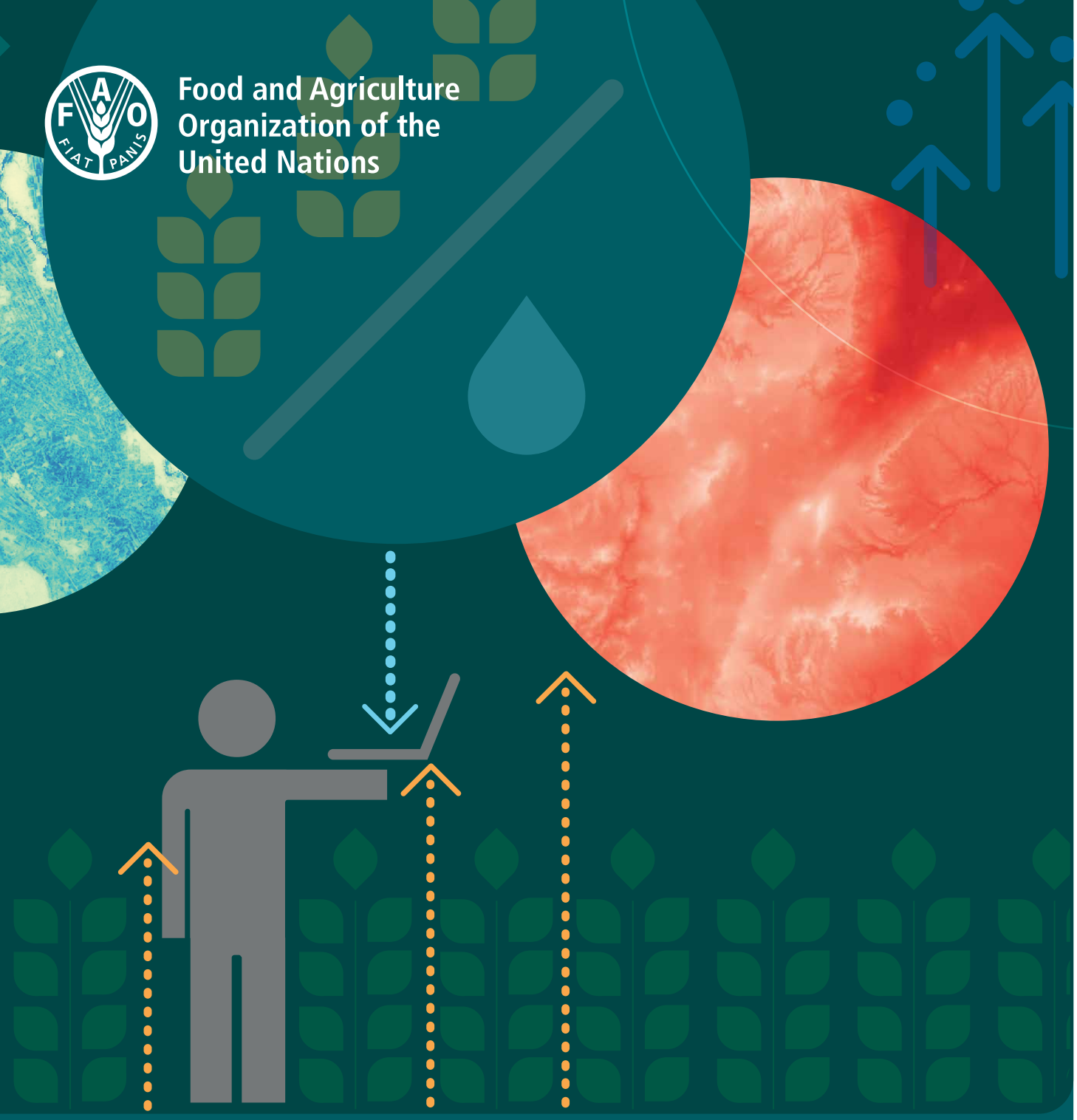


Food and Agriculture Organization of the United Nations

WaPOR DATABASE METHODOLOGY

Version 2 release

April 2020

WaPOR Database methodology

Version 2 release

April 2020

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 2020

Required citation:
FAO. 2020. *WaPOR database methodology: Version 2 release, April 2020*. Rome. https://doi.org/10.4060/ca9894en

ISBN 978-92-5-132981-8

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, as of today, access to 11 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. WaPOR version 1 became available starting from June 2018 and WaPOR version 2 was launched in June 2019. Each version of the data was improved based on extensive internal and external validation and quality assessment. This document describes the methodology used to produce version 2 of the data at the 250m (level 1), 100m (level 2) and 30m (level 3) resolution distributed through the WaPOR portal.

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 two outputs of the programme: the development of an operational methodology and the development of an open-access database to monitor land and water productivity.

The methodology was developed by the FRAME[1] consortium, consisting of 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, during the second beta methodology review workshop, in January 2018 and the project evaluation meeting, March 2019. Participants in these workshops were: Pauline Jaquot, 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, Marloes Mul, Jonna Van Opstal (IHE-Delft), 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 person. FRAME Contract Manager and Managing Director: Steven Wonink (steven.wonink@eleaf.com).

Abbreviations and acronyms

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
TBP	Total biomass production
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

Example of TBP data component at Level 2 (Season 2, 2015). Source: wapor.apps.fao.org

1. Introduction

This report outlines the methodology applied to produce the different data components of WaPOR, the Food and Agriculture Organization of the United Nations (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 and following a further expert review, WaPOR 2.0 was launched in June 2019. This document describes the methodology applied to produce the database at levels 1 (250 m), 2 (100 m) and 3 (30 m), as made available through WaPOR version 2.0 release.

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.

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°)	Countries[1]: Morocco, Tunisia, Egypt, Ghana, Kenya, Niger, Sudan, South-Sudan, Mali, Benin, Ethiopia, Rwanda, Burundi, Mozambique, Uganda, Iraq, 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, Lebanon, Kenya, Mozambique and Sudan.

[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 Non-Self Governing Territories and Sovereignty unsettled territories

[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.

The data components that are produced for the WaPOR database are listed in Table 2. Water productivity, evaporation, transpiration, interception, net primary productivity, total biomass production and land cover classifications are produced at all three levels. Phenology is delivered for levels 2 and 3 and Harvest Index for Level 3 only (thereby allowing for the calculation of yield in combination with other crop parameters), 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.

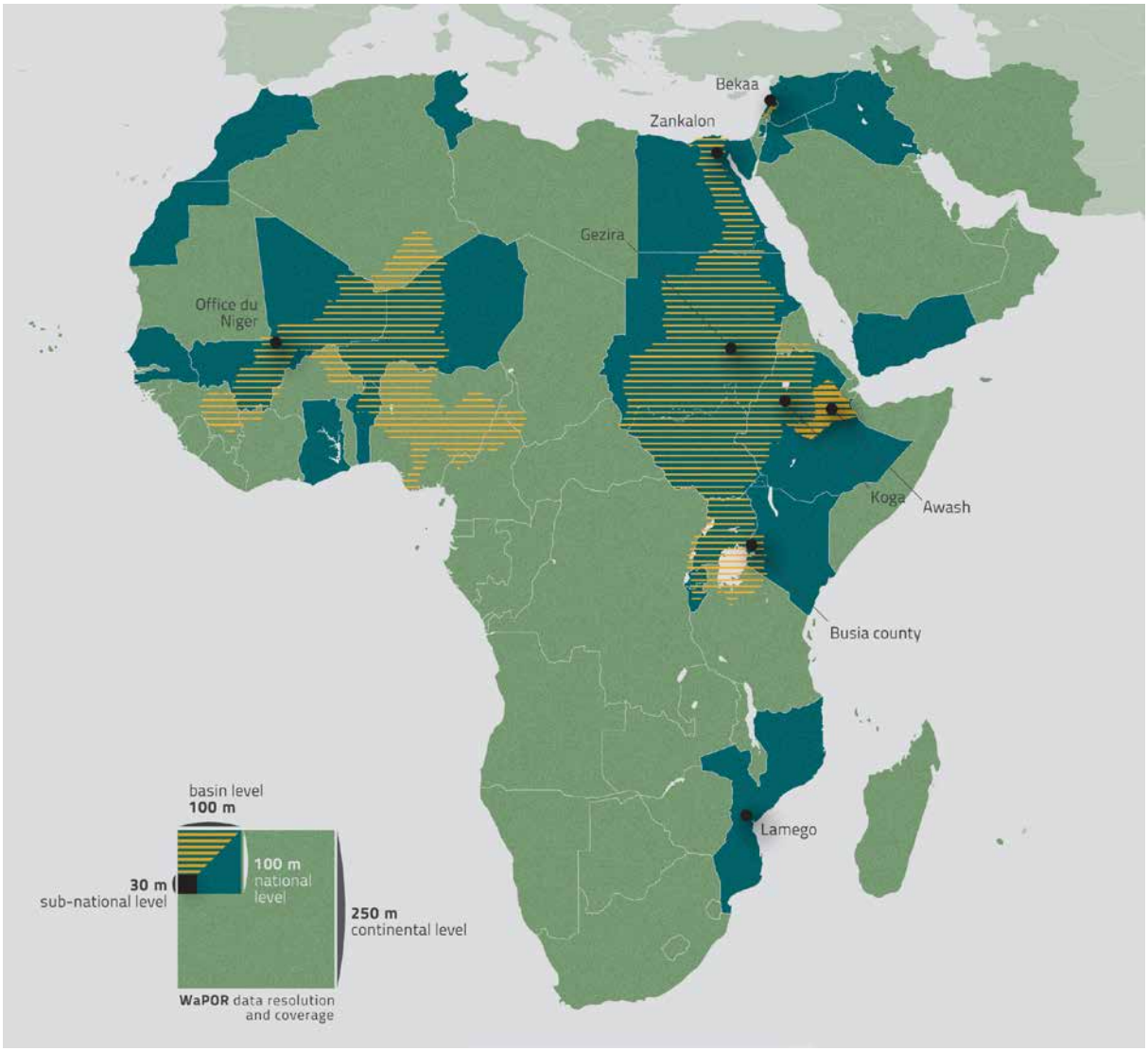
Additional complementary data layers are listed in Table 3. These include layers that can be applied by the user to add value to the WaPOR data components, or to inform the user about the quality of input data. Details are given in Chapter 2.

[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 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° N/S	Lat/Lon (m) at 40° 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

Figure 1:

WaPOR data coverage at all three levels of spatial resolution and areas of interest



Source: http://www.fao.org/in-action/remote-sensing-for-water-productivity/en/

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/ Annual	Dekadal/ Annual	
Interception (I)	Dekadal/ Annual	Dekadal/ Annual	Dekadal/ Annual	

Actual evapotranspiration and interception (ETIa)	Dekadal/ Annual	Dekadal/ Annual	Dekadal/ Annual	
Net primary production (NPP)	Dekadal	Dekadal	Dekadal	
Total biomass production (TBP)	Annual	Seasonal	Seasonal	
Phenology		Seasonal	Seasonal	
Harvest Index (HI)			Seasonal	
Reference Evapotranspiration (RET)	Daily/ Dekadal/ Annual			Different resolution: 20km
Precipitation	Daily/ Dekadal/ Annual			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			Seasonal	Adjust NPP using updated LUE at the end of the season.
NDVI quality layer	Dekadal/ Long-term aggregate	Dekadal/ Long-term aggregate	Dekadal/ Long-term aggregate	Indicates quality of external data used to produce NDVI.
Land surface temperature (LST) quality layer	Dekadal	Dekadal		Indicates the quality of the Land surface temperature (LST) input data
Land cover classification quality	Annual	Annual	Seasonal	Indicates the quality of the Land cover classification.

1.2 Structure of the database methodology document

This document describes the characteristics and the methodology applied to produce the data published on WaPOR as of June 2019 (version 2.0). The methodology applied throughout all three levels is given in this document.

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, 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 the production of levels 1, 2 and 3 until December 2020 is provided in Annex 1.

1.3 Related documents

This document focuses on the core theory that underlies the methodology applied for the production of version 2.0 of the WaPOR data components. Related, more detailed, information can be found in the following accompanying documents:

- The data manuals that accompany WaPOR 2.0 contain a detailed discussion of the processing chain of each dataset related to evapotranspiration and biomass production. The data manual includes details on external data sources used, such as satellite sensors, meteorological data and static data sources at various resolutions. Differences in the processing chain due to different input data sources and resolutions 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 reports (FAO and IHE-Delft, 2019 and FAO and FRAME, 2020).

BOX 1:
Data structure

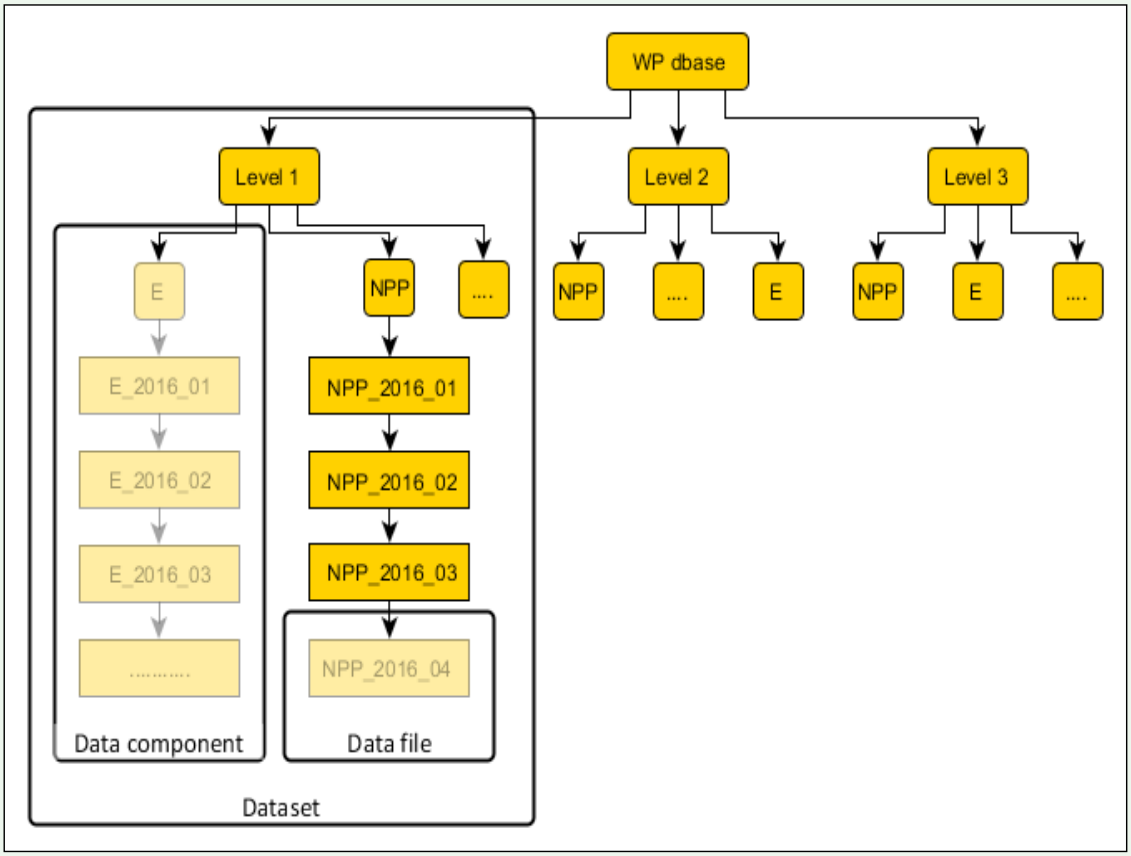
The term ***data*** is frequently used throughout this document. The following definitions explain the different uses of the term within WaPOR :

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.

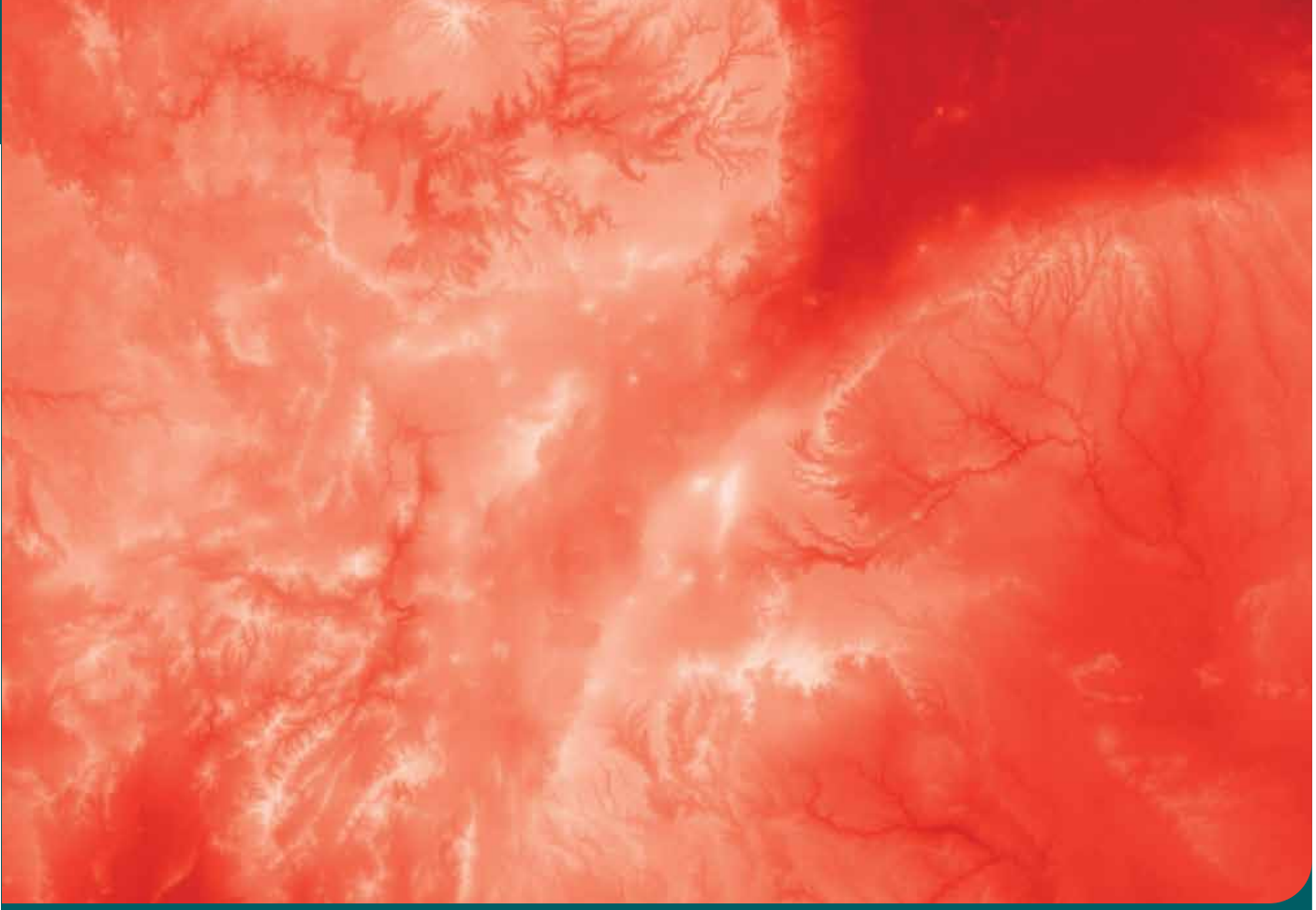
[3] A few data components that are also intermediate data components are distributed through WaPOR, these will be noted in the text.

- ***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 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, which 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.



Example of ETIa data component at Level 2 (2016, dekad 36). Source: wapor.apps.fao.org

2. Methodology for the production of the data components

As shown in Table 2 and 3, the 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 several 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:

- 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.
- Different approaches to the pre-processing of external data sources can easily be incorporated without changing the overall processing structure of the data components.
- 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.

- All input data is converted to the required resolution prior to the processing of the data components.
- Improved processing efficiency is ensured, as the intermediate data components are produced only once and are used as input in various data components.
- 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:

- 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.
- 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 and 2, the numerator is total biomass production (TBP), as no information on crop is available for those data components. At level 3, WP can be calculated using yield as numerator, as crop-specific information is available for higher resolution data.
- Phenology is only produced at level 2 and 3 in order to produce data components with a degree of seasonality, harvest index is only available at level 3 as it relies on detailed crop information.[4]

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 phenology data component also requires studying the NDVI intermediate data

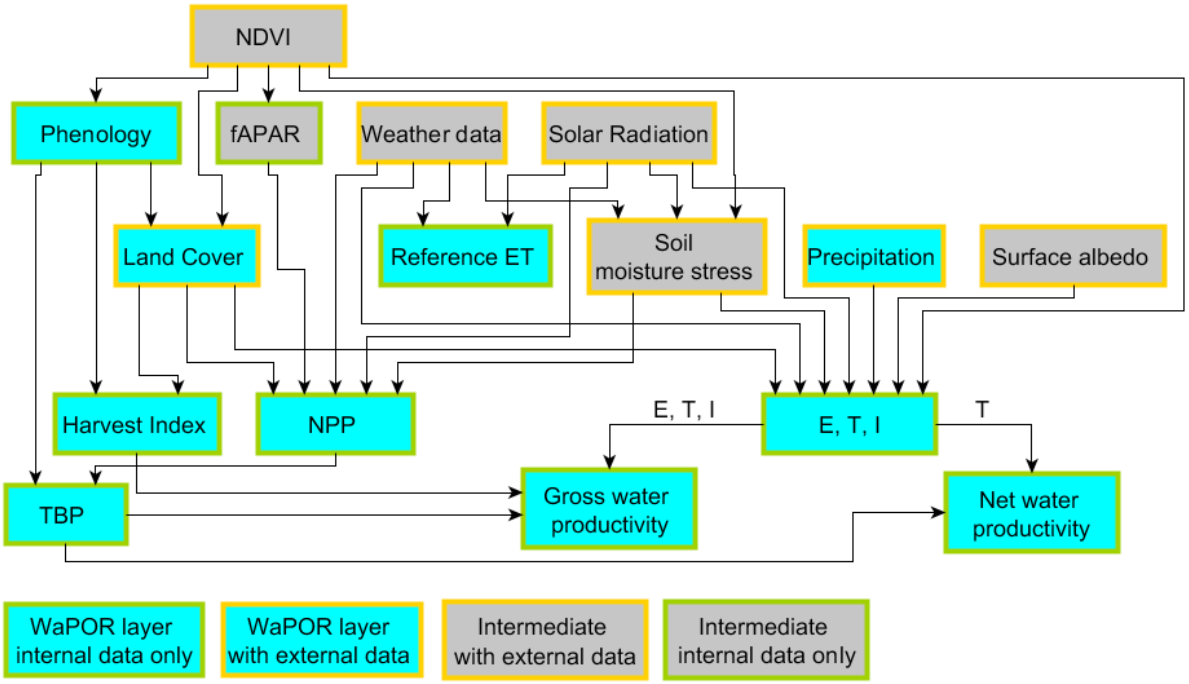
[4] The quality and usefulness of the Harvest Index as a spatial data component, is still being assessed (as of March 2020). Harvest Index is a fixed crops-specific value, or a range of values, found in literature and made available to WaPOR users for specific applications.

component. For the 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 manuals.

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



It should be noted that the (intermediate) data components are produced in two distinct processing phases, i.e. historical data processing which produced data from 2009 up to April 2019, followed by a phase of continuous near real time (NRT) processing. Data components produced in NRT are released approximately 5 days after the end of a dekad and should be seen as preliminary data components. A higher quality version of the data component is produced and delivered after 6 dekads have passed. This final version of the dekadal dataset has a higher

quality because gap filling and interpolation processes, where needed, have been based on more data observations. This implies that other temporal aggregations (monthly, seasonal, annual), and layers that depend on those, are updated as well. Practically this means that a final annual aggregation of the most recent full calendar year can only be produced after the end of February. Likewise, the final monthly aggregation of the most recent calendar months can only be produced 2 full months later.

2.1 WaPOR data components

This section describes the methodology applied to derive the data components as published through WaPOR (WaPOR v 2.0) at https://wapor.apps.fao.org/home/WAPOR_2/1

2.1.1 Water Productivity

2.1.1.1 Gross biomass water productivity

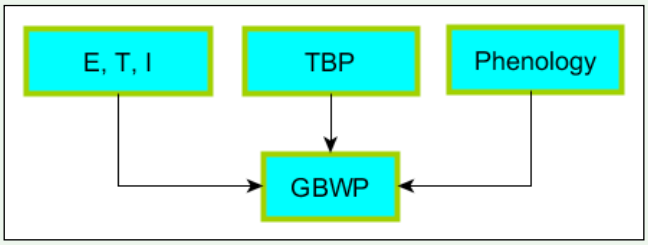
Description

The Gross biomass water productivity expresses the quantity of output (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 2.1.5.), this indicator provides insights on the impact of vegetation development on consumptive water use and thus on water balance in a given domain.

Gross biomass water productivity is calculated and made available through WaPOR on seasonal basis at level 2 and 3. However, as the input data (NPP) are also available on dekadal basis, user-defined temporal aggregations are possible.[5]

Methodology

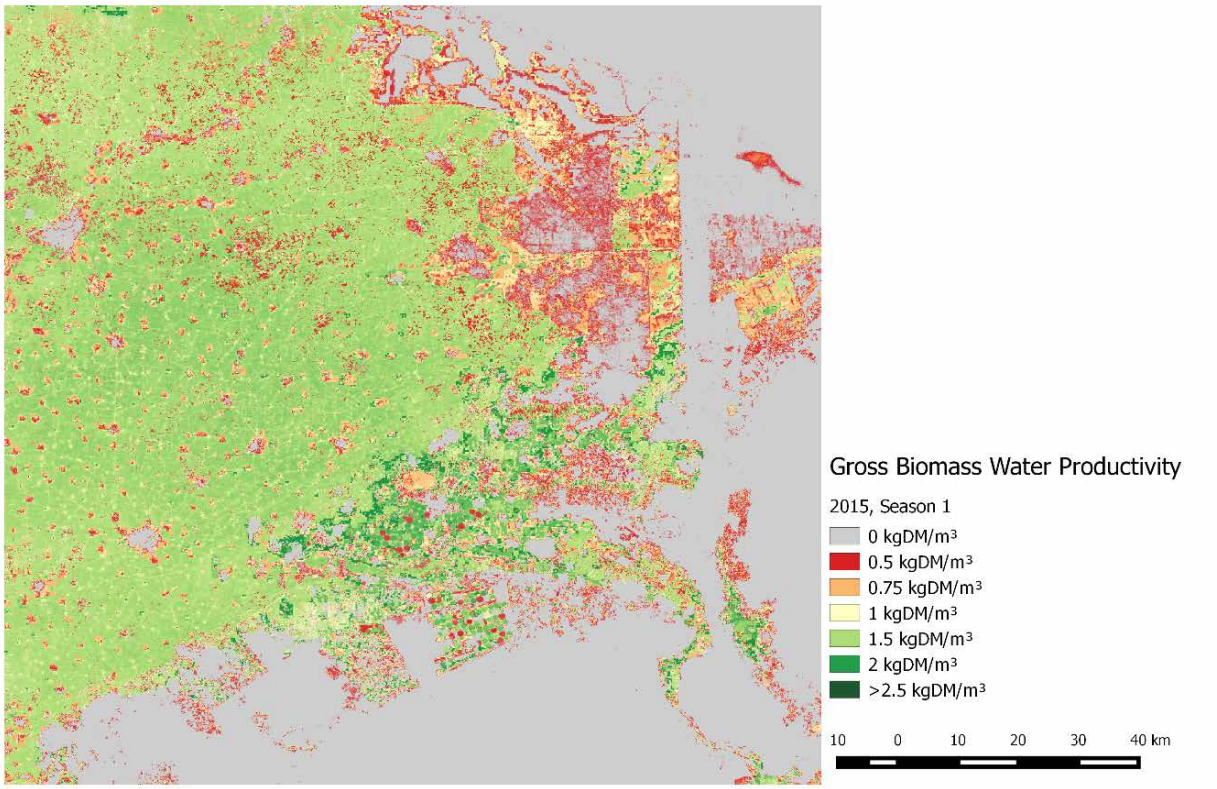
Box 2:
Gross biomass water productivity in relation to other data components



- Calculating GBWP requires input from total 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.

[5] The functionalities for computing GBWP and NBWP over user-defined areas and time are available in WaPOR through the "Area Water Productivity" analysis tool, or through specific APIs.

Figure 3:
Example of seasonal Gross biomass water productivity in the Nile Delta, season 1 of 2015



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

The calculation of Gross biomass water productivity is as follows:

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

Where TBP is total biomass production in kgDM/ha, E is evaporation, T is transpiration and I is interception, all in mm. The following data is used for calculating GBWP: TBP, E, T, I, and phenology if calculated on seasonal time step.

2.1.1.2 Net biomass water productivity

Description

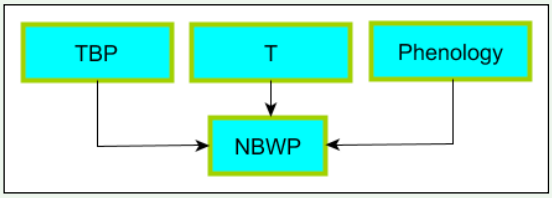
The Net biomass water productivity expresses the quantity of output (total biomass production) in relation to the volume of water beneficially consumed (through canopy transpiration) in the year, and thus net of soil evaporation.

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 seasonal basis at level 2 and 3. However, as the input data are also available on dekadal basis, user-defined temporal aggregations are possible.

Methodology

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



- Calculating NBWP requires input from total 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:

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

Where TBP is total biomass production in kgDM/ha and T is transpiration in mm. The following data is used for calculating NBWP: TBP, 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[1]	Measures quantity of dry biomass output in relation to consumptive water use	Seasonal (further aggregated to user-defined)
NBWP	kg/m³	0 to 6	Measures quantity of dry biomass output in relation to transpiration (or beneficial water consumption)	Seasonal (further aggregated to user-defined)

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

2.1.2 Crop water productivity

2.1.2.1 Gross crop water productivity

Description

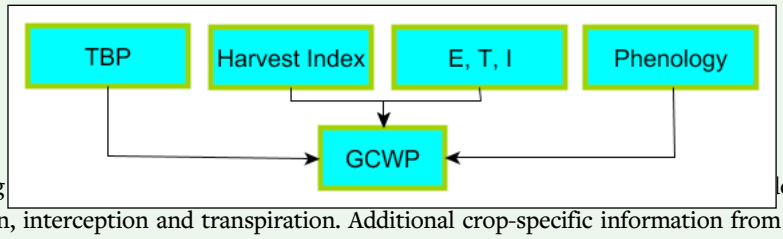
The gross crop water productivity expresses the quantity of the crop yield in relation to the total volume of water consumed in a given period (FAO, 2016). By relating the yield to total evapotranspiration (sum of soil evaporation, canopy transpiration and interception), this indicator provides insights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

Gross Crop water productivity is calculated through WaPOR on a seasonal basis at level 3 depending on availability of crop specific information. However, as the input data are also available on dekadal basis, user-defined temporal aggregations are possible.

Methodology

BOX 4:

Gross Crop water productivity in relation to other data components



- Calculating lex as well as actual evaporation, interception and transpiration. Additional crop-specific information from literature or field measurements is also required to relate biomass to yield The output is not used in any other data component.

The calculation of Gross Crop water productivity is as follows:

$$GCWP = \frac{TBP * HI * AoT/(1 - \theta)}{E + T + I} \qquad (3)$$

Where TBP is seasonal total biomass production in kgDM/ha, HI is described in section 2.1.3 and it represents the harvestable part of the crop. AoT is the ratio between above-ground and total biomass production, θ is the moisture content in the harvested product, both values are found in literature for most common crops and are made available through WaPOR resources. Additional multiplier is needed for accounting for different LUE of C4 crops. E is evaporation, T is transpiration and I is interception, all in mm. The following data is used for calculating GCWP: phenology, seasonal TBP, HI and seasonal E, T and I.

2.1.2.2 Net crop water productivity

Description

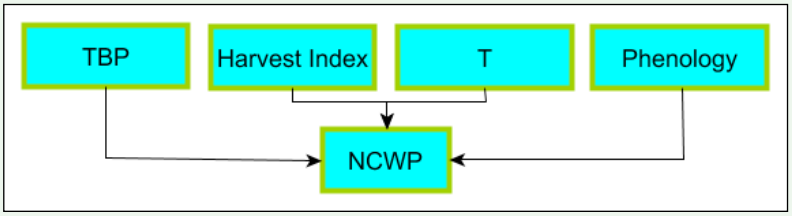
The Net Crop water productivity expresses the quantity of the major crop yields in relation to the volume of water beneficially consumed (through canopy transpiration) in the season, and thus net of soil evaporation. Contrary to Gross Crop water productivity, Net Crop water productivity is particularly useful in monitoring how effectively crops use water to develop yield.

Net Crop water productivity is calculated and made available through WaPOR on a seasonal basis at level 3. However, as the input data are also available on dekadal basis, user-defined temporal aggregations are possible.

Methodology

BOX 5:

Net crop water productivity in relation to other data components



- Calculating NCWP requires input from phenology, total biomass production, harvest index and transpiration. Additional crop-specific information from literature or literature is also required to relate biomass to yield
- The output is not used in any other data component.

The calculation of Net Crop water productivity is as follows:

$$NCWP = \frac{TBP * HI * AoT/(1-\theta)}{T} \qquad (4)$$

Where TBP is seasonal above ground biomass production in kgDM/ha, HI is described in section 2.1.3 and it represents the harvestable part of the crop yield. AoT is the ratio between above-ground and total biomass production, θ is the moisture content in the harvested product, and both values are found in literature for most common crops and are made available through WaPOR resources. Additional multiplier is needed for accounting for different LUE of C_4 crops. T is seasonal Transpiration in mm. The following data is used for calculating NCWP: seasonal TBP, HI and seasonal T.

Table 5

Overview of net water productivity data component

Data component	Unit	Range	Use	Temporal resolution
NCWP	kg/m^3	0 to 6[1]	Measures quantity of crop yield output in relation to transpiration (or beneficial water consumption)	Seasonal (or aggregated to user-defined)
GCWP	kg/m^3	0 to 6	Measures quantity of crop yield output in relation to consumptive water use	Seasonal (or aggregated to user-defined)

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

2.1.3 Harvest index

Description

The harvest index is used to separate biomass production into harvestable and a non-harvestable fraction. The harvest index makes it possible to calculate crop yield and Crop water productivity.

In general, the harvest index is used to assess the productivity of a specific plant variety, indicating how much of the biomass production contributes to the harvestable fraction of a crop (yield). It is expressed as the ratio of weight of dry grains over the total dry matter.[6] Yield formation of a crop is a complex process, involving internal (plant characteristics) and external factors such as weather conditions, soil moisture content and management practices. Timing is particularly important since biomass production and yield formation are affected by external factors in different ways during different growth stages of the crop.

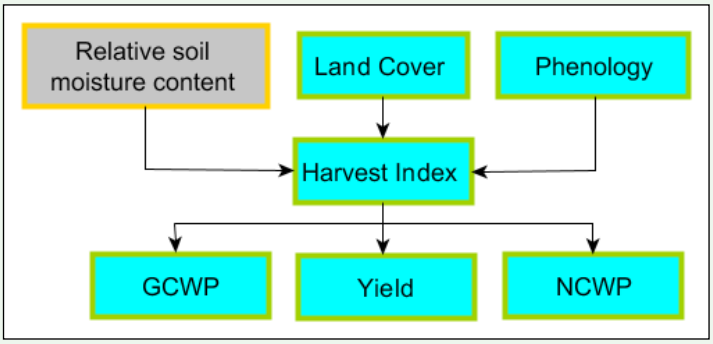
The harvest index is delivered seasonally at level 3, so it represents the harvest index at the end of the growing season. The index ranges from 0 (no harvestable fraction) to a theoretical maximum of 1 when all the biomass (TBP) is harvested. The harvest index is delivered for the main crops: wheat, maize, potatoes, grapes, rice and sugarcane.

The harvest index can only be properly interpreted with land cover information as well as crop phenology information. The former relates the harvest index to the specific crop and the latter to a specific growing season. Typical reference harvest index values for the level 3 crops are for instance: rice (0.43), wheat (0.48), maize (0.48), sugarcane (0.35) (FAO's AquaCrop model, 2017). These values change during the course of the growing season. At first the harvest index will be low, because most of the biomass production contributes to the growth of the crop canopy (vegetative stage). The harvest index increases during the growing season while the crop produces the harvestable material (yield formation). In general, only the final value of the harvest index is relevant as this is related to the amount of material that can be harvested.

Methodology

BOX 6:

Harvest index in relation to other data components



- Calculation of the harvest index requires input from dekadal relative soil moisture content and seasonal phenology.
- Seasonal crop classification information from the land cover data component is needed to identify the specific crops.
- No external data source is required.
- The output is the seasonal harvest index and it is used to calculate the yield and the gross/net crop water productivity.

[6] Definition of harvest index in AQUASTAT (http://www.fao.org/nr/water/aquastat/main/index.stm)

When direct field measurements of crop yield are not available, estimates of the harvest index can be obtained through modelling or methods based on remote sensing. Whereas remote sensing can be used to derive water use and biomass production relatively accurately, the allocation of biomass within the crop cannot be directly monitored. This requires the use of basic crop specific assumptions about growth stages and stress response. Given the complex interplay between the external and internal factors, this is challenging. Accounting for those external influences on the harvest index can be done in different ways. In this project, the methodology adjusts the harvest index for the effect of soil moisture stress.

Central to the methodology for estimating the harvest index is the adjustment of a reference harvest index , based on soil moisture stress experienced by the crop throughout the growing season. Soil moisture stress during flowering and yield formation are considered important. Flowering is difficult to pinpoint as it spans a relatively short period, therefore this methodology only adjusts the on the basis of water stress during yield formation, i.e. from MOS to EOS.

This requires dekadal relative soil moisture content. This approach is a simplification of the method used by FAO's AquaCrop model, which combines the effects of multiple stress factors to adjust a reference harvest index. It is explained in detail in FAO's Irrigation and Drainage Paper 66 (Steduto *et al.*, 2012).

In order to adjust the reference harvest index based only on soil moisture content, the following equations are relevant:

$$D_r = W_{r,fc} - W_r \qquad (5)$$

$$TAW = W_{r,fc} - W_{r,wp} \qquad (6)$$

$$p = \frac{D_r}{TAW} = 1 - s_e \qquad (7)$$

Where:

D_r = Soil water depletion [mm]

$W_{r,fc}$ = Root zone soil moisture content at field capacity [mm]

$W_{r,wp}$ = Root zone soil moisture content at wilting point [mm]

W_r = Root zone soil moisture content [mm]

TAW = Total available soil water [mm]

p = Fractional depletion of TAW

The fractional depletion is used to provide thresholds at which stress will occur (see also Figure 4). The only stress factor taken into account is the stress due to stomata closure (i.e. when p is large). For example, for wheat, the p_{upper} threshold is 0.65 and the p_{lower} threshold is 1 (permanent wilting point). In Figure 4, this is indicated by the relative stress:

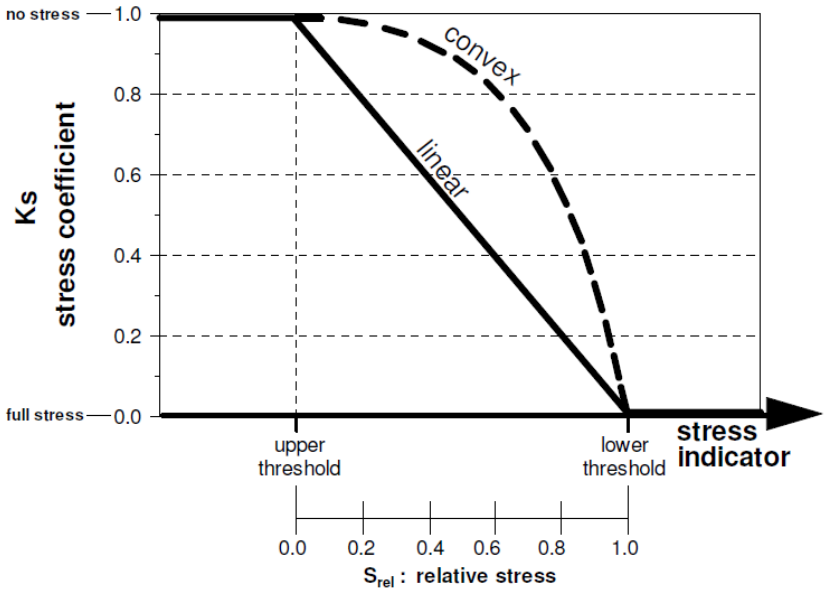
$$S_{rel} = \frac{p_{upper} - p}{p_{upper} - p_{lower}} \qquad (8)$$

The valid range for this relative stress is between 0 and 1. The stress coefficient is then calculated as:

$$K_s = 1 - \frac{e^{S_{rel} f_{shape}} - 1}{e^{f_{shape}} - 1} \qquad (9)$$

Figure 4:

Stress coefficient (Ks) in relation to various degrees of stress



The f_{shape} factor determines the shape of the curve. This stress coefficient is calculated daily and used to limit the harvest index by factor $f_{post\downarrow}$:

$$f_{post\downarrow} = \frac{\sum_{1=1}^{n(yield)} K_s\left(1 - \frac{1-K_s}{b}\right)}{n(yield)} \qquad (10)$$

Where b is a crop-specific parameter (for wheat b =7) indicating the crop response to moisture stress. The parameter $n(yield)$ indicates the number of days for the yield formation, which is derived from phenology data (for wheat $n(yield)$ is approximately 67 days).

The f_{shape} factor for maize, rice and wheat for instance is respectively 6, 3 and 2.5 for the stress related to stomatal closure. These values were obtained from the AquaCrop Reference Manual.

Finally the harvest index can be determined by adjusting the reference harvest index HI_o:

$$HI_{adj} = f_{post\downarrow} HI_0 \qquad (11)$$

HI_o (see Table 6) can be derived from literature for specific crops.

Table 6:
Reference harvest index values used at level 3

Crop	Reference Harvest Index (%)[1]
Wheat	0.48
Maize	0.48
Potatoes	0.75
Grapes	0.5
Rice	0.43
Sugarcane	0.35

[1] Reference Harvest index values taken from the AquaCrop version 6.0 crop files and from Principles of agronomy for sustainable agriculture (Villalobos & Fereres, 2016). Note that the value of reference harvest index is chosen as the middle high end of HI values reported for the majority of the given crop species or class. As a guide, reference HI can be 50% or slightly higher for modern high-yielding cultivars of grain crops, but considerably lower for earlier cultivars and land races. Since the 1980s only marginal improvements have been made in the HI of the major crops. Because it takes approximately 2.5 times as much assimilate to make a gram of oil compared to sugar or starch, HI for oil seed crops are substantially lower than for grain crops, between 0.25 and 0.4. HI for root crops, on the other hand, are usually much higher, with a range of 0.7 to 0.8 common for high-yielding cultivars of potato, sweet potato, and sugar beet, presumably because strong stems are not required to support the harvestable product (Steduto et al., 2012).

Table 7:
Overview of harvest index data component

Data component	Unit	Range	Use	Temporal resolution
HI	-	0 to 1	It is used to calculate crop yield and the Crop water productivity	Seasonal

As alternative to the HI provided as spatial data component in the database, users have access to ranges of HI values as found in literature and chose a suitable value according to the prevailing practices and varieties.

2.1.4 Phenology

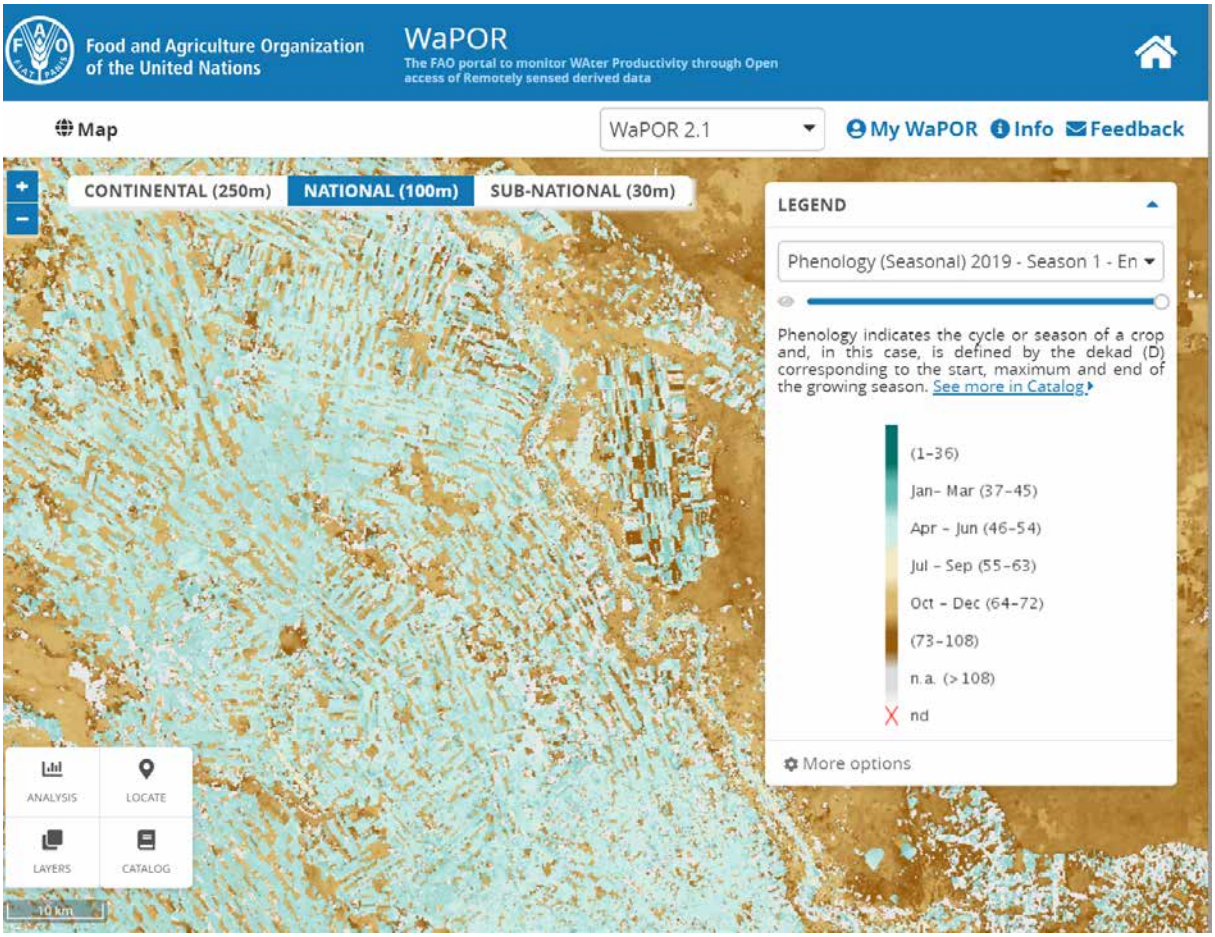
Description

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.

Phenology is produced at levels 2 and 3 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 5 shows an example of the phenology data component (end of season 1, 2015) at level 2.

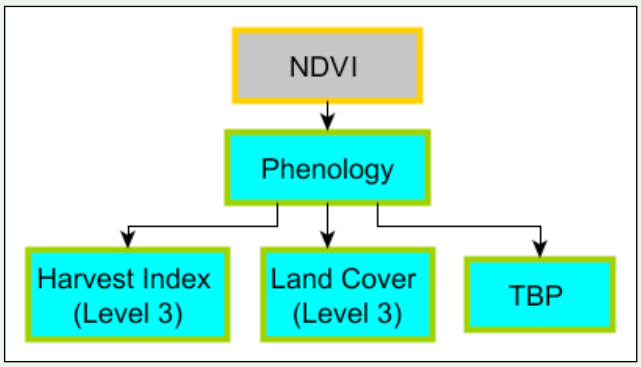
Figure 5:
Example of phenology data at level 2, showing the end of season 1 (2019)



Source: wapor.apps.fao.org

Methodology

BOX 7:
Phenology in relation to other data components



- 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 total biomass production (level 2 and 3) and is used as input to the land cover classification and harvest index (Level 3).

To determine the start, maximum and end of up to two seasons for a given calendar year (January - December) WaPOR applies a methodology that is based on the methods described by Van Hoolst *et al.* (2016).[7] 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 first year, 37-72 for the target year, 73-108 for the next year[8]).

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 the 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 EOS occurs in it. Difficulties arise when the EOS occurs close to the start of a calendar year as it will be an incorrect representation of when the season took place. To circumvent this, a growing season is attributed to a calendar year only if the EOS 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 whereas if the EOS is recorded in dekad 4, it will be attributed to 2018.

If only one season occurs, SOS1 < MOS1 < EOS1. By definition EOS1 lies in the target year but SOS1 and MOS1 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 SOS2 and MOS2.

In the case of perennials such as orchards, it is likely that a SOS, MOS and EOS will not be identified since orchards cannot be directly distinguished in their phenological information. Where a clear growing season can be distinguished, such as for apples, this will be represented in the data. However, if no growing season can be distinguished as for most perennials, a 'no season' label is applied.

Table 8:
Overview of the phenology data component

Data component	Unit	Range	Use	Temporal resolution
Phenology	dekad	1-108[1]	It is used to calculate TBP (Level 2 and 3) and as input for the land cover classification and the harvest index (level 3)	Seasonal

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

[7] 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.

[8] For example, if the target year is 2016, dekad 37 represents 1-10 January 2016, dekad 1 represents 1-10 January 2015, dekad 73 represents 1-10 January 2017. Dekad 30 represents 21-31 October 2015.

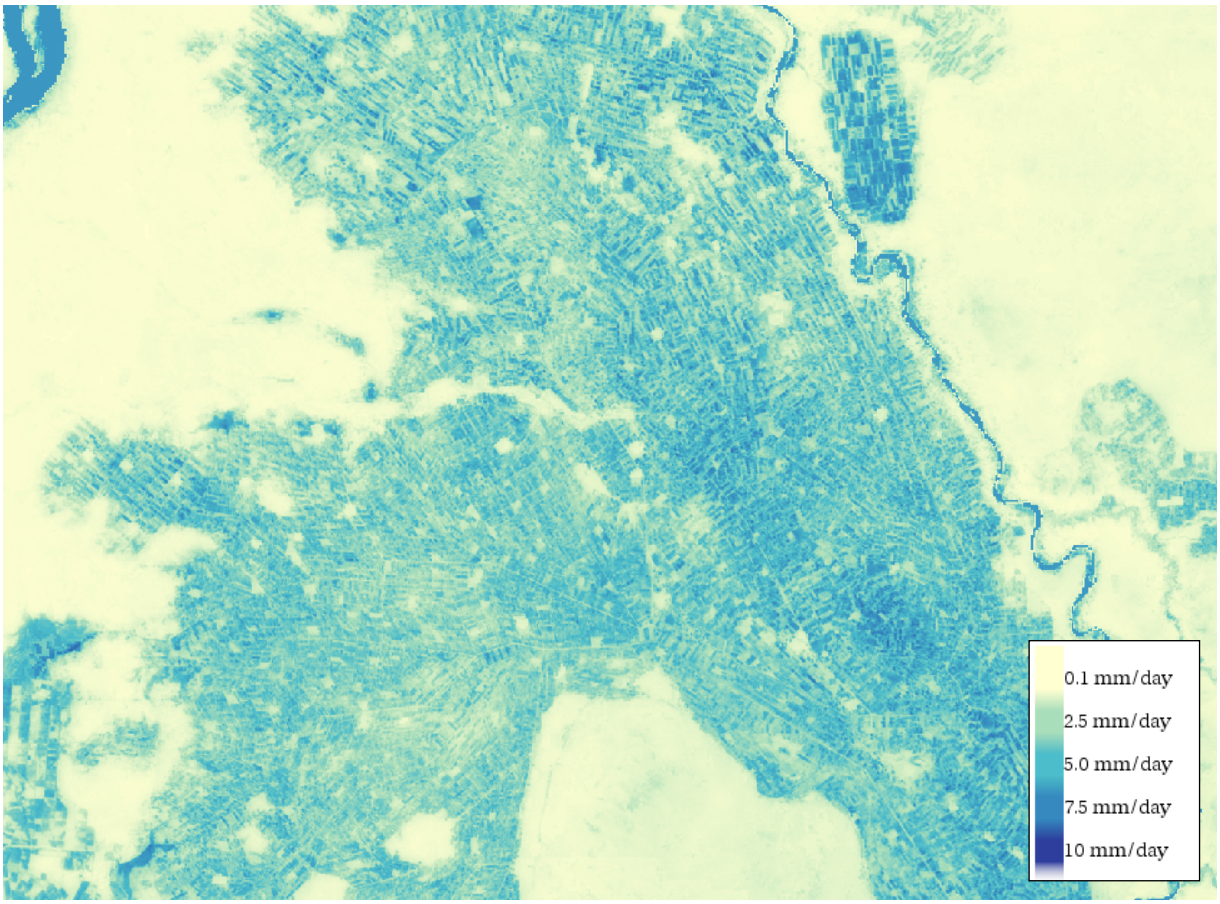
2.1.5 Evaporation, transpiration and interception

Description

Evapotranspiration is the sum of the soil evaporation (E), evaporation from canopy through 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 at all three levels on a dekadal basis, where pixel values represent the average daily E, T, I and ETIa values[9] for that specific dekad in mm/day. Accumulation on monthly and annual basis can also be found on the WaPOR portal for these four parameters at all levels. Figure 6 shows an example of the ETIa data component at level 2.

Figure 6:
Example of ETIa data component at level 2 (2016, dekad 36)



Source: wapor.apps.fao.org

[9] Average daily E, T, I and ETIa values can be converted into volume for a specific area, e.g. 1 mm = 1 l/m² or 1 mm = 10 m³/ha.

Of all data components, E and T require the largest number of inputs to calculate (see Figure 2 and the summary in Box 8). 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 (significantly) lower resolution.[10] The spatial variability of these data sources is therefore more limited, thereby affecting the resulting E and T data component.

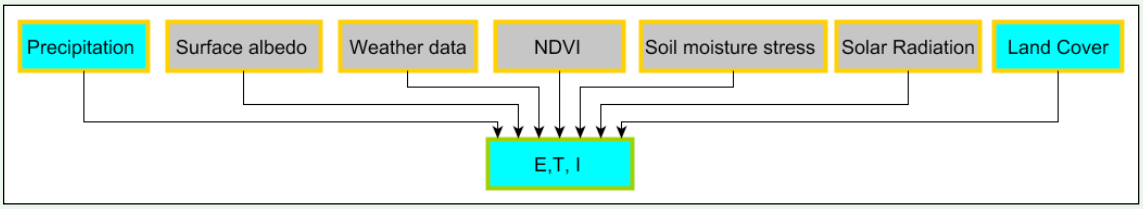
The collection of optical satellite data can be hampered by the presence of clouds, reducing the information on temporal variability. Although both aspects are accommodated for within the data processing chain, its implications should be understood when considering the results: the quality of the E, T, and I data component is a combination of the accuracy of the algorithms and the quality of the external data. The NDVI quality layer is provided to indicate the quality of the external optical satellite data (described in Section 2.2.1). The LST quality layer is provided to indicate the quality of the external thermal satellite data which is used to compute the soil moisture stress (described in section 2.2.3).

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[11]. 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 8:

Evaporation, transpiration and interception in relation to other data components



- Calculating E, T and I 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 and land cover classification.
- 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

[10] For example, temperature data has a spatial resolution of 0.25 degrees (~25 km) and atmospheric transmissivity has a spatial resolution of 4 km.

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

(final or intermediate) data components (see Box 8). 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.

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}]

P_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

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

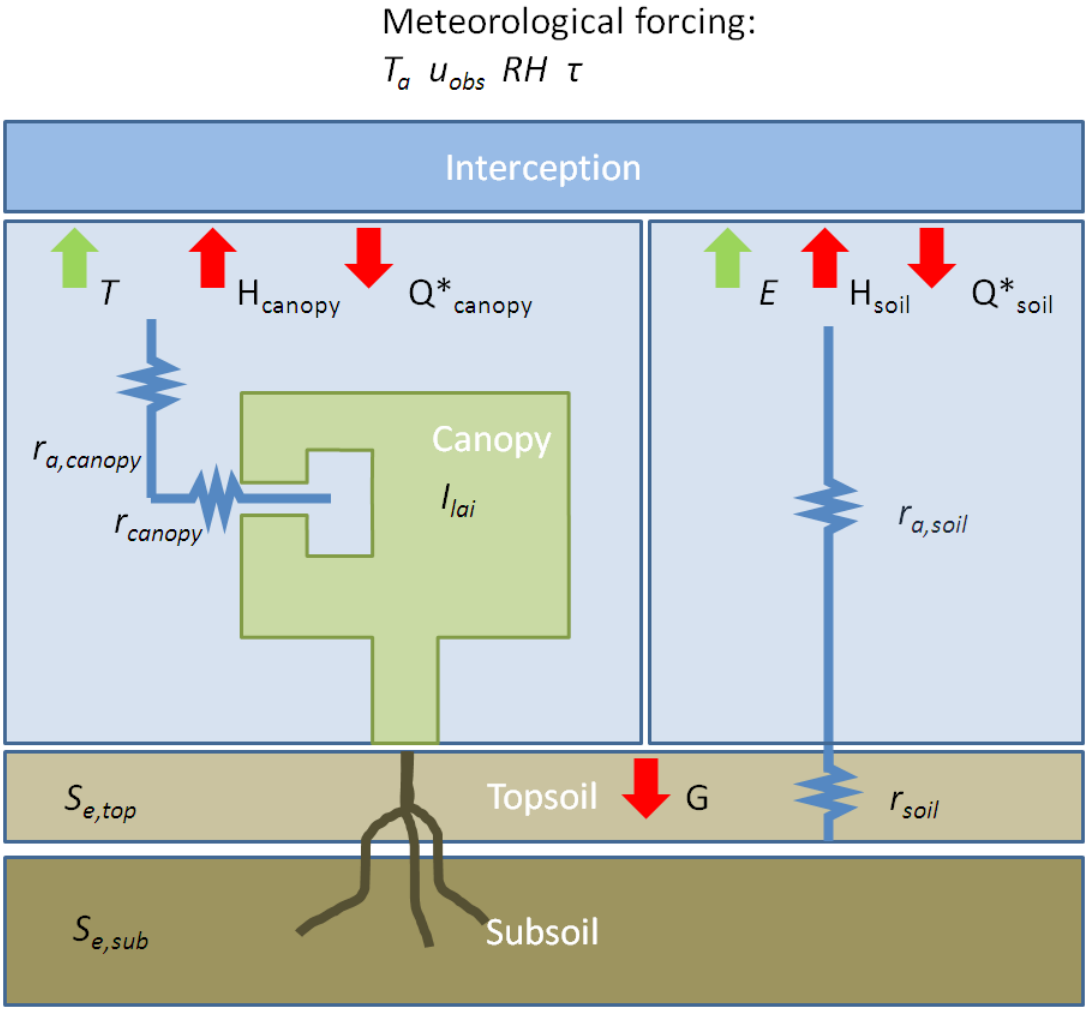
The two equations differ with respect to the net available radiation ($R_{n,\,soil}$ and $R_{n,\,canopy}$) as well as the aerodynamic and surface resistance ($r_{a,\,soil}$, $r_{s,\,soil}$ and $r_{a,\,canopy}$, $r_{s,\,canopy}$). Furthermore, the soil heat flux (G) is not taken into account for transpiration.

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 7.

Figure 7:

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 a_o 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 , 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:

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

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

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.

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

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

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

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 [mm day^{-1}] is a function of the vegetation cover, LAI and precipitation (P), expressed as (Hoyningen-Hüne, 1983, Braden, 1985):

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

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.

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

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 $r_{canopy,\,min}$ 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}]
$l_{lai,eff}$	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_{tr}	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, Reginato and Idso, 1986; Holtslag, 1984) with:

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

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

Where:

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

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

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

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:

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

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 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 is as follows:

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

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

Table 9:

Overview of E, T, I and ETIa data components

Data component	Unit	Range	Use	Temporal resolution
Evaporation	mm/ day	0-10	Measures soil evaporation in a dekad.	Dekadal
Transpiration	mm/ day	0-10	Measures canopy transpiration in a dekad.	Dekadal
Interception	mm/ day	0-10	Measures canopy interception in a dekad.	Dekadal
Actual evapotranspiration and interception (ETIa)	mm/ day	0-12	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

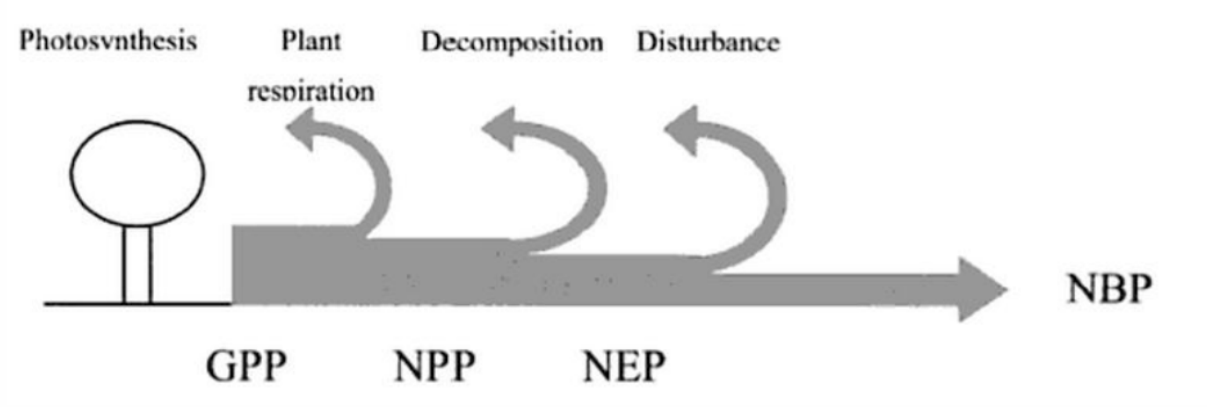
2.1.6 Net primary production

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 8 shows a schematic overview of carbon fluxes.

Figure 8:

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, Sabbe and Eerens (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. Two additional changes were made, which were requested during the methodology review: (i) a reduction factor for soil moisture stress that accounts for short-term water deficiency was added, and (ii) the application of light use efficiencies specific to the type of natural vegetation and the type of crops classified within WaPOR: for level 1 and level 2 the land cover classes produced for WaPOR are used, but because the land cover maps for level 3 are produced at a later stage, the L2 WaPOR land cover classes are first applied. Once the level 3 land cover maps are produced, a raster layer with correction factors containing LUE values specific for each type of natural vegetation (e.g. woodland, shrubland and grassland) and crop type (C3 and C4) (see 2.1.10) is supplied. The user can correct the preliminary NPP data components by multiplying with the LUE correction factor data layer.

[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.

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/m2/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 be converted to DMP using a constant scaling factor of 0.45 gC/gDM (Ajtay, 1979). Therefore 1 gC/m^2/day (NPP) = 22.222 kgDM/ha/day (DMP). Typical values for NPP within the region vary between 0 and 5.4 gC/m^2/day (NPP), or 0 to 120 kgDM/ha/day (DMP), although higher values can occur (theoretically up to 320 kgDM/ha/day). Figure 9 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.

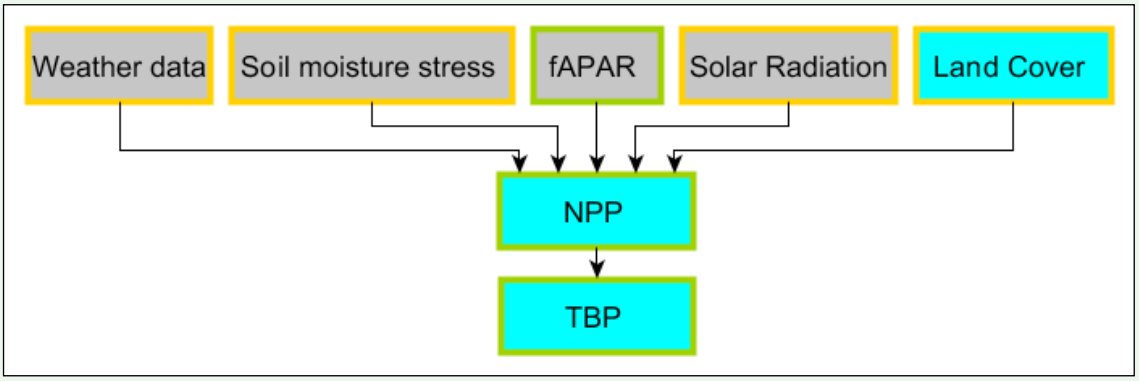
Figure 9:
Example of NPP data component at level 2 (2016, dekad 36)

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

Methodology

BOX 9:

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 total 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), fAPAR (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 2.1.10). Since land cover is only produced at the end of the season, a complementary LUE correction factor data layer is produced for level 3, for which crop information is available, to allow the user to adjust for the correct land cover after the end of season (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, Sabbe and Eerens, 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:

$$\begin{aligned} NPP = Sc.R_s.\ \varepsilon_p.fAPAR.SM.\ \varepsilon_{LUE}\ .\ \varepsilon_T.\ \varepsilon_{CO2}.\ \varepsilon_{AR} &= Sc.\ fAPAR.SM.\ \varepsilon_{LUE}\ R_s.\varepsilon(T,CO_2) \\ &= fAPAR.SM.\ \varepsilon_{LUE}\ NPP_{max}\ [\varepsilon_{RES}] \end{aligned} \quad (35)$$

With: $\varepsilon(T, CO_2) = \varepsilon_P.\varepsilon_T.\varepsilon_{CO2}.\varepsilon_{AR}$.

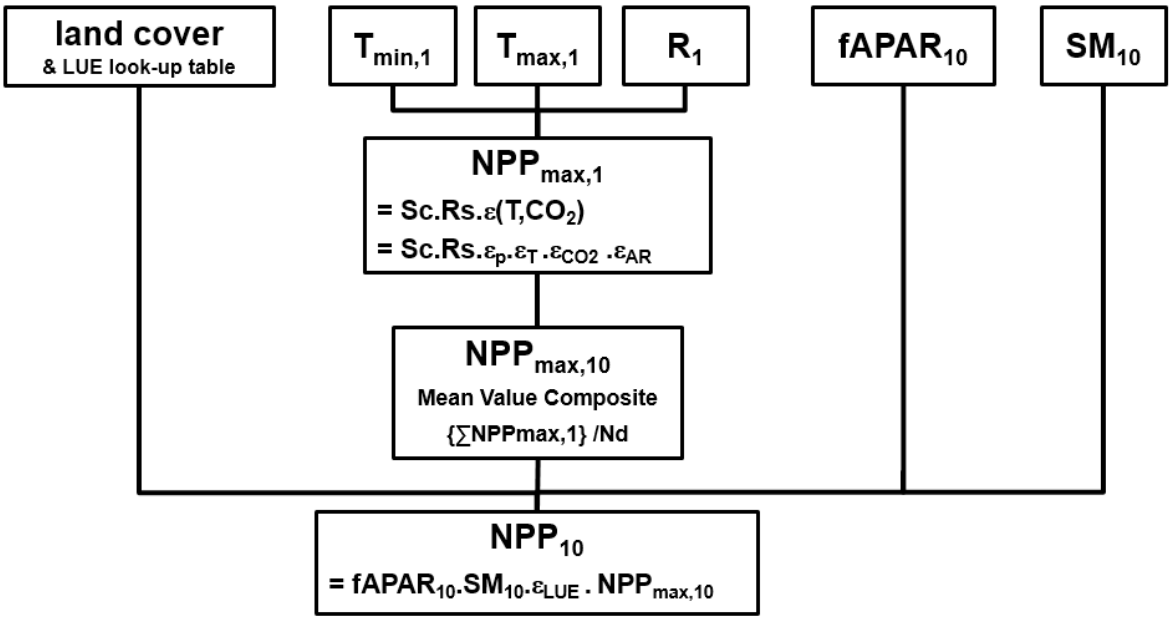
This formulation better highlights the fact that, within the limits of the described model, NPP is only determined by six basic factors: *fAPAR*, soil moisture stress, radiation, temperature, land cover specific Light use efficiency and CO_2. However, in practice the CO_2 level is mostly considered as 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, fAPAR 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 10 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 10,000m^2, 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 10:

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 NPPmax with the fAPAR, soil moisture stress and the land cover dependent Light use efficiency



- 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.

- 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.

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):

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

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

Table 10:
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; Used to calculate TBP data component	Dekadal

[1] Typical range in the region of interest
[2] Theoretical range for NPP

2.1.7 Total biomass production

Description

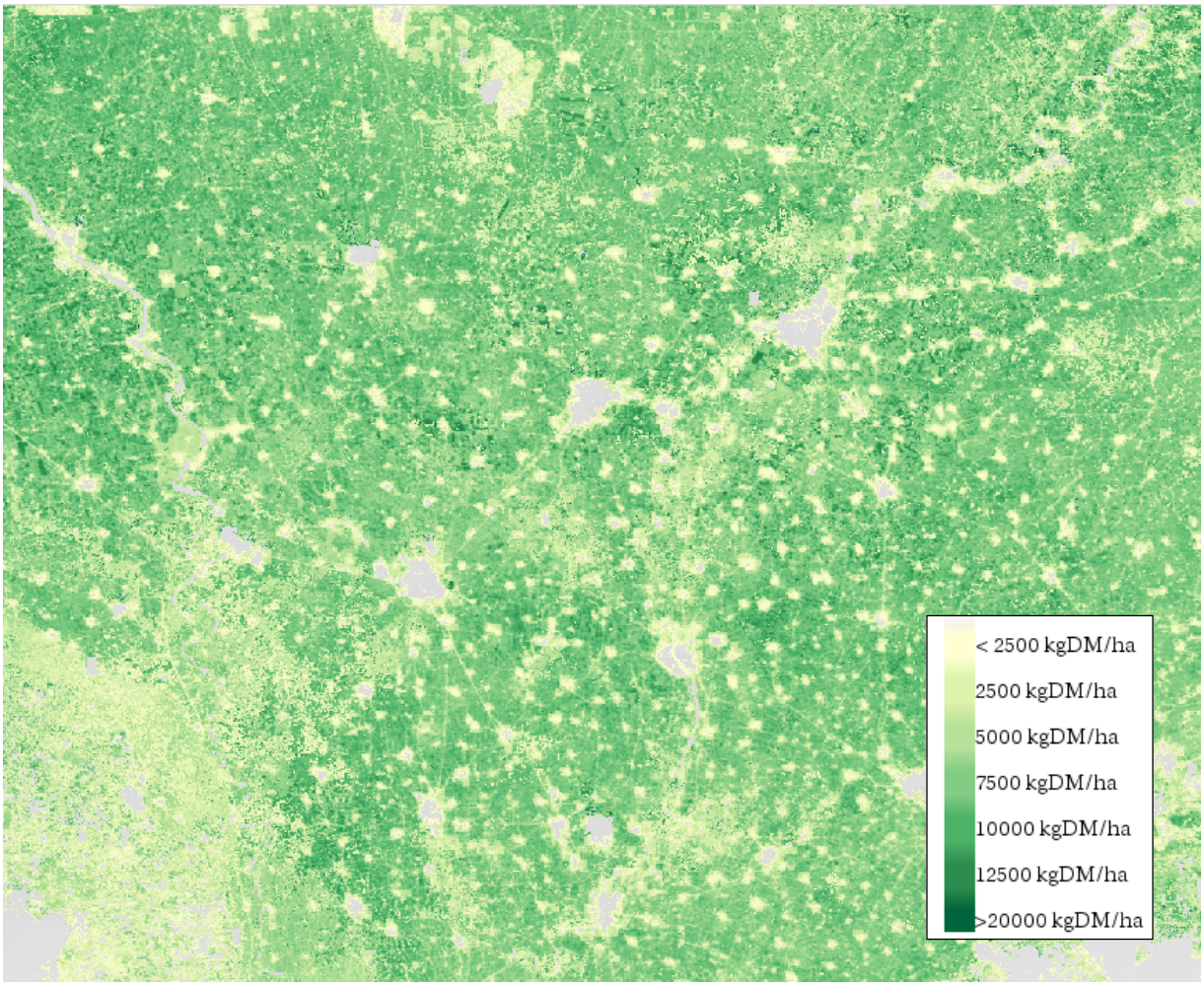
Total biomass production (TBP) is defined as the sum of the dry matter produced during the crop growing season. Hence, TBP steadily increases between the start (SOS) and end of season (EOS).

TBP is a good indicator for crop yield forecasting/estimation because it integrates three important aspects: the current vegetation status (via fAPAR), the meteorological influences (via DMP) and the 'history' (via the summation over the course of the season). TBP, expressed in kgDM/ha, typically ranges between 0 and 45, although higher values are possible. As the TBP is an integration of the DMP over time, its accuracy is closely related to the accuracy of the NPP, which is discussed in Section 2.1.6. Figure 11 shows an example of the TBP data component at level 2.

TBP is delivered for levels 2 and 3, 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 seasonal TBP is the dependency on phenological information, meaning that TBP 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.

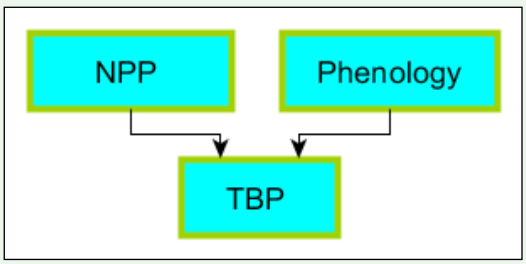
Figure 11:
Example of TBP data component at level 2 (Season 2, 2015)



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

Methodology

BOX 10:
Total biomass production in relation to other data components



- Calculating TBP requires input from NPP for dekadal biomass production and phenology for demarcating the growing season (at level 2 and 3).
- No external data source is required.
- The output is seasonal for level 2 and 3 and annual for level 1.

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. TBP 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).

$$TBP_s = \sum_{i=SOS}^{EOS} N_d(\mathrm{i}) * \mathrm{DMP(i)} \qquad (39)$$

Where:

- DMP(i) is the dry matter production at dekad *i*, expressed in kgDM/ha/day.
- N_d(i) is the number of days within each dekad, varying between 8 (end February) and 11.
- The first term, ΣN_d(i), is needed to obtain the TBP sum in terms of kgDM/ha. Without it, one would obtain the mean.

Table 11:
Overview of TBP data component

Data component	Unit	Range	Use	Temporal resolution
TBP	kgDM/ha	0-20,000 for seasonal, 0-35,000 for annual	Total dry matter produced. It can be used to derive yields if information on phenology and crop parameters are available.	Seasonal, annual, or user-defined (cumulated NPP on period of interest)

2.1.8 Reference ET

Description

Reference evapotranspiration (RET) is defined as the evapotranspiration from a hypothetical reference crop. It simulates the behaviour of a well-watered grass surface and can be used to estimate potential ET for different crops by applying predefined crop coefficients. This information can be used in the design of irrigation schemes. Together with estimates of the evaporation, transpiration and interception, crop coefficients may be derived as the ratio between ETIa and RET. This information may be combined with land cover maps, to infer crop coefficients during the growing season for different type of crops. RET is not influenced by land cover and can be calculated using standard weather measurements and solar radiation.

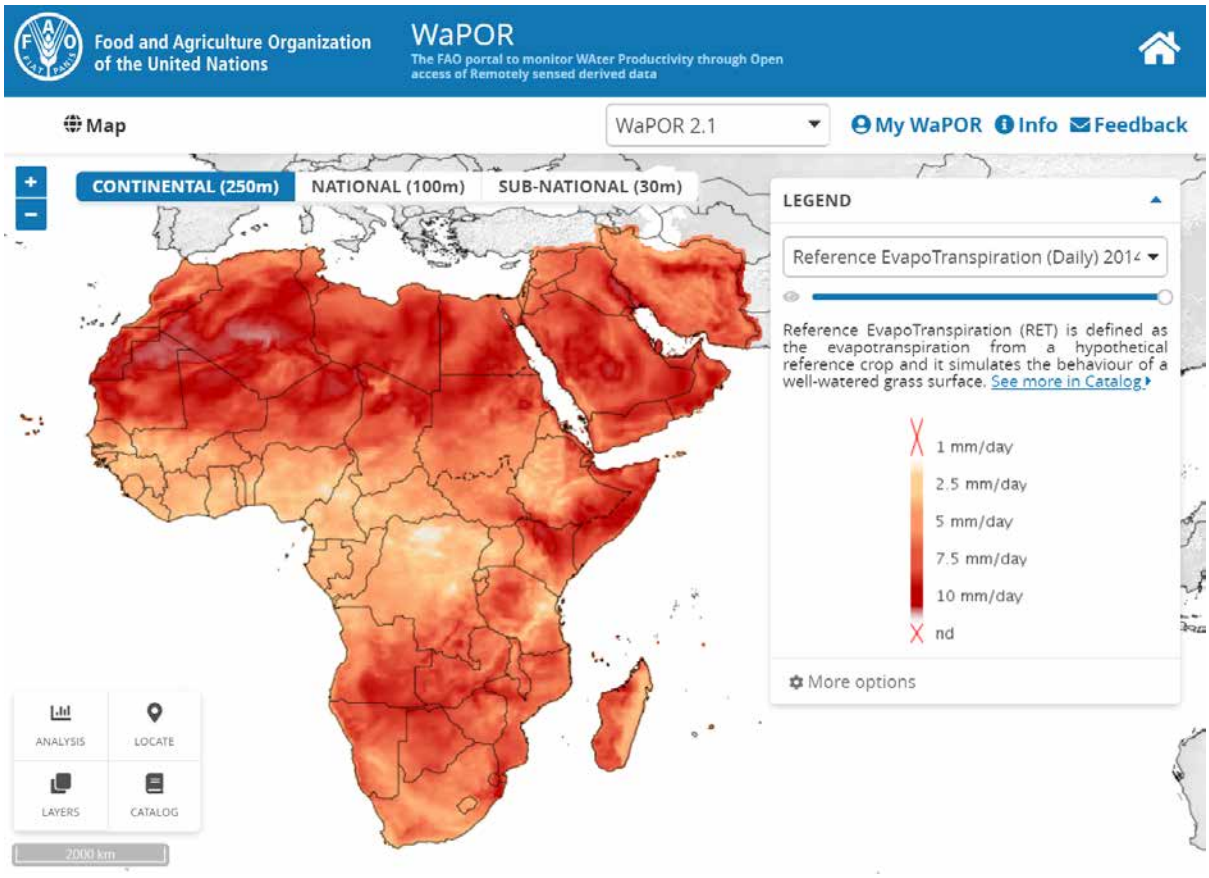
RET is delivered at level 1 with a resolution of 20 km. It is delivered on a daily basis, thus pixel values represent the daily reference ET[17] in mm/day. Figure 12 shows an example of the RET data component. The highest reference evapotranspiration values can be found in the Sahara desert and on the southern part of the Arabian Peninsula where RET values can reach 16 mm/day. Here, temperature and incoming solar radiation are high as it is close to the equator and the relative humidity is low. This creates potential for high evapotranspiration, though little water is actually available to fill this potential. The equatorial rainforests have a lower reference evapotranspira-

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

[17] Average daily ET values can be converted into volume for a specific area, e.g. 1 mm = 1 l/m² or 1 mm = 10 m³/ha.

tion compared to the desert area as the relative humidity is high, supressing the potential for evapotranspiration. In this area, the RET ranges from 1 to 5 mm/day.

Figure 12:
Example of RET data component at level 1, 20 km resolution (15-09-2014)

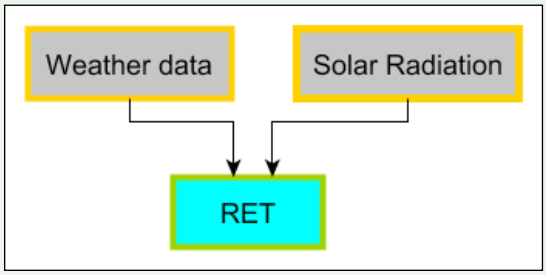


Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor

Since RET is produced at a resolution of 20 km, it has limited information on spatial variation compared to the other data components at level 1 (at 250 m).

Methodology

BOX 11:
Reference evapotranspiration in relation to other data components



- Calculating daily RET requires input from daily weather data and solar radiation.
- No external data source is required to calculate RET.

Reference evapotranspiration is calculated in a similar way as evaporation and transpiration, applying the Penman-Monteith equation. The main differences are that for calculating RET some of the variables are predefined (i.e. crop height, bulk surface resistance and albedo) and E and T are not calculated separately (see Eq 31). The theoretical background behind the Penman-Monteith equation is given in paragraph 2.1.5.

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \qquad (31)$$

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}]
- p_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]
- 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}]

The soil heat flux G is considered to be net 0 for the whole day.

The aerodynamic equation for the reference crop is parametrized taking into account the crop height of 0.12m,

$$r_a = \frac{208}{u_{obs}} \qquad (32)$$

where u_{obs} is the wind speed [ms^{-1}] at observation height of 10m.

- r_s bulk surface resistance [s m^{-1}]

The resistance to vapour flow from the transpiring reference crop is set to 70 sm^{-1} ϱ_a. Δ and γ are a function of air temperature and elevation.

As previously defined in section 2.1.5, the net radiation R_n is solved using the radiation balance:

$$R_n = (1 - \alpha_0)R_s - L^* - I \qquad (33)$$

where α_o is the surface albedo [-] (a fixed albedo of 0.23 is used for the reference crop), R_s is incoming solar radiation [W m^{-2}], L^* is net long wave radiation [W m^{-2}], I is the energy needed for interception [W m^{-2}], which is set at 0 for calculating RET.

For more information on the parameterization, see FAO report 56, page 21 (Allen, 1998).

Table 12:
Overview of RET data component

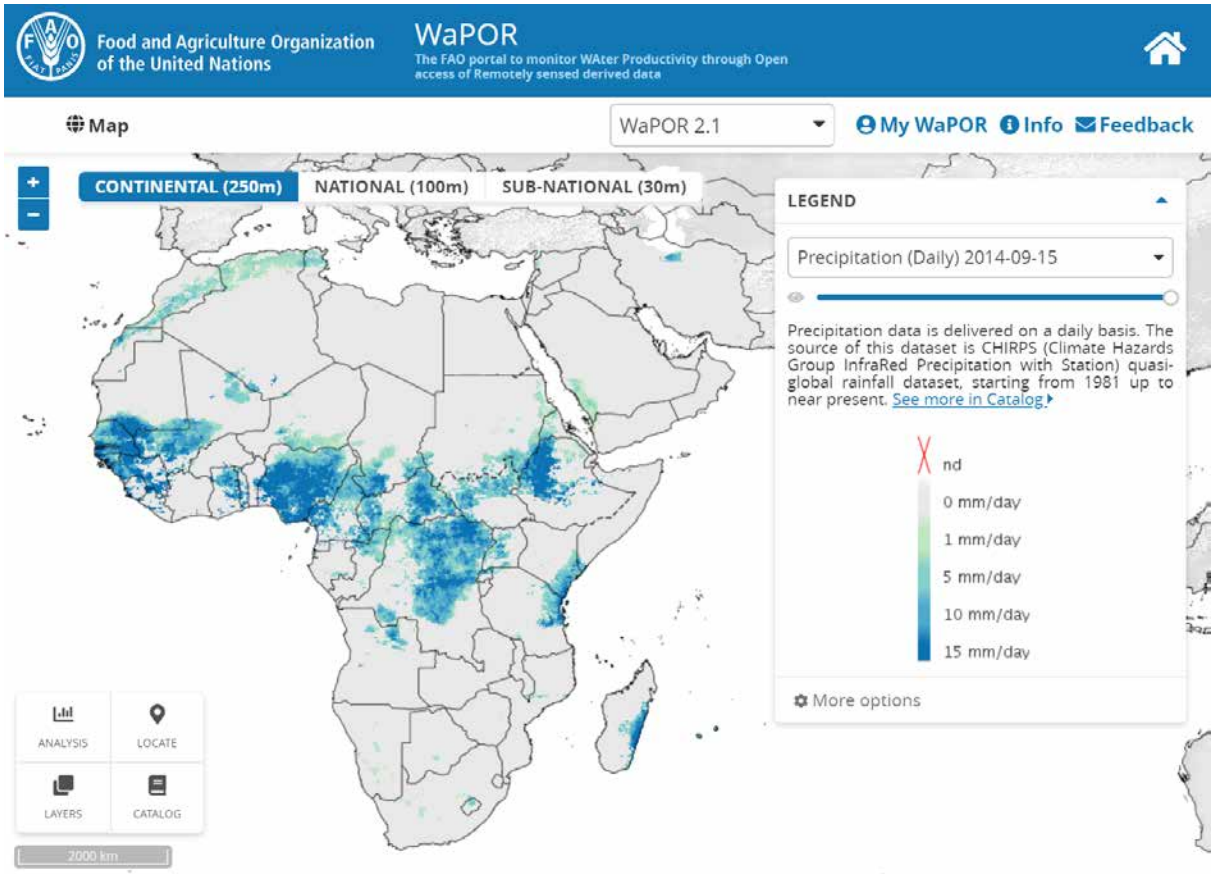
Data component	Unit	Range	Use	Temporal resolution
Reference ET	mm/day	0.5-16	Can be used to estimate potential ET for different crops by applying predefined crop coefficients	Daily (with aggregate at dekadal, monthly and annual time steps)

2.1.9 Precipitation

Description

Daily total precipitation (in mm) is provided using CHIRPS[18] data, an existing external data source that combines satellite observations with global models and measurements at local stations.[19] The daily precipitation data has a resolution of approximately 5km (0.05°). Figure 13 shows an example of the precipitation data component used for WaPOR.

Figure 13:
Example of precipitation data component at level 1, 5 km resolution (15 September 2014)



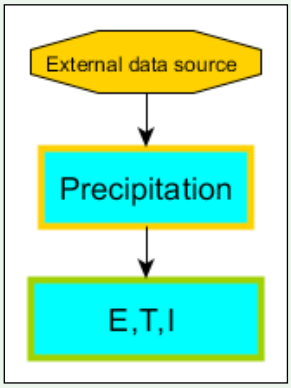
Source: FAO WaPOR, http://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor

[18] http://chg.geog.ucsb.edu/data/chirps/

[19] Specifics on data sources can be found in the data manual.

Methodology

BOX 12:
Precipitation in relation to other data components



- External data are used to produce the precipitation data component
- Precipitation is used as input to determine evaporation, transpiration and interception.

Precipitation data is obtained from an external source and delivered as a data component. It is also used as input to the E, T and I data components. An external large scale peer reviewed data source is used to obtain precipitation data. The data are checked for data gaps, and where these occur in areas that are important to WaPOR, e.g. agricultural areas, such gaps are filled with another suitable and comparable dataset.

Table 13:
Overview of Precipitation data component

Data component	Unit	Range	Use	Temporal resolution
Precipitation	mm/day		Provides rainfall estimates.	Daily

2.1.10 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 at level 3 identifies the main crop types of wheat, rice and maize as well as additional crops that cover more than 10% of the level 3 study areas (see Table 14).

The classes used for land cover mapping at levels 1, 2 and 3 as shown in Table 14 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.

The land cover maps at all three levels distinguish between irrigated and rainfed cropland. Data on agricultural land cover are important for evaluation of current land use practices as it can be coupled with water productivity data, enabling the comparison between different crops within a region, or the same crop between different regions. The level 1 and 2 land cover maps are annual and the level 3 land cover maps are produced for every dekad in the year.

Table 14:

Overview of land cover classes per level

Level 1	Level 2	Level 3
Cropland rainfed	Cropland rainfed	Maize
		Rice
		Wheat
		Specific crops (covering more than 10% of the area)
Cropland irrigated	Cropland irrigated	Maize
		Rice
		Wheat
		Specific crops (covering more than 10% of the area)
Cropland fallow	Cropland fallow	Cropland fallow
Tree cover	Tree cover	Tree cover
Tree cover: closed needleleaved, evergreen (>40%)	Tree cover: closed needleleaved, evergreen (>40%)	
Tree cover: closed broadleaved, evergreen (>40%)	Tree cover: closed broadleaved, evergreen (>40%)	
Tree cover: closed needleleaved, deciduous (>40%)	Tree cover: closed needleleaved, deciduous (>40%)	
Tree cover: closed broadleaved, deciduous (>40%)	Tree cover: closed broadleaved, deciduous (>40%)	
Tree cover: closed, mixed leaf type (broadleaved and needleleaved)	Tree cover: closed, mixed leaf type (broadleaved and needleleaved)	Tree cover (closed)
Tree cover: closed unknown forest type	Tree cover: closed unknown forest type	
Tree cover: open needleleaved, evergreen (15-40%)	Tree cover: open needleleaved, evergreen (15-40%)	
Tree cover: open broadleaved, evergreen (15-40%)	Tree cover: open broadleaved, evergreen (15-40%)	
Tree cover: open needleleaved, deciduous 15-40%)	Tree cover: open needleleaved, deciduous 15-40%)	

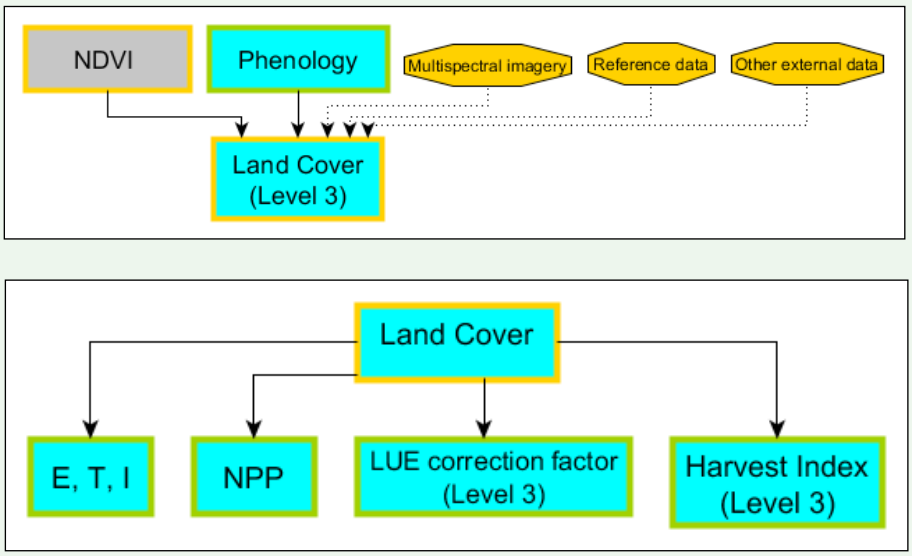
Tree cover: open broadleaved, deciduous 15-40%)	Tree cover: open broadleaved, deciduous 15-40%)	
Tree cover: open, mixed leaf type (broadleaved and needleleaved)	Tree cover: open, mixed leaf type (broadleaved and needleleaved)	Tree cover (open)
Tree cover: open unknown forest type	Tree cover: open unknown forest type	
Shrubland	Shrubland	Shrubland
Grassland	Grassland	Grassland
Wetland	Wetland	Wetland
Artificial	Artificial	Artificial
Bare soil	Bare soil	Bare soil
Water body	Water body	Water body

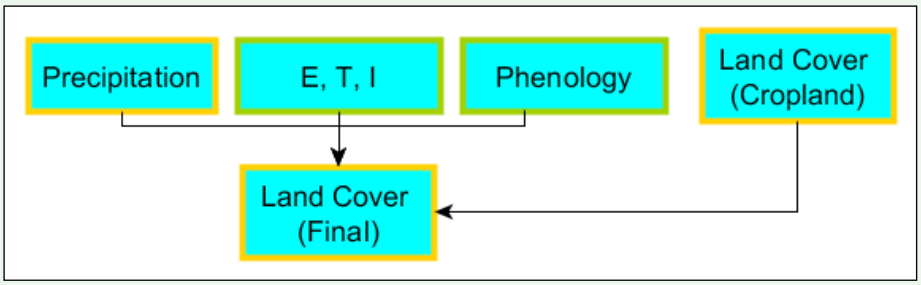
The result of a Land cover classification can be evaluated in several ways, where the use of confusion matrix is commonly applied. However, the development of methods for the accuracy assessment of products derived from moderate to low spatial resolution data is still being researched (Foody, 2002). 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).

Methodology

BOX 13:

Land cover classification in relation to other data components





- At level 1 and 2, the Copernicus Global Land Service 100m land cover product is used as base layer
- At level 3, Land cover classification makes use of dekadal NDVI time series and seasonal phenology information.
- At level 3, external data is required in the form of multispectral satellite imagery and other external ancillary datasets.
- Classifying land cover and crops requires a substantial amount of reference data. For level 3, this static input is derived from field work activities in the level 3 study areas.
- Final land cover map including rainfed/irrigated cropland classes is derived from combining phenology with Precipitation and ETI for the cropland areas (derived from an intermediate Land cover map).
- The LCC output is annual for level 1 and 2, and dekadal for level 3
- At level 3, the LCC output is used to derive the harvest index for specific crops.
- LCC output is used to determine the seasonal Light use efficiency (LUE) correction factor for use with NPP.

Level 1 and 2 land cover mapping: The Copernicus Global Land Service: land cover 100m

At level 1 and 2, the Copernicus Global Land Service 100m land cover product is used as base layer.

Within the Copernicus Global Land Service (CGLS), the dynamic land cover map at 100 m resolution (CGLS-LC100) was released in the spring of 2019. The product was generated by the Global component of the Land Service of Copernicus, the Earth Observation programme of the European Commission. The research leading to the current version of the product has received funding from various European Commission Research and Technical Development programs. The product is based on PROBA-V data provided by Belgian Science Policy Office (BELSPO) and distributed by VITO.

The CGLS Land Cover product provides a primary land cover scheme at three levels, 12 classes at level 1 up to 23 classes at level 3, according to the LCCS scheme. Next to these discrete classes, the product also includes continuous field layers for all basic land cover classes that provide proportional estimates for vegetation/ground cover for the land cover types. The map was provided for the reference year 2015 on a global scale, using the PROBA-V 100 m time-series, as well as a database of high quality training points and several ancillary datasets, reaching an accuracy of 80 % at the level 1 classification scheme. For more details on the CGLS-LC100m product, please refer to the Product User Manual.[20]

[20] https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_LC100_V2_I2.10.pdf

This CGLS-100m product was used as the base layer for the level 1 and 2 LC products.[21] For level 1, the maps were resampled to 250m using the *mode*-method. The "level 3" classes from the CGLS-100m were used, depicting more detail in the forest classes, i.e. making the distinction between open- and closed forest, as well as broad- and/or needle-leaf trees.

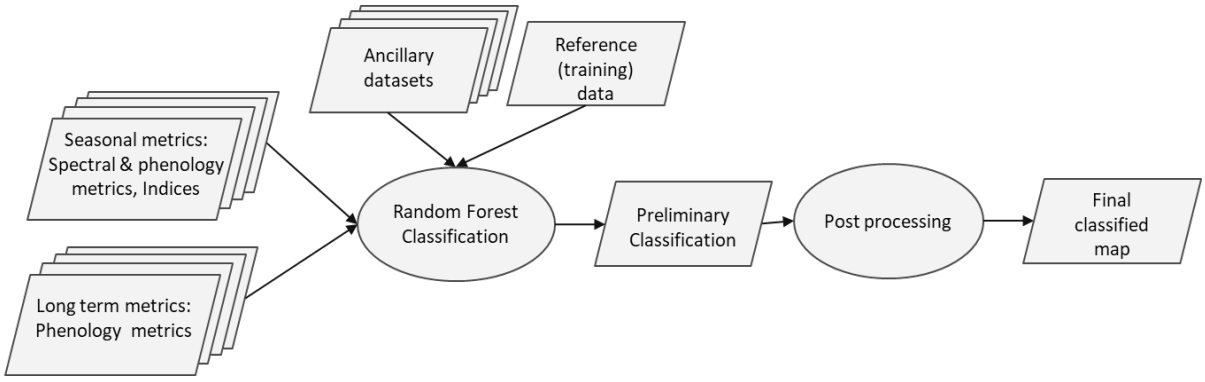
Level 3 land cover

The production of the Land cover classification data component at level 3 requires input from the phenology data component to demarcate the season, as well as dekadal NDVI data and spectral reflectance data. External reference (training) data are an important component of Land cover classification.

The general workflow for level 3 Land cover classification is shown in Figure 14. A supervised classification methodology is applied to assign a specific class to each pixel of the image. Training data consist of seasonal and long term metrics derived from dekadal NDVI time series, phenology and spectral reflectance data combined with reference data denoting the exact location of each of the classes specified in Table 14. The different components of the classification processing chain are discussed in the sections below.

Figure 14:

Schematic overview of the Land cover classification processing chain at level 3. Different types of reference data as well as dekadal NDVI and multispectral remote sensing inputs are used to train a machine learning classifier



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 for at least one moment in time during a growing season and distributed relatively evenly. Since the level 3 Land cover classifications are delivered on a dekadal basis[22], a large amount of reference data is required. The gathering of suitable reference data is 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 absence or 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.

[21] The global CGLS-100m land cover map for 2015 served as base layer for both Level 1 and 2, whereas the cropland class was further divided into irrigated, rainfed and fallow, on an annual basis.

[22] The dekadal land cover data is derived from seasonal land cover maps.

Fieldwork was conducted in each of the level 3 areas to collect reference data. In some areas additional existing historical reference data from external sources was used to add to the reference dataset.

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 metrics used at level 3 are listed in Annex 2.

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[23] when faced with large data volumes and complex feature spaces. Many of the current global land cover maps have been produced with ML, e.g. Globeland30, GlobCover, CCI. The classifier applied in this project, Random Forest, is therefore a machine learning algorithm. To increase the accuracy of the classifier for identifying urban areas and water, existing peer-reviewed specialised external datasets are included in the training dataset. These datasets are not solely based on spectral satellites, but combine for example, digital elevation models and radar data. Further details are provided in the data manual.

Post processing

After the ML classifier has produced an initial preliminary land cover map (on a seasonal basis), several post-processing steps are carried out to check and correct the land cover map. First, pixels in the Land cover classification maps for season 2 that coincide with an "out of season" classification in the phenology data are relabelled as "fallow". For all seasons, pixels classified as one of the main crops that have a low (<70%) concomitant land cover classification quality value are relabelled as "Mixed crop". (A description of the land cover quality layer is given below). Finally, temporal corrections are applied to the seasonal land cover for the entire time period. These rule-based corrections are based on the fact that certain land cover types are unlikely to change more than once over the entire historical time period, e.g. perennial crops and natural woodland can be seen as semi-permanent land cover which will not change into a seasonal crop and then later back into the original perennial crop or natural woodland. Pixels that exhibited such short-term temporary changes were relabelled to ensure a consistent pattern over time for semi-permanent land cover types. However if a change occurs and persists to the end of the historical time period, it is assumed that the vineyard, orchard or woodland was removed and replaced by a new land cover type. Similar rules were applied for urban areas. In a final step, the post-processed seasonal land cover maps are converted to dekadal land cover maps using the phenology.

[23] For example Maximum Likelihood

Irrigation mapping

As the CGLS-LC100m product does not make the distinction between rainfed and irrigated croplands, a separate workflow was designed based on the available information in WaPOR on the growing seasons, evapotranspiration, and precipitation. This methodology is also applied to the level 3 land cover maps. Important to note is that the basis of the irrigation mapping is a labelling methodology, meaning that the delineating of the LC cropland areas and the land use irrigated/rainfed are two separate and independent steps.

The basis of the irrigation mapping is the water deficit index (WDI), which compares within a growing season the available water inputs from precipitation, and the water outputs from the transpiration, evaporation and interception. Based on the phenology-information, the total seasonal ETIa and P are calculated. The WDI is then defined as the ratio of the seasonal P (Ps) and the seasonal ETIa ($ETIa_s$). For Ps, the accumulation starts one month before the start of the growing season to account for water already available in the soil at the beginning of the growing season.

$$WDI = \frac{P_s}{ETIa_s} = \frac{\sum_{SOS-3}^{EOS} P}{\sum_{SOS}^{EOS} ETI}$$

A WDI of 1 or larger means that water made available through P sufficed to sustain the water that was consumed by the crops ($ETIa_s$). A WDI<1 means a larger water consumption was observed in relation to the available water from precipitation over the growing season. For the labelling of irrigated areas, a threshold of WDI<0.9 was used to determine irrigation.

At level 1 and 2 a few additional steps were needed to go from a seasonal irrigation detection based on the WDI to the annual land cover maps where croplands were further specified in irrigated and rainfed croplands:

- Only areas where the ETIas was above 100mm were considered, to account for fallow areas.
- For a given target year, the croplands were labeled as irrigated if in a five-year window in at least two years an irrigated season was detected. For example, for the target year 2015, all WDI's in the period 2013 -2017 were evaluated, and needed to be below 0.9 in at least two years.
- This evaluation was also performed for fallow areas (i.e. ETIas <100mm), to label fallow croplands.

Complementary data layer: light use efficiency (LUE) correction factor (level 3 only)

This additional raster layer is delivered to the WaPOR database at level 3 to enable users to recalculate NPP and TBP at the end of the season when the correct land cover for the season is known. It contains light use efficiency (LUE) correction factors for all vegetated areas, where the (preliminary) NPP and TBP were calculated using the LUE values according to the land cover classes of the level 2 WaPOR land cover map. The LUE correction factor is the ratio between the actual LUE and the LUE applied for the NPP data at level 3. Once the seasonal Land cover classification for the level 3 area is known and the LUE correction factor is calculated, the user can multiply the (preliminary) NPP or TBP data with the LUE correction factor. The correction factor is 1 when the actual LUE is equal to the preliminary LUE used. In other cases it can either be higher or lower.

Complementary data layer: land cover classification quality layer

This additional raster layer is delivered to 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. All land cover maps contain a fraction of falsely classified pixels. The Land cover classification quality layer is mainly the result of the machine learning classifier quality output, depicting the probability of the classified pixel. 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. Additional flags are foreseen, which may be different between the levels. At level 1 and 2, pixels for which external data sources were used (and thus not specifically classified) are labelled as no-data (255), including permanent waters, urban areas and permanent snow and ice. At level 3, flags are used that indicate when pixels were reclassified during post-processing: a flag value of 241 indicates that a pixel was reclassified during post-processing whilst a flag of 240 indicates that a pixel was reclassified using a fixed mask (used in isolated cases).

Table 15:
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 14.	Annual (level 1 and 2) dekadal (level 3)
LUE correction factor	-	-	Used to recalculate NPP (and TBP) at the end of the season when the correct land cover for the season is known.	Seasonal (level 3 only)
Land cover classification quality	%	0-100, flags 240,241	Indicates the quality of the original ML classifier and whether a pixel was relabelled during post-processing.	Seasonal (level 3) Annual (level 1 and 2)

2.2 Intermediate data components

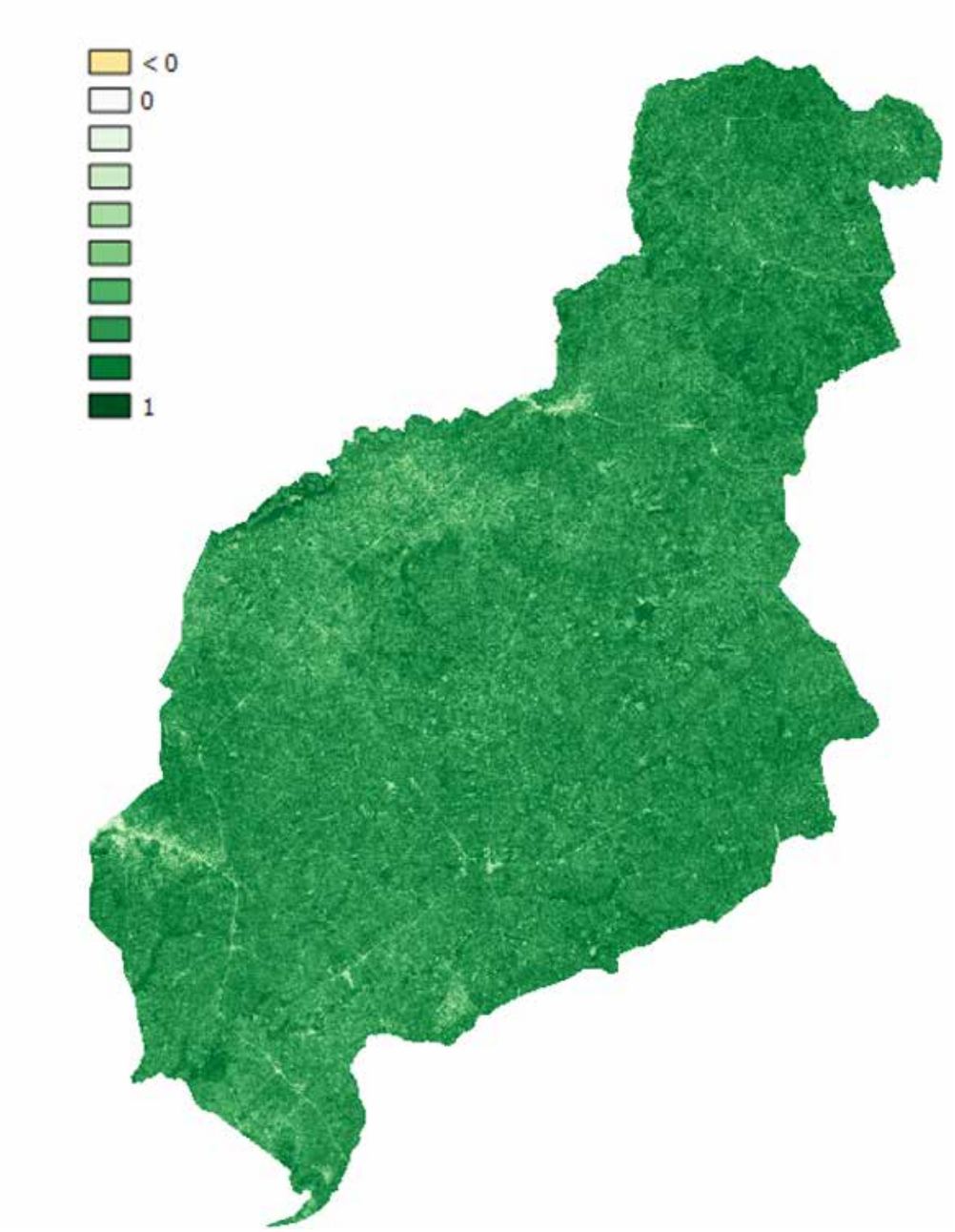
2.2.1 NDVI

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 and T. NDVI values range between -1 and 1. Vegetated areas have positive values closer to 1, bare soil/artificial surfaces have values of around 0, and water has negative NDVI values. Figure 15 shows an example of the NDVI intermediate data component.

Figure 15:

Example of NDVI intermediate data component at level 3 (composite for dekad 20, 2019). This intermediate data component is not published through WaPOR



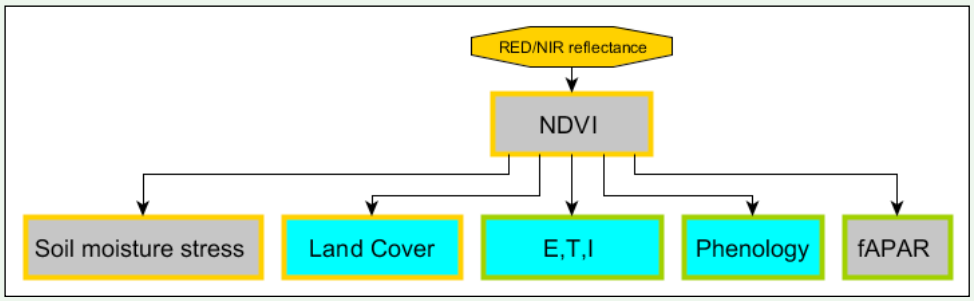
Source: this study

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).

Methodology

BOX 14:

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. 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.

At level 1 and 2 frequent satellite-based reflectances are converted to dekadal NDVI composites in the following way: First pixels that cannot be used for NDVI calculations are flagged as water, sea, cloud, and error pixels. Next, a "tile-based" dekadal synthesis is produced using a constrained[24] Max-NDVI compositing rule so that the dekadal NDVI comprises the "best" observation extracted from the available scenes within the dekad.

At level 1 and 2 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 16, 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.

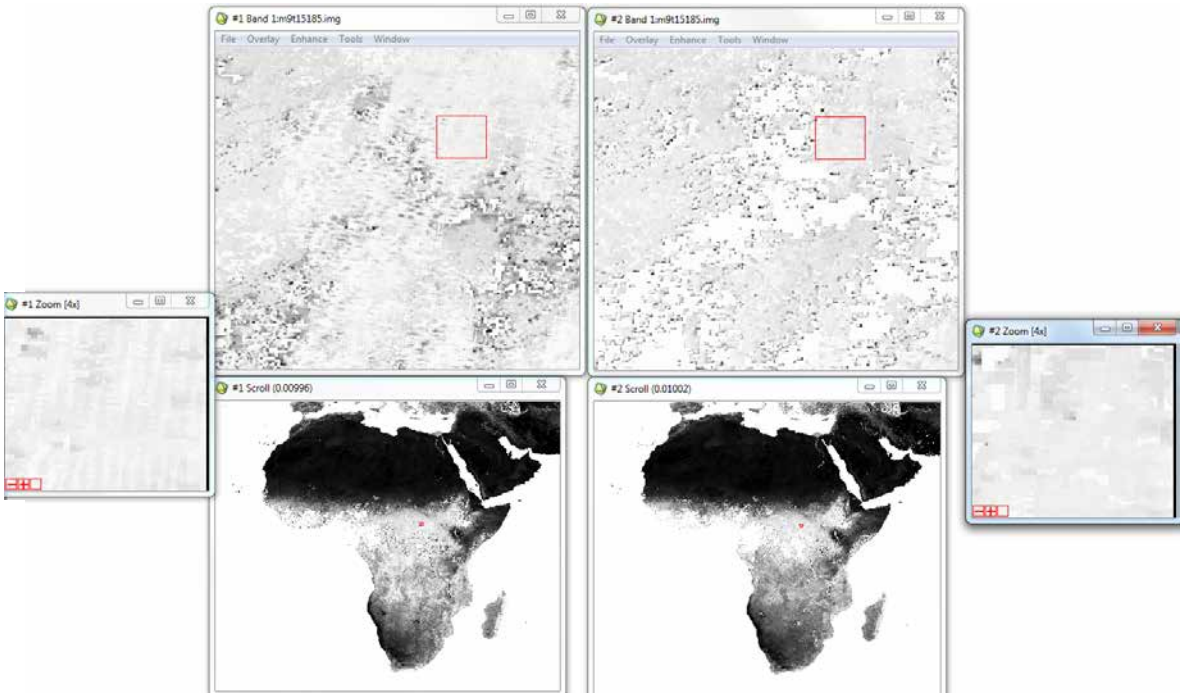
The pixels with missing and/or unreliable values in the dekadal NDVI series are then replaced by more plausible data through a process of interpolation based on the methodology[25] explained in Swets *et al.* (1999). The resulting images have no data gaps, see Figure 17 for an example.

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

[25] 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 16:

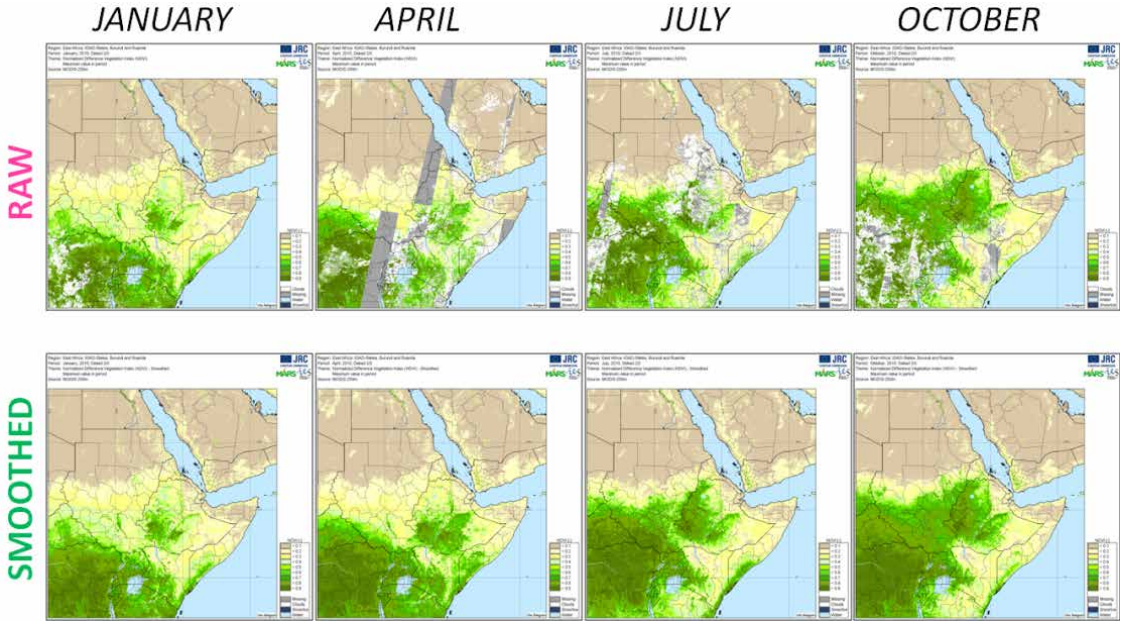
Examples of NDVI composite without (left) and with (right) view zenith angle constraint. Three different zoom-levels are shown for the same area. As can be seen in the image on the right, the angle constraint decreases the occurrence of artefacts, but increases the number of pixels without valid observation



Source: JRC-MARS (https://ec.europa.eu/jrc/en/mars)

Figure 17:

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



Source: JRC-MARS (https://ec.europa.eu/jrc/en/mars)

At level 3, NDVI is first calculated using instantaneous satellite data.[26] As at levels 1 and 2, 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 gap filling are less accurate. This is more likely to happen at level 3 due to the larger gap in satellite observations that occur within a dekad. As mentioned above, the larger the gaps due to cloud coverage, the lower will be the accuracy of the prediction as a date as close as possible from the missing data should be used for resconstruction to ensure similar spectral characteristics. This is not always the case in level 3 areas where cloud occurrence is high.

Data layers that indicate the quality of each of the dekadal NDVI data composites are produced for all three levels (see description of the methodology below).

Complementary data layer: NDVI quality layer

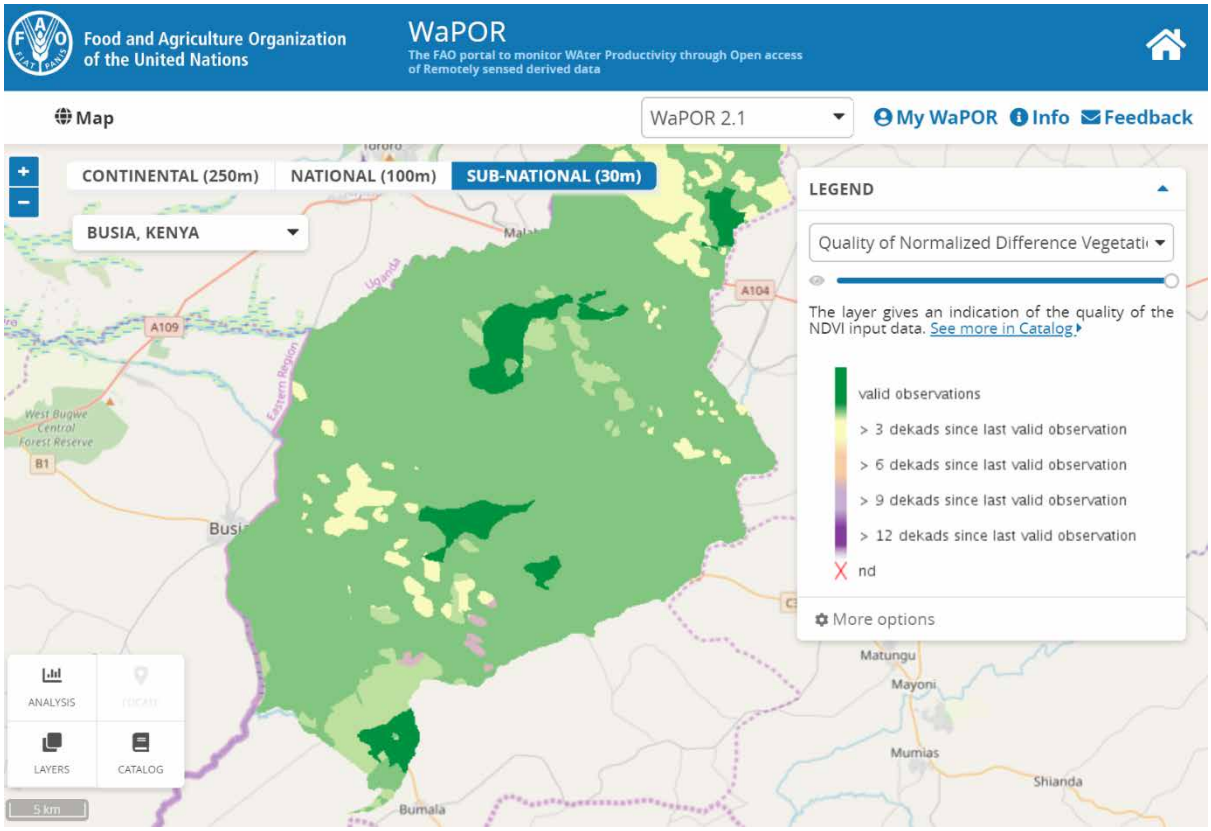
The quality layer is produced during the compositing of the NDVI. The quality index (QI) for every pixel in each dekad depicts the number of dekads since the last observation date used for reconstruction. When a number of consecutive dekads have not had direct observations for calculating the NDVI, the quality layer pixel value of the consecutive dekadal NDVI quality layers will represent the cumulative number of dekads since a valid observation. The user can use consecutive NDVI quality layers to calculate the length of the gap between valid observations. In the case that the pixel observation is usable and no reconstruction was needed, the QI is set to 0 (ideal situation).

An additional long-term NDVI quality layer is produced as a summary of the quality for the entire time period (2009 to last completed year, or 2014 to 2019 for Proba-V at level 2). Pixel values indicate the percentage of good NDVI observations over the entire time period. Whereas the dekadal NDVI quality layer shows the quality of the data for a specific time of year, the long-term NDVI quality layer has been added to give users a quick overview of the areas that tend to suffer more cloud cover.

The NDVI quality layer not only represents the quality for the NDVI, but also for evapotranspiration, fAPAR and NPP as these are calculated using the NDVI composites as input. Figure 18 shows an example of the NDVI quality data layer at level 3.

[26] Whereas satellite observations used at level 1 and 2 are made frequently (daily), higher resolution satellite data used for NDVI at level 3 have a much lower temporal resolution, with image acquisitions taking place approximately every 5-16 days.

Figure 18:
Example of the NDVI quality layer at level 3



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

Table 16:
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	dekads		Indicates quality of NDVI composite for a specific dekad.	Dekadal
NDVI quality layer (combined)	%	0-100	Indicates quality of NDVI for the entire time period of the data component.	One-off

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.[27]

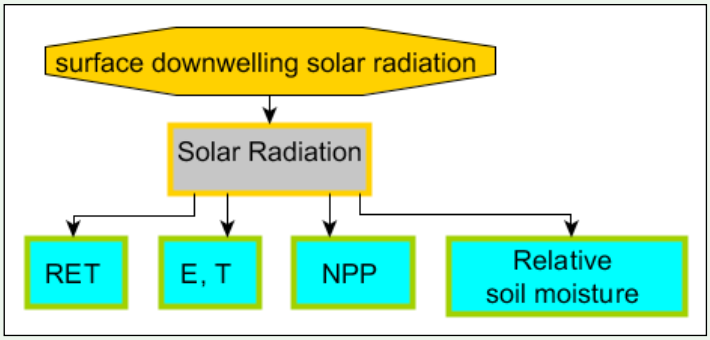
[27] Also referred to as Top of Canopy (TOC).

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 15:

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[28] all determine the incidence angle of the sun at a specific location. Topographical features such as slope and aspect can be extracted from a digital elevation model (DEM) and 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

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

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, Trezza and Tasumi (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 17:
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, 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.

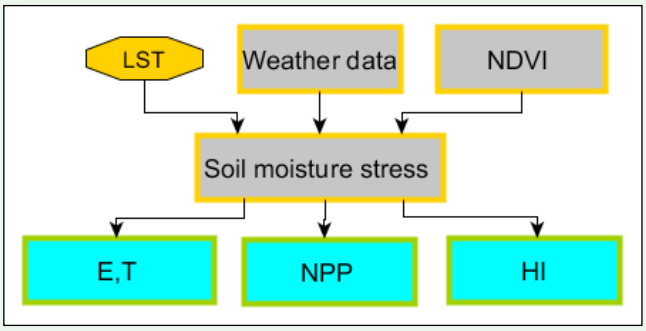
Soil moisture content varies strongly in time and place. Within the WaPOR area of interest extremes occur in northern Africa and the Middle East where relative soil moisture content is very low throughout the year (with the exception of areas close to rivers) and the equatorial region which is characterised by high relative 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 no stress and 1 maximum 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 levels 1 and 2.

Methodology

BOX 16:

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 at all levels.
- Soil moisture stress is incorporated in the calculation of NPP at all levels.
- Soil moisture stress is incorporated in the calculation of Harvest Index (Level 3 only).

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[29] (Carlson, 2007). External input data required are visual/NIR and thermal imagery.

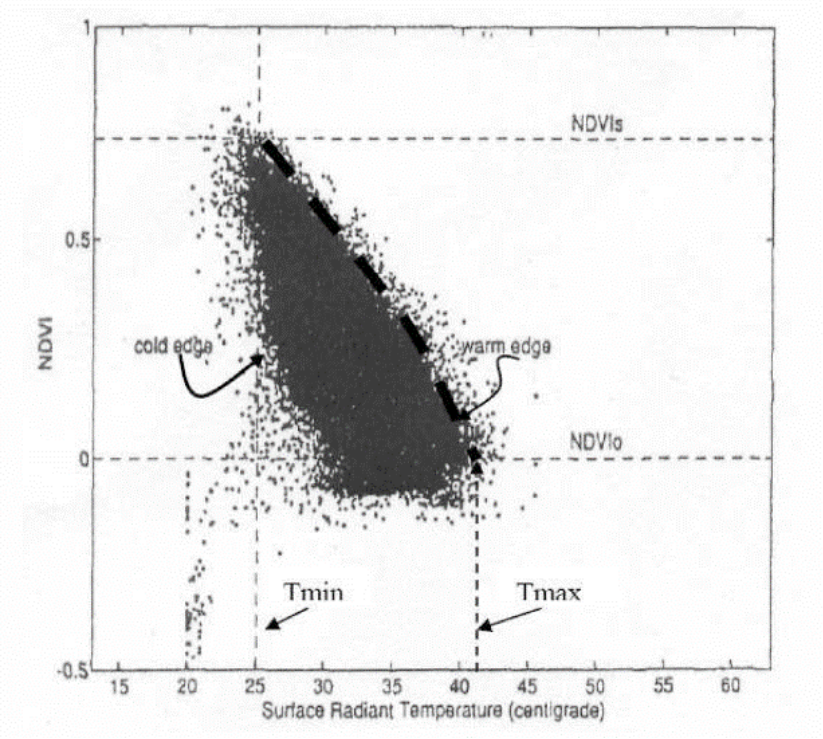
[29] An alternative approach is based on the use of radar imagery from ASCAT. WaPOR data production partners apply the LST method as it has a higher resolution and therefore provides a better representation of the spatial variability of soil moisture content. It is also a better indicator for the water 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.

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 19). 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 19:

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



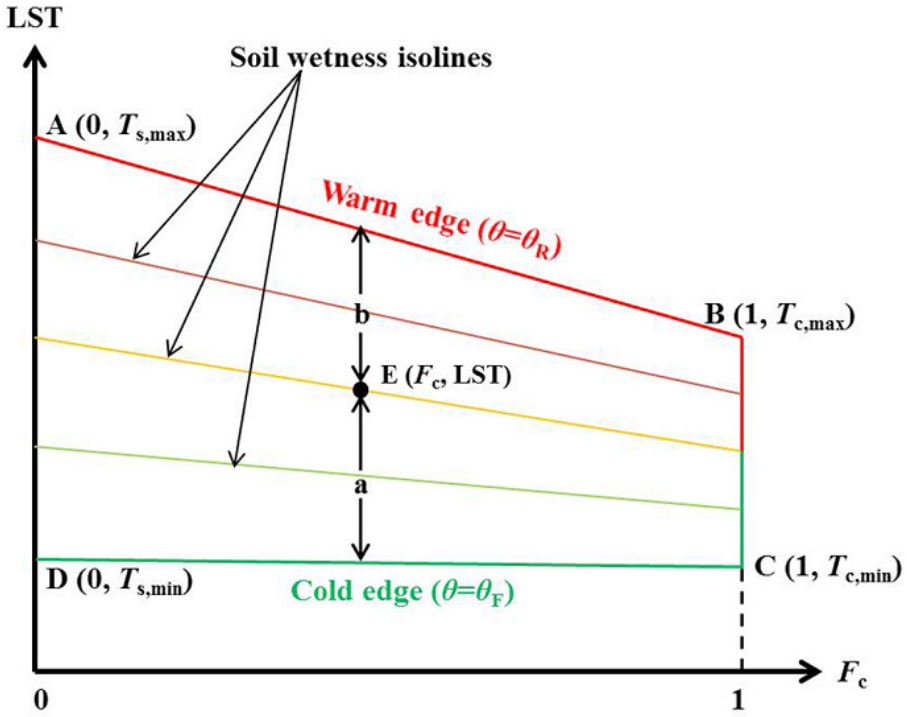
A drawback of this method is that it requires calibration by manual selection of references pixels for each thermal image. This introduces subjectivity through the selection process and makes it difficult to operationalize for a larger area. This problem was overcome by the method developed by Yang *et al.* (2015). The original triangle method was modified by introducing the effect of stomatal closure of 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 20, taken from the improved trapezoid method[30] of Yang *et al.* (2015).

[30] 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.

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 20.

Figure 20:

The trapezoidal vegetation coverage (Fc) / 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.

Assuming no sensible heat flux, the temperature of the cold edge (C and D) T_{min} is approximated by the wet-bulb temperature (T_{wet}) for bare soil (corner point D) and air temperature for full vegetation cover (corner point C) at around the same time as when the LST is measured. A linear relationship between the temperatures at corner point D and corner point C provide for all other vegetation covers. 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 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 (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 Sánchez, Caselles and Kustas (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 relative 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) has 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 relative soil moisture content is determined for both the top soil and the root zone. Therefore the same relative 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 relative 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 relative 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.

Relative soil moisture content composites

At level 3, relative soil moisture is first determined on instantaneous basis i.e. on the dates of the Landsat acquisitions from instantaneous LST images with cloud covered parts masked out. Corresponding gaps in the soil moisture time series are filled using a smoothing method based on the Savitzky-Golay filter (Chen *et al.* 2004). The method was originally developed by the authors to smooth out noise in NDVI time series and deal with cloud contamination by approaching the data to the higher NDVI enveloppe. To smooth the soil moisture time series, a modified and more 'neutral' version of the method is applied: the linear interpolation is still followed by a Savitzky-Golay smoothing but the iterative process by which the values are approached to a higher enveloppe is skipped. It should be noted that the results of the smoothing are used on masked pixels only whereas unmasked pixel values are kept to maximize accuracy. The instantaneous soil moisture is then converted to dekadal output.

Soil moisture stress

The relative soil moisture content determines the availability of water for evaporation and transpiration. Whether this is reduced due to a shortage can be calculated 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.

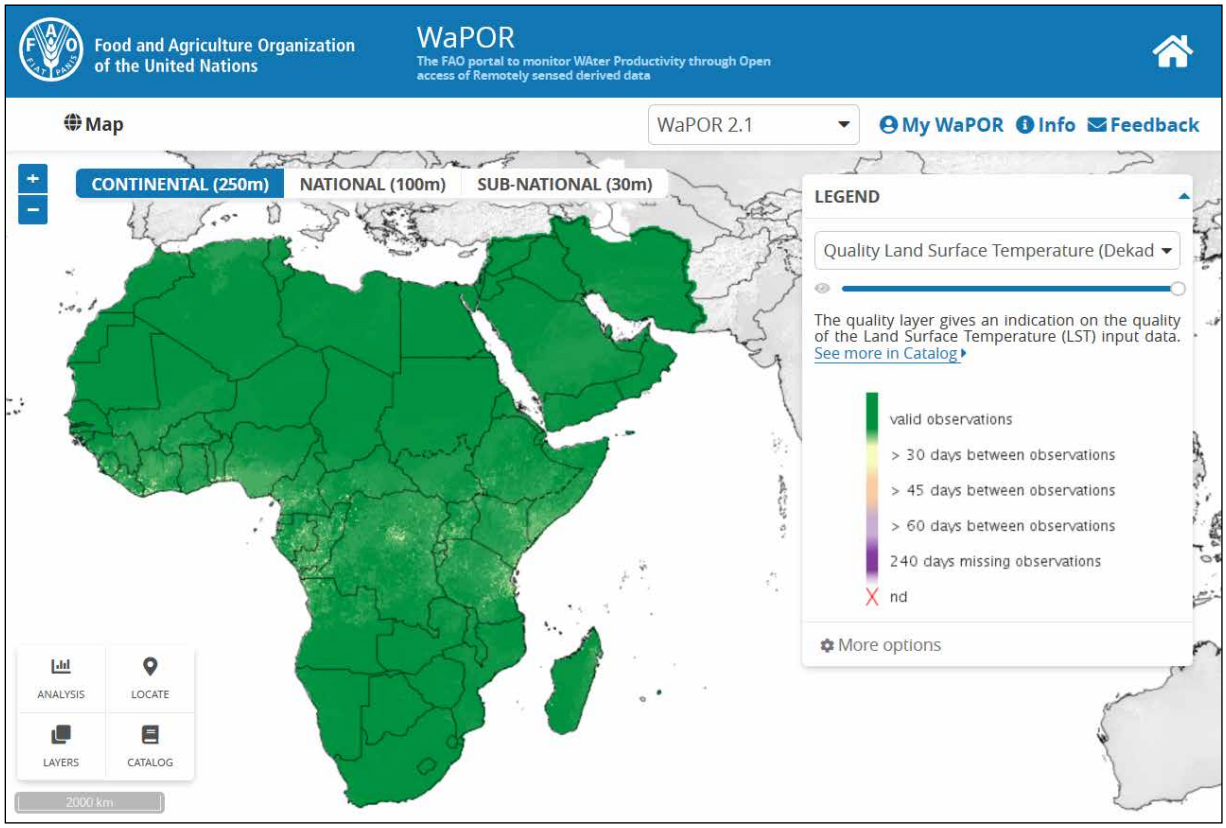
This soil moisture stress factor, ranging between 1 and 0, is used as input for the E and T to reduce evapotranspiration.

Complementary data layer: land surface temperature quality layer (level 1 and 2)

Cloud cover causes data gaps in the input data required for the calculation of relative 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 land surface temperature quality layer indicates the number

of days since the last observation, given on a pixel-by-pixel basis. Figure 21 shows an example. The LST quality layer is only produced for level 1 and 2 because at level 3 land surface temperature is obtained from instantaneous Landsat data and its quality can thus inferred from the L3 NDVI quality layer.

Figure 21:
Example of Land surface temperature data quality layer at level 1 (2020, dekad 12)



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

Table 18:
Overview of the (intermediate) data components related to soil moisture

Data component	Unit	Range	Use	Temporal resolution
Relative 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

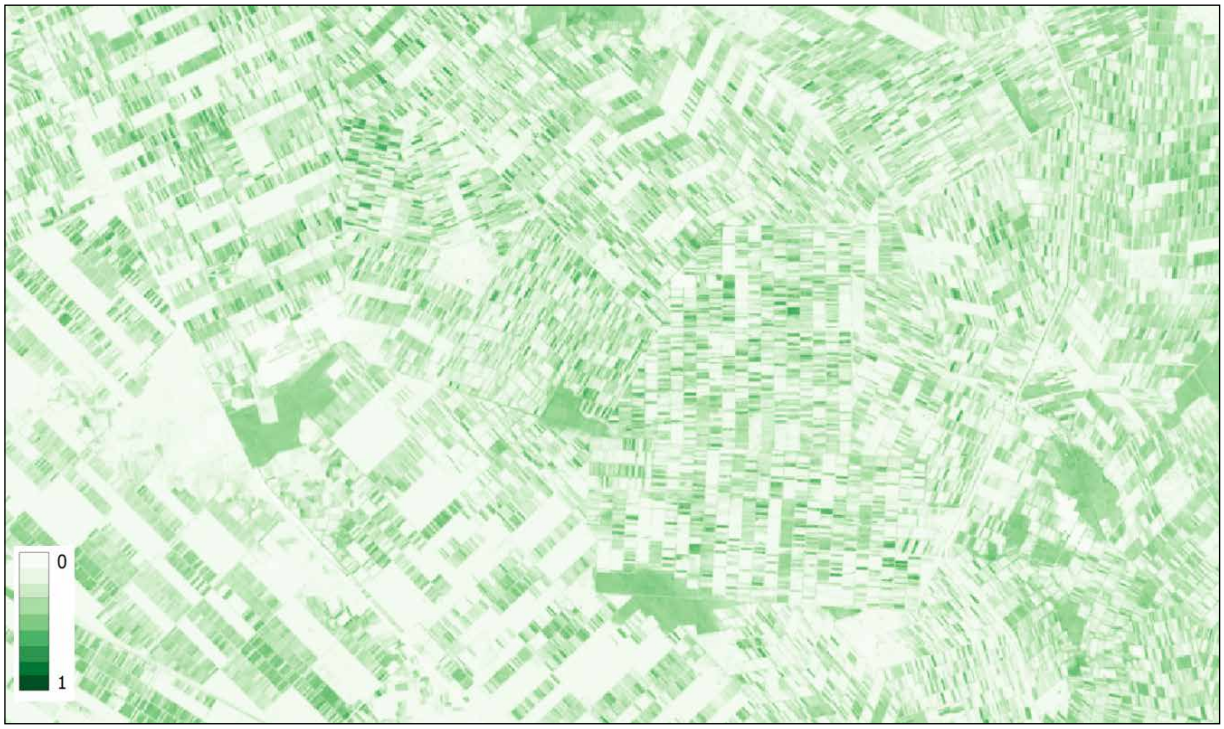
Description

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) and E and T (albedo). Figure 22 shows an example of the fAPAR intermediate data component at level I. fAPAR values range from 0 to 1. Figure 23 shows an example of the albedo intermediate data component at level I. 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.

Figure 22:

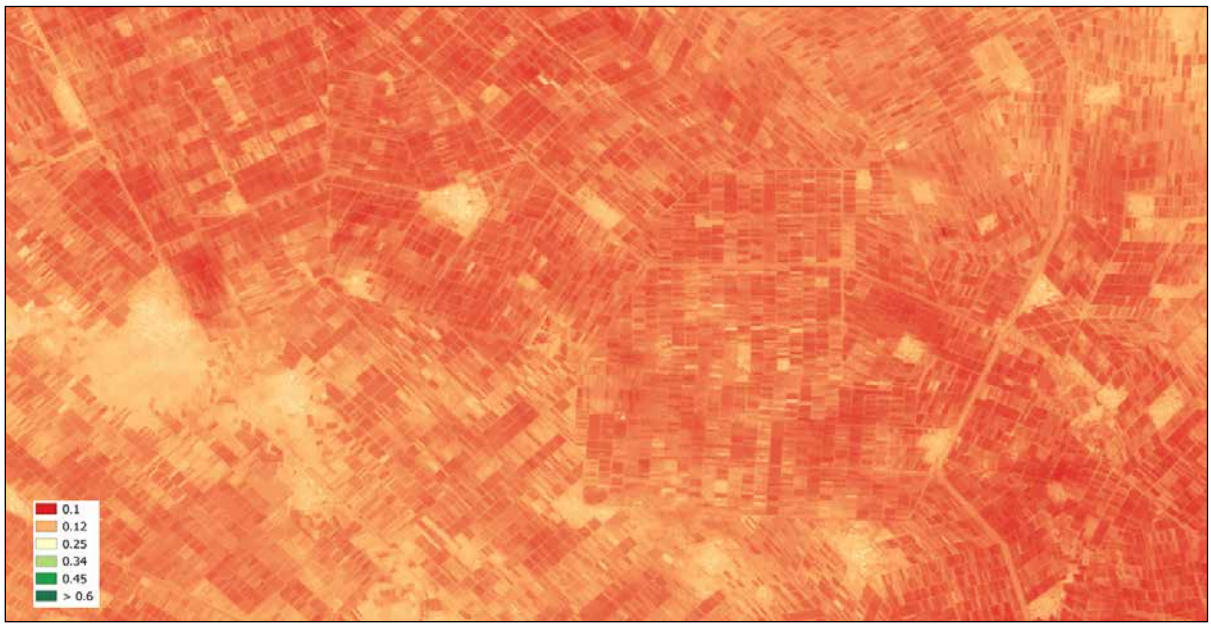
Example of fAPAR intermediate data component at level 3 (2014, dekad 25). This intermediate data component is not published through WaPOR



Source: this study

Figure 23:

Example of albedo intermediate data component at level 3 (2019, dekad 18). This intermediate data component is not published through WaPOR

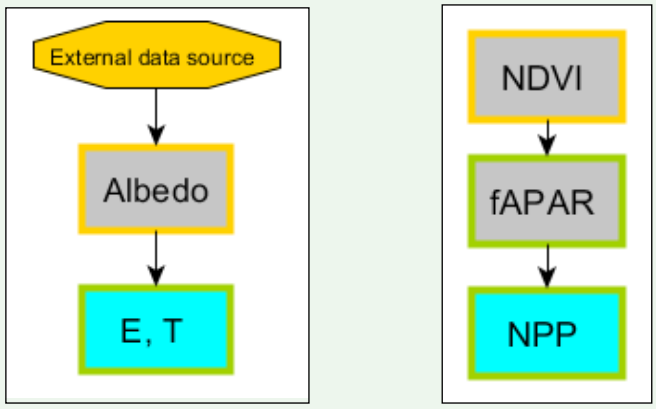


Source: this study

Methodology

BOX 17:

fAPAR and albedo in relation to other data components



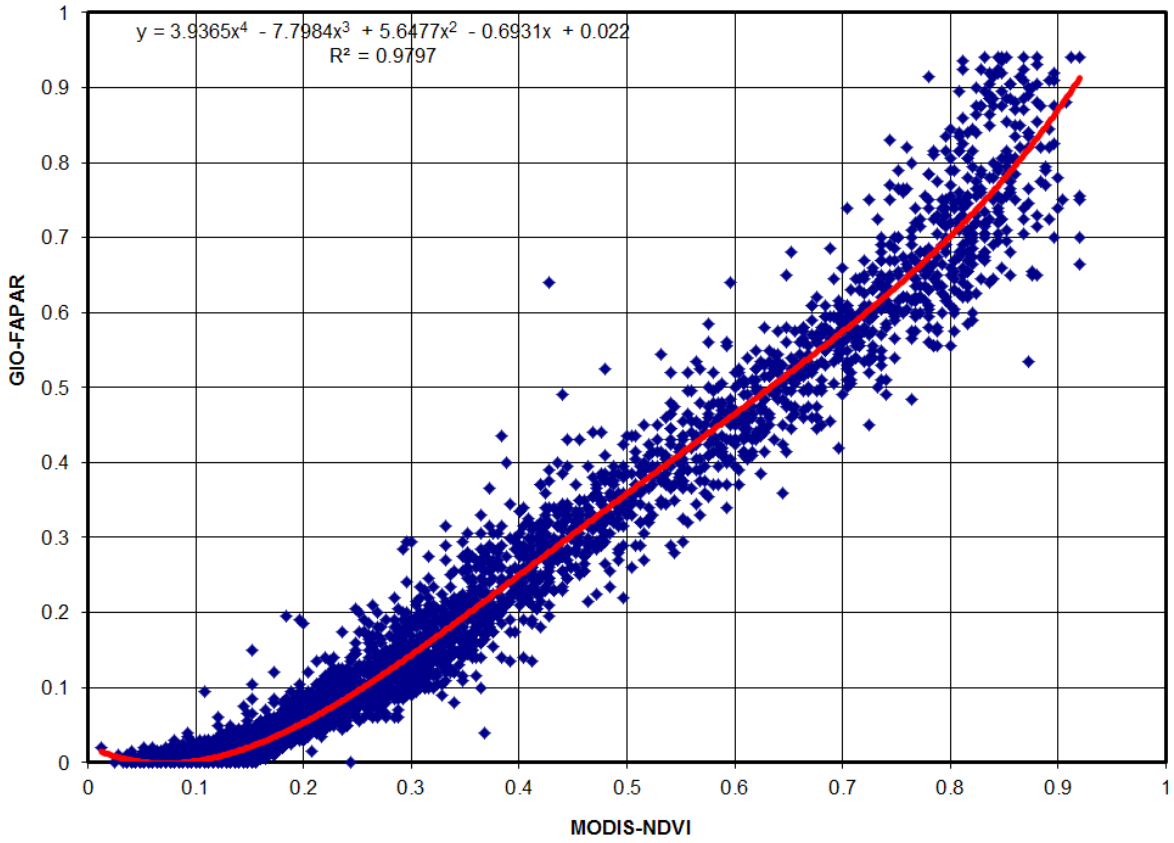
- External data sources are used as input.
- fAPAR is used as input to various data components, e.g. NPP, E, T and intermediate data components such as soil moisture.
- Surface albedo is used as input to produce E and T

fAPAR

fAPAR at level 1 and 2 is estimated by using a direct relationship between the NDVI and a global fAPAR product, Figure 24 shows an example. To ensure consistency between the levels, the fAPAR for level 3 is derived using the same direct relationship as for level 2 (see Figure 24). Further details of the processing are given in the data manual.

Figure 24:

Example of the relation between MODIS NDVI and the Copernicus* fAPAR product with data from nine dates between 2014 and 2016 (dekads 4, 16 and 28 from 2014-16)



*http://land.copernicus.eu/global/products/fapar

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 $r_o = \Sigma w_i . r_i$ (summation over the i bands), with r_i and w_i the spectral reflectance and weight of the i-th band. Note that $\Sigma w_i = 1$.

For level 3, gaps due to cloud coverage and other sources of missing data are filled using a smoothing method based on the Savitzky-Golay filter (Chen *et al.* 2004). The method was originally developed by the authors to

smooth out noise in NDVI time series and deal with cloud contamination by approaching the data to the higher NDVI enveloppe. The method used here was modified to approach the lower enveloppe of the albedo time series as cloud occurrence lowers the NDVI but increases the albedo. Smoothing also allows to correct for atmospheric variabilty and incidence angle. After smoothing, the instantaneous albedo is converted to dekadal output.

Table 19:
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 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.

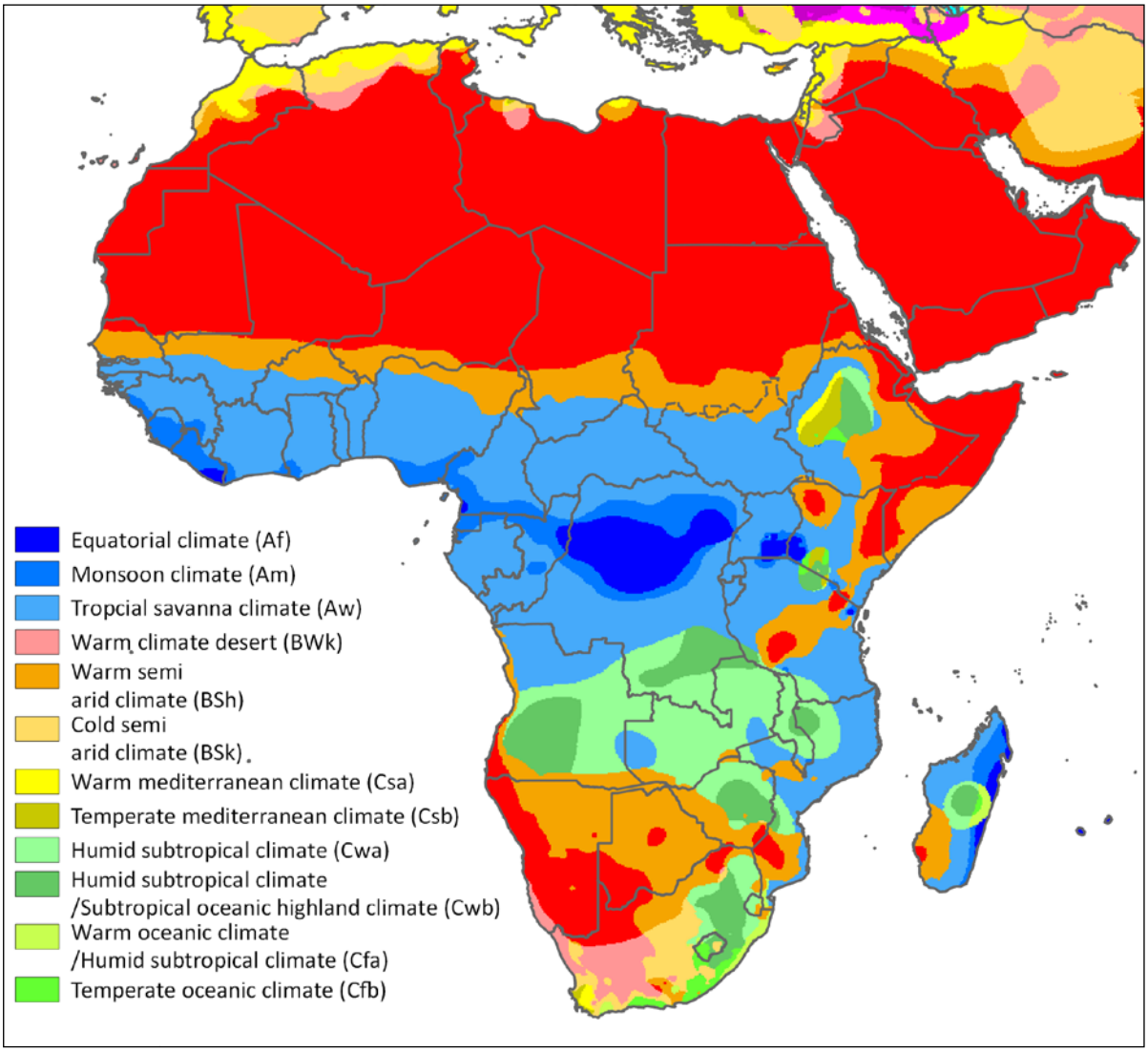
Transmissivity is discussed in section 2.2.2 and precipitation is discussed in section 2.1.9. The acquisition of temperature, wind speed and humidity data is discussed below. Although these parameters are routinely measured by most meteorological stations around the world the number of meteorological 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. Figure 25 shows the different climate zone according to Köppen, with the desert areas in Northern Africa and the Arabian Peninsula and the tropical climate zone around the equator forming extremes. Typical values for temperature, humidity and wind therefore vary greatly within the area.

Air temperature, humidity and wind speed 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.

Figure 25:

Köppen climate classification for (most of) the area of interest



Source: https://commons.wikimedia.org/ wiki/File:Africa_Koppen_Map.png

Figure 26:

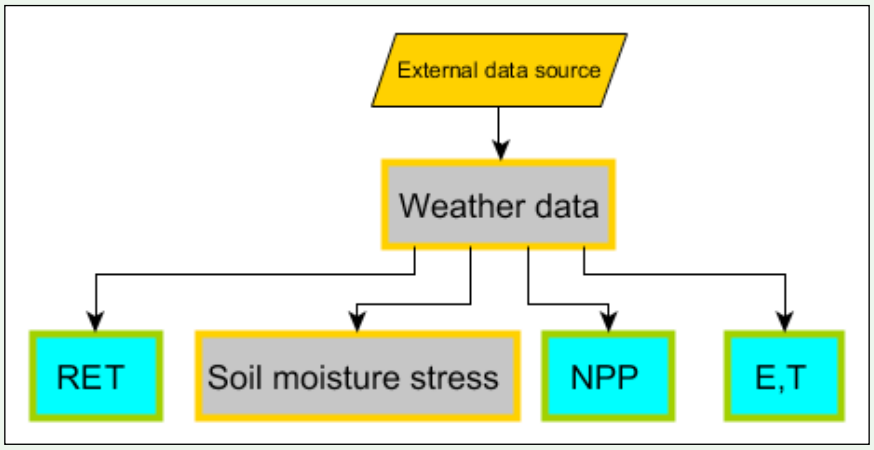
Example of weather data at level 1 for maximum daily air temperature on 30-07-2019, these are intermediate data components are not published through WaPOR

Source: this study

Methodology

BOX 18:

Weather data in relation to other data components



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

Temperature, specific 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.

Weather data used as input to the evapotranspiration calculations are aggregated to an average daily value and instantaneous weather data used as input for soil moisture is derived by linear interpolation of hourly data. The coarse resolution weather data are resampled to match the resolution of the level. In order to produce smooth meteorological data fields, specific 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 in 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 27 shows an example where a DEM was used to resample coarse resolution global temperature data. The Beqaa 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 27:
Example of coarse resolution global temperature data resampled for the Beqaa valley (circled) using a DEM. This example uses GEOS-5 temperature data

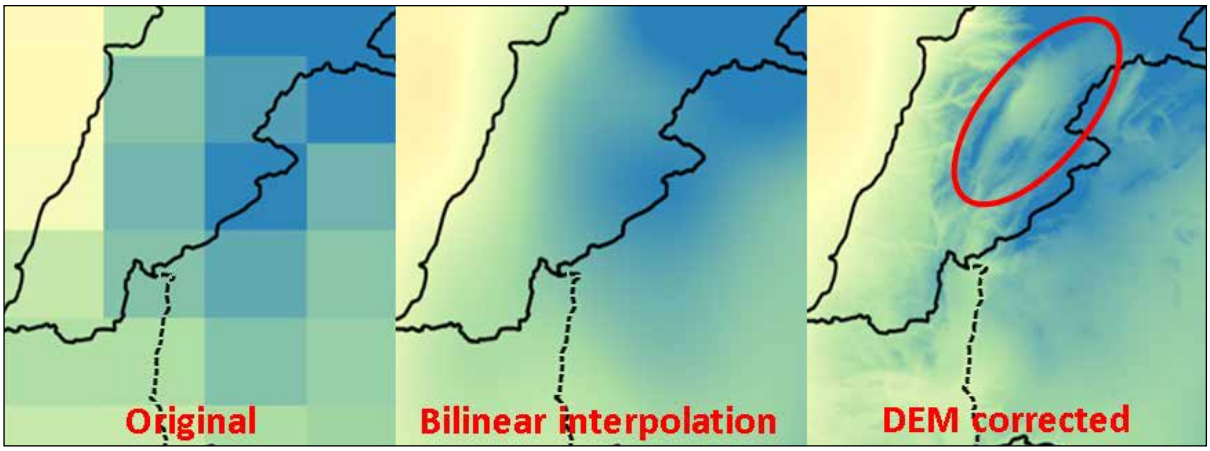


Table 20:

Overview of intermediate data components related to weather

Data component	Unit	Range	Use	Temporal resolution
$T_{min,}$ T_{max} and T_{ave}	K		Used to calculate E, T, RET, NPP.	Daily
T_{hourly}	K		Hourly temperature data is interpolated to obtain instantaneous temperature for calculating soil moisture.	Daily
Specific humidity	g/kg of air		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 tables of sensors used in WaPOR v2.0, L1, L2, L3

L1 Data component	Input data components	Sensor	Data product	Comment
Evaporation, Transpiration, Interception	Precipitation		CHIRPS v2, CHIRP	
	Surface albedo	MODIS	MOD09GA, MOD09GQ	
	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	NDVI	MODIS	MOD09GQ	
	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	MODIS	MOD09GQ	
	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	Precipitation		CHIRPS v2, CHIRP	
	Land Cover		WaPOR LCC product	
Reference ET	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	Solar radiation		SRTM	DEM
		MSG		Transmissivity
Precipitation	-	-	CHIRPS v2, CHIRP	
NDVI Quality layer	NDVI	MODIS	MOD09GQ	
LST Quality layer	-	MODIS	MOD11A1, MYD11A1	
Land Cover map			Based on Copernicus Land cover map, 2015	Resampled to L1 resolution, Irrigation mapping added

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. Sentinel 2 data resampled to 100 m replace PROBA-V, starting in 2020.
	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. Sentinel 2 data resampled to 100 m replace PROBA-V, starting in 2020.
	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	

L2 Data component	Input data components	Sensor	Data product	Comment
Phenology	NDVI	PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, resampled to 100m. Sentinel 2 data resampled to 100 m replace PROBA-V, starting in 2020.
Land cover classification			Based on Copernicus Land cover map, 2015	Irrigation mapping added
NDVI Quality layer	NDVI	PROBA-V		PROBA-V data are available from March 2014, data for earlier dates uses MODIS MOD09GQ, resampled to 100m. Sentinel 2 data resampled to 100 m replace PROBA-V, starting in 2020.
LST Quality layer	-	MODIS	MOD11A1, MYD11A1	

L3 Data component	Input data components	Sensor	Data product	Comment
Evaporation, Transpiration, Interception	Precipitation		CHIRPS v2	
	Surface albedo	Landsat-5,7, 8		Landsat data is available over the whole time period.
	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	NDVI	Landsat-5,7, 8		Landsat data is available over the whole time period.
	Soil moisture stress	Landsat-5,7, 8		Land surface temperature and NDVI from Landsat data available over the whole time period.
		MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	Solar radiation		SRTM	DEM
		MSG		Transmissivity
	Land Cover	Landsat-5,7, 8		NDVI from Landsat data available over the whole time period.

L3 Data component	Input data components	Sensor	Data product	Comment
NPP	Solar radiation		SRTM	DEM
		MSG		Transmissivity
	Soil moisture stress	Landsat-5,7, 8		Land surface temperature and NDVI from Landsat data available over the whole time period.
		MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	fAPAR	Landsat-5,7, 8		NDVI from Landsat data available over the whole time period.
	Weather data (temp, specific humidity, wind speed, air pressure)	MERRA/ GEOS-5		MERRA used up to start of GEOS-5 (21-2-2014)
	Precipitation		CHIRPS v2	
	Land Cover	Landsat-5,7, 8		NDVI from Landsat data available over the whole time period.
Phenology	NDVI	Landsat-5,7, 8		Landsat data is available over the whole time period.
Land cover classification	Phenology	Landsat-5,7, 8		Input parameters based on Landsat NDVI time series data, reflectance and phenology. (See Annex 2)
NDVI Quality layer	NDVI	Landsat-5,7, 8		

Annex 2: Input metrics used for level 3 Land cover classification

1. Seasonal metrics (defined by phenology data for the target season):

- Derived from the dekadal NDVI dataset:
  - Percentiles: p5, p10, p25, p50, p75, p90, p95
  - IQR = (p75-p25)
  - Average,
  - Standard deviation,
  - Coefficient of variation
  - Minimum NDVI (that is not <0)
  - Dekad number of min NDVI
  - Maximum NDVI
  - Range (Max - min)
  - Dekad number of max NDVI
  - NDVI gradient up (first min value of the season to max NDVI)
  - NDVI gradient down (max NDVI to last min value of the season)

2. Spectral reflectances for the Landsat image with the max NDVI over the season:

  - All bands excluding thermal bands
  - NDWI = (NIR-SWIR)/(NIR+SWIR)

3. Long-term metrics:

- Derived from the dekadal NDVI dataset:
  - Percentiles over time: p5, p10, p25, p50, p75, p90, p95
  - IQR = (p75-p25)
  - Average,
  - Standard deviation,
  - Coefficient of variation

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 ETo 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).

Braden, H., *Energiehaushalts- und Verdunstungsmodell für Wasser- und Stoffhaushalts-untersuchungen landwirtschaftlich genutzter Einzugsgebiete*. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft, (1985), 42, 254-299.

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.

Copernicus Service information (2019): https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_PUM_LC100_V2_I2.10.pdf

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.

FAO and IHE Delft (2019). *WaPOR quality assessment. Technical report on the data quality of the WaPOR FAO database version 1.0*. Rome. 134 pp.

FAO and FRAME (2020). *WaPOR V2 quality assessment. Technical report on the data quality of the WaPOR database version 2* (in press).

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.

von Hoyningen-Hüne, J., *Die Interception des Niederschlags in landwirtschaftlichen Beständen*. Schriftenreihe des DVWK, 1983, 57, 1-53.

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.

Monteith, J., & Unsworth, M. (2013). *Principles of environmental physics.* Academic Press.

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.

Šúri, M., & Hofierka, J. (2004). A new GIS-based solar radiation model and its application to photovoltaic assessments. *Transactions in GIS*, 8(2), 175-190.

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).

Swinnen, E., Van Hoolst, R., Eerens, H., (2018). Gio Global Land Component - Lot I "Operation of the Global Land Component", *Algorithm Theorethical Basis Document, Dry Matter Productivity (DMP) Version 2.*

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 Environmen*t, 83(3), 376-399.

Villalobos, F. J., & Fereres, E. (**Eds.**). (2016). *Principles of agronomy for sustainable agriculture*. Springer.

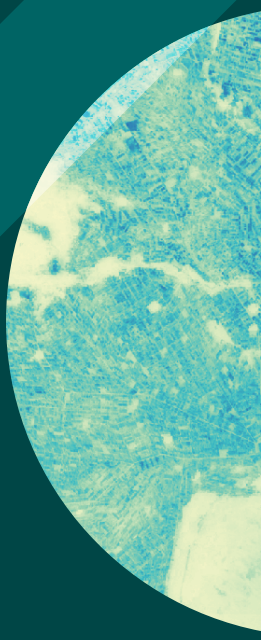
Wallace, J. S., Gash, J. H. C., McNeil, D. D., & Sivakumar, M. V. K. (1986). *Measurement and prediction of actual evaporation from sparse dryland crops. Scientific Report on Phase II of ODA Project*, 149.

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.

WaPOR Database methodology

Version 2 release



The FAO portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR) provides, as of today, access to 11 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. WaPOR Version 1 became available starting from June 2018 and WaPOR Version 2 was launched in June 2019. Each version of the data was improved based on extensive internal and external validation and quality assessment. This document describes the methodology used to produce Version 2 of the data at the 250m (Level 1), 100m (Level 2) and 30m (Level 3) resolution distributed through the WaPOR portal.

ISBN 978-92-5-132981-8

9 789251 329818

CA9894EN/1/09.20