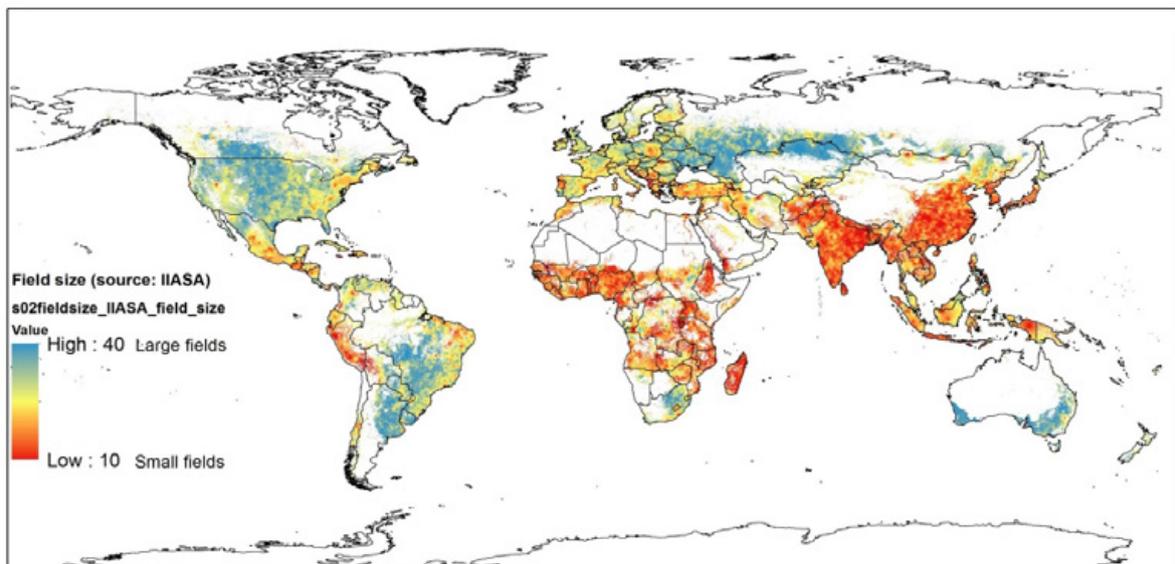
### 3.9 Field size

The field size data layer was produced by IIASA and is based on crowd-sourcing achieved through the GEOWIKI approach, in which people were asked to make an estimation of the field size on basis of photo interpretations in Google Earth. Geo-Wiki is a platform that provides citizens with the means to engage in environmental monitoring of the earth by providing feedback on existing spatial information overlaid on satellite imagery, or by contributing entirely new data. Data can be inputted via the traditional desktop platform or mobile devices, with campaigns and games used to incentivize input. Resulting data are available without restriction (<http://www.geo-wiki.org/>).

During the first Human Impact campaign, IIASA gathered additional information, alongside land cover. If the volunteer selected the land cover type 'Cropland' or 'Mosaic of cropland / natural vegetation', an additional box appeared on Geo-Wiki, requesting that the user determined and entered the field size using categories that range from very small to large. As training, volunteers could click on a link to see examples of how different field sizes appear on Google Earth. These data on field size were then interpolated using a simple inverse distance weighted procedure to produce the first global map of field size. Although there are clearly gaps and interpolation artefacts as a result of the low density of validation points in places, the overall pattern corresponds very well to what we would expect, i.e. large fields with intensive agriculture in the USA, Canada and Europe on one end of the spectrum and very small fields in Western Africa, parts of China, etc. on the other. Interest in developing this product further has already been expressed by members of the Group on Earth Observation (GEO) Global Agricultural Monitoring and Early Warning System Task (AG-01) [13], who recognize the value in such a product for monitoring purposes, i.e. larger field sizes can be monitored by coarser resolution imagery, whilst small fields require higher resolution.

The data range of the field size layer is between 10 and 40, which can be translated as an approximation as follows:

- 1-10 → very small fields (< 0.5 ha).
- 11-20 → small fields (0.5 – 2 ha).
- 21-30 → medium fields (2 – 20 ha).
- 31-40 → large fields (> 20 ha).



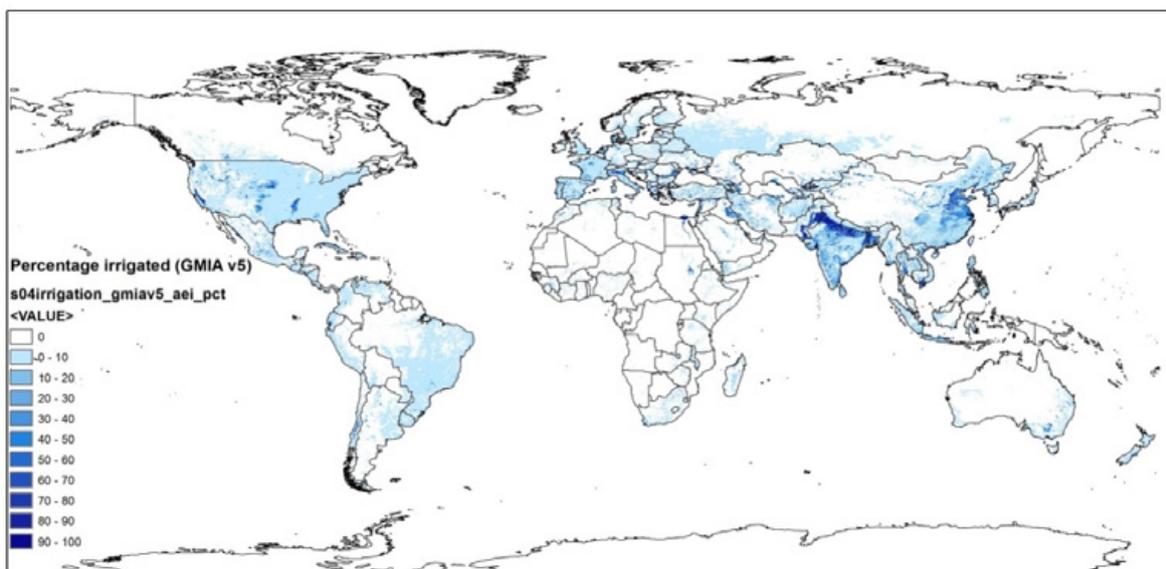
**Figure 22** Field size with a relative scale range from value 10 (very small fields) to value 40 (very large fields). Through Geo-Wiki, users were asked to determine and enter the field size using categories ranging from very small to large. These point observations were interpolated using a simple inverse distance weighted procedure for the complete crop land area

## 3.10 Irrigated agriculture

Many data sets are available for the global distribution of irrigated areas, such as Global Map of Irrigation Areas (GMIA version 5.0, (Rheinische Friedrich-Wilhelms-University, Bonn, Germany / FAO) and Global Irrigated Area Mapping (GIAM version 2.0, source: IWMI). The GMIA was identified as most detailed and accurate database for irrigated agriculture (percentage irrigated per grid cell) and was therefore selected. The following section will give some background information about GMIA.

### 3.10.1 Global Map of Irrigation Areas (GMIA)

The selected data layer from GMIA\_v5 (2005) shows the fraction of area equipped for irrigation on a raster with a resolution of 5 arc minutes, approximately 10 kilometres (Siebert et al., 2015). Additional map layers show the percentage of the area equipped for irrigation that was actually used for irrigation and the percentages of the area equipped for irrigation that was irrigated with groundwater, surface water or non-conventional sources of water. Please note that information for the additional layers on area actually irrigated or on the water source for irrigation was derived from statistical survey data (e.g. census reports). Therefore all grid cells belonging to the same statistical unit will have the same value. Consequently, the accuracy at pixel level will be very limited, depending upon the size of the statistical unit (Siebert et al., 2015). For more information, visit <http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm>.



**Figure 23** Global Map of Irrigation Areas version 5. Source: Rheinische Friedrich-Wilhelms-University, Bonn, Germany and FAO, Rome, Italy

## 3.11 Crop land mask

The crop land mask is based on the integration of IIASA cropland mask and FAO GLCshare. These are explained in the next sections.

### 3.11.1 IIASA hybrid crop land layer

IIASA has developed a global cropland extent map within the framework of the Geo-Wiki project. The Geo-Wiki Project is a global network of volunteers who wish to help improve the quality of global land cover maps. Since large differences occur between existing global land cover maps, current ecosystem and land-use science lacks crucial accurate data (e.g. to determine the potential of additional agricultural land available to grow crops in Africa). Volunteers are asked to review hotspot maps of

global land cover disagreement and determine, based on what they actually see in Google Earth and their local knowledge, if the land cover maps are correct or incorrect. Their input is recorded in a database, along with uploaded photos, to be used in the future for the creation of a new and improved global land cover map. Global and Regional Land Cover and Cropland Data Sets that have been used for the creation of the IIASA global hybrid crop land layer and was built from:

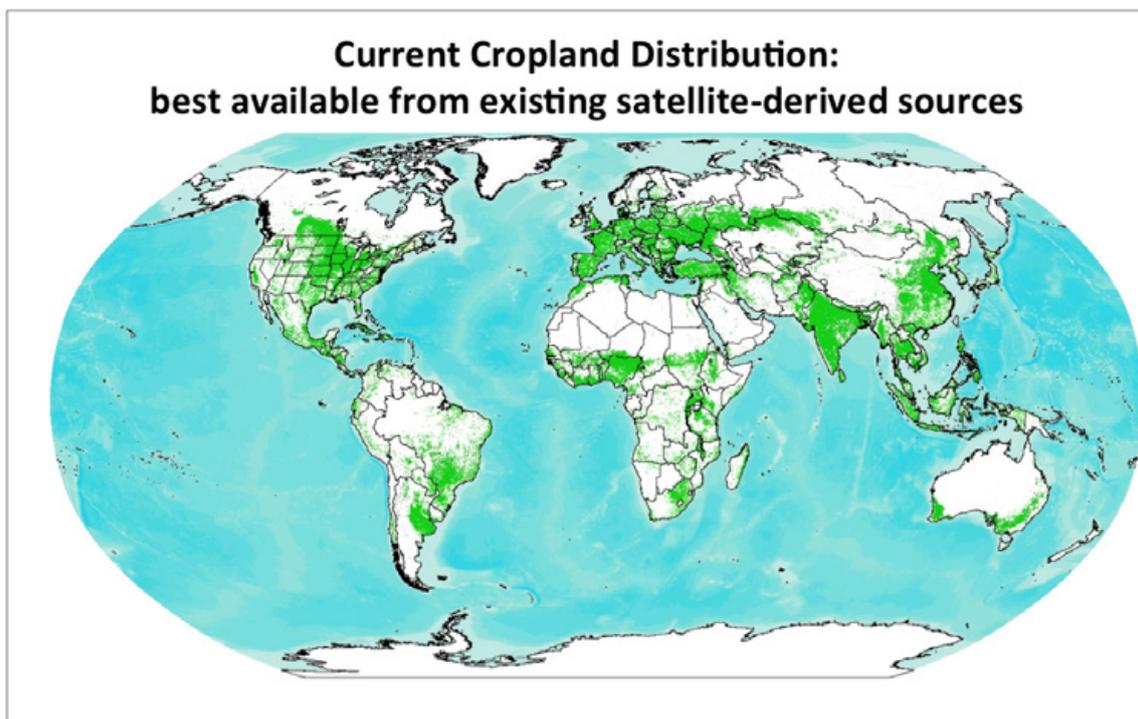
- MODIS Land Cover product (2005).
- GlobCover (2005).
- Global Land Cover 2000 (GLC-2000).
- Global Cropland Extent – South Dakota State University.
- Composite Africa cropland map (composed of eight individual maps), JRC, Italy.

Alongside the above-mentioned databases, national maps in the public domain were also used (see table below).

<b>Europe</b>		
• EU 27 States		
<b>North America</b>		
• USA		
<b>South America</b>		
• Summer crops Brazil		
<b>Africa (AfriCover countries + contributions made)</b>		
• Burkina Faso	• Kenya	• Somalia
• Burundi	• Lesotho	• Sudan
• DRC	• Mali	• Swaziland
• Egypt	• Uganda	• Tanzania
• Eritrea	• Rwanda	• Tunisia
• Gambia	• Senegal	
	• South Africa	
<b>Asia</b>		
• China	• Northern Part of India	
<b>Australia &amp; the Pacific</b>		
• Australia		

**Figure 24** National- and other land cover data sources also used in the production of the IIASA hybrid crop land layer

The newest hybrid cropland data layers can be downloaded from Beta-hybrid.geo-wiki.org. We used Version 8 (IIASA\_cropland\_hybrid\_14052014v8).



**Figure 25** IIASA hybrid crop land layer based on a multitude of EO data sources

### 3.11.2 FAO GLC-SHARE

The new FAO land cover product: the “Global Land Cover-SHARE (GLC-SHARE)” has been recently published. It brings together global land cover data for the first time and represents the most reliable global view of planetary land cover assembled to-date. This new product collects previously scattered and unharmonized land cover information from around the globe into one centralised database, marking a major improvement in information regarding the physical characteristics of the Earth’s surface. Up until now, one of the major challenges for countries and organisations was to identify, measure and efficiently record a good global overview of land cover. The GLC-SHARE database has been compiled, just as the IIASA hybrid cropland layers, from multiple sources, which were quality controlled and harmonised using internationally accepted definitions and standards, with the aim to bring a wealth of country-level information into one consolidated dataset that spans the entire planet. Applications of the database include assessment, monitoring and reporting of the distribution of the major land cover classes, land suitability evaluation, land accounting, environmental accounting, climate change impact assessments in productivity and yields, land use planning and sustainable development addressing food security and environmental threats (<http://www.glcn.org>).

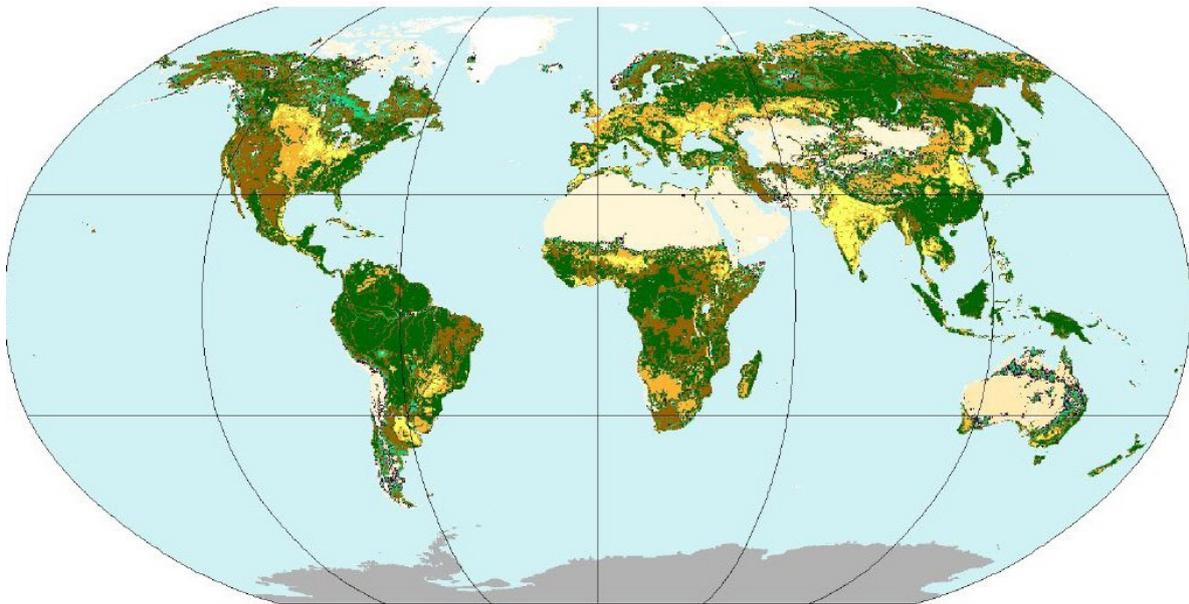


Figure 3 – Distribution of dominant GLC-SHARE Land Cover Database.

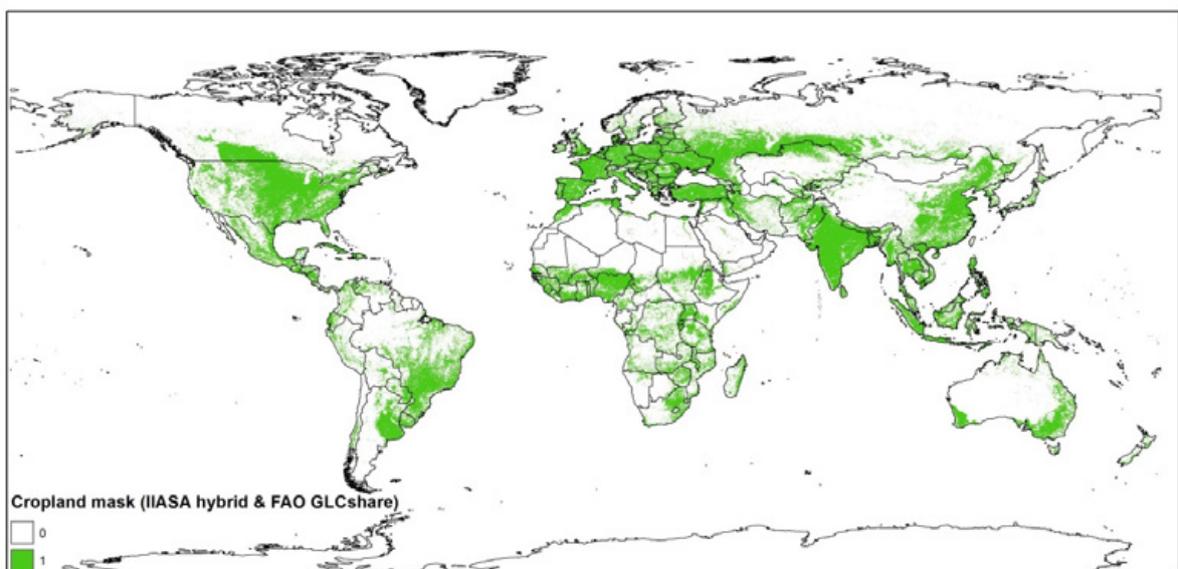


**Figure 26** GLC-SHARE land cover database initiated by FAO

For our purposes, we used the most recent version 11 of the FAO GLC share (FAO\_GLC\_share\_02\_v11\_update).

### 3.11.3 Integration of IIASA crop land mask and FAO GLC-SHARE

Since IIASA hybrid layer and FAO GLC share show a different crop land extent for, not all, but many regions, it was decided to use a pragmatic approach in this stage. A new crop land layer was produced, in which the two data sources were both included to create maximum extent of the current crop land.



**Figure 27** New crop land mask with a 1km spatial resolution based on the integration of GLC-SHARE and IIASA hybrid crop land layer

---

## 4 Global stratification methodology

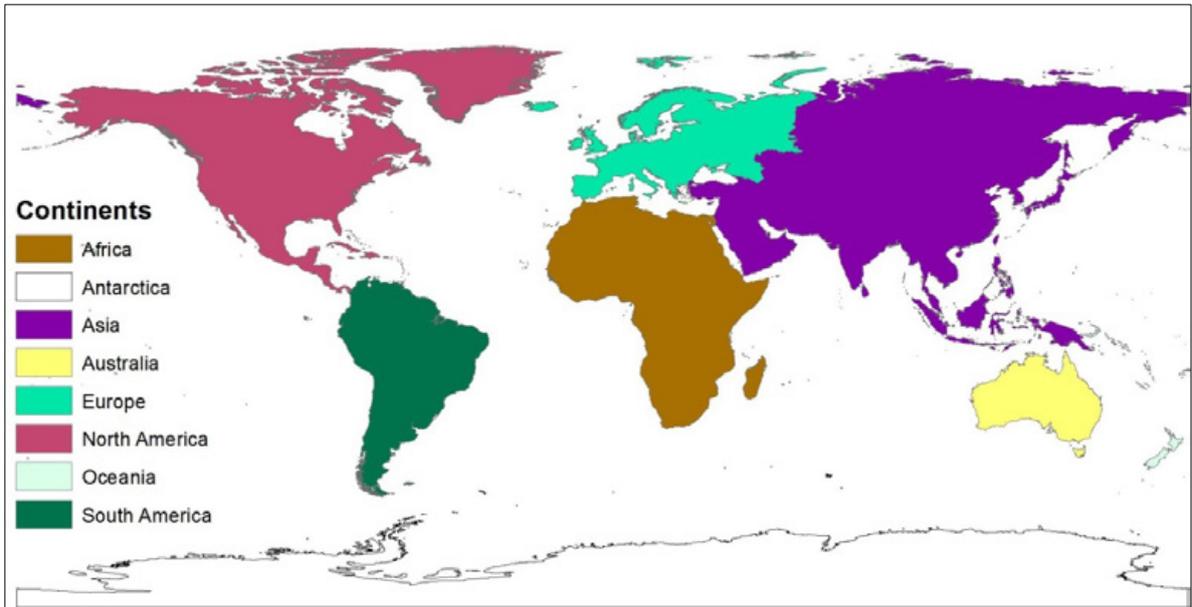
The global stratification methodology of dividing the globe into relatively homogenous agro-ecological units by segmentation techniques is explained by using examples for the African continent. The selection of the input global data layers for the segmentation is crucial for the resulting output at various hierarchical scales. Therefore, the final 13 selected global environmental input layers for the segmentation process are also highlighted here. The 13 data layers were selected on basis of their relevance for agricultural production and SIGMA's cropland classification using GAES as a stratification tool. The segmentations at four hierarchical scales were implemented using eCognition on a continental basis. In other words, GAES is implemented on four nested hierarchical spatial scales, with Level 4 as the most detailed spatial scale. Clustering techniques provided less useful results due to the fragmented 'salt and pepper' effect, and so, providing scattered clusters that are difficult to handle as strata for decision-making.

### 4.1 Global input data layers

Table 9 gives an overview of the input global data layers, for segmentations. Before performing any segmentation, the data layers required normalizing (range from 0-100). Normalization is necessary to give each data layer an equal weight in the segmentation process. Normalization of the data layers was implemented on a continent-by-continent basis (Fig. 27). The 13 input layers were selected according to the needs of WP3, WP4 and WP5, upon discussions of a workshop at FAO in April 2014.

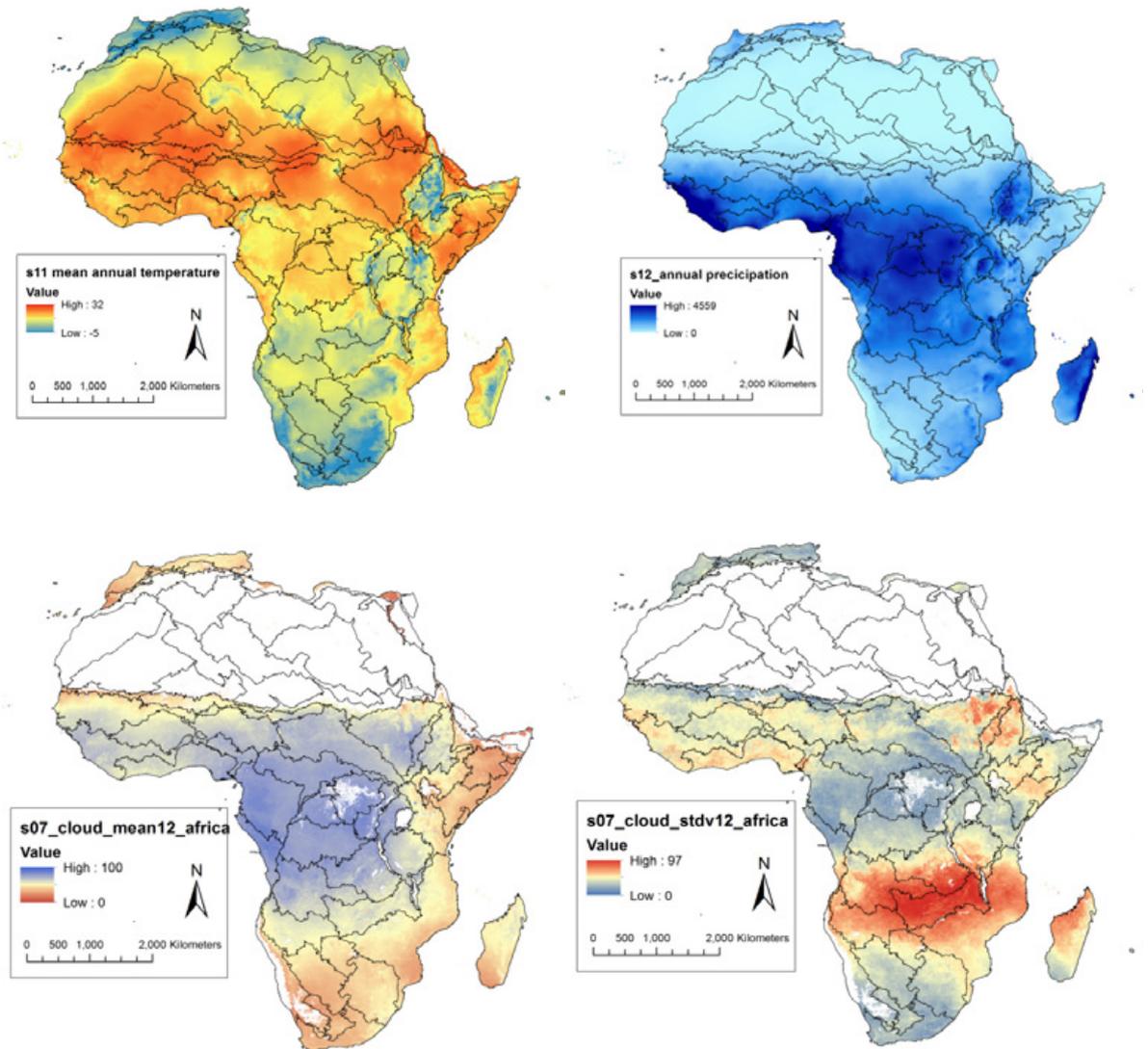
**Table 10** Summary of all 13 global input data layers that have been used as an input for segmentation at various spatial scales

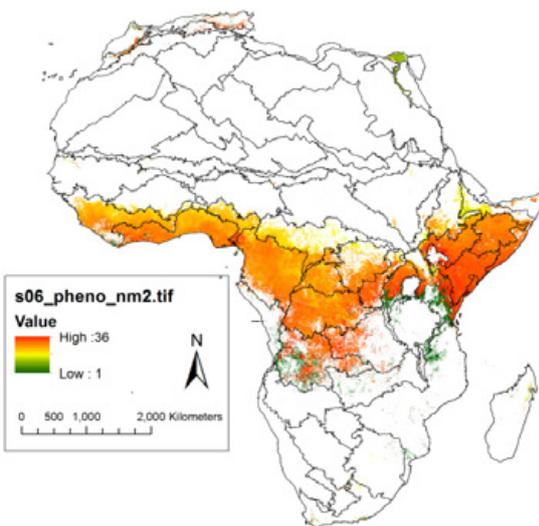
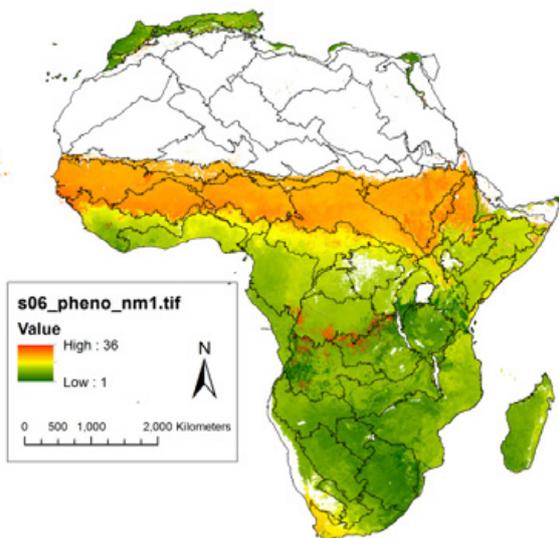
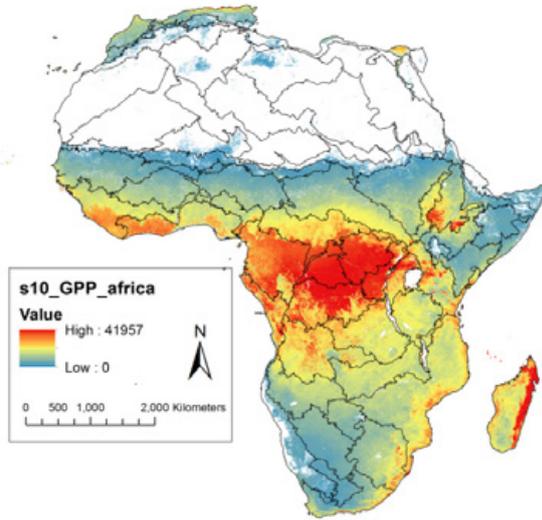
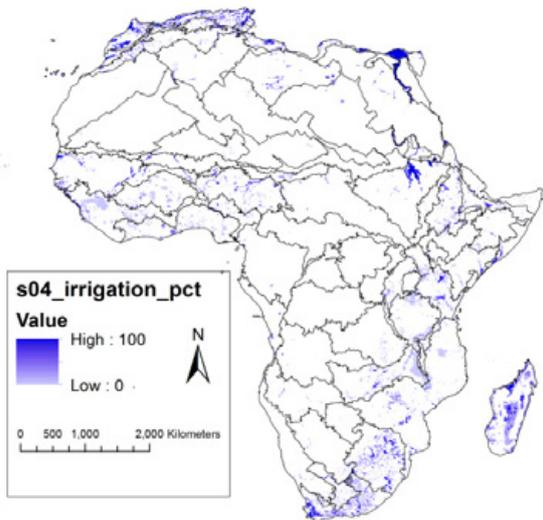
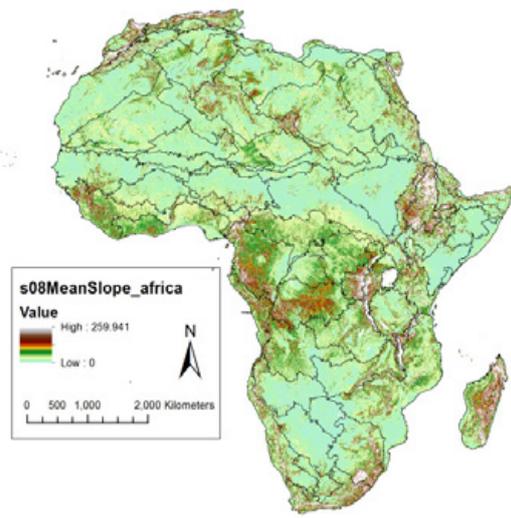
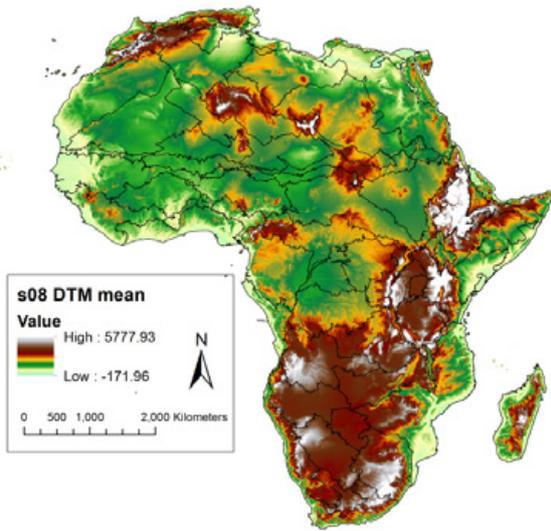
Nr	Theme	Parameter layers
1	Climate	Annual mean temperature
2	Climate	Annual total precipitation
3	Climate	Mean cloud fraction over the growing season
4	Climate	Standard deviation cloud fraction over the growing season
5	Altitude	DTM mean altitude
6	Altitude	DTM mean slope
7	Irrigation	Percentage irrigated
8	Production	Gross Primary Production (GPP)
9	Phenology	Decade when maximum biomass is reached in growing season 1
10	Phenology	Decade when maximum biomass is reached in growing season 2
11	Growing cycles	0, 1, 2 or more growing cycles
12	Crop types	Number of crop types
13	Field size	Size of field parcels

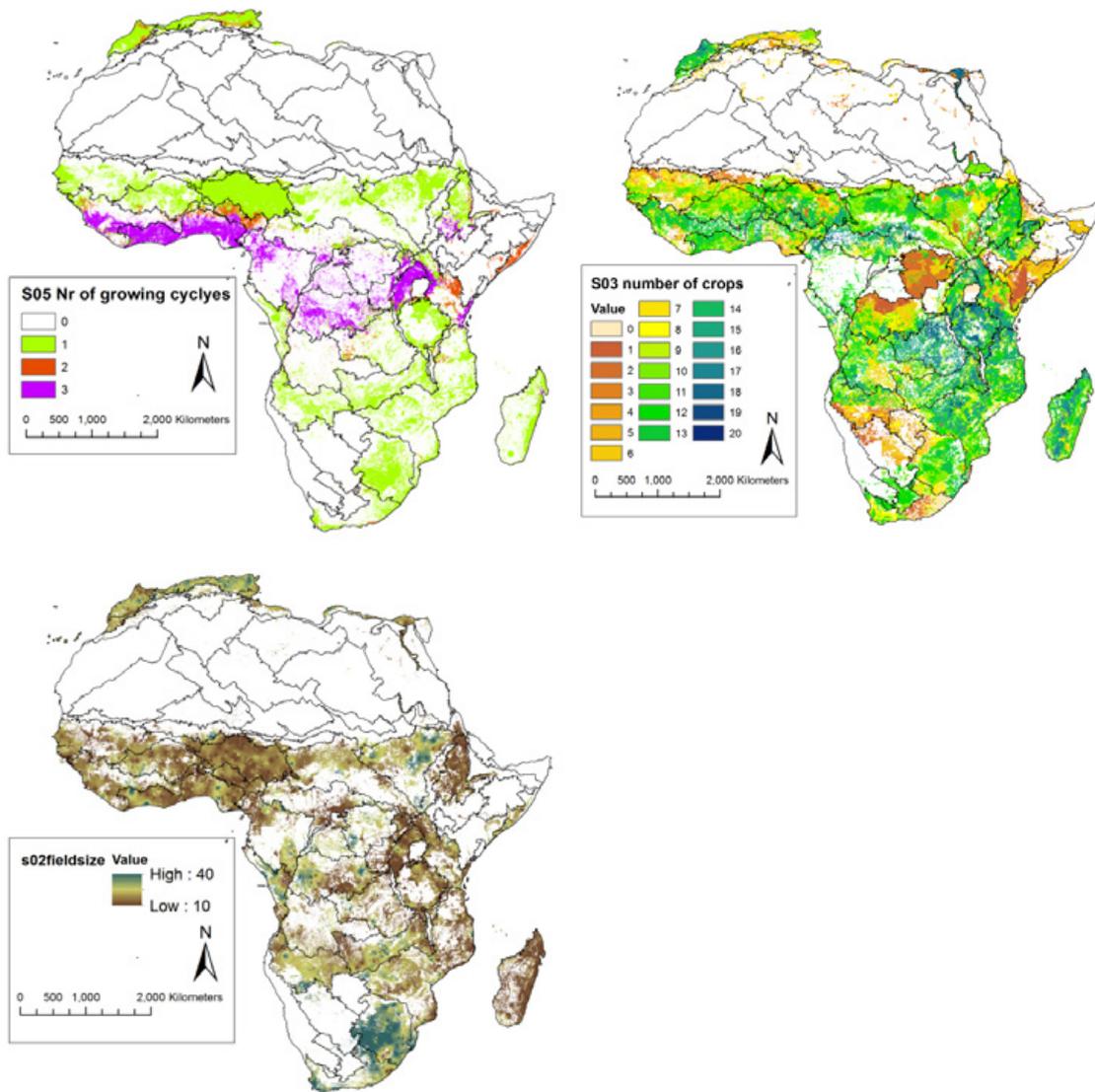


**Figure 28** Segmentation has been performed per continent. Only Antarctica and Oceania were not segmented

The methodology will be explained further on basis of examples for the African continent.





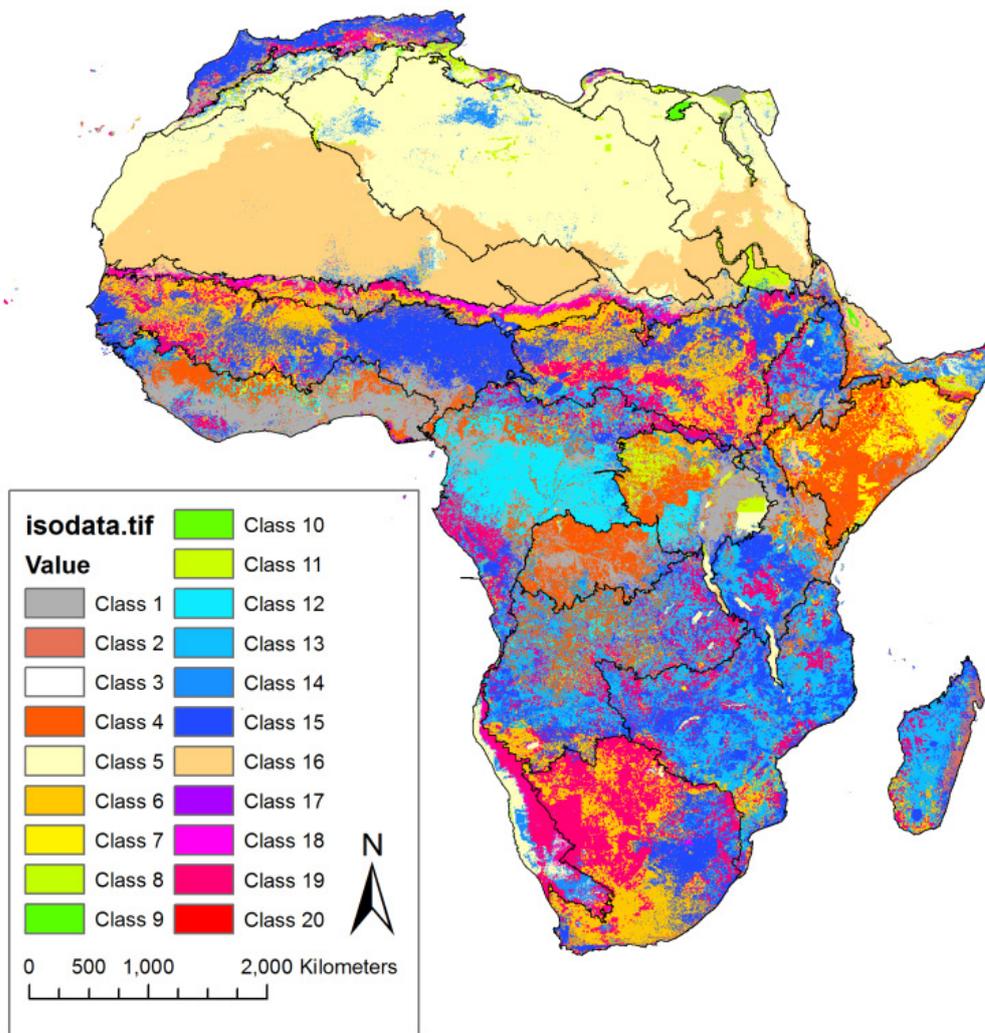


**Figure 29** Overview of the 13 input data layers for the African continent overlaid with the resulting segmentation (3rd hierarchical level). The 13 layers used are : 1) Annual mean temperature, 2) Annual total precipitation, 3) Mean cloud fraction over the growing season, 4) Standard deviation cloud fraction over the growing season, 5) DTM mean altitude, 6) DTM mean slope, 7) Percentage irrigated, 8) Gross Primary Production (GPP), 9) Decade when maximum biomass is reached in growing season 1, 10) Decade when maximum biomass is reached in growing season 2, 11) 0, 1, 2 or more growing cycles, 12) Number of crops types, and 13) Size of the field parcels

## 4.2 Isodata clustering

An isodata clustering was performed in ENVI 5.2 on Africa, on the 13 normalized data layers using the following parameters:

- Minimum number of classes: 10.
- Maximum number of classes: 20.
- Maximum number of iterations: 10.
- Minimum number of pixels in a class: 10,000 (km<sup>2</sup>).



**Figure 30** Example of an isodata clustering overlaid with GAES Level 2 segmentation

As shown in Figure 30, isodata clustering results are less useful due to the fragmented 'salt-and-pepper' effect of the individual grid cells that provides scattered clusters that are difficult to handle as strata for decision-making. Segmentation is believed to provide better results for a global stratification, as the resultant strata are simpler and more compact.

### 4.3 Segmentation

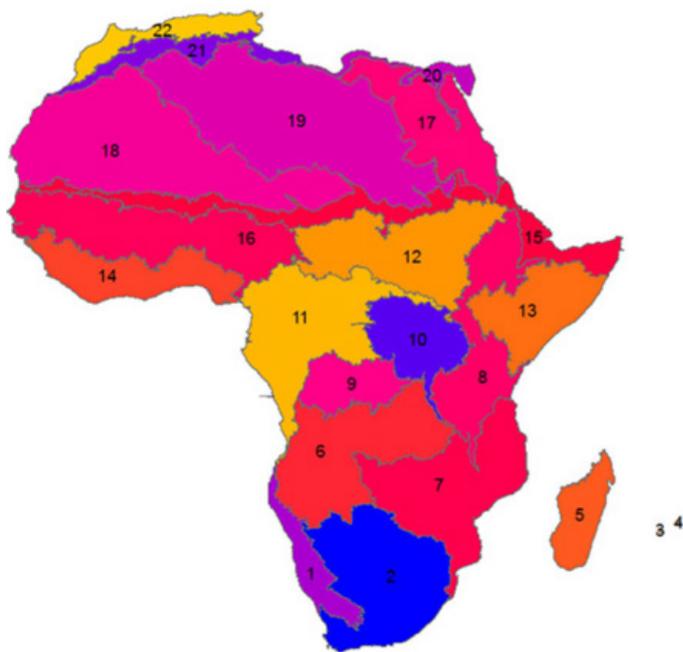
It is interesting to see that the isodata clustering provides clusters that are very much fragmented and dispersed ('salt-and-pepper' effect), but at the same resembles to a large degree the segmented strata. A segmentation with large polygons as strata are much easier to handle in a strategy towards crop land mask classification than the dispersed isodata clusters. This is the main reason that segmentation of relatively homogeneous strata is preferred above pixel-based clustering techniques. A multi-resolution segmentation has been performed in such a way that more compact and simpler strata can be identified. Segmentation is performed in Ecognition Developer 9.0 (Trimble), with the following parameters:

- Scale parameter: 4000 (Level 1); 2000 (Level 2); 1000 (Level 3); 500 (Level 4).
- Shape parameter: 0.2.
- Compactness parameter: 0.0.

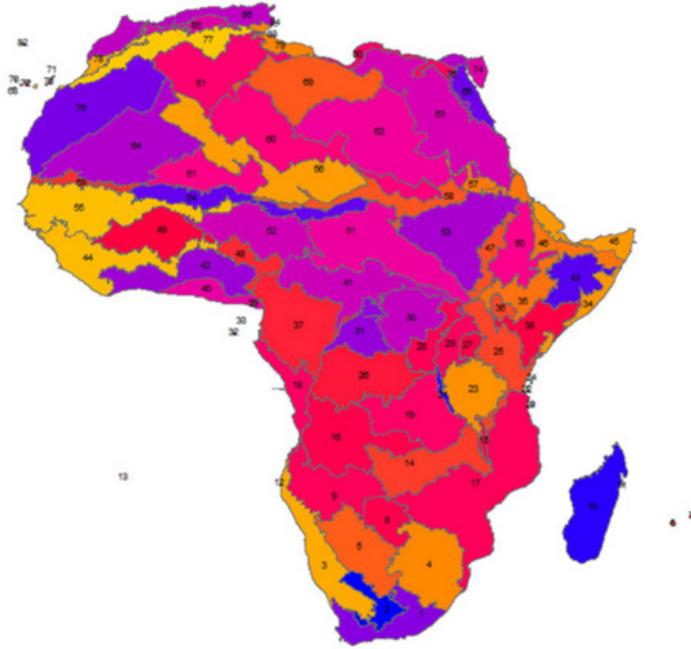
The figures below represent the segmentations at four different hierarchal and nested levels. Level 4 is the most detailed, whilst Level 1 shows only eight strata for the entire African continent.



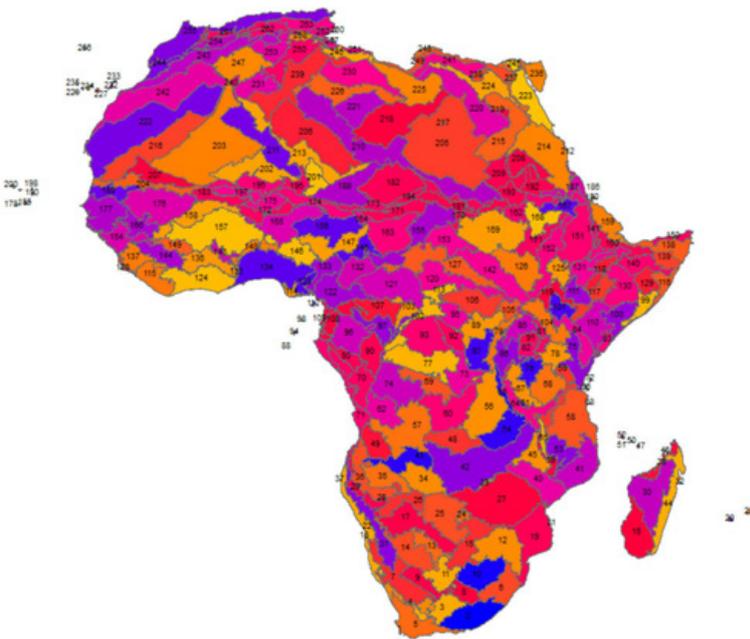
**Figure 31** Example Level 1 GAES for Africa (eight strata)



**Figure 32** Example Level 2 GAES for Africa (22 strata)



**Figure 33** Example Level 3 GAES for Africa (86 strata)

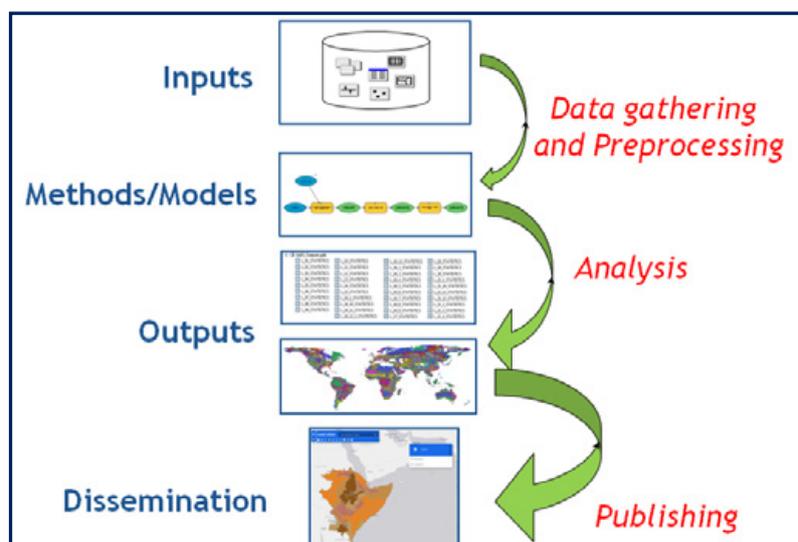


**Figure 34** Example Level 4 GAES for Africa (263 strata)

The next chapter describes the further characterisation and description of the four hierarchical stratifications, as obtained by segmentation, resulting in the GAES database.

## 5 Results: a Global Agro-Environmental Stratification (GAES)

This chapter finalises the report with the further description and characterisation of segments in the GAES database as a follow-up of the global segmentation per continent. As described in the former Chapter, GAES has four hierarchical spatial levels, of which Level 4 is the most detailed stratification. The population of the GAES strata with external environmental data sets at all four hierarchical levels in terms of their major characteristics for each strata, resulted in more than 80 attributes. This process included a number of tasks, such as additional data collection and analysis, and finally the dissemination of the outputs. Large part of the data had been, of course, already collected for the segmentation itself. A simple typology of the Agro-Environmental (AE) strata is finally based on four attributes: 1) climate; 2) altitude; 3) parent material; 4) land cover.



**Figure 35** Workflow and artefacts of GAES

Final GAES products are:

- The personal geo-database file GAESv01a.gdb with all four hierarchical levels of the stratification as feature classes (GAES\_L1\_01a; GAES\_L2\_01a; GAES\_L3\_01a; GAES\_L4\_01a). All four hierarchical levels with all attributes are also available as shape files. In total, each GAES level has 92 variables, including a narrative field "Descriptor" which provides a typology as a small narrative to describe the most relevant biotic/abiotic characteristic of each polygon.
- GAES.gdb: This geo-database includes all the input data layers and the developed spatial tools for running the characterisation and harmonisation.
- Two geo-processing tools: GAES and Characterization. Such tools are python scripts stored in a tool box within the GAES.gdb. They are used to run characterisation and harmonisation of results through a user-friendly interface.
- One GAES web map service (WMS) published via FAO's GIS server site. The WMS can be used by any GIS client and be embedded in other GIS projects for display or analysis (desktop/webgis/mobile/table).
- One GAES Web Application published via FAO's GIS server site (<http://hqfao.maps.arcgis.com/apps/Viewer/index.html?appid=27f8cd872dc4488c82140636153b2adc>). This webapp can be approached by web browser and can be used to visualize the result of GAES. It facilitates the exploration of layers, variables and statistics for the 70 variables within each strata globally.

## 5.1 GAES.gdb: data gathering, input data and preprocessing

Data gathering was performed after an intense data planning phase carried out jointly by FAO and Wageningen Environmental Research (Alterra). During the planning phase, descriptive variables required for the characterisation of the stratification were first identified, based on their relevance for a global agro-environmental stratification.

Subsequently, data sources were identified and the appropriate analytical methods and the scripting/spatial tools were identified and agreed.

**Table 11** Layers and sub-layers used as an input for characterisation of strata

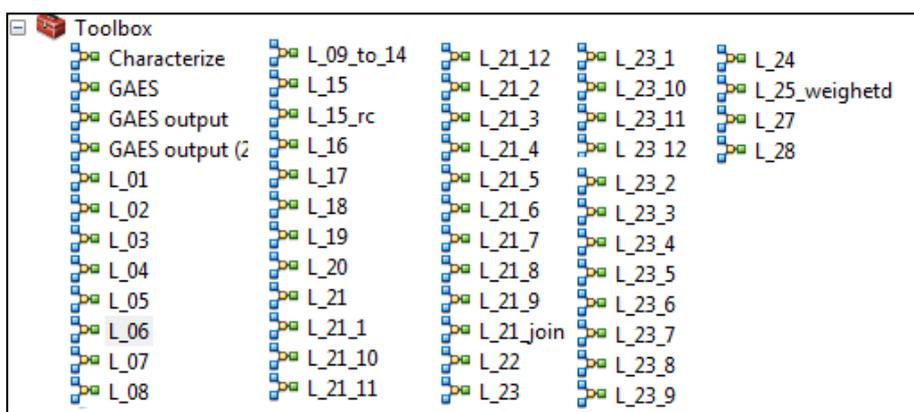
Layer name	Theme	Script/method	Variable type	Source
L1	Climatic zones	Reclassification using Metzger 2013 criteria Zonal Stat. majority	Qualitative	GENZ
L2	Elevation zones	Reclassification using LANMAP (Mücher et al, 2010) Zonal Stat. majority	Numeric categorical	GTOPO30
L3	Parent material	Zonal stat. majority	Qualitative	HWSD (FAO)
L4	Land cover (dominant)	Zonal stat. majority	Qualitative	GLC-SHARE
L5	Cropping area	Calculate real surface Zonal stat. sum	Numeric categorical	IIASA, FAO
L6	Field size	Zonal stat, min max mean	Numeric categorical	IIASA
L7	Crop type (dominant)	Zonal majority	Qualitative	IFPRI
L8	Irrigation	Calculate real surface Zonal stat. sum	Numeric continuous	Siebert et al., 2015
L9	Growing cycles 1 season start	Zonal stat. min, max	Numeric categorical	VITO, ALTERRA
L10	Growing cycles 1 season end	Zonal stat. min, max	Numeric categorical	VITO, ALTERRA
L11	Growing cycles 1 season max	Zonal stat. min, max	Numeric categorical	VITO, ALTERRA
L12	Growing cycles 2 season start	Zonal stat. min, max	Numeric categorical	VITO, ALTERRA
L13	Growing cycles 2 season end	Zonal stat. min, max	Numeric categorical	VITO, ALTERRA
L14	Growing cycles 2 season max	Zonal stat. min, max	Numeric categorical	VITO, ALTERRA
L15	Area under 1/2/3 growing seasons	Calculation or real surface/Cross tabulation/Zonal stat. sum	Numeric continuous	VITO, ALTERRA
L16	Monthly mean cloud fraction	Sum decadal layers to monthly mean Zonal stat mean for each month layer	Numeric continuous	ESA Cloud CCI
L17	Mean cloud fraction over growing season	Zonal stat. mean within crop mask	Numeric continuous	ESA Cloud CCI
L18	Digital Terrain Model (DTM-Elevation)	Zonal min, max, mean, std	Numeric continuous	SRTM
L19	Digital Terrain Model (DTM-Elevation)	Zonal min, max, mean, std	Numeric continuous	SRTM
L20	GPP (Kg C/m3)	Zonal std	Numeric continuous	MODIS product

Layer name	Theme	Script/method	Variable type	Source
L21	Monthly mean temperature (Monthly/Celsius)	Zonal stat. mean	Numeric continuous	WorldClim
L22	Temperature (Year/Celsius)	Zonal stat. mean	Numeric continuous	WorldClim
L23	Monthly total rainfall (Monthly/mm)	Zonal stat. mean	Numeric continuous	WorldClim
L24	Rainfall (Year/mm)	Zonal Stat mean	Numeric continuous	WorldClim
L25	Soil productivity	Weighted overlay of soil quality indices; Zonal stat. mean	Numeric categorical	GAEZ (FAO)
L26	Road Density	Density from line features, cell 1 km, radius 25km <sup>2</sup> zonal stat. mean	Numeric continuous	ESRI/FAO
L27	Population density (num people/sq km)	Zonal stat. mean	Numeric continuous	LANDSCAN 2012
L28	Land cover	Calculate real surface per land cover class form GLC; Zonal stat. sum	Numeric continuous	GLC-SHARE (FAO)

The identified data sets were imported within a file geo-database called GAES.gdb and named according to field "Layer Name". All layers have a resolution of 1 km (at equator) and are in the WGS 84 coordinate system. Pre-processing was required to some degree, particularly for specific layers in order to make layers suitable for subsequent analysis.

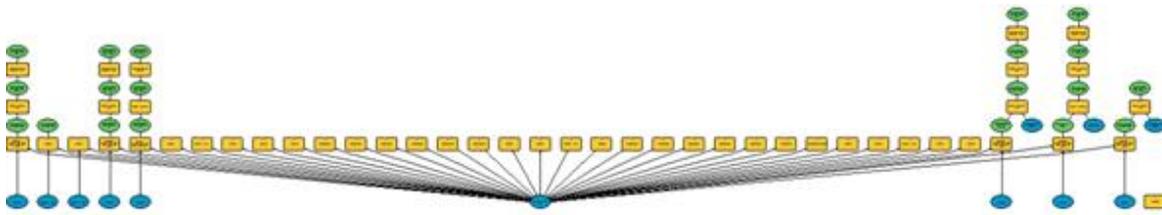
## 5.2 Analysis, models and tools

Statistics were calculated for each spatial layer listed in Table 11 within the zones defined by the strata shape files. Spatial models were built to run the analysis, according to the specific requirements for each layer. Models were constructed using a combination of tools from ArcGIS Spatial Analyst extension, compiling them within Model Builder for subsequent process automatization. In total, 30 models were created.



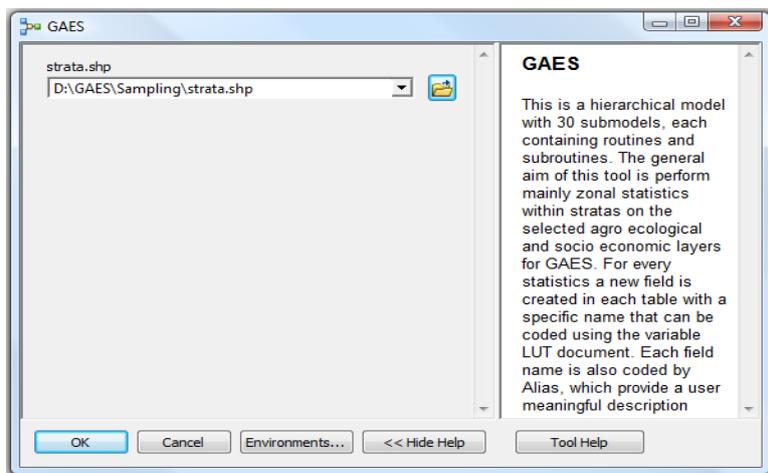
**Figure 36** Individual models created with Model Builder and saved as tools

Subsequently, individual models were grouped within a unique nested model called **GAES** (Figure 36). The model is able to perform all the individual models in sequence and save the full characterisation output database in batch mode.



**Figure 37** GAES nested model

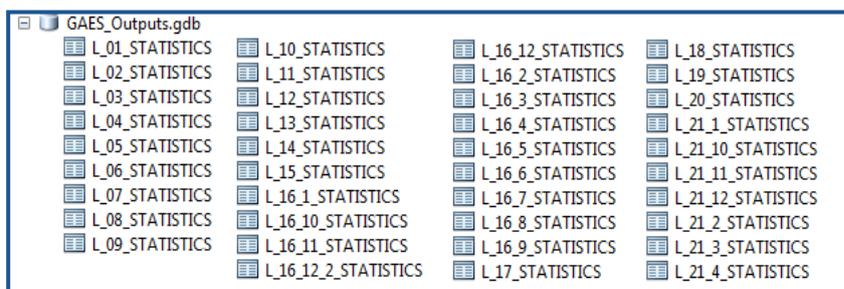
To facilitate the reuse and sharing of the model, it was saved as a tool within an ArcGIS toolbox inside the GAES.gdb. Such a tool can be run with a one-click step by the end-user and will perform the full analysis automatically and populate the output GAES database. The user has an option to select a stratification layer to be used to define zones, within which to perform the analysis, as shown in Figure 38.



**Figure 38** GAES tool interface

### 5.3 GAES statistical outputs

Results from GAES Model are written as Statistical tables within a geo-database called GAES\_Outputs.gdb. This is the main output of the GAES WP3.2. Each table is named with prefix from layer name and suffix "\_STATISTICS". For instance the results of analysis performed on Climatic Zones from the layer called L\_01 is "L\_01\_STATISTICS".



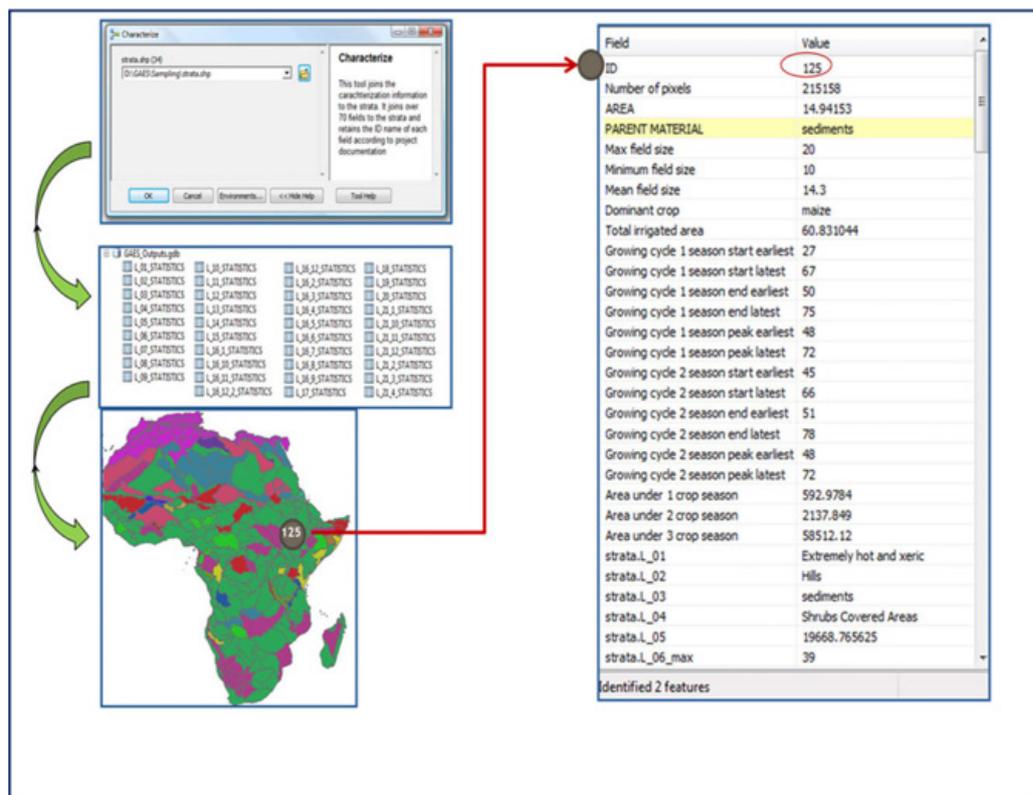
**Figure 39** GAES output geo-database populated with statistical tables

After each run of the GAES tool, the statistical tables are overwritten and replaced. Inside the table, each row represents a polygon from the strata layer and is coded using the polygon ID. The calculated statistics are written within the field, which has the same name of the layer. E.g. for L\_01\_STATISTICS, the field containing the statistics is called "L\_01". The statistics fields also have an alias, which is designed to provide a meaningful name.

### 5.3.1 Joining outputs to strata

Output STATISTICS tables can be used as they are or can be joined to the strata shape file for further GIS visualization and analysis. Once joined, the strata shape file can be colour-coded in GIS based on any field, and can be queried against all variables at once. Joined fields can be renamed to meaningful names using the alias created by default by the GAES tool. The process of manually joining over 30 tables and 70 fields, filtering out unnecessary fields, and formatting the table is a very time consuming task within the GAES workflow. This has been identified as a potential limitation for reuse of GAES tool/product. In order to overcome this, a tool called 'Characterize' has been developed to automate the process. The tool was created in the Model Builder and is saved as a tool in the toolbox within the GAES.gdb. The user selects the strata and the tool links this to all statistics from the output geo-database. The user can visualize the results in desktop GIS environment by opening the strata shape file.

The Characterize tool, adds an additional field to the strata shape file called "Descriptor". This file stores a descriptive text that summarizes the strata characteristics with a short descriptive narrative. The tool generates such text by concatenating the values from the first four layers, hence Climate, Elevation, Parent Material and Land Cover. More details on this are provided in Section 5.4.



**Figure 40** Joining output tables to strata shape file and Characterize tool for automatization

## 5.4 GAES final results

The GAES database (version 01a) is a newly developed Global Agro-Environmental Stratification that has been designed in the first instance to support the production of a global cropland mask within the EU project SIGMA based on a strategy of exploiting a range of Earth Observation (EO) data.

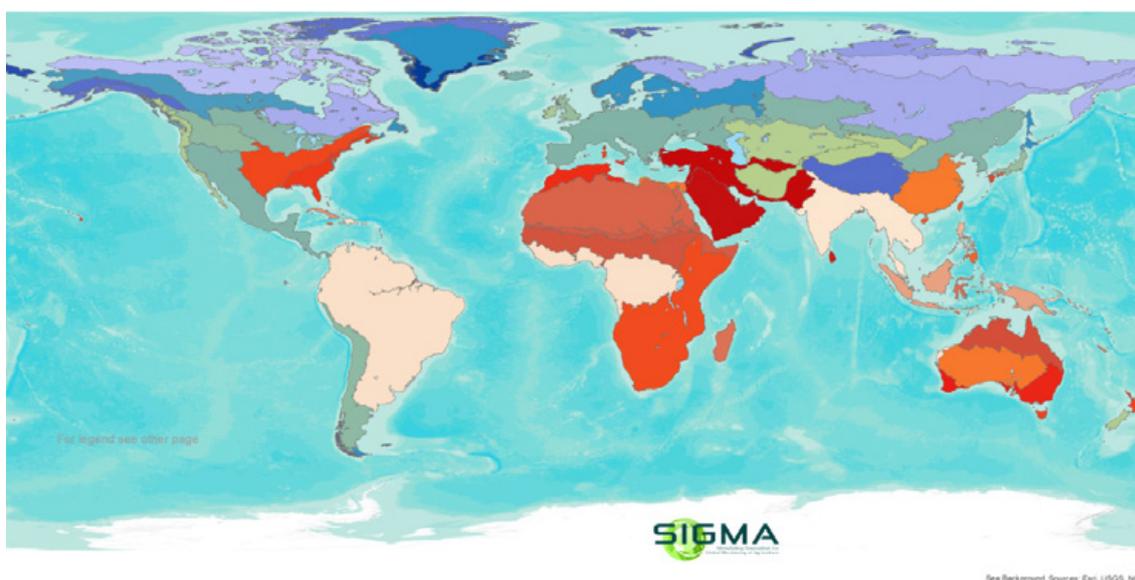
Nevertheless, it is anticipated that the GAES can be exploited for a wider range of applications, including others within SIGMA to make a data gap analysis to identify those strata, for which no field reference data are available on agricultural production, to assess uncertainties in existing global cropland classifications, and to assess the EO derived soil moisture products. The GAES database has four hierarchical layers with 92 attributes. A simplified legend (attribute descriptor) has been based four main characteristics:

- Climate (16 classes: arctic, cold and mesic, cold and wet, cool temperate and dry, cool temperate and moist, cool temperate and xeric, extremely cold and mesic, extremely cold and wet, extremely hot and arid, extremely hot and moist, extremely hot and xeric, hot and arid, hot and dry, hot and mesic, warm temperate and mesic, warm temperate and xeric).
- Altitude (five classes: very high mountains, high mountains, mountain, hills, lowlands).
- Parent material (five classes: rocks, sediments, organic material, glaciers, water bodies).
- Land cover (11 classes: artificial surfaces, cropland, grassland, tree covered areas, mangroves, herbaceous vegetation, shrubs covered areas, bare soil, sparse vegetation, snow and glaciers, water bodies).

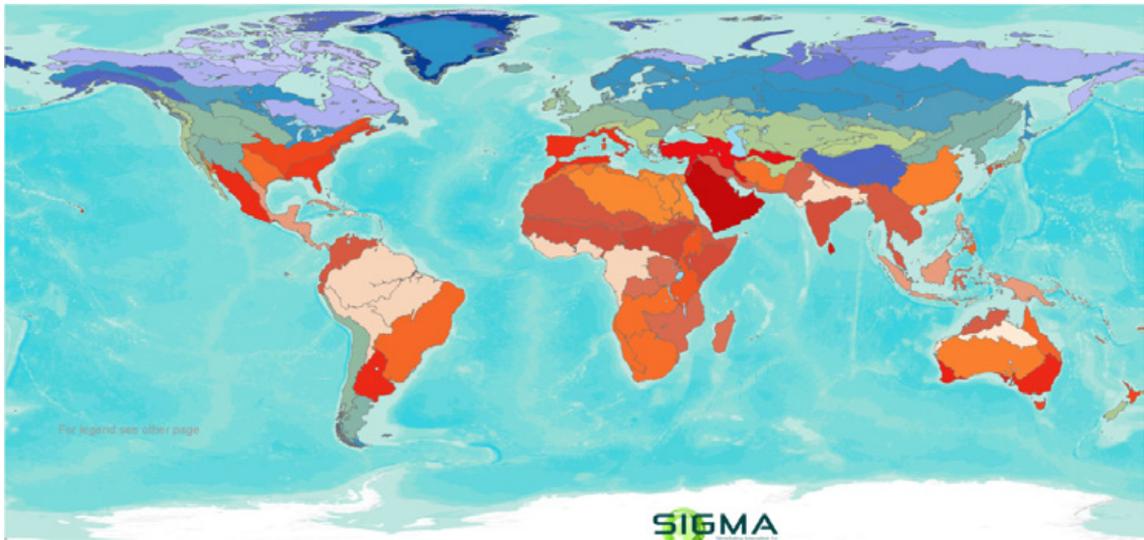
This could lead in principle to 4400 types ( $16 \times 5 \times 5 \times 11$ ). However, not all combinations exist and the number of combinations is also depending on the spatial scale of the stratification. GAES Level 1 has 194 agro-environmental (AE) types (818 strata); GAES Level 2 has 300 AE types (1688 strata); GAES Level 3 has 374 AE types (2087 strata); GAES Level 4 has 516 agro-environmental AE types (3208 strata). Figure 45 shows the naming of the 194 AE types present at the GAES Level 1. The name is a combination of temperature, altitude, parent material and land cover characteristics.

**Table 12** Summary statistics of the four hierarchical strata of GAES

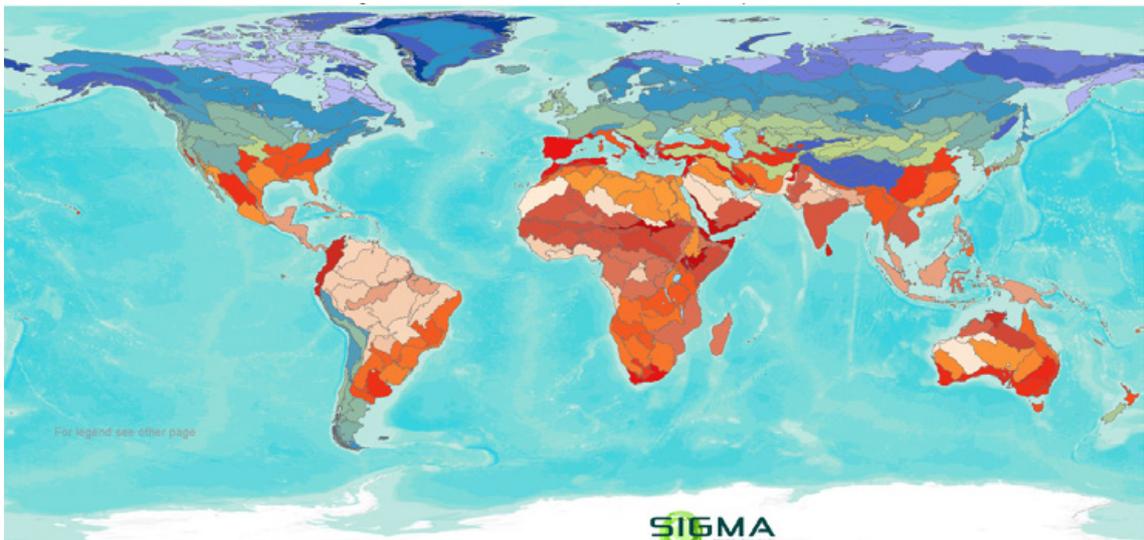
Statistics	Level 1	Level 2	Level 3	Level 4
Nr of strata	818	1688	2087	3208
AE types	194	300	374	516
Min area (km <sup>2</sup> )	26	11	9	9
Max area (km <sup>2</sup> )	15449045	4699599	1605143	740882
Mean area (km <sup>2</sup> )	163409	79249	64108	4170
SD (km <sup>2</sup> )	929563	374725	190128	80597



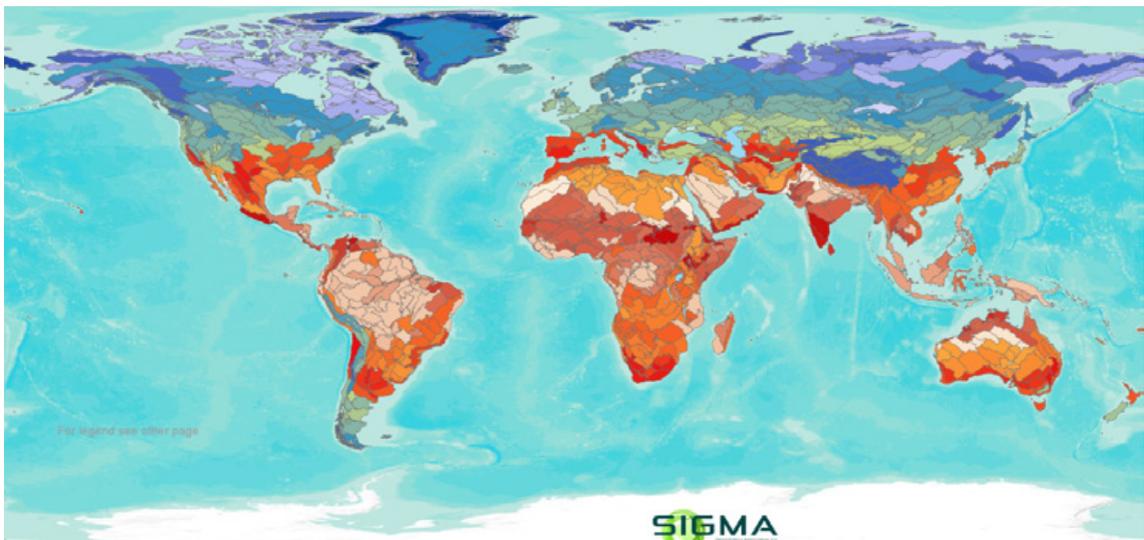
**Figure 41** Global Agro-Environmental Stratification (GAES) – Level 1



**Figure 42** Global Agro-Environmental Stratification (GAES) – Level 2



**Figure 43** Global Agro-Environmental Stratification (GAES) – Level 3



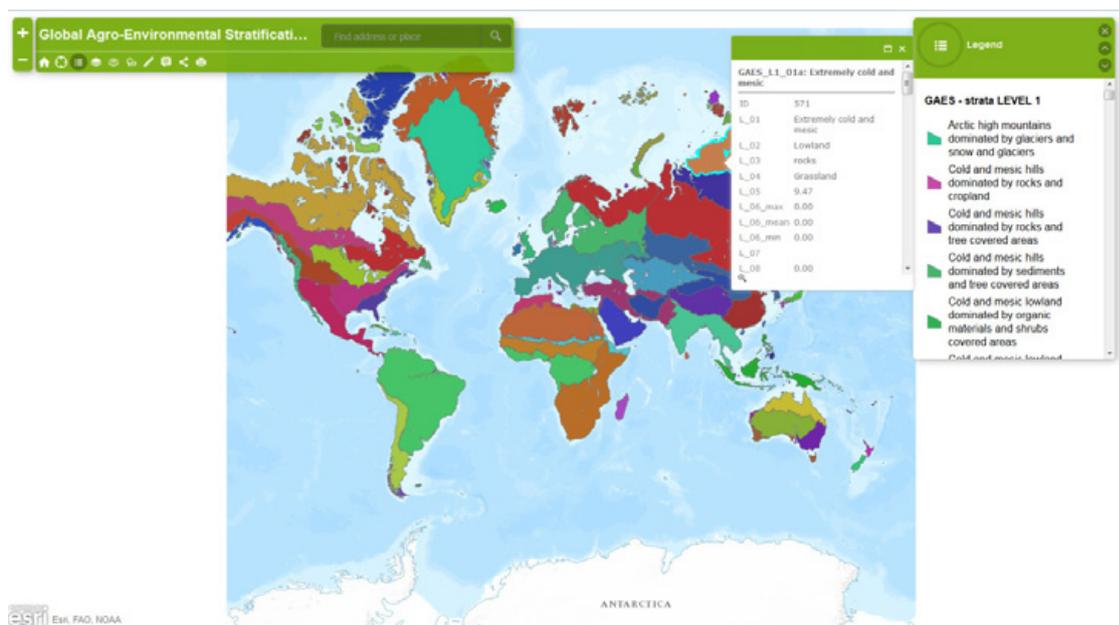
**Figure 44** Global Agro-Environmental Stratification (GAES) – Level 4



## 5.5 GAES dissemination and outreach

Dissemination of GAES outputs is achieved through a pervasive articulation of delivery channels:

1. GAES strata (four levels) will be available for download through FAO's Geo network and SIGMA data portal.
2. GAES strata are published as WebMapServices through FAO ArcGIS Server Site and available at <http://hqfao.maps.arcgis.com/apps/Viewer/index.html?appid=27f8cd872dc4488c82140636153b2adc>. They can be used by any GIS client.
3. GAES WebAPP: a WebApp will enable end user to browse, overlay and query GAES outputs inside a custom app that can be opened with any HTML viewer on desktop, tablet and mobile device.



**Figure 46** WebMapServices through FAO ArcGIS Server Site

Such constellation of delivery channels ensures both a pervasive dissemination and the perpetuation of outputs, being both Geonetwork and the ArcGIS Server site maintained by the FAO Team.

---

# References

- Balkovic, J., van der Velde M., Skalsky, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov A., Mueller, N.D., Obersteiner, M., 2014. Global wheat production potentials and management flexibility under the representative concentration pathways. *Global and Planetary Change*, 122:107-121.
- Balkovic, J., van der Velde, M., Schmid, E., Skalsky, R., Khabarov, N., Obersteiner, M., Sturmer, B., Xiong W., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. *Agricultural Systems*, 120:61-75
- Baruth., B., Royer, A., Klisch A., Genovese G., 2008. The use of remote sensing within the MARS crop yield monitoring system of the European Commission. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B8. Beijing 2008.
- Batjes, N.H., 2009. Harmonized soil profile data for applications at global and continental 12 scales: updates to the WISE database. *Soil Use and Management*, 25, 124–127.
- Carter S., Stuver J, 2014, Slope derived from SRTM data, v4, Wageningen University & CIFOR.
- Cohn, A.S., Mosnier, A., Havlik, P., Valin, H., Herrero, M., Schmid, E., O'Hare, M., Obersteiner, M., 2014. Cattle ranching intensification in Brazil can reduce global greenhouse gas emissions by sparing land from deforestation. *PNAS*, 111(20):7236-7241.
- Elliott, J., Deryng, D., Mueller, C., Folberth, C., Khabarov, N., Rosenzweig, C., Wisser, D., 2014. Constraints and potentials of future irrigation water availability on agricultural production under climate change. *PNAS*, 111(9):3239-3244.
- FAO, 1996. *Agro-ecological Zoning - Guidelines*. Rome, Italy.
- FAO & IIASA, 2007. *Mapping biophysical factors that influence agricultural production and rural vulnerability*. Environmental and Natural Resources Vol. 11. ISBN 978-92-5-105689-993 pp.
- Fischer, G., E. Hizsnyik, et al., 2009. *Biofuels and Food security*. OFID/IIASA. Vienna, Austria.
- Fischer, G., F.O., Nachtergaele, et al., 2012. *Global Agro-ecological Zones (GAEZ v3.0) - Model Documentation*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.
- Folberth, C., Gaiser, T., Abbaspour, K., Schullin, R., Yang, H., 2012. Regionalization of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields. *Agr. Ecosyst. Environ.* 151, 21–33.
- Frank, S., Bottcher, H., Havlik, P., Valin, H., Mosnier, A., Obersteiner, M., Schmid, E., Elbersen, B., 2013. How effective are the sustainability criteria accompanying the European Union 2020 biofuel targets? *GCB Bioenergy*, 5(3):306-314 (May 2013) (Published online 9 July 2012).
- Frere, M., Popov, G.F., 1979. *Agrometeorological crop monitoring and forecasting*. FAO Plant Production and Protection paper No. 17, by M. Frère and G.F. Popov. FAO, Rome, Italy.
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: The challenge of feeding 9 billion people. *Science* 327(5967): 812-818.

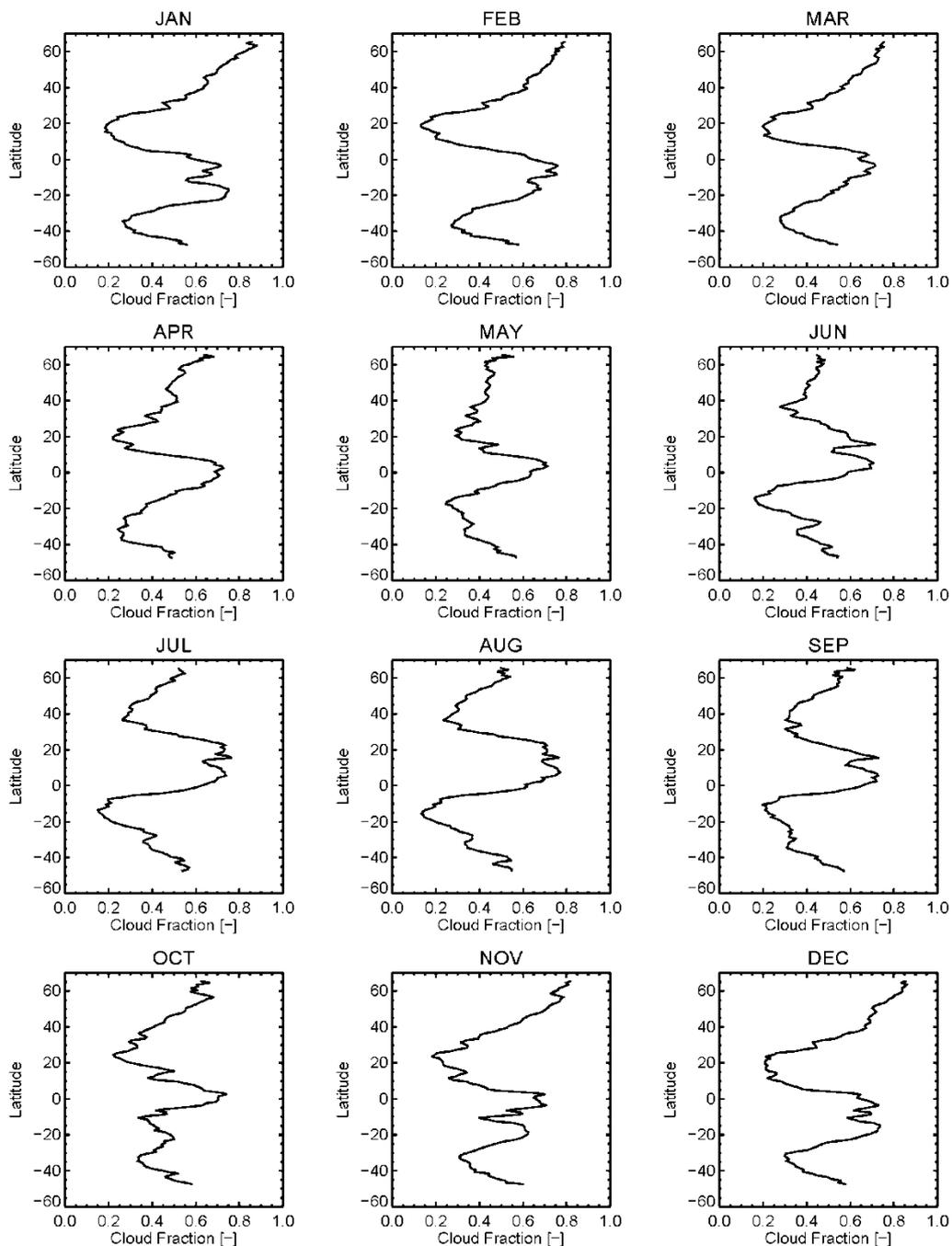
- 
- Gommes R., 1993. FAOINDEX, Version 2.1, Agrometeorology Group FAO, Rome.
- Hasegawa, T., Fujimori, S., Shin, Y., Takahashi, K., Masui, T., Tanaka, A., 2014. Climate Change Impact and Adaptation Assessment on Food Consumption Utilizing a New Scenario Framework, *Environmental Technology & Science*, 48, (1), 438-445.
- Havlik, P., Schneider, U.A., Schmid, E., Bottcher, H., Fritz, S., Skalsky, R., Aoki, K., De Cara, S., Kindermann, G., Kraxner, F., Leduc, S., McCallum, I., Mosnier, A., Sauer, T., Obersteiner, M., 2011. Global land-use implications of first and second generation biofuel targets. *Energy Policy*, 39(10):5690-5702.
- Havlik, P., Valin, H., Herrero, M., Obersteiner, M., Schmid, E., Rufino, M.C., Mosnier, A., Thornton P.K., Bottcher, H., Conant, R.T., Frank, S., Fritz, S., Fuss, S., Kraxner, F., Notenbaert, A., 2014. Climate change mitigation through livestock system transitions. *PNAS*, 111(10):3709-3714.
- Hazeu, G.W., Metzger, M.J., Múcher, C.A., Perez-Soba, M., Renetzeder, C., Andersen, E. 2011. European environmental stratifications and typologies: An overview. *Agric. Ecosyst. Environ.*, 142 (2011) 29–39.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.
- Hutchinson, M., Johnson, F., 2014. Application of ANUSPLIN to produce new intensity-frequency-duration (IFD) index rainfalls across Australia', 35th Hydrology and Water Resources Symposium, HWRS 2014, Conference Organising Committee, TBC, pp. 557-564.
- Klijn, J.A., 1995. Hierarchical concepts in landscape ecology and its underlying disciplines; the unbearable lightness of a theory. Report 100, The Winand Staring Centre. 144 pp.
- Leclere, D., Havlik, P., Fuss, S., Schmid, E., Mosnier, A., Walsh, B., Valin, H., Herrero, M., Khabarov, N., Obersteiner, M., 2014. Climate change induced transformations of agricultural systems: insights from a global model. *Environmental Research Letters*, 9(12):124018.
- Leip, A., Marchi, G., Koehler, R., Kempen, M., Britz, W., Li, C., 2008. Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe. *Biogeosciences* 5, 73–94.
- Liu, J., 2009. A GIS-based tool for modelling large-scale crop-water relations. *Environ. Modell. Softw.* 24, 411–422.
- Liu, J., Williams, J.R., Zehnder, A.J.B., Yang, H., 2007. GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale. *Agr. Syst.* 94, 478–493.
- Licker, R., M. Johnston, et al., 2010. Mind the gap: How do climate and agricultural management explain the 'yield gap' of croplands around the world? *Global Ecology and Biogeography* 19(6): 769-782.
- Massart, M., Rembold, F., Rojas, O., Leo, O. 2010. The Use of Remote Sensing Data and Meteorological Information for Food Security Monitoring, Examples in East Africa. In: Chuvieco, E., Li, J., Yang, X. (Eds.) (2010): *Advances in Earth Observation of Global Change*. Springer. P. 201-216.
- Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Múcher, C.A., Watkins, J.W., 2005. A climatic stratification of the environment of Europe. *Glob. Ecol. Biogeogr.* 14 (6), 549-563.

- 
- Metzger, M.J., R.G.H. Bunce, et al., 2011. Top-level tiers for Global Ecosystem Classification and Mapping Initiative (GEOSS Task ED-06-02). EBONE-WP3 Deliverable report D3.1.
- Metzger, M.J., A.D. Shkaruba, et al., 2012. Descriptions of the European Environmental Zones and Strata. Wageningen Environmental Research (Alterra) Report. A. Wageningen: 152.
- Metzger, M.J., D.J. Brus, et al., 2013a. Environmental stratifications as the basis for national, European and global ecological monitoring. *Ecological Indicators* 33: 26-35.
- Metzger, M.J., R.G.H. Bunce, et al., 2013b. A high-resolution bioclimate map of the world: A unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography* 22(5): 630-638.
- Monfreda, C., N. Ramankutty, et al., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000." *Global Biogeochemical Cycles* 22(1): GB1022.
- Mosnier, A., Havlik, P., Valin, H., Baker, J.S., Murray, B.C., Feng, S., Obersteiner, M., McCarl, B.A., Rose, S.K., Schneider, U.A., 2012. The Net Global Effects of Alternative U.S. Biofuel Mandates. Report NI R 12-01, Nicholas Institute for Environmental Policy Solutions, Duke University, Durham, NC, USA.
- Mücher, C.A., 1992. A discussion on land use classifications. An amalgam of methods. Internal report, Department of Tropical Crop Science, Wageningen Agricultural University, The Netherlands.
- Mücher, C.A., R.G.H. Bunce, R.H.G. Jongman, J.A. Klijn, A. Koomen, M.J. Metzger and D.M. Wascher, 2003. Identification and Characterisation of Environments and Landscapes in Europe. Wageningen Environmental Research (Alterra) rapport 832, Wageningen Environmental Research (Alterra), Wageningen.
- Mücher, C.A., Klijn, J.A., Wascher, D.M., Schaminée, J.H.J., 2010. A new European Landscape Classification (LANMAP): A transparent, flexible and user-oriented methodology to distinguish landscapes. *Ecol. Indicat.*, 10 (1), 87–103.
- Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A. (2012) Closing yield 15 gaps through nutrient and water management. *Nature*, 490, 254–257.
- Mueller, L., Schindler, U., et al., 2010. Assessing the productivity function of soils. A review. *Agronomy for Sustainable Development* 30(3): 601-614.
- Mueller, N.D., Gerber, J.S., et al., 2012. Closing yield gaps through nutrient and water management." *Nature* 490(7419): 254-257.
- Pini, G., Kayitakire, F., Leo, O., Rembold, F., Boogaard, H., Van Kraalingen, D., Van Der Wijngaart, R., Rojas, O., 2015. The Global Water Satisfaction Index (MARS GWSI). A MARS application for the global monitoring of rainfed agriculture. In preparation.
- Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000-Global monthly irrigated and rainfed 1 crop areas around the year 2000: A new high-resolution data set for agricultural and 2 hydrological modeling. *Global Biogeochemical Cycles*, 24, GB1011.
- Prentice, I.C., W. Cramer, et al., 1992. Special Paper: A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate. *Journal of Biogeography* 19(2): 117-134.
- Quiroz, R., Zorogastúa, P., et al., 2000. Toward A Dynamic Definition of Agroecological Zones Using Modern Information Technology Tools. *Natural Resource Management*. C. P. R. 2000. CIP, Lima, Peru.

- 
- Ramankutty, N., Foley, J.A., et al., 2002. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change." *Global Ecology and Biogeography* 11(5): 377-392.
- Rojas, O., Rembold, F., Royer, A. and Negre, T., 2005. Real-time agro meteorological crop yield monitoring in Eastern Africa. *Agron. Sustain. Dev.* 25 (2005) 63-77
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Mueller, C., Arneth, A., Boote, K.J., Folberth, C., Khabarov, N. et al., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *PNAS*, 111(9):3268-3273
- Schneider, U.A., Havlík, P., Schmid, E., Valin, H., Mosnier, A., Obersteiner, M., Böttcher, H., Skalský, R., Balkovič, J., Sauer, T., Fritz, S., 2011. Impacts of population growth, economic development, and technical change on global food production and consumption. *Agr. Syst.* 104, 204-215.
- Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-Year High-Resolution Global 13 Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate*, 14 19, 3088-3111.
- Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B.R. (2015). A global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences*, 19(3), 1521-1545.
- Skalsky, R., Tarasovicova, Z., Balkovic, J., Schmid, E., Fuchs, M., Moltchanova, E., Kindermann, G. and Scholtz, P., 2008. Geo-bene global database for bio-physical modeling v. 1.0. Concepts, methodologies and data. Technical Report, IIASA, Laxenburg, 57 pp. Available from: <  
<http://www.geo-bene.eu/?q=node/1734>>
- Stehfest, E., Heistermann, M., Priess, J.A., Ojima, D.S., Alcamo, J., 2007. Simulation of global crop production with the ecosystem model DayCent. *Ecol. Model.* 109, 203-219.
- Stomph, T.J., Múcher, C.A., Fresco, L.O., 1997. Environmental impact of land use: a new basis for analysis. *The Land: (1.2)*, 29-142.
- Tan, G., Shibasaki, R., 2003. Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. *Ecol. Model.* 168, 357-370.
- Tatsumi, K., Yamashiki, Y., et al., 2011. Estimation of potential changes in cereals production under climate change scenarios. *Hydrological Processes* 25(17): 2715-2725.
- Tóth, G., Kozłowski, B., et al., 2012. Global Agro-ecological Zones (GAEZ v3.0) – GAEZ Data portal. IIASA, Laxenburg, Austria and FAO, Rome, Italy.
- Van Ittersum, M.K., K.G. Cassman, et al., 2013. Yield gap analysis with local to global relevance-A review. *Field Crops Research* 143: 4-17.
- United Nations, 2015. World Population Prospects. Key findings & advance tables. 2015 Revision.
- van Oijen, M., Balkovic, J., Beer, C., Cameron, D.R., Ciais, P., Cramer, W., Kato, T., Kuhnert, M., Martin, R., Myneni, R., Rammig, A., Rolinski, S., Soussana, J.F., Thonicke, K., van der Velde, M., Xu, L., 2014. Impact of droughts on the carbon cycle in European vegetation: A probabilistic risk analysis using six vegetation models. *Biogeosciences*, 11(22):6357-6375.
- Van Wart, J., van Bussel, L.G.J., Wolf, J., Licker, R., Grassini, P., Nelson, A., Boogaard, H., Gerber, J., Mueller, N.D., Claessens, L., van Ittersum, M.K., Cassman, K.G., 2013. Use of agro-climatic zones to upscale simulated crop yield potential. *Field Crops Research* 143: 44-55.

- 
- White, D.H., Lubulwab, G.A., Menzc, K., Zuod, H., Winte, W., Slingenbergh, J., 2001. Agro-climatic classification systems for estimating the global distribution of livestock numbers and commodities. *Environment International* 27(2-3): 181-187.
- Werner, M., 2001. Shuttle Radar Topography Mission (SRTM) Mission Overview. *Frequenz*, 18 55, 75-79.
- Whitcraft, A.K., Vermote, E.F., Becker-Reshefa, I., Justicea, C.O., 2015. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sensing of Environment*, Volume 156, January 2015, 438-447.
- Williams, J.R., 1995. The EPIC model. In: Singh, V.P. (Ed.), *Computer models of watershed hydrology*, pp. 909-1000, Water resources publisher, Colorado, USA.
- Wood, S., Sebastian, K., et al., 2010. Spatial Perspectives. *Research Futures: Projecting Agricultural R&D Potentials for Latin America and the Caribbean*. P. G. Pardey, S. Wood and R. Hertford. Washington, DC, International Food Policy Research Institute.
- You, L., S. Wood, U. Wood-Sichra, 2006. Generating global crop maps: from census to grid. Selected paper, IAAE (International Association of Agricultural Economists) Annual Conference, Gold Coast, Australia.
- You, L., S. Crespo, Z. Guo, J. Koo, K. Sebastian, M.T. Tenorio, S. Wood, U. Wood-Sichra, 2009a. Spatial Production Allocation Model (SPAM) 2000 Version 3 Release 6.
- You, L.S., Wood, S., Wood-Sichra, U., 2009b. Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. *Agricultural Systems* 99 (2009) 126-140
- You, L.S., Wood, S., Wood-Sichra, U., Wu, W., 2014. Generating global crop distribution maps: From census to grid. *Agricultural Systems* 127 (2014) 53-60.

# Annex 1 Monthly cloud fraction across latitudes



Monthly cloud fraction across latitudes per 1x1 degree grid box, average over all 0.1x0.1 degree pixels that have an agricultural land use  $\geq 20\%$  of the grid area.



---

Wageningen Environmental Research  
P.O. Box 47  
6700 AA Wageningen  
The Netherlands  
T +31 (0)317 48 07 00  
[www.wur.nl/environmental-research](http://www.wur.nl/environmental-research)

Wageningen Environmental Research  
Report 2761  
ISSN 1566-7197

---

The mission of Wageningen University and Research is "To explore the potential of nature to improve the quality of life". Under the banner Wageningen University & Research, Wageningen University and the specialised research institutes of the Wageningen Research Foundation have joined forces in contributing to finding solutions to important questions in the domain of healthy food and living environment. With its roughly 30 branches, 5,000 employees and 10,000 students, Wageningen University & Research is one of the leading organisations in its domain. The unique Wageningen approach lies in its integrated approach to issues and the collaboration between different disciplines.





To explore  
the potential  
of nature to  
improve the  
quality of life



---

Wageningen Environmental Research  
P.O. Box 47  
6700 AB Wageningen  
The Netherlands  
T +31 (0) 317 48 07 00  
[www.wur.eu/environmental-research](http://www.wur.eu/environmental-research)

Report 2761  
ISSN 1566-7197

The mission of Wageningen University and Research is "To explore the potential of nature to improve the quality of life". Under the banner Wageningen University & Research, Wageningen University and the specialised research institutes of the Wageningen Research Foundation have joined forces in contributing to finding solutions to important questions in the domain of healthy food and living environment. With its roughly 30 branches, 5,000 employees and 10,000 students, Wageningen University & Research is one of the leading organisations in its domain. The unique Wageningen approach lies in its integrated approach to issues and the collaboration between different disciplines.

