Food and Agriculture Organization
of the United Nations

USING REMOTE SENSING IN SUPPORT OF SOLUTIONS
TO REDUCE AGRICULTURAL WATER PRODUCTIVITY GAPS

DATABASE METHODOLOGY: LEVEL 2 DATA

WaPOR Version 1 | October 2018

WaPOR Database methodology: Level 2 data

V1 release

October 2018

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2018

Required citation: FAO 2018. WaPOR Database Methodology: Level 2. Remote Sensing for Water Productivity Technical Report: Methodology Series. Rome, FAO. 63 pages. Licence: CC BY-NC-SA 3.0 IGO.

ISBN 978-92-5-130057-2

Contents

Figures

Tables

Boxes

Preface

Achieving Food Security in the future while using water resources in a sustainable manner will be a major challenge for current and future generations. Increasing population, economic growth and climate change all add to increasing pressure on available resources. Agriculture is a key water user and careful monitoring of water productivity in agriculture and exploring opportunities to increase it is required. Improving water productivity often represents the most important avenue to cope with increased water demand in agriculture. Systematic monitoring of water productivity through the use of Remote Sensing techniques can help to identify water productivity gaps and evaluate appropriate solutions to close these gaps.

The FAO portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR) provides access to 10 years of continued observations over Africa and the Near East. The portal provides open access to various spatial data layers related to land and water use for agricultural production and allows for direct data queries, time series analyses, area statistics and data download of key variables to estimate water and land productivity gaps in irrigated and rain fed agriculture.

The *beta* release of WaPOR was launched on 20 April 2017 covering the whole of Africa and the Near East region with a spatial resolution of 250 m and, eventually complemented with data at 100 m resolution for selected countries and river basins. WaPOR Version 1 became available starting from June 2018. This document describes the methodology used to produce the data at 100 m resolution (Level 2) distributed through WaPOR Version 1.

Acknowledgements

FAO, in partnership with and with funding from the Government of the Netherlands, is developing a programme to monitor and improve the use of water in agricultural production. This document is part of the first output of the programme: the development of an operational methodology to develop an open-access database to monitor land and water productivity.

The methodology was developed by the FRAME[1] consortium, consisting of the eLEAF, VITO, ITC, University of Twente and Waterwatch foundation, commissioned by and in partnership with the Land and Water Division of FAO.

Substantial contributions to the eventual methodology were provided during the first Methodology Review workshop, held in FAO Headquarters in October 2016 and during the second *beta* methodology review workshop, in January 2018. Participants in these workshops were: Henk Pelgrum, Karin Viergever, Maurits Voogt and Steven Wonink (eLEAF), Sergio Bogazzi, Amy Davidson, Jippe Hoogeveen, Michela Marinelli, Karl Morteo, Livia Peiser, Pasquale Steduto, Erik Van Ingen (FAO), Megan Blatchford, Chris Mannaerts, Sammy Muchiri Njuki, Hamideh Nouri, Zeng Yijan (ITC), Lisa-Maria Rebelo (IWMI), Job Kleijn (Ministry of Foreign Affairs, the Netherlands), Wim Bastiaanssen, Gonzalo Espinoza, Jonna Van Opstal (UNESCO-IHE), Herman Eerens, Sven Gilliams, Laurent Tits (VITO) and Koen Verberne (Waterwatch foundation).

[1]For more information regarding FRAME, contact eLEAF (http://www.eleaf.com/). Contact persons. FRAME project manager: Steven Wonink (steven.wonink@eleaf.com). Managing Director: Maurits Voogt (maurits.voogt@eleaf.com)

Abbreviations and acronyms

AGBP	Above Ground Biomass Production
DEM	Digital Elevation Model
DMP	Dry Matter Productivity
E	Evaporation
EOS	End of Season
ESU	Elementary Surface Area
ET	Evapotranspiration
ETIa	Actual EvapoTranspiration and Interception
FAO	Food and Agriculture Organization of the United Nations
FRAME	Consortium consisting of eLEAF, VITO, ITC and the Waterwatch Foundation
GBWP	Gross Biomass Water Productivity
I	Interception
LAI	Leaf Area Index
LST	Land Surface Temperature
LUE	Light Use Efficiency
MOS	Maximum of Season
MS	Multi-Spectral
NBWP	Net Biomass Water Productivity
NDVI	Normalised Difference Vegetation Index
NIR	Near Infrared
NPP	Net Primary Production
NRT	Near Real Time
PHE	Phenology
RET	Reference Evapotranspiration
ROI	Region of Interest
SMC	Soil Moisture Content
SOS	Start of Season
T	Transpiration
TIR	Thermal Infrared
TOC	Top of Canopy
VI	Vegetation Index
VNIR	Visible and Near Infrared
WaPOR	FAO portal to monitor Water Productivity through Open access of Remotely sensed derived data

1 Introduction

This report outlines the methodology applied to produce the different data components of WaPOR, the FAO portal to monitor Water Productivity through Open access of remotely sensed derived data. This data is mainly derived from freely available remote sensing satellite data. The aim of this document is to provide the theory that underlies the methods used to produce the different data components. References are included throughout the document so that additional information on specific aspects of the methodology can be found. Detailed information on the processing chain, data sources and processing steps are provided in the Data Manual.

The beta release of WaPOR, was launched on April 20, 2017. Based on the methodology review process, a new version WaPOR 1.0 became available in June 2018, focusing first on the coarser resolution level (Level 1), covering the whole of Africa and the Near East at 250 m ground resolution and then the national / river basin level (Level 2) at 100 m resolution. This document describes the methodology applied to produce the database at Level 2 (100 m), as made available through WaPOR Version 1.0 release, starting in June 2018.

1.1. Characteristics of the datasets

Each dataset (also called 'level') is defined by a unique region of interest and a specific spatial resolution. Table 1 specifies the resolution and area covered by the different levels while Figure 1 shows the extent of Level 2 data on a map.

Table 1: Spatial resolution and Regions of Interest of the different datasets (levels).

Dataset	Resolution	Region of Interest
Level 1	~250m (0.00223°)	Africa and Near East (bounding box 30W, 40N, 65E, 40S)
Level 2	~100m (0.000992°)	*Counties*[1]*:* Morocco, Tunisia, Egypt, Ghana, Kenya, South Sudan, Mali, Benin, Ethiopia, Rwanda, Burundi, Mozambique, Uganda, West Bank and Gaza Strip, Yemen, Jordan, Syrian Arab Republic and Lebanon. *River basins*[2]*:* Niger, Nile, Awash and Jordan and Litani.
Level 3	~30m (0.000268°)	Irrigation schemes and rainfed areas in Egypt, Ethiopia (2 areas), Mali and Lebanon.

[1] The boundaries of the countries are derived from the latest version (2014/2015) of the Global Administrative Unit Layers (GAUL), http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691 and they include also the Non-Self Governing Territory Western Sahara and the Sovereignty unsettled territories: Hala'ib Triangle and Ma'tan al-Sarra.
[2] The boundaries of the river basins are derived from HydroSHEDS (http://www.fao.org/geonetwork/srv/en/metadata.show?id=37038).

The pixel resolutions (in m) shown in Table 1 are approximate values. The data is delivered in a geographic coordinate system that measures coordinates in latitude and longitude. The pixel size, when expressed in meters, will therefore vary with latitude[2]. The resolution remains the same when expressed in degrees, regardless of latitude.

[2] When resolution is expressed in meters, higher latitudes (further from the equator) have a higher resolution in an east-west direction. It should therefore be noted that, as a result, the raster values should be converted into areal quantities by first calculating the exact size of

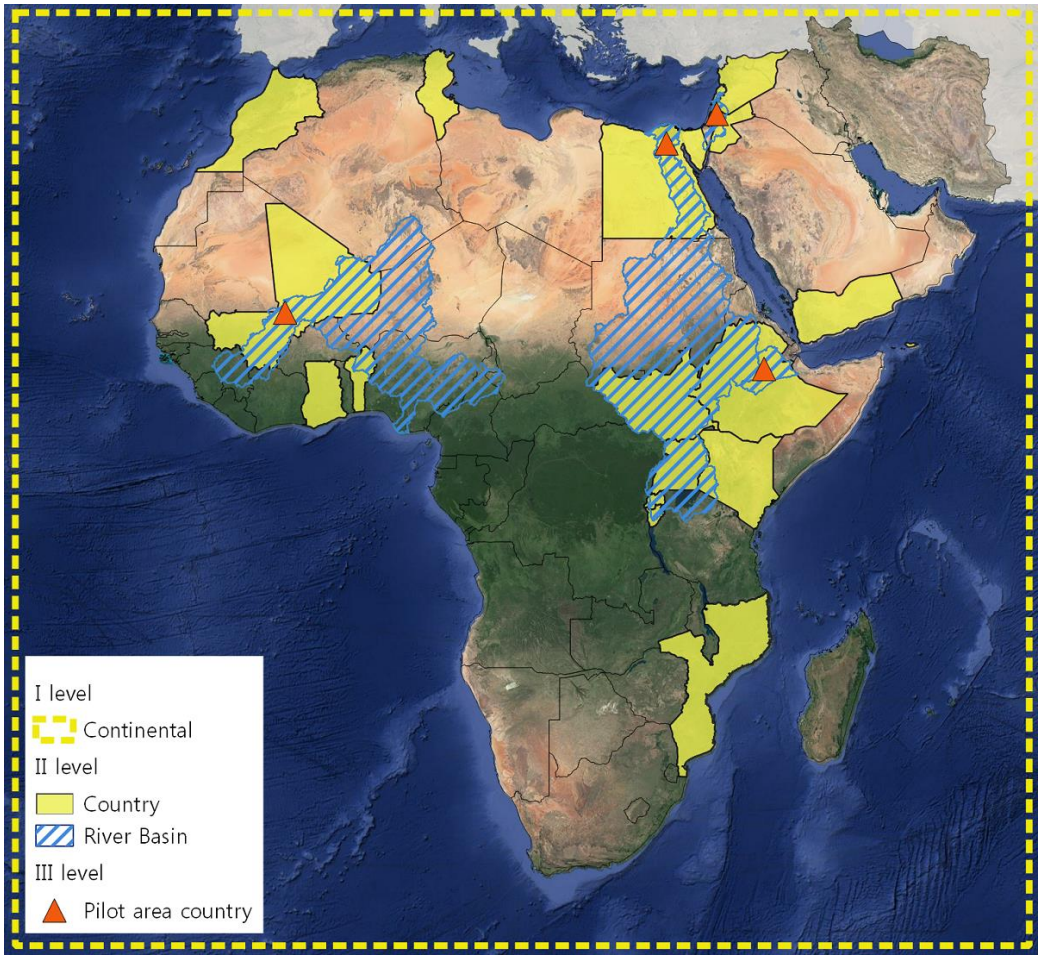


Figure 1: WaPOR data coverage at the national and river basin level (Level 2) refers to the areas shown in yellow (countries) and blue diagonal lines (river basins).

Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor.

The data components that are produced for the WaPOR database are listed in Table 2. Water Productivity, Evaporation, Transpiration, Interception, Net Primary Productivity, Above Ground Biomass Production and Land Cover Classifications are produced at all three levels. Phenology is delivered for Levels 2 and 3 and HI for Level 3 only (thereby allowing for the calculation of crop yield). Reference Evapotranspiration and Precipitation are only produced at Level 1 and it should be noted that these two data components have a much lower spatial resolution than the other Level 1 data components and that they are both produced daily. Details of the methodology can be found in Chapter 2 of the Level 1 Data Methodology document.

a specific pixel (in meters) before calculating the area it covers. The table below shows how the pixel size (expressed in m) varies with increasing latitude.

Dataset	Degrees	Equator Lat/Lon (m)	Lat/Lon (m) at 20^{o} N/S	Lat/Lon (m) at 40^{o} N/S
Level 1	0.00223	246.6/248.2	246.9/233.4	247.6/190.4
Level 2	0.000992	109.7/110.4	109.8/103.8	110.1/84.7
Level 3	0.000268	29.6/29.8	29.7/28.0	29.8/22.9

Additional complementary data layers are listed in Table 3. These include layers that can be applied by the user to add value to the WaPOR data components, or to inform the user about the quality of input data. Details with regard to Level 2 layers are given in Chapter 2.

Table 2: Overview of the WaPOR data components, per Level, with temporal and spatial resolutions specified.

Data components	Level[1] 1 (~250m)	Level 2 (~100m)	Level 3 (~30m)	Remarks
Water Productivity (WP)	Annual[2]	Dekadal[3]/ Seasonal[4]	Dekadal/ Seasonal	*Level specific calculations*
Evaporation (E)	Dekadal/Annual	Dekadal/Annual	Dekadal/Annual	
Transpiration (T)	Dekadal/Annual	Dekadal/Seasonal/Annual	Dekadal/Seasonal/Annual	
Interception (I)	Dekadal/Annual	Dekadal/Annual	Dekadal/Annual	
Actual Evapotranspiration and Interception (ETIa)	Dekadal/Annual	Dekadal/Seasonal/Annual	Dekadal/Seasonal/Annual	
Net Primary Production (NPP)	Dekadal	Dekadal	Dekadal	
Above ground biomass production (AGBP)	Annual	Seasonal	Seasonal	Yield at Level 3 for selected crops
Phenology		Seasonal	Seasonal	
Harvest Index			Seasonal	
Reference Evapotranspiration (RET)	Daily			*Different resolution: 20km*
Precipitation	Daily			*Different resolution: 5km*
Land cover classification	Annual	Annual	Dekadal	*Level specific classes*

[1] Level 1: Continental, Level 2: Country/River basin, Level 3: Irrigation scheme/sub-basin.
[2] Annual as standard product, with possibility of calculating on user-defined intervals.
[3] Dekadal refers to a period of approximately 10 days. It splits the month in 3 parts, where the first and second dekads consist of 10 days each and the duration of the last dekad ranges between 8 and 11 days.
[4] Seasonal refers to the growing season. The length and number may vary, with a maximum of 2 growing seasons per year.

Table 3: Overview of additional data layers, specifying the levels, temporal and spatial resolutions and what these additional data layers can be used for.

Complementary data layers	Level 1 (~250m)	Level 2 (~100m)	Level 3 (~30m)	Use
LUE correction factor	Annual[1]	Annual	Seasonal	*Adjust NPP and AGBP using updated LUE at the end of the season.*

Above-ground Over Total (AOT) biomass production ratio correction factor			Seasonal	*Adjust NPP and AGBP using updated AOT ratios at the end of the season.*
NDVI quality layer	Dekadal	Dekadal	Dekadal	*Indicates quality of external data used to produce NDVI.*
Land Surface Temperature quality layer	Dekadal	Dekadal		*Indicates quality of external data used to produce Soil moisture stress – L1 and L2 data at the same resolution.*
Land Cover Classification Quality	Annual	Annual	Seasonal	Land Cover Classification Quality

[1]Only produced for 2017 and 2018 as the land cover produced for 2016 was used preliminarily. Changes in land cover in years following 2016 will be taken into account with the LUE correction factor.

1.2. Structure of the database methodology document(s)

This document describes the characteristics and the methodology applied to produce the data published on WaPOR as of June 2018 (version 1.1). It refers to the country / river basin level (Level 2) datasets, as shown in Figure 1 and detailed in Box 1 and Table 1.

Although similar across all levels, the methodology is split in level-specific documents for easier reference. The assumption is that users will more likely access data at the specific level that best suit their needs, rather than switching between different levels. The level-specific documentation will thus provide a practical instrument to understand the data of interest, without the need to go looking through the documentation of the whole database.

Chapter 1 contains information on the characteristics of the datasets. As illustrated in Box 1 the data structure is made up of three different datasets (also called 'levels'), each comprising a number of data components. The 'level' of the dataset determines the characteristics (such as spatial resolution and region of interest) of the data components.

Chapter 2 sets out the methodology for the production of the different data components. The underlying body of scientific knowledge is summarised, citing references where the reader can find more detailed information if needed. The methodology description is split in two parts: Part 1 describes the methodology applied for the data components that are made accessible through WaPOR. Part 2 of Chapter 2 describes the methodology applied for the production of intermediate data components that are not distributed through WaPOR[3]. Intermediate data components convert external data sources into common inputs for the production of the data components, for example the NDVI, which is used as input to produce the Evaporation, Transpiration, Interception, fAPAR, Land Cover Classification and Phenology data components. Details of the specific data sources of satellite, static and meteorological data are addressed in the Data Manual, while a summary table with sensors used in production of Level 2 is provided in Annex 1.

It should be noted that the (intermediate) data components are produced in two distinct processing phases, i.e. *historical data processing* which produces data from 2009 up to a point in time in 2018, followed by a phase of continuous *near real time (NRT) processing*, starting in 2018 where the historical processing left off, continuing up to 2019. In some cases, the different processing phases necessitate differences in processing approaches. These are also addressed in the Data Manual.

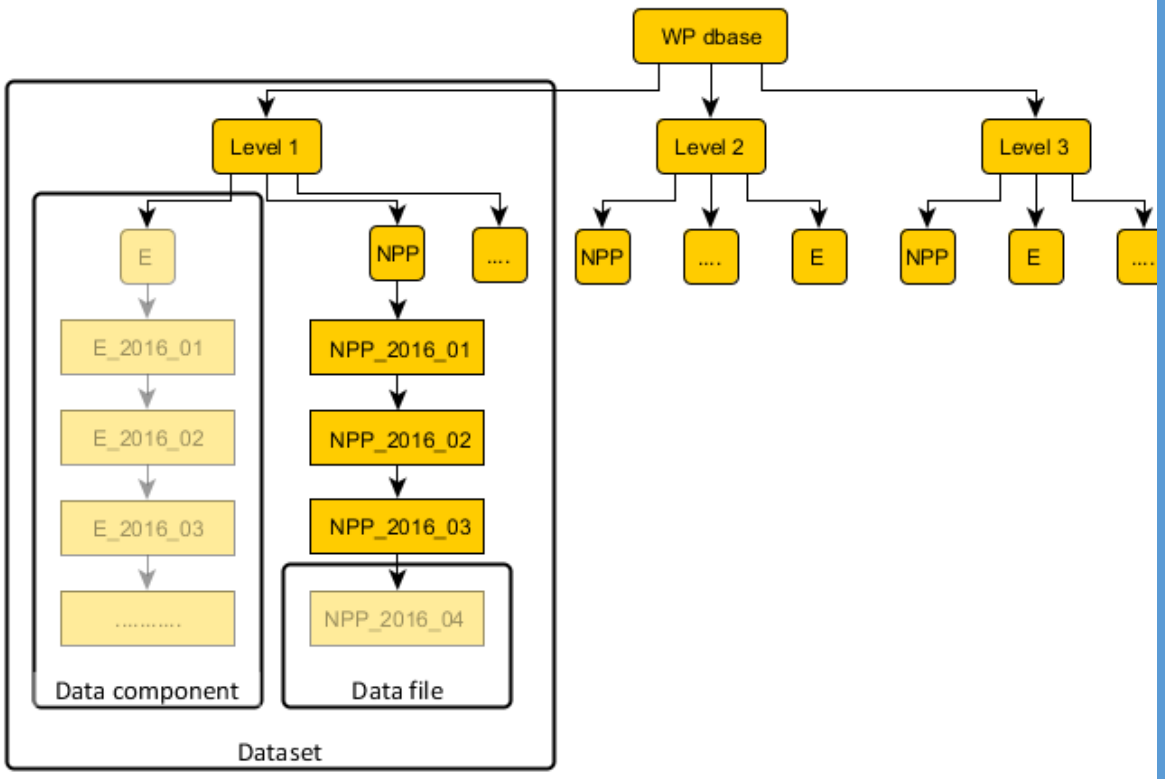
[3] A few data components that are also intermediate data components are distributed through WaPOR, these will be noted in the text.

Box 1: Data Structure

The term ***data*** is frequently used throughout this document. The following definitions explain the different uses of the term within WaPOR :

The following definitions are used in relation to the Water Productivity database:

- ***Data (file)***: raster data in GeoTIFF format, containing coordinate reference system (CRS) information in line with the OGC and ISO TC211 specifications.
- ***Data component***: A time series of similarly structured data files containing one specific type of information (e.g. Evaporation). Each individual data file contains information on the data component for a different time period.
- ***Dataset***: A set of related data components which cover the same Region of Interest (ROI) and time period (though not necessarily with the same temporal and spatial resolution). For example, the continental dataset (Level 1) contains, amongst others, Evaporation, Transpiration, Interception and Net Primary Productivity data components.



The term ***data*** is also used in relation to external sources, e.g. data used as input to produce or to validate the different data components. The following data sources can be distinguished:

- Regularly updated data includes satellite imagery and meteorological data, used for the production of all data components.
- Static data, such as elevation and soil type, do not change within the time period of the datasets.
- Reference data refers to ground or field observations or measurements which are used in most cases to validate the data components. Reference data is also used for the production of the land cover data component.

1.3. Related documents

This document focuses on the core theory that underlies the methodology applied for the production of the data components. Related, more detailed, information can be found in the following accompanying documents:

- Other Level-specific methodology documents related to Level 1 and Level 3.
- The Data Manual contains a detailed discussion of the processing chain of each dataset, i.e. at Level 1, 2 and 3. The Data Manual include details on external data sources used, as satellite sensors, meteorological data and static data sources at various resolutions. Differences in the processing chain due to different in input data sources, resolutions and processing phase (historic or NRT) are explained.
- Reports on Validation results are delivered at different stages. Quality assessment is an important part of WaPOR, therefore independent internal quality control procedures have been set up to validate the data components. The methodology for validation and quality control is detailed in these validation reports.

2 Methodology for the production of the data components

As shown in Table 2 and 3, WaPOR database consists of several data components related to water productivity, biomass production, evapotranspiration and land cover, as well as several complementary data layers, containing additional information. Part 1 of this chapter sets out the method by which these data components and complementary data layers are produced.

Part 2 of this Chapter describes the methodology of eleven intermediate data components. The intermediate data components are used to standardise the processing chain, converting external data sources into the standardised input data required for the data components. The processing structure based on the production of intermediate data components, was designed because it has the following advantages:

1. Flexibility and adaptability are ensured. NDVI and weather data, for example, can be obtained from many different sources. External data sources can be changed easily by defining standardised inputs in the form of the intermediate data components.
2. Different approaches to the pre-processing of external data sources can easily be incorporated without changing the overall processing structure of the data components.
3. Consistency between data components is higher with the use of common standardised inputs. This is important as many data components are closely related to each other, e.g. biomass production and Evaporation, Transpiration, Interception.
4. All input data is converted to the required resolution prior to the processing of the data components.
5. Improved processing efficiency is ensured, as the intermediate data components are produced only once and are used as input in various data components.
6. Quality checks can be done on the intermediate data components. In fact, two data layers are delivered that contain information on the quality of the remote sensing observations used to produce the intermediate data components: NDVI and Land Surface Temperature.

The following two remarks about resolution should be noted:

1. The method to produce the data components is independent of spatial resolution. Each pixel is considered a closed system in relation to adjacent pixels. Although in reality exchange of energy and matter takes place between adjacent pixels, these exchanges are considered negligible when considering the spatial and temporal resolution of the datasets. Therefore, all variables referred to in the methodology description can be interpreted as a point representing the average for the area covered by the pixel, whether at 250m, 100m or 30m resolution.
2. The temporal resolution of the data components can vary, i.e. daily, dekadal, seasonal and annual. When data components with a different temporal resolution are combined, the component with the highest temporal resolution will determine the output temporal resolution. For example, when dekadal NDVI is combined with daily weather data, processing takes place on a daily basis followed by an aggregation to dekadal values again. This ensures that information is retained at the highest level of detail for as long as possible during processing.

In general, the same methodology is applied across different levels to produce a data component. For example, Evaporation, Transpiration, Interception and Net Primary Production are produced at all three levels (see Table 2 for an overview of data components in the different levels) applying the same methodology. Some specific exceptions exist:

- Land cover classifications are specific for each level due to differences in the input data sources used and the level of land cover detail required.
- Water Productivity reflects the level of detail of the numerator of the equation. At Level 1, the numerator is annual above ground biomass production (AGBP), as no information on crop nor seasonality is available for the 250 m resolution data components. At Level 2 the numerator is seasonal above ground biomass production, as phenology information is available that makes possible calculating seasonal aggregates. At Level 3, WP is calculated when applicable using yield as numerator, as crop-specific information is available for higher resolution data.
- Phenology is only produced at Level 2 and 3, for which information on seasonality is available.
- Harvest Index is only produced at Level 3 as crop-specific land cover information is available at this level.

Figure 2 shows the relationship between the data components. This flow chart can be used as a reading guide. Each component is discussed in a separate section of this chapter. By following the arrows in the opposite direction all relevant information for the production of a specific data component can be obtained. For example, understanding the full processing chain of the AGBP data component also requires studying the NDVI intermediate data component. For Evaporation and Transpiration, seven other data components, of which five are intermediate data components, should be studied to understand all aspects of the production process. External data sources are not listed in this flow chart, nor are they discussed in this document. Details on the external data sources used can be found in the Data Manual.

The sections for each of the data components follow the same structure. A description of the data component includes information on the typical value range and a figure showing an example of the data component. The theory that underlies the methodology of the data component is then described. This starts with a box denoting the relationship between the data component under study and the other components. At the end of every methodology description, a table summarises the characteristics of the specific data components. Where relevant, a short discussion on challenges and limitations related to the data component is included.

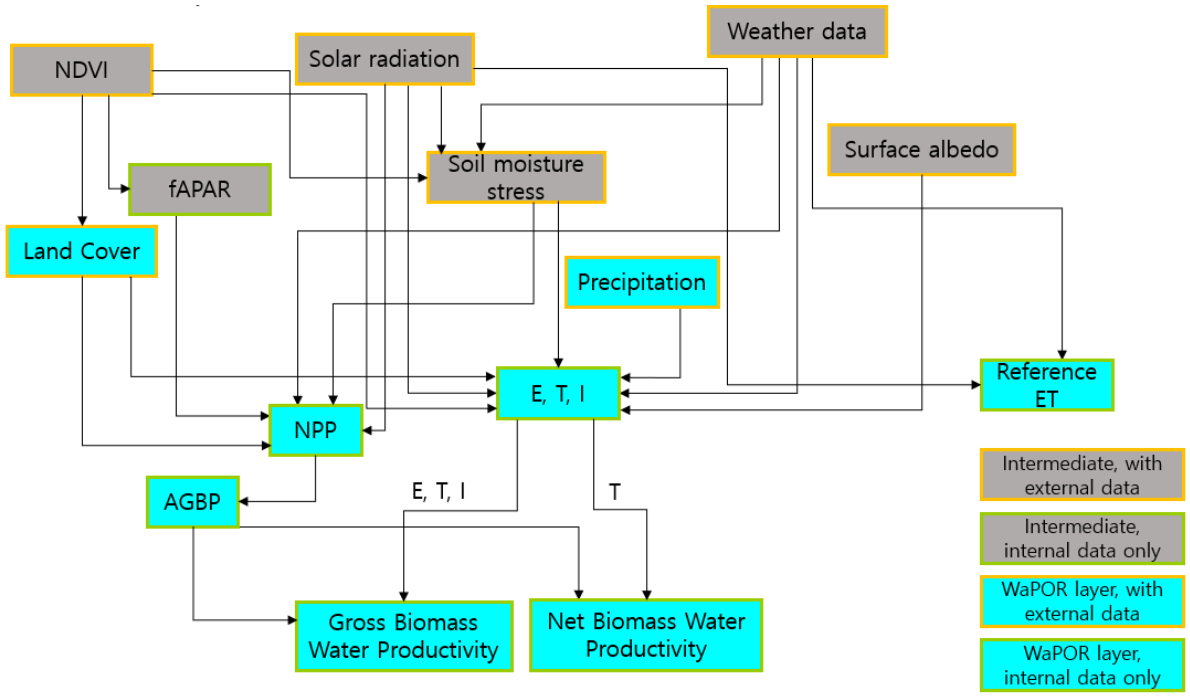


Figure 2: Data component flow chart. The grey boxes represent intermediate data components that convert external data into standardised input. Green outlines represent data components that are derived solely from other data components. Boxes with orange outlines represent data components that require external data sources that are not shown in the flow chart. Blue boxes represent data variables that are distributed through WaPOR.

2.1. WaPOR data components

This section describes the methodology applied to derive the data components as published through WaPOR v 1.0 at https://wapor.apps.fao.org

2.1.1. Water Productivity

2.1.1.1. Gross Biomass Water Productivity

Description

The gross biomass water productivity expresses the quantity of output (above ground biomass production) in relation to the total volume of water consumed in a given period (FAO, 2016). By relating biomass production to total evapotranspiration (sum of soil evaporation, canopy transpiration, and interception, section 2.1.3), this indicator provides insights on the impact of vegetation development on consumptive water use and thus on water balance in a given domain.

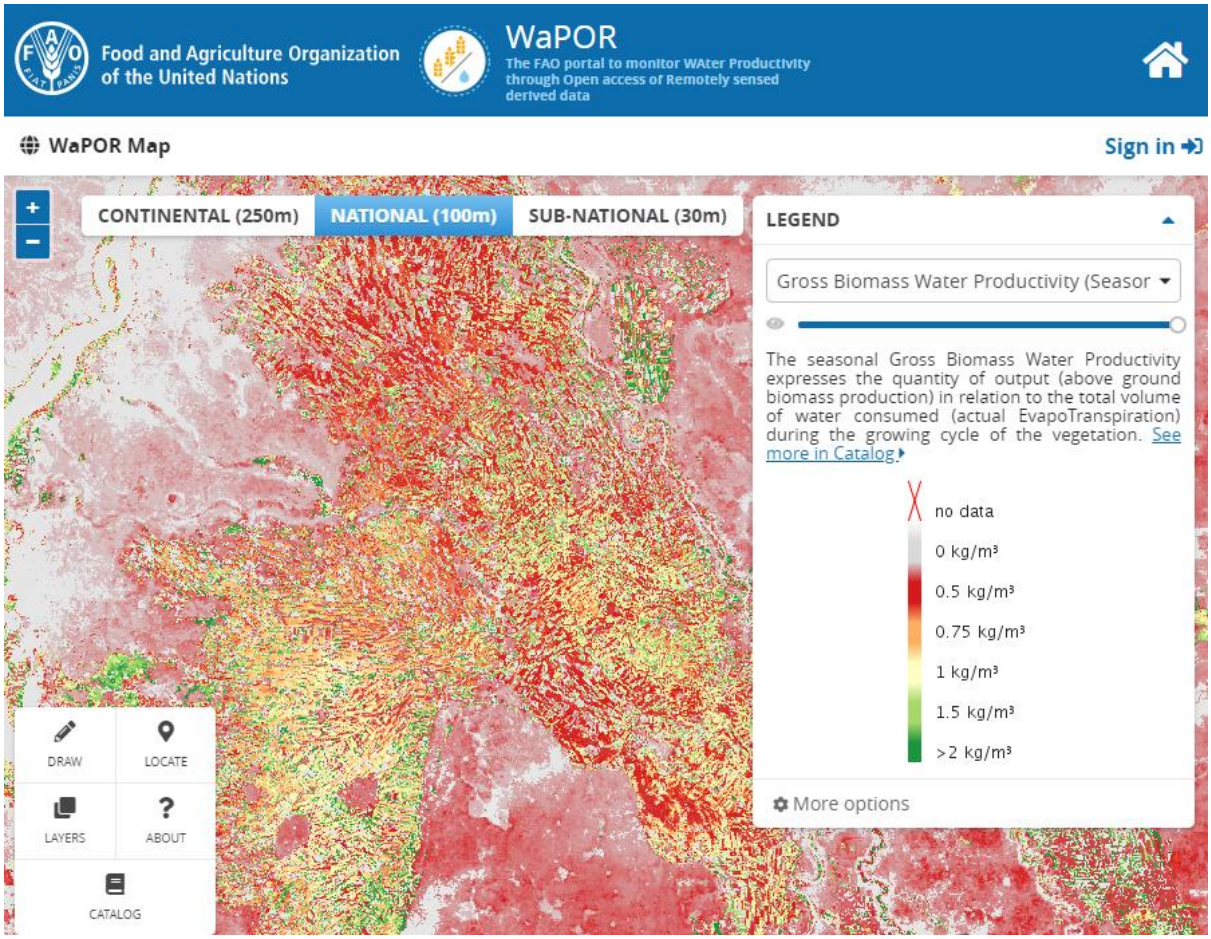
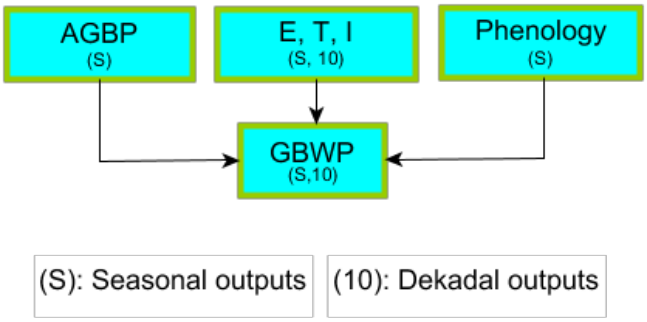


Figure 3: Example of seasonal gross biomass water productivity in the Nile Delta, Season 1 of 2017.
Source: FAO WaPOR, https://wapor.apps.fao.org

Gross biomass water productivity is calculated and made available through WaPOR on seasonal basis at Level 2. However, as the input data are also available on dekadal basis, user-defined temporal aggregations are possible.

Methodology

Box 2: Gross biomass water productivity in relation to other data components.



- ✓ Calculating GBWP requires input from above ground biomass production, evaporation, transpiration and interception and phenology if calculated on seasonal time step.
- ✓ No external data source is required to calculate GBWP.
- ✓ The output is not used in any other data component.

The calculation of gross biomass water productivity is as follows:

$$GBWP = \frac{AGBP}{E + T + I} \qquad (1)$$

Where AGBP is above ground biomass production in kgDM/ha. E is evaporation, T is transpiration and I is interception, all in m^3. The following data is used for calculating GBWP: AGBP, E, T, I and phenology if calculated on seasonal time step.

2.1.1.2. Net Biomass Water Productivity

Description

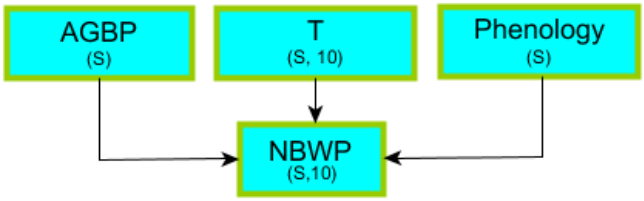
The net biomass water productivity expresses the quantity of output (above ground biomass production) in relation to the volume of water beneficially consumed (by canopy transpiration) in the year, and thus net of soil evaporation.

Contrary to gross water productivity, net water productivity is particularly useful in monitoring how effectively vegetation (and, more importantly, crops) uses water to develop biomass (and thus yield).

Net biomass water productivity is calculated and made available through WaPOR on a seasonal basis at Level 2. However, as the input data are also available on dekadal basis, user-defined temporal aggregations are possible[4].

Methodology

Box 3: Net biomass water productivity in relation to other data components.



- ✓ Calculating NBWP requires input from above ground biomass production, transpiration and phenology if calculated on seasonal time-step.
- ✓ No external data source is required to calculate NBWP.
- ✓ The output is not used in any other data component.

The calculation of net biomass water productivity is as follows:

$$NBWP = \frac{AGBP}{T} \qquad (2)$$

Where AGBP is above ground biomass production in kgDM/ha and T is transpiration in m³ (section 1). The following data is used for calculating NBWP: AGBP, T and phenology if calculated on seasonal time-step.

Table 4: Overview of Biomass Water productivity data components

Data component	Unit	Range	Use	Temporal resolution
GBWP	kg/m³	0 to 6[4]	Measures quantity of biomass output in relation to consumptive water use	Seasonal (further aggregated to user-defined)
NBWP	kg/m³	0 to 6[5]	Measures quantity of biomass output in relation to transpiration (or beneficial water consumption)	Seasonal (further aggregated to user-defined)

2.1.2. Phenology

Description

Phenology indicates the cycle or season of a crop and, in this case, is defined by the dekad corresponding to the start, maximum and end of the growing season. This information can be derived from satellite-based vegetation index time series.

At Level 2 Phenology is available for a maximum of two growing seasons annually. The phenology for one growing season is delivered as three raster files of which the pixel values are expressed in dekad numbers. The first raster indicates the Start of Season (SOS), the second the Maximum of Season (MOS) and the third represents the End of Season (EOS). With a maximum of 2 growing seasons annually, a full year is therefore described by 6 raster files.

Figure 4 shows an example of the Phenology data component (Start of Season 1, 2015) at Level 2.

[4] Range observed in WaPOR area, but theoretical range could go up to 25.

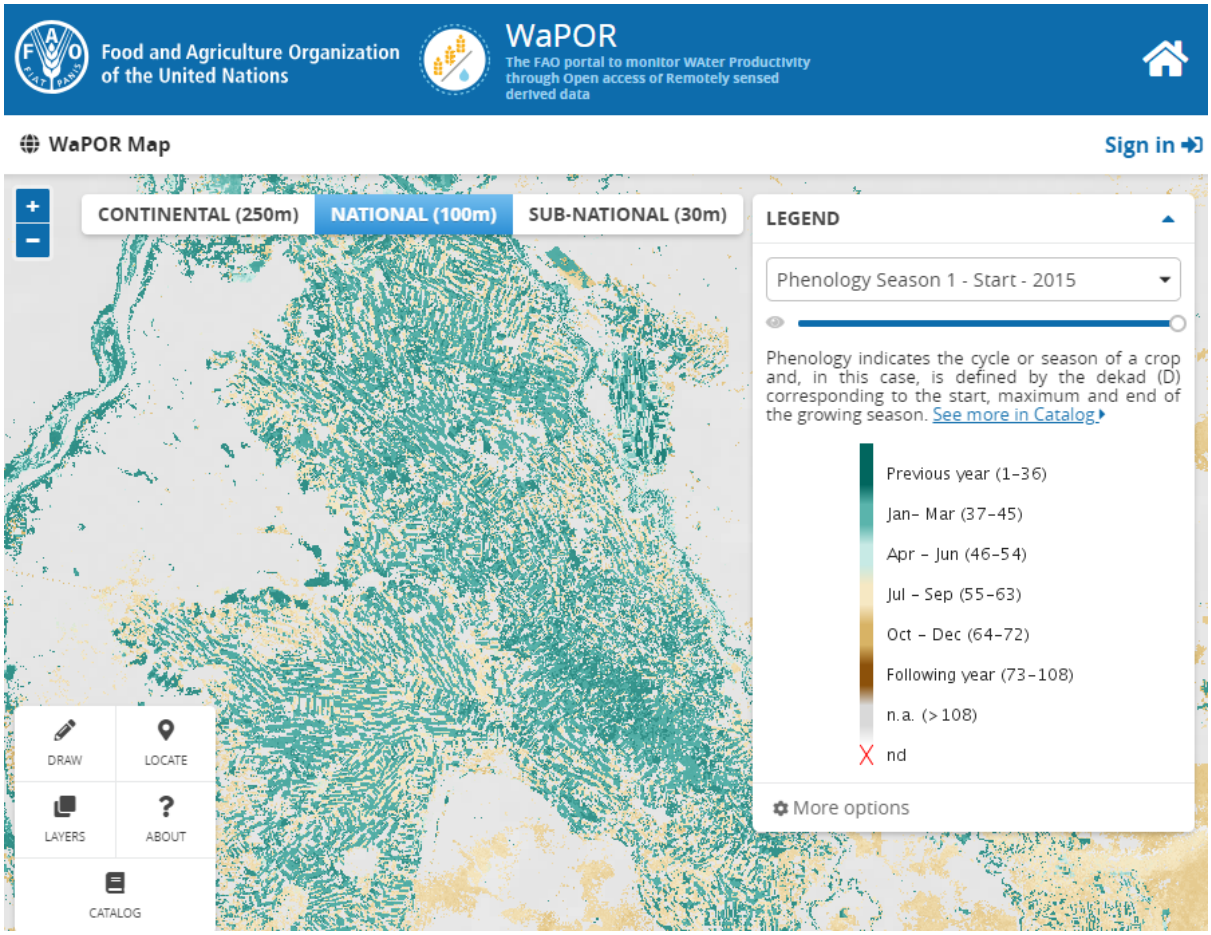
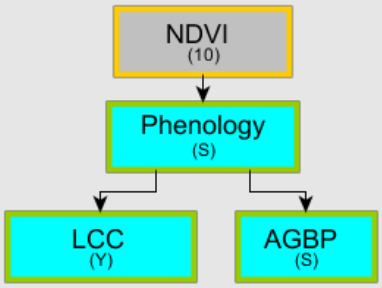


Figure 4: Example of Phenology data at Level 2, showing the Startof Season 1 (2015).
Source: FAO WaPOR, https://wapor.apps.fao.org

Methodology

Box 4: Crop Calendar in relation to other data components.



(Y): Annual outputs

- ✓ Calculating the Phenology only requires input from NDVI time series based on dekadal values.
- ✓ No external data source is required.
- ✓ The output is used to calculate Above Ground Biomass Production and is used as an input to the Land Cover Classification.

To determine the Start, Maximum and End of up to two seasons at Level 2 for a given calendar year (January - December) WaPOR applies[5] the methodology described by Van Hoolst et al. (2016)[6]. This methodology can derive phenological information from a time series of dekadal vegetation index composites (NDVI). The input dekadal NDVI time series covers exactly three calendar years (3 x 36 = 108 dekads), with the target year in the middle. The output "dates" of the Start, Maximum and End of season are expressed in dekads, numbered from the start of the time series spanning 3 years (1-36 for the previous year, 37-72 for the target year, 73-108 for the next year[7]).

Phenology outputs can be prone to some variability due to the inherent structure of the data and methodology. The quality of the NDVI time series plays a determining role in the outcome of the Phenology data component. Noise in the data can create local maxima/minima which can be mistaken for separate growing seasons. Phenology parameters also strongly depend on the definitions of the start/end of growing season. This makes comparison with other data sources on start/end of growing season difficult. A growing season is included in a calendar year if the End of season occurs in it. Difficulties arise when the End of season occurs close to the start of a calendar year as it will be a wrong representation of when the season took place. To circumvent this, a growing season is counted in a target year only if the End of season falls after the first 3 dekads of this calendar year. For instance if an EOS is recorded in dekad 2 of 2018, the growing season will be attributed to 2017 instead. In the case of evergreens, it is likely that a SOS, MOS and EOS will not be identified .If no growing season can be distinguished, a 'no season' label will be applied, as can clearly be seen in the tropical forests.

By definition the EOS lies in the target year or in the first three dekads of the next year. The SOS, however, can be situated in either the target year or the previous year. If two seasons occur, SOS1 < MOS1 < EOS1 < SOS2 < MOS2 < EOS2. Since EOS1 and EOS2 are by definition situated in the target year, this also holds for the intermediate MOS2 and SOS2. SOS1 and MOS1 can be situated in the previous or the target year.

[5] Using PHENOdef, developed in GLIMPSE (Global Image Processing Software).

[6] The publication describes the methodology as applied to the FAO-ASIS project. This was applied using SPOT-VEGETATION (1km) data. For WaPOR the methodology is applied to higher resolution input data that provide more spatial detail and are less influenced by mixing effects. As a consequence, it is expected that the estimated results (SOS, EOS) will be more precise and land cover specific.

[7] For example, if the target year is 2016, dekad 37 represents 1-10 January 2016, dekad 1 represents 1-10 January 2015. Dekad 30 represents 21-31 October 2015.

Table 5: Overview of the Phenology data component

Data component	Unit	Range	Use	Temporal resolution
Phenology	dekad	1-75[1]	It is used to calculate AGBP and as input parameter for the land cover.	Seasonal

[1]Where 36+3<EOS<72+3 and SOS1<MOS1<EOS1<SOS2<MOS2<EOS2

2.1.3. Evaporation, Transpiration and Interception

Description

Evapotranspiration (ET) is the sum of the soil evaporation (E), canopy transpiration (T) and interception (I). The interception describes the rainfall intercepted by the leaves of the plants that will be directly evaporated from their surface. This concept will be further explained below. The Evaporation, Transpiration and Interception are limited by climate (wind speed, radiation and air temperature) and soil conditions (soil moisture content). The sum of all three parameters i.e. the Actual EvapoTranspiration and Interception (ETIa) can be used to quantify the agricultural water consumption. In combination with biomass production or yield it is possible to derive the agricultural water productivity.

Evaporation, transpiration, interception and ETIa are delivered for Level 2 on a dekadal basis, where pixel values represent the average daily E, T, I and ETIa values[8] for that specific dekad in mm/day. Accumulation on seasonal or annual basis can also be found on the WaPOR portal for these four parameters. Figure 5 shows an example of the ETIa data component at Level 2.

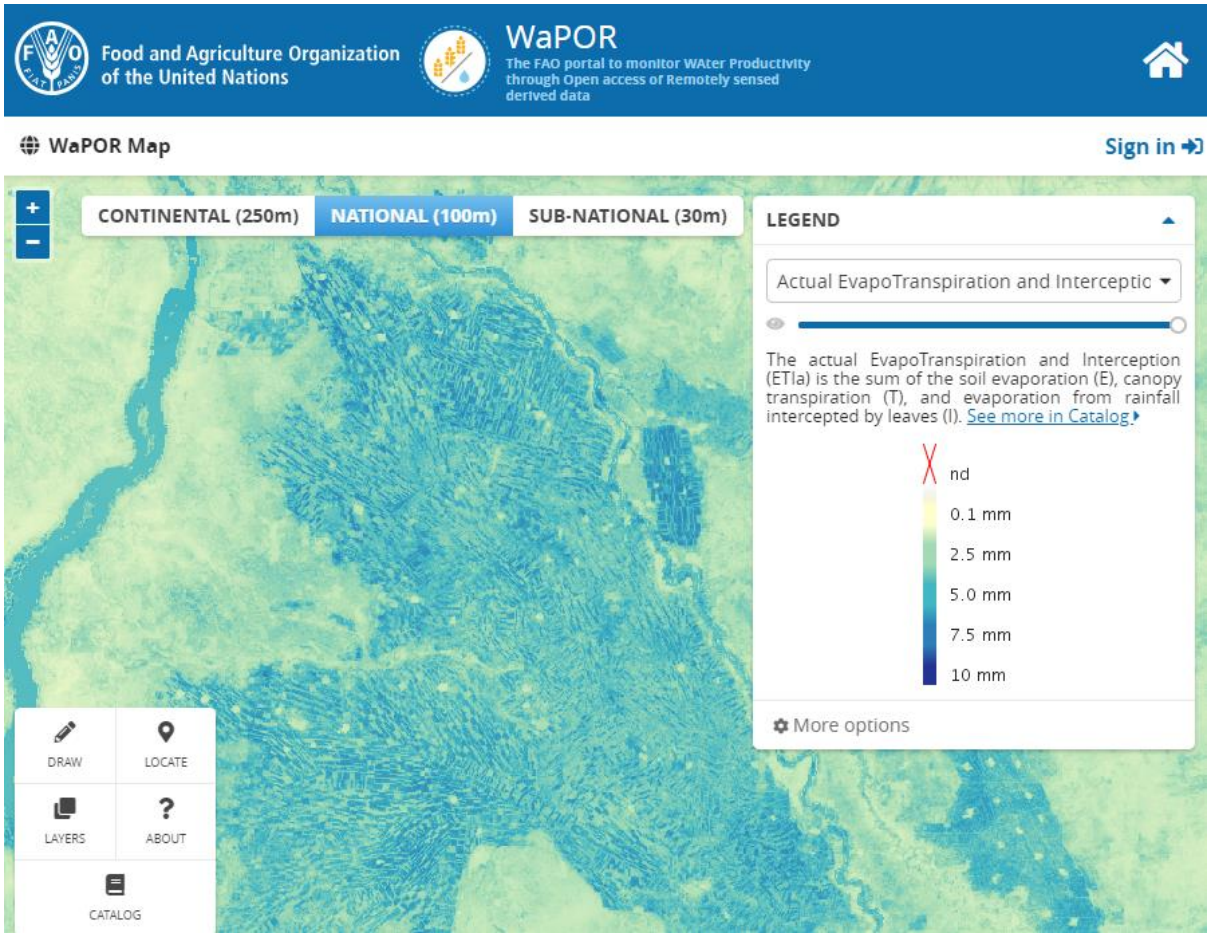


Figure 5: Example of ETIa data component at Level 2 (2017, dekad 27).
Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor

Of all data components, E and T require the largest number of inputs to calculate (see Figure 2 and the summary in Box 5). Only the external optical satellite data is available at the three resolutions of Levels 1 (250 m), 2 (100 m) and 3 (30 m) whilst the other external input data sources all have a

[8] Average daily E, T, I and ETIa values can be converted into volume for a specific area, e.g. 1 mm = 1 l/m^2 or 1 mm = 10 m^3/ha.

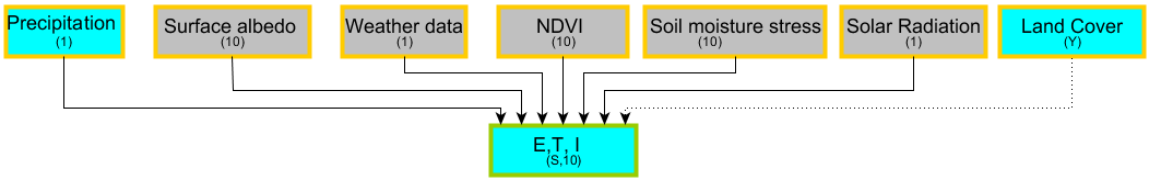
(significantly) lower resolution[9]. The spatial variability of these data sources is therefore more limited, thereby affecting the resulting E and T data component.

The collection of optical satellite data can be hampered by the presence of clouds, reducing the information on temporal variability. Although both aspects are accommodated for within the data processing chain, its implications should be understood when considering the results: the quality of the E, T, and I data component is a combination of the accuracy of the algorithms and the quality of the external data. One additional data layer is provided that indicates the quality of the input data for NDVI (described in Section 2.2.1).

Methodology

The method to calculate E and T is based on the ETLook model described in Bastiaanssen et al. (2012). It uses the Penman-Monteith (P-M) equation, adapted to remote sensing input data. The Penman-Monteith equation (Monteith, 1965) predicts the rate of total evaporation and transpiration using commonly measured meteorological data (solar radiation, air temperature, vapour pressure and wind speed). It has become the FAO standard for calculating the actual and reference evapotranspiration. FAO irrigation and drainage paper 56 (Allen et al., 1998) describes the method in detail[10]. The reader is advised to consult this document for detailed information on the use of the P-M equation and guidelines regarding the calculation of evapotranspiration.

Box 5: Evaporation, transpiration and interception in relation to other data components.



- ✓ Calculating E and T requires input from seven data components. Solar radiation, Weather data and Precipitation are daily inputs. Soil moisture stress, NDVI and Surface albedo are dekadal inputs. I only requires input from NDVI and Precipitation.
- ✓ Land Cover input is used to derive surface roughness and minimum stomatal resistance.
- ✓ No external data sources are used to calculate E, T and I.
- ✓ E, T and I are used as input to Water Productivity.
- ✓ E, T and I are calculated on a dekadal basis.

This section considers the P-M equation from a remote sensing perspective, i.e. implementation in an operational environment. This is done by dissecting the P-M equation to the level of the input data, consisting of 7 (final or intermediate) data components (see Box 5). In order to understand the processing chain for the E, T and I data components, the reader is advised to consult the relevant sections in this chapter for explanations of all the input data components.

[9] For example, temperature data has a spatial resolution of 0.25 degrees (~28 km) and atmospheric transmissivity has a spatial resolution of 4 km.

[10] FAO irrigation and drainage paper 56 (Allen et al. 1998) can be found on the FAO website: www.fao.org/docrep/X0490E/x0490e00.htm.

Penman-Monteith equation (ET)

The Penman-Monteith equation is also known as the combination-equation because it combines two fundamental approaches to estimate evaporation (Allen et al., 2005). These are the surface energy balance equation and the aerodynamic equation. The Penman-Monteith equation is expressed as:

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \qquad (12)$$

where:

λ latent heat of evaporation [J kg^{-1}]
E evaporation [kg m^{-2} s^{-1}]
T transpiration [kg m^{-2} s^{-1}]
R_n net radiation [W m^{-2}]
G soil heat flux [W m^{-2}]
ρ_a air density [kg m^{-3}]
c_p specific heat of dry air [J kg^{-1} K^{-1}]
e_a actual vapour pressure of the air [Pa]
e_s saturated vapour pressure [Pa] which is a function of the air temperature
Δ slope of the saturation vapour pressure vs. temperature curve [Pa K^{-1}]
γ psychrometric constant [Pa K^{-1}]
r_a aerodynamic resistance [s m^{-1}]
r_s bulk surface resistance [s m^{-1}]

The ETLook model solves two versions of the P-M equation: one for the soil evaporation (E) and one for the canopy transpiration (T):

$$\lambda E = \frac{\Delta(R_{n,soil} - G) + \rho_a c_p \frac{(e_s - e_a)}{r_{a,soil}}}{\Delta + \gamma(1 + \frac{r_{s,soil}}{r_{a,soil}})} \qquad (13)$$

and

$$\lambda T = \frac{\Delta(R_{n,canopy}) + \rho_a c_p \frac{(e_s - e_a)}{r_{a,canopy}}}{\Delta + \gamma(1 + \frac{r_{s,canopy}}{r_{a,canopy}})} \qquad (14)$$

The two equations differ with respect to the net available radiation ($R_{n,soil}$ and $R_{n,canopy}$) as well as the aerodynamic and surface resistance ($r_{a,soil}, r_{s,soil}$ and $r_{a,canopy}, r_{s,canopy}$). Furthermore, the soil heat flux (G) is not taken into account for transpiration.

The Net Radiation and the Aerodynamic and Surface Resistance are discussed in more detail below. The other parameters of the equation are not taken into further consideration, as these are constants or variables that can be derived directly from mathematical relationships.

The main concepts of the ETLook model are illustrated in a schematic representation in Figure 6.

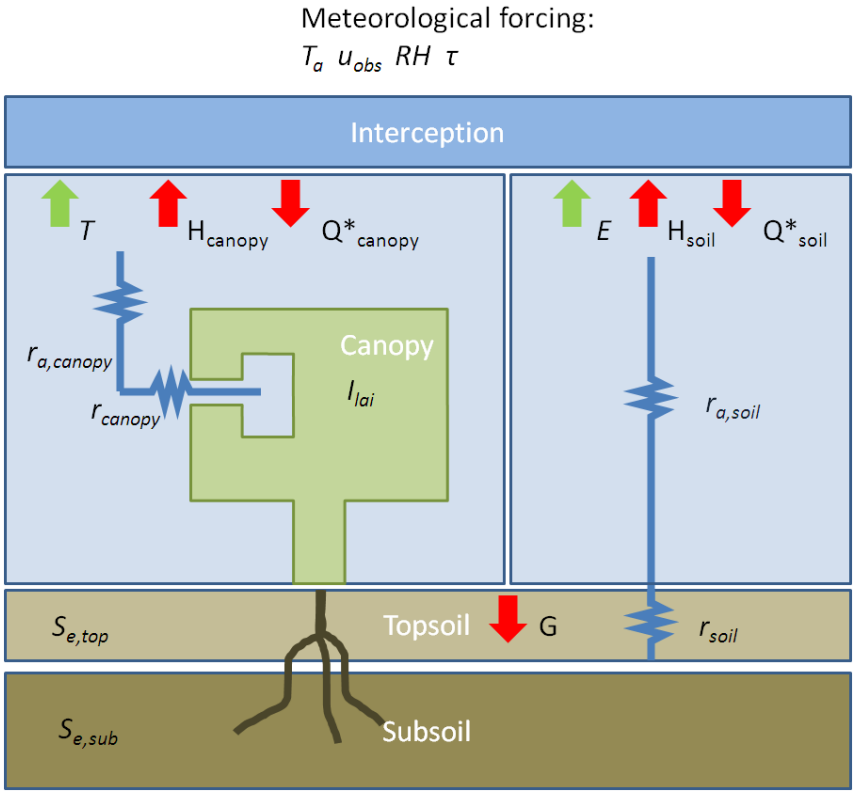


Figure 6: Schematic diagram illustrating the main concepts of the ETLook model, where two parallel Penman-Monteith equations are solved. For transpiration the coupling with the soil is made via the subsoil or root zone soil moisture content whereas for evaporation the coupling is made via the soil moisture content of the topsoil. Interception is the process where rainfall is intercepted by the leaves and evaporates directly from the leaves using energy that is not available for transpiration.

Net radiation (R_n)

The net radiation R_n represents the available energy at the earth's surface, which can be described by the radiation balance:

$$R_n = (1 - \alpha_0)R_s - L^* - I \quad (15)$$

where α_0 is the surface albedo [-], R_s is incoming solar radiation [W m^{-2}], L^* is net long wave radiation [W m^{-2}], I represents energy dissipation due to interception losses [W m^{-2}].

The net radiation is derived differently for the soil and canopy. Leaf area index I_{lai}, a measure of canopy density, is used to separate the net radiation into soil net radiation and canopy net radiation. An increase in leaf area index results in an exponential decrease in the fraction of the radiation available for the soil as more is captured by the canopy. The division is calculated using Beer's law (which describes the attenuation of light through a material), leading to the following descriptions of soil and canopy net radiation:

$$R_{n,soil} = R_n \exp(-aI_{lai}) \quad (16)$$

$$R_{n,canopy} = R_n\left(1 - \exp(-aI_{lai})\right) \quad (17)$$

where a is the light extinction factor for net radiation [-].

The leaf area index (LAI) I_{lai} [m^{2}m^{-2}] describes the amount of green leaf area per unit of soil area. A leaf area index equal to zero indicates that there is no vegetation present, a leaf area index larger than zero indicates the presence of green leaves. The NDVI I_{ndvi} [-] is used to derive I_{lai}. This is done in

two steps. First, NDVI is used to calculate vegetation cover c_{veg}, which is subsequently converted into leaf area index. The two equations below describe this conversion for a specific range of the NDVI value.

$$\begin{cases} c_{veg} = 0 & I_{ndvi} \leq 0.125 \\ c_{veg} = 1 - \left(\frac{0.8 - I_{ndvi}}{0.8 - 0.125}\right)^{0.7} & 0.125 < I_{ndvi} < 0.8 \\ c_{veg} = 1 & I_{ndvi} \geq 0.8 \end{cases} \quad (18)$$

The second step is the conversion from vegetation cover to leaf area index I_{lai} according to the following relationships:

$$\begin{cases} I_{lai} = 0 & I_{ndvi} \leq 0.125 \\ I_{lai} = \frac{\ln\left(-\left(c_{veg} - 1\right)\right)}{-0.45} & 0.125 < I_{ndvi} \leq 0.795 \\ I_{lai} = 7.63 & I_{ndvi} > 0.795 \end{cases} \quad (19)$$

This relationship has been derived using a large number of LAI functions compiled from literature (e.g. Carlson and Ripley, 1997; Duchemin, et al., 2006). The above relationship represents the average from these compiled relationships.

Interception is the process where rainfall is intercepted by the leaves. This evaporates directly from the leaves and requires energy that is not available for transpiration. Interception I [mm day^{-1}] is a function of the vegetation cover, LAI and precipitation (*P*), expressed as:

$$I_{mm} = 0.2 I_{lai} \left(1 - \frac{1}{1 + \frac{c_{veg} P}{0.2 I_{lai}}}\right) \quad (20)$$

Interception is relatively high with a small amount of precipitation, with the fraction intercepted decreasing quickly as precipitation increases. The maximum interception is determined by the LAI. The energy I needed to evaporate I_{mm} is calculated as follows:

$$I = I_{mm} \frac{\lambda}{86{,}400} \quad (21)$$

where:

λ latent heat of evaporation [J kg^{-1}]

The net long wave radiation L^*, i.e. the difference between the incoming and outgoing long wave radiation, is computed using the formulation described in FAO report no 56 (Allen et al., 1998). This is a function of the air temperature (T_a), actual vapour pressure (e_a) and transmissivity (τ).

As indicated above, the total evapotranspiration is obtained by summing the soil evaporation and canopy transpiration calculated from the Penman-Monteith equation and the interception by the leaves.

Soil heat flux (G)

The soil heat flux G is required to calculate evaporation from the soil surface. It is calculated according to FAO report no 56 (Allen et al., 1998). For northern latitudes, the maximum value for G is recorded in May. For southern latitudes this occurs in November. For northern latitudes it is calculated with the equation below. $-\pi/4$ is replaced by $3\pi/4$ for southern latitudes.

$$G = \frac{\sqrt{2}A_{t,year}k\sin(\frac{2\pi J}{p} - \frac{\pi}{4})}{z_d}\exp(-aI_{lai}) \quad (22)$$

where:

$A_{t,year}$	yearly air temperature amplitude [K]
k	soil thermal conductivity [W m^{-1} K^{-1}]
J	day of year [-]
p	number of days in year [-]
z_d	damping depth [m]
I_{lai}	leaf area index [-]
a	light extinction factor for net radiation [-] (same as in (16) and (17))

The damping depth (z_d) and the soil thermal conductivity (k) depend on soil characteristics. Usually these are taken as constants. The yearly air temperature amplitude is derived from climatic data.

Surface resistances (r_s)

The surface resistances in the Penman-Monteith equations describe the influence (resistance) of the soil and the canopy on the flow of vapour in relation to evaporation and transpiration.

The soil resistance $r_{s,soil}$ is modelled using the minimal soil resistance $r_{soil,min}$ and relative soil moisture content S_e by means of a constant power function (Camillo and Gurney, 1986; Clapp and Hornberger, 1978; Dolman, 1993; Wallace et al., 1986):

$$r_{s,soil} = r_{soil,min}(S_e)^{-2.1} \quad (23)$$

The canopy resistance is a function of the leaf area index, minimum stomatal resistance $\boldsymbol{r_{canopy,min}}$ and a number of reduction factors (Jarvis, 1976; Stewart, 1988). The Jarvis-Stewart parameterization describes the joint response of soil moisture and LAI on transpiration considering meteorological conditions (solar radiation, temperature and relative humidity ϕ):

$$r_{s,canopy} = \left(\frac{r_{canopy,min}}{I_{lai,eff}}\right)\left(\frac{1}{S_t S_v S_r S_m}\right) \quad (24)$$

where:

$r_{canopy,min}$	minimum stomatal resistance [s m^{-1}]
$I_{lai,eff}$	effective leaf area index [-]
S_t	temperature stress [-], a function of minimum, maximum and optimum temperatures as defined by Jarvis (1976)
S_v	vapour pressure stress induced due to persistent vapour pressure deficit [-]
S_r	radiation stress induced by the lack of incoming shortwave radiation [-]
S_m	soil moisture stress originating from a lack of soil moisture in the root zone [-]

The minimum stomatal resistance $r_{canopy,min}$ can have different values for different types of vegetation. This is derived from land cover information. The canopy resistance equation is based on a single leaf layer, therefore effective leaf area index has to be calculated as follows (Mehrez et al., 1992; Allen et al., 2006a):

$$I_{lai,eff} = \frac{I_{lai}}{0.3I_{lai} + 1.2} \quad (25)$$

Aerodynamic resistance (r_a)

The aerodynamic resistance has to be calculated for both neutral and non-neutral conditions. Neutral conditions exist when turbulence is created by shear stress (wind) only. Buoyancy (thermal rise of air) causes unstable non-neutral conditions. Under neutral conditions the aerodynamic resistance for soil ($r_{a,soil}$) and canopy ($r_{a,canopy}$) can be computed (Allen et al., 1998; Choudhury et al., 1986; Holtslag, 1984) with:

$$r_{a,soil} = \frac{\ln\left(\frac{z_{obs}}{z_{0,soil}}\right)\ln\left(\frac{z_{obs}}{0.1z_{0,soil}}\right)}{k^2 u_{obs}} \quad (26)$$

$$r_{a,canopy} = \frac{\ln(\frac{z_{obs} - \mathrm{d}}{z_{0,canopy}})\ln(\frac{z_{obs} - \mathrm{d}}{0.1z_{0,canopy}})}{k^2 u_{obs}} \quad (27)$$

Where:

k	von Karman constant [-]
u_{obs}	wind speed at observation height [m s^{-1}]
d	displacement height [m]
$z_{0,soil}$	soil surface roughness [m]
$z_{0,canopy}$	canopy surface roughness [m]
z_{obs}	observation height [m]

The soil and canopy surface roughness are derived from land cover and NDVI. Land cover classes are used to assign the obstacle height from which surface roughness to momentum ($z_{0,m}$) is derived. To account for seasonal variation during the growing season, NDVI is used to scale the obstacle height for vegetation.

Under non-neutral conditions also the turbulence generated by buoyancy should be included. The Monin-Obukhov similarity theory (Monin and Obukhov, 1954) is used to describe the effect of buoyancy on the turbulence by means of stability corrections:

$$r_{a,soil} = \frac{\ln\left(\frac{z_{obs} - d}{0.1z_{0,soil}}\right) - \psi_{h,obs}}{ku_*} \quad (28)$$

$$r_{a,canopy} = \frac{\ln(\frac{z_{obs} - d}{0.1z_{0,m}}) - \psi_{h,obs}}{ku_*} \quad (29)$$

Where $\psi_{h,obs}$ is the stability correction for heat which is a function of z_{obs}, d and L, the Monin-Obukhov length defined as:

$$L = \frac{-\rho c_p u_*^3 T_a}{kgH} \quad (30)$$

Where:

T_a	air temperature [K]
u_*	friction velocity [m s^{-1}]
H	sensible heat flux (see text below)

The Monin-Obukhov length can be thought of as the height in the boundary layer at which the contribution of shear stress to turbulence is equal to the contribution of buoyancy to turbulence.

Both the aerodynamic resistance under non-neutral conditions and the sensible heat flux, the source of this non-neutral condition, are unknown variables. They can only be solved through an iterative process. A first estimate of the sensible heat flux H using the definitions for $r_{a,soil}$ and $r_{a,canopy}$ under neutral conditions provides a first estimate for the Monin-Obukhov length. The stability corrections $\psi_{h,obs}$ are then introduced in an iterative approach. When the iterations are converging, final values of evaporation and transpiration can be calculated. Iterations typically converge after only a small number of iterations (usually approximately 3).

ET conversion to mm

When the aerodynamic resistances are solved, evaporation and transpiration can be calculated. At this stage of the calculations they are still expressed as the available energy for evaporation and transpiration [W m^{-2}], hence the notation: λET, λE, λT in the P-M equation. These are then converted to mm:

$$E = \lambda E \left(\frac{t_{day}}{\lambda}\right) = \lambda E \left(\frac{86{,}4000}{2{,}453{,}780}\right) \approx 0.035 \lambda E \qquad (31)$$

Where t_{day} is the number of seconds in a day (86,400) and λ is the latent heat of evaporation which is a function of temperature, λ at 293 K is equal to 2,453,780.

A similar equation can be used for λET, λT. The equation for λ is as follows:

$$\lambda = \lambda_0 + c * T \qquad (32)$$

Where c = -2,361 J/kg/C and λ_0 = 2,501,000 J/kg

Table 6: Overview of E, T, I and ETIa data components

Data component	Unit	Range	Use	Temporal resolution
Evaporation	mm /day	0-10[11]	Measures soil evaporation in the period of a dekad	Dekadal
Transpiration	mm /day	0-2	Measures canopy transpiration in the period of reference	Dekadal
Interception	mm /day	0-2	Measures canopy interception in the period of reference	Dekadal

[11] Range of values for evaporation on land. On open water, we find values up to 15 mm/day.

Actual evapotranspiration and Interception (ETIa)	mm /day %	0-12	It can be used to quantify the agricultural water consumption. In combination with biomass production or yield, it is possible to derive the agricultural water productivity.	Dekadal

2.1.4. Net Primary Production

Description

Net Primary Production (NPP) is a fundamental characteristic of an ecosystem, expressing the conversion of carbon dioxide into biomass driven by photosynthesis. NPP is part of a family of definitions describing the carbon fluxes between the ecosystem and the atmosphere. Gross Primary Production (GPP) represents the carbon uptake by the standing biomass due to photosynthesis. NPP is the GPP minus autotrophic respiration, the losses caused by the conversion of basic products (glucose) to higher-level photosynthates (starch, cellulose, fats, proteins) and the respiration needed for the maintenance of the standing biomass. NEP or Net Ecosystem Production also accounts for the contribution of soil respiration, i.e. the re-conversion to CO_2 of leaf and other litter by soil micro-flora. Finally, subtracting the losses due to disturbance and anthropogenic removals gives the Net Biome Production (NBP). Figure 7 shows a schematic overview of carbon fluxes.

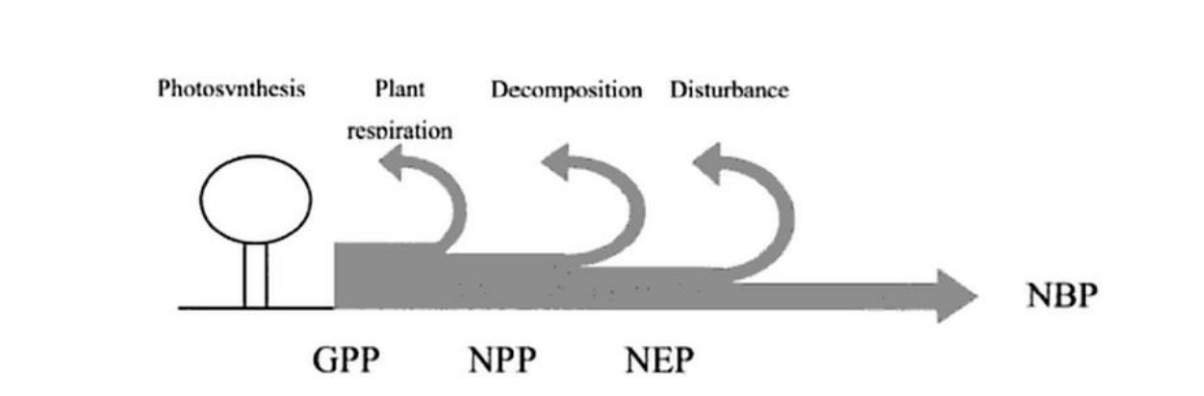


Figure 7: The component fluxes and processes in ecosystem productivity. GPP: Gross Primary Production, NPP: Net Primary Production, NEP: Net Ecosystem Production, NBP: Net Biome Production (Valentini, 2003)

NPP is derived from satellite imagery and meteorological data. The core of the methodology has been detailed in Veroustraete et al. (2002), whilst the practical implementation[12] is described in Eerens et al. (2004). These methodologies were improved within the framework of the Copernicus Global Land Component[13], the most important change being the incorporation of biome-specific light use efficiencies (LUEs). WaPOR applies this updated methodology. At Level 2, crops are not identified and therefore only one generic cropland LUE value of 2.49 is used. A reduction factor for soil moisture stress that accounts for short-term water deficiency is also added.

NPP is delivered for all three levels on a dekadal basis, where pixel values represent the average daily net primary production for that specific dekad in $gC/m^2/day$. In some cases, such as for agricultural purposes, it is more appropriate to measure Dry Matter Production (DMP, in kgDM/ha/day). NPP can

[12] The practical implementation was developed for the MARS Crop Yield Forecasting System (Eerens *et al.*, 2004)

[13] More information, including the validation report can be found at http://land.copernicus.eu/global/products/dmp.

be converted to DMP using a constant scaling factor of 0.45 gC/gDM (Ajtay et al., 1979). Therefore 1 gC/m²/day (NPP) = 22.222 kgDM/ha/day (DMP). Typical values for NPP vary within the region between 0 and 5.4 gC/m²/day (NPP), or 0 to 120 kgDM/ha/day (DMP), although higher values can occur (theoretically up to 320 kgDM/ha/day). Figure 8 shows an example of the NPP data component at Level 2.

It should be noted that the effects of several potentially important factors, such as nutrient deficiencies, pests and plant diseases are omitted in the calculation of the NPP product. However, it might be argued that the adverse effects of diseases and shortages of nutrients are manifested (sooner or later) via the remote sensing-derived fAPAR.

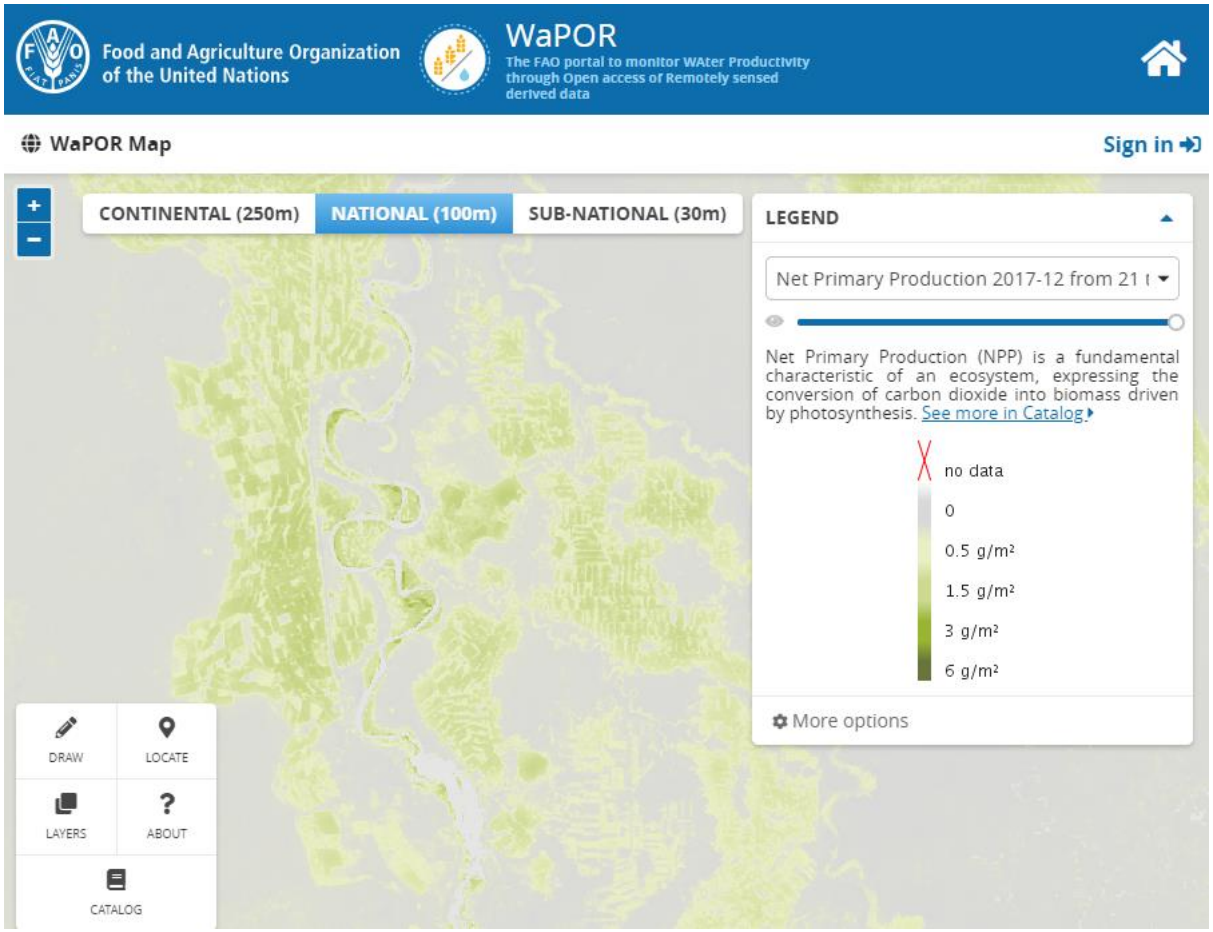
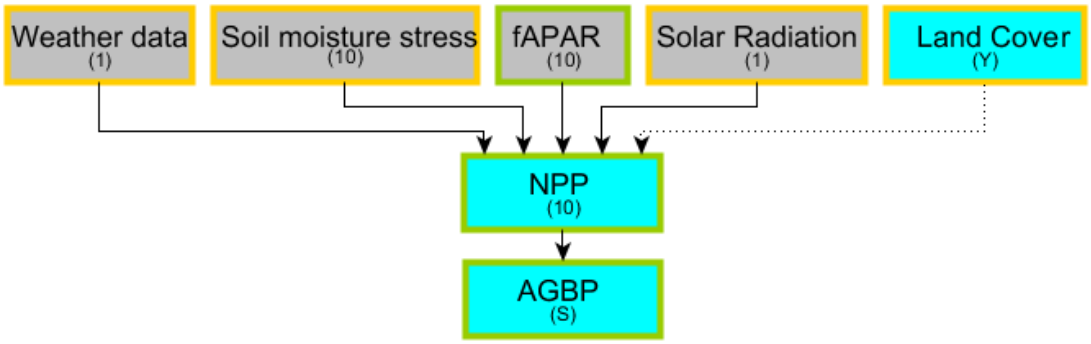


Figure 8: Example of NPP data component at Level 2 (2017, dekad 35).
Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor

Methodology

Box 6: Net Primary Production in relation to other data components.



- ✓ Calculating Net Primary Production requires daily input from Weather data and Solar radiation and dekadal input from fAPAR and Soil moisture stress.
- ✓ Seasonal or annual land cover is an indirect input as light use efficiencies are dependent on land cover.
- ✓ A soil moisture stress reduction factor is incorporated to adjust for water stress.
- ✓ No external data source is required to calculate Net Primary Production.
- ✓ NPP is produced on a dekadal basis.
- ✓ Dekadal NPP is used as input to calculate Above Ground Biomass Production.

Calculating NPP requires daily input from Weather data (T_{min}/T_{max}) and Solar radiation, as well as dekadal inputs from fAPAR and Soil moisture stress. Land Cover is an indirect input as Light Use Efficiency (LUE) is land cover specific.

The method to compute Net Primary Production is based on Monteith (1972), which describes ecosystem productivity in response to solar radiation. The equation is expressed as follows:

$$NPP = Sc\ R_s \varepsilon_p\ fAPAR\ SM\ \varepsilon_{lue}\ \varepsilon_T\ \varepsilon_{CO2}\ \varepsilon_{AR}\ [\varepsilon_{RES}] \qquad (34)$$

Where:

Sc	Scaling factor from DMP to NPP [-]
R_s	Total shortwave incoming radiation [GJ_T/ha/day]
ε_p	Fraction of PAR (0.4 – 0.7µm) in total shortwave 0.48 [JP/JT]
$fAPAR$	PAR-fraction absorbed (PA) by green vegetation [JPA/JP]
SM	Soil moisture stress reduction factor
ε_{lue}	Light use efficiency (DM=Dry Matter) at optimum [kgDM/GJPA]
ε_T	Normalized temperature effect [-]
ε_{CO2}	Normalized CO_2 fertilization effect [-]
ε_{AR}	Fraction kept after autotrophic respiration [-]
ε_{RES}	Fraction kept after residual effects (including soil moisture stress)[-]

The following are obtained from intermediate data components: incoming solar (shortwave) radiation[14] R_s (see section 2.2.2), *f*APAR (see section 2.2.4) and soil moisture stress (see section 2.2.3).

The fraction ε_p of PAR (Photosynthetically Active Radiation, 400-700 nm) within the total shortwave (200-4000 nm) varies slightly around the mean of ε_p=0.48, denoting that 48% of all incoming solar radiation is situated in the 400-700nm region. Although small variations occur, this value is kept constant.

Light Use Efficiency (ε_{LUE}) is a coefficient for the efficiency by which vegetation converts energy into biomass. It is a land cover specific variable and is derived from the last known land cover (see section 2.1.6). For the period 2009-2015, the correct LC was known for the production of NPP, and the correct LUE values were used. However, for 2016 onwards, the correct land cover is only known at a later stage, and changes in land cover will result in changes in LUE values and thus also in NPP values. To correct for these changes, a complementary LUE correction factor data layer is produced for to allow the user to adjust for the correct land cover (see relevant methodology documents).

The effect of temperature (ε_T), atmospheric CO_2 concentration (ε_{CO2}) and autotrophic respiration[15] (ε_{AR}) is simulated via rather complex biochemical equations (see Veroustraete et al., 2002). However, the influencing factors driving these biochemical processes are temperature (T) and CO_2 concentration. The CO_2 concentration is assumed to be constant over the globe, as well as within a year. The overall increasing trend in CO_2 concentrations, resulting in the greening effect of CO_2, is included by adjusting the CO_2 concentration with a linear function over time. This function was derived from the annual 'spatial' average of globally-averaged marine surface (CO_2) data from the NOAA-ESRL cooperative air sampling network of the last 15 years.

The factor ε_{RES} (residual) is added in the above equation to emphasize the fact that some potentially important factors, such as the effect of droughts, nutrient deficiencies, pests and plant diseases, influence NPP. The factor includes the effect of soil moisture stress.

Given the simple elaboration of the epsilons, equation 34 can be rewritten as follows:

$$\text{NPP} = \text{Sc.R}_s.\ \varepsilon_p.f\text{APAR.SM}.\ \varepsilon_{LUE}.\ \varepsilon_T.\ \varepsilon_{CO2}.\ \varepsilon_{AR} = \text{Sc.}\ f\text{APAR.SM}.\ \varepsilon_{LUE}\text{R}_s.\varepsilon(\text{T},\text{CO}_2)$$

$$= f\text{APAR.SM}.\ \varepsilon_{LUE}\text{NPP}_{max}\ [\varepsilon_{RES}\ ]$$

$$= f\text{APAR.SM}.\ \varepsilon_{LUE}\text{NPP}_{max}\ [\varepsilon_{RES}\ ] \qquad (35)$$

With: $\varepsilon(\text{T}, \text{CO}_2) = \varepsilon_P.\varepsilon_T.\varepsilon_{CO2}.\varepsilon_{AR}$.

This formulation better highlights the fact that, within the limits of the described model, NPP is only determined by six basic factors: *f*APAR, soil moisture stress, radiation, temperature, land cover specific light use efficiency and CO_2. However, in practice the CO_2 level is mostly considered as a global constant. At the same time, the above equation provides a practical method to bypass the differences in temporal (and spatial) resolution between the inputs. The meteorological inputs (R_s, T_{min}, T_{max}) are provided on a daily basis, *f*APAR and SM are derived from the dekadal data components and the final NPP product has a dekadal frequency.

In practice the procedure according to Eerens et al. (2004) and as illustrated in Figure 9 is applied.

[14] Solar radiation is mostly reported in terms of kJ_T/m^2/day with variations between 0 and 32,000. This corresponds with 320 GJ_T/ha/day (1 hectare is 10,000m², and 1 GJ is 1,000,000 kJ).

[15] The autotrophic respiration is calculated as a simple fraction of NPP and is therefore assumed to have the same ecophysiological behaviour. It is not considered as an independent component.

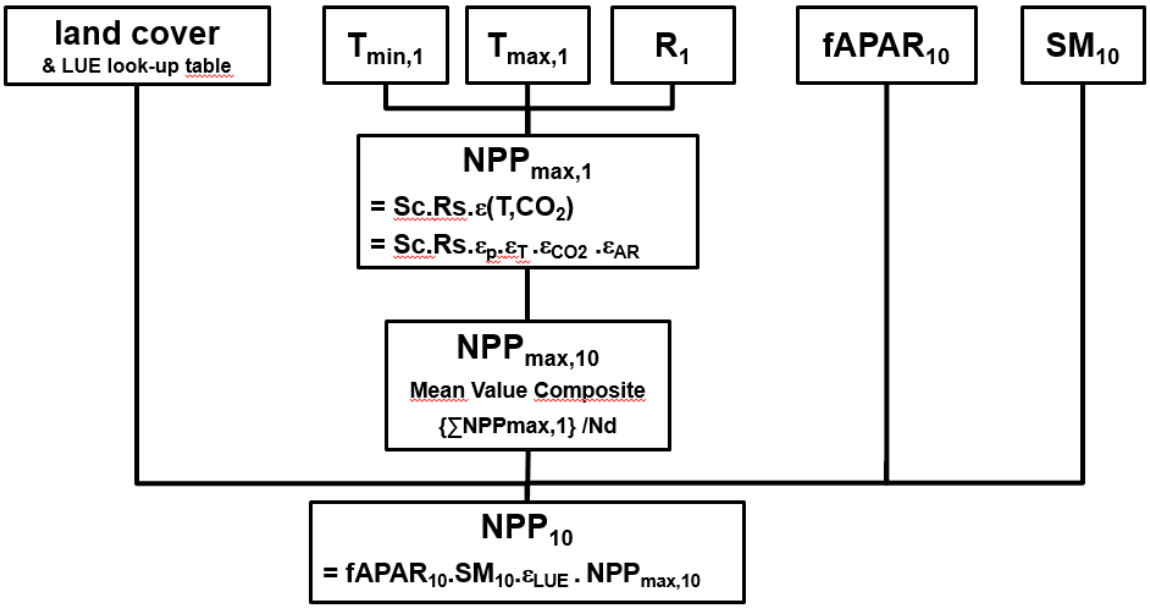


Figure 9: Detailed process flow of NPP. Daily NPP_{max} is estimated based on meteorological data. At the end of each dekad, a mean value composite of these NPP_{max} images is calculated. The final NPP_{10} product is retrieved by the simple multiplication of the mean value composite NPP_{max} with the fAPAR, soil moisture stress and the land cover dependent light use efficiency. (Eerens, 2004)

- Based on the meteorological inputs (R_s, T_{min}, T_{max}), the yearly fixed value of the CO_2 level and the above-mentioned variant of the Monteith equation, data are generated with:

$$NPP_{max} = Sc.R_s.\varepsilon(T,CO_2) = Sc.R_s.\varepsilon_P.\varepsilon_T.\varepsilon_{CO2}.\varepsilon_{AR} \quad (36)$$

- NPP_{max} represents the maximum obtainable NPP, for the (virtual) cases where fAPAR would be equal to one. LUE is assumed to be 1 at this stage.
- At the end of every dekad, a new data layer is computed with the mean of the daily $NPP_{max,1}$ scenes. Next, $NPP_{max,10}$, fAPAR and SM are simply multiplied to retrieve the final image with the NPP estimates. The LUE corresponding to the land cover is also applied in this step.

This practical approach can be formulated as follows (the subscripts 1 and 10 indicate daily and dekadal products, N_d is the number of days in each dekad):

$$NPP_{10} = fAPAR_{10} . SM. \varepsilon_{LUE}.NPP_{max,10} \quad (37)$$

$$\text{with} \quad NPP_{max,10} = \{\Sigma NPP_{max,1}\} / N_d \quad (38)$$

Table 7: Overview of NPP data component

Data component	Unit	Range	Use	Temporal resolution
Net primary Production (NPP)	$gC/m^2/day$	0-5.4[1] 0-13.5[2]	Indicates the conversion of carbon dioxide into biomass driven by photosynthesis;	Dekadal

[1]Typical range in the ROI
[2]Theoretical range for NPP

2.1.5. Above Ground Biomass Production

Description

Above Ground Biomass Production (AGBP) is defined as the sum of the above-ground dry matter produced during the growing season. Hence, AGBP steadily increases between the start (SOS) and end of season (EOS).

AGBP, expressed in kgDM/ha/day, typically ranges between 0 and 45, although higher values are possible. As the AGBP is an integration of the DMP over time, its accuracy is closely related to the accuracy of the NPP, which is discussed in Section 2.1.4. Figure 10 shows an example of the AGBP data component at Level 2.

At Level 2, AGBP is delivered on a seasonal basis. The seasonal value represents the total accumulated biomass during one growing season, from SOS to EOS.

A limitation for the derivation of AGBP is the dependency on phenological information, meaning that AGBP can only be derived for areas where seasonality is detected. For ecosystems, such as tropical forests or deserts, that experience almost no seasonality, the start of season is theoretically set at January 1st and end of season is set at December 31st.

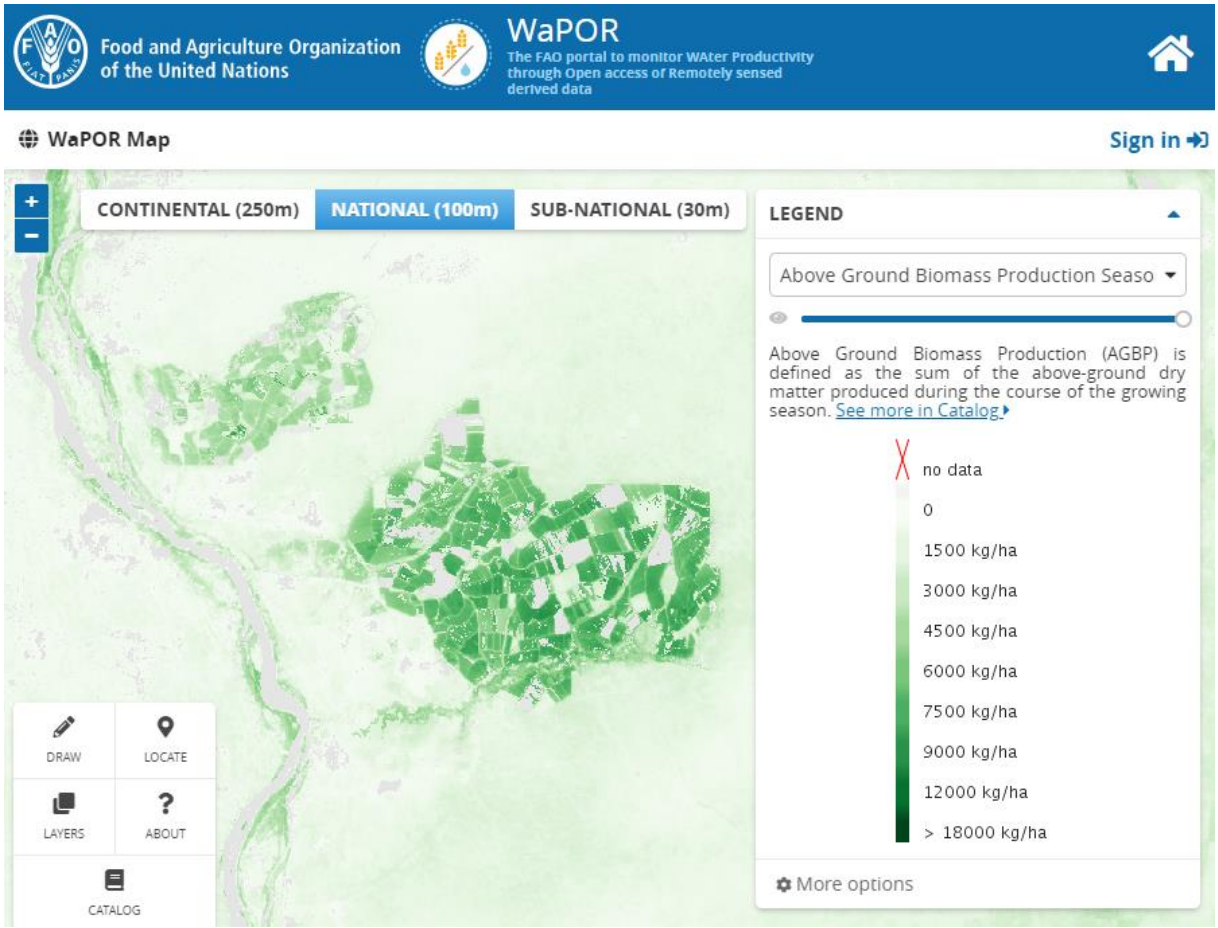
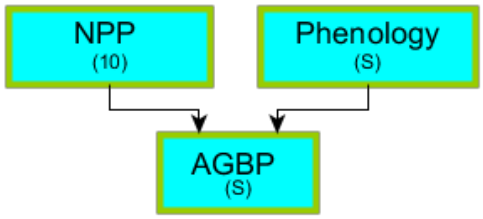


Figure 10: Example of AGBP data component at Level 2 (Season 1, 2017).

Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor

Methodology

Box 7: Above Ground Biomass Production in relation to other data components.



- ✓ Calculating AGBP requires input from NPP for dekadal biomass production and Phenology for demarcating the growing season.
- ✓ No external data source is required.
- ✓ The output is seasonal for Level 2.

To derive the accumulation in biomass production over or during a growing season, first the start and the end of the growing season need to be identified using the phenology data component. AGBP is then calculated as the sum of NPP, converted[16] into DMP units (kgDM/ha), between the start of the season (SOS) and the end of the season (EOS).

In addition, a factor (*F*) is included to account for the division between the above and below-ground components, here referred to as the AGBP Over Total (AOT) biomass production correction factor. According to literature, the above-ground fraction F is approximately 0.65 (see, for instance, Trischler et al., 2014).

The equation to compute the seasonal AGBP for a given pixel thus becomes:

$$AGBP_s = \sum_{i=SOS}^{EOS} N(\mathrm{i}) * \mathrm{DMP}(\mathrm{i}) * AOT \qquad (39)$$

Where:

- DMP(i) is the Dry Matter Production at dekad *i*, expressed in kgDM/ha/day.
- *N*(i) is the number of days within each dekad, varying between 8 (end February) and 11.
- The first term, $\Sigma N_d(i)$, is needed to obtain the AGBP sum in terms of kgDM/ha. Without it, one would obtain the mean.
- *AOT* accounts for the fractioning between the above and below-ground biomass (i.e. the AGBP Over Total (AOT) biomass production correction factor, where the default is AOT=0.65).

Table 8: Overview of AGBP data component

Data component	Unit	Range	Use	Temporal resolution

[16] Where 1 gC/m²/day (NPP) = 22.222 kgDM/ha/day (DMP), see Section 2.17.

AGBP	kgDM/ha	0-20,000 for seasonal	Above-ground dry matter produced. It can be used to derive yields if information on phenology and harvest index is available.	Seasonal

2.1.6. Land Cover Classifications

Description

Land cover can be defined as the observed (bio)-physical cover on the earth's surface, encompassing vegetation, bare rock and soil as well as human-made features. Land use, on the other hand, can be derived from the land cover, combined or linked with the activities or actions of people in their environment (Di Gregorio, 2005). WaPOR land cover mapping focuses on agricultural land cover and distinguishes between irrigated and rain fed cropland at Level 1 and 2, and introduces crop information at Level 3. Data on agricultural land cover are important for the evaluation of current land use practices as it can be coupled with water productivity data, enabling the comparison between different farming systems within a region, or across different regions.

For Level 1 and 2, the annual LC maps denote the broad LC classes as shown in Table 9. Important to note is that prior to 2015, most input data at level 2 is derived from resampled level 1 data, as no 100m satellite data was available. The land cover for the period 2009-2014 is therefore resampled from Level 1, and thus only included the land cover classes of Level 1, as depicted in Table 9.

The classes in Table 9 are compatible with the Land Cover Classification System (LCCS) that was developed by FAO and UNEP (Di Gregorio, 2005). This ensures that the land cover data created at all resolution levels is standardised, making it compatible with and easily compared, correlated and harmonized with other land cover data using this system. For Level 1 and 2 the classification efforts were streamlined with the Copernicus[17] Land service activities.

The result of a land cover classification can be evaluated in several ways, where the use of confusion matrix is commonly applied. However, the development of methods for the accuracy assessment of products derived from moderate to low spatial resolution data is still being researched (Foody, 2002). Landscape characteristics such as land cover heterogeneity and patch size impact on classification accuracy at coarser resolutions, with the probability of a correct classification decreasing with decreasing patch size and increasing heterogeneity (Smith et al, 2003). The land cover classifications are independently validated and calibrated where necessary (see Reports on Validation results).

Table 9: Overview of land cover classes per Level. For Levels 1 and 2 the classes used in the annual maps are shown.

Level 1 (annual)	Level 2 (annual)	Level 3
Cropland rainfed	Cropland rainfed	Maize
		Rice
		Wheat
		Crop (covering more than 10% of the area)

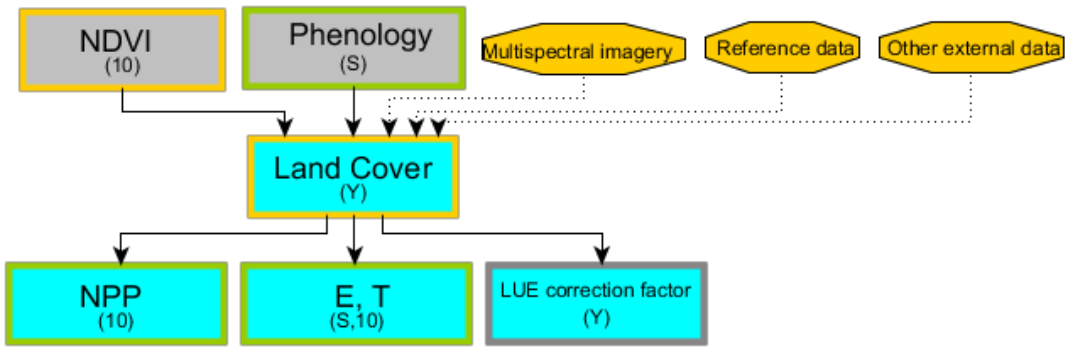
[17] Framework contract (199494) for the operation, evaluation and evolution of the Global Land component of the Copernicus land service, Lot1: operation of the Global Land component, thematic domain vegetation and energy (Link: http://land.copernicus.eu/global).

Cropland irrigated	Cropland irrigated	Maize
		Rice
		Wheat
		Crop (covering more than 10% of the area)
Natural vegetation	Tree cover[1]	Tree cover
	Shrubland[1]	Shrubland
	Grassland[1]	Grassland
	Wetland[1]	Wetland
Artificial	Artificial[1]	Artificial
Bare soil	Bare soil[1]	Bare soil
Water body	Permanent[1]	Permanent
	Seasonal[1]	Seasonal

[1]Detailed land cover classes at Level 2 that are only available on the land cover maps for 2015 onwards. Level 2 land cover maps prior to 2015 are derived from resampled level 1 data, as no 100m satellite data was available.

Methodology

Box 8: Land Cover classification in relation to other data components.



- ✓ Land Cover Classification makes use of dekadal NDVI time series and seasonal phenology information.
- ✓ External data is required in the form of multispectral satellite imagery.
- ✓ Other external data includes vegetation indices such as EVI.
- ✓ Classifying land cover requires a substantial amount of reference data. This static input is provided in the framework of the collaboration with the Copernicus Land Service activities..
- ✓ The seasonal LCC output is used to determine the light use efficiency (LUE) correction factor for use with NPP and AGBP.
- ✓ Additionally, Precipitation and ET are use to distinguish croplands that is irrigated from rainfed cropland.

The production of the land cover classification data component requires input from the Phenology data component to demarcate the growing seasons, and the dekadal NDVI data as well as the original reflectance data on which the NDVI is based. External reference data are an important component of land cover classification. For the irrigation labelling of the croplands, additional information on the precipitation as well as the ET is used.

Although the methodology may vary slightly between the different levels, the general workflow is shown in Figure 11. A supervised classification is applied to assign a specific class to each pixel of the image. Training data consists of dekadal NDVI and reference data denoting the exact location of each of the classes specified in Table 9. The different components of the classification processing chain are discussed in the sections below. The discussions are general to provide an understanding of the methodology applied. Technical details are provided in the Data Manual.

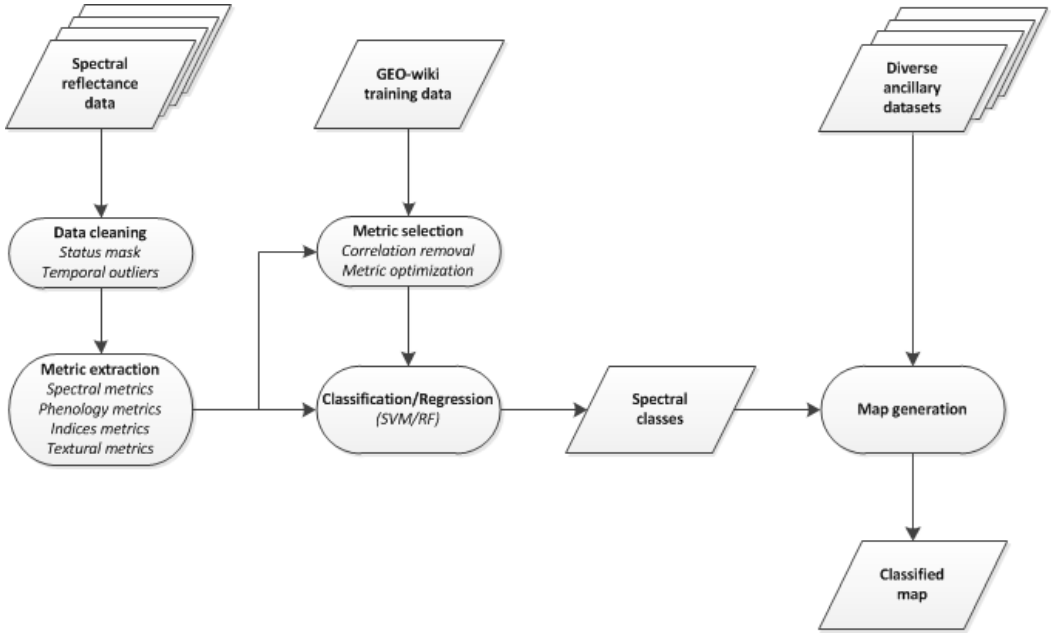


Figure 11: Schematic overview of the land cover classification processing chain. Different types of reference data as well as dekadal NDVI and multispectral remote sensing inputs are used to train a machine learning classifier. The input data vary across the different Levels. Source: this study.

Reference data

A key component for the production of accurate land cover classifications is a sufficient amount of high quality reference data encompassing all the required classes at the various levels for at least one moment in time and distributed relatively evenly. Since the land cover classifications are delivered annually on a seasonal basis, a huge amount of reference data is required. The gathering of suitable reference data is therefore one of the main challenges for the production of the Land cover classification data component.

The accuracy of land cover mapping products strongly relies on the quality, quantity and accuracy of the reference data available. It should be noted that an over or under representation of a class and differences in sampling density between different classes within a reference dataset can greatly influence the classification outcome. For example, a relatively large amount of training points on forest cover is likely to result in an over-classification of forest cover.

The generation, by means of fieldwork, of a reference dataset that is suitable for the extent of the Level 1 products requires significant efforts and was therefore not feasible within the framework of this project. Several additional external sources were used to collect as much as possible good quality reference data suitable for use at the various levels. Some sources of reference data were applicable across all levels whilst others were level-specific.

Reference data suitable for use mainly at Levels 1 and 2 was obtained through the C-GLOPS initiative. Figure 12 shows the reference data points that match the Level 1 and 2 classes.

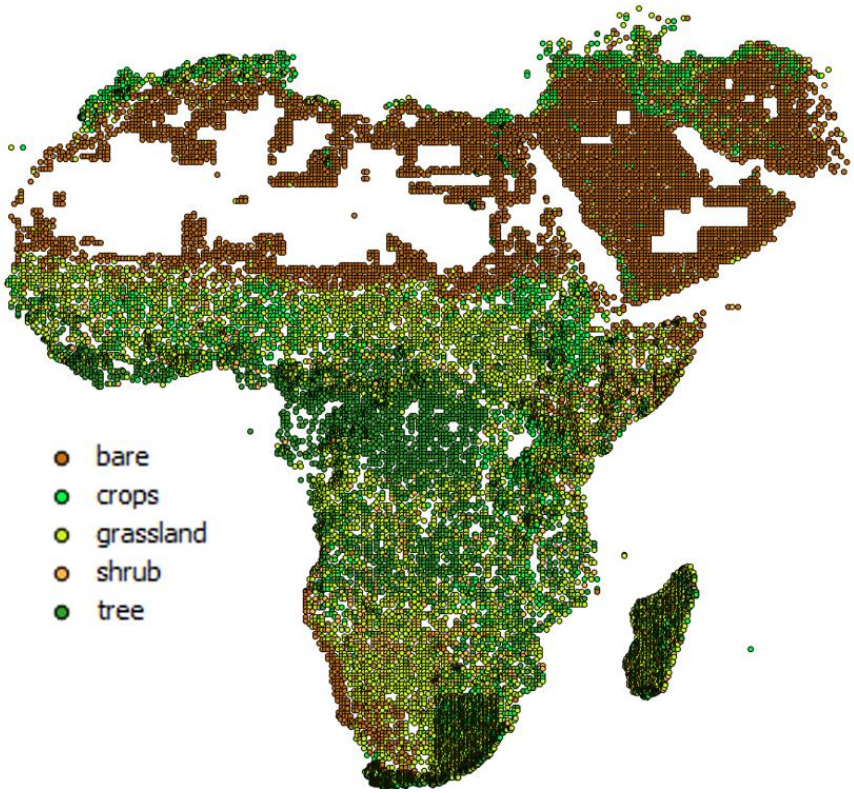


Figure 12: Map showing the reference data points obtained from C-GLOPS initiative applied at Level 2. The points are denoted per land cover class, but for clarity of the figure no distinction is made between irrigated and rain fed croplands. Source: C-GLOPS

Classification metrics

In addition to reference data depicting exact locations of the different classes, a classifier needs input variables which can aid in the differentiation between the different land cover types. These metrics are typically descriptors of the spectral behaviour of the different classes through time, exploiting the differences in phenology. The metrics describe the temporal behaviour of the individual spectral bands, a selection of vegetation indices and phenological descriptors. For these variables, descriptive statistics are extracted for the reference year as well as for the vegetation season and off-season within that reference year using phenological parameters (start- and end of season). The Data manual contains details on the statistical descriptors used at each level.

Classifier

A wide variety of classification algorithms have been used to map land cover from remotely sensed data. In the early stages of remote sensing, unsupervised classification and cluster labelling was the common method for large area land cover mapping (see Wulder et al., 2004). However, machine learning (ML) algorithms have since proven to be more accurate and efficient alternatives to conventional parametric algorithms[18] when faced with large data volumes and complex feature spaces. Many of the current global land cover maps have been produced with ML, such as Globeland30, GlobCover, CCI. The classifier applied in this project is therefore a machine learning algorithm, i.e. Random Forest and Random Forest Regression. Technical details are provided in the

[18] For example Maximum Likelihood

Data manual. It is important to note that the application of the classifier on a continental or regional level requires an approach adapted to the conditions. Throughout Africa, for example, large differences occur in the physical conditions that are reflected in the phenology. Training classifiers on a continental level will disregard these temporal-spatial variations within the land cover class. By using a local classifier, these differences could be accounted for, thereby increasing its accuracy. At Level 1 and 2, the classifiers are trained and applied per ecozone[19]. The classifier are not trained to generate information on the Urban and Water classes. Separate workflows are developed for these classes, as much better products are available which are not solely based on spectral satellites. For example, global urban datasets exist, most of which are produced with radar data. Water on the other hand, can much better be detected when a digital terrain model is included. Information on the slopes can provide very valuable information on where water can potentially be present, significantly decreasing the number of falsely classified water pixels. Further details are provided in the Data manual.

In addition to a per-year classification, a temporal consistency check is performed over the LC products over the years. Variations in intermediate products and final land cover maps over the years may be due to real changes on the ground, as well as noise in the reflectance data. To account for random fluctuation due to noise, a temporal smoothing is performed on continuous intermediate land cover products, such as the crop cover layer (depicting the percentage of cropland cover per pixel, which is used to determine the cropland areas). More information on these intermediates as well as the temporal smoothing can be found in the Data manual.

For example, the crop cover layer, depicting the percentage coverage of cropland areas within one pixel, is used as an input to determine the cropland areas.

Complementary data layer: Light Use Efficiency (LUE) Correction factor

This additional raster layer is delivered to the WaPOR database to enable users to recalculate NPP and AGBP when the correct land cover for a specific year is known. This layer is produced for those years where the final land cover was not known at the time of data production. In that case, the last known land cover was used. Once the correct land cover is known, a Light Use Efficiency (LUE) correction factor is derived for all vegetated areas, to account for changes in LUE where the land cover changed. For level 1, the LUE corrections will be provided for 2016 onwards. The LUE correction factor is the ratio between the actual LUE and the LUE applied previously. When the land cover classification is known and the LUE correction factor is calculated, the user can multiply the available NPP or AGBP data with the LUE correction factor. The correction factor is 1 when the actual LUE is equal to the applied LUE. In other cases it can either be higher or lower.

Complementary data layer: Land Cover Classification Quality layer

This additional raster layer is delivered through the WaPOR database to inform users about the quality of the land cover classification. A combination of factors influences the accuracy of the classification across a land cover classification map, such as amount of cloud cover, similarity of the land cover classes, amount of training data and complexity of the landscape. All land cover maps contain a fraction of falsely classified pixels, and this fraction will differ substantially between regions as a result of the abovementioned influencing factors. For example, classifications in central Africa are typically less reliable due to the very high cloud cover. As a per-pixel validation is not feasible, these regional differences will be quantified through an evaluation of how well the original training points have been classified. For each pixel in the land cover map, all training points in a predefined window around this pixel will be evaluated, providing for that pixel an indication of how well the classification has been

[19] http://www.fao.org/nr/gaez/en/

performed in the surrounding area. Although it is not a direct measure of the reliability of the classification of that pixel, it will provide an indication of how well the classification was able to separate the different classes in that area. A quality value close to 100 represents more certainty regarding the classification, whilst pixel values close to 0 indicate pixels for which the classification is less accurate.

Table 10: Overview of Land Cover data component

Data component	Unit	Range	Use	Temporal resolution
Land Cover Classification	-	-	Qualitative maps that show land cover according to the land cover classification scheme shown in Table 9.	Seasonal
LUE correction factor	-	-[1]	Used to recalculate NPP (and AGBP) at the end of the season when the correct land cover for the season is known.	Seasonal
Land Cover Classification Quality	%	0-100	Quality of the classification in a region, to provide an indication of the reliability of the land cover product in the area of interest.	

[1]A value of 1 indicates that the actual LUE is equal to the default LUE.
[2]A value of 1 indicates that the actual AOT ratio is equal to the default AOT ratio.

2.2. Intermediate data components

2.2.1. NDVI

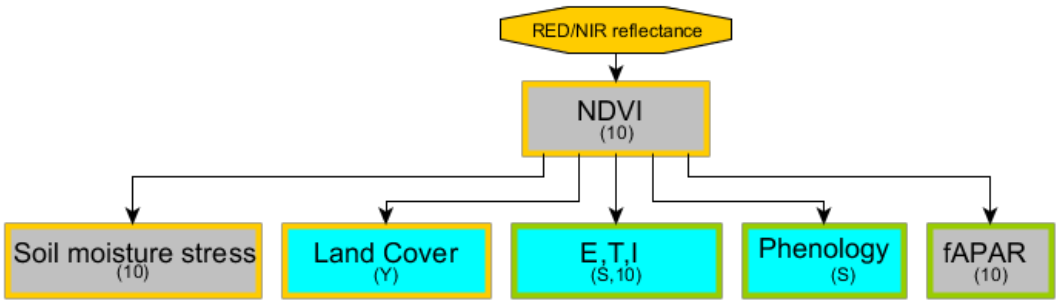
Description

The Normalized Difference Vegetation Index (NDVI) correlates well with photosynthetically active vegetation and is therefore a measure of the greenness of the earth's surface. Since it only requires a red and NIR band, the NDVI is a commonly used vegetation index that can easily be derived using most multispectral sensors. Dekadal NDVI composites are produced and used internally as input for the computation of various data components, such as fAPAR E, T and I. NDVI values range between -1 and 1, where vegetated areas have positive values closer to 1, bare soil/artificial surfaces have values of around 0, and water has negative NDVI values.

One of the main challenges when producing NDVI time series is the high cloud cover that occurs over certain areas. NDVI composites are produced to fill gaps and missing data that occur in the input satellite imagery. When an insufficient number of data observations are available within a composite period, the results of smoothing and gap filling are less accurate. Data layers that indicate the quality of each of the dekadal NDVI data composites are produced (see description of the methodology below). For an example of NDVI map, see Methodology document for Level 1.

Methodology

Box 9: NDVI in relation to other data components.



- ✓ Red and NIR reflectances are required to calculate NDVI.
- ✓ The output is used in various data components, directly and indirectly.

NDVI is based on the spectral reflectance of the red and near-infrared wavelengths of multispectral satellite data. It is calculated as follows:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

The following steps are followed at all three levels to produce dekadal NDVI composites and a concomitant quality layer:

1. Composites are made to reduce gaps due to clouds and other missing data
2. Leftover gaps and anomalies (unreliable values) filled by smoothing

3. Information from the results of steps 1 and 2 is combined to produce a data layer that indicates the quality of the NDVI input data.

Frequent satellite-based reflectances are converted to dekadal NDVI composites through the following procedures. First pixels that cannot be used for NDVI calculations are flagged as water, sea, cloud, and error pixels. Then a "tile-based" dekadal synthesis is produced using a constrained[20] Max-NDVI compositing rule so that the dekadal NDVI comprises the "best" observation extracted from the available scenes within the dekad.

The viewing angle has an important effect on the NDVI in that increasing view zenith angles tend to result in higher NDVI values. As the dekadal composites are produced using the max-NDVI criterion, the compositing step is more likely to select pixels with a high viewing zenith angle. As shown in Figure 13, this results in artefacts. To minimize this effect, a maximum viewing zenith angle is imposed in the compositing step. However, this also reduces the number of available observations within a dekad, resulting in more no-data pixels.

The pixels with missing and/or unreliable values in the dekadal NDVI series are then replaced by more plausible data through a process of interpolation based on the methodology[21] explained in Swets et al. (1999). The resulting images have no data gaps, see Figure 14 for an example.

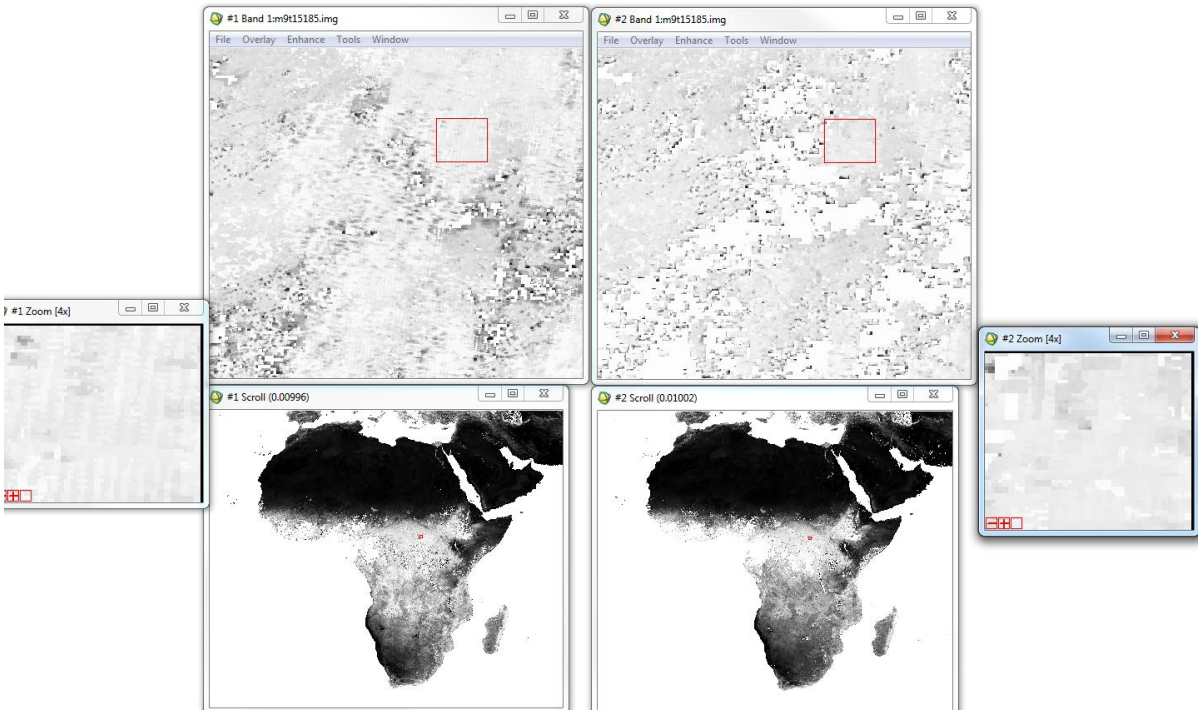


Figure 13: Examples of NDVI composite without (left) and with (right) view zenith angle constraint. Three different zoom-levels are shown for the same area. As can be seen in the image on the right, the angle constraint decreases the occurrence of artefacts, but increases the number of pixels without valid observation.

[20] "Constrained" means that previously flagged observations are not included in the selection.

[21] This methodology was also applied in the EU-MARS and FAO-ASIS projects. To this end VITO has developed dedicated programs (GLIMPSE, SPIRITS) which analyse a time series of dekadal composites of any vegetation index to detect unreliable observations (mostly local minima) and replace them by means of interpolation so that the resulting images have no data gaps.

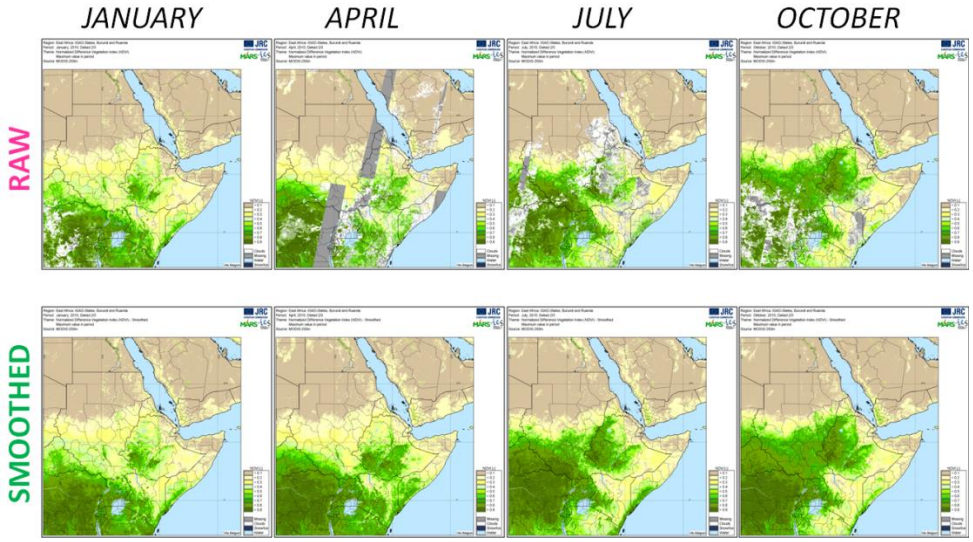


Figure 14: Example of original and smoothed NDVI for four dekads in 2010 from MODIS-250m over the Horn of Africa. Smoothing replaces all clouds and missing values with appropriate values.

Complementary data layer: NDVI Quality layer

The quality layer is produced during the smoothing of the NDVI. The quality index (QI) for every pixel in each dekad depicts if and how a new value was created for that pixel. The procedure can be described as follows:

1) First all the valid observations (not flagged) are treated. If the final estimate of the smoothed NDVI is very close to the (pre-cleaned) NDVI value, the QI is set to 0 (ideal situation). The resemblance is dictated by the user-specified tolerance. Where a valid observation was present but it was adapted by the smoothing, the QI is set to 250.
2) For the remaining (flagged) observations, the QI is set to the number of days in-between the surrounding valid observations (i.e. with QI=0 or 250). If the length of the data gap exceeds 240 days, it is saturated to 240. The fundamental idea is that the longer the gap, the less reliable the smoothing is. For the observations at the profile edges, it is assumed that the (a priori unknown) observations preceding the first valid observation and following the last valid observations are “good”.

This quality layer depicts the quality for the NDVI, fAPAR, albedo and NPP, as all these data components rely on the same input, i.e. the spectral reflectance data. Furthermore, the length of the data gap is the same. Figure 15 shows an example of the NDVI quality data layer at Level 2.

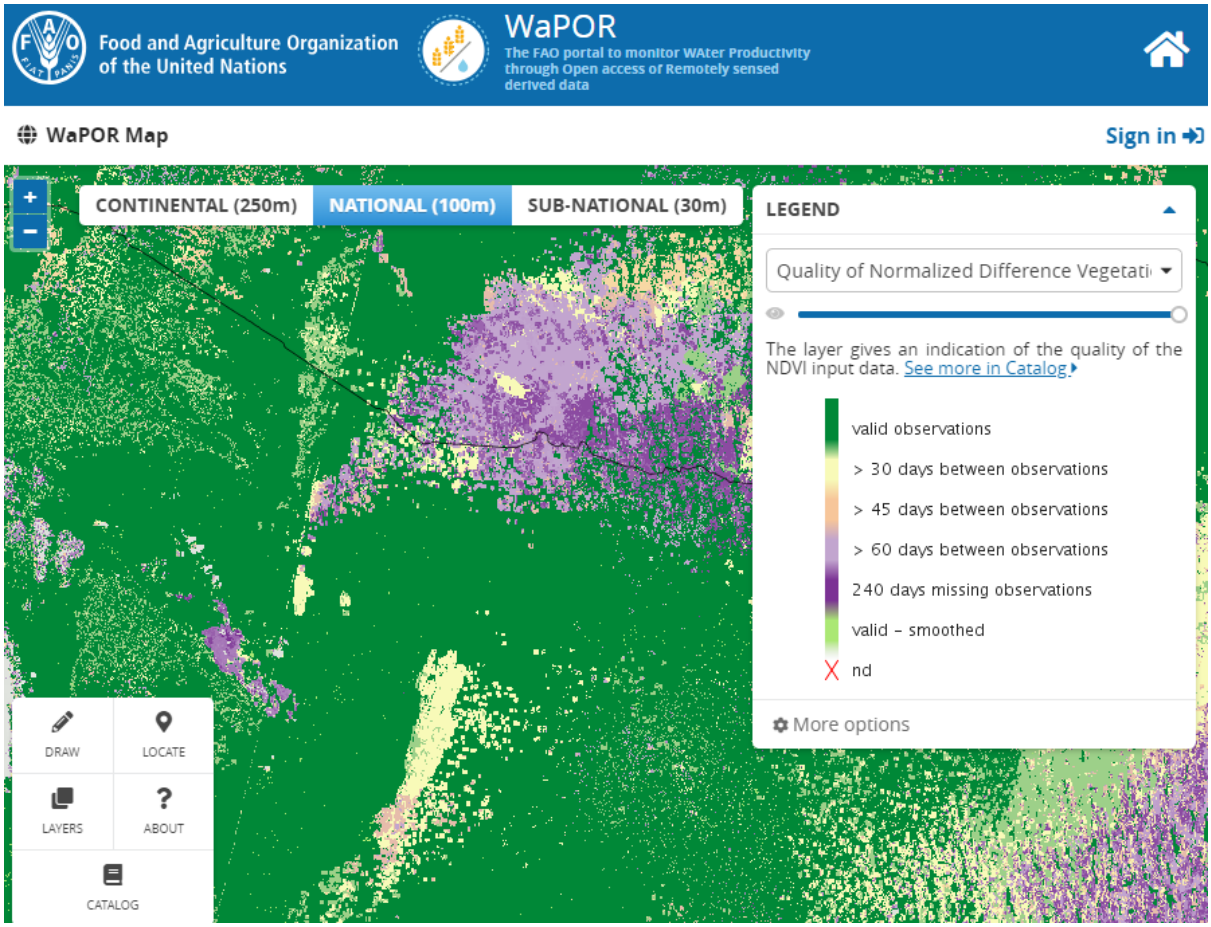


Figure 15: Example of the NDVI Quality Layer at Level 2.
Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor

Table 11: Overview of NDVI intermediate data component and complementary quality layer

Data component	Unit	Range	Use	Temporal resolution
NDVI	-	-1 to 1	Measure of greenness of vegetation.	Dekadal
NDVI Quality layer	days		Indicates quality of NDVI composite.	Dekadal

2.2.2. Solar radiation

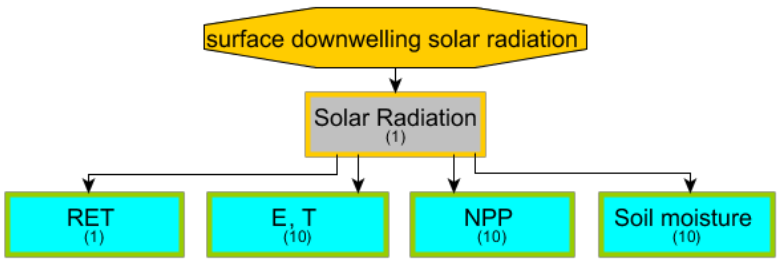
Description

The availability of solar energy is the main driver for evapotranspiration and biomass production. Unless water availability is limited, places that receive more solar radiation (through latitudinal location, sun angle and/or number of sunny days) are likely to have higher crop yields. Atmospheric conditions determine how much of the solar radiation that reaches the top of the earth's atmosphere reaches the land surface[22].

This intermediate data component calculates the amount of solar radiation (expressed in $Wm^{-2}d^{-1}$) that reaches the land surface of a specific location on a specific day, based on the combined effect of location, date, local topography and atmospheric conditions. It is delivered on a daily basis for all three levels. Solar radiation values typically range from around 50 (when transmissivity is very low) to around 300 $Wm^{-2}d^{-1}$. In addition to the daily solar radiation, another data component, the instantaneous solar radiation, is calculated separately. This data component calculates the amount of solar radiation (in Wm^{-2}) at time of satellite overpass and is used as input to compute the soil moisture.

Methodology

Box 10: Solar Radiation in relation to other data components.



- ✓ Surface downwelling solar radiation is required to calculate Solar Radiation.
- ✓ A DEM is used to calculate the solar zenith angle to the land surface.
- ✓ Solar Radiation is used for calculating SMC, E, T, RET and NPP..

The amount of solar radiation that reaches the land surface is determined by a combination of factors. Latitudinal position, day of the year and local topography[23] all determine the incidence angle of the

[22] Also referred to as Top of Canopy (TOC).

[23] For example, in the northern hemisphere, south facing slopes are warmer than north facing slopes.

sun at a specific location. Topographical features such as slope and aspect can be extracted from a digital elevation model (DEM) are used to calculate the solar zenith angle to the surface. All these factors are combined to calculate the potential solar radiation for any location on the land surface at a given day.

However, not all the potential solar radiation reaches the land surface. To determine the actual solar radiation reaching the earth's surface, the potential solar radiation is adjusted for atmospheric transmissivity, a measure of the amount of solar radiation that is propagated through the atmosphere. The transmissivity is derived from surface downwelling solar (sds) radiation measurement which are regularly made during the day by geostationary meteorological satellites. Atmospheric transmissivity can be calculated by comparing the calculated solar radiation at the top of atmosphere with the measured sds radiation.

The atmosphere causes the scatter of a part of the incoming solar radiation. This effect increases as the transmissivity decreases. Under clear atmospheric conditions most of the solar radiation reaches the surface directly, as can be seen by the sharp shade of sunlit objects. Under hazy or cloudy conditions, shades are less sharply delineated as the scattering of solar radiation cause the radiation to come in from different directions. This effect has to be taken into account: the total available solar radiation that reaches the land surface is the sum of the direct and indirect (diffuse) solar radiation. Both are calculated with the transmissivity determining the ratio between them. A diffusion index is calculated which is provided as a function of the transmissivity. The diffusion index is 1 when transmissivity is low, indicating that no direct solar radiation is available, the diffusion index is 0 when transmissivity is high, indicating that no diffuse solar radiation is available. The next step involves the calculation of the solar radiation during different moments of the day. This requires complicated geometry mathematics, particularly for slopes. More detail on this part of the methodology can be found in Allen et al. (2006b).

Although the transmissivity and DEM input data are the same resolution (approximately 5 km and 90 m respectively) at all three levels, solar radiation is calculated separately for all three levels as the inputs are resampled for each level.

The method to produce the instantaneous solar radiation (used as input in the soil moisture processing chain) is also applied at all three levels but differs from the one of the daily solar radiation described above. It is based on the implementation of the Solar Radiation Model *r.sun* whose detailed equations can be found in Suri and Hofierka (2004).

Table 12: Overview of Solar Radiation data component

Data component	Unit	Range	Use	Temporal resolution
Solar radiation	$Wm^{-2}d^{-1}$	50-300[1]	Estimates daily solar radiation that reaches land surface at a specific location, used to calculate RET, E, T and NPP.	Daily
	Wm^{-2}	0-1000	Estimates solar radiation that reaches land surface at a specific location and specific date and time, used to calculate SMC.	Instantaneous

[1] These values are typical low and high values and do not indicate maximum and minimum values.

2.2.3. Soil moisture stress

Description

Soil moisture availability is one of the most important parameters governing biomass production and evapotranspiration. Lack of soil moisture can seriously hamper biomass growth by reducing vegetation transpiration. Soil moisture is directly released to the atmosphere from the top soil through evaporation and from the vegetation cover through transpiration.

Evaporation reduces as vegetation cover increases. Soils fully covered by vegetation experience very little evaporation as nearly all of the available energy is captured by the vegetation cover and used for transpiration. Transpiration drives the transport of soil moisture from the sub soil through plant roots. The root zone may hold more water and enables the plant to continue with transpiration even when the top soil is dry.

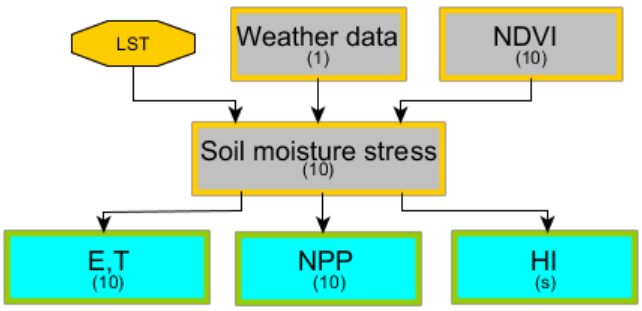
Relative soil moisture content and stress is produced at all three levels at a dekadal temporal resolution. These are intermediate data components that are used as input to other data components and are not published through WaPOR.

Soil moisture content varies strongly in time and place. Within the WaPOR area of interest extremes occur in northern Africa and the Middle East where soil moisture content is very low throughout the year (with the exception of areas close to rivers) and the equatorial region which is characterised by high soil moisture content throughout the year. Other areas generally show more seasonal variation in soil moisture content. Pixel values of relative soil moisture content range between 0 and 1, where 0 is equal to the soil moisture content at wilting point and 1 is equal to the soil moisture content at field capacity. Soil moisture stress values also range between 0 and 1, where 0 means maximum stress and 1 means no stress.

Data layers that indicate the quality of the input data used to produce each of the dekadal Soil moisture stress data composites are produced for Level 2. For an example of a Soil moisture map, see Methodology document for Level 1.

Methodology

Box 11: Soil moisture stress in relation to other data components.



- ✓ Calculating Soil Moisture Stress requires Weather data input as well as NDVI intermediate data components.
- ✓ Land Surface Temperature (LST) is required as external data source.
- ✓ Soil moisture stress is used as input to calculate E and T.
- ✓ Soil moisture stress is incorporated in the calculation of NPP.

The methodology applied for calculating relative soil moisture content and soil moisture stress is based on the correlation between Land Surface Temperature (LST, derived from thermal infrared imagery), vegetation cover (derived from the NDVI) and soil moisture content. This is also known as the triangle method[24] (Carlson, 2007). External input data required are visual/NIR and thermal imagery.

The triangle method is named after the shape of the scatter plot that emerges when all pixels in an image are plotted with NDVI on one axis and temperature on the other axis. Discarding outliers, a triangle shape appears, delineated by two marked boundaries (see Figure 16). These boundaries represent two physical conditions of water availability at the land surface, called the cold edge and the warm edge. At the cold edge, water is readily available and the soil moisture content is at field capacity. Evapotranspiration takes place at maximum rate, with the latent heat flux at its maximum and the sensible heat flux at zero. In this situation, the LST is close to the ambient air temperature. At the warm edge no soil moisture is available and evapotranspiration and the latent heat flux are equal to zero.

Incoming radiation increases LST. This increase depends on the vegetation cover (NDVI). The LST increase is highest when no vegetation is present and smallest when vegetation fully covers the land surface. Therefore, the difference between the cold and the warm edge is largest for bare soil and smallest for fully vegetated surfaces. In general, LST is lower when the soil moisture content and/or the vegetation cover are higher.

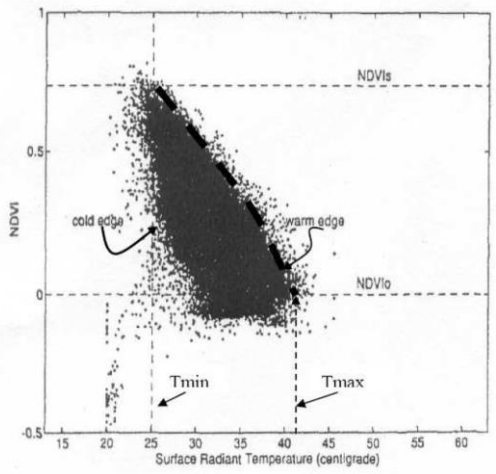


Figure 16: An example of a scatter plot of NDVI versus surface radiant temperature taken from Carlson (2007). The cold edge on the left side and the warm edge on the right side of the point cloud are clearly distinguishable.

A drawback of this method is that it requires calibration by manual selection of references pixels for each thermal image. This introduces subjectivity through the selection process and makes it difficult to operationalize for a larger area. This problem was overcome by the method developed by Yang et al. (2015). The original triangle method was modified by introducing the effect of stomatal closure of

[24] An alternative approach is based on the use of radar imagery from ASCAT. WaPOR data production partners apply the LST method as it has a higher resolution and therefore provides a better representation of the spatial variability of soil moisture content. It is also a better indicator for the water content in the root zone in the sub-soil than radar methods which are only able to observe soil moisture content in the top layer of the soil. The moisture content of these two soil layers is not necessarily correlated. The results based on radar also tend to be less accurate for areas with moderate to dense vegetation cover. eLEAF (leading partner of FRAME Consortium) has applied the LST method with good results in South Africa, Russia and Ukraine.

vegetation under dry condition as a result of water stress (Moran et al., 1994). As a result, the temperature of the warm edge at a fully vegetated surface becomes higher than under wet conditions. This results in a trapezoid shape as depicted in Figure 17, taken from the improved trapezoid method[25] of Yang et al. (2015).

The trapezoid, corners numbered A, B, C, D, are defined by the linear relationship between LST and vegetation cover under the two extreme conditions of the cold edge and the warm edge. The top line segment (A – B) shows this relationship under completely dry conditions (no available soil moisture). Point A represents bare soil. Point B represents full vegetation cover. The bottom line segment (D – C) represents soil moisture at field capacity. Again, on the left side (D) for bare soil and on the right side (C) for full vegetation cover. This linear relationship between LST and vegetation cover (under equal soil moisture conditions) is not only true for the extreme conditions but for each value of the soil moisture content, as shown by the soil wetness isolines in Figure 17.

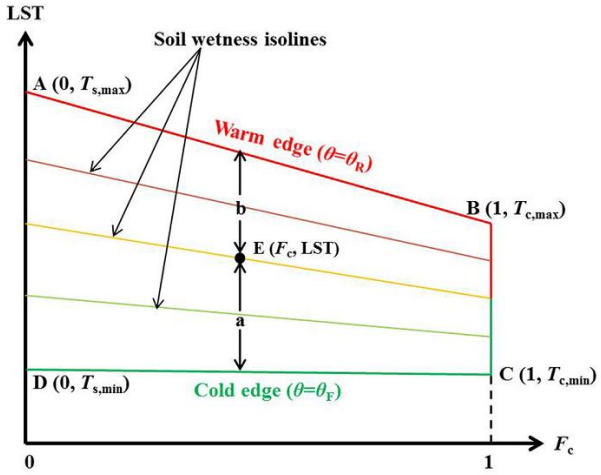


Figure 17: The trapezoidal vegetation coverage (F_c) / land surface temperature (LST) space (transposed axis). Points A, B, C and D are estimated for each separate pixel using modified Penman/Monteith equations.
Source: Yang et al., 2015.

The relative soil moisture content of a specific location (e.g. point E) can be derived from its relative distance to the cold edge (a) and warm edge (b) using:

$$S_e = \frac{b}{a+b} \quad (40)$$

Where:

$$a = LST - T_{min} \quad (41)$$

$$b = (1 - F_c)(T_{s,max} - T_{c,\max}) + T_{c,max} - LST \quad (42)$$

Solving these equations in order to derive the relative soil moisture content first requires calculation of the four corner points of the trapezoid (A – D) as well as information on vegetation cover and LST of point E. The NDVI intermediate data component is used to derive vegetation cover whilst LST is derived from thermal satellite imagery.

[25] Yang et al. (2015) report that their method is able to reproduce spatial and temporal patterns of observed surface soil moisture with an RMSE of 0.06 $m^3 \cdot m^{-3}$ at the field scale and 0.03 $m^3 \cdot m^{-3}$ at the regional scale. The approach has not been tested on a continental scale.

Assuming no sensible heat flux, the temperature of the cold edge (C and D) is approximated by the wet-bulb temperature ($\boldsymbol{T_{wet}}$) at around the same time as when the LST is measured. The wet bulb temperature is defined as the minimum temperature which may be achieved by bringing an air parcel to saturation by evaporation in adiabatic conditions (Monteith & Unsworth, 2013). Thus the cold edge conditions are considered to be such that there is enough soil moisture and a sufficient evaporation rate to reach saturation of the cooling air and therefore for the temperature to approach $\boldsymbol{T_{wet}}$. Compared to the cold edge, calculating the corner points A and B of the warm edge requires more effort. This is done with the Penman-Monteith equation rewritten to yield T_{max} at point A and B. We provide an overview of the steps below, more detail can be found in Yang et al. (2015).

At the warm edge, a large part of the incoming radiation is used for heating the land surface, thus increasing LST. The amount of energy available depends on the incoming solar radiation ($\boldsymbol{R_s}$) and net long wave radiation (L^*). The surface albedo (a) is an important factor in determining how much of this energy is retained to heat the land surface. This requires the deduction of two theoretical albedo values, one for bare soil (point A) and one for full vegetation cover (point B). Soils generally have a higher albedo, reflecting more of the incoming radiation than vegetated cover. Theoretical values can be derived from the land cover class and soil type maps. Here it is derived from the surface albedo intermediate data component.

Part of the warming of the land surface is lost again through the sensible heat flux (H). The sensible heat flux depends on the aerodynamic resistance to heat transfer determined by soil and canopy characteristics. Bare soils have a higher resistance than vegetation due to the lower surface roughness, resulting in a lower sensible heat flux. Surface roughness is derived from the land cover class. The method to calculate the aerodynamic resistance is based on Sanchez et al. (2008).

For bare soil, the soil heat flux (G) also has to be included, assuming a fixed fraction of the net radiation of 0.35. Soil heat flux does not need to be included for a fully vegetated surface as the soil surface is not directly heated by incoming radiation.

This method is applied on a pixel-by-pixel basis with no spatial dependencies, making it possible to apply the same methodology for different regions in a consistent manner. However, parameterising the soil moisture algorithm on a continental scale is challenging, particularly for the Level 1 area of interest where soil moisture content, vegetation cover and weather conditions vary greatly (e.g. the dry Saharan desert and the wet tropical rainforests present extreme opposites). A specific challenge lies in the determination of the reference values for the corner points of the warm edge. Calculation of these hypothetical values depends on a number of assumptions under extreme conditions which can be challenging to estimate. The surface albedo intermediate data component is used to provide the minimum and maximum surface albedo which is input to the Yang algorithm. The surface albedo for point A (high surface albedo) and point B (low surface albedo) have been determined with the use of the albedo time series for each pixel, obtained from the albedo intermediate data component. By using these values instead of constant values, it is ensured that the theoretical maximum LST is being derived using realistic surface albedo values.

The soil moisture content is determined for both the top soil and the root zone. Therefore the same soil moisture content is used for the determination of evaporation and transpiration, albeit in different formulations. Soil moisture stress limits transpiration by means of the canopy resistance. For evaporation the soil moisture content is used to model the soil resistance. The vegetation cover determines the route of the water flow, i.e. through transpiration or evaporation.

By using the soil moisture content the model is able to separate between evaporation and transpiration. Some studies use the triangle/trapezoid method to calculate the evaporative fraction

directly, but then it is not possible to make the distinction between transpiration and evaporation. Hence the need for the ETLook model.

Soil moisture stress

The soil moisture content determines the availability of water for evaporation and transpiration. Whether this is reduced due to a shortage can be calculate with a stress factor. This stress factor for transpiration (S_m) can be derived using the following relationship as defined in American Society of Civil Engineers (ASCE, 1996):

$$S_m = K_{sf} S_e - \frac{\sin(2\pi S_e)}{2\pi} \quad (43)$$

The tenacity factor K_{sf} ranges from 1 for drought-sensitive plants to 3 for drought-insensitive (tenacious) plants. A default value of 1.5 is chosen when no crop information is available.

This soil moisture stress factor, ranging between 1 and 0, is used as input for the E and T to reduce evaporation and transpiration.

Complementary data layer: Land Surface Temperature Quality layer

Cloud cover causes data gaps in the input data required for the calculation of soil moisture content and soil moisture stress. Daily soil moisture is determined from daily LST images with cloud covered parts masked out. These daily images are then composited into dekadal data, taking into account the quality of the input LST layer (i.e. viewing angle and proximity to clouds). The soil moisture stress quality layer indicates the number of days since the last observation, given on a pixel-by-pixel basis. For an example of a Land Surface Temperature Quality map, see Methodology document for Level 1.

Table 13: Overview of the (intermediate) data components related to Soil Moisture

Data component	Unit	Range	Use	Temporal resolution
Soil Moisture Content	-	0-1	Used to calculate E and T	Dekadal
Soil Moisture Stress	-	0-1	Used to adjust NPP for the effect of soil moisture stress.	Dekadal
Land Surface Temperature Quality layer	Days	1-365	Indicates the quality of the Soil Moisture Stress intermediate data component which is used as an input to produce NPP, E and T	Dekadal

2.2.4. fAPAR and Albedo

Description

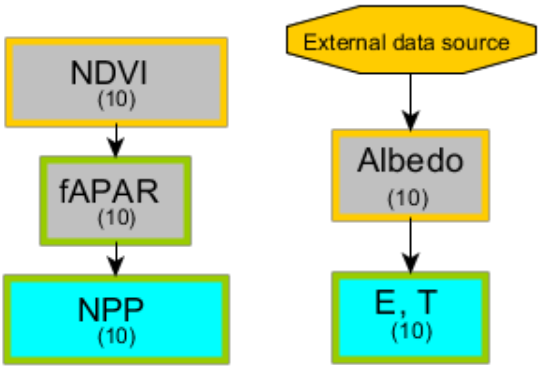
fAPAR and albedo both play an important role in the radiative energy balance of ecosystems and in the estimation of the carbon balance. fAPAR is the fraction of photosynthetically active radiation (400-700nm) that is absorbed by the vegetation canopy (when only absorption by live leaves is taken into account, it is referred to as 'green' fAPAR). Albedo from the land surface is the ratio of the radiant flux over the shortwave spectrum (approximately 200-3000nm) reflected from the earth's surface to the incident flux. Similar to the different definitions of the "spectral reflectance" (BRDF, R-factor, hemispherical reflectance), the integrated albedo also comes in different versions, but for this project it suffices to find the hemispherical albedo.

Both these intermediate data components are produced at all three levels with a dekadal temporal resolution. They are not published through WaPOR, but are used as input for the calculation of NPP

(fAPAR) as well as E and T (albedo). fAPAR values range from 0 to 1. Surface albedo varies in space and time as a result of processes such as changes in solar position, snowfall and changes in vegetation cover. A typical range for albedo of land areas is 0.1 to 0.4.

Methodology

Box 12: fAPAR and albedo in relation to other data components.



- ✓ External data sources are used as input.
- ✓ fAPAR is used as input to various data components, e.g. NPP, E, T and intermediate data components such as soil moisture.
- ✓ Surface albedo is used as input to produce E and T

fAPAR

fAPAR at Level 1 and 2 is estimated by using a direct relationship between the NDVI and a global fAPAR product. The fAPAR for Levels 1 and 2 is derived using the same method to ensure consistency between the levels. Further details of the processing are given in the Data Manual.

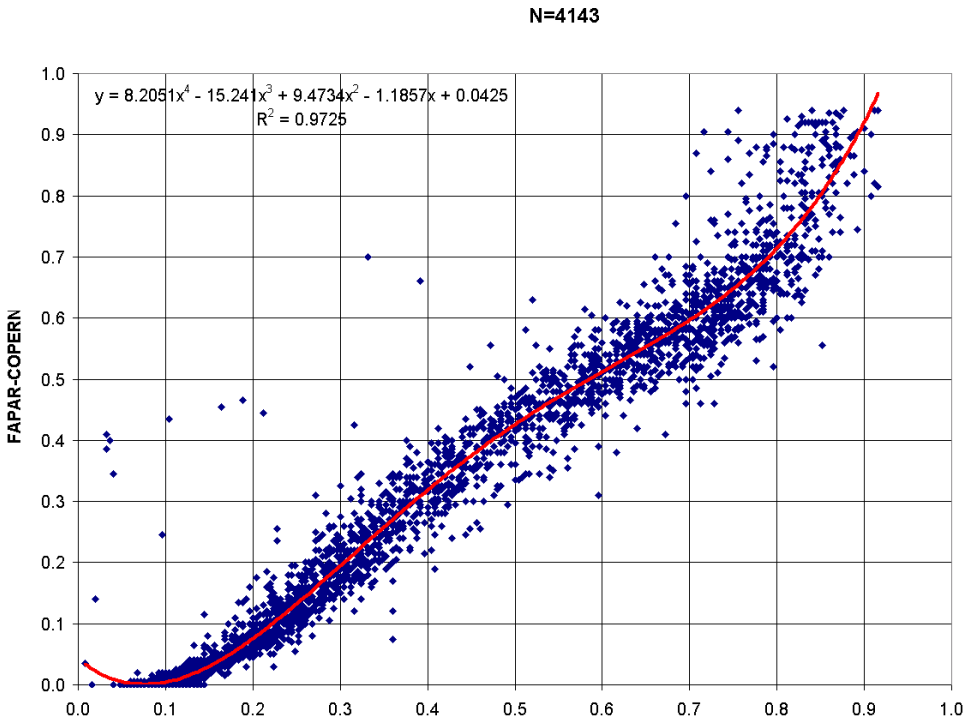


Figure 18: Example of the relation between MODIS NDVI and the Copernicus[26] fAPAR product with data from nine dates between 2014 and 2016 (dekads 4, 16 and 28 from 2014-16).

Albedo

The method applied to calculate the albedo assigns a specific weight w_i to each available spectral band i. The assigned weights compensate for the uneven distribution of the incoming solar radiation over the spectrum and depend on the sensor of the input data (details are provided in the Data Manual). The final albedo is computed as $r_0 = \Sigma w_i.r_i$ (summation over the i bands), with r_i and w_i the spectral reflectance and weight of the i-th band. Note that $\Sigma w_i = 1$.

Table 14: Overview of the intermediate data components related to fAPAR and albedo

Data component	Unit	Range	Use	Temporal resolution
fAPAR	-	0-1	Used as input to NPP	Dekadal
Surface Albedo	-	0.1-0.4	Used as input to produce E and T.	Dekadal

2.2.5. Weather data

Description

Biomass production and evapotranspiration are driven by meteorological conditions. The transmissivity (see section 2.2.2) of the atmosphere affects the available solar radiation at the land surface and precipitation, temperature, wind speed and relative humidity are important factors for evapotranspiration.

The acquisition of temperature, wind speed and relative humidity data is discussed below. Although these parameters are routinely measured by most meteorological stations around the world the number of meteorological stations in the area of interest is relatively small. WaPOR therefore uses a global atmospheric model to supply this data. The advantages of these models are a good coverage of the whole project area and a high consistency. Drawback is the relatively low resolution of these data sources. Therefore, temperature data is adjusted for orography to improve results in mountainous areas, as explained below.

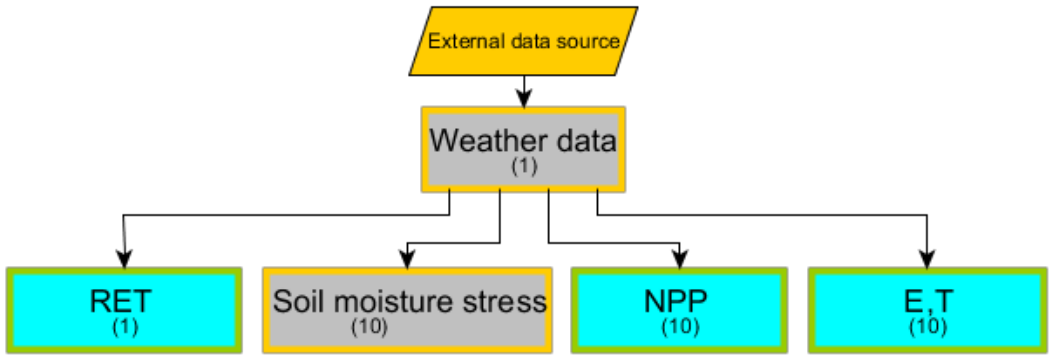
WaPOR area covers various climate zones. For a map of the climate zones according to Köppen, see Methodology document for Level 1 .

Air temperature (T_{min} and T_{max}, in Kelvin), relative humidity (in %) and wind speed (in ms^{-1}) are produced for all three levels. These intermediate data components are produced as daily meteorological grids that are used as input to calculate E, T, RET, NPP and soil moisture stress. These intermediate data components are not published through WaPOR. The quality and resolution of the input data has a strong impact on the output data. Although some adjustments can be made to improve input meteorological data, they are generally based on coarse resolution products.

Methodology

Box 13: Weather data in relation to other data components.

[26] http://land.copernicus.eu/global/products/fapar



- ✓ Weather data refers to air temperature, relative humidity and wind speed are derived from an external data source.
- ✓ Weather data is an important element for calculating biomass production, evaporation and transpiration, it is indirectly connected to most data components.

Temperature, relative humidity and wind speed are derived from a global atmospheric model which uses both synoptic observations and global climate models to produce hourly grids for a large number of atmospheric variables.

These data are resampled to match the resolution of the Level. In order to produce smooth meteorological data fields, relative humidity and wind speed are resampled using a bilinear interpolation method, and temperature is additionally resampled using information on elevation.

Weather shows large variation over short distances, particularly in mountainous areas. Characterising this variability is difficult without detailed monitoring with many ground stations. Temperature is strongly affected by elevation. In general, temperature decreases 6˚C for every km of increasing elevation. The average input data temperature values are at 0.25 degrees resolution (i.e. pixel values representing the average temperature within an area of approximately 25km) do not sufficiently take the effect of topography and elevation into account in mountainous areas. The temperature data is therefore resampled on the basis of elevation. This is done two steps:

1. The average elevation of the input pixel is calculated by resampling the DEM to 0.25 degrees. The input temperature data is then assumed to be representative for this elevation.
2. The temperature of every pixel at Level 1, 2 and 3 is recalculated on the basis of its elevation difference with the average elevation using the temperature lapse rate of 6˚C/km.

Figure 19 shows an example where a DEM was used to resample coarse resolution global temperature data. The Bekaa valley is not visible in the original and bilinear resampled data. Resampling based on the elevation makes the valley visible, with cold mountain ranges on both sides and a relatively warm valley floor. The effect of aspect was not taken into account as this would introduce additional uncertainties that could not be quantified within the scope of the current exercise.

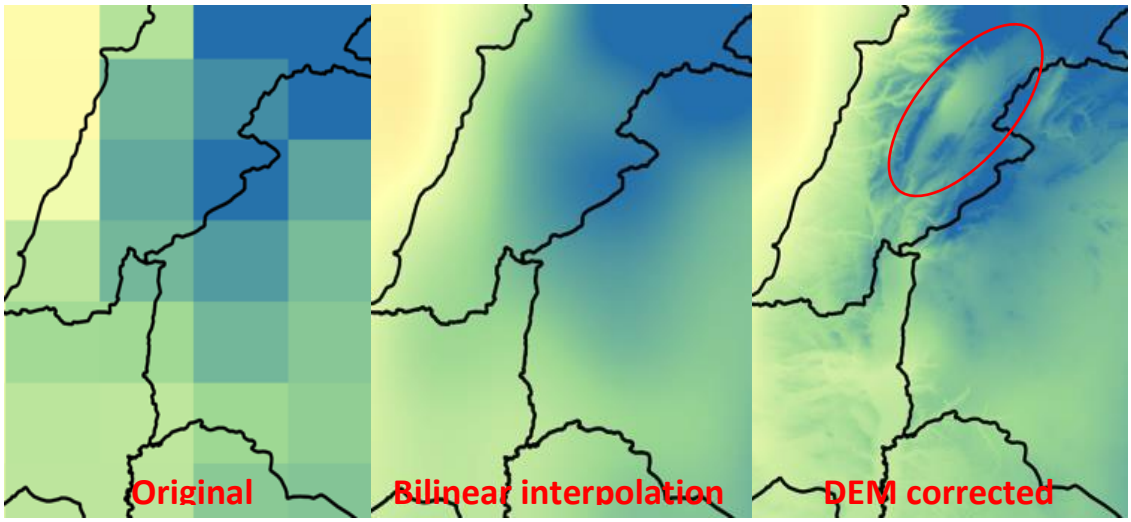


Figure 19: Example of coarse resolution global temperature data resampled for the Bekaa valley (circled) using a DEM. This example uses GEOS-5 temperature data.

Table 15: Overview of intermediate data components related to weather

Data component	Unit	Range	Use	Temporal resolution
T_{min} and T_{max}	K		Used to calculate E, T, RET, NPP and soil moisture.	Daily
Relative humidity	%		Used to calculate E, T, RET, NPP and soil moisture.	Daily
Wind speed	ms^{-1}		Used to calculate E, T, RET, NPP and soil moisture.	Daily

Annex 1: summary table of sensors used in WaPOR v1.0, L2

L2 Data component	Input data components	Sensor	Data product	Comment
Evaporation, Transpiration, Interception	Precipitation		CHIRPS v2, CHIRP	
	Surface albedo	PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, MOD09GA, resampled to 100m
	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	NDVI	PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, resampled to 100m
	Soil moisture stress	MODIS	MOD11A1, MYD11A1	Land Surface Temperature
	Solar radiation		SRTM	DEM
		MSG		Transmissivity
	Land Cover		WaPOR LCC product	
NPP	Solar radiation		SRTM	DEM
		MSG		Transmissivity
	Soil moisture stress	MODIS	MOD11A1, MYD11A1	Land Surface Temperature
	fAPAR	PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, resampled to 100m
	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-V		MERRA used up to start of GEOS-V (21-2-2014)
	Precipitation		CHIRPS v2	
	Land Cover		WaPOR LCC product	
Phenology	NDVI	PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, resampled to 100m
Land cover classification[27]		PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, resampled to 100m

Annex 1: Summary table of sensors and products used for L2 v1.0 release

[27] Note that this data component is not distributed through WaPOR as of September 2017.

References

Ajtay, G. L. (1979). Terrestrial primary production and phytomass. *The global carbon cycle, SCOPE 13*, 129-181.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO, Rome*, *300*(9), D05109.

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (2005). Penman-Monteith Equation. *Encyclopedia of Soils in the Environment. Elsevier/Academic Press, Oxford/Boston*, 180-188.

Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., ... & Smith, M. (2006a). A recommendation on standardized surface resistance for hourly calculation of reference ET_o by the FAO56 Penman-Monteith method. *Agricultural Water Management*, *81*(1), 1-22.

Allen, R. G., Trezza, R., & Tasumi, M. (2006b). Analytical integrated functions for daily solar radiation on slopes. *Agricultural and Forest Meteorology*, *139*(1), 55-73.

ASCE (1996), Hydrology Handbook, American Society of Civil Engineers Task Committee on Hydrology Handbook, American Society of Civil Engineering, Publ. 28, 784.

Bastiaanssen, W. G. M., Cheema, M. J. M., Immerzeel, W. W., Miltenburg, I. J., & Pelgrum, H. (2012). Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model. *Water Resources Research*, *48*(11).

Camillo, P. J., & Gurney, R. J. (1986). A resistance parameter for bare-soil evaporation models. *Soil Science*, *141*(2), 95-105.

Carlson, T. (2007). An overview of the" triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors*, *7*(8), 1612-1629.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*(3), 241-252.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote sensing of Environment*, *91*(3), 332-344.

Chen, X., Cui, Z., Fan, M., Vitousek, P., Zhao, M., Ma, W., ... & Deng, X. (2014). Producing more grain with lower environmental costs. *Nature*, *514*(7523), 486-489.

Choudhury, B. J., Reginato, R. J., & Idso, S. B. (1986). An analysis of infrared temperature observations over wheat and calculation of latent heat flux. *Agricultural and Forest Meteorology*, *37*(1), 75-88.

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties. *Water resources research*, *14*(4), 601-604.

Di Gregorio, A. (2005). *Land cover classification system: classification concepts and user manual: LCCS* (No. 8). Food & Agriculture Organization.

Dolman, A. J. (1993). A multiple-source land surface energy balance model for use in general circulation models. *Agricultural and Forest Meteorology*, *65*(1-2), 21-45.

Duchemin, B., Hadria, R., Erraki, S., Boulet, G., Maisongrande, P., Chehbouni, A., ... & Khabba, S. (2006). Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices. *Agricultural Water Management*, *79*(1), 1-27.

Eerens, H., Piccard, I., Royer, A., & Orlandi, S. (2004). Methodology of the MARS crop yield forecasting system. Vol. 3: Remote sensing information, data processing and analysis. *Eds. Royer A. and Genovese G., EUR*, *21291*.

FAO (2016). Water Accounting and Auditing: A Sourcebook. *FAO Water Reports*, no. 43.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote sensing of environment*, *80*(1), 185-201.

Hazaymeh, K., & Hassan, Q. K. (2015). Spatiotemporal image-fusion model for enhancing the temporal resolution of Landsat-8 surface reflectance images using MODIS images. *Journal of Applied Remote Sensing*, *9*(1), 096095-096095.

Holtslag, A. A. M. (1984). Estimates of diabatic wind speed profiles from near-surface weather observations. *Boundary-Layer Meteorology*, *29*(3), 225-250.

Jarvis, P. G. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *273*(927), 593-610.

Katerji, W. and Sheng, T. (2012), Litani River Basin Management Support Program, Land Use And Crop Classification Analysis For The Upper Litani River Basin (May 2011 – October 2011) . USAIDS Projects Reports.

Mehrez, M. B., Taconet, O., Vidal-Madjar, D., & Valencogne, C. (1992). Estimation of stomatal resistance and canopy evaporation during the HAPEX-MOBILHY experiment. *Agricultural and forest meteorology*, *58*(3-4), 285-313.

Monin, A. S., & Obukhov, A. M. F. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, *151*(163), e187.

Monteith, J. L. (1965, July). Evaporation and environment. In *Symp. Soc. Exp. Biol* (Vol. 19, No. 205-23, p. 4).

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *Journal of applied ecology*, *9*(3), 747-766.

Moran, M. S., Clarke, T. R., Inoue, Y., & Vidal, A. (1994). Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote sensing of environment*, *49*(3), 246-263.

Sánchez, J. M., Caselles, V., & Kustas, W. P. (2008). Estimating surface energy fluxes using a micro-meteorological model and satellite images. *Tethys J. Weather Clim. West. Mediterr*, *5*, 25-36.

Smith, J. H., Stehman, S. V., Wickham, J. D., & Yang, L. (2003). Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment*, *84*(3), 342-349.

Steduto, P., Hsiao, T. C., Fereres, E., & Raes, D. (2012). I&D Paper 66 *Crop yield response to water*. Roma: FAO.

Stewart, J. B. (1988). Modelling surface conductance of pine forest. *Agricultural and Forest meteorology*, *43*(1), 19-35.

Swets, D. L., Reed, B. C., Rowland, J. D., & Marko, S. E. (1999, May). A weighted least-squares approach to temporal NDVI smoothing. In *Proceedings of the 1999 ASPRS Annual Conference: From Image to Information, Portland, Oregon* (pp. 17-21).

Trischler, J., Sandberg, D., & Thörnqvist, T. (2014). Estimating the annual above-ground biomass production of various species on sites in Sweden on the basis of individual climate and productivity values. *Forests*, *5*(10), 2521-2541.

Valentini, R. (2003). *Fluxes of carbon, water and energy of European forests* (Vol. 163). Springer Science & Business Media.

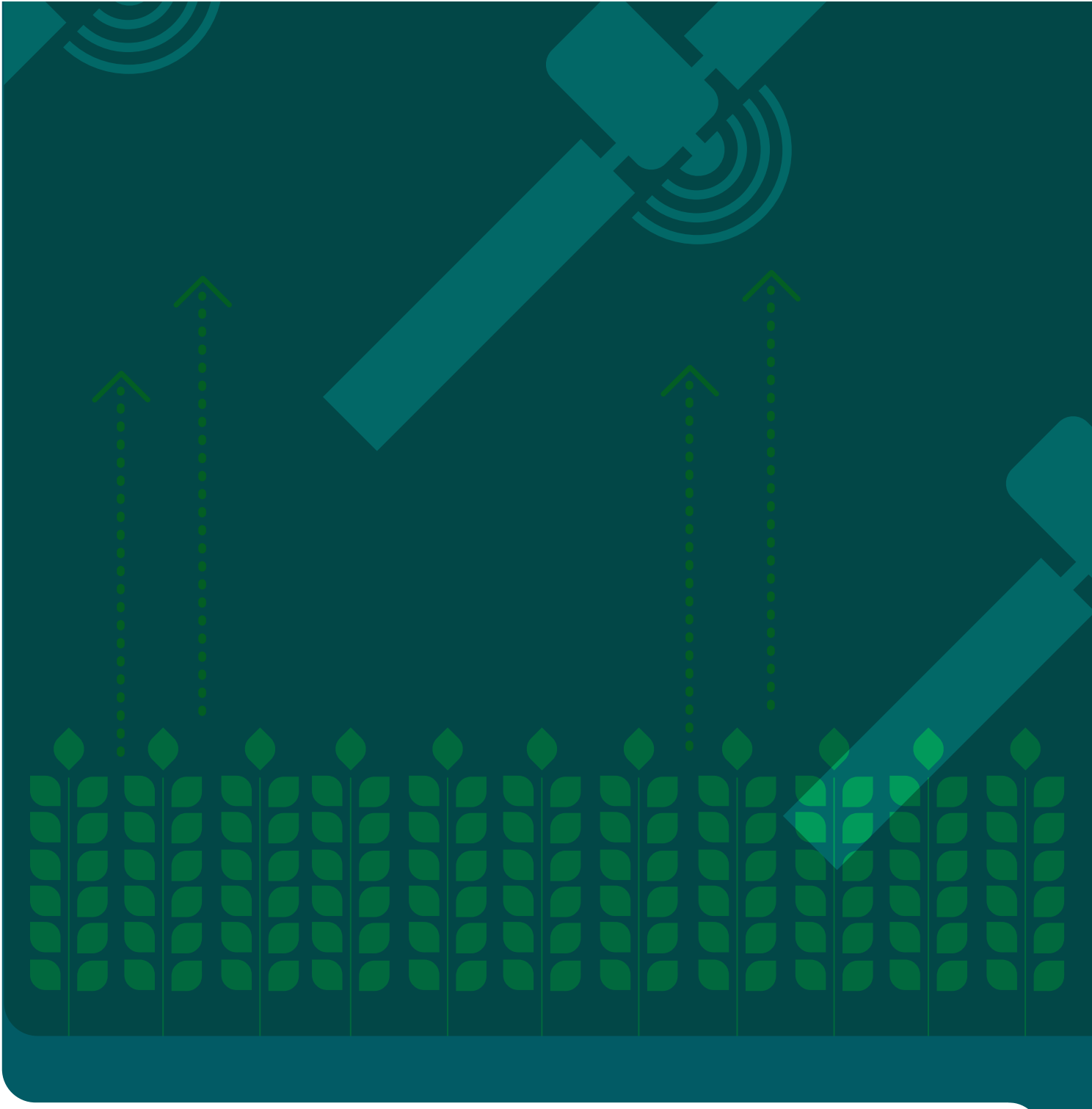
Van Hoolst, R., Eerens, H., Haesen, D., Royer, A., Bydekerke, L., Rojas, O., ... & Racionzer, P. (2016). FAO's AVHRR-based Agricultural Stress Index System (ASIS) for global drought monitoring. *International Journal of Remote Sensing*, *37*(2), 418-439.

Veroustraete, F., Sabbe, H., & Eerens, H. (2002). Estimation of carbon mass fluxes over Europe using the C-Fix model and Euroflux data. *Remote Sensing of Environment*, *83*(3), 376-399.

Wallace, J. S., Gash, J. H. C., McNeil, D. D., & Sivakumar, M. V. K. (1986). Measurement and prediction of actual evaporation from sparse dryland crops. *Scientific Report on Phase II of ODA Project*, *149*.

Wulder, M. A., Hall, R. J., Coops, N. C., & Franklin, S. E. (2004). High spatial resolution remotely sensed data for ecosystem characterization. *BioScience*, *54*(6), 511-521.

Yang, Y., Guan, H., Long, D., Liu, B., Qin, G., Qin, J., & Batelaan, O. (2015). Estimation of surface soil moisture from thermal infrared remote sensing using an improved trapezoid method. *Remote Sensing*, *7*(7), 8250-8270.



Funded by:

Ministry of Foreign Affairs of the Netherlands

Partners:

UNESCO-IHE Institute for Water Education

IWMI International Water Management Institute

Frame consortium:

eleaf FEED THE WORLD

ITC

vito vision on technology

UNIVERSITY OF TWENTE.

WATERWATCH FOUNDATION

ISBN 978-92-5-130057-2

9 789251 300572

I8225EN/1/11.17