EVALUATING THE QUALITY OF REMOTE SENSING-BASED AGRICULTURAL WATER PRODUCTIVITY DATA

Megan Blatchford

EVALUATING QUALITY OF REMOTE SENSING-BASED AGRICULTURAL WATER PRODUCTIVITY DATA

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof.dr.ir. A.Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Thursday 07 October 2021 at 14.45 hrs

by

Megan Blatchford

born on 16 May 1989.

in Nambour, Australia

Graduation committee

This thesis has been approved by

A/Prof. dr. ir. Chris Mannaerts, supervisor
Dr. Yijian Zeng, co-supervisor

ITC dissertation number 401
ITC, P.O. Box 217, 7500 AE Enschede, the Netherlands

ISBN 978-90-365-5243-1
DOI 10.3990/1.9789036552341

Cover designed by Job Duim
Printed by CTRL-P Printing

Graduation committee:

Chairman/Secretary
Prof.dr.ir. F.D. van der Meer

Supervisor
A/Prof. dr. ir. Chris Mannaerts, supervisor

Co-supervisor
Dr. Yijian Zeng, co-supervisor

Members

Prof.dr.ir. Charlotte de Fraiture	UN-IHE/WUR
Prof.dr.ir.Nick van de Giesen	TUDelft
Prof.dr. Dominique Mazvimavi	UWC, Cape Town, South Africa
Prof.dr. Andy Nelson	ITC-UTwente
Assoc.prof.dr.ir. Christiaan van der Tol	ITC-UTwente

Acknowledgements

I would like to firstly thank my supervisor Chris Mannaerts, for his patience and support along with his confidence in my capabilities. I tend to be little on edge and pester close to deadlines. He was always able to quell any anxiety. My husband and I look forward to more beers and mussels with him and his beautiful wife, Ines, in the future.

Next, I would like to thank my mentors, Yijian Zeng, Hamideh Nouri and Poolad Karimi, and supervisor. They were always available for banter, coffee and guidance.

This research has been possible due to the opportunity and financial support through the UN-FAO WaPOR project. I also kindly want to thank those at eleaf and Vito on who I worked in the FRAME consortium with. They were very supportive and collaborative.

Thank you to the staff in the ITC department, especially Prof. Bob Su, a busy but funny man, and Lindy Snijders the effervescent and the always helpful and approachable.

Further, so many thanks to my office mates. I felt a real comradery with them. Particularly, Sammy Njuki of whom I stole many potatoes from, and Rousha Zeng who always warmed my soul with her nurturing ways.

I would like to thank my husband, and best friend, Andres Cabrera Flamini who bared my endless thesis discussion (which was most often a monologue) and fed my belly and soul when I was when I needed it most. Finally, I would like to express my thanks to my parents, Rob and Donna Blatchford, who not only tolerated my long absence during my PhDs degree, but made the degree possible for me in more ways than they probably know.

Megan Blatchford
January 2021
Enschede, The Netherlands

Table of Contents

List of figures

List of tables

List of Abbreviations

AGBP	Above Ground Biomass Production
DMP	Dry Matter Productivity
CV	Coefficient of Variation
E	Soil Evaporation
EC	Eddy Covariance (flux tower)
EOS	End of Season
ETa	Actual Evapotranspiration
ETIa	Actual Evapotranspiration and Interception
ETa-EC	ETa from eddy covariance flux tower measurements
ETa-Fu	ETIa estimated from Fu
ETa-MOD16	ETIa and interception obtained from MOD16 dataset
ETa-MSG	ETIa obtained from geostationary Meteosat dekadal GPP product from the EUMETSAT LandSAF (LSASAF)
ETa-WB	ETIa obtained from the physical water balance
ETIa-WPR	ETIa obtained from WaPOR database and methodology
fAPAR	Fraction of Absorbed Photosynthetically Active Radiation
GPP	Gross Primary Productivity
HI	Harvest Index
I	Interception of rainfall
L1	Level 1 (250m resolution)
L2	Level 2 (100m resolution)
L3	Level 3 (30m resolution)
LAI	Leaf Area Index
LCC	Land Cover Classification
LE	Latent Energy (latent heat of evaporation)
LST	Land Surface Temperature
LUE	Light Use Efficiency
METE	Metehara
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalised Difference Vegetation Index
NPP	Net Primary Productivity
NPP-MOD17	NPP obtained from MOD17 dataset
NPP-MSG	NPP obtained from geostationary Meteosat dekadal GPP product from the EUMETSAT LandSAF (LSASAF)

NPP-WPR	NPP obtained from WaPOR database
ODN	Office du Niger
PCP	Precipitation
PHE	Phenology (crop)
PSN	Net Photosynthesis
Q	Surface run-off
QAQC	Quality assurance/quality control
r	Correlation Coefficient
R^2	Coefficient of Determination
ETo	Reference Evapotranspiration (FAO-56 definition)
ETo-EC	ETo from eddy covariance flux tower meteorological data
ETo-WPR	ETo from WaPOR database (GEOS-5 model meteorological data)
RMSE	Root Mean Square Error
SGD	United Nations - Sustainable Development Goals
SMC	Relative Soil Moisture Content (wetness indicator)
SOS	Start of Season
SR	Solar Radiation
T	Transpiration (by vegetation)
TBP	Total Biomass Production
V1	WaPOR version 1.1
V2	WaPOR version 2.1
WaPOR	The FAO portal to monitor WAter Productivity through Open access of Remotely sensed derived data
WON	Wonji
WP	Water Productivity
ZAN	Zankalon
ß	WaPOR beta version
ε_a	Autotrophic (plant) Respiration (fraction)

Chapter 1

Introduction

The subject of this thesis is on understanding the quality of remote sensing-based crop water productivity datasets and their implications on interpreting agricultural productivity. In this introduction, first a brief review of crop water productivity and why the push for large scale monitoring. Next, the approaches to monitoring are briefly described, followed by a discussion on the need to understand dataset accuracy. Finally, the motivation, objectives and structural overview of the remaining chapters is given.

1. Why monitoring crop water productivity?

There is mounting pressure on the agricultural sector to increase crop production between 60 and 110% by 2050 (Alexandratos and Bruinsma, 2013), to ensure future food security based on the demands of a growing population. However, while there is increasing demand of agricultural output, agriculture at a global scale is facing multiple limitations on inputs, in particular, water and arable land. The consumption of land and water resources are occurring at a faster rate than the global regeneration (Steffen et al. 2015; Rockstrom, Lannerstad, and Falkenmark 2007; Falkenmark, Rockström, and Karlberg 2009; Conijn et al. 2018).

Crop water productivity (CWP), an indicator for water use efficiency (WUE), aims to integrate land productivity in conjunction with water productivity. CWP evolved from the terms WUE and water productivity (WP). Water use requirements of plants were first studied in the early 1900s by weighing containers (Briggs and Shantz 1914). This was then constrained to the ratio of plant production to evapotranspiration for a unit area (Viets, 1962). In the 1950s and 1960s 'water use requirement' was scaled to field and termed WUE (Hanks et al. 1969; Hanks and Tanner 1952; Bierhuizen and Slayer 1965). Tanner and Sinclair (1983) summarised the literature and defined WUE as the biomass of water accumulated per unit of water transpired (T) and evaporated (E) per unit crop area. Modern WUE is commonly used by irrigation engineers to assess the efficiency of water delivery or supply, whereas WP distinctly refers to the water consumed through actual evapotranspiration (ETa). This WP term was further expanded to consider beneficial outputs, e.g., physical or economic yield, rather than purely the plant production. Finally, CWP, was developed to specifically define the physical crop production in terms of fresh yield to the ETa: CWP (kg m^{-3}) = Yield (kg ha^{-1}) / ETa (mm $season^{-1}$) × 10. This definition considers only actual ETa which accounts for water that is consumed, and therefore no longer available to other uses. ETa consists of soil evaporation (E) and plant transpiration (T).

Since food security remains the primary goal of the agricultural sector, measures to limit water consumption in agriculture must not come at the cost of curbing food production. Hence, improving CWP, and subsequently WUE, is gaining attention in addressing the issue of increasing food demand with increasing water limitations (Hoogeveen et al. 2015; Kijne et al. 2009). This has been recognised by the United Nations (UN) Sustainable Development Goals (SDGs) which stipulate that agricultural productivity should be doubled by 2030 (SGD2.3) and that WUE must be substantially increased (SDG6.4) (UN 2016). The UN Food and Agricultural Organisation (FAO) and International Water Management Institute (IWMI) have already initiated programs aimed at measuring and monitoring CWP—Water Productivity Open-access Portal (WaPOR) (https://wapor.apps.fao.org/home/1).

Monitoring CWP helps evaluate where CWP is on track, and where improvements or changes need to be made. Monitoring and evaluation forms the basis for modification of land and water management interventions and can narrow in on high and low performance areas to assess the efficiency of management activities. Monitoring and evaluation of CWP helps with identifying the most valuable and efficient use of water resources (and other resources) in crop growth.

2. How to monitor crop water productivity?

CWP has seldom been measured at in-situ level for the assessment of agricultural performance. Rather, in-situ measurements to monitor its components, yield and evapotranspiration, have been a common way to understand, monitor, and benchmark implications of the environment (Arora et al. 2011; Pinter et al. 2003; Ali and Talukder 2008) and climate conditions (Sadras, Grassini, and Steduto 2007) as well as on farm management practices (Rockström and Barron 2007; Geerts and Raes 2009; Arora et al. 2011; Noellemeyer, Fernández, and Quiroga 2013) on crop growth and water consumption. In-situ measurements are limited in their ability to monitor crop conditions over extensive areas due to their considerable costs and resource demands.

Remote sensing allows monitoring of various aspects of agricultural production. Open access satellite imagery now provides near real-time data at varying spatial and temporal resolutions including: 10 m with < 10-day return period (Sentinel 2), 30 m with 16-day return period (Landsat), 100 m with near daily return period (Proba-v), and 250 m with a 1 to 2-day return period (MODIS, Sentinel 3). Higher resolutions are available for paid products including e.g.: Planet (3–5 m), GeoEye (1 m), and Pleiades-1A (2 m). These data sources provide a spatially and temporally extensive option to estimate

agriculture indices over large areas and time periods. For instance, the UN Food and Agricultural Organization (FAO) has released the Water Productivity Open-access portal (WaPOR) database in 2018 providing open access to remote sensing CWP for Africa and the Middle East.

The application of remote sensing in estimating agricultural performance indicators is increasing as it offers a cost effective reproducible method for measurement that can cover larger physical areas as compared to in-situ methods, such as field water balances or ground measurements (Sadras et al. 2015). Remote sensing can be used to estimate evaporation, transpiration and biomass product of plants (yield), rainfall and land use, and therefore CWP (Hatfield et al. 1984; Wiegand et al. 1991). It can be a valuable tool in determining these variables when ground based observations are not available for large spatial scales. Glenn et al. (2020) proposes that "Remote sensing is perhaps the only feasible means of estimating ETa over wide areas of mixed landscape types". The estimations of ETa and yield can then be converted to estimate the CWP.

The estimation of ETa through remote sensing is divided into methods that use a thermal-based (surface energy balance) approach and those that use a vegetation index-based approach. Thermal based approaches estimate the ET through energy balance, where the latent energy or heat flux is the residual in the energy balance and is converted to ET dividing it by the latent heat of vaporization. Current source algorithms include amongst several others: SEBAL (Bastiaanssen et al. 1998), METRIC (Allen 2007), SEBS (Su 2002), S-SEBI (Roerink et al.2000) and ALEXI (Anderson et al. 2011). The main distinction between most thermally based methods for calculating ETa is the determination of the sensible heat flux and partitioning of the turbulent i.e. sensible and latent heat fluxes (Courault et al. 2005; Kalma et al. 2008; Li et al. 2009).

The underlying principle of many remote sensing-based estimates of biomass production, which is also used in agriculture, is that the relationship between the absorbed light and the carbon assimilation in most plants is relatively constant (Monteith 1977; 1972). This ratio, termed light use efficiency (LUE), is used to convert remote sensing-based estimation of light absorption to gross primary productivity (GPP). In agricultural applications the GPP is then converted to dry matter productivity typically through static conversion factors. The crop yield is then derived using the harvest index (HI), the above ground biomass fraction (f) and the moisture content (θ) of the harvestable product, which are all crop-specific factors. The main differences among the remote sensing-based biomass estimates are the LUE stress factors (or scalars) (Song et al. 2013) and the fAPAR function. Only a few studies have compared variations in these

algorithms with no definitive conclusions on which is preferred for agricultural applications.

3. Need for understanding accuracy of remote sensing observations

Validation is the process of understanding and assessing the degree in which the model represents the 'real' world. Despite significant improvements in remote sensing technology over the last decades, there is still significant uncertainty in estimating evapotranspiration and biomass at continental scale. This is due to the difficulty in fully representing each land cover class in such a diverse set of environments. Further complexity is added in assessing performance on continents such as Africa and Middle East due to the scarcity of ground data, resulting in under representation of calibration and validation datasets on these continents.

Validation of remote sensing products is an essential step in understanding their applicability and characterizing their uncertainty. This uncertainty can guide the end-user if a remote sensing-based product is suitable as input into different water management activities along with the associated risk when making a decision based on the product. Many studies exist that attempt to validate large remote sensing-based evapotranspiration and primary productivity datasets. Most studies are focused on one or two validation methods at one scale. The most common validation methods are either point or pixel scale against ground-truth data, like eddy covariance (EC) measurements (e.g., Mu, Zhao, and Running 2011; Ardö 2015; Wang et al. 2013), or spatial inter-comparison of a product over regions, land cover classes, biomes (e.g., Mueller et al. 2011). Some authors validate multiple products, based on remote sensing and other models or methods, against each other for spatial and temporal patterns and against ground-truth data (e.g., Hu, Jia, and Menenti 2015; Nouri et al. 2016; Ardö 2015; Liu, Wang, and Wang 2014). However, other than a few occasions, for example, Velpuri et al. (2013), these validation efforts often failed to evaluate the product at multi-scale, from pixel to basin or region.

Best-practice validation strategies of large remote sensing datasets have been proposed by (Zeng et al. 2019; 2015). They recommend multi-stage validation activities that include combinations of direct validation, physical validation and cross-comparisons. In practice, many developers of remote sensing products include all or at least a combination of these activities during their validation. To name a few, these include the MODIS MODLAND product (Morisette et al. 2002; Morisette et al. 1998); Copernicus Global Land Service products

Dry Matter Productivity; and ASTER land surface temperature (Schneider et al. 2012).

In regions such as Africa, where little observational data is available, validation should utilise all available avenues for ascertaining product quality, with a multi-step and -phase validation strategy that includes direct validation (with ground measurements), physical consistency check and cross-comparisons. As such, the limitations due to the sparseness of available data are reduced, and the product quality is understood from a multi-scale perspective, by using validation best-practice and combining multiple validation techniques.

4. Objectives and organization of thesis

The main objective of this thesis is to determine the suitability of large remote sensing-based datasets for monitoring and evaluating of CWP. Efforts were placed on the main CWP components, evapotranspiration, and biomass (as described in Chapter 2). The research is focused on understanding the data quality in the context of their applications. This thesis aims to do this via the following steps:

1. Benchmark remote sensing-based CWP accuracy standards based on attainable in-situ CWP accuracy (Chapter 3)
2. Quantify accuracy of a large remote sensing-based evapotranspiration dataset (Chapter 4)
3. Quantify the suitability of varying remote sensing-based resolutions for application in agricultural productivity (Chapter 5)
4. Quantify the sample size to improve processing time of validation activities of large continental CWP datasets (Chapter 6)
5. Reflect on the implications of the CWP concept (Chapter 7)

Chapter 2

Introduction to the The FAO portal to monitor WAter Productivity through Open access of Remotely sensed derived data dataset

This chapter describes the FAO portal to monitor WAter Productivity through Open access of Remotely sensed derived data (WaPOR) and its datasets used in this thesis. CWP was derived from the ETIa and NPP datasets and was not taken directly from the portal.

1. Dataset extent and resolution

The WaPOR database provides evapotranspiration and interception (ETIa) and net primary productivity (NPP) in three spatial resolutions dependent on the location and extent. The products available are shown in Table 2.1 and are available online on the WaPOR portal (https://wapor.apps.fao.org/home/WAPOR_2/1).

The extent of the lowest resolution dataset (L1) of the database in shown in Figure 2.1.

Table 2.1: Description of the WaPOR V2.0 ETIa and NPP data products, available on the WaPOR portal, used for validation in Africa (FAO, 2020b).

	Spatial resolution	Temporal resolution *	Spatial extent (in Africa)	Satellite (spatial resolution | return period)
Level I (L1)	250m	Dekadal	Continental Africa	MODIS (250m|1-day)
Level II (L2)	100m	Dekadal	Morocco, Tunisia, Egypt, Ghana, Kenya, Niger, Sudan, Mali, Benin, Ethiopia, Rwanda, Burundi, Mozambique, Uganda.	**MODIS (250m|1-day) **PROBA-V (100m|2-day)
Level III (L3)	30m	Dekadal	Awash, Ethiopia Koga, Ethiopia Office du Niger (ODN), Mali Zankalon, Egypt	Landsat (30m|16-day)

*Dekadal is approximately ten days. It splits the month into three parts, where the first and second dekads are ten days and the third dekad covers the remaining days in the month.
**MODIS is resampled to 100 m up to March 2014 and Project for On-Board Autonomy - Vegetation (Proba-V) is used from March 2014.

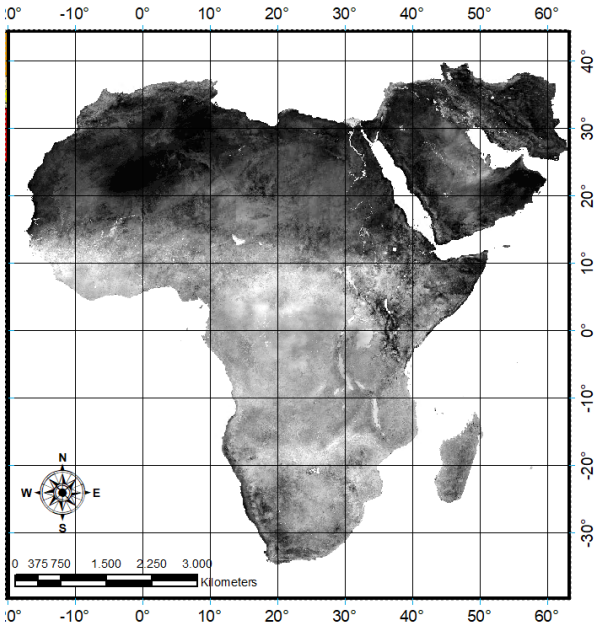

Figure 2.1: The extent of the images used in the progressive sampling.

2. Evapotranspiration and Interception

The analysis dataset is the ETIa V2.0 products available on the WaPOR portal. The ETIa is based on a modified version of the ETLook model (ETLook-WaPOR) described in Bastiaanssen et al. (2012). The ETLook-WaPOR model uses Penman-Monteith to estimate ETa adapted to remote sensing input data (FAO, 2018, 2020a). The Penman-Monteith approach uses the combined approaches of the energy balance equation and the aerodynamic equation and is described in the FAO-56 drainage paper (Allen et al. 1998). The ETIa defines soil evaporation and transpiration separately using Equation 2.1 and Equation 2.2. The interception is a function of the vegetation cover, leaf area index (LAI) and PCP. The ETI-WaPOR is then calculated as the sum of evaporation, transpiration and interception.

$$\lambda E = \frac{\Delta(\mathrm{Rn,soil} - \mathrm{G}) + \frac{\rho_{air}\, C_P (e_{sat} - e_a)}{r_{a,soil}}}{\Delta + \gamma\left(1 + \frac{r_{s,soil}}{r_{a,soil}}\right)} \quad (2.1)$$

$$\lambda T = \frac{\Delta(\mathrm{Rn,canopy}) + \frac{\rho_{air}\, C_P (e_{sat} - e_a)}{r_{a,canopy}}}{\Delta + \gamma\left(1 + \frac{r_{s,canopy}}{r_{a,canopy}}\right)} \quad (2.2)$$

Where E and T (mm day^{-1}) are the evaporation and transpiration respectively and λ is the latent heat of vaporisation. R_n (MJ m^{-2} day^{-1}) of the soil ($R_{n,soil}$) and canopy ($R_{n,canopy}$) is the net radiation and G (MJ m^{-2} day^{-1}) is the ground heat flux. ρ_{air} (kg m^{-3}) is the density of air, C_P (MJ kg^{-1} °C^{-1}) is the specific heat of air,($e_{sat} - e_a$) (kPa) is the vapour pressure deficit (VPD), r_a (s/m) is the aerodynamic resistance, r_s (s/m) is the soil resistance, or canopy resistance when using the Penman-Monteith-model to estimate evaporation or transpiration respectively. $\Delta = d(e_{sat})/dT$ (kPa °C^{-1}) is the slope of the curve relating saturated water vapour pressure to the air temperature, and γ is the psychometric constant (kPa °C^{-1}). This approach partitions the ETIa to evaporation and transpiration using the modified versions of Penman-Monteith, which differentiate the net available radiation and resistance formulas based on the vegetation cover according to the ETLook model (Bastiaanssen et al. 2012). A major difference between ETLook-WaPOR and ETLook is the source of remote sensing data for the soil moisture. In the original ETLook soil moisture is derived from passive microwave, and in the WAPOR approach soil moisture is derived from Land Surface Temperature (LST).

Interception (I) is the process where the leaves intercept rainfall. Intercepted rainfall 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 PCP.

$$I = 0.2\, I_{lai}\left(1 - \frac{1}{1+\frac{c_{vegPCP}}{0.2\, I_{lai}}}\right) \qquad (2.3)$$

C_{veg} is the vegetation cover and is calculated from the normalised difference vegetation index (NDVI) and I_{lai} is the leaf area index converted from c_{veg}.

3. Net primary productivity and above ground biomass production

NPP is a fundamental characteristic of an ecosystem, expressing the conversion of carbon dioxide into biomass driven by photosynthesis. NPP is the gross primary productivity (GPP) minus autotrophic respiration, the losses caused by the conversion of basic products (glucose) to higher-level photosynthesis (starch, cellulose, fats, proteins) and the respiration needed for the maintenance of the standing biomass. The NPP, as defined in WaPOR, is expressed as:

$$\mathrm{NPP} = Sc \cdot Rs \cdot \varepsilon_p \cdot fAPAR \cdot SM \cdot \varepsilon_{lue} \cdot \varepsilon_T \cdot \varepsilon_{CO_2} \cdot$$

$$\varepsilon_{AR}[\varepsilon_{RES}] \quad (2.4)$$

Where Sc [-] is the scaling factor from dry matter productivity (DMP) to NPP, Rs is the total shortwave incoming radiation (MJT ha^{-1} day^{-1}), ε_p is the fraction of photosynthetically absorbed radiation (PAR) (0.4–0.7µm) in total shortwave with a value of 0.48 (JPar/JTotal-sw). fAPAR [-] is the PAR-fraction absorbed by green vegetation. SM [-] is the soil moisture stress reduction factor. ε_{lue} (-) is the light use efficiency (LUE) (DM=Dry Matter) at optimum (kgDM/GJPA), ε_T (-) is the normalized temperature effect, ε_{CO2} (-) is the normalized CO_2 fertilization effect, the ε_{AR} (-) is the fraction kept after autotrophic respiration and ε_{RES} (-) is the fraction kept after residual effects (including soil moisture stress). A look-up table, based on the land use classification is used to determine the LUE for a given pixel.

When total biomass (TBP) or above ground biomass productivity (AGBP) is derived from the continental NPP data (without prior information on crop type), the following conversions are used in the WaPOR database (FAO 2018).

$$TBP\ \left(\frac{kg\ DM}{ha.day}\right) = 22.22\ \cdot \mathrm{NPP} \quad (2.5)$$

$$AGBP\ \left(\frac{kg\ DM}{ha.day}\right) = 0.65 \cdot\ 22.22\ \cdot \mathrm{NPP} \quad (2.6)$$

Where 0.65 is the conversion fraction from total to above ground biomass and 22.22 is the conversion from NPP in gC m^{-2} day^{-1} to DMP (above and below ground dry biomass) in kg ha^{-1} day^{-1}, assuming a carbon fraction of 0.45 in the organic matter.

4. Intermediate datasets

Datasets (including intermediate datasets) available for the validation include relative soil moisture content (SMC) – a wetness indicator, NDVI, solar radiation (SR), NDVI quality layer, LST quality layer, PCP and reference evapotranspiration (ETo) (Table 2.2). The producers of the datasets in the WaPOR portal – the FAO's Remote sensing-based database for the monitoring of agricultural water and land productivity in Africa and the Middle East (FRAME) Consortium, led by eLEAF and comprised of The Flemish institute for technological research (VITO), International Institute for Geo-Information Science and Earth Observation at the University of Twente (ITC-UTWENTE) and WaterWatch – provided the SMC and NDVI layers for the validation. All other layers are available on the WaPOR portal. The NDVI quality layer and the LST quality layer are indicators of the quality of the input satellite data. The NDVI quality layer provides the gap, in days, to the nearest valid observation for that variable. The LST quality layer

provides the number of the days between the date of the data file and the earlier remote sensing observation on which the data is based.

Table 2.2: Description of the intermediate and product datasets used for the evaluation of ETIa (FAO, 2020a).

Dataset	Spatial | Temporal resolution/s	Data product/s*	Sensor/s**
NDVI	Available for L1, L2 and L3 (per Table 2.1)	MOD09GQ[1], PROBA-V[2], Landsat 5,7,8[3]	MODIS[1], PROBA-V[2], Landsat[3]
SMC		MOD09GQ[1], PROBA-V[2], Landsat 5,7,8[3]	MODIS[1], PROBA-V[2], Landsat[3]
SR		SRTM (DEM)	MSG
LST quality layer	As for L1; Table 2.1	MOD11A1, MYD11A1	MODIS
NDVI quality later	As for L1; Table 2.1	MOD09GQ[1], PROBA-V[2], Landsat 5,7,8[3]	MODIS[1], PROBA-V[2], Landsat[3]
PCP	5km|daily	CHIRPS v2, CHIRP	TRMM, GPM
ETo	25km|daily	SRTM (DEM)	MSG, MERRA/GEOS-5

* MOD09GQ - MODIS/Terra Surface Reflectance Daily L2G Global 250 m SIN Grid; MOD11A1 - MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1 km SIN Grid; MYD11A1 - MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1 km SIN Grid; SRTM - Shuttle Radar Topography Mission (DEM – Digital Elevation Model - 90m); CHIRPS - Climate Hazards Group InfraRed Precipitation with Station data

** MODIS - Moderate Resolution Imaging Spectroradiometer; PROBA-V - Project for On-Board Autonomy - Vegetation; Landsat – Landsat Satellite 5, 7 and 8; MSG - Meteosat Second Generation (used for transmissivity); TRMM- Tropical Rainfall Measuring Mission; GPM – Global Precipitation Measurement; MERRA- Modern-Era Retrospective analysis for Research and Applications ; GEOS-5 – Goddard Earth Observing System

[1] L1 data product and sensor; [2] L2 data product and sensor; [3] L3 data product and sensor

Reflectance bands for the continental (L1), national (L2) and sub-national (L3) datasets are currently retrieved from the MODIS, MODIS/Project for On-Board Autonomy—Vegetation (PROBA-V), PROBA-V, and Landsat satellites respectively (Table 2.1). Information is embedded in visible, near-infrared, and thermal (Landsat only) infrared bands, which were used to retrieve the land surface temperature (LST), vegetation indices, and atmospheric temperature.

WaPOR ETIa and NPP further relies on input from weather, digital elevation, precipitation and transmissivity data from other sources. The weather data (i.e., air temperature, relative humidity

wind speed) is obtained from Modern-Era Retrospective analysis for Research and Applications (MERRA) prior to 21 February 2014 and the Goddard Earth Observing System (GEOS-5) after 21 February 2014 (Rienecker et al. 2011). The weather data is resampled using a bilinear interpolation method to the 250 m resolution. The temperature data is additionally resampled based on the DEM (90 m) of the SRTM. Atmospheric transmissivity is taken from the Meteosat Second Generation (MSG) (FAO 2018). Precipitation data are used as a limiting factor of soil moisture availability and interception and is sourced from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset (Funk et al. 2015). ETo is estimated using the weather, atmospheric transmissivity, and solar radiation data and has no satellite data input. Detailed information on the methodology and processing can be found in the WaPOR methodology documents (FAO 2018).

Chapter 3

Status of accuracy in remotely sensed and in-situ agricultural water productivity estimates: A review[1]

[1] This chapter is based on: Blatchford, M.L., Mannaerts, C.M., Zeng. Y., Nouri, H., Karimi, P., 2019. Status of accuracy in remotely sensed and in-situ agricultural water productivity estimates: A review, Remote Sensing of Environment 234, 111413

Abstract

The scarcity of water and the growing global food demand has fevered the debate on how to increase agricultural production without further depleting water resources. Crop water productivity (CWP) is a performance indicator to monitor and evaluate water use efficiency in agriculture. Often in remote sensing datasets of CWP and its components, i.e. crop yield or above ground biomass production (AGBP) and evapotranspiration (ETa), the end-users and developers are different actors. The accuracy of the datasets should therefore be clear to both users and developers. We assess the accuracy of remotely sensed CWP against the accuracy of estimated in-situ CWP. First, the accuracy of CWP based on in-situ methods, which are assumed to be the user's benchmark for CWP accuracy, are reviewed. Then, the accuracy of current remote sensing products are described to determine if the accuracy benchmark, as set by in-situ methods, can be met with current algorithms. The percentage error of CWP from in-situ methods ranges from 7% to 67%, depending on method and scale. The error of CWP from remote sensing ranges from 7% to 22%, based on the highest reported performing remote sensing products. However, when considering the entire breadth of reported crop yield and ETa accuracy, the achievable errors propagate to CWP ranges of 74% to 108%. Although the remote sensing CWP appears comparable to the accuracy of in-situ methods in many cases, users should determine whether it is suitable for their specific application of CWP.

1 Introduction

Over the past decades, the use of crop water productivity (CWP) as an agricultural performance indicator has increased. This indicator is specified in the United Nations (UN) Sustainable Development Goals (SDGs), which stipulate that agricultural productivity should be doubled by 2030 (SGD2.3) and that water use efficiency must substantially increase (SDG6.4) (UN, 2016).

CWP, as an indicator, is a measurable property that allows users to monitor and evaluate agricultural water productivity. CWP provides a way to benchmark and define goals, objectives or gaps for management and decision making (Hellegers et al. 2009). It can also be used to analyse and evaluate the impacts of alternative management strategies (Kijne, 2003), as it is influenced by on-farm management (Geerts and Raes, 2009).

Remote sensing can currently be used to measure agricultural performance at high spatial and temporal resolutions. The application of remote sensing in estimating agricultural performance indicators is increasing as it offers a cost-effective reproducible method for measurement that can cover larger physical areas as compared to in-situ methods, such as field water balances or ground measurements (Sadras et al. 2015).

Remote sensing allows monitoring of various aspects of agricultural production. Open access satellite imagery now provides near real-time data at varying spatial and temporal resolutions including: 10 m with less than 10-day return period (Sentinel 2), 30 m with 16-day return period (Landsat), 100 m with daily return period (Proba-v), and 250 m with a 1 to 2-day return period (MODIS, Sentinel 3). Higher resolutions are available for paid products including: Planet (3-5 m), GeoEye (1 m), and Pleiades-1A (2 m). These data sources provide a spatially and temporally extensive option to estimate agriculture indices over large areas and time periods, even at a global scale. For instance, the UN Food and Agricultural Organisation (FAO) is currently releasing the Water Productivity Open-access portal (WaPOR) database, providing open access to remote sensing CWP for Africa and the Middle East. This database includes actual evapotranspiration (ETa), above ground biomass production (AGBP) and gross biomass water productivity (GBWP) at spatial scales varying from 100 m to 250 m, depending on location, at a 10-day temporal resolution (FAO, 2019).

The accuracy requirements of remote sensing products have been specified for certain applications. The Global Climate Observing System (GCOS) has defined observation requirements for essential climate variables (ECVs) (WMO, 2011), which includes AGBP. The Copernicus Global Land Service defined three accuracy levels for dry matter productivity (DMP): threshold, target and optimal absolute

accuracy at 10, 7 and 5 t ha^{-1} $year^{-1}$, respectively (Swinnen et al. 2015). As these accuracy requirements are defined for their intended use – GCOS for climate modelling and GL for land surface monitoring (Su et al. 2018; Zeng et al. 2015) – they are not necessarily relevant to agriculture. However, they are currently the only existing standards.

Accuracy standards for remotely-sensed datasets have not been specifically established for applications in agriculture. Given the increasing research and application of remote sensing in agriculture and the introduction of open-access datasets, such as the WaPOR database, it is essential to define these end-user requirements. These accuracy standards set the quality standards of the datasets for the producers and allow users to verify if a dataset meets their needs. Thus, the accuracy of the remote sensing dataset should be high enough that the indicators derived from them can serve their intended purpose: to improve the agricultural system.

This review first benchmarks the accuracy of CWP based on in-situ methods. In-situ methods are those that have been used in agricultural performance assessment in the field. Second, the reported accuracy and potential of remote sensing-based CWP are critically reviewed. From this, the current reported accuracy of CWP remote sensing variables is discussed to identify if they can meet the standards of in-situ methods.

2. Definitions of crop water productivity and its components

2.1 Crop water productivity

Irrigation performance indicators came to prominence in the 1980s as a tool to monitor and evaluate the efficiency of irrigation systems (Abernethy, 1990; Bos and Nugteren, 1974; Seckler et al. 1988). Water use efficiency (WUE) is a commonly used indicator in irrigation performance. WUE is defined as the relation between a unit of crop yield and a unit of water applied or diverted. This indicator is primarily geared towards irrigation engineers (van Dam et al. 2006). This definition focuses on the efficiency of engineering infrastructure and design, but it does not consider the productivity potential of the applied water. This definition was extended to water productivity (WP) or CWP, which is dynamic and dependent on the user. The CWP indicator specifically focuses on the crop yield per unit of water consumed by the crop (Zwart and Bastiaanssen, 2004):

$$CWP\ (kg\ m^{-3}) = \frac{Crop\ yield\ (kg\ ha^{-1})}{10 \times \sum_{i=SOS}^{EOS} ET_a\ (mm)} \quad (3.1)$$

The crop yield is defined as the seasonal crop yield and the ETa is taken as the accumulated crop ETa, from start of season (SOS) to end of season (EOS). The conversion factor, 10^{-1}, converts ETa from mm to m3 ha^{-1}. By using ETa it considers all the water used by the crop, including rainfall and groundwater inputs to the agricultural cropping system, rather than just irrigation water. Therefore, CWP as an indicator is equally valid for irrigated and rainfed systems (Bossio et al. 2008).

Based on the CWP definition (equation 3.1), CWP is estimated on a seasonal basis, and therefore the accuracy requirements are relevant to the crop growing season. CWP has also been applied to assess variation within a field (Hellegers et al. 2009), among fields (Jiang et al. 2015; Zwart and Leclert, 2010) and blocks within an irrigation scheme (Ahmed et al. 2010; Conrad et al. 2013; Zwart and Leclert, 2010), and among schemes (Awulachew and Ayana, 2011). Therefore, the spatial resolution that is required for CWP is dependent on the scale of the performance assessment. CWP has also been used as an indicator to assess trends over time (El-Marsafawy et al. 2018; Wang et al. 2018). Generally, CWP is applied in a relative manner, rather than an absolute manner. That is, the CWP is compared to other users or the same user over time, rather than applied as an absolute value.

2.2 Crop yield

Early work in the 1980s on understanding crop yield variability noted the usefulness of vegetation indices (VI) for vegetation characterisation (Tucker and Sellers, 1986). Typically, a linear regression is assumed between spectral vegetation indices and crop yield, as estimated through in-situ methods. Some authors have claimed that up to 80% of in-field crop yield variability can be explained by VI (Shanahan et al. 2001; Tucker et al. 1980; Wiegand and Richardson, 1990). Although these empirical approaches show good agreement for many crops in a local setting (e.g. wheat), they are unique to the crop and location and therefore lack the physical basis to extend to other crops or locations (Lobell, 2013).

The underlying principle of many remote sensing-based estimates of biomass production, which is also used in agriculture, is that the relationship between the absorbed light and the carbon assimilation in most plants is relatively constant (Monteith 1977, 1972). This ratio, termed light use efficiency (LUE), is used to convert remote sensing-based estimation of light absorption to gross primary productivity (GPP) (Zhang et al. 2015):

$$GPP\ (gC\ m^{-2}\ day^{-1}) = \varepsilon\ LUE_{max}\ (gC\ MJ^{-1})\ \sum PAR\ \ (MJ\ m^{-2}\ day^{-1}) \times fAPAR\ (-) \quad (3.2)$$

Where ε is a scalar to account for various stress factors, LUE is the Light Use Efficiency, PAR is the Photosynthetically Active Radiation, and fAPAR is the fraction of Absorbed Photosynthetically Active Radiation and GPP is the total amount of CO_2 that is fixed by the plant in photosynthesis. The maximum LUE (LUEmax) is commonly scaled to account for deficiencies due to environmental stress. These are varied between models and often include at least one of the following: soil moisture stress, vapour pressure deficit or heat stress (Bloom et al. 1985). While crop models, such as Aquacrop (Raes, 2017), and carbon assimilation models, such as SCOPE (van der Tol et al. 2009), often incorporate a nitrogen stress factor, it not frequently incorporated into remote sensing approaches.

The PAR is taken as the spectral range of solar radiation that is available to the plant for photosynthesis (Asrar et al. 1992). The fAPAR has been identified as a suitable integrated indicator of the status of the plant canopy (Gobron et al. 2000). There are a number of available satellite based fAPAR products currently available at the global scale. The currently available products include the MODIS Terra FAPAR (operational) (Myneni et al. 2002), the COPERNICUS 1-km (GEOV2) fAPAR product (operational) (Verger et al. 2017) and the Quality Assurance for Essential Climate Variables (QA4ECV) FAPAR product (1982-2016) (Pinty et al. 2006) amongst others. The products vary in retrieval methods, fAPAR definitions and satellite platforms.

The net primary productivity (NPP) is defined as the net amount of primary production after carbon lost to autotrophic respiration (AR) is considered:

$$NPP\ (gC\ m^{-2}\ day^{-1}) = GPP\ (gC\ m^{-2}\ day^{-1}) - AR\ (gC\ m^{-2}\ day^{-1}) \quad (3.3)$$

The distinction between the GPP, NPP, DMP, AGBP and yield are shown in Figure 3.1. The LUEmax and AR are often specified for vegetation type in global models. For example, MODIS (Running et al. 2004; Zhao et al. 2005) and Copernicus (Swinnen and Van Hoolst, 2018) GPP, NPP and DMP global products use look-up tables containing LUEmax for different vegetation types, including cropland. In agricultural applications, the LUEmax is not solely defined for cropland, but for specific crop type.

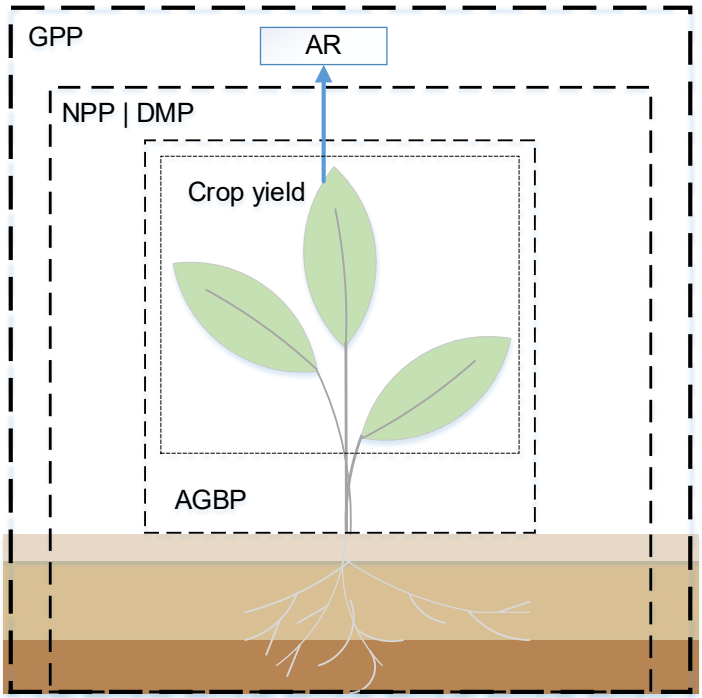

Figure 3.1: Distinction between GPP, NPP, DMP, AGBP and crop yield products, where each box compares the plant parts associated with each product.

In agricultural applications the NPP is then converted to DMP (2.3), typically through static conversion factors, before being converted to crop yield (2.4):

$$DMP\ (kg\ ha^{-1}) = \frac{1}{0.045} NPP\ (gC\ m^{-2}) \quad (3.4)$$

Where 0.045 is the conversion factor from organic carbon to dry organic biomass. The crop yield is then derived using the harvest index (HI), above ground fraction (f) and the moisture content (θ) of the harvestable product (Prince et al. 2001):

$$Crop\ yield\ (kg\ m^{-2}) = f\ \cdot\ HI(-)\ \sum_{SOS}^{EOS} DMP\ (kg\ ha^{-1})\ /\ (1 - \theta) \quad (3.5)$$

The HI definition varies from crop to crop. For example, for cereals it is defined as the ratio of grain yield to total seasonal AGBP (Donald, 1962), and for potato it is defined as the ratio of tuber to total seasonal below and AGBP. HI and θ are not well defined through remote sensing for a diverse variety of crops and are often taken as

standard values, as Bastiaanssen and Steduto (2017) did for a global Earth observation study of CWP. Remote sensing uses crop specific (and sometimes location specific) constants of LUEmax, HI and θ (Zwart et al. 2010).

2.3 Evapotranspiration

ETa is the process of water transferring from land to the atmosphere and is comprised of evaporation from the Earth's surface and transpiration from plants. These processes are typically estimated together due to the difficulty in partitioning them. Remote sensing-based ETa estimates first appeared in the 1970s (Li et al. 2009). Since then, a number of approaches have been developed including surface energy balance approaches such as Surface Energy Balance System (SEBS) (Su, 2002), Surface Energy Balance Algorithm for Land (SEBAL) (Bastiaanssen et al. 1998), Surface Energy Balance Index (SEBI) (Menenti and Choudhury, 1993), Simplified Surface Energy Balance Index (S-SEBI) (Roerink et al. 2000), Enhancing the Simplified Surface Energy Balance (SSEB) (Senay et al. 2007), Operational Simplified Surface Energy Balance (SSEBop) (Senay et al. 2013), Mapping EvapoTranspiration at high Resolution with Internalized Calibration (METRIC) (Allen et al. 2007), Atmosphere-Land Exchange Inversion model (ALEXI) and disaggregated ALEXI (DisALEXI) (Anderson et al. 2011), Penman-Monteith based models (PM-models) (Mu et al. 2007), and simplified empirical regression methods, such as VI-based methods (Glenn et al. 2011). Although there is no consensus on the best algorithm or approach, the surface energy balance and PM-models are more frequently used for large scales as they offer generalised approaches and reduce the need of calibration and parametrization. The surface energy balance estimates the latent energy as the residual of the surface energy balance:

$$Rn\ (W\ m^{-2}) = LE\ + H\ + G \tag{3.6}$$

Where, LE (W m^{-2}) is the latent heat flux, Rn is the net radiation, H (W m^{-2}) is the sensible heat flux and G (W m^{-2}) is the ground heat flux. The LE is converted to ETa by LE/ λ, where λ is the latent heat of vaporization. Several surface energy balance algorithms exist that vary in complexity and data requirements. Two prominent types of surface energy balance approaches are the single-source (e.g. SEBS and SEBAL) and two-source models (ALEXI and DisALEXI). The WaPOR database (FAO, 2018) calculates ETa based on the ETLook model (Pelgrum et al. 2012) and is defined as:

$$LE = \frac{\Delta(\mathrm{Rn} - \mathrm{G}) + \rho_{air} \times \mathrm{C}_P(\mathrm{e}_{sat} - \mathrm{e}_a)/\mathrm{r}_a}{\Delta + \gamma\,(1 + \frac{\mathrm{r}_s}{\mathrm{r}_a})} \quad (3.7)$$

Where Δ = d(e_sat)/dT (kPa °C^{-1}) is the slope of the curve relating saturated water vapor pressure to air temperature (T°C). ρ_{air} (kg m^{-3}) is the density of air, c_p (MJ kg^{-1} °C^{-1}) is the specific heat of air, (e_{sat} - e_a) (kPa) is the vapour pressure deficit, r_a (s m^{-1}) is the aerodynamic resistance, r_s (s m^{-1}) is the surface resistance or canopy resistance when using the PM-model to estimate canopy or crop ETa, and γ (kPa °C^{-1}) is the psychometric constant. This approach further partitions ETa to evaporation and transpiration using modified versions of Penman-Monteith, which differentiate the net available radiation and resistance formulas based on the fractions of vegetation and bare soil. The accuracy of this approach is highly dependent on the accurate estimation of the canopy resistance (or the inverse – canopy conductance) (Raupach, 1998).

2.4 Accuracy metrics

Accuracy refers to the closeness of a measurement, observation, or estimate to a true value. The accuracy of the in-situ and remote sensing estimate of CWP can be expressed through a number of metrics. The percentage (or relative) error allows for standardisation as the accuracy becomes comparable, even if values are significantly different in size. The relative error is defined as:

$$Relative\ Error\ (\%) = \frac{|absolute\ error|}{accepted\ value} \times\ 100 \quad (3.8)$$

The absolute error is defined as:

$$absolute\ error = \ experimental\ value - accepted\ value \quad (3.9)$$

The accepted value is user defined. Often, the field or in-situ measurement or estimate is taken as the accepted value and the remote sensing value is considered the experimental value. When in-situ methods are validating other in-situ methods, the method considered most accurate is typically considered the accepted value. Otherwise, for field measurements with no comparison to other methods, the error is taken as the variation in repeated measurements. Where possible, the relative error is taken directly from the literature. If the relative error is not reported, but the absolute error or deviation and the mean errors are stated, the relative error is calculated using equation 3.7 and 3.8. If the metrics of relative errors are not reported

in the literature in a way which allows calculating the relative error, the errors are taken directly from the literature in the form of the root mean square error (RMSE) or the coefficient of determination (R^2).

In terms common to error propagation, the absolute error is defined as:

$$absolute\ error = \Delta x \tag{3.10}$$

This is equivalent to absolute uncertainty, which is typically expressed as x±Δx. For CWP, the error can be determined through simple error propagation in the multiplication of uncertainties (BIPM et al. 2008; Taylor, 1997):

$$R = \frac{X}{Y} \tag{3.11}$$

$$\delta R \approx |R| \cdot \sqrt{\left(\frac{\delta X}{X}\right)^2 + \left(\frac{\delta Y}{Y}\right)^2} \tag{3.12}$$

Where, in this case, R represents the CWP, δR represents the uncertainty of CWP, |R| represents the absolute value of the mean, and δR/|R| represents the relative uncertainty or percent error. Similarly, X in this case represents the crop yield and Y represents the ETa.

When possible, the error associated with different methods to estimate yield, ETa, and CWP, is categorised. The categories are expert error, typical error and novice error, which is based on the categories defined by Allen et al. (2011). The expert error refers to the maximum error derived from the scientific literature, the typical error range is cited as the range of error associated with larger studies where scientific experts were not present in the data collection, and the novice error is defined as the lowest reported accuracy for that approach.

3. In-situ methods accuracy for crop water productivity assessment

CWP, in the form of equation 3.1, has seldom validated in irrigation performance assessment. Therefore, focus is given to the errors associated with the components of CWP in order to derive the CWP uncertainty associated with the combination of field methods to estimate yield and ETa. These methods have historically been accepted as standards in estimating crop yield and ETa and therefore will be

considered as benchmarks for the accuracy of remote sensing products.

3.1 Crop yield

Methods for estimating crop yield and biomass include physical measurements, personal estimates and micrometeorological measurements. Physical measurements comprise whole-plot harvest, crop-cutting over sub plots (Verma et al. 1988), and sampling of harvest units such as sacks, baskets and bundles. Personal estimates include expert assessments and farmers' estimates, both predictive and recall, and daily records. Micrometeorological measurements primarily include eddy covariance (EC) and chamber techniques to measure carbon fluxes. Crop-cuts and farmer estimates are the two most commonly used methodologies by scientists and statisticians to estimate crop production.

Commonly accepted in-situ methods for accuracy (where literature is available) include: whole-plot harvest, crop-cutting, and both recall and predictive farmer estimates. Crop-cutting, whole-plot harvest and models estimate the biological yield as they do not take into account post-harvest losses. Farmer estimates measure the economic yield, therefore the post-harvest losses are typically accounted for (Fermont and Benson, 2011). Micrometeorological measurements are less common for estimating crop yield, as compared to other methods. They measure GPP, NPP or net ecosystem exchange (NEE) rather than directly measuring crop yield (Moureaux et al. 2012).

The whole-plot harvest method to estimate crop yield is generally undertaken in demonstration plots in on-farm trials (Norman et al. 1995). This method requires a clear delineation of the plot boundary before harvest. The harvest is typically dried and weighed post-harvest. When the plot requires multiple harvests, the drying and weighing is done separately and added. This method is determined as the standard to estimate crop yield and biomass (Casley and Kumar, 1988) and is suggested to provide the highest accuracy. The error typically arises from an error in crop area estimation, the irregular shape of fields, the inclusion of areas not planted and/or not having proper supervision (Murphy et al. 1991). This method is suggested to be almost bias free as it avoids error from on-field variability (Sud et al. 2016). This method is most suitable to fields that are less than 0.5ha, as crop-cutting and whole-plot harvest take a similar time at this field size (Casley and Kumar, 1988).

The crop-cutting method to estimate biomass and crop yield uses sampling on sub-plots. The production is taken as the sum of the sub-plot production over the sum of the sub-plot areas. This method, developed in the 1940s in India (Mahalanobis and Sengupta, 1951;

Sukhatme, 1947), was recommended as the standard method to estimate crop production in the 1950s (FAO, 1982). The sub-plot's size and shape is known to greatly influence the bias of the plot, where decreasing sub-plot size corresponds to increasing bias, indicating a trade-off between resources required and degree of accuracy.

The following examples of crop-cutting errors have been found in the literature. FAO (1982) reported over-estimation for irrigated and non-irrigated wheat yield ranging from 4.8%-11% for triangular plots of 11 m^2 and 15.7-23.4% for triangular plots of 2.7 m^2 when compared to a whole-plot harvest estimate on a 44 m^2 plot. Fielding and Riley (1997) found a difference in yield estimates of broccoli from small plots to be 36-82% greater than large plots. Poate (1988) suggests that the effect of bias is essentially eliminated for plot sizes greater than 40 m^2, yet bias of 14% with 60 m^2 triangular sub-plots has been found in other studies (Casley and Kumar, 1988). FAO (1982) suggests that the sub-plot size can be smaller for more densely plotted fields and up to 100 m^2 for mixed cropping. Bias of 28% for sorghum and 17% for yam was found in plot sizes of 50 m^2 and 100 m^2. The bias was not reduced until plot sizes increased to 200 m^2 (Poate and Casley, 1985). The bias reduced to 8-10% when re-analysed using a variant of the standardised method. Other research has found overestimation of crop-cutting to be 37-86% as compared to farmer estimates (Minot, 2008, as cited in Fermont and Benson, 2011), greater than 20% as compared to other crop-cut methods (Casley and Kumar, 1988) and 14-38% as compared to whole-plot harvest (Verma et al. 1988).

The error of cross-cut is primarily a result of on-field variability, which is commonly 40-60% (Casley and Kumar, 1988; Fielding and Riley, 1997; Poate, 1988). Other contributing sources of error, with an upward bias in parenthesis if known, include: calculation of plot area (5%), focus effect (<5%), border bias (<5%) and edge effect (2-3%) (Verma et al. 1988). Although each of these biases is small individually, they can accumulate to large upward biases (Diskin, 1999). The highest biases are often attributed to fields that have small, irregular shapes with uneven planting density and mixed cropping (Murphy et al. 1991), where crop-cutting was poorly executed (Rozelle, 1991). Undertaking crop-cutting under controlled conditions, where enumerators follow the rules precisely, can significantly increase reliability (Poate and Casley, 1985).

Farmer surveys are commonly accepted as reasonable estimates for crop yield. Farmer estimates can be either recall or predictive. Recall estimates are suggested to have higher accuracy, particularly when farmers are surveyed close to post-harvest. However, recall periods across literature range from weeks up to three to six seasons. Predictive estimates are obtained on a plot by plot basis, based on either farmer or expert experience (Sud et al. 2016). Studies

in the 1980s comparing crop-cutting to farmer estimates showed that the crop-cutting method reported consistently higher crop yields than farmer estimates. A study in Zimbabwe showed upward bias of 27-82% (Casley and Kumar, 1988) and a study in Ethiopia showed a 31-46% upward bias (Minot, 2008, as cited in Fermont and Benson, 2011) as compared to farmer recall. Studies in Asia showed a high fit (R^2>0.85) between crop-cutting and farmer predictions (David, 1978; Singh, 2003), yet the bias was as still as high as 25-37% (David, 1978). However, a study in Sweden showed no bias of farmer recall as compared to crop-cutting with a range of -4.9-9% at the country level, which may be a result of expert crop-cutting.

A study across five countries in Africa (Verma et al. 1988) showed that farmer estimates of production, both recall (taken either immediately after harvest or within three weeks after harvest) and predictive (taken 2 and 4 weeks pre-harvest), were frequently less biased than crop-cutting when compared to whole-plot harvest. The crop-cutting method (25 m^2) sub-plot showed an average upward bias of 34%, while pre-harvest and recall farmer estimates had an average upward bias of 9% and 3% respectively. This suggests that farmer recall estimates were the most accurate method of the three in estimating production. There is evidence that in some countries, such as Malawi, Philippines, and Nepal, farmers are not familiar with their cropped area, which can lead to error in estimating crop yield per hectare (Rozelle, 1991). On the other hand, farmers in China and Indonesia were very familiar with their area. Therefore, supporting farmers in their estimation area can improve the accuracy, while surveys should be undertaken where the cropped area is well known (Poate and Casley, 1985). Further, to increase the reliability of farmer estimates, surveys should be as close as possible from harvest date (Malik, 1993), and care should be taken with conversion to standard units from local units (Diskin, 1999). It is suggested that farmer estimates may be just as accurate, if not more accurate, as crop-cutting methods, at least for estimating total production (Murphy et al. 1991; Poate, 1988; Verma et al. 1988).

Yield can also be estimated in field by in-situ measurements of carbon fluxes. GPP and NPP is first estimated and then can be converted to yield estimates through crop and location specific conversion factors, as per equations 3.2-3.4. The two predominant methods to estimate carbon fluxes are EC and chamber methods. The EC method continuously measures spatially averaged carbon fluxes for an area of a few hectares (Baldocchi, 2003), while the chamber method measures only the change in gas concentrations of the area covered by the chamber. EC and chamber methods have been widely compared to each other (Dugas and Bland, 1989; Kutzbach et al. 2007) in a number of ecosystems. Chamber methods vary and are also well

compared to each other (Pumpanen et al. 2003; Rochette and Hutchinson, 2005). However, scarce research reports on the accuracy of these methods in agricultural land classes. Further, no studies were found that compared EC to methods that estimate crop yield, i.e. whole-plot harvest, crop-cut or farmer estimates. The limited available research specific to cross-comparison of these methods in cropped areas or grassland is included here. It should be noted that the reported accuracies here relate to carbon fluxes and do not consider errors introduced converting these measurements to crop yield.

EC measurements of carbon fluxes were compared to automatic chamber techniques in cotton and wheat fields (Wang K. et al. 2013). The difference in NEE between the two systems was -9-7%. Riederer et al. (2014) compared EC and chamber measurements in a grassland site. The results were comparable (R^2=0.78); however, they suggested EC is preferable as it is more sensitive to atmospheric conditions. Steduto et al. (2002) compared the carbon flux from closed-system canopy-chamber chamber measurements to the pattern of flux measurements by Bowen ratio energy balance (see section 3.2 of this Chapter) for sugarbeet and marjoram crops. The overall maximum deviation was approximately 6-8%. Dugas et al. (1997) found that the canopy chamber method underestimated carbon uptake as compared to the leaf chamber and micrometeorological methods in grasslands, which was similar to comparisons reported in other environments. It is noted that the leaf chamber method has the least precision due to scale, while the micrometeorological methods are prone to error due to error in input data. The reported agreement in measurements between the two methods in non-agricultural lands vary significantly, from 8-26% (Dore et al. 2003) and up to >60% (Fox et al. 2008). Other studies have used EC (Buysse et al. 2017; Miyata et al. 2000; Suyker and Verma, 2010; Zanotelli et al. 2013) or chamber measurements (Langensiepen et al. 2012; Maljanen et al. 2001; Wagner and Reicosky, 1992) at field level in a cropped area but have not compared the measurements to other in-situ carbon measurement methods. EC faces spatial representation issues. The EC footprint defines the field of representation of the measured flux, which is influenced by wind speed and direction. Therefore, ideally EC stations should be placed on flat, homogenous terrain. Authors attempt to deal with the footprint issue through footprint modelling (Schmid, 2002); however, in remote sensing comparisons, many authors simply compare point-to-pixel, and the footprint is neglected (Turner et al. 2005).

The errors associated with crop yield per hectare estimated from these methods, as derived from the literature discussed here, are summarised in Figure 3.2. Where known, the accuracy is divided into novice error, typical error and expert error. The expert error ranges

are defined as the highest cited accuracy, associated with a carefully planned and executed approach (Poate and Casley, 1985; Verma et al. 1988). The typical error is cited as the range of error associated with larger studies where enumerators are not present for the entire data collection period (David, 1978), and the novice error is defined as the lowest reported accuracy for that approach (Casley and Kumar, 1988; Fermont and Benson, 2011). This applies even to farmer estimates, where the error can be reduced by an expert supporting farmers in their estimate of the cropped area. In Figure 3.2, the y-axis is the suggested relative error range, as defined in equation 3.7 and the x-axis are the in-situ methods. The expert error is shown with the most saturated colour, and the novice error is shown with the least saturation. This division acknowledges that the error is minimised when an expert in the field carries out the estimate of that in-situ approach. This was only applied where known; if unknown, only the typical error is displayed. This is based on the approach taken by Allen et al. (2011) in defining the accuracy of methods to estimate ETa.

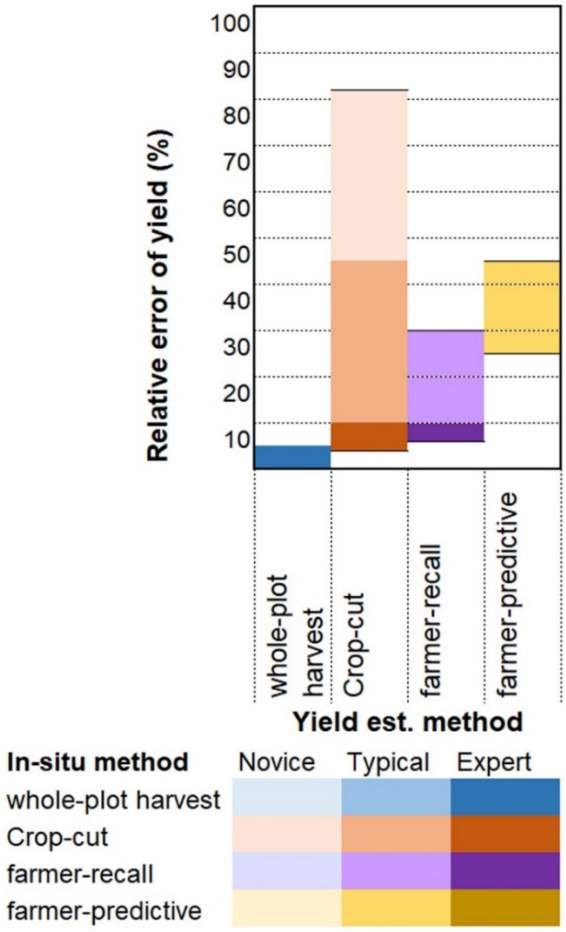

Figure 3.2: Relative error associated with in-situ methods of crop yield estimation. All methods provide estimates for at field scale for cropping season.

Our literature review reveals that the whole-plot harvest has the highest accuracy and is typically used as the reference for estimating the error of other in-situ methods, with a relative error typically less than 5%. The crop-cutting method shows to have the next highest accuracy, if carried out by an expert. However, if the enumerator is not carefully guided, this method shows the lowest accuracy with a cited relative error of up to 82%. The recall farmer estimates did not reach accuracies as high as the crop-cut when undertaken by an expert. However, the typical error was less. Due to the limited available literature, the predictive farmer estimates only show a typical range. Compared to the expert and typical ranges of the other in-situ methods, predictive farmer estimates have the highest associated error. EC and chamber method estimates are not included in Figure 3.2, as currently there is insufficient evidence to pertain to the accuracy or uncertainty of deriving crop yield from these methods.

Other methods to estimate crop yield and biomass include daily recording, crop cards, purchase records from the agro-industry, and crop models (Fermont and Benson, 2011). The accuracy of these estimates, with the exception of models, are not well reported. Crop models are useful tools in estimating crop yield and biomass under various conditions. The complexity of crop models varies extensively with different specific applications (Boote et al. 1996; Jin et al. 2018). Although they are useful in prediction and scenario analysis, the accuracy of these methods will not be included here as they are not considered standards in reporting or measuring of biomass or crop yield. Further, the calibration and validation of crop models are typically carried out using crop-cutting and farmer estimates.

3.3 Evapotranspiration

Several in-situ measurement systems exist to determine ETa. These measurement systems can be categorised in hydrological methods (such as soil water balance and lysimeters), micro-meteorological methods (such as EC, Bowen ratio energy balance (BREB), and the scintillometer method), and plant physiology methods (such as sap flow) (Rana and Katerji, 2000). These methods, and their accuracies, have been comprehensively discussed by Allen et al. (2011) and are summarised in Figure 3.3. Thus, only accuracies reported in crop and grass systems published after 2011 are included. Due to the limited data availability on in-situ measurement uncertainty in agricultural lands, uncertainty observed in grasslands is also

included as grasslands are similar to crops in height and in their low sensitivity to night-time fluxes (Wohlfahrt et al. 2012). However, it must be acknowledged that they are typically more spatially heterogeneous as compared to croplands, and often have a larger aerodynamic roughness due to plant density (Moureaux et al. 2012). It should be noted that the ETa error reported post-2011 is considered expert error, as the literature cited here was undertaken by scientists.

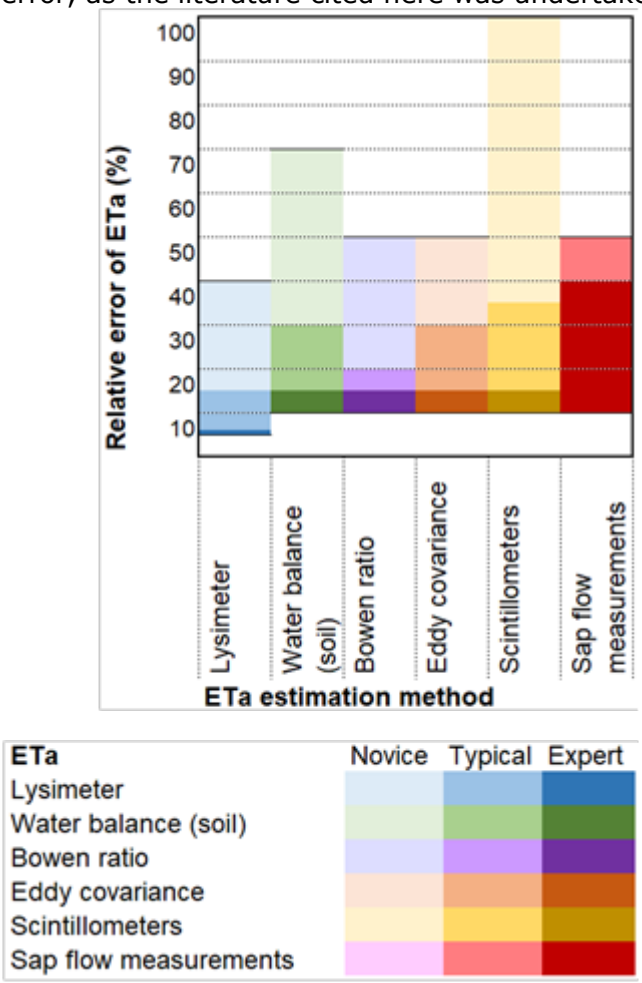

Figure 3.3: Relative error associated with in-situ methods of ETa estimation used for irrigation performance, adapted from Allen et al. (2011).

Lysimitry has the lowest expert, typical and novice error. In line with previously reported accuracy, several authors have more recently asserted the accuracy of the lysimeter is within 5-25%. Gebler et al. (2015) looked at the variation between six lysimeters in a grass site in

close proximity (within 50 m of each other) with similar soil properties and reported a resulting relative error of 8%. The variation was mainly attributed to non-homogenous harvest management. Evett et al. (2012) compared lysimeter measurements to the soil water balance in an irrigated cotton field and found a relative error range of 5-18%. Wind speed has the largest effect on lysimeter accuracy as it affects scale performance (Howell et al. 1995). Increasing the measurement frequency can help reduce wind speed effects (Dugas and Bland, 1989). Using this approach in an irrigated almond orchard, Lorite et al. (2012) found that up to 97% of the observed variability from a one-tree weighing lysimeter was caused by wind speed. Lysimitry, along with sap flow measurements, have the least spatial coverage. This means the selection of a suitable field or plot, in which the lysimeter can appropriately represent the vegetation and soil dynamics, is essential to retain the expert level accuracy. This is combined with the need to ensure the equipment is properly installed and calibrated. Lysimitry is often used for the validation of other in-situ ETa methods as it is generally accepted to be the most accurate method to estimate ETa.

The soil water balance was compared to EC in rainfed wheat fields by Imukova et al. (2016) with Gaussian error propagation law to determine the uncertainty. The resulting uncertainty ranged from ±0.3-0.5mm day-1 with resulting error ranging from 24-48% (Imukova et al. 2016). The accuracies of EC were highly dependent on the energy balance closure method. The method for energy balance closure and the related accuracy has been investigated by number of authors. Both Sánchez et al. (2016) and Hirschi et al. (2017) found that forced energy balance closure using the Bowen ratio approach was the most successful when compared to the residual (of the energy balance) approach and the direct measurement approach. The Bowen ratio approach ensures scalar similarity in closing the energy balance, while the residual attributes the proportion of the closure to either the latent heat flux, the sensible heat flux, or both. The Bowen ratio approach found differences of 3-7% at seasonal scale in a drip irrigated vineyard (Hirschi et al. 2017) and 23% at daily scale (Sánchez et al. 2016) in a short grassland as compared to lysimeters. The residual approach had errors of 1-13% at seasonal scale (Hirschi et al. 2017) and 29% at daily scale (Sánchez et al. 2016). Mauder et al. (2018) evaluated energy balance closure methods in two grassland sites. They found that the Bowen ratio approach had better comparability with the lysimeter, but a higher bias, as than the residual approach. Similar results were observed by Gebler et al. (2015) who reported relative errors of 3.8% and 8% for annual and monthly scales respectively, as compared to a lysimeter, using the Bowen ratio approach to closure.

No literature since 2011 was identified that reports on the accuracy of the BREB method to estimate ETa. The accuracy of the BREB method is highly dependent on the accuracy of net radiation and ground heat flux measurements. Additionally, the errors in temperature and vapour pressure gradients can have a significant impact on ETa estimations (Cellier and Olioso, 1993). Irmak et al. (2014) looked at studies that compared the BREB method on multiple sites, including in agricultural sites, to other ETa measurement methods. Results varied considerably. On an annual scale in a lentil field, BREB overestimated ETa by 10-43% as compared to lysimeter ETa (Prueger et al. 1997). On a daily scale, Todd et al. (2000) noted differences between BREB and lysimeter to be 5-15% during the day and 25-45% at night in an irrigated alfalfa field. When BREB was compared to EC without forced energy balance closure, EC was reported within 67-77% of BREB ETa estimates. These discrepancies suggest that estimates of the scalar turbulent fluxes of H and LE are underestimated and/or that Rn is overestimated (Wilson et al. 2002).

Moorhead et al. (2017) reported surface layer scintillometer errors of 14% for a daily scale and 31% for an hourly scale as compared to lysimeter in irrigated sorghum fields. The error reported for large aperture scintillometers was higher at 52% (Moorhead, 2015). Yee et al. (2015) compared the latent and sensible heat fluxes of two large aperture scintillometers and two microwave scintillometers to EC estimates in a grassland site. The root mean deviations of latent heat fluxes between the scintillometers and EC ranged between 40.7-164.3 W m^{-2}, equivalent to 1.4-5.8 mm day^{-1}. When the scintillometers were compared to each other, the latent energy flux root mean square deviations (RMSD) ranged between 18.5-88.8 W m^{-2}, equivalent to an ETa RMSD of 0.65-3.1 mm day^{-1}. Beyrich et al. (2012) compared five side-by-side scintillometer systems and reported relative deviations of 5% within the sensible heat fluxes. However, the relative variation of the latent energy fluxes or ETa were not reported. The footprint consisted of more than 90% agricultural fields.

Sap flow ETa measurement uncertainty in cotton was estimated to be 0.03-0.5 mm hr^{-1}, based on repeated measurements (Uddin et al. 2013). In maize fields, pre-calibration sap-flow transpiration measurements over-estimated transpiration rates by 30-40%, which was reduced by half after calibration (Wang Y. et al. 2017). The difficulty in using sap-flow measurements as a stand-alone method to estimate ETa is that it actually measures transpiration, not ETa. Further, the measurements are at plant scale and errors typically occur at upscaling to the canopy, rather than the measurements themselves (Zhang et al. 2014). Therefore, representative soil evaporation measures are required in parallel for a valid comparison against ETa measurements.

It is also worth noting that the crop coefficient (Kc) is a widely accepted method to estimate ETa from reference evapotranspiration (ETo) in agricultural applications (Allen et al. 2011, 1998; Doorenbos and Pruitt, 1977), such as for estimating crop water demand. The Kc method considers the evapotranspiration under standard conditions as the ETo multiplied by a Kc. To obtain ETa a soil water coefficient needs to be incorporated to account for water stress. A number of Kc values have been defined based on crop, crop phenology (crop curve) and climate. The dual crop coefficient is more complicated and splits the Kc based on crop transpiration (basal crop coefficient, Kcb) and soil evaporation (Ke) (Allen et al. 1998, 1996). Despite the wide application of the Kc to estimate ETa in research (Guerra et al. 2015), it is difficult to determine the accuracy of this method. This is further complicated by the range in Kc values, as defined by FAO (Allen et al. 1998). The Kc values are empirically derived and not universal due to variations in a number of factors including climate, cultivar, soil type and agronomic practices. Anderson et al. (2017) found that the Kc and Kcb maximum values for various crops, when derived from EC, were similar to previous studies; however, the Kcb seasonal trends were different to those in literature. Howell et al. (2015) found that the accuracy of the ETa estimated by Kc varied considerably between years as compared to lysimeters. Liu and Luo (2010) found that the Kc approach showed reasonable seasonal ETa with 10% relative error for winter wheat and summer maize as compared to lysimitry. However, peak ETa was underestimated and the mean relative error of ETa from the Kc approach for developmental stages ranged between 6.1% (mid-season) to 18.5% (end of season) for wheat and 5.4% (development) to 33.1% (initial-stage) for maize. Similarly, Guodong et al. (2016) found the Kc approach was sufficient in estimating seasonal ETa of cherry trees, with relative error of less than 5% when compared to the soil water balance method. However, the relative difference on a daily scale was 12.5 to 50%. These examples of the Kc method show mixed accuracy and typically require local calibration for Kc.

The appropriate in-situ method to estimate ETa is highly dependent on the resources available, the physical characteristics of where the measurements are taken, and the required measurement scale. Each method offers different advantages and disadvantages. Each method also has a different scale of representation, from leaf to plant scale (sap flow measurement), sample scale (lysimitry), plot scale (soil water balance and sap flow measurements), field scale (Bowen ratio and EC), and several hectares (scintillometers).

3.4 Crop water productivity

The current accuracy of the CWP from in-situ measures were derived as a combination of in-situ measures for estimating both crop yield and ETa through simple error propagation, using equations 3.13-3.14. The relative error ranges were derived by applying the error propagation equation to the maximum (novice) and minimum (expert) error associated with each crop yield and ETa in-situ measurement. These derived errors, however, do not take into account spatial scale differences between the crop yield and ETa measurements. Figure 3.4 shows the CWP relative error for each combination of the previously described crop yield and ETa in-situ techniques. The relative error is plotted on the y-axis, the ETa methods are plotted on the x-axis, and the crop yield methods are colour coded. The colour saturation is then used to distinguish if the in-situ methods are novice, typical or expert.

The relative error of the CWP field measurement, when estimates are undertaken by an expert, ranges from less than 5% (combination of lysimeter and whole-plot harvest) and up to 40% (combination of sap flow measurement and whole-plot harvest). For the crop-cutting method, the relative error ranges between 6-11% when combined with lysimeter, between 10-18% when combined with scintillometers, and can reach up to 41% when combined with sap flow measurements by experts. The relative error ranges for crop-cutting are comparable with the farmer estimates (recall). The typical errors are higher and range between 11-42% for the combination of lysimeter and farmer estimates (recall) to greater than 60% for the sap flow measurements and farmer estimates (predictive). The error ranges highlight the importance of the in-situ measurements being undertaken with due diligence; otherwise, the typical errors frequently exceed 40%, irrelevant of the method, while novice errors frequently exceed 50-60%.

In terms of setting conventional standards for the acceptable accuracy of CWP, the error for an expert should be used as the target. Excluding sap flow measurements (the least accurate ETa method), the target relative error is therefore in the range of 2% (lysimitry combined with whole-plot harvest) up to 18%. The acceptable error, however, may be taken as the typical error. The typical error ranges from 11% and up to 60%. This upper bound is too high to be suitable, particularly when CWP is being applied to estimate absolute values and not just spatial variability.

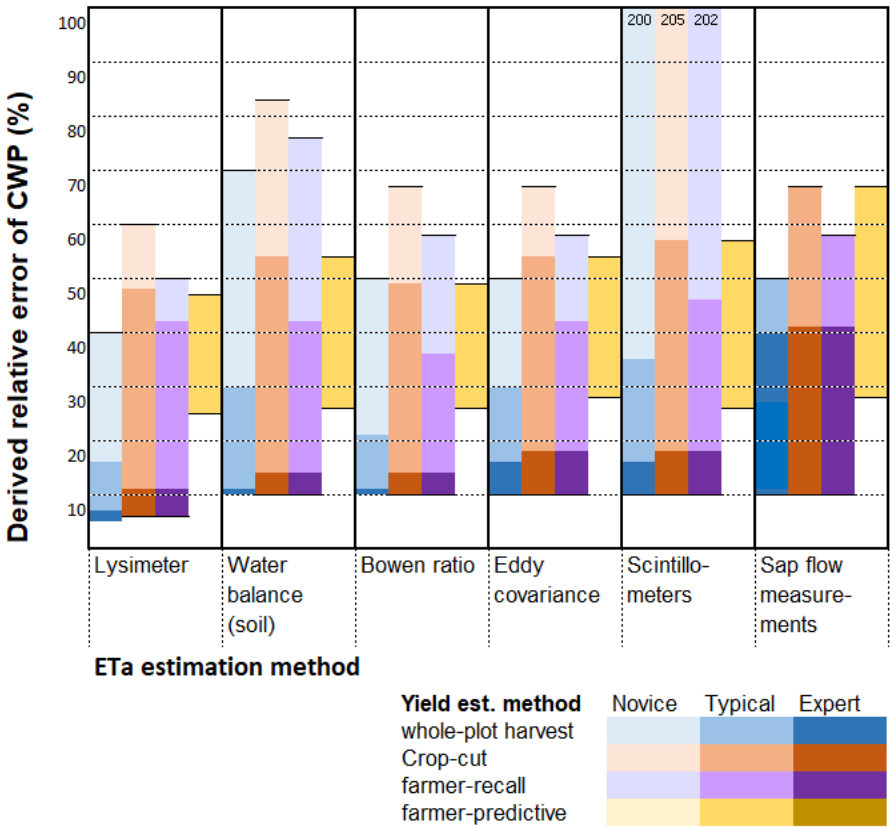

Figure 3.4: Relative error associated with CWP derived from in-situ methods of estimating ETa and crop yield. When numbers are located at the top of y-axis, they indicate value of relative error (when it goes) beyond 100%.

4. Accuracy of remote sensing-based approaches to assess crop water productivity

The potential of remote sensing to study irrigation and agricultural performances was first suggested in the late 1970s and early 1980s. The first applications estimated ETa to quantify crop water stress (Idso et al. 1977; Jackson et al. 1983), relative water supply (Menenti et al. 1992) and water deficit index (Moran et al. 1994). Then, remotely sensed ETa was used to assess the evaporative fraction (Bastiaanssen, 1998; Su, 2002), spatial distribution represented through the coefficient of variation (CV) of ETa (Bastiaanssen, 1998), CV of depleted fraction (Roerink et al. 1997) and water use efficiency (Menenti et al. 1989). Meanwhile, vegetation indices were being

applied to assess the performance of productivity indicators such as crop yield over applied water (Thiruvengadachari and Sakthivadivel, 1997) and spatial distribution and variation of crop yield (Bastiaanssen et al. 1999). These products, ETa and crop yield, were first combined to assess CWP in 1999 (Bastiaanssen et al. 1999). Several authors have used remote sensing to estimate CWP since.

As there exists only one direct validation of remote sensing CWP, the accuracy of ETa and crop yield as individual components of CWP, estimated by remote sensing, is summarised here.

4.1 Crop yield

To assess the overall error in remote sensing derived crop yield products, a comprehensive literature review was conducted and reported errors in croplands by various authors were synthesized (Table 3.1). This literature synthesis encompasses generalised approaches, with validation in croplands that do not include calibration. Generalised approaches are those that do not require calibration or parametrization. As such, it excludes regression models as these are typically specific to location, climates or crop, along with complex assimilation and forcing models.

Global and continental models for GPP and NPP were not originally designed for applications in agricultural performance and monitoring. However, more recently, these products have been tested or applied in agricultural land use classes. Further, based on the same underlying concept described in equation 1, the FAO has released a remotely sensed dataset of NPP for Africa and the Middle East with the specific purpose of monitoring and evaluating CWP (FAO, 2018). Therefore, validation on remote sensing-based GPP, NPP, AGBP (or DMP) and crop yield estimates were all considered, as long as they apply a generalised approach. Correction factors, relevant to crop and location, are often applied to retrieve crop yield from NPP and GPP (equations 3.3-3.7). Though these corrections are simple, they can impose significant errors. The implications of validating crop yield intermediates are discussed in section 4.1.1 of this Chapter.

Table 3.1: Stated accuracy of remote sensing derived crop yield, ordered by publication year.

Author	Location	Study size	Years	Main crops	Variable	Sensor (Spatial|temporal resolution)	Accuracy of study[1,2]	Method of validation
Campos et al. (2018b)	Nebraska, USA		2002-2011	Wheat	Crop yield	Landsat-5 TM ,7 ETM+ (30 m| 16-day)	RMSE (soybean yield) = 0.27;0.35 ton ha^{-1} RMSE (maize yield) = 0.81; 1.06 ton ha^{-1}	Crop-cut
	Central Asia - Fergana Valley	363,000ha	2010-2014	Mixed	Crop yield	Landsat-5 TM (30 m| 16-day)	RE = 10% R^2 = 0.709	Farmer reported yields - recall
Löw et al. (2017)	Global model, tested globally	Global	2004-2014	not specified	GPP	MODIS (500 m|8-day)	R^2=0.34 RMSE=94% RE = 33% Bias = -0.22 ton ha^{-1} day^{-1}	EC
	Saudi Arabia	50ha	2015	Maize	GPP	Landsat 8 (30 m|16-days)	RE = 5.8-6.2%	EC
Wang L. et al. (2017)	Awash, Ethiopia	14,000ha	2014-2016	Sugarcane	ABGP	Landsat 8 (30 m|16-days)	RE = 8.7-14.7% r=0.75 R^2=0.37-0.57	Crop-cut and Farmer reported yield - recall

Author	Location	Study size	Years	Main crops	Variable	Sensor (Spatial|temporal resolution)	Accuracy of study[1,2]	Method of validation
	Global model, tested globally		2001-2011	GPP: 36 cropped sites Yield: 12 cropped sites	GPP, Yield	MODIS (1 km|8-day)	Yield: R^2=0.61; RE=30-61% GPP: R^2=0.90; RMSE = 0.02-0.11 ton ha^{-1} day^{-1}; bias = 0-0.07 ton ha^{-1} day^{-1}; RE=0.5-88% (median=11.9%)	EC
Madugundu et al. (2017)	Global model, tested in USA	Global	2004-2005	Maize/soybean	GPP	MODIS (1 km|8-day)	bias < 49.8%	EC
	Global model, tested globally	Global	2000-2010	Mixed	GPP	MODIS (1 km|8-day)	MODIS GPP: r = 0.86; RMSE= 0.06 ton ha^{-1} day^{-1}; bias= -0.00 ton ha^{-1} day^{-1} TEC GPP model: r=0.77; RMSE= 0.08 ton ha^{-1} day^{-1}; bias= -0.02 ton ha^{-1} day^{-1}	EC

Author	Location	Study size	Years	Main crops	Variable	Sensor (Spatial|temporal resolution)	Accuracy of study[1,2]	Method of validation
Yilma (2017)	Global model, tested in USA	Global	2001-2005	Soybean	GPP	MODIS (1 km|16-day & 1 km|8-day)	MODIS GPP: R^2=0.64-0.67; RMSE=0.09-0.10 ton ha^{-1} day^{-1}; bias=negligible VPM: R^2=0.5-0.79; RMSE=0.08-0.10 ton ha^{-1} day^{-1}; bias=0.02-0.04	EC
	USA - Nebraska		2007-2010	Maize	Crop yield	MODIS (250 m|daily)	R^2 irrigated = 0.22 R^2 rainfed = 0.09 R^2 all = 0.52	Farmer reported yields - recall
Yuan et al. (2016)	Global model, tested in Africa	Global	2005-2006	Millet/grassland	GPP	MODIS (1 km|8-day)	r= 0.71 and 0.8	EC
	Global model, tested globally	Global	2008-2009	Maize, orchard	GPP	MODIS (1 km|8-day)	RE (maize)= -69.2% to -78.4% RE (orchard) = -74.1%	EC

Author	Location	Study size	Years	Main crops	Variable	Sensor (Spatial|temporal resolution)	Accuracy of study[1,2]	Method of validation
Tang et al. (2015)	Mexico - Yaqui River coastal plain	225,000ha cultivated area	1999-2000	Wheat	AGBP, crop yield	NOAA-AVHHR (1 km|Monthly) Landsat TM	Bias (yield) = +0.5 t ha^{-1} RE = 9.1%	Farmer reported yields - recall; crop cuts; and ministry statistics
	Global model, tested in USA	Global	2000	Corn/soy	NPP	MODIS (1 km|8-day)	RMSE = 2 ton ha^{-1} $year^{-1}$ RE = 20%	EC
Yan et al. (2015)	Global model, tested in USA	>12,000ha/county	2001-2002	Wheat	Crop yield	MODIS (1 km|8-day)	RE (state) = 5% R^2 (county) = 0.01-0.46 R^2 (climate) = 0.33-0.67	Ministry statistics based on farmer reported yields - recall

Author	Location	Study size	Years	Main crops	Variable	Sensor (Spatial|temporal resolution)	Accuracy of study[1,2]	Method of validation
	India - Sirsa			Wheat	Crop yield	Landsat ETM (30 m|16-day) | NOAA (1.1km|daily)	RE (regional) = 6% RE (field) = not reported	Regional scale: Regional statistics Field scale: crop-cut
Yuan et al. (2015)	Pakistan – Indus Basin		1993-1994	Wheat, rice, cotton, sugar-cane	AGBP/ Crop yield	AVHRR (1.1 km|monthly)	RE = 22%-42%	Regional statistics
	Mexico - Yaqui Valley		1993-1994 1999-2000 2000-2001	Maize, wheat, soybean	Crop yield	Landsat 5 TM | Landsat 7 ETM+ (30 m|16-day)	RE (regional) = 20% RE (field - wheat) = 4%	Whole-plot harvest
Sibley et al. (2013)	Sri-Lanka	1,752,100ha	1999-2000	Tea, coconut, rice, rubber	AGBP/ Crop yield	NOAA-AVHRR images (1.1 km|10-day)	R^2=0.47	Regional statistics

Author	Location	Study size	Years	Main crops	Variable	Sensor (Spatial|temporal resolution)	Accuracy of study[1,2]	Method of validation
	Mexico-Yaqui Valley		1993-1994 1999-2000 2000-2001	Wheat	Crop yield	Landsat 7 ETM+ (30 m|16-day) With CASA model (no calibration)	R^2 =0.78 RMSE = 0.37 ton ha^{-1} RE=5.9%	Farmer reported yield

[1]Abbreviations accuracy metrics used in this table: R^2 – coefficient of determination; r - correlation coefficient; RE – Relative Error (or percentage error); RMSE – root mean square error

[2] GPP (gC $m^{-2}.day^{-1}$) units are converted to ton $ha^{-1}.day^{-1}$ using equation 2 to ease comparison between GPP and AGBP and yield errors.

The main differences in the remote sensing models are the LUE stress factors (or scalars) (Song et al. 2013) and the fAPAR function. A few studies have compared variations in these algorithms with no definitive conclusions on which is preferred for agricultural applications. Yuan et al. (2015) compared the EC-LUE model (Yuan et al. 2010, 2007), MODIS-GPP -MOD17- algorithms (Running et al. 2004) and the vegetation production model -VPM- (Xiao et al. 2004) to EC GPP estimates at 3 adjacent corn and soybean fields in the USA. The MODIS-GPP typically underestimated GPP by -0.06 to -0.41 gC m^{-2} day^{-1}, the EC-LUE had a positive bias of 0.16-0.37 gC m^{-2} day^{-1}, and the VPM had a positive bias of 1.02-1.70 gC m^{-2} day^{-1}. Madugundu et al. (2017) compared the GPP derived from VPMs, one based on the enhanced vegetation index (EVI), one based on the normalized difference vegetation index (NDVI) and one based on the Land Surface Water Index (LSWI), for irrigated maize to EC GPP. The temporal resolution was 7-8 days as the site covered two Landsat-8 satellite paths. The mean average percentage error (MAPE) between the GPP from EC and GPP from the EVI VPM was 6.2%. The MAPE between GPP from EC and GPP from the NDVI VPM was 5.8%.

Yuan et al. (2016) compared GPP and yield estimated from EC-LUE model against GPP and yield estimated from EC at 36 cropped sites. The yield was derived by multiplying the EC-LUE GPP by the HI, the f and the autotrophic respiration. The EC-LUE had good agreement with the GPP at most sites with an overall R^2 of 0.9 and a RMSE and bias ranging between 1.75-5 gC m^{-2} day^{-1} at EC sites and 0.03-3.34 gC m^{-2} day^{-1} at yield sites. The sites showed no distinction in performance between irrigated (16 sites) and rainfed (9 sites) sites. The yield had a significantly poorer performance. The estimated crop yield accounted for approximately 61% of the variation in crop yield over a total of 26 site-years. The model underestimated yield between 61% and 32% at several sites, while three sites overestimated crop yield by 34% to 55%. The difference in accuracies between crop yield and GPP was primarily attributed to the uncertainty in the HI estimation method.

Global models have not been designed specifically for croplands, yet studies do not consistently find croplands to be performing better or worse than forest, grassland or other sites. Sjöström et al. (2013) compared MODIS GPP to GPP at 12 EC sites, including one cropped site in Africa. The correlation (r), RMSE and bias values for sites was 0.74, 2.13 gC m^{-2} day^{-1} and 1.18 gC m^{-2} day^{-1}, respectively. The r, RMSE and bias at the cropped site for 2005 and 2006 was 0.71 and 0.8, 0.97 gC m^{-2} day^{-1} and 0.73 gC m^{-2} day^{-1}, and -0.59 gC m^{-2} day^{-1} and -0.32 gC m^{-2} day^{-1}, respectively. As seen, the performance at the cropped site was better than the average for all sites in Africa. Yan et al. (2015) compared a generalised remote sensing derived GPP (TEC GPP model) and the generalised MODIS GPP

product to EC GPP at 18 sites, including six cropped sites across the globe. The TEC GPP model differentiated for C4 and C3 plants and introduced a water stress factor dependent on remotely sensed precipitation products. The TEC GPP model had an r, RMSE and bias of 0.86, 2.82 gC m^{-2} day^{-1}, and -0.16 gC m^{-2} day^{-1}, respectively, across cropped sites. The MODIS products had an r, RMSE and bias of 0.77, 3.38 gC m^{-2} day^{-1}, and -0.76 gC m^{-2} day^{-1}, respectively, across cropped sites. TEC GPP and the MODIS GPP performance was comparable at cropped and non-cropped sites, with average r-values across all sites of 0.85 and 0.73, respectively. The TEC GPP model did perform better than MODIS GPP at water stressed sites. Both models performance also increased at an annual time scale.

Turner et al. (2005) considered the MODIS NPP product to EC NPP at six sites (1 cropped) in the USA. They found RMSE of 91 gC m^{-2} $year^{-1}$ and 105 gC m^{-2} $year^{-1}$ for soybean and corn respectively, corresponding to over 2 ton ha^{-1} $year^{-1}$ of DMP. The cropped site performed similar to the forested sites, but not as well as the grassland sites. The RMSE was 8 gC m^{-2} $year^{-1}$ and 34 gC m^{-2} $year^{-1}$ for the cropped sites and grassland site respectively. The EC GPP and NPP were scaled to 5km x 5km grid using the Biome-B GC model. The error appeared to be lower for longer timescales and larger extents.

In a global study that compared MOD17A2H GPP to the EC GPP at 18 sites across the globe (including 3 cropped sites), it was found that croplands were not performing as well as forested sites (Wang L. et al. 2017). The R^2, RMSE and bias at the cropped sites was 0.34, 94%, and -10 gC m^{-2} day^{-1}, respectively. The cropped sites, similar to the grassland sites, had a significantly lower agreement to flux data as compared to the forested sites. The main possible sources of error were identified as the fAPAR MODIS product, land cover classification, and the LUE_{max}. The GPP estimates were improved when the MODIS fAPAR product was replaced with fAPAR derived from the Generation and Applications of Global Products of Essential Land Variables (GLASS) leaf area index (LAI) dataset (the R^2 for all sites increased to 0.79).

Similarly, a study in China found that without calibration of LUE_{max} the performance of MODIS GPP performed much worse in croplands compared to other vegetation (Wang X. et al. 2013). MOD17 was compared to 10 EC sites, including four maize sites and an orchard. The RMSE over the maize sites ranged between 59.7-89.4 gC m^{-2} 8-day^{-1} . The relative errors ranged between -69.2% to -78.4%. The RMSE at the orchard site was 51.2 gC m^{-2} 8-day^{-1} and the relative error was -43.3%. The cropped sites were typically performing worse than the non-cropped sites. The remote sensing product consistently understated the EC GPP. However, after LUE_{max} was adjusted for, the results improved considerably for all sites. The maize sites RMSE

reduced to 14.6-17.8 gC m^{-2} 8-day^{-1} and the relative error reduced to 3.1-11.5% (Wang X. et al. 2013).

Similar to NPP and GPP, significant differences in accuracy have been observed in literature for crop yield and AGBP. Positive results were found at the district level by Löw et al. (2017), who reported R^2 of 0.71 and an average overestimation of approximately 10% when compared to reported cotton, rice and wheat yields. Similar error was reported for wheat grain at a regional scale (±6%) by Bastiaanssen et al. (2003). However, when they considered a plot-to-plot comparison of remote sensing crop yield to crop-cutting, there was almost no correlation. Yilma (2017) reported total biomass errors of 8.7-14.7% against crop-cuts of sugarcane using different methods to calculate the vapour stress. When compared on a scheme level, the R^2 was 0.37 and 0.57 for all sugarcane varieties for a single variety of sugarcane respectively.

Campos et al. (2018a) estimated crop yield from remote sensing using LUE, WUE and normalised CWP, models. The results were compared to irrigated soybean and irrigated maize yields estimated from crop-cuts throughout the season until harvest. The LUE AGBP, as compared to crop-cuts, had an R^2 of 0.98. The RMSE values for different fields ranged between 1.39-2.18 ton ha^{-1} for each field over the growing season. WUE and CWP based approaches showed similar results for R^2. The CWP model had the lowest RMSE values (1.07-1.58 ton ha^{-1}). The SD (accuracy) of the crop-cut measurements was <5%. Sibley et al. (2013) compared MODIS (LUE model) derived crop yields to 134 irrigated and 94 rainfed maize fields in Nebraska and to a Hybrid-Maize model, with Landsat and MODIS used for model calibration. The APAR method was not as accurate as the Landsat crop-model based regression in terms of R^2 but was comparable with the Landsat calibrated crop-model. The RMSE was the highest for the APAR method in both irrigated and rainfed areas 2-3.2 ton ha^{-1}, while the Landsat crop-model based regression had RMSE values of just over 2 ton ha^{-1}.

Lobell et al. (2003) estimated wheat, soybean and maize yields in the Yaqui Valley, Mexico. The wheat yields were compared to whole-plot harvest measurements of grain and biomass, which also gave the HI. Intermediate data on APAR and moisture content were also taken in field. The regional wheat yield estimates varied up to 20% while field-based estimates indicated errors in regional wheat yields of less than 4% for both years of data. Lobell et al. (2002) compared remote sensing-based (CASA model) wheat yield estimates to farmer reported yields and found an R^2 of 0.78 and a RMSE of 0.37 ton ha^{-1}.

Crop yield is sometimes compared to regional statistics or values from literature. Zwart and Bastiaanssen (2007) compared remote sensing-based estimates of crop yield and biomass to both the

mean values and the distribution of local statistics and farmer reported crop yields, as the location of the fields where the measurements were derived were not available. They found that the crop yield from remote sensing LUE based approach was within 0.5 ton ha^{-1} to farmer reported wheat yields in Mexico. Similarly, Bastiaanssen and Ali. (2003) also compared remote sensing-based yield estimates of wheat, rice, cotton, and sugarcane in the Indus Basin, Pakistan. The average values per crop and per district were compared against regional statistics. The MAPE values per crop were 22% for wheat, 23% for sugarcane, 29% for rice and up to 42% for cotton. The RMSE for wheat, rice, cotton, and sugarcane were 0.53 ton ha^{-1}, 0.62 ton ha^{-1}, 0.55 ton ha^{-1}, and 13.5 ton ha^{-1}, respectively. Potential sources of error included sensor resolution as compared to plot size, land use patterns or rotations, and accuracy of secondary reported data.

Similarly, Samarasinghe (2003) estimated yields of tea, rubber, coconut and rice from remote sensing in Sri-Lanka and compared them to district level statistics of crop yield. The R^2 values ranged from 0.25 for rubber and up to 0.52 for tea. The author concluded that the monthly yield of tea, rubber and coconut could not be predicted from monthly biomass production. However, the model predicted rice yields better. The R^2 was 0.47 and the RMSE was 0.43 ton ha^{-1}. The model was suggested to perform better for rice due to prior knowledge on crop season. Reeves et al. (2005) found percentage errors of -4% to 5% at the state level. However, the error substantially increased at county and climate zones scales with R^2 values of 0.33-0.46 and 0.33-0.67, respectively, for varying years. The authors attributed this to high intra- and inter-annual variability in observed crop yield at county level. Further issues identified were smaller spatial aggregation, aberrant precipitation leading to a widely ranging wheat yield, difficulty relating estimates of above ground GPP to wheat yield, and the presence of other crops in pixels classified as wheat.

Yield and AGBP are often validated at different spatial and temporal scales to GPP and NPP. GPP and NPP are typically validated at the resolution of the image return period, while crop yield and AGBP are validated at seasonal or annual scales. Further, GPP and NPP are often validated using EC towers, typically a point-to-pixel comparison, whereas crop yield data is compared to in-situ data at the field or plot scale.

It difficult to assign an accuracy to the remote sensing of crop yield as there is a vast difference in reported accuracy. Reported relative GPP errors in croplands range from as little 5% after LUE_{max} adjustment (Wang X. et al. 2013) and up 70% and even 90% (Wang L. et al. 2017). This also highlights that a priori knowledge of the crop type has a significant influence on the accuracy of the remote sensing data by ensuring that LUEmax values are accurately allocated.

Reported errors of remote sensing estimates of crop yield and GPP have a similar range, from a few percent at a regional scale (Reeves et al. 2005), and as low as 10% (Löw et al. 2017) and up to 80% (Bastiaanssen and Ali, 2003) at field scale.

Figure 3.5 shows the relative error ranges of both remote sensing and in-situ measurements reported in, or derived from literature. Distinction between validation products, GPP or NPP and crop yield or AGBP, are made. The remote sensing values are taken from Table 3.1. The in-situ values are taken from Figure 3.1. The figure is a stacked column chart. The mean reported (or derived) relative error from each study, where available, is included. The highest reported error range is less than 5%, which was reported by one study (Lobell et al. 2003). Five studies, one validating GPP and four validating yield, have reported errors in the range of 5-10%. Three of these studies were validated at field scale (i.e. validated by EC, farmer reported yield or crop-cut) and two were validated at a regional scale against statistics. The GPP and crop yield do not seem to be attributed with higher or lower errors, despite findings by Yuan et al. (2016). This may be a result of higher prior knowledge of local HI, f and θ. The highest reported accuracy has the same relative error as the whole-plot harvest in-situ method. Five studies have a reported accuracy with the same relative error (expert) as the crop-cut and farmer recall methods. Seven studies report accuracies within the typical accuracy for crop-cut or farmer recall. Only three studies do not meet the typical or expert error of any in-situ method.

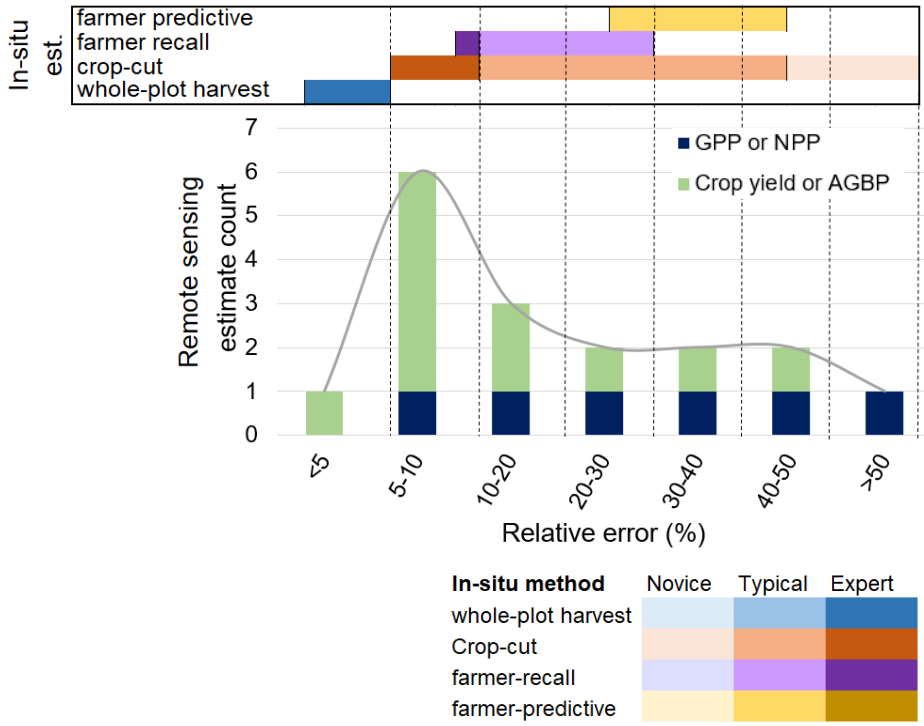

Figure 3.5: Count of relative error ranges of remote sensing-based GPP, NPP, AGBP and yield reported in, or derived from, literature compared to in-situ relative errors.

Integration of remote sensing into crop models through data assimilation methods is becoming more prevalent, including models such as the Simple Algorithm for Yield estimated (SAFY) (Battude et al. 2016) and Simulateur mulTIdisciplinaire pour les Cultures Standard (STICS) (Brisson et al. 2003; Duchemin et al. 2008). The integration of remote sensing and models have been well synthesised previously by Delécolle et al. (1992) and more recently by Jin et al. (2018). Further research is being undertaken to integrate remote sensing derived canopy state variables at larger scales (Jin et al. 2018; Kasampalis et al. 2018). Another promising approach being developed is the generalised regression based model. This model relates the seasonal VI peak to crop yield. However, the regression currently utilises a crop specific slope (e.g. wheat) and is only suitable at administrative unit or county scale (Azzari et al. 2017; Becker-Reshef et al. 2010; Franch et al. 2015).

4.2 Error introduced to account for crop type

However, in remote sensing, the AGBP, GPP or NPP is more commonly available than crop yield. The accuracy of the AGBP should therefore be high enough to meet the crop yield user requirements after the HI, f and biomass moisture content (θ) is applied. The HI varies with the environment (Hay, 1995), cultivar (Ismail, 1993), breeding and agronomic practices (Sinclair, 1998).

Uncertainty of HI has not been established. Ranges of HI vary significantly for crop types and varieties (Hay, 1995). In an Australian literature review large ranges in HI were reported for grain crops; for example wheat, barley and maize HI were found to range between 0.08-0.56, 0.09-0.57 and 0.41-0.62 respectively (Unkovich et al. 2010). In a global review of various crops Hay (1995) also reported large HI ranges; for example rice, chickpea and potato HI was reported between 0.35-0.62, 0.28-0.36 and 0.47-0.62 respectively. Additionally, variability in moisture content will introduce some error, and many reported HI do not indicate the moisture content. Various models have been developed to estimate HI, but most pertain to grain crops (Fereres and Soriano, 2007; Kemanian et al. 2007; Sadras and Connor, 1991). Moisture content can vary significantly with crops; for example, a typical moisture content of wheat, rice and potato yields are 11% (Unkovich et al. 2010), 21% (Unkovich et al. 2010) and 79% (Rees et al. 2012), respectively. It is most common to adapt the HI and the θ to the local application, as applied by Zwart and Bastiaanssen (2007), Bastiaanssen and Ali (2003) and Singh et al. (2006). Alternatively, the provider can compute CWP as a function of AGBP where local users apply HI and θ to estimate CWP as a function of crop yield. This will minimise the error introduced from these factors, particularly between cultivars. Yuan et al. (2016) showed significant reductions in accuracy in estimated crop yield from remote sensing as compared to GPP when using the EC-LUE method. They attributed the reduction in certainty to HI. This again highlights the error these factors introduce. FAO (Raes et al. 2018) includes values for the HI within the Aquacrop model, with a set upper bound and empirical relations to stress factors such as temperature and moisture deficit. This has not yet been applied in remote sensing; however, it may provide insight for developments in remotely sensed crop yield algorithms.

Additionally, several authors have identified the need to distinguish LUEmax based on crop type. Xin et al. (2015) identified a large variation in GPP LUE for different crops, highlighting the importance of correcting generalised datasets for factors including not only HI and moisture content, but also maximum LUE. Bastiaanssen and Ali (2003) compiled LUEmax values from literature, which varied significantly between crops, particularly between C3 and C4 crops. The importance of distinguishing LUEmax between C3 and C4 crops was

also highlighted by the work of Yan et al. (2015) and Yuan et al. (2015). Other authors have incorporated lookup tables for LUEmax, based on land cover type and crop type, into their generalised approaches (Bastiaanssen and Steduto, 2017; FAO, 2018).

Without integrated physical approaches to estimate HI, f, θ, and LUEmax, accurate land classification is important to ensure that the appropriate crop specific conversion factors or look-up tables for the AGBP fraction, HI and LUEmax are used. This is particularly difficult in areas with small plot sizes and mixed cropping patterns.

4.3 Evapotranspiration

The accuracy of ET_a is better described and summarised in literature than that of crop yield. Several methods have been developed over the past decades to estimate ET_a with the most common being the surface energy balance approach. The WaPOR database estimates ET_a based on a remote sensing Penman-Monteith approach. Like in-situ ET_a methods, significant work has been done in summarising the accuracy of ET_a through remote sensing. Therefore, the expected accuracies and uncertainties from remote sensing ET_a are only briefly described.

Karimi and Bastiaanssen (2015) compiled 33 research papers to investigate the error associated with remote sensing-based ETa estimation on an annual or seasonal scale. They demonstrated that the absolute error of ETa varied from 1% to 20% and the MAPE was 5.4% with a standard deviation of 5.0%. The MAPE increased slightly when considering only the error of ETa estimates over cropped areas, with 60% of the studies achieving an error of ≤5%. However, these errors were often associated with algorithms that have been both developed and tested where local parametrisation and calibration were possible. This is consistent with more recent studies, such as Yilma (2017), who reported a mean differences between SEBAL estimated ETa and lysimeter ETa of 9.3 and 15.4% for onion and potato fields respectively. The range of errors were reported to be 1.3-23.4%.

Kalma et al. (2008) assessed 30 published literature, with 20 covering cropped areas, on various methods to estimate ETa from remote sensing. The time-steps in the review ranged from instantaneous (during overpass) to 16-day averages, while the spatial resolution ranged from 30 m to 1 km. A typical error of 15-30% was reported when compared to in-situ measurements. Similarly, Verstraeten et al. (2008), Glenn et al. (2007) and Jiang et al. (2004) found typical errors of 20-30% for various methods with similar ranges in time-steps. The authors did not identify a link between accuracy and spatial resolution. However, Kalma et al. (2008) noted that temporal resolution and scaling did have a large impact on uncertainty. This is due to the strong bias of surface temperature values on minimal cloud

cover days when scaling from daily to weekly or monthly values, and the influence of nocturnal transpiration when scaling from instantaneous, 30 min or hourly estimates to daily estimates. Glenn et al. (2010) reported that the heterogeneity in a pixel attributes to lower accuracy. Lower spatial resolution should therefore reduce accuracy due to the higher the chance for heterogeneity within a pixel.

Validation has been undertaken on current operational global ET_a models MOD16 (Mu et al. 2011, 2007) and EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA-SAF MSG ET) (Trigo et al. 2011). MOD16 (1km, daily resolution) is based on the PM model and LSA-SAF MET (3 km at the nominal position at 0° longitude, 30-minute) ET is based on a simplified soil-vegetation-atmosphere transfer scheme (SVAT). MOD16 and a further improved version, which included the addition of soil heat flux, simplification of the vegetation cover fraction, and improved estimated of stomatal conductance and boundary layer resistance, were compared to 46 EC sites in the USA (7 being cropped sites). The improved version had a mean daily bias of 0.31 mm day^{-1} and had values within 10-30% of the tower values (Mu et al. 2011). The difference in the total annual ET_a at cropped sites between EC and MOD16 was 11.8%. The mean average error of the improved algorithm at cropped sites ranged from 0.16-0.48 mm day^{-1} or 9-30% with a mode error of 0.3 mm day^{-1} or 20%. The MOD16 errors in croplands ranged from 36-53%. The improved version saw the larger improvements in cropped and grass sites as compared to forest sites. The authors found MOD16 underestimated ET_a in croplands.

The performance of global models, on the 8-day time step, is consistent with accuracy reported from literature (as discussed above). More recently, Hu et al. (2015) compared both MOD16 (1 km, 8-daily) and LSA-SAF MET ET to 15 EC sites (2 cropped sites) in Europe. LSA-SAF MET ET performed better in terms of r, RMSE and bias in all sites, including cropped sites. Specifically in the 2 cropped sites, the LSA-SAF MET ET had R^2 of 0.93 and 0.92, RMSE of 0.52 mm day^{-1} for both sites, and bias of -0.10 and 0.27 mm day^{-1}. MOD16 had R^2 of 0.90 and 0.91, RMSE of 0.72 and 0.47 mm day^{-1}, and bias of -0.39 and 0.26 mm day^{-1}, respectively. The high agreement is in spite of the site heterogeneity, as the pixel extends beyond the cropped site for both MOD16 and LSA-SAF MET and includes mixed cropping patterns and urban area at both sites. LSA-SAF MET ET set quality criteria as error < 25% when ET_a is > 0.4 mm day^{-1} and <0.1 mm day^{-1} when ET_a is < 0.4 mm day^{-1}. This criterion was met in 70% of instances for 15 stations in Europe.

Ershadi et al. (2014) compared SEBS, PM, advection-aridity (AA) model and a modified Priestley–Taylor (PT-JPL) approach at 20 FLUXNET stations across the USA, including four cropped sites. The relative error at cropped sites were 38%, 56%, 61% and 38% for the SEBS, AA, PM and PT approaches respectively. The grass sites showed

similar results. The AA methods performed best in grassland (relative error = 73%). None of the approaches performed consistently in croplands. The R^2 was highest in the crop and grass sites for both SEBS, at 0.76 and 0.78 respectively, and PT-JPL approaches, at 0.74 and 0.77 respectively.

Various surface energy balance models show inconsistent results when modelled at different locations. Singh and Senay (2016) compared four energy balance methods for estimating ET_a, SEBAL, METRIC, SEBS and SSEBop with EC in three cropland sites in a humid continental climate in the USA. METRIC and SSEBop had the best performance, with relative errors (daily) between sites of 2.5-13.7% and 7.1-12.6% respectively. The SEBAL and SEBS models performed considerably worse; they typically understated ET_a, especially on days when ET_a was high, with relative errors of 39.6-42.6% and 25-31.1% respectively. The authors attributed higher errors in SEBAL and SEBS to the method of upscaling of instantaneous to daily ET_a and suggested that the daily net radiation equation in SEBAL should be calibrated to local atmospheric conditions. These remote sensing energy balance methods, along with S-SEBI, were also validated against EC at 4 sites, including a grassland and citrus orchard, in the USA (Bhattarai et al. 2016). Overall, SEBAL had the lowest percent bias (1%), followed by SEBS (3%), S-SEBI (-8%), METRIC (16%), and SSEBop (36%). SEBS had the lowest RMSE (0.74 mm day^{-1}) and SSEBop had the highest RMSE (1.67 mm day^{-1}). The performance at all sites, except the lake which performed worse, were comparable for the SEBS, S-SEBI, and METRIC models. SSEBop had the worst performance of five surface energy balance models.

Most recently, Khan et al. (2018), used triple collocation to provide mutually uncorrelated absolute and relative error structure among MOD16, Global Land Evaporation and Amsterdam Model (GLEAM), and Global Land Data Assimilation System (GLDAS) ET_a products. The three products performed well in nine EC sites (AsiaFlux), including three rice paddy and three grassland, with RMSE ranging between 3.69-12.98 mm 8-day^{-1} in the rice paddy and grassland sites. However, all four datasets, including the EC data, had relative uncertainties exceeding 25%.

Karimi et al. (2019) compared ETa from a SSEBop and CMRSET Ensemble product, downscaled with 250 m MODIS NDVI to the gross inflows (effective precipitation plus irrigation withdrawals) in irrigated sugarcane in Swaziland. The annual ETa from the Ensemble had relative bias of -5%, a RMSE of 9% and a relative error of 7%, as compared to the net inflows. This can be attributed to the groundwater table being assumed to be steady, as the water table depth influences soil moisture content. Therefore, errors may be higher than reported here.

Other authors have compared remote sensing-based ET_a to basin scale water balances. For example, Senay et al. (2011) compared annual ET_a estimates derived from SSEB to watershed water balances around the globe. The agreement between SSEB ET_a and water balance ET_a was very high. The R^2 was 0.9 and the mean annual bias was -67 mm, or 11%. Senay et al. (2016) compared SSEBop to the water balance from Colorado River Basin, USA and to the ET_a estimated from two EC stations. SSEBop ET_a showed relative bias, on an annual scale, of 7.3%, 10% and -0.5% for the total, upper and lower part of the basin. SSEBop also showed a lower agreement at the EC stations. The R^2 values were 0.82 at both EC stations on a daily scale and 0.92 and 0.95 at each station on a monthly scale. However, the relative bias was varied, with daily relative bias of -22.1% and 13.1% at daily scale and -34.7 and 2.4% on a monthly scale.

The Kcb remote sensing-based approach is not a generalised approach; however, it is discussed here briefly due to its popularity and potential. Kc and Kcb have been empirically related to VI in remote sensing for more than 30 years (Bausch and Neale, 1987; Neale et al. 1989). The Kcb remote sensing approach estimates ETa based on the Kcb-VI relationship. The relationship has been described by various empirical equations including a power function (Nagler et al. 2013), as a scalar (Glenn et al. 2011) and as a linear function (Choudhury et al. 1994; Melton et al. 2012; Nagler et al. 2013). Accuracy from such methods have been reported to be as high as 5-15% (Duchemin et al. 2006; Hunsaker et al. 2007; Nagler et al. 2013). A review by Glenn et al. (2010) found RMSE in the range of 10-30% across biomes and vegetation including, wheat, corn and cotton.

Several authors have been able to extrapolate Kcb-VI relationships between crops (Campos et al. 2013, 2010; Odi-Lara et al. 2016), and a generalised approach has been suggested for major crop categories, i.e. vegetables, tubers, legumes, fibres, oils, cereals (Melton et al. 2012). On the other hand, Calera et al. (2017) summarised Kcb-NDVI relationships found in literature for different crops. Each study had unique Kcb-NDVI relationship, whether for the different crops or the same crop in different locations. Mateos *et al,* (2013) validated a synthetic crop coefficient approach (Kcs) for estimated ET_a under non-stressed conditions in Spain. The approach was then applied at basin scale in the Guadaliquivir Basin (González-Dugo et al, 2013). The approach required prior information of crop location and the crop-growing cycle. The approach performed well for annual and tree crops (except olive), however, was less successful for seasonal crops. The overall RMSD was 0.75 mm day^{-1}. The authors suggest the weakest part of the model is the soil evaporation component and that further work on the Kcb-VI relationship is required for more crops. Therefore, the Kcb-VI relationship cannot always be

extrapolated directly to new locations. However, it has been shown that once the relationship is developed for a specific crop and location, it can be a very reliable method for that area.

Like remote sensing-based crop yield estimates, remote sensing-based ETa estimates show a large range of reported errors. Locally parameterised and calibrated ETa models have been validated numerous times, however, the validation of global models in crop areas is less common and more difficult. There is scarce ground data for cropped areas when compared to the spatial extent of the global models. The reported accuracy of remote sensing-based ETa varies widely, between location and models. Karimi and Bastiaanssen (2015) suggest remote sensing-based ETa error on an annual scale can be as low as 5%, which is the same accuracy associated with lysimeters, while Kalma et al. (2008) suggest accuracy in the range of 15-30%, which is in the same range as expert and typical errors associated with the soil water balance, Bowen ratio, and EC. Reported errors of generalised models varies considerably. Some models have reported errors of less than 15% (Bhattarai et al. 2016; Singh and Senay, 2016) while other models have reported errors of greater than 40% (Bhattarai et al. 2016; Ershadi et al. 2014). The latter is within the range of in-situ lysimeter, Bowen ratio, and EC measurements when performed by a novice.

4.4 Crop water productivity

Remote sensing-based estimates of CWP error are derived from the reported error of remote sensing-based crop yield (and GPP) and ET_a errors. The lowest reported remote sensing-based crop yield and ET_a errors are in the range of 5-10% and 5-20% respectively. This corresponds to a best case scenario of a CWP relative error of 7.1%-22.4%. Other case studies reported errors up to 70-90% for crop yield and 25-60% for ET_a. This propagates to CWP ranges 74.3%-108.2%. This corresponds well with the only cited literature on validating remote sensing-based CWP in croplands (through EC GPP and ET_a), which reported errors of 82.3% and 14.7% on an annual scale for soybean and 21.2% and 30.9% on an annual scale for maize (Tang et al. 2015). This suggests that under the right conditions, remote sensing-based CWP can be estimated to a similar accuracy of the combination of field-based measurement techniques, like farmer recall combined with EC or soil water balance.

The greatest challenge in synthesising the accuracy of remote sensing datasets for CWP was the heterogeneity of error reporting. This was also noted by other authors who reviewed accuracy of remote sensing products (Karimi and Bastiaanssen, 2015). Reporting a number of accuracy metrics is crucial for reporting and understanding scientific results. It was identified that relative error was frequently not

explicitly stated (or able to be derived). While relative error may not provide the complete picture or error characterisation, it does allows for: (i) comparison between products, for example yield and GPP, and (ii) error propagation, which is required to ascertain the achievable accuracy of CWP.

The identified crop yield, ET_a, and CWP error estimates are valid and based on an exhaustive literature review. However, they do not comprehensively consider the errors within the validation process itself, which can include: comparing the remote sensing value to field measurements with their own inherent error, error characterisation, issues with spatial and temporal scaling between remote sensing and in-situ products, and scale issues between the resolution of the remote sensing products and the scale in which they are required by the user (Zeng et al. 2015).

Remote sensing estimates are comprised of both random and systematic errors. Random errors are caused by unknown and unpredictable changes and are always present; systematic errors are consistent and introduced by the inaccuracy inherent to the system. Random errors are typically normally distributed and can be represented by the standard deviation of their distribution (Povey and Grainger, 2015). In certain applications of CWP, a systematic error will have a lower impact on the analysis. For example, when undertaking a comparative assessment – of one user to another or the same user over time (all estimated under the same model) – a systematic bias should not influence the result. However, in estimating absolute values of CWP and comparing to other studies or literature, systematic bias could significantly misinform the user. Many of the studies reported on bias, which can help the user identify if the errors in the remote sensing dataset is dominated by systematic error.

The point spread function (PSF) effect describes the response of the imaging system to the point source or object (Mira et al. 2015; Van der Meer, 2012). This effect means that the signal for a given pixel is a weighted combination of contributions from within the pixel and also contributions from neighbouring pixels, based on the across-track and along track directions. This effect introduces the greatest uncertainty in heterogeneous landscapes (Duveiller et al. 2015, 2011).

Field based observations have their own uncertainties, and remote sensing-based estimates are being compared to field methods which frequently have errors exceeding 20% (Nagler et al. 2013; Nouri et al. 2016). All the literature cited reported the remote sensing-based errors against the value of the field observation, thus accepting the field observation as the true value. However, as discussed in Section 3 of this Chapter, the field observations are associated with their own, often significant errors. Triple collocation attempts to deal with this issue by characterising error, both systematic bias and random error,

through observing the spatial and temporal difference in three independent datasets. However, triple collocation requires multiple datasets with large numbers of coincident data points, including in-situ (Su et al. 2014), that are not frequently available for ETa and crop yield. The only cited literature using this method for ETa found relative uncertainties exceeding 25% in both the remote sensing-based data and the in-situ observations (Khan et al. 2018). Ultimately, the actual accuracy of remote sensing is constrained by the accuracy of the field measurements they are compared to (Glenn et al. 2011).

In-situ measurements not only have their own sources of error, they can prove difficult for comparison with remote sensing data due to spatial and temporal scaling issues and the difficulties of identifying the area of representation. Scaling issues when comparing remote sensing and in-situ measurements arise from: (i) the local and point scale measurement being compared with a spatially continuous dataset (Ran et al. 2016), (ii) the sparsity and availability of point measurements in both time and space, (iii) vegetation heterogeneity within a pixel (Clark et al. 2001; Foken and Leclerc, 2004; Stoy et al. 2013), (iv) geolocation errors, and (v) systematic errors, e.g. Foken (2008) suggested the main cause of errors in EC are due to the different spatial and temporal scales in the energy balance components.

User requirements will vary dependent on the specific application of CWP. The user could be estimating a time series of a single user (inter- or intra-seasonally), a comparison of users in a season within an irrigation scheme, comparing an irrigation scheme to another irrigation scheme, assessing whether the CWP meets local or national targets, setting CWP targets, or considering the CWP for a basin scale to assess user demand. Each of these applications may require not only a different accuracy, but also a different spatial resolution. Reported accuracies in this review cover a large range of sensors with varying resolutions; for example, the Landsat sensor has a spatial resolution of 30 m while the MOD16 has a spatial resolution of 1km. The spatial resolution of the dataset not only influences the dataset accuracy – as pixel heterogeneity has a significant influence on accuracy (Liu et al. 2016) – but the applicability of the product. For example, a 30 m product may be used to estimate in-field variability (Kharrou et al. 2013) while a 1km product may be limited to estimate inter-scheme variability or inter-annual variability at scheme level (Al Zayed et al. 2016).

Reported errors are related to not only to specific spatial resolutions, but also temporal resolutions. While some authors reported error on a seasonal scale, others reported error on a daily scale (e.g. Yan et al. 2015) or at the resolution of the satellite return period (e.g. Wang X. et al. 2013). This creates a temporal scale

mismatch between the satellite derived product and the field observation. The scale mismatch requires either aggregation of the high resolution of dataset to the low resolution dataset, averaging over the same period or disaggregation of the low resolution dataset. Further, it can be difficult to compare the accuracy of remote sensing products that are reported with different temporal and spatial scales. However, it is important to provide accuracies at all available scales. CWP is a seasonal product with an associated error at a seasonal scale; however, the user often aggregates daily values to a seasonal product that may range from a few months up to two years.

The accuracy of remote sensing is typically derived from comparison to in-situ measurements or estimates. CWP relative error derived from in-situ measurements are low, 7-11%, when undertaken by an expert (Figure 3.4). However, the typical errors have large ranges from 7-36% (not including sap flow). Though this reportedly aligns with the accuracy of CWP from remote sensing, the application must be considered as even the accuracy of in-situ methods may still not be suitable for all user applications.

When CWP values are being used as absolute values, rather than relative to other users, the scale of error may be related to the required precision. For example, the ETa precision required for irrigation can be low quite low for some irrigation methods. This is reflected in potential distribution uniformity, which ranges from 60% for furrow irrigation to 90% for drip irrigation (Brouwer et al. 1989). The actual CWP required for the purpose of understanding consumption and efficiency may be different than the precision useful for a farmer to make yield or ETa improvements. There is no use stipulating an accuracy or precision requirement for a farmer, if the farmer cannot achieve that accuracy with their inputs such including irrigation application.

With the onset of the WaPOR database, the continental dataset is expected to be more frequently applied in local settings. Therefore, the accuracy of the global models should be carefully validated and reported so the user can determine if it meets their requirements. The accuracy of remote sensing should at least be comparable to the accuracy between the expert and typical ranges of ground measurements. This can be difficult to prove and quantify for large scale remote sensing datasets. Any developments and improvements in remote sensing will be difficult to prove without improvement of the in-situ measurements they are validated against (Glenn et al. 2007). It is also essential to provide an accuracy metric (such as a relative error) that the user can clearly understand in order to determine if the dataset is useful for their application.

5. Conclusions

The main objective of this research was to assess the accuracy of remotely sensed and in-situ CWP products. Remote sensing provides a tool to estimate CWP at much larger extents and in areas where field measurements are not available. CWP datasets are typically not provided as a remote sensing product; however, its two main constituents, yield and ETa, are. The accuracy of CWP was therefore derived by propagating the reported accuracy of both remote sensing and in-situ ETa and crop yield. The in-situ methods were first described. In-situ methods have commonly been used to understand crop performance and are typically used as the reference value for remote sensing estimates to quantify their accuracy. They are ascribed as the benchmark for accuracy of CWP. The reported accuracy of remote sensing-based methods were then synthesised and compared to the benchmark, or the error accuracy with in-situ products.

The error associated with in-situ methods for estimating crop yield ranges from <5% (whole-plot harvest) to 45% (crop-cutting and farmer surveys), while for estimating ETa it ranges from 5-15% (lysimeter) to 50% (sap flow measurements). This propagates to CWP errors from field measurements that range from 7-67%. Based on remote sensing reported accuracy of ETa and yield (or GPP), the best case scenario of a CWP relative error from remote sensing is in the range of 7.1-22.4%. Other case studies reported errors up to 70-90% for crop yield and 25-60% for ETa, which propagates to CWP ranges of 74.3%-108.2%.

The literature review revealed that remote sensing can estimate CWP within the error range from in-situ methods. However, the review also revealed a great deal of heterogeneity in the reporting of both errors and uncertainty. The characterisation of error, e.g. random error or systematic bias, will define if the data products are suitable for different applications of CWP. Further research should describe the way in which these errors are reported to ensure that end-user requirements are met. It was also identified that the gap between remote sensing estimates of GPP and crop yield needs further development, as large uncertainty lies with the intermediates that convert GPP to yield.

Chapter 4

Evaluation of WaPOR V2 evapotranspiration products across Africa[2]

[2] This chapter is based on: Blatchford, M.L., Mannaerts, C.M., Njuki, S.M., Nouri, H., Zeng. Y., Pelgrum, H., Wonick, S., Karimi, P., 2020. Evaluation of WaPOR V2 evapotranspiration products across Africa, Hydrological Processes - Special Issue in honour of Ed Glenn (2020) 34:3200-3221.

Abstract

The Food and Agricultural Organization of the United Nations (FAO) portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR) offers continuous actual evapotranspiration and interception (ETIa-WPR) data at a 10-day basis across Africa and the Middle East from 2009 onwards at three spatial resolutions. The continental level (250 m) covers Africa and the Middle East (L1). The national level (100 m) covers 21 countries and four river basins (L2). The third level (30 m) covers eight irrigation areas (L3). To quantify the uncertainty of WaPOR version 2 (V2.0) ETIa-WPR in Africa, we used a number of validation methods. We checked the physical consistency against water availability and the long-term water balance and then verify the continental spatial and temporal trends for the major climates in Africa. We directly validated ETIa-WPR against in-situ data of 14 eddy covariance stations (EC). Finally, we checked the level consistency between the different spatial resolutions. Our findings indicate that ETIa-WPR is performing well, but with some noticeable overestimation. The ETIa-WPR is showing expected spatial and temporal consistency with respect to climate classes. ETIa-WPR shows mixed results at point scale as compared to EC flux towers with an overall correlation of 0.71, and a root mean square error of 1.2 mm day^{-1}. The level consistency is very high between L1 and L2. However, the consistency between L1 and L3 varies significantly between irrigation areas. In rainfed areas, the ETIa-WPR is overestimating at low ETIa-WPR and underestimating when ETIa is high. In irrigated areas, ETIa-WPR values appear to be consistently overestimating ETa. The SMC, the input of quality layers and local advection effects were some of the identified causes. The quality assessment of ETIa-WPR product is enhanced by combining multiple evaluation methods. Based on the results, the ETIa-WPR dataset is of enough quality to contribute to the understanding and monitoring of local and continental water processes and water management.

1. Introduction

The Actual evapotranspiration (ETa) is the second-largest process in the terrestrial water budget after precipitation (PCP). ETa is also an essential component of plant growth and, therefore, the carbon cycle. Available water resources are becoming, or are already scarce, in many basins worldwide (Degefu et al. 2018). The acceleration of the water cycle from a climate change perspective will further influence water availability not only for human consumption but also our food sources (Rockström et al. 2012). For this purpose, accurate estimates of ETa are required for several management tasks, including, but not limited to, water accounting, water footprint, basin-wide water balances, irrigation, crop management and monitoring of climate change and its impact on crop production. These activities require ETa at varying extents and spatio-temporal resolutions.

Remote sensing from satellites is perhaps the only feasible means for quantifying and monitoring ETa for wide-areas (Glenn et al. 2007). Several remote sensing approaches exist to estimate ETa which include, surface energy balance methods (e.g. Bastiaanssen et al. 1998; Su, 2002; Allen et al. 2007), Penman-Monteith methods (FAO, 2020a) and more empirical vegetation indices based methods (Glenn et al. 2008; Nagler et al. 2013). Currently, there are a number of open-access remote sensing-based ETa products based on remote sensing data at the continental and global scale. Global products include: the Moderate Resolution Imaging Spectroradiometer (MODIS) based ETa (MOD16) product (Mu et al. 2011), generated every 8-days at 250 m; the Global Land Evaporation Amsterdam Model (GLEAM) ETa (Miralles et al. 2011), generated daily at 0.25°; the operational Simplified Surface Energy Balance (SSEBop) ETa (Senay et al. 2011), generated monthly at 1km; and the Land Surface Analysis-Satellite Applications Facility (LSA-SAF) Meteostat Second Generation (MSG) ETa (Ghilain et al. 2011), generated daily at approximately 3km.

Validation of these remote sensing products is an essential step in understanding their applicability and characterize uncertainty. This uncertainty can guide if the ETa product is suitable as input into different water management activities along with the associated risk when making a decision based on the product. Many studies exist that attempt to validate large remote sensing-based ETa datasets. Most studies are focused on one or two validation methods at one scale. The most common validation methods are either point or pixel scale against ground-truth data, like eddy covariance (EC) measurements (e.g., Mu et al. 2011), or spatial inter-comparison of a product over regions, land classes, biomes (e.g., Mueller et al. 2011). Some authors validate multiple products against each other for spatial and temporal patterns and against ground-truth data (e.g., Hu et al. 2015; Nouri et al. 2016). Liu et al. (2016) evaluated basin-scale ETa estimates against the water

balance method. Velpuri et al. (2013) compared MOD16 (1km) at point scale to EC and at basin scale to the water balance. Other than a few occasions, for example, Velpuri et al. (2013), these validation efforts often failed to evaluate the product at multi-scale, from pixel to basin or region.

The best-practice validation strategies of big remote sensing datasets have been proposed by (Zeng et al. 2019; 2015). They recommend multi-stage validation activities that include combinations of direct validation, physical validation and cross-comparisons. In practice, many developers of remote sensing products include all or at least a combination of these activities during their validation. To name a few, these include the MODIS MODLAND product (Morisette et al. 2002; Morisette et al. 1998); Copernicus Global Land Service products Dry Matter Productivity (Swinnen et al. 2015); and ASTER land surface temperature (Schneider et al. 2012).

In regions such as Africa, where little observational data is available, validation should utilise all available avenues for ascertaining product quality, with a multi-step and -phase validation strategy that includes direct validation (with ground measurements), physical consistency check and cross-comparisons. As such, the limitations due to the sparseness of available data are reduced, and the product quality is understood from a multi-scale perspective, by using validation best-practice and combining multiple validation techniques.

The latest available database of continental products, released in 2019, for Africa and the Middle East, is now available on FAO portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR) (https://wapor.apps.fao.org/home/WAPOR_2/2). It provides the highest available spatial resolution for an operational open-access ETa and interception (ETIa) product at the continental scale. This paper presents a multi-scale validation of the version 2 (V2.0) ETIa. The results from each validation procedure were analysed individually and then as a whole to determine trends and draw conclusions of the product quality.

2. Data and methods

2.1 The dataset

The analysis dataset is the ETIa V2.0 products available on the WaPOR portal and is described in Chapter 2 - section 2. The extent of and the resolution of the dataset is also described in Chapter 2 - section 1. Finally, the WaPOR datasets, including intermediate datasets, input sensors and input data products, used in the validation are provide in Chapter 2 – section 4.

As this chapter contains multiple ETIa datasets, the ETIa sourced from the WaPOR portal is referred to as ETIa-WPR for the rest of this chapter.

2.2 Validation approach and workflow

The validation approach comprises three components, physical validation, direct validation and level consistency (Figure 4.1). The physical validation and direct validation were undertaken on the L1 product for the period 2009-2018. The physical validation (section 4.3) includes an assessment of the water balance and water availability (4.3.1) and a spatial and temporal consistency check (4.3.2) for the extent of Africa. The water balance utilises other existing continental datasets to complete the water balance and is therefore also considered cross-validation. The spatial and temporal consistency checks if spatial and temporal patterns were being captured. The direct validation (section 4.4) involves a comparison to ETa estimations from EC stations. The level consistency (section 4.5) checks for the consistency between levels and therefore indicates if the quality of the L1 product is representative of the L2 and L3 products.

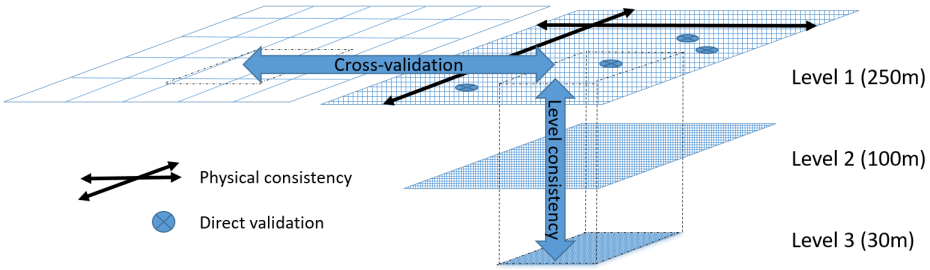

Figure 4.1: Validation approach used in in the validation of the ETIa-WPR product in Africa

2.3 Physical consistency

2.3.1 Water balance and water availability

The basin-scale performance of ETIa-WPR is analysed for 22 major hydrological basins of Africa (Lehner & Grill, 2013) through three approaches (Figure 4.2). First, the ETIa-WPR was compared to the PCP on an annual basis to analyse the water consumed through ETIa to the water available from PCP.

Second, the basin-scale water balance approach compared the long-term ETIa-WPR product to the long-term ETa derived from the water balance (ETa-WB). In many studies, the long-term water balance (>1 year) for large basins assume a negligible change in storage (Hobbins et al. 2001; Wang & Alimohammadi, 2012; Zhang et al. 2012). The long-term water balance, taken from 2009-2018 in this case, is therefore defined using equation 4.4.

$$ETa\text{-}WB = PCP - Q \quad (4.1)$$

Where PCP is the long-term precipitation and Q is the long-term basin run-off or streamflow, and the ETa-WB is the long-term ETa derived from the water balance averaged to mm $year^{-1}$. The PCP product found in the WaPOR portal was obtained from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset (Funk et al. 2015). The long-term Q was obtained from the Global Streamflow Characteristics Dataset (GSCD) (Beck et al. 2015). The GSCD consists of global streamflow maps, including percentile and long-term mean Q at a 0.125° resolution, providing information about runoff behaviour for the entire land surface including ungauged regions.

Third, the ETIa-WPR and PCP annual values were compared to the average ETa from MODIS Global Evapotranspiration Project (ETa-MOD16) for the period 2000-2013 (Mu et al. 2007; Mu et al. 2013) and to values from the literature for basins where data is available. The ETa-MOD16 product is also based on the Penman-Monteith equation and considers the surface energy partitioning process and environmental constraints on ETa. The algorithm uses both ground-based meteorological observations and remote sensing observations from MODIS. Basins were not included in the comparison if the ETa-MOD16 data covered less than 80% of the basin area.

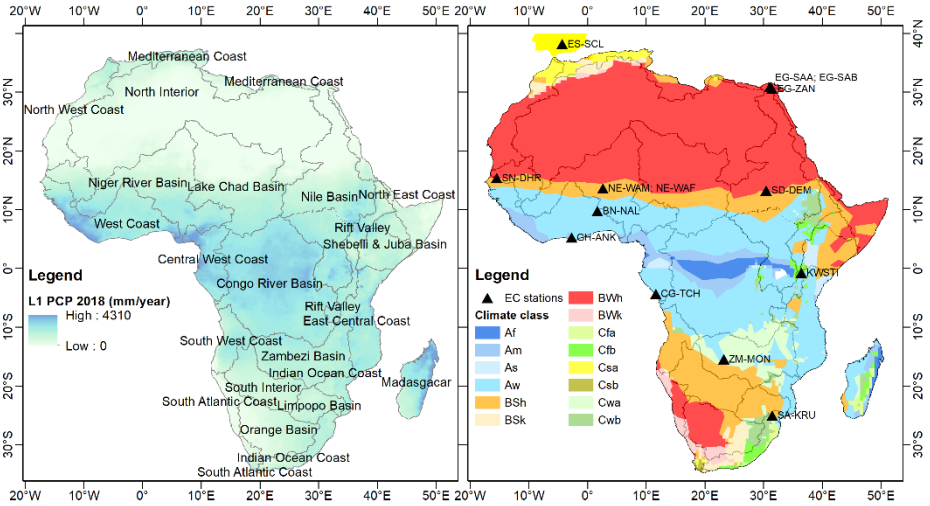

Figure 4.2: The left – 22 major hydrological basins of Africa used in the water balance approach and right – Koppen-Geiger climate classification and eddy-covariance stations where Af – tropical rainforest, Am – tropical monsoon, As – tropical dry savanna, Aw – tropical wet savanna, BSh – arid hot steppe, BSk, arid steppe cold, BWh – arid hot desert, BWk – arid cold steppe, Cfa – temperate without dry season hot summer, Cfb – temperate without dry season warm summer, Csa – temperate dry summer hot summer, Csb – temperate dry summer warm summer, Cwa – temperate dry winter hot summer, Cwb – temperate dry winter warm summer. Note some stations are in close proximity and are there for represented by one point on the map (e.g. NE-WAM and NE-WAF).

2.3.2 Spatial and temporal consistency

The temporal and spatial trends were observed over the African continent in space and time by observing mean ETIa-WPR, SMC and NDVI for all climate zones during the study period on a dekadal basis. The Koppen-Geiger classification (Figure 4.2) is used to consider the mean dekadal values for the main climatic zones in Africa (Kottek et al. 2006). A sample size of 30,000 stratified random pixels is used to represent the continental. This corresponds to less than 0.01% of the total image, however, is considered suitable to represent seasonal trends for the major climate zones. The arid or desert class – B – dominates Africa (57.2%), followed by the tropical class - A (31.0%) and then warm temperate - C (11.8%). The largest sample count corresponds to the largest climatic zones, with a linear 1:1 line representing area to count. The data is further disaggregated based on the northern and southern hemispheres to account for opposite seasonal patterns.

2.4 Direct Validation

The ETIa-WPR is compared to the in-situ ETa from EC fluxes (ETa-EC) at a dekadal scale using 14 locations (13 across Africa and 1 in the Spain extension area) (Figure 4.2). The country, station code, vegetation, climate zones and available data for comparison – for both WaPOR and the local site, are shown in Table 4.1. The majority of EC sites are in shrubland or savannas. Egypt stations (EG), the NG-WAM station and GH-ANK station which are located in an irrigated area, agricultural land and forested areas respectively.

The SA-SKU, SNDHR, GH-ANK, SD-DEM, CG-TCH, ZM-MON and ES-SCL EC sites were obtained from the global Fluxes Database Cluster Dataset (FLUXNET). The FLUXNET 2015 (https://fluxnet.fluxdata.org/) dataset consist of open-source high-quality data products collected from multiple regional networks. The NE-WAM, NE-WAF and BN-NAL sites were obtained from the African Monsoon Multidisciplinary Analysis—Coupling the Tropical Atmosphere and the Hydrological Cycle (AMMA-CATCH) project, aiming at establishing long-term observations on the climate and the environment over Western Africa. KWSTI is operated by the ITC-UTWENTE in partnership with Water Resources Management Authority (WRMA), the Kenya Wildlife Services (KWS) and Egerton University. The EG-ZAN, EG-SAA and EG-SAB sites were operated through the University of Tsukuba, in partnership with Cairo University, National Water Research Center, Delta Barrage, Qalubia, Egypt and the Agriculture Research Center, Giza, Egypt in the Nile Delta. These irrigated sites in the Nile Delta, were under rotation with three major summer crops – rice, maize and cotton – and four major winter crops – wheat, berseem, fava beans and sugar beet.

Table 4.1: EC site data and descriptions

Site	Country	Ecosystem	Climate	Data-years used
SA-SKU	South Africa	Savannas wooded grassland	BSh	2009; 2011
SN-DHR	Senegal	Savannas	BWh	2010-2013
SD-DEM	Sudan	Savannas	BWh	2009
NE-WAM	Niger	Crops (millet, bare soil, tiger bush)	BSh	2009-2012
NE-WAF	Niger	Crops (fallow; shrubs)	BSh	2010-2011
ES-SCL	Spain	Pasture and Scatter oak trees	Csa	2016-2017
GH-ANK	Ghana	Evergreen broadleaf forests	Am	2011-2014
BN-NAL	Benin	Guinean savanna vegetation	Aw	2009

KWSTI	Kenya	Open shrubland	Cfb	2012-2014
CG-TCH	Republic of Congo	Savanna grassland	Aw	2009
ZM-MON	Zambia	Savanna woodland	Cwa	2009
EG-ZAN	Egypt	Irrigated agriculture	BWh	2011-2013
EG-SAA	Egypt	Irrigated agriculture	BWh	2011-2013
EG-SAB	Egypt	Irrigated agriculture	BWh	2011-2013

ETIa-WPR for L1 (250 m) were spatially averaged over a 3x3 pixel window surrounding the EC station, based on the assumption that the window represents the measurement footprint of the EC station. The ETa-EC data was derived from LE flux and then aggregated temporally to dekadal averages to match the temporal resolution of the ETIa-WPR products. Intermediate products, including WaPOR NDVI, SMC and the NDVI and LST quality layers were analysed along with the ETa trends to identify possible sources of error. Reworking the LE flux data to daily values was done (accounting for NaN, non-removed spikes, early morning (dawn) and evening (day-night inversions), dew spiking, etc.) which are not necessarily removed by the standard Eddy Covariance pre-processing software's (converting very high frequency sonic 30-sec and gas analyzer measurements to 30-minute interval fluxes).

The EC method, as a validation method for remote sensing, contains its own inherent errors of up to 10-30%. This uncertainty is related to a number of causes included scale mismatch (where the area of the footprint compared to the remote sensing area compared only partially overlaps), canopy heterogeneities, and measurement problems (Allen et al. 2011). However, it is fairly common for authors to use EC in heterogeneous landscapes in both validating and driving large remote sensing-based studies (e.g. ETa - Mu, Zhao and Running 2011; ETa - Velpuri et al. 2013; Sjöström et al. 2013).

The ETIa-EC was also compared against in-situ VPD (VPD-EC) and ETo (ETo-EC). In WaPOR, the VPD and ETo are estimated using GEOS-5 (VPD and ETo) and MSG (ETo only), as compared to being derived from satellite images. GEOS-5 and MSG are available daily and satellite image gaps do not influence the quality of the VPD and ETo quality. The ETo-EC was estimated using the same method adopted by WaPOR (FAO, 2020a), which is based on FAO-56 (Allen et al. 1998), and was derived from in-situ (EC) meteorological data.

$$ETo = \frac{\Delta(\mathrm{Rn}-\mathrm{G})+\frac{\rho_{air}\,\mathrm{C}_P(\mathrm{e}_{sat}-\mathrm{e}_a)}{\mathrm{r}_a}}{\Delta+\gamma\,(1+0.34\cdot\frac{\mathrm{r}_s}{\mathrm{r}_a})} \qquad (4.2)$$

Where rs is taken as 70 s m^{-1}, ra is taken as 208/u_{obs} and u_{obs} is the observed wind speed (m s^{-1}) at 10m.

2.5 Level consistency

L3 and L2 ETIa-WPR were compared to the L1 data for the period of 2009-2018 on a dekadal basis. A bilinear resampling method was used to spatially aggregate the high-resolution L3 and L2 layers to the resolution of the coarse L1 layer. A random stratified sample of 30,000 points over the entire L2 extent is used for the comparison of the L1 and L2. The L1 and L3 were compared over the entire L3 extent of the Awash, Zankalon, Office du Niger (ODN) and Koga L3 irrigation areas for all pixels. Table 4.2 shows the description of each L3 irrigated area. The EC station at Zankalon is located in a L3 area. Therefore, as part of the level consistency, all three levels were also compared to the ETa-EC at this station. The method described in section 2.4 was used to extract the L3 and L3 ETIa-WPR at the station.

Table 4.2: Description of L3 irrigated scheme areas used in the product evaluation

	Awash	Koga	Zankalon	ODN
Average plot size of irrigated area (ha)	10.40	0.24	0.21	5.93
SD plot area (ha)	6.24	0.12	0.13	0.46
Major crops in the irrigated area	Major: sugarcane. Minor: haricot, crotalaria	Wheat, maize, potato, onion, cabbage, barley	Wheat, rice, maize, cotton, sugar beet, berseem, fava bean, tomato, potato	Rice, sugarcane
Vegetation in the non-irrigated area	Rainfed agriculture, bare/natural vegetation	Rainfed agriculture: maize, millet, teff, barley	NA	Sparse arid

3. Results

3.1 Physical consistency

3.1.1 Water balance and water availability

The annual ETIa-WPR divided by the annual PCP (ETIa/PCP) during 2009-2018 for Africa is shown in Figure 4.3. The annual ETIa-WPR exceeds the annual PCP (ETIa/PCP >1) on 55% occasions for all basins over the ten years study period. The highest number of exceedances occur in 2014 and 2016 (64%), and the lowest number of exceedances occur in 2018 (27%). The majority of these exceedances, 66%, are by less than 10%. The average ETIa-WPR to PCP ratio for the continent of Africa is 0.93. The lowest ratio is in 2010, 0.87, and the highest is in 2015, 0.97. These ratios are significantly higher than the suggested average, 0.65, of ETa to PCP ratio over the global terrestrial surfaces (McDonald, 1961). This ratio is expected to be lower in dry regions or parts of the continent. Except for Lake Chad Basin, basins in the Central, North and West of Africa have ETIa-WPR less than PCP. Most of the exceedances (ETIa>PCP) occur in the South of Africa and on the Horn of Africa.

The basins have the highest ETIa-WPR/PCP ratio in 2015, particularly in Southern Africa. All basins south of Zambezi Basin show a significant decrease in PCP from 2014 to 2015, including a 246, 98 and 238 mm $year^{-1}$ drop in Limpopo, Orange and the South Interior respectively. In the same timeframe, the largest ETIa-WPR change is in Limpopo, with a 17 mm $year^{-1}$ increase, followed by the South Atlantic Coast with a 35 mm $year^{-1}$ decrease. The decrease in PCP is attributed to the drought in this region during this period as a result of the El Nino climatic event (USAID, 2016). However, ETIa-WPR does not seem to respond appropriately to these extreme drops in PCP, which is likely because the SMC does not show any significant response to reduced PCP in this period. The PCP drop in 2015 in drought affected basins ranged from 16.8-39.1% of the 2009-2018 average while the SMC drop only ranged from 2.2-6.0%. Therefore, the ETIa-WPR is not being properly limited by reduced water availability in the soil.

The average (av.), minimum (min) and maximum (max) annual ETIa-WPR and PCP values for the 2009-2018 period are shown in Table 4.3. Where literature values were available, annual estimates of ETIa-WPR and PCP are compared with historical estimates on annual ETa and PCP, with ETa from MODIS Global Evapotranspiration Project (ETa-MOD16) and with the ETa-WB. In most cases, the ETIa-WPR is larger than the ETa values in literature, from the water balance and from MOD16. The PCP falls within the range of literature for all but three basins. The PCP is less than that found in literature in the Limpopo and Orange Basin, which is also likely due to the drought in this region which occurred after the estimates as reported in the literature. It is also important to note that the Congo River Basin, Central West Coast and west coast basins have vast areas of low-quality NDVI and LST layers for much of the year. They are making the annual mean ETIa-

WPR values derived from remote sensing much less reliable in these basins.

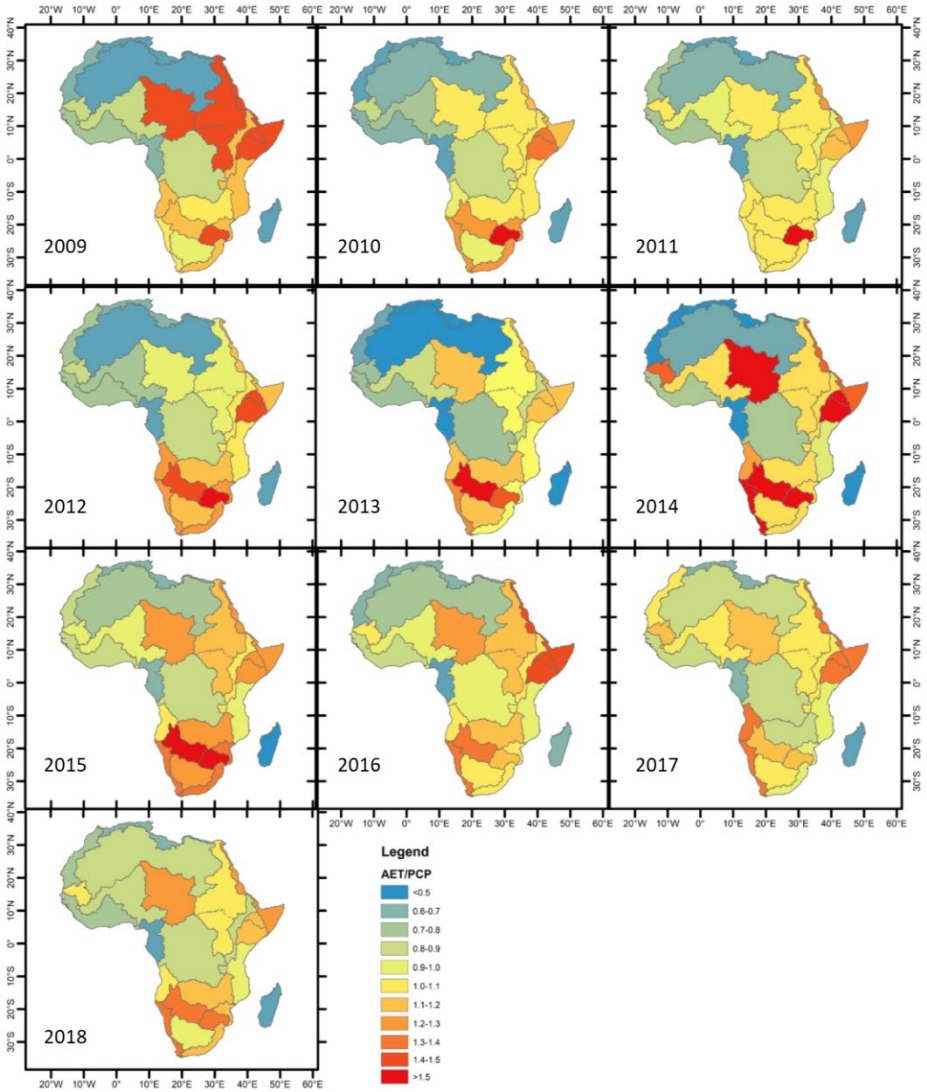

Figure 4.3: Annual ETIa-WPR/PCP in L1 for the 22 major hydro-basins in Africa for the period 2009-2018.

The ETIa-WPR and ETa-MOD16 are plotted against the ETa-WB in Figure 4.4. The relationship between both the ETIa-WPR and ETa-MOD16 products show strong linear relationships with ETa-WB. While the ETa-WPR product has a better R^2, the ETa-MOD16 has a lower bias. The ETIa-WPR shows a slightly positive bias, which is increasing with

increasing ETa-WB. The absolute difference between the ETIa-WPR and the ETa-WB is typically increasing with increasing ETa-WB. The relative differences between ETIa-WPR and ETa-WB are lower at high ETa values. The absolute difference and relative difference between ETIa-WPR and ETa-MOD16 were greater at lower ETa-MOD16. The absolute relative difference, between ETIa-WPR and ETa-WB typically decreased with increasing PCP. The long-term ETIa-WPR is larger than the ETa-WB on 13 out of 22 basins. The Q represented from 4.4% (South Interior) up to 47.0% (Central West Coast), with a median of 18.6%, of the long-term PCP. The Q is greater in basins with greater ETIa-WPR and PCP. In basins where the long-term average Q is less than 150 mm year^{-1} (18 basins), the relative difference between ETa estimates ranged from -20% to +70%. When the long-term average Q is greater than 200 mm year^{-1}the relative difference ranged from -12% to +20%.

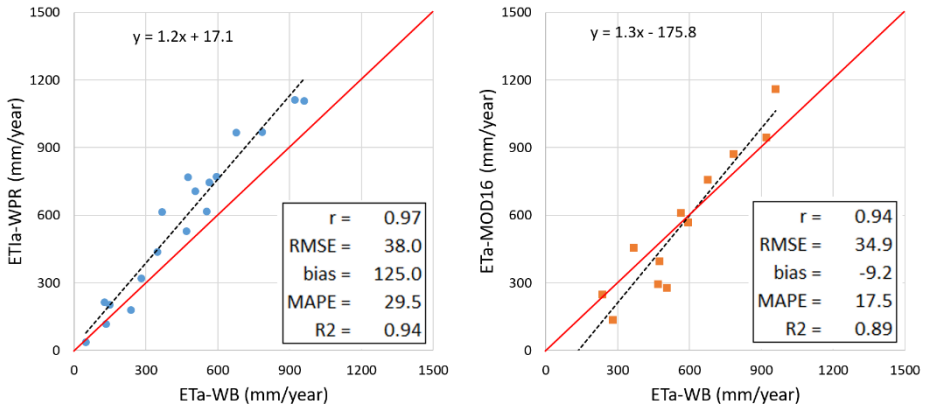

Figure 4.4: The relationship between long-term average annual ETIa-WPR (mm year^{-1}) (left) and the ETa-MOD16 (right) plotted against average annual ETa-WB (mm year^{-1}) for the 22 major hydrological basins of Africa. The black dotted line is the linear regression, and the red line is the 1:1 line.

Table 4.3: The annual PCP and ETIa (min and max) of major basins derived from the WaPOR database for the period 2009-2018 compared against the available values in literature and the ETa-WB (all values are mm year^{-1}).

Basin	PCP_{WaPOR} av. (min|max)	PCP Literature	ETIa-WPR av. (min|max)*	ETa-MOD16*	ETa Literature	ETa-WB
Lake Chad Basin	374 (322 |442)	236-451[1-3]	437 (399 |471)	-	216-363[1,3]	346
Nile Basin	649 (538 |706)	512-693[1,2,4]	714 (685 |737)	-	416-515[1,4]	-
Senegal River Basin	548 (472 |630)	252-550[1,2]	529 (475 |589)	-	258-323[1]	468
Rift Valley	762 (682 |887)	650[2]	771 (727 |803)	568	-	591
Niger River Basin	679 (625 |754)	423-740[1-3]	618 (583 |665)	-	329-410[1,3]	553
Shebelli & Juba Basin	474 (400 |602)	435-518[2,5]	615 (559 |698)	455	504	367
Central West Coast	1847 (1598 |1908)	1785[2]	1108 (1046 |1177)	1159	-	959
Congo River Basin	1517 (1452 |1600)	1165-1689[1,2]	1318 (1253 |1401)	949	1004-1098[1,6]	-
East Central Coast	966 (876 |1135)	960[2]	970 (928 |1038)	872	-	784
South West Coast	861 (697 |984)	940[2]	968 (886 |1078)	758	-	676
Zambezi Basin	928 (772 |1094)	732-1016[1,2,7]	1006 (942 |1069)	627	637-798[1,7]	-
Limpopo Basin	519 (326 |683)	530-648[1,8]	770 (662 |845)	396	516-569[1]	474
Orange Basin	303 (213 |368)	325-393[1,2]	320 (272 |388)	-	306-335[1]	280

[1]Voisin et al. (2008); [2]FAO, (1997); [3]Li, Coe, & Ramankutty, (2005); [4]The Nile Basin Initiative Secretariat, (2014); [5]Sebhat & Wenninger, (2014); [6]Chishugi & Alemaw, (2009); [7]Matondo & Mortensen, (1998); [8]LBPTC, (2010).
*av(min|max) are the yearly average, minimum and maximum for that basin.

The long-term (2009-2018) ETIa-WPR for basins in Africa is estimated to be 590.6 mm year^{-1}, which is 12.2% larger than the long-term ETa-WB, estimated to be 518.7 mm year^{-1}. The 2010 ETa average for the entire WaPOR extent is compared against ETIa-WPR V1 and other models in Figure 4.5. These values are sourced from the WaPOR V1 validation report (FAO and IHE Delft 2019) and include three remote sensing-based surface energy balance models - Atmosphere-Land Exchange Inverse (ALEXI), Surface Energy Balance System (SEBS), and SSEBop v4, a remote sensing-based Penman-Monteith approach - MOD16, a remote sensing-based artificial neural network product - Water, Energy, and Carbon with Artificial Neural Networks (WECANN), a hybrid remote sensing-based model – ETMonitor, a land surface models with remote sensing data assimilation - Global Land Data Assimilation System (GLDAS), a Priestley-Taylor approach driven by meteoroidal data - GLEAM v3.2, and, an up-scaled FLUXNET product - Multi-Tree Ensemble (MTE). The ALEXI and SSEBop v4, both remote sensing-based surface energy balance models, have a similar performance, 519 and 497 mm year^{-1} respectively. All other approaches, including SEBS, MTE, ETMonitor, WECANN, MOD16, GLEAM v3.2 and GLDAS, report a lower average annual ETa in 2010, ranging from 11% lower (GLDAS) to 38% lower (GLEAM). As compared to the CHIRPS PCP product, ETa as estimated from these products are consuming 54% (GLEAM) to 78% (GLDAS) of the PCP. Compared to the models with higher ETa that are consuming 83% (SSEBop) to 87% (ALEXI).

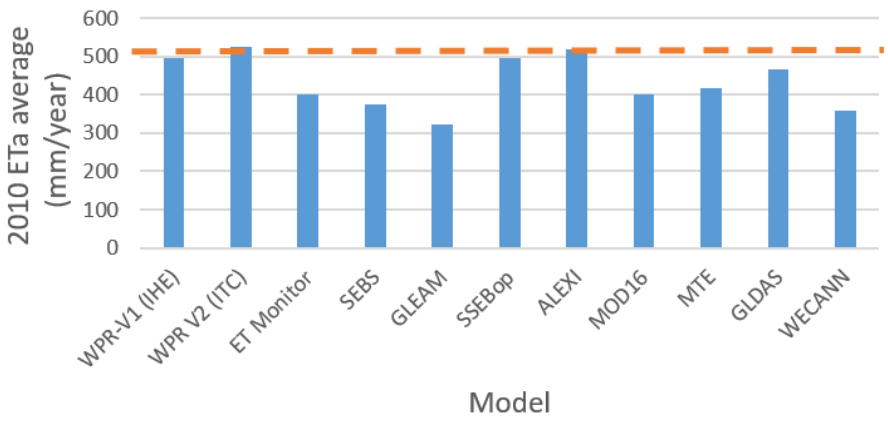

Figure 4.5: 2010 continental ETa of various models (values taken from FAO 2019) and ETIa-WPR. The orange dotted line represents the ETIa-WPR and was used for reference to other datasets.

3.1.2 Spatial and temporal consistency

The mean ETIa-WPR, SMC and NDVI were plotted for all climate zones for the northern and southern hemisphere. Figure 4.6 shows some examples of the largest sub-zones per main climate; wet tropical-savanna (Aw), arid-desert-hot (Bwh) and temperate-dry winter-warm summer (Cwb). The average ETIa-WPR (y-axis on the left), and SMC and NDVI (y-axis on the right) are reported from dekad 0901 (2009 - dekad 1) to 1836 (2018 – dekad 36).

The temporal trend for each climate zone is inversed between hemispheres, reflecting the opposite seasons between hemispheres. For example, peak ETIa-WPR values occur around dekad 19 and trough values occur around dekad 01 in the northern hemisphere. Conversely, in the southern hemisphere, peak ETIa-WPR values occur around dekad 01 and trough values occur around dekad 19. The inverse pattern highlights the need to separate climate zones based on hemisphere, as these trends would otherwise cancel out and flatten out temporal trends.

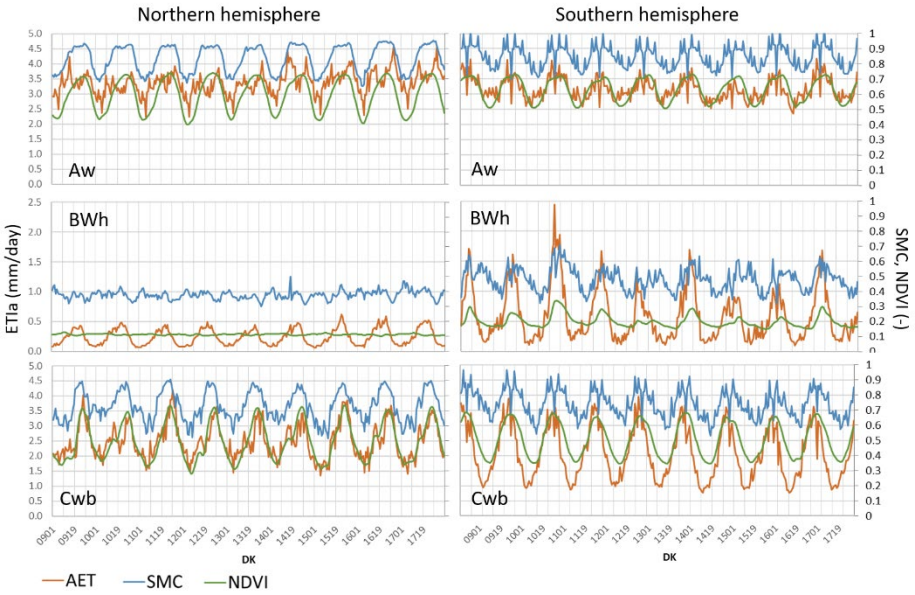

Figure 4.6: Times series of average ETIa-WPR (orange line), SMC (blue line) and NDVI (green line) in tropical wet savanna (Aw), hot arid desert (BWh) and sub-tropical highland climate classes (Cwb) in the northern hemisphere (left) and southern hemisphere (right). Note that BWh has a different ETIa-WPR y-axis range to Aw and Cwb.

The Aw zones are maintaining the highest ETIa-WPR values and shows the lowest relative variability throughout the year. The BWh zones consistently have lower ETIa-WPR values. The BWh in the

southern hemisphere is higher than in the northern hemisphere, and the relative intra-annual variation is greater. The ETIa-WPR in these zones follows a clear seasonal pattern, that is not evident from the NDVI or the SMC. The ETIa-WPR is predominantly governed by evaporation in these arid zones, which is indicated by the low NDVI all-year-round. The temperate zone, Cwb, shows the greatest intra-annual variability in ETIa-WPR, which reflects the more dramatic climatic seasonal variation in these years. ETIa-WPR in Cwb in the northern hemisphere shows two peaks per year. The two seasons are consistent with the zones' location in the Rift Valley of Eastern Africa. The Rift Valley experiences two wet seasons as influenced by the intertropical convergence zone (Hills, 1978). The wet months are March through May and October through December with higher PCP in the March through May period.

ETa is either controlled by available energy or available water. All zones, other than BWh and Aw in the northern hemisphere, show a clear relationship between the ETIa-WPR and the NDVI and SMC. The Aw zone in the northern hemisphere shows two ETIa-WPR peaks a year, however, the SMC and NDVI show one. Despite water being available through SMC in this zone during peak NDVI, there is a drop in ETIa-WPR, suggesting that during this period the ETIa-WPR is limited by solar radiation or available energy. Although not shown here – ETIa-WPR in BWh in the northern hemisphere follows the same seasonal trend as radiation. In the Aw zone in the northern hemisphere, the net radiation peaks several dekads before the NDVI and SMC, resulting in a double-peaked ETIa-WPR. The ETIa-WPR in BWh zone shows a clear seasonal trend, despite no clear seasonal NDVI or SMC trend. Therefore, it is governed by the amount of solar radiation which has a clear yearly trend at the latitudes within the BWh zone.

The SMC appears high in the arid zones, particularly considering such low NDVI in these regions. For example, in BWh in the northern and southern hemisphere, the mean SMC for the climate zone, across all dekads in the study period, never drops below 0.3 and 0.32 respectively. These regions have high potential energy and are typically water constrained. As the SMC is high in these areas with high energy availability, it is resulting in a higher than expected ETIa-WPR in these zones. The SMC, NDVI or ETIa-WPR do not seem to be responding to the drought in the region, where decreasing PCP values should result in reduced SMC and ETIa-WPR during the 2014-2015 period. The low NDVI values indicate that it is the evaporation component (driven by SMC, solar radiation and soil resistance) that is being overestimated in these dry regions.

3.2 Direct validation

The agreement between ETIa-WPR and ETa-EC is shown in Figure 4.7 and Table 4.4. Figure 4.7 shows the time series of ETIa-WPR and ETa-EC for all available in-situ data from all EC stations. Table 4.4 shows the corresponding metrics for each station, including correlation (r), root mean square error (RMSE), bias, mean average percent error (MAPE) the coefficient of determination (R^2) and the average NDVI and LST quality for the comparison period. A good overall correlation (r=0.71) is found between all sites and observations. Substantial variations existed between sites. Consistency in results is seen between years for most sites. The ETIa-WPR typically captured seasonality at most sites.

The best-performing sites in terms of correlation and R^2 are SN-DHR, SD-DEM, EG-ZAN, EG-SAA and EG-SAB. These sites are characterised by arid or semi-arid climates and short vegetation. SN-DHR and EG-SAB also have the best performance in terms of MAPE. The ETIa-WPR closely follows the ETa-EC at the SN-DHR and SD-DEM site, and both respond quickly to rainfall events. At each of these sites the WaPOR SMC and NDVI are well related to both the ETa-EC and ETIa-WPR. For example, the R^2 for the SMC or NDVI and ETa-EC or ETIa-WPR ranges between 0.82-0.87 at SN-DHR and 0.69-0.86 at SD-DEM. SD-DEM does overestimate ETIa-WPR when ETa-EC is low and NDVI is low. These sites are also associated with having high-quality LST and NDVI layers (the average LST quality for the comparison period is equal to or less than 1).

The next best performing sites, in terms of correlation and R^2, are ES-SCL, ZM-MON and CG-TCH. Excluding CG-TCH, these sites also have good quality NDVI and LST quality layers. The reasonable performance at the CG-TCH station is because the variation in ETa-EC and ETIa-WPR is well related to the VPD derived from the EC station and ETo, with R^2=0.62 and 0.66 respectively. The VPD and ETo are derived from GEOS-5 (VPD and ETo) and MSG (ETo only), as compared to being derived from satellite images. GEOS-5 and MSG are available daily and satellite image gaps do not influence the quality of the VPD and ETo quality. The irrigated agriculture sites, EG-ZAN and EG-SAB, despite high correlation and R^2, are systematically larger than the ETa-EC during both high and low ETa-EC, as indicated by the average daily bias (Table 4.4). This was less evident at EG-SAA. The seasonal values ETIa-WPR and ETa-EC for the summer maize 2012 crop at EG-ZAN are 682 mm and 424 mm, respectively. Compared to ETa from a lysimeter (ETa-lys), 543 mm, as cited in literature (Atta et al. 2015), at EG-ZAN for the same crop and period. It, therefore, suggests that the ETa at the irrigated sites fall somewhere between the ETa-EC and L1 ETIa-WPR. The overestimation is likely directly related to the net radiation difference between the EC and WaPOR datasets as inferred from the ETo estimated from the EC data and compared to the WaPOR ETo. The

WaPOR ETo has a high linear agreement with the EC ETo (R^2=0.93). However, the bias of WaPOR ETo is consistently 50% greater than the EC ETo.

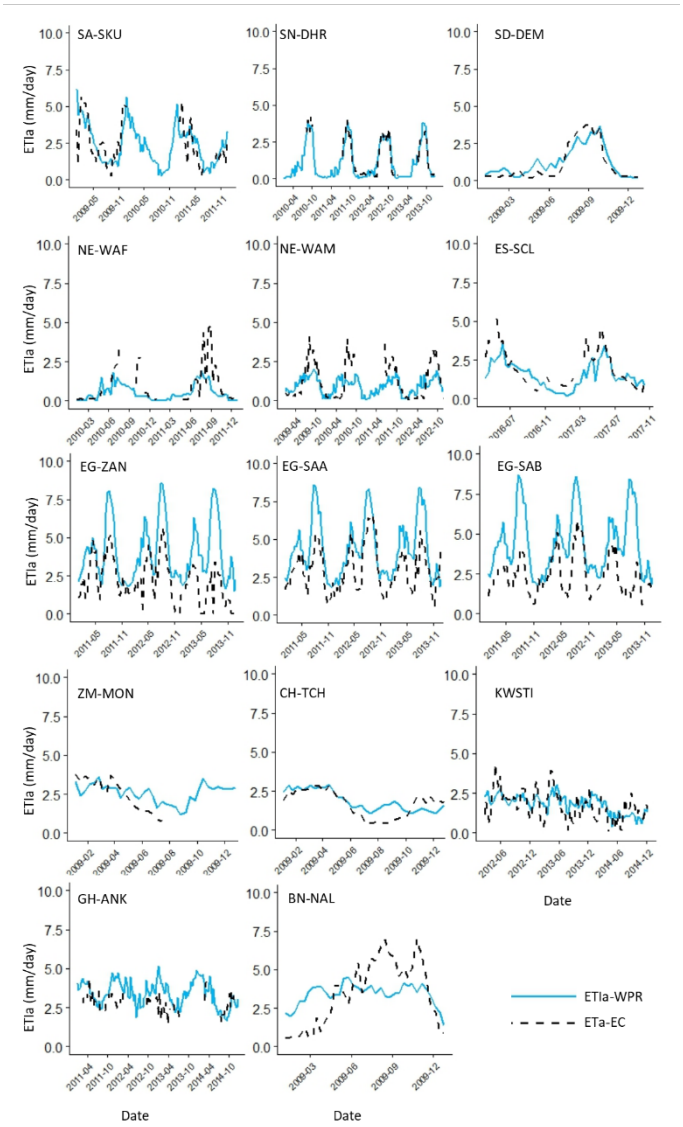

Figure 4.7: Time series of dekad ETIa-WPR (mm day^{-1}) (solid blue line) and dekad ETa-EC (mm day^{-1}) (dashed black line) for the available periods which varies for different sites. Note that the dates are reported in YYYY-MM format.

Table 4.4: Statistics comparing dekadal ETIa-WPR with ETa-EC in 14 locations; more information about sites is available in Table 4.1.

	Dekad Count	RMSE (mm day^{-1})	Bias (mm day^{-1})	MAPE (%)	R^2	r ETa-EC	r VPD-EC	r ETo-EC	NDVI QUAL*	LST QUAL*
SA-SKU	63	1.1	0.1	36.3	0.47	0.46	-	-	5.5	0.9
SN-DHR	72	0.4	0.0	17.2	0.92	0.96	-0.43	-0.59	2.0	0.9
SD-DEM	33	0.6	0.3	48.4	0.80	0.90	-0.47	-0.70	1.7	0.5
EG-ZAN	95	2.2	1.7	68.9	0.69	0.68	0.43	0.37	1.3	0.2
EG-SAA	108	0.9	0.8	16.5	0.72	0.75	0.39	0.47	1.4	0.3
EG-SAB	104	1.3	1.6	59.9	0.54	0.58	0.46	0.41	1.3	0.3
NE-WAF	49	1.12	-0.5	67.2	0.31	0.56	-0.45	-	7.4	1.3
NE-WAM	118	0.9	-0.2	58.6	0.40	0.63	-	-	6.3	1.3
ES-SCL	45	0.9	-0.3	34.0	0.52	0.72	-0.47	-	NA	NA
GH-ANK	80	1.0	0.6	28.3	0.12	0.34	0.35	-0.36	99.5	18.0
BN-NAL	36	1.8	0.0	44.9	0.27	0.52	-0.22	-0.82	11.3	2.1
CG-TCH	36	0.6	0.2	27.3	0.55	0.74	0.79	0.95	227.0	23.8
ZM-MON	20	0.8	0.2	27.3	0.48	0.69	-0.59	-0.64	7.0	1.0
KWSTI	98	0.8	0.1	37.7	0.26	0.53	-0.15	-	1.5	0.8
Overall	**957**	**1.2**	**0.5**	**40.4**	**0.54**	**0.71**	**-**	**-**	**-**	**-**

*The NDVI quality layer provides the gap, in days, to the nearest valid observation for that variable. The LST quality layer provides the number of the days between the date of the data file and the earlier remote sensing observation on which the data is based.

ETIa-WPR and ETa-EC show a weak correlation at NE-WAF and NE-WAM. The ETIa-WPR begins increasing earlier in the season, particularly at NE-WAM, and although the ETIa-WPR is capturing the seasonal trend, it is not capturing the magnitude of the ETa-EC summer values. The difference is likely related to the low-quality NDVI and LST layers during the summer (average annual values LST and NDVI gaps appear low in Table 4.4, however major gaps are concentrated in the summer season). These sites are not highly correlated with the site VPD or ETo and therefore the lower quality LST and NDVI is expected to have a great impact on the quality of ETIa-WPR here. The ETIa-WPR is strongly related to the SMC at these sites (e.g. R^2=0.73 at NE-WAM); however, the ETa-EC shows no relationship with the WaPOR SMC (R^2=0.37 at NE-WAM). Both of these sites are dominated by evaporation (in WaPOR) for most of the year – as indicated by low NDVI all year.

The ETIa-WPR performance at BN-NAL is not capturing the site seasonality. BN-NAL ETIa-WPR and ETa-EC show annual values ranging from 1.4-4.5 mm mm day^{-1} and 0.6-6.9 mm mm day^{-1} respectively. The ETIa-WPR at BN-NAL does not appear to capture the rainy period in July-September where the highest gaps in the NDVI exist (low NDVI quality). At this site, the WaPOR SMC and NDVI layers have a stronger relationship with the ETa-EC than the ETIa-WPR. For example, the R^2 between the WaPOR NDVI and the ETa-EC and the WaPOR NDVI and the ETIa-WPR are 0.87 and 0.56 respectively. This is, therefore, pointing to an overestimation of the evaporation component when NDVI is low and an underestimation of the transpiration component when the transpiration is high.

The ETIa-WPR has the lowest performance at the GH-ANK and KWSTI in terms of both the regression and the temporal trends. The GH-ANK site is characterised by a tropical climate and high vegetation height (evergreen forest). Further, the ETa-EC is not strongly related to the VPD or the ETo at both GH-ANK and KWSTI. The VPD at this site ranges from 0.07-0.81 with high relative humidity. The KWSTI site is located in the Rift Valley, between the Aberdares Ranges to the east and the Mau escarpment to the west. This setting creates a complex micro-climate with significant diurnal variation in temperature and wind speed, among other meteorological variables. This site has an inferior NDVI quality layer and a very low correlation with VPD. As a result, errors in the input meteorological data may highly influence ETa-EC estimates at the site.

The results show noticeable improvement for all metrics on average across all sites on a monthly scale (Figure 4.8 and Table 4.5). The overall correlation improved by 0.1, the overall RMSE reduced by 0.6 mm day^{-1}, the bias reduced by 0.2 mm day^{-1}, the MAPE reduced by 14.1% and the R^2 increased by 0.06. The correlation improved at

each site and the bias decreased at each site except NE-WAM and ZM-MON. The increase at NE-WAM may be a result of missing EC data during the dry season, creating a bias in favour of the wet season. The RMSE increased at the Egypt sites and at BN-NAL, likely due to the strong bias they displayed at a dekadal scale in both wet and dry seasons in favour of ETIa-WPR which aggregated at a monthly scale. Other sites showed underestimations in the wet season and overestimations in the dry season, flattening out bias over time. The MAPE increased significantly at EG-SAA. EG-SAA followed the temporal profile more closely and showed the least bias compared to EG-ZAN and EG-SAB. Overall, at a monthly scale ETIa-WPR is still overestimating ETa-EC at low ETa and at high ETa the linear regression conforms with the 1:1 line. However, if the Egypt sites are excluded from the analysis, the ETIa-WPR is overestimating ETa-EC when ETa-EC is less than 1.6 mm day^{-1} and underestimating ETa-EC when ETa-EC is greater than 1.6 mm day^{-1}.

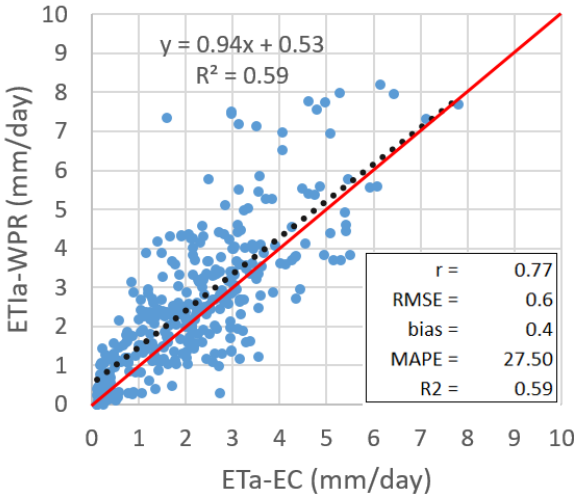

Figure 4.8: The relationship between monthly mean daily ETIa-WPR (mm day^{-1}) plotted against monthly mean daily ETa-EC (mm day^{-1}). Only months with valid observations for all dekads within that month are included. The dotted black line represents the linear regression, and the red line represents the 1:1 line.

Table 4.5: Statistics comparing monthly ETIa-WPR with ETa-EC in 14 locations; more information about sites is available in Table 4.1.

	Month Count	RMSE (mm day^{-1})	Bias (mm day^{-1})	MAPE (%)	r	R^2
SA-SKU	22	0.9	0.1	30.4	0.76	0.57
SN-DHR	28	0.3	0.0	13.9	0.98	0.95
SD-DEM	11	0.5	0.3	44.8	0.92	0.85
EG-ZAN	33	2.1	1.7	69.5	0.75	0.57
EG-SAA	36	1.4	0.8	31.7	0.77	0.6
EG-SAB	32	1.7	1.3	45.8	0.76	0.58
NE-WAF	19	1.0	-0.5	59.5	0.73	0.53
NE-WAM	41	0.9	-0.3	54.9	0.70	0.49
ES-SCL	19	0.8	-0.3	36.4	0.77	0.6
GH-ANK	34	0.9	0.6	26.9	0.45	0.2
BN-NAL	12	1.7	0.0	42.9	0.58	0.34
CG-TCH	12	0.6	0.2	25.9	0.76	0.58
ZM-MON	6	0.7	0.1	19.8	0.67	0.45
KWSTI	35	0.7	0.1	31.7	0.54	0.3
Overall	**340**	**0.6**	**0.4**	**26.3**	**0.77**	**0.59**

*The NDVI quality layer provides the gap, in days, to the nearest valid observation for that variable. The LST quality layer provides the number of the days between the date of the data file and the earlier remote sensing observation on which the data is based.

3.3 Level consistency

The consistency between the ETIa data products for the L1 and L2 data products is high. The ETIa-WPR RMSE, between L1 and L2, for each dekad for the 2009-2018 period ranged from 0.0 to 0.1 mm day^{-1}, while the correlation ranged from 0.95 to 1.00 with a median of 0.98. The median R^2 over the period is 0.96 while the median bias is 7%. The consistency between layers dropped slightly after 2014, coinciding with the introduction of PROBA-V in March 2014. The median correlation dropped from an approximately perfect positive linear correlation (~1.00) to 0.96, and the median RMSE increase was negligible (<0.1 mm day^{-1}). A slight positive systematic bias, in favour of L2, is evident after 2014, with median bias increased from 4% to 9%.

The L1 and L3 ETIa-WPR products have a lower consistency as compared to the L1 and L2 products in the four irrigation areas. The mean ETIa-WPR values for all dekads in the Zankalon and Awash schemes are shown in Figure 4.9. The Awash area has the highest consistency of all scheme areas, reflected in the highest average correlation and R^2 across dekads, 0.84 and 0.71 respectively. The ETIa-WPR RMSE between L1 and L3 in the Awash ranges from 0.42-1.0 mm day^{-1}, while the correlation ranges from 0.63-0.92. The median correlation for all dekads in the study period is 0.84, and the median R^2 is 0.84. The RMSE is highest when the ETIa-WPR is highest. The RMSE temporal trend is in line with the seasonal trend in the Awash and displays the two seasons associated with the intertropical convergence zone. The correlation is above 0.73 on 95% of dekads, and lowest on dekads when the mean ETIa-WPR is highest.

The Koga has the lowest consistency of the schemes. Although the RMSE between L1 and L3 is lower, ranging from 0.3-0.7 mm day^{-1}, the median correlation is 0.67, and the median R^2 is 0.45. Zankalon performed slightly better, with a median correlation of 0.71 and a median R^2 of 0.51. The RMSE is higher in Zankalon than the Koga, but this reflects the higher ETIa-WPR values found in the area. The ODN had the same RMSE (0.6 mm day^{-1}) as Zankalon and the highest range of RMSE (0.2-1.6 mm day^{-1}). The correlation and R^2 are also similar, with median values of 0.73 and 0.53 respectively. All schemes show similar per cent bias medians (9-12%). The only scheme that shows a systematic bias is ZAN, where the L1 is consistently higher ETIa-WPR values than L3.

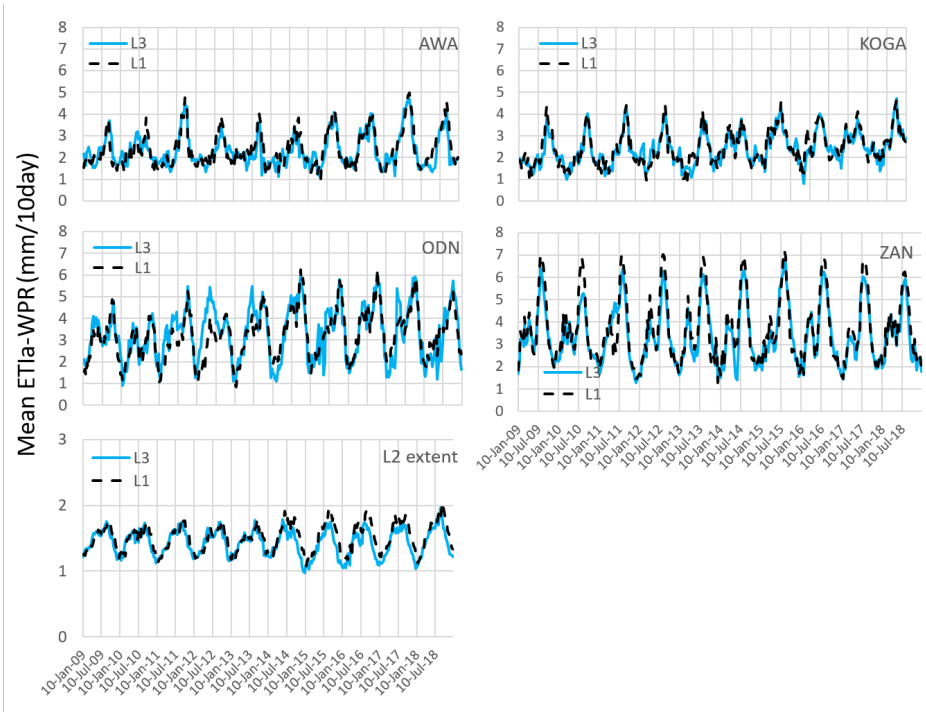

Figure 4.9: Mean ETIa-WPR per dekad (mm day^{-1}) in continental level - L1 (blue line) and irrigation scheme level - L3 (black line) for 2009-2018 period in each of the L3 extents (Chapter 2 - Table 2.1) and the continental level - L1 (blue line) and basin and country level – L3 (black line) in the L2 extent. Note that the date is reported in YYYY-MM.

The 10-daily average ETa-EC and ETIa-WPR for all three spatial resolutions at EG-ZAN are shown in Figure 4.10. The L1 and L2 ETIa-WPR show high consistency with each other. The L3 ETIa-WPR is consistently sitting between the ETa-EC and the L1 and L2 ETIa-WPR. All levels capture the overall ETa-EC seasonal trends. The L3 data shows a slightly lower R^2 (L3=0.66 and L1=0.69) and correlation (L3=0.53 and L1=0.68), but a much lower bias (L3=1.1 mm day^{-1} and L1=1.7 mm day^{-1}) and a lower RMSE (L3=1.0 mm day^{-1}and L1=2.2 mm day^{-1}) when compared with ETa-EC. The better R^2 and correlation reflect the L1 and L2 ETIa-WPR ability to capture the temporal fluctuations of ETa-EC better than L3 ETIa-WPR. An example of this is at dekad 1117, where L1 and L2 ETIa-WPR capture the ETa-EC dip, whereas L3 ETIa-WPR stays flat. The L3 ETIa-WPR have a better seasonal agreement with the ETa-lys for the summer maize crop in 2012 (L3=487 mm, L1=682 mm and ETa-lys=543 mm).

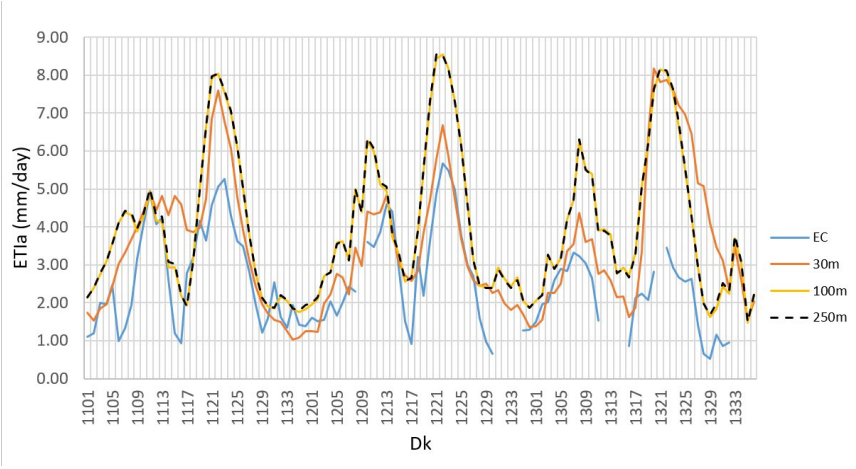

Figure 4.10: Level consistency validation of ETIa-WPR for three levels of L1, L2 and L3 ETIa-WPR in comparison with ETa-EC per dekad (Dk) for the 2011-2013 period at EG-ZAN EC station.

The NDVI and ETIa-WPR for the 250 m buffer are shown in Figure 4.11 for the three spatial resolutions. The 30 m level is picking up more spatial variation (standard deviations: L3=0.05, L2=0.02; L1=0.02) at the site and has a lower mean NDVI for the site as compared to L2 and L1 (mean: L3=0.74; L2=0.82 and L1=0.83). This reflects the lower ETIa value for this dekad, which is more similar to the ETIa-EC (Figure 4.10) and shows some limitations in comparing L1 data to EC in a heterogeneous landscape.

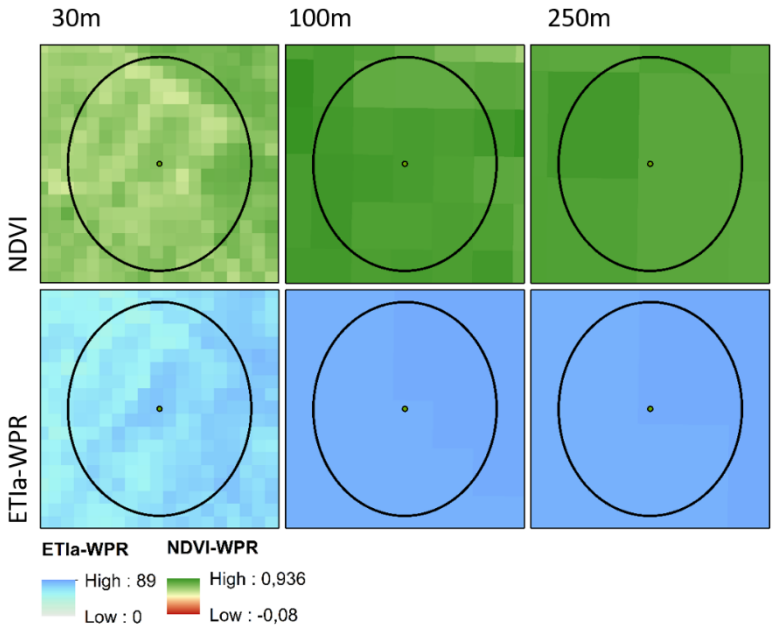

Figure 4.11: NDVI and ETIa-WPR for the EG-ZAN site for all three spatial resolutions (L3=30m, L2=100m and L1 = 250m) on dekad 1222 (1st dekad of Aug 2012). The point is the station location; the circle is the buffer used for data extraction to compare to the ETa-EC.

4. Discussion

4.1 Product accuracy

The ETIa-WPR results are comparable the improved MODIS global terrestrial ETa algorithm, MAPE of 24.6% as compared to EC measurement, when driven by the tower meteorological data (Mu et al. 2011). The ETIa-WPR error estimates, on average, are also close the average errors in EC measurements as EC measurements typically have errors of 20-30% (Allen et al. 2011; Blatchford et al. 2019), however, it appears that the ETIa-WPR is regularly overestimating ETIa, which is evident at local to basin level. Figure 4.12 shows the bias and number of observations between ETIa-WPR and ETa-EC for all EC observations disaggregated based on 0.5 mm day^{-1} ETa-EC increments. The results are further defined based on non-irrigated sites, irrigated agriculture, and all stations. For non-irrigated sites, there is a positive bias (ETIa-WPR>ETa-EC) when the ETa-EC is less than 2.5 mm day^{-1}and becomes negative when the ETa-EC is greater than or equal to 2.5 mm day^{-1} (this reduced to 1.6 mm day^{-1}at a monthly scale). This bias increases, both positive and negative, as the

ETa-EC deviates from 2.5 mm day^{-1}. The underestimation is further exacerbated by the fact that ETa-EC estimations can lead to underestimation of the latent energy or ETa-EC by 20% (Wilson et al. 2002; Glenn et al. 2007). Underestimation bias is larger than overestimation bias and increases with increasing ETIa-WPR. However, Africa as a continent is dry with long-term (2010-2015) average daily ETIa-WPR for the continent being 1.5 mm day^{-1}. Therefore, the ETIa-WPR frequently overestimates at the annual, basin scale. The irrigated sites (EG-SAA, EG-SAB and EG-ZAN) are overestimated for nearly all ETa-EC. The irrigated sites strongly influenced the overall bias, as these sites have many observation points. When irrigated and non-irrigated results are combined, the changing point where ETIa-WPR is greater than ETa-EC occurs when ETa-EC exceeds 3.5 mm day^{-1}.

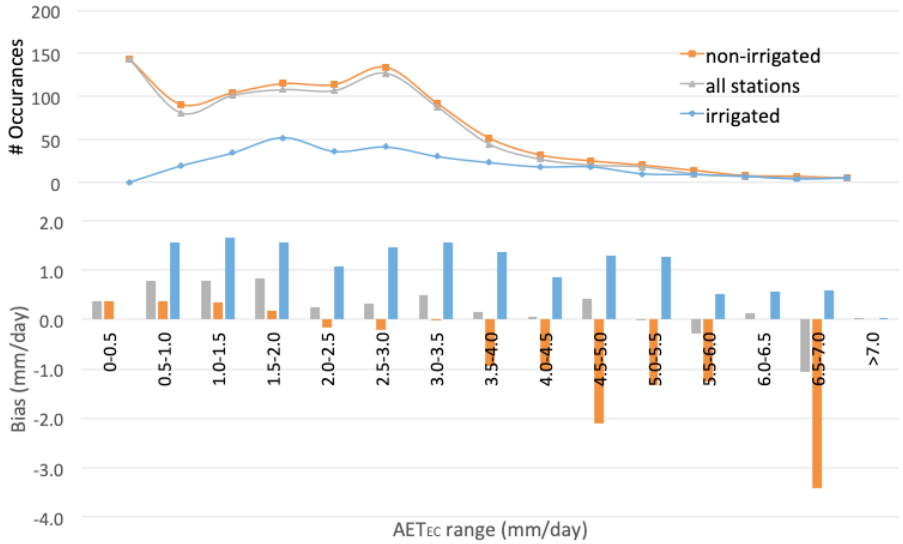

Figure 4.12: Upper – number of observations for a given ETa-EC range. Lower – bias of dekadal ETIa-WPR (mm day^{-1}), as compared to ETa-EC, plotted against the increasing ranges of ETa-EC (mm day^{-1}) for observations at natural vegetation sites (orange bar), irrigated agriculture sites (blue bar) and all sites (grey bar).

Why is WaPOR overestimating when ETIa is low?

ETIa-WPR is overestimating ETa in dry, hot, water-stressed conditions (e.g., water-limited). The ETIa-WPR estimates for prolonged dry weather and the dry seasons of WaPOR are usually higher than the observed values (flux towers, field). These overestimations are small in terms of absolute values (mm day^{-1}) but can lead to overestimation of results in higher annual ETIa-WPR when compared to water mass balance checks of river basins. The overestimation in dry regions is

likely to be primarily due to the functioning of the SMC constraint or the too high SMC in dry regions.

The WaPOR SMC is considered, on average, high in arid regions (e.g., Figure 4.6) and therefore, ETIa-WPR is likely not effectively accounting for soil moisture limitations. The high SMC is resulting in an overestimation of the evaporation component in particular, as NDVI is low and therefore the region is dominated by the evaporation component of ETIa-WPR. Arid regions should be largely regulated by water availability rather than energy. Conversely, under well-water conditions, the Penman-Monteith method is primarily driven by Rn (e.g. energy limited) (Rana & Katerji, 1998). As Penman-Monteith is a linearised approximate solution, problems may occur in extreme conditions and errors in the soil evaporative term (Leca et al. 2011). Majozi et al. (2017b) noted that Penman-Monteith methods need to include a SMC constraint. Though the ETIa-WPR methodology does include a SMC constraint, overestimations in SMC are reducing its functionality. The SMC is estimated using the trapezoidal method (function of NDVI and LST) (FAO, 2018). Where the NDVI is low, the LST component could be the primary contributing factor to SMC errors.

For water-stressed crops, crop resistance errors can attribute to the large error in ETa estimations, while for tall crops, the VPD can have a large influence on the error (Rana & Katerji, 1998). Extreme conditions include when aerodynamic resistance is high, >50 m s^{-1} (Paw, 1992). High aerodynamic resistance can occur in sparse vegetation, when surface temperature is much greater than air temperature (e.g. water-stressed conditions) and when wind speed is very low (Paw, 1992; Dhungel et al. 2014). Cleverly et al. (2013) and Steduto et al. (2003) found when the standard aerodynamic resistance values were used the Penman-Monteith method over- and underestimated ETo when ETo is low and high respectively and suggested the aerodynamic resistance should vary with climatic variables as it is responsive to relative humidity gradients.

It is recommended to further verify the behaviour of the SMC. The SMC relative moisture index is derived from LST and vegetation cover (NDVI) data. Therefore, verification against highest available physically based satellite soil moisture data (e.g., active microwave sensors on-board Sentinel-1A, Metop, etc.) is advised. It may be helpful to use SMC for transpiration and passive microwave sensors for evaporation.

The main source of error in the ET-WB method is the uncertainty in PCP. Studies on the CHIRPS PCP product shows high correlations, at monthly and regional scales, in Eastern Africa (r = 0.7-0.93) (Dinku et al. 2018; Gebrechorkos et al. 2018) and Burkino Faso (r = 0.95) (Dembélé and Zwart, 2016) with little to no bias. Muthoni et al. (2018) reported that CHIRPS v2 slightly over-estimated low-intensity rainfall

below 100 mm and slightly under-estimated high-intensity rainfall above 100 mm compared in Eastern and Southern Africa. On an annual, basin-scale, the CHIRPS PCP product does not show significant bias, except for in largely ungauged tropical basins (e.g. Congo) (Liu et al. 2016).The Q component is less than 25% of the PCP in all but three basins used in the comparison, Central West Coast, West Coast and North East Coast (though in the North East Coast ETIa-WPR > PCP). In basins where Q is a significant component of the water balance, its uncertainty is going to have the largest influence on the uncertainty of the ETa-WB. The R^2 values of modelled GRDC Qmean against streamflow data were >0.9 (Beck et al. 2015). Therefore, the Qmean is expected to be high in gauged basins. Unguaged basins, in the analysis, have higher uncertainty and introduce higher uncertainty into ETa-WB. Basins with no streamflow data include North Interior, North East Coast, Shebeli & Juba Basin and Limpopo. Of these basins only the North Interior has ETIa-WPR<PCP. If basins are removed from the analysis with missing streamflow data the regression between ETIa-WPR and ETa-WB only marginally improves (R^2 = 0.96 compared to R^2 = 0.94), suggesting the quality of Qmean is appropriate for the water balance check. Therefore, the large overestimations of ETIa-WPR should not be attributed to the simplified water balance approach.

Wetland and irrigated areas are expected to have ETIa greater than PCP. Wetland and irrigated areas represent 1% and <2% respectively of land cover in Africa and is suggested to have little impact on the overall water balance for most basins. The basins with the greatest irrigated land cover and the highest fraction of ETIa-WPR from irrigation, are Limpopo Basin (6.4% of land cover), Orange Basin (4.3% of land cover) and Indian Ocean Coast (6.7% of land cover). ETIa-WPR in these basins contribute to 6.0%, 7.6% and 8.7% of the total evapotranspiration. For each of these basins, ETIa-WPR is greater than PCP by more than the fraction of ETIa-WPR from irrigation. Basins with large wetlands and high ground water availability include the Niger (Niger Delta), the Nile (the Nile Delta and Sudd wetland), the South Interior (Okavango Delta) (FAO and IHE 2019). However, large areas of shrubland and deciduous tree cover also have ETIa greater than PCP, when compared with the WaPOR land cover dataset (available on the WaPOR portal). The overestimation of ETIa compared to PCP on an annual basin appears to be more closely related to climate. Climate zones BSh, BWh, CWa and Aw have large areas with ETIa greater than or approximately equal to PCP at an annual scale. These zones are largely associated with basins with ETIa greater PCP.

Why is WaPOR overestimating ETIa in irrigated fields?

ETIa-WPR is overestimating ETa dry, hot, non-water-stressed conditions (e.g., irrigated fields). These errors might lie in the FAO-

Penman-Monteith method's and may be associated with local advection effects. Local advection may increase ETa over a water-limited field by up to 30% (De Bruin et al. 2016; Trigo et al. 2018). There is an underlying assumption of no advection in the ETo definition for a reference grass field (Allen et al. 1998). However, in small fields, under arid conditions with high temperatures, local advection effects may occur when warm, dry air formed over an upwind, adjacent field is advected horizontally over the well-watered fields (De Bruin & Trigo, 2019). This horizontal advection of sensible heat increases the ETa of water from well-watered areas, where SMC is high and not limiting, but will result in the overestimation of ETa in water-limited fields or areas. While the Egypt fields are well irrigated (Sugita et al. 2017) with SMC ranging from 0.6-1 throughout the irrigation season, surrounding fields are not, and frequently have low SMC or water limiting condition, which can potentially drive up the ETIa-WPR estimates. The Zankalon irrigated area, where EG-ZAN is located, has small fields, ~0.2ha (Table 4.2), as does the EG-SAA and EG-SAB. Therefore, these sites may be particularly influenced by this effect as 0.2ha is 3% of an L1 -250 m pixel, 20% of an L2 -100 m pixel and 200% of an L3 -30 m pixel (e.g., see Figure 4.10).

Why is WaPOR misrepresenting ETIa when ETIa is high in humid conditions?

ETIa-WPR is not representing ETa well in water unlimited conditions with high humidity. The Penman-Monteith method is not suitable for very low VPD (or high humidity) (Paw & Gao, 1988). Further, for tall crops, the VPD can have a considerable influence on the error (Rana & Katerji, 1998). It is not suitable in these conditions because of the linear assumption of saturated vapour pressure and air temperature. Paw, (1992) advised that the use of non-linear equations should be used in extreme conditions to maintain errors of less than 10-15%.

Quality of input data is likely affecting the quality of the ETIa-WPR in these regions. Low-quality data or missing RH data means VPD is calculated from Tmin. In humid climates condensation occurs during the night, which leads to an overestimation of VPD (Allen et al. 1998), which is found when Penman-Monteith is applied without RH data in humid regions of Ecuador (Córdova et al. 2015). In water unlimited regions, the overestimation of VPD can lead to higher ETa, as it is easier for the flux to occur when there is less moisture in the air. Further, these regions frequently contain low-quality NDVI and LST layers in these regions. This is resulting for example, in overestimation of radiation at GH-ANK skewing results at this location. The NDVI and LST-quality layers are therefore a good indicator of the quality of the ETIa in these regions.

4.2 Product consistency

There is very high consistency between L1 and L2 products. The high consistency is partly explained by the use of a downscaled MODIS product before the introduction of PROBA-V in 2014 and the SMC component, which is based on MODIS for both L1 and L2 for the entire database period. The high consistency suggests that at a given scale, for example basin scale, the 100 m product provides no additional value to the 250m resolution. However, at higher resolution applications, the product does show spatial variation not captured by the L1 product (e.g. Figure 4.11) and may provide better insight into intra- and inter-field level variations.

The consistency between the L1 and L3 products is mixed. The Awash and ODN L3 areas show high consistency between L1 and L3. In the Koga, there is a strong positive bias for L1 ETIa-WPR, while the agreement between L1 and L3 in the Koga and in Zankalon is lower. These errors are likely largely attributed to the different input temporal and spatial resolutions available from the satellite platform combined with high spatial and temporal heterogeneity in the area (e.g. Koga and Zankalon have much smaller irrigated fields and higher crop diversity than the Awash and ODN–see Table 4.2). All levels have a dekadal time-step. However, the satellite revisit period varies, having revisits of 1-day, 2-days and 16 days for MODIS (L1), PROBA-V (L2) and Landsat (L3), respectively, with daily meteorological data input. The variation in the revisit period can lead to differences when interpolating images to a dekadal timescale, particularly in rainy periods and during the growing season (Gao et al. 2006). Uncertainty of up to 40% has attributed to the difference in a 16-day revisit as compared to 4-day revisit, depending on climate and season (Guillevic et al. 2019), though this was without daily meteorological data as a tool for interpolation. Conversely, the L3 dataset can capture more spatial variability for a given image as compared to the L1 and L2 data, which is highly important when using non-linear models. Therefore, the L3 dataset is expected to perform better in areas of higher spatial heterogeneity (Sharma et al. 2016).

5.Conclusion

The WaPOR products for Africa and the Middle East provide the highest resolution continuous near-real-time products available so far to monitor ETIa. Current validation efforts need to be continued and intensified to confirm the suitability of these products for various uses. However, significant issues with the sparseness of available ground-truth measurements make direct validation to in-situ, insufficient as a

sole means to validate the ETIa product over continental Africa. To compensate for insufficient ground-truth locations, we added physical consistency and level consistency checks as part of the validation analyses.

The ETIa-WPR product is responsive to general trends in the magnitude of ETIa for most climates and shows good correlations at both local (EC) and basin (WB) scales. In dry irrigated areas, WaPOR appears to be overestimating ETIa, particularly the coarse resolution. The overall ETIa-WPR MAPE of 26.3% on a monthly, point scale, 40.4% on a daily, point scale and 29.5% on an annual, basin scale. These are promising results considering that WaPOR presents a continental almost near real time open-access dataset. Analysis of the intermediate data components provide insights into some of the possible causes of the over- and underestimation of ETI-WPR, which appear to be primarily driven by an overestimation of the SMC which is driving overestimation of evaporation. Users should also be cautious in applying the dataset in very hot, arid conditions, in high canopy (e.g. forests) and areas with large gaps in the NDVI- and LST-quality layers. Further validation activities are suggested as new ground-data become available, particularly in cropped and irrigated areas.

Chapter 5

Influence of spatial resolution on remote sensing-based irrigation performance assessment using WaPOR data[3]

[3] This chapter is based on: Blatchford, M.L., Mannaerts, C.M., Zeng. Y., Karimi, P., 2020. Influence of spatial resolution on remote sensing-based irrigation performance assessment using WaPOR data, Remote Sensing (2020) 12, 2949.

Abstract

This paper analyses the effect of the spatial assessment scale on irrigation performance indicators in small and medium-scale agriculture. Three performance indicators—adequacy (i.e., sufficiency of water use to meet the crop water requirement), equity (i.e., fairness of irrigation distribution), and productivity (i.e., unit of physical crop production/yield per unit water consumption)—are evaluated in five irrigation schemes for three spatial resolutions—250 m, 100 m, and 30 m. Each scheme has varying plot sizes and distributions, with average plot sizes ranging from 0.2 ha to 13 ha. The datasets are derived from the United Nations Food and Agricultural Organization (FAO) water productivity through open access of remotely sensed–derived data (the Water Productivity Open Access Portal—WaPOR) database. Irrigation indicators performed differently in different aspects; for adequacy, all three resolutions show similar spatial trends for relative evapotranspiration (ET) across levels for all years. However, the estimation of relative ET is often higher at higher resolution. In terms of equity, all resolutions show similar inter-annual trends in the coefficient of variation (CV); higher resolutions usually have a higher CV of the annual evapotranspiration and interception (ETIa) while capturing more spatial variability. For productivity, higher resolutions show lower crop water productivity (CWP) due to higher aboveground biomass productivity (AGBP) estimations in lower resolutions; they always have a higher CV of CWP. We find all resolutions of 250 m, 100 m, and 30 m suitable for inter-annual and inter-scheme assessments regardless of plot size. While each resolution shows consistent temporal trends, the magnitude of the trend in both space and time is smoothed by the 100 m and 250 m resolution datasets. This frequently results in substantial differences in the irrigation performance assessment criteria for inter-plot comparisons; therefore, 250 m and 100 m are not recommended for inter-plot comparison for all plot sizes, particularly small plots (<2 ha). Our findings highlight the importance of selecting the spatial resolution appropriate to scheme characteristics when undertaking irrigation performance assessment using remote sensing.

1. Introduction

Irrigation is typically performed in areas with arid climates, low precipitation, and/or frequent droughts, which makes water management both complex and important. On a continental scale, irrigation consumes approximately 86% of freshwater withdrawals in Africa (Frenken 2005). This is already higher than the global average and is expected to increase with increasing prosperity and therefor food demands. Water governance in Africa must be performed in such a way to provide for increasing demands in food and simultaneous increasing demands from industry, the environment, and municipalities.

Quantifying water balance components and land productivity for irrigation schemes scales has a wide variety of applications. This includes but is not limited to initiating and evaluating water conservation practices, evaluating equitable water distribution (Kharrou et al. 2013; Alexandridis et al. 1999; Bastiaanssen and Bandara 2001; Nouri et al. 2020a), assessing water and land productivities (Roerink et al. 1997; Blatchford et al. 2018), input to water policy and resource management (P. Karimi et al. 2012; Hellegers et al. 2010) and improving irrigation management and systems (Kharrou et al. 2013; Bastiaanssen et al. 2001).

Remote sensing is a powerful tool to understand agricultural performance at high spatial and temporal resolutions. The application of remote sensing in estimating agricultural performance indicators is becoming more prolific as it provides more information, in both time and space, than can be provided by traditional methods, such as water balance or ground measurements (Blatchford et al. 2020). Remote sensing can provide insight into various aspects of agricultural production, including estimation of actual evapotranspiration (ETa) and biomass production. With the increase in open access to satellite imagery and retrieval algorithms, remote sensing provides effective yet spatially and temporally extensive options to estimate agricultural indices, which are especially beneficial for evaluating irrigation performance in data-scarce regions, like Africa.

The use of remote sensing to estimate and evaluate various indicators for irrigation performance at the irrigation scheme level (i.e., the irrigation perimeter) has been tested and reported for multiple spatial and temporal resolutions (e.g., spatial resolution refers to the pixel size, and temporal resolution refers to the satellite revisit time). These include studies on indicators including equity (Kharrou et al. 2013; Roernick et al. 1997; Bastiaanssen et al. 1996)), adequacy (Roerink et al. 1997; Taghvaeian et al. 2018; Elnmer et al. 2018), sustainability and water productivity (Karimi et al. 2019; De Teixeira and Bassoi 2009; Zwart and Leclert 2010). These studies used datasets with input sensors of varying resolution, including Landsat TM with a resolution of 30 m and 16-day revisit and moderate resolution imaging

spectroradiometer (MODIS) Terra with resolutions of 250 m and 1-day revisit or 500 m and 8-day revisit. The accuracy of these studies is often not reported, and when it is, varies considerably (Blatchford et al. 2019). Studies utilising global datasets show a large range in errors in ETa, net primary productivity (NPP), or biomass production and crop water productivity (CWP) (Blatchford et al. 2019). Local studies, with parametrised models typically have a higher accuracy. However, it is not clear how the resolution of the dataset influences the accuracy or the potential application of irrigation performance indicators. For example, at what resolution can different indicators be assessed for a given scale of application (e.g., inter- or intra-scheme) or a given scale of the irrigation scheme (e.g., plots size).

The variation in the satellite revisit period can lead to different irrigation performance indicator values when interpolating images, particularly in rainy periods and during the growing season (Gao et al. 2006). Uncertainty in ETa of up to 40% has been attributed to the difference in a 16-day revisit as compared to 4-day revisit, depending on climate and season with no assimilation of daily meteorological data (Guillevic et al. 2019). Likewise, spatial resolution is highly important when using non-linear models in heterogeneous areas due to pixel purity (Durgam et al. 2020; Duveiller et al. 2011). Therefore a higher spatial resolution, when analysing irrigation performance indicators, is expected to improve the assessment in areas of higher spatial heterogeneity (Sharma et al. 2016).

Several studies have investigated the relationship between spatial resolution or temporal resolution on accuracy of different input parameters (Zeng et al. 2016; Zhuang et al. 2020) used to estimate irrigation performance by remote sensing; for example normalized difference vegetation index (NDVI) (Jiang et al. 2005; Teillet, Staenz, and Williams 1997), energy balance components (Ramírez-Cuesta et al. 2019), ETa (Nouri et al. 2020b) or NPP or gross primary productivity (GPP) (Vanikiotis et al. 2018). Further, much effort has been placed on continuously improving the resolution of these products (Hatfield et al. 2016) or aggregating images with varying resolutions to maximise information (i.e. aggregating high spatial and low temporal resolution images with low spatial and high temporal resolution images) (Ershadi et al. 2013). A recent study compared the accuracy of wheat yield with varying spatial and temporal resolution and found that, for yield (or NDVI trends), a higher spatial resolution was preferred to a higher temporal resolution (Durgam et al. 2020). Another study suggested at a regional scale Landsat ETM+ 30m and MODIS 250m data can both estimate agricultural area and NDVI to a reasonable level of accuracy (Pax-Lenney and Woodcock 1997).

However, despite some investigations into these parameters and spatial resolution, there is gap between research projects and the

impact of image resolution on the quality and accuracy of irrigation performance indicators at medium – to – high resolution scale, for a given irrigation scheme or farm scale.

The recent online United Nations Food and Agricultural Organization (FAO) portal to monitor water productivity through open access of remotely sensed-derived data (the Water Productivity Open Access Portal—WaPOR) avails evapotranspiration and interception (ETIa) and yield related datasets in Africa and the Middle East at dekadal time steps at three different spatial resolutions (250 m, 100 m, and 30 m), depending on geographic location. The availability of these datasets provides an opportunity to monitor farming practices and irrigation management from space and is thus ideal for regions with limited in-situ or local datasets. In this study, we utilize WaPOR datasets to derive three irrigation performance indicators—namely, adequacy, equity, and productivity—in five irrigation schemes—Wonji, Metehara, Koga, Zankalon, and Office du Niger (ODN)—in Africa. We evaluate the impact on each of the indicators at inter-plot and scheme levels and suggest the best resolution for each based on farm plot size.

2. Materials and methods

2.1 Scheme descriptions

Five irrigation schemes in Africa were selected—the Wonji (Ethiopia), the Metehara (Ethiopia), the Koga (Ethiopia), Zankalon (Egypt), and ODN (Mali). The Wonji and Metehara are located in the Awash Basin, the Koga scheme is located in the Upper Blue Nile, the Zankalon scheme is located in the Nile Delta Basin, and the ODN scheme is located in the Niger Basin (Figure 5.1). The description of each irrigation scheme, standard deviation (SD) of plot sizes used in the evaluation, the major crops cultivated, the scheme area used in the evaluation, and the minimum and maximum elevation of each scheme are given in Table 5.1. The minimum and maximum elevation was taken from the digital elevation model (DEM) (90 m) of the Shuttle Radar Topography Mission (SRTM).

In the Wonji and the Metehara plot boundaries were provided by the irrigation scheme managers in AutoCAD and converted to shapefile. The Koga, ZAN and ODN plots were digitally delineated using google earth. Statistics were derived using R software and vizualizations were done in ArcMap Thematic Mapping.

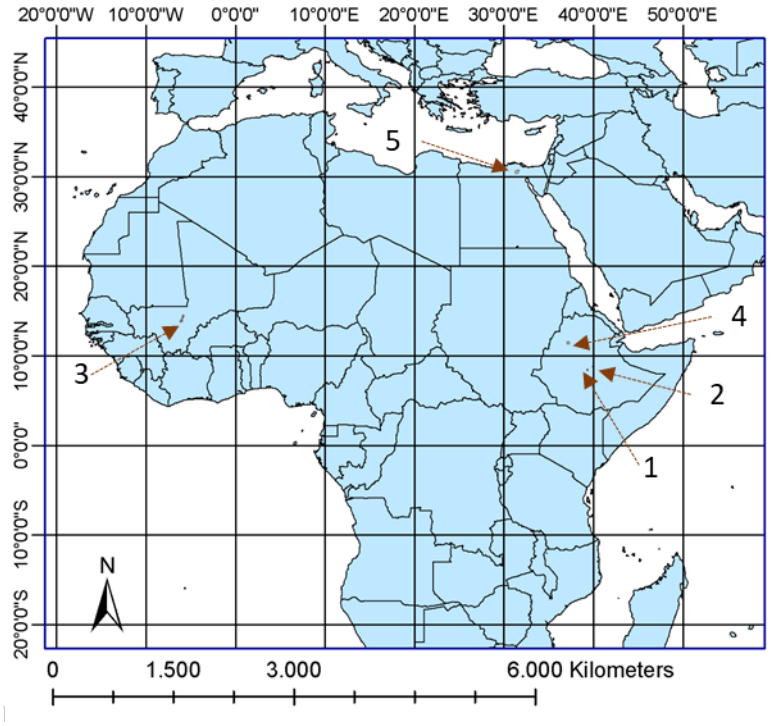

Figure 5.1: Location of irrigation schemes: 1) Wonji, 2) Metehara, 3) ODN, 4) Koga, 5) Zankalon

Table 5.1: Description of irrigated scheme areas used in the product evaluation.

	Wonji	Metehara	Office du Niger (ODN)	Koga	Zankalon
Average plot size in irrigated area (ha)	13.04	8.83	5.93	0.24	0.21
SD plot area (ha) | CV	6.41 | 0.49	5.58 | 0.63	0.46 | 0.08	0.12 | 0.50	0.13 | 0.62
Major crops in irrigated area	Sugarcane	Sugarcane	Rice, sugarcane	Wheat, maize, potato, onion, cabbage, barley	Wheat, rice, maize, cotton, sugar beet,

					berseem, favabean, tomato, potato
Area (ha)	6130	6954	1773	145	126
Min|Max elevation (msl)	1539|1546	950|982	278|286	2009|2051	6|15

2.2 Input datasets

The analysis dataset is the ETIa V2.0 and NPP V2.0 products available on the WaPOR portal and are described in Chapter 2 – section 2-3. The extent of and the resolution of the dataset is also described in Chapter 2 - section 1. Finally, the WaPOR datasets, including intermediate datasets, input sensors and input data products, used in the validation are provide in Chapter 2 – section 4.

Three input datasets to assess the irrigation performance were collected in the five study areas. The three input datasets were the actual evapotranspiration or ETIa, net primary productivity or NPP, and reference ET or ETo. The dataset was assessed from a near 10-day dekadal to annual temporal scales over a 10-year period (2009–2018).

The WaPOR database provides ETIa and NPP in three spatial resolutions (Chapter 2 - Table 2.1) for the five selected schemes (Table 5.1).

The ETo is provided at one resolution (25km|daily) and is described in Chapter 2 – Table 2.2.

2.2.1 Evapotranspiration and interception

The ETIa in the WaPOR portal is based on a modified version of the ETLook model (ETLook-WaPOR) (Pelgrum et al. 2012). The ETIa dataset is described in Chapter 2 – section 2.

2.2.2 Above ground biomass productivity

NPP is a fundamental characteristic of an ecosystem, expressing the conversion of carbon dioxide into biomass driven by photosynthesis. The NPP in the WaPOR portal, and the conversion of NPP to AGBP, is described in Chapter 2 – section 3.

2.2.3 Reference evapotranspiration

The ETo, based on FAO-56 (1996), was derived from the same gridded meteorological data, transmissivity data, and digital elevation data described for the ETIa and NPP. The ETo in the WaPOR portal is defined as:

$$ETo = \frac{\Delta(R_n - G) + p \cdot c_p \cdot \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + r_s/r_a)} \quad (4.1)$$

Where p is air density (kg m^3), c_p is the specific heat (J $°C^{-1}$), r_a is the aerodynamic resistance (s m^{-1}), r_s is the bulk surface resistance (s m^{-1}), Δ is the slope vapour pressure curve [kPa $°C^{-1}$] and γ psychrometric constant [kPa $°C^{-1}$]. rs is taken as 70 s m^{-1} and the r_a is taken as $208/u_{obs}$. u_{obs} is the observed wind speed (m s^{-1}) at 10m. The ETo is provided at a 25km, daily resolution in the WaPOR portal.

2.3 Performance indicators

Three performance indicators (Table 5.2) were studied at different spatiotemporal resolutions. These indicators are selected based on the data available directly from the WaPOR portal:

- Adequacy – the sufficiency of water use to meet the crop water requirement (CWR) or potential evapotranspiration;
- Equity –the fairness of irrigation water distribution;
- CWP – the unit of physical crop production or yield per unit water consumed.

The indicators are applied to the scheme based on available data relating to crop information. The crop coefficient is only known for two schemes, and therefore, that indicator is only applied in those two schemes. In plots where the pixel size is less than the plot size, only pixels fully contained within the plot are considered. In plots where the pixel size is greater than the plot size, the data is extracted as point data to a point in the middle of the plot.

Table 5.2: Irrigation performance indicators and their properties

Criteria	Performance Indicator	Definition	Applied irrigation schemes
Adequacy	Relative evapotranspiration	$\frac{ETIa}{CWR}$	Metehara, Wonji
Equity	CV of evapotranspiration	$\frac{Standard\ deviation\ ETIa}{Mean\ ETIa}$	All
Productivity	CWP	$\frac{beneficial\ biomass}{ETIa}$	All

Adequacy was assessed using the relative evapotranspiration (relative ET) indicator. Relative evapotranspiration is estimated through the ratio of seasonal ETIa to the seasonal CWR (Kharrou et al. 2013; Karatas et al. 2009) and is assessed on an annual basis. The CWR was taken at the seasonal relative ET multiplied by the average seasonal crop coefficient (Kc). Due to data limitations in the WaPOR database, there is insufficient information on crop growth cycle to consider Kc at a higher temporal resolution. The single Kc method, rather than the dual crop coefficient method, was used, as the purpose of this manuscript is not to test the actual adequacy but the ability of datasets with different resolutions to capture the variation. This indicator was assessed in the Metehara and the Wonji, where the schemes are heavily dominated by one major crop type at two levels, between irrigation plots and between irrigation schemes. The average annual sugarcane Kc for a sugarcane ratoon crop was taken as approximately 0.95 (Allen et al. 1998).

Equity is a measure of irrigation uniformity (Bastiaanssen et al. 1999; Abernethy 1990). It is assessed through estimating the coefficient of variation (CV) of ETIa at two levels—between irrigation plots and between irrigation schemes. The ETIa is assessed at annual and dekadal scale.

CWP is defined here as the ratio of beneficial biomass to ETIa. The CWP can vary based not just on management and irrigation practices (Pinter et al. 2003; Oweis and Hachum 2003; Bossio et al. 2008), but on climate (Sadras et al. 2007; French and Schultz 1984), crop cultivar (van Wart et al. 2013) and other environmental factors (Hatfield, et al. 1984; Ali and Talukder 2008; Kijne et al. 2003) therefore it is assessed at only one level, between irrigation plots. The beneficial biomass is taken as the AGBP. The ABGP is a reasonable indicator of yield. The CWP is assessed at an annual scale.

2.4 Validation

Plot data were collected at two irrigation schemes—AGBP data in the Wonji, and ETa data in the Zankalon. The observed AGBP (AGBPa) data was collected from farmers for the period 2009–2016 and were provided by the sugarcane estate managers in the Wonji. In-situ ETa data was collected through eddy covariance (EC) observations, which was operated through the University of Tsukuba in partnership with Cairo University, the National Water Research Center, Qalubia, Egypt, and the Agriculture Research Center, Giza, Egypt (Sugita et al. 2017).

Farmer estimates of fresh AGBPa were available for 66 plots in the Wonji within the WaPOR data timeframe. The reported AGBPa (with planting and harvest dates) were used for a direct NPP comparison. The average dekad value for each NPP pixel falling within the plot was

aggregated over the growing season for each plot. The WaPOR NPP was converted to WaPOR AGBP (dry matter) using Equation (2.6) and then converted to a fresh AGBP (AGBPe) using conversion factors shown in Table 5.3 (i.e., AGBPe = WaPOR AGBP multiplied by each of the conversion factors). The reported yield is fresh matter (compared to WaPOR, which is dry); therefore, a conversion for moisture content was also required.

Table 5.3: Conversion factors used to convert WaPOR dry AGBP to sugarcane for comparison against farmer reported fresh AGBPa.

Factor	WaPOR AGBP	$AGBP_a$	Conversion factor*
LUE (gC/MJ)	2.7	2.6 (Blatchford et al. 2020)	0.96
Moisture content (-)	-	0.65 (Yilma 2017)	$\frac{1}{1-0.65}$
Above ground fraction (-)	0.65	0.8 (Yilma 2017)	1.23
HI	-	0.95	0.95

* The conversion factor is the factor applied to the dry WaPOR AGBP to make it comparable with the $AGBP_a$; the conversion factor is a result of reducing the difference between the values associated with the WaPOR AGBP and the $AGBP_a$.

The EC site is an irrigated site in the Nile Delta and is under rotation with three major summer crops—rice, maize and cotton—and four major winter crops—wheat, berseem (Trifolium alexandrinum), fava beans, and sugar beet (Sugita et al. 2017). The field is 200 m × 200 m, surrounded by agricultural land, extending at least 800 m in the dominant (northwest) wind direction. Suitable data for validation covers 2010–2013. The WaPOR-derived ETIa dataset was spatially averaged over a 750 m × 750 m pixel window covering the EC station, based on the assumption that the window represents the measurement footprint of the EC station. The WaPOR ETIa for the in-situ comparison was taken as the sum of soil evaporation, plant transpiration, and interception. The ETa from the EC (ETa-EC) data were derived from the latent heat flux (LE) measurements and aggregated temporally to dekadal averages to match the temporal resolution of the WaPOR ETIa products. Footprint data was unavailable, which may introduce bias to the spatial analyses of ETa. At L1 resolution, plot size is smaller than the pixel resolution, and a single pixel extends beyond the experimental field. However, due to the high correlation between L1 and L2 (which fits within the field), we decided to include EC data for the validation at both levels.

3. Results

3.1 ETIa and AGBP

The mean annual ETIa and AGBP for each resolution and each scheme are shown in Table 5.4. The Wonji and Metehara have the highest AGBP; this is to be expected as both schemes are dominated by sugarcane. ODN has the highest ETIa, which is expected as it is dominated by rice. The ETIa has a difference of less than 6% between levels in the Wonji, the Metehara, and Zankalon; 5–9% in ODN; and 11–17% in the Koga. The DMP has a difference of less than 6% between levels in the Metehara, ODN and Zankalon, 8% in the Wonji, and 26–29% in the Koga.

Table 5.4: Mean annual ETIa and DMP for irrigated schemes for 2009-2018 (scheme coefficient of variation (CV) reported in brackets)

Product	Dataset	Wonji	Metehara	ODN	Koga	Zankalon
ETIa	L3	1498	1648	1832	793	1394
(mm		(0.17)	(0.34)	(0.14)	(0.19)	(0.12)
year^{-1})	L2	1433	1557	1664	884	1368
		(0.12)	(0.34)	(0.12)	(0.15)	(0.09)
	L1	1480	1591	1736	926	1406
		(0.11)	(0.33)	(0.13)	(0.16)	(0.08)
AGBP	L3	37.2	46.2	16.4	2.3	2.3
(ton ha^{-1})		(0.21)	(0.16)	(0.27)	(0.19)	(0.16)
	L2	40.1	49.1	16.1	2.9	2.4
		(0.14)	(0.16)	(0.19)	(0.15)	(0.11)
	L1	40.0	48.7	16.1	(0.16)	2.4
		(0.14)	(0.15)	(0.20)		(0.10)

3.2 Adequacy

Each resolution estimates a similar annual mean relative ET across the scheme for all years and all levels; it follows the similar inter-annual trends in both the Wonji and the Metehara (Figure 5.2). In the Wonji, the annual relative ET, averaged across all years (2009–2018), ranged from 0.61–0.78 for L3, 0.63–0.76 for L2, and 0.67–0.75 for L1. The Metehara had higher mean relative ET and a higher variation in relative ET as compared to the Wonji for all levels. In the Metehara, the mean annual relative ET across all years ranged from 0.88–0.94 for L3, 0.81–0.90 for L2, and 0.80–0.88 for L1.

Across levels, the L3 mean relative ET is frequently higher than the L2 and the L1 ETo in both the Wonji and the Metehara. The L3 relative ET variation is consistently higher in the Wonji, which is reflected in the larger quartile range shown in Figure 5.2. The variation between plots was the highest in 2018 for both the Wonji and the Metehara.

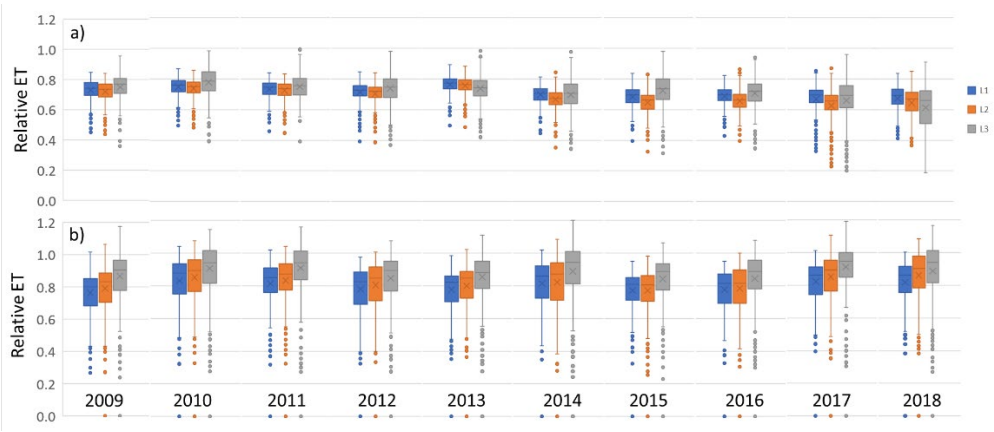

Figure 5.2: Box and whisker plot showing annual mean, quartiles and range relative ET of the a) Wonji and b) Metehara schemes for the L1, L2, and L3 resolutions.

Figure 5.3 shows the relative ET in 2018 for sugarcane plots in the Wonji and in the Metehara derived from WaPOR. All three levels can pick up spatial trends in relative ET. All levels show lower relative ET in the north-eastern part of the Wonji Scheme and the higher relative ET in the eastern and the north-western parts of the Wonji. In the Metehara, all levels showed that the centre of the scheme has higher relative ET, and the southern part of the scheme has lower relative ET. The higher resolutions can capture plot-to-plot variability better. A clear example is in the western part of the Wonji scheme. L1 shows all plots in this area ranging from 0.55–0.80, whereas in the same region, L3 is identifying plots to range from 0.35–0.85.

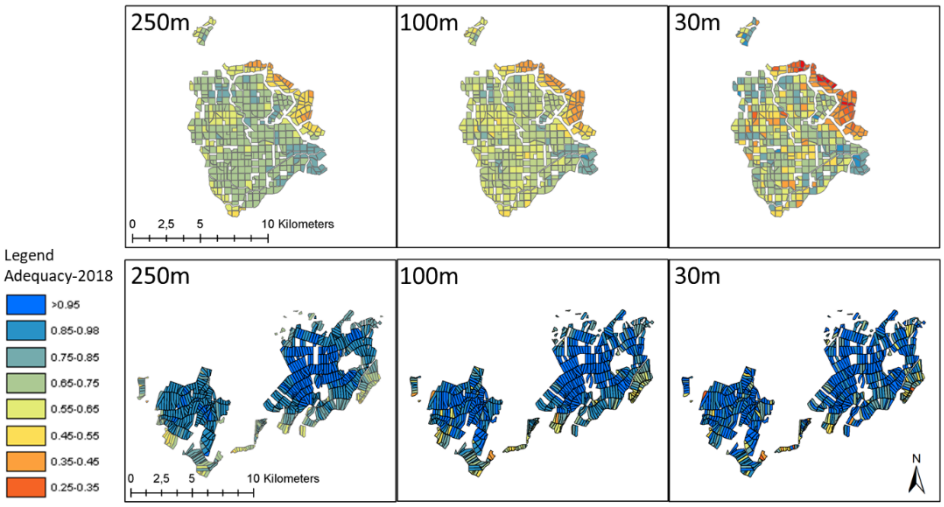

Figure 5.3: Average plot adequacy (relative ETa) in 2018 for the (upper) Wonji and (lower) Metehara schemes for L1, L2, and L3 resolution.

3.3 Equity

The 2009–2018 annual scheme ETIa CV was always highest for the L3 dataset (Table 5.5). The L3 annual scheme ETIa CV was higher for all schemes in all years on all but two occasions (Figure 5.4). All levels show similar inter-annual trends in CV; however, the magnitude of interannual variability was greater for the L3 dataset. The Koga was an exception: while L2 and L1 followed the same interannual trend, the L3 CV showed an opposite trend in years 2010–2011 and 2015–2018. Each level in the Koga showed similar mean annual ETIa trends; however, the L1 and L2 SD decreased in those years where L3 SD either increased or remained similar. This may be a result of varying year-to-year agricultural practices that are better picked up by L3. Alternatively, it may be a result of processing issues: for example, a wetter year can increase cloud cover and therefore data gaps, particularly of L3, where the revisit time is already lower than for L1 and L2.

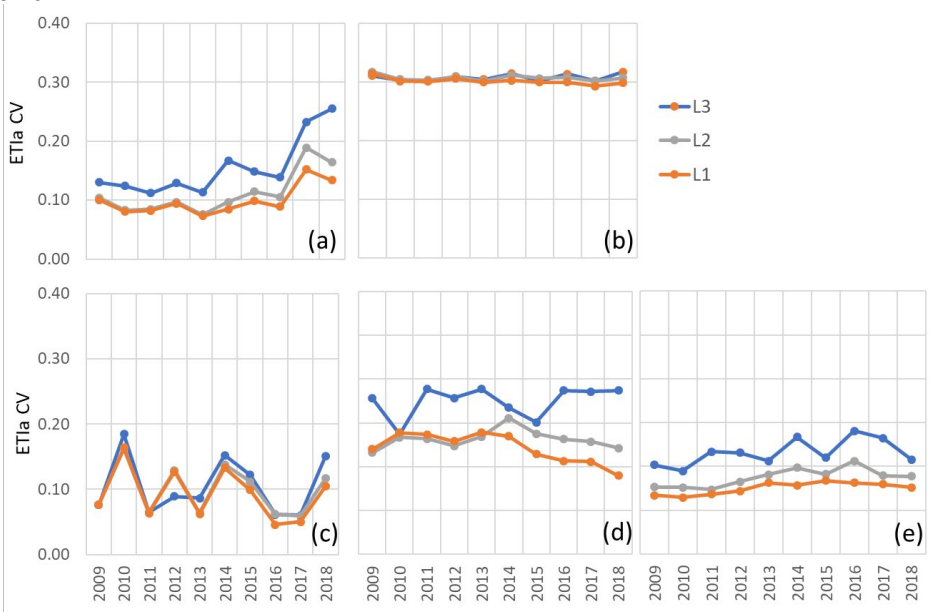

Figure 5.4: Annual intra-plot ETIa CV in each scheme (a) Wonji, (b) Metehara, (c) ODN, (d) Koga and (e) Zankalon.

Figure 5.5 shows the 2018 annual evapotranspiration in Zankalon and Metehara. The L3 dataset is capturing more spatial

variability than the L1 and L2 datasets in both schemes. The plot size is significantly smaller in the Zankalon, and the L3 is not capturing plot ETIa differences as well as in the Metehara. This is shown by the plot clusters of similar values. In the Metehara, despite having similar annual CV in 2018, it is evident that L2 is picking up more spatial variations than L1.

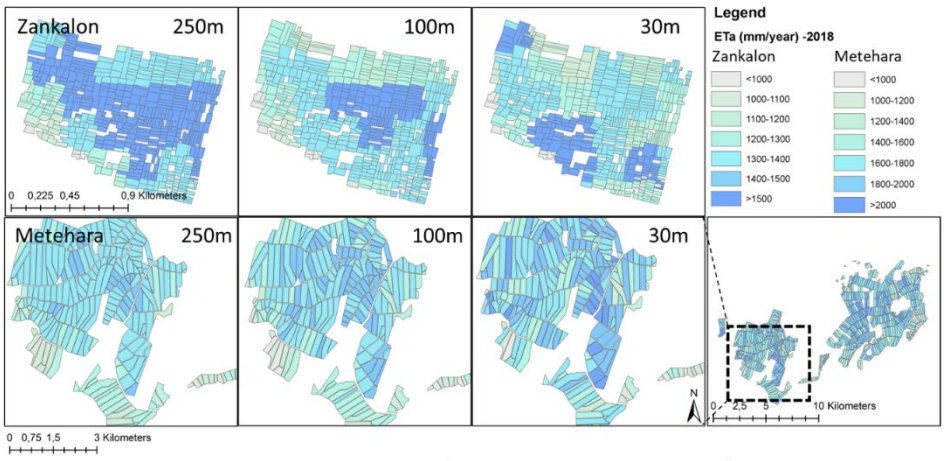

Figure 5.5: Average plot ETIa in 2018 for the (upper) Zankalon and (lower) Metehara schemes to compare three levels of spatial resolution (L1, L2, L3).

Figure 5.6 shows the dekadal ETIa CV for each scheme from 2009–2018. In most cases, a greater dekadal variation in the L3 CV compared to L1 CV and L2 CV suggests that L3 captures seasonal variation better. The L3 dataset reports the largest intra-annual variation among three datasets, particularly in schemes with smaller plots. This suggests that the lower resolution is doing a better job at picking up variation in the larger plots as compared to smaller plots, i.e., L1 may be suitable at identifying variation or equity in the Wonji but not the Koga, where differences between L1 and L3 are most significant. The magnitude of differences in CVs between datasets was less for schemes with larger plots (e.g., Wonji and Metehara). However, the Wonji has the least consistent difference in the CV at a dekadal scale. The L2 mimics the L1 CV trend up to 2013, where it then deviates in four of the five schemes. This corresponds with the introduction of the PROBA-V sensor into the L2 dataset. The L3 dataset captured the magnitude of variability better than the L1 and L2 datasets. However, the L3 dataset may not have captured the dekad-to-dekad changes as well. This is noted by a frequently smoother L3 trend line as compared to L1 and L2.

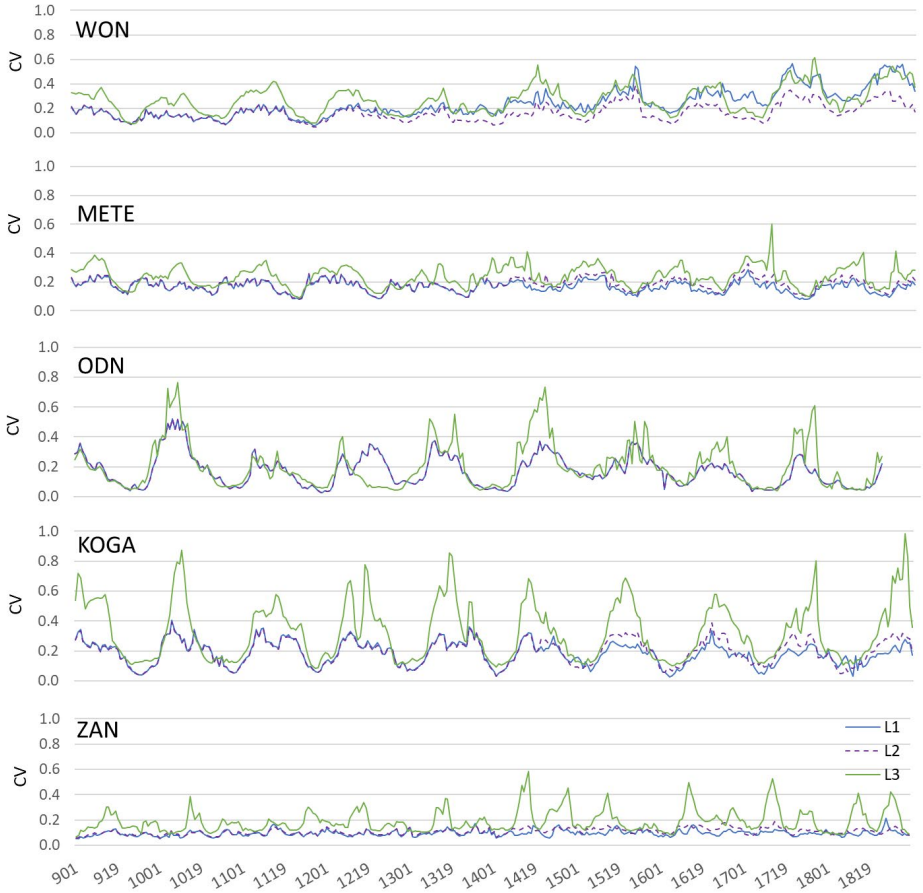

Figure 5.6: Dekadal intra-plot CV of ETIa for each irrigation scheme.

3.4 Productivity

The mean annual CWP values of 2009–2018 for each scheme are shown in Table 5.5. The L3 CWP is frequently lower than the L1 and L2 CWP due to the higher L1 and L2 AGBP estimates. The L3 CWP has the greatest level of difference in the Koga. The L3 CWP CV is always greater than the L1 and L2 CV. The scheme level CWP interannual variation is consistent between levels, i.e., the scale, magnitude and direction of change in CWP between years is the same for all levels. The CWP CV is smaller than the ETIa CV and DMP CV.

The examples of the plot level CWP in 2018 for the Wonji, the Koga, and the ODN are shown in Figure 5.7. While all levels capture the scheme average CWP similarly, the plot-to-plot variation is considerably different between levels. L3 captures the most variation

between plots. In the Wonji, all datasets capture scheme areas with higher or lower CWP well—for example, each level shows that the north-eastern part of the scheme has the lowest CWP and that the eastern tip has the highest. However, the L3 dataset clearly captures more variation with distinct plot-to-plot difference. The ODN shows distinct differences in plot-to-plot CWP variation between datasets. However, the L3 dataset can pick up more outlier plots. Zankalon, not shown here, showed similar results to the Koga, and the Metehara showed results more like the Wonji (also like ETIa spatial distribution patterns shown in Figure 5.5).

Table 5.5: Mean annual scheme CWP for 2009-2018 for all schemes (scheme CV in brackets).

Product	Dataset	Wonji	Metehara	ODN	Koga	Zankalon
CWP	L3	3.7	2.4	1.4	3.5	3.0
(kg m^{-3})		(0.08)	(0.10)	(0.16)	(0.09)	(0.05)
	L2	4.3	2.8	1.5	4.5	3.0
		(0.05)	(0.07)	(0.05)	(0.09)	(0.04)
	L1	4.2	2.7	1.4	4.4	3.1
		(0.05)	(0.07)	(0.05)	(0.09)	(0.03)

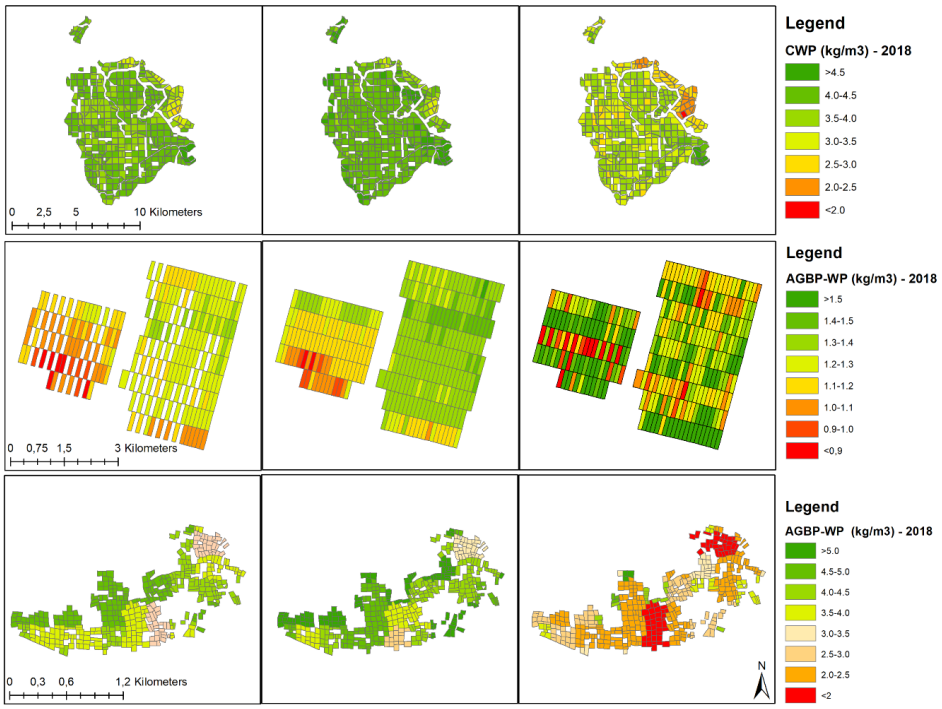

Figure 5.7: Average plot CWP (2018) in the (upper) Wonji, (middle) ODN, and (lower) Koga (lower) for the L1, L2 and L3 resolutions.

3.5 Validation-evaluation of the WaPOR dataset

The AGBPe is plotted against the farmer reported (or AGBPa) in Figure 5.8. The L3 dataset has the highest performance when compared to the in-situ data in terms of both coefficient of determination (R^2) and root mean square error (RMSE). The L3, L2, and L1 R^2 are 0.7, 0.4, and 0.5, respectively, and the L3, L2, and L1 RMSE are 33.9 ton ha^{-1}, 31.11 ton ha^{-1}, and 48.2 ton ha^{-1}, respectively, which equates to a normalized RMSE (NRMSE) of 22.5%, 21.8%, and 33.8%.

The AGBPa shows more variation between plots and years, with an AGBPa SD of 47 ton ha^{-1} compared to an AGBPe of 26 ton ha^{-1}. The AGBPe is comparable to the average global biomass yield which is 69.8 ton ha^{-1} (Steduto et al. 2012). FAO WATER (Steduto et al. 2012) reports that the cane yield varies from 50 ton ha^{-1} to 150 ton ha^{-1} depending on the variety and ratooning stages, which suggests that the WaPOR-derived values are within the reported range, while the farmer estimates are on the high side. The harvested biomass is related to the growing period. The average total growing period of sugarcane in the Awash L3 area is 585 days with a range of 305–1037 days or 0.8–2.8 years. WaPOR give reasonable results for sugarcane, but a gap exists in linking biomass to yield in remote sensing-based estimates (Blatchford et al. 2019). Often, a priori knowledge of harvest index is the most useful way to link biomass and yield.

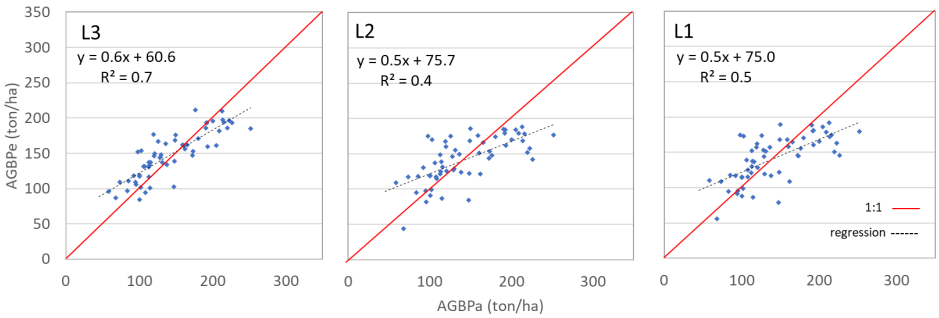

Figure 5.8: WaPOR-derived AGBPe for L3 (left), L2 (middle), and L1 (right) compared against AGBPa.

ETIa was compared to the EC in the Zankalon irrigation scheme. The 10-day daily average ETa-EC and WaPOR ETIa for all three spatial resolutions at the EC site are shown in Figure 5.9. The L1 and L2 ETIa show high consistency with each other. The L3 ETIa consistently sits between the ETa-EC and the L1 and L2 ETIa. All levels capture the

overall ETa-EC seasonal trends. The L3 data shows a slightly lower R^2 (L3 = 0.36, L2 = 0.60, and L1 = 0.61) and noticeably lower bias (L3 = 1.06 mm day^{-1}, L2 = 1.7 mm day^{-1}, and L1 = 1.7 mm day^{-1}) and a lower RMSE (L3 = 1.0 mm day^{-1}, L2 = 2.2 mm day^{-1}, and L1 = 2.2 mm day^{-1}) when compared with the ETa-EC. The higher R^2 associated with the L1 and L2 ETIa reflects their ability to capture the temporal fluctuations of ETa-EC better than L3 ETIa. An example of this is at dekad 1117 or 2011–17, where L1 and L2 ETIa capture the ETa-EC dip, whereas L3 ETIa stays flat. These findings are consistent with other validations that compare WaPOR L1 ETIa data with EC data in other agricultural fields in both the Nile Delta and other dry, arid regions (Blatchford et al. 2020). It should be noted that the same study found lower suitability of WaPOR data in very wet, non-agricultural areas.

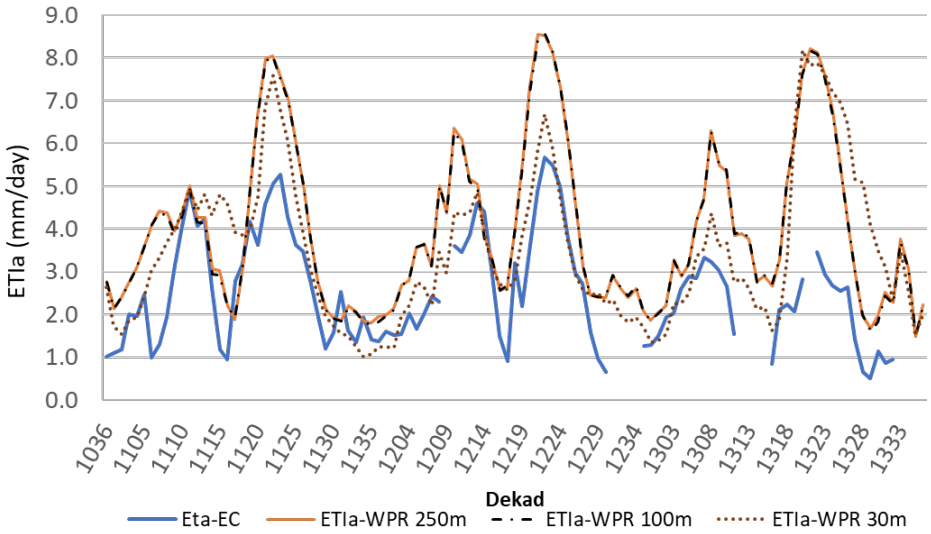

Figure 5.9: WaPOR dekadal L3, L2 and L1 ETIa (mm day^{-1}) compared to in-situ dekadal ETa (mm day^{-1}).

4. Discussion

This study shows the influence of spatial resolution on irrigation performance indicators, relevant to different plot sizes and schemes. All resolutions captured the seasonal trends. This was evident when observing indicators over time, i.e., equity (e.g., Figure 5.6). The higher resolution showed a higher accuracy, in terms of R^2 and RMSE, for AGBP compared to in-situ data. The relative AGBP errors (21%, 28%, and 28% for L3, L2, and L1, respectively) is in the upper 50th percentile of reported error ranges for non-parametrized remote-sensing-based estimates and is within the ranges of error associated

with farmer reported yields (Blatchford et al. 2019). Some users may prefer a locally parameterized model to achieve dataset accuracy.

The L3 ETIa had lower bias: the R^2 was lower than the L2 or L1 datasets when compared to EC data. While the L1 and L2 data had better R^2, they had a higher bias. Further, the L1 and L2 data have a higher temporal resolution in terms of spectral input, and when comparing to daily flux data, the temporal gaps of the L3 data may be influencing the accuracy more than the gain in spatial resolution. A study on a 22 ha olive farm showed that datasets with resolution coarser than L3 will result in great discrepancies as compared to actual evapotranspiration values due to the aggregation of non-linear components and to the inclusion of non-agricultural areas in such aggregation (Ramírez-Cuesta et al. 2019). Similarly, a study in wheat found that the 30 m NDVI dataset, as compared to 250 m and 1 km NDVI datasets, provided the most accurate yield estimates (Durgam et al. 2020). This is absolute error, not relative error in space and time, looking at variation.

The level consistency of ETIa and AGBP does not seem related to the plot size at annual, scheme level. However, the L3 dataset, at the scheme level, was able to capture the spatial variation better and provide more information on scheme level equity. Validation showed that although differences in datasets were evident, the spatial and temporal trends were consistent. However, the magnitude of seasonal and spatial variability varied between datasets and that the L1 and L2 may be in some cases overestimating the ETIa. However, it should be noted that the schemes with smaller plots (i.e., Zankalon and Koga) also had the highest crop diversity. This will drive more variation in the field yields and water consumption and also in satellite interpretation, as they are less homogenous. However, this is considered to reflect reality, as small holder agriculture is often associated with more diversity, for example in South Africa (Fanadzo et al. 2010).

Adequacy was captured well by all resolutions at an annual, whole of scheme level in the Wonji and the Metehara. While the L3 data showed higher relative ET CV, each level was able to capture the inter-annual trends. It is suggested that all resolutions are suitable to assess adequacy at interannual or inter-scheme level for the schemes assessed. Inter-plot spatial variability of adequacy was captured well by all in the Wonji and the Metehara, however, the L3 dataset captured variability the best. It is suggested that all resolutions are suitable to assess adequacy at interannual scheme level for the schemes assessed, though L3 is preferred for the scheme with smaller plots, i.e., Koga and Zankalon. This research is limited to assessing adequacy at annual scale, and in practice adequacy should be assessed at various stages of critical crop growth. Further research should consider the impact of resolution on a shorter timescale.

Inter-annual equity was captured reasonably well by all levels in the Wonji, Metehara, ODN and Zankalan. However, L3 was distinctly different to L1 and L2 in the Koga, suggesting that all levels are suitable for larger plots with mixed results for smaller plots. This may be influenced by the heterogenous landscape in the Koga, which is mixed with irrigated pixels (Ramírez-Cuesta et al. 2019). L1 and L2 are not suitable to assess equity at plot-to-plot scale in the Koga and Zankalon. Plot-to-plot equity was also not well captured by the L3 dataset in these schemes, which is identified by the plot clusters of similar values (Figure 5.5). However, the L3 and L2 datasets may be used to compare scheme plots or divisions.

Equity was also assessed on dekadal scale. All datasets captured some level of seasonal variability in each scheme. The L3 dataset captured the magnitude of variability better than the L1 and L2 datasets. The difference in CV, or equity, between resolutions decreased with increasing plot size, this is with the expectation of the Wonji, the scheme with the largest plots. The Wonji has the least inconsistent difference in CV of ETIa between spatial resolutions over space (e.g., adequacy in Figure 5.3) and time (e.g., equity in Figure 5.6). This is likely due to the introduction of Proba-V in March 2013, where the 100 m data is no longer resampled from the 250 m NDVI (or L1) data. In terms of ETIa, this only has a large effect in the Wonji. The other schemes only see a light deviation in results from the introduction of Proba-V. However, the L3 dataset did not capture the dekad-to-dekad changes as well, which was identified by a sometimes smoother L3 trendline. This may be a result of the revisit time of the L1 and L2 satellite data input of one or two days, compared to the L3 dataset of 16 days, providing less NDVI inputs and smoothing the results. This may become particularly influential during raining seasons, where cloud cover reduces the satellite data availability further for satellites with lower return periods and for users wishing to look beyond a seasonal scale. This affect may be being reduced by the daily meteorological data used to interpolate image gaps.

Inter-annual scheme CWP was captured by all levels for all schemes irrelevant of dataset and therefore suggested that all resolutions are suitable for scheme-scale inter-annual comparison. This is compared to another study which showed that at a regional scale, the L3 resolution provided little extra value in determining agricultural boundaries or NDVI trends (Pax-Lenney and Woodcock, 1997).

The three levels are derived from different sensors and therefore have different nominal band wavelengths, Landsat being the most dissimilar to the other levels. The primary use of optical bands in WaPOR is to estimate various vegetation indices (primarily NDVI). Studies in marshes and agricultural lands have shown that nominal band wavelengths limit differences between MODIS and Landsat do not

greatly impact vegetation indices such as the enhanced vegetation index (EVI) or NDVI (Jarchow et al. 2018; Albarakat and Lakshmi, 2019). However, the WaPOR products are not currently inter-calibrated, and the impacts of varying nominal wavelengths in WaPOR is not well known. The product producer (WaPOR) is currently undertaking inter-calibration to improve dataset consistency, which is suggested for best practice quality assessment (Zeng et al. 2015; Zeng et al. 2019).

Though it has been suggested that Landsat 30 m data is a valuable asset to water management, which may be true, it still may not be of high enough resolution for monitoring and evaluating small-scale agriculture (Anderson et al. 2012). Pixel purity represents the relative contribution of the surface of interest to the signal detected by the remote sensing instrument (Duveiller, 2012). For smaller plots, the signal will be mixed by varying crops or non-agricultural area. Therefore, when the plot size it smaller than the pixel size, pixel mixing will arise. We found that at the scheme level, for absolute indicators, such as productivity or adequacy, the lower resolution products are still suitable to assess scheme-level performance over time. However, for relative indicators (i.e., equity), the spatial variation is integral, and the higher resolution product is required.

Therefore, while accuracy is lower and variation is not as well captured at the plot level, at the scheme level, L1 and L2 resolutions appear suitable and save time and processing costs. However, for higher-level inter-plot comparison, the L3 data provided value in all schemes for all indicators at both a dekadal and an annual scale. However, for patchwork agricultural landscapes, with smaller plots, i.e., <2 ha, utilisation of Sentinel (10 m visible bands), should improve interpretation and monitoring inter-plot irrigation performance. This is recommended, as the L3 data had much more variation in the scheme with small plots (Koga and Zankalon), and therefore, a higher resolution baseline or reference dataset would provide added insight to the suitability of the L3 data. As WaPOR moves toward utilization of the Sentinel 2 and 3 platform, the relevant datasets will become readily available for comparison and application in assessing irrigation performance.

5. Conclusion

There is an increasing ability to monitor irrigation performance from space with available open-source datasets. It is important that the effectiveness of these datasets be understood based on irrigation scheme characteristics and spatial resolution of the dataset itself. As the methodology for all datasets is the same, the resolution is a unique varying factor to compare applicability in irrigation performance assessment. The following is a brief summary and discussion of this

study, presented here to understand and characterize the application of WaPOR datasets with 30 m, 100 m, and 250 m spatial resolutions in irrigation performance assessment for schemes with varying plot sizes of small (<2 ha) to medium (>10 ha) plots.

The following conclusions can be drawn from this study on comparison of varying remote-sensing-based resolution datasets in irrigation performance assessment:

- Spatial resolutions of 250 m, 100 m, and 30 m are suitable for inter-annual and inter-scheme assessments for adequacy, equity, and CWP, regardless of plot size;
- Spatial resolutions of 250 m and 100 m should not be used for inter-plot comparison for adequacy, equity, or CWP on plots <2 ha. The 30 m resolution may also be too coarse, and Sentinel-2 application should be considered;
- Spatial resolutions of 250 m and 100 m show general spatiotemporal trends for adequacy, equity, and CWP within a scheme, but not the full extent of plot-to-plot variation for all plot sizes tested.

It is suggested that these conclusions can be applied generally to irrigation areas that have similar plot sizes. However, further investigation into the resolution requirements to suitably undertake irrigation performance assessment in small plots, i.e., ≤2 ha, is needed.

Chapter 6

Determining representative sample size for validation of continuous, large continental remote sensing data[4]

[4] This chapter is based on: Blatchford, M.L., Mannaerts, C.M., Zeng. Y., 2021. Determining representative sample size for validation of continuous, large continental remote sensing data, International Journal of Applied Earth Observation and Geoinformation. JAG 94 (2021) 102235.

Abstract

The validation of global remote sensing data comprises multiple methods including comparison to field measurements, cross-comparisons and verification of physical consistency. Physical consistency and cross-comparisons are typically assessed for all pixels of the entire product extent, which requires intensive computing. This paper proposes a statistically representative sampling approach to reduce time and efforts associated with validations of remote sensing data having big data volume. A progressive sampling approach, as typically applied in machine learning to train algorithms, combined with two performance measures, was applied to estimate the required sample size. The confidence interval (CI) and maximum entropy probability distribution were used as indicators to represent accuracy. The approach was tested on 8 continental remote sensing-based data products over the Middle East and Africa. Without the consideration of climate classes, a sample size of 10,000-100,000, dependent on the product, met the nominally set CI and entropy indicators. This corresponds to <0.01% of the total image for the high-resolution images. All continuous datasets showed the same trend of CI and entropy with increasing sample size. The actual evapotranspiration and interception (ETIa) product was further analysed based on climate classes, which increased the sample size required to meet performance requirements, but was still determined to be significantly less than the entire dataset size. The proposed approach can significantly reduce the processing time while still providing a statistically valid representation of a large remote sensing dataset. This can be useful as more high-resolution remote sensing data becomes available.

1. Introduction

Global and continental remote sensing data sets are increasingly available in terms of access and dataset type. Satellite temporal and spatial resolution, i.e. the pixel size and return period, are also increasing as satellite technology improves. As the dataset resolution improves, the processing time increases not only for production but also for validation.

A common part of validation procedures for continental (and larger) datasets is verification of spatial and temporal consistency and cross-comparison to other datasets (Loew et al. 2017; Zeng et al. 2019, 2015). Spatial and temporal consistency is a physical consistency test that considers the variation and relative uncertainty of the product over space and time, while cross-comparison compares the product directly to a reference product developed by a different producer, or using a different satellite, algorithm etc. Physical consistency analyses variation of the product, dependent on factors such as climate and season. This is seen as a component of the validation strategies of various producers including the Copernicus Global Land Surface Products, which includes vegetation (e.g. dry matter productivity and leaf area index), energy (e.g. surface albedo), water (e.g. water bodies) and cryosphere products (Smets et al. 2013), Advanced Along-Track Scanning Radiometer (AATSR) Land Surface Temperature (LST) Validation Strategy (Schneider et al. 2012), the FAO Water Productivity Open-access portal (WaPOR) of Remotely sensed derived data validation methodology, which includes vegetation (e.g. net primary productivity - NPP) and water (precipitation and evapotranspiration - ETa) products (FAO, 2018) and Moderate-resolution Imaging Spectroradiometer (MODIS) land validation strategy (MODLAND), which includes vegetation (e.g. vegetation indices and NPP) and energy (e.g. LST) (Morisette et al. 2002).

Typically, physical consistency and cross-comparisons are evaluated over the entire extent of the dataset on a pixel-by-pixel basis to determine spatial and temporal trends and differences. This not only requires substantial computational costs as spatial resolution increases but can also be excessive to understand the performance of the dataset. Alternatively, samples can provide enough insight to accuracy with less computation to evaluate these arbitrarily large datasets. In several fields, including in land use classification (Heydari and Mountrakis, 2018), machine learning, bioinformatics (Kim, 2009), clinical studies (Gupta et al. 2016; Kirby et al. 2002; Lachin, 1981) and classifier design studies (Fukunaga and Hayes, 1989), scaling-down techniques are used to approach the problem of training large datasets by selecting a sample of the data which is meant to accurately represent the entire dataset. However, determining the appropriate

sample size with large datasets is not always obvious and has not been applied in validation.

Machine learning frequently deals with developing and training algorithms for large databases. In machine learning the primary categories of scaling-down sampling methods are random selection, active learning techniques and progressive sampling (ElRafey and Wojtusiak, 2017). Random sampling uses passive learning, active learning uses semi-supervised machine learning to choose data from which it learns. Active learning and passive learning methods typically use an arbitrary, predefined sample size (Warmuth et al. 2003). Active learning algorithms seek to select the most informative cases for training while progressive sampling aims to minimize the amount of computation for a given performance target.

Progressive sampling incrementally increases the sample size until the accuracy of the algorithm no longer improves, or converges (Luo, 2016) and is designed to efficiently produce models with high accuracy (Meek et al. 2002). This prevents processing the entire database, which may be resources heavy (Ng and Dash, 2006). Progressive sampling helps balance the prediction accuracy and the data processing effort (Sarkar et al. 2016) to determine the optimal sample size (Gu et al. 2001).

Progressive sampling is sparsely identified in literature in remote sensing applications. It has been applied to learn neural network ensembles of arbitrarily large datasets (Peng et al. 2004), digital terrain modeling (DEM) data acquisition (Chen and Li, 2013; Makarovic, 1973), and has been integrated into digital image mapping (Rauhala, 1989). Most commonly, progressive sampling has been used in clinical studies (Figueroa et al. 2012) and in training algorithms in machine learning and association rules in data mining (Last, 2009; Ng and Dash, 2006; Umarani and Punithavalli, 2011; Zeng and Luo, 2017). Examples in other fields include field sample design for ecological studies (Stein and Ettema, 2003) and argo-ecological characterization (Steduto and Todorovic, 2001).

The accuracy or performance of a product can be estimated through many metrics. For discrete variables, such as land classification or in machine learning applications, the accuracy is often taken as the number of correct predictions over the total number of data predictions (Congalton and Green, 2009; Foody, 2002). For continuous datasets, the accuracy metrics used are more diverse. Regression metrics are commonly used when true value is known. However, physical consistency and cross-comparisons is performed in the absence of known true values and is a method to observe comparative performance of the data over space and time rather than provide absolute accuracy.

The coefficient of variation (CV) is an index of reliability or relative variability, which is commonly used in several fields of science (Payton, 1996; Reed et al. 2002; Schectman, 2013). The confidence interval (CI) is expressed in terms of the variation around the expected value in terms of the CV or standard deviation. The CI can be used to express relative accuracy (Burt et al. 1997; Young and Lewis, 1997). The Principle of Maximum Entropy states that the distribution with the maximum entropy best matches the current state of knowledge and provides a measure of the amount of information needed to represent an outcome from a probability distribution for a random variable. Entropy was first formulated to understand the diversity and uniformity of discrete variables otherwise known as the Shannon's Index (Shannon, 1948). It was later generalised to the differential entropy and to continuous random variables (Jaynes, 1957; Santamaría-Bonfil et al. 2016). It has recently been used as an indicator, among others, to evaluate satellite based soil moisture retrievals as compared to ground-truth measurements (Kumar et al. 2018). These indicators may be useful for assessing the representative sample size as they reflect both mean and standard deviation, along with probability distributions without prior information (Cover and Thomas, 1991; Sim and Reid, 1999).

The purpose of this manuscript is to estimate the sample size required to accurately represent the modelled or estimated dataset. The purpose of this manuscript is not to estimate the sample size required to determine the 'true' value, or to determine the accuracy of the dataset as compared to a 'true' value.

It is proposed that validation of large remote sensing datasets, such as physical consistency and cross-comparison, can be analyzed through a representative sample size, which can be determined by the performance requirements of the dataset. This paper proposes that the CI and the maximum entropy probability of the sample dataset can define the threshold of the required sample dataset size to run these validation activities. A simple progressive sampling approach, used in machine learning and algorithm training, is adapted and used to determine the sample size to yield statistically significant results.

2. Materials and Methods

2.1 The dataset

The approach was applied to six remote sensing-based datasets that cover continental Middle East and Africa (Chapter 2 – Figure 2.1). Remote sensing-based products include actual evapotranspiration and interception (ETIa), net primary productivity (NPP), solar radiation (SR), reference evapotranspiration (ETo), relative soil moisture index

(SM) and normalized difference vegetation index (NDVI). The data used covers two spatial resolutions and two temporal resolutions. The resolution, image date, pixel count, image CI and sensor or data product used as input, for each image used is shown in Table 6.1.

Table 6.1: Dataset properties for images used

Data	Resolution	Image date	Pixel count	CI*	Data product/s	Sensor/s
ETIa	250m|10-day	10 Apr 2009	1.22E+09	1.78	MOD09GQ	MODIS; MERRA; MSG
SM	250m|10-day	10 Apr 2009	1.22E+09	0.84	MOD09GQ	MODIS
NPP	250m|10-day	10 Apr 2009	1.22E+09	2.81	MOD09GQ	MODIS; MERRA; MSG
NDVI	250m|10-day	10 Apr 2009	1.22E+09	1.27	MOD09GQ	MODIS
SR	250m|1-day	18 Nov 2009	1.22E+09	0.66	STRM (DEM)	MSG
ETo	25km|1-day	15 Dec 2009	2.38E+05	0.92	STRM (DEM)	MERRA; MSG

*CI is confidence interval as defined in Equation 6.2.

All data was sourced from the Level 1 continental products in the WaPOR database version 1 (FAO, 2017). The ETIa, NPP and ETo are sourced directly from the WaPOR portal (https://wapor.apps.fao.org/home/WAPOR_2/1), and the NDVI, SR and SM were provided by the WaPOR dataset producers (see Chapter 2 for more information).

Further, the CI performance criterion was applied to the ETIa product, and tested for different climate classes using a Köppen-Geiger classification (Kottek et al. 2006). It is suggested that this approach can be used in evaluating the dataset for physical consistency, convergence should be achieved, or quantified, for each class or characteristic in which the physical consistency test is applied. For example, division of regions or classes can be based on hemisphere, climate or land use. The four major climate classes were: Arid (2.6E09 km2), Equatorial (1.1E09 km^2), warm temperate (0.3E09 bil km^2) and humid continental (360,450 km^2). The class arid, equatorial, warm temperate and humid continental are represented with all classes starting with B, A, C and D respectively.

2.2 Sampling schedule

The sample schedule in this research followed a geometric approach, due to the large size of the data (Estrada and Morales, 2004), using one geometric constant – a = 10, and two starting sample sizes, n0 = 100 and n0 = 300:

$$S_i = [\, a \cdot n0,\ a^2 \cdot n0,\ a^3 \cdot n0, \ldots) \quad \text{where} \quad n0 = 100 \text{ and } n0 = 300 \tag{6.1}$$

where Si is the sampling schedule, n0 is the starting sample size, a is the geometric factor and Ni is the sample size. Each sample increment, or sample size, is referred to as Ni. Each sample size was extracted 10 times (a sample set), with samples within a sample size referred to as Nj, Nj+1, Nj+2… Nj+9. This allows repetition of the test. The constant, a, was selected so the sample size can be identified quickly, considering the large size of the dataset (Chapter 2 - Table 2.2). The starting sample sizes were selected randomly as a way to reduce the aggressiveness of the approach. The sampling schedule tentatively reached Ni = 3,000,000, which can be extended if convergence is not met. A random sampling method was used for all sampling schedules to ensures consistency. It was assumed that the results achieved for a single image can be extrapolated over time.

2.3 Sample extraction

All spatial samples were randomly generated in R software. One sample was generated for each test. Therefor there are 100 random samples per dataset (NPP, ETIa etc), i.e. 10 randomly generated samples for each of the 10 sample sets. These random spatial point datasets were then used to extract the dataset values for each of the datasets. Where climate classification is considered, the climate class associated with each spatial point in the random spatial dataset was over-layed to extract the feature (climate class) of that point.

2.4 Performance measures

This study will use the CI and the differential entropy, or maximum entropy distribution, of the sample dataset (x) as indicators or performance criterion. Nominally, the acceptable performance for this case study was taken as 5% ($\Delta CI_{i,j}=5\%$). The acceptable entropy was defined as the entropy where the dependence on Nj is negligible. It is set nominally as $\Delta H(x)=0.05$.

The definition of the CI used as an indicator in this research is taken as (Clemmens and Burt, 1997):

$$CI_x = \pm 2CV_x \tag{6.2}$$

Where, CI_x is the CI and CV_x is the coefficient of variation. This CI definition is commonly applied in hydrology. It gives a measure of the CI relative to the magnitude of the expected value rather than the actual value which is found when using the z-coefficient (i.e. number of standard deviations) (Clemmens and Burt, 1997). The CV_x is defined as:

$$\mathrm{CV_x} = \mathrm{sd_x}/\bar{\mathrm{m}}_\mathrm{x} \tag{6.3}$$

Where sd_x is the standard deviation of the sample dataset and $\bar{m}_x$ is the mean of the sample dataset. The sd_x is the standard statistical measure of variability. Although the CI is formally taken as the mean ±2sdx, the CI defined in equation (6.4) provides a relative accuracy, with no units and often expressed as a percent.

The Principle of Maximum Entropy for continuous distributions, the differential entropy, is defined as (Jaynes, 2003, 1957):

$$\mathrm{H(x)} = -\int_{-\infty}^{\infty} \mathrm{p(x)} \ln \mathrm{p(x)} \mathrm{dx} \tag{6.4}$$

Where H(x) is the differential entropy and p(x) is the probability density function. This function applies to any probability density function that can be defined. The base of log is not important as long as it is uniform, as changing the base simply changes the scale of the entropy (Rajan et al. 2017). This requires a randomly generated sample dataset. The natural logarithm (ln) is used in this case. The entropy of the dataset will increase with an increasing population. The higher the entropy, the more information is given to that distribution. Therefore, when the marginal increase in entropy is negligible or minimal, little to negligible information can be gained by increasing the population size.

2.5 Detecting convergence

Detecting convergence requires statistical judgement on what performance is suitable. Once the required performance is determined, the suitability of the sample size can be determined. The Probably Close Enough Criteria (PCE) is a deduction procedure used in machine learning. It outputs an expression that has a high likelihood of closely approximating the expression to be learned (Valiant, 1984). Meaning that there is only a small chance that the mining algorithm could do better in training the algorithm using the entire database instead of the defined sample size. The PCE defines the suitable sample size as (John et al. 1996; Provost et al. 1999):

$$(acc(N_{i+1}) - acc(N_i) > \Delta \mathrm{E}\,) \leq \delta \tag{6.5}$$

Where acc(N) refers to the accuracy (acc) of the sample size (Ni). ΔE refers to the acceptable increase in accuracy (or marginal increase) and δ is the probability that the maximum accuracy will be exceeded on any run, therefore, satisfying the accuracy requirement for each run of a sample size and any increase in sample size.

The PCE criterion is adapted to determine the suitable sample size with the selected performance measures. The marginal increase in performance, referred to as ΔE in equation 6.5, is determined by calculating the statistical variation in CI, or the differential entropy with increasing sample size steps (Ni):

$$(\mathrm{CI}_{i+1,j} - \mathrm{CI}_{i,j} \leq \Delta\mathrm{CI}_i) \quad \text{and} \quad (\mathrm{H(x)}_{i+1,j} - \mathrm{H(x)}_{i,j} \leq \Delta\mathrm{H(x)}_i) \quad (6.6)$$

Where acc(N) refers to the accuracy (acc) of the sample size (Ni). ΔE refers to the acceptable increase in accuracy (or marginal increase) and δ is the probability that the maximum accuracy will be exceeded on any run, therefore, satisfying the accuracy requirement for each run of a sample size and any increase in sample size.

The PCE criterion is adapted to determine the suitable sample size with the selected performance measures. The marginal increase in performance, referred to as ΔE in equation 6.5, is determined by calculating the statistical variation in CI, or the differential entropy with increasing sample size steps (Ni):

$$(\mathrm{CI}_{i,j+1} - \mathrm{CI}_{i,j} \leq \Delta\mathrm{CI}_j) \quad \text{and} \quad (\mathrm{H(x)}_{i,j+1} - \mathrm{H(x)}_{i,j} \leq \Delta\mathrm{H(x)}_j) \quad (6.7)$$

Where the ΔCIj and ΔH(x)j is the range of performance of samples Nj for sample size Ni. If the difference in the CI is greater than the acceptable CI (ΔCIj) for any sample, Nj, the sample size is rejected and sample size is increased. Both equations 6.8 and 6.9 need to be met for the sample size to be considered suitable. This was undertaken for the entire sample data set.

The expected trend of the CI and the entropy for an infinitely large, positive, continuous dataset, with increasing sample size, is shown in Figure 6.1. The black crosses represent the CI or differential entropy for each sample Nj for a given sample size Ni, within the sampling schedule, Si. The black lines show the expected maximum (plotted for CI and entropy) and minimum (plotted for CI only) values of the sample set. The decreasing range in performance metric values for increasing sample size reflects convergence. The CI is expected to converge at a mean value, while entropy is expected to converge to a maximum value.

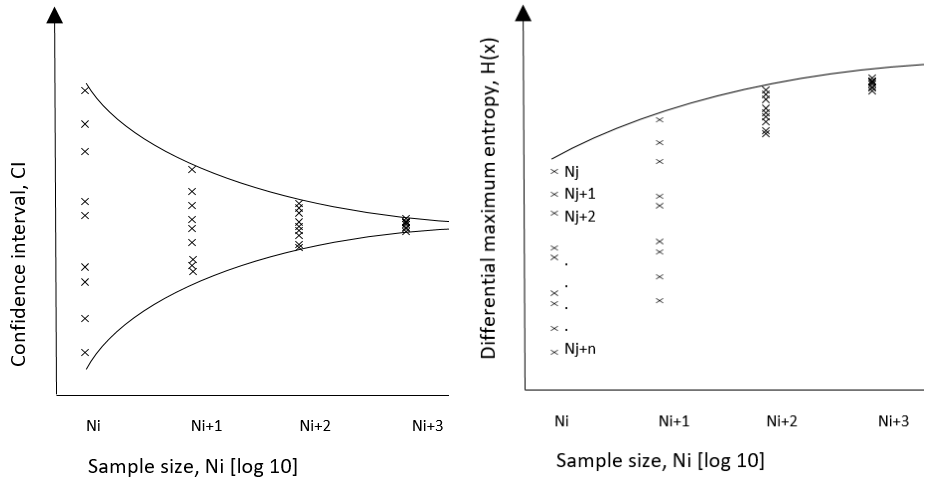

Figure 6.1: Expected learning curve trend of the CI (a) and the entropy (b) for repeated tests and increasing sample size. Black lines indicate trends. The black line shows the expected trend of variation for increasing sample size.

2.1 Detecting convergence

An internal cross-compared was undertaken to determine if relationships between datasets remained with increasing sample size. There is a strong link between ETIa and NPP (Blatchford et al. 2019), which is therefor used to assess the relationship variation with increasing sample size. Correlation was selected as the metric. The correlation was estimated for all samples in all sample sets.

3. Results

The CI for the full set of sampling schedules and for all datasets, are shown in Figure 6.2. The sample size is plotted on the x-axis, using a logarithmic scale, and the CI is plotted on the y-axis. The CI shows the most variation at low sample sizes and converges as the sample size increases. The range of CI is decreasing with increasing sample size for all samples to the CI value seen in Table 6.1. The rate of convergence is also greatest when the sample size is low, and decreases with increasing sample size. There are 3 occasions where the range in CI increases with an increasing step size: the ETIa CI range is greater when Ni = 100,000 than when Ni = 30,000, the ETo CI range is greater at Ni = 1000 than when Ni = 300 and the NPP when Ni=100.000 and Ni=300.000. This type of variation before convergence is expected as extreme values are expected to have a greater influence in the data distribution and CI when the sample size is small.

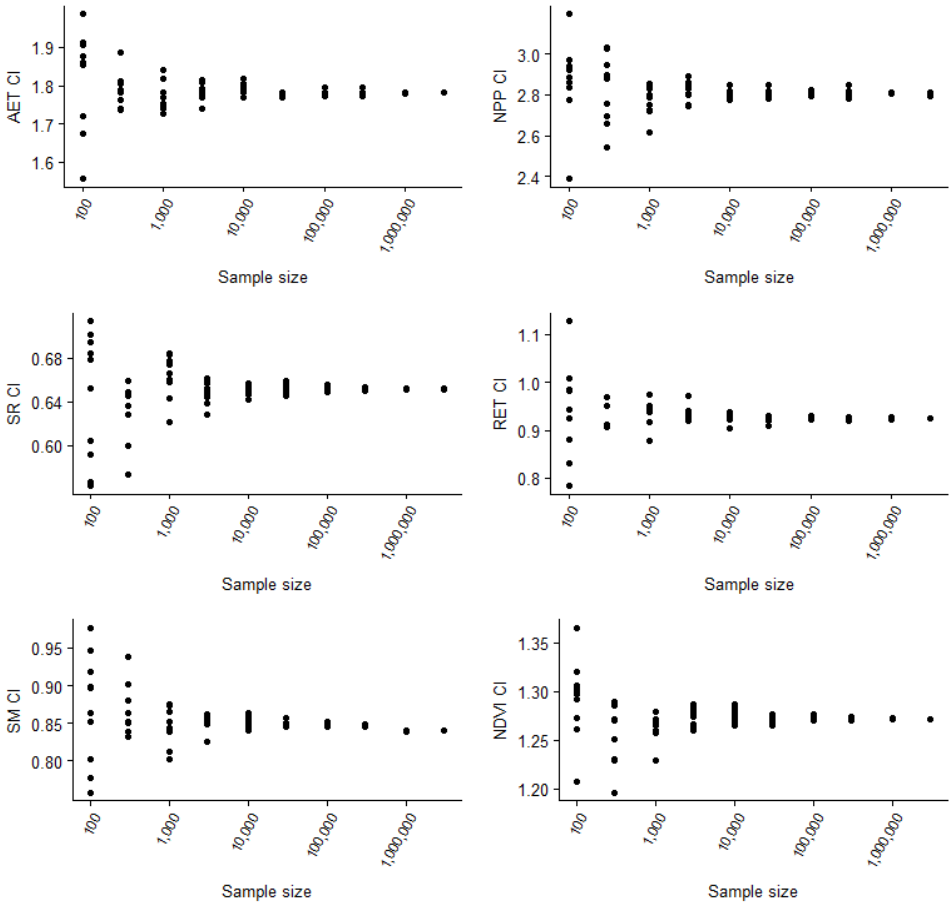

Figure 6.2: CI plotted against sample size for the sampling schedule for each iteration (x10). Note that each plot has a different scale for Y-axis. AET is ETIa and RET is ETo.

The smallest CI range (ΔCIj) occurs at Ni = 3,000,000 for all datasets, which is less than 0.05% for all datasets. The SR, SM and NDVI datasets reached a CI range of less than 5% at Ni = 3000, with CI ranges of 3.34%, 3.57% and 2.51% respectively. The ETIa and ETo datasets both reached a CI range of less than 5% at Ni = 10,000, with CI ranges of 4.63% and 3.41% respectively. The NPP reached a CI range of less than 5% at Ni = 100,000 or 3.16% of total sample size.

The CI range with increasing step size (ΔCIi) is less than 5% at a sample size of 30,000 for NPP. The ΔCIi condition is met at the same sample size that ΔCIj is met for ETIa, SR, SM, NDVI and ETo. Therefore for ΔCIi,j range of 5% is met at Ni =3000 for SR and SM and at Ni = 10,000 for ETIa, ETo and NDVI. While for NPP it is met at Ni = 100,000.

The distribution of the ETIa samples sets Ni=300, Ni=3000 and Ni=30,000, and the differential entropy plotted against the sample size is shown in Figure 6.3. The 1—day ETIa is plotted on the x-axis and the density for a given ETIa is plotted on the y-axis. Each line represents a single sample Nj, for a given sample size, Ni. The variation in the distribution functions is greatest when the sample size is low, Ni=300. The range in the amplitude of both peaks varies, and with high ETIa values. The greatest variation for Ni=3000 is seen at the peaks and troughs. The variation for Ni=30,000 is only visible at the second peak, but the variation is less visible than the smaller sample sizes. The higher sample sizes have a greater convergence. This is represented by the merging distribution functions. While different datasets had a different distribution function, all showed converging distribution functions with increasing sample size.

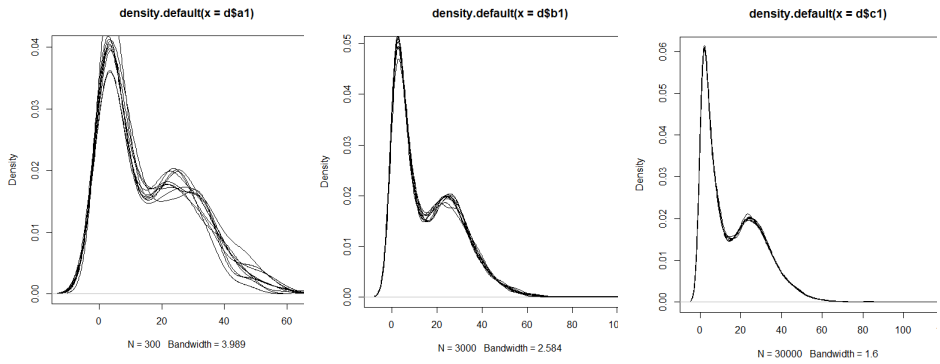

Figure 6.3: Density functions of ETIa sample size sets, a) Ni = 300, b) Ni = 3000 and c) Ni = 30,000.

Figure 6.4 shows the entropy of all samples for all datasets. The entropy values converge with increasing sample size for all datasets. All datasets show increasing entropy with increasing sample size. Entropy values are sometime higher at lower sample sizes as compared to higher sample sizes. However, the values are converging to higher mean values for each increasing sample size. The range of CI values are highest for all datasets at Ni = 100 or Ni = 300. The rate of convergence is greatest when the sample size is low, and decreases with increasing sample size.

There are only 2 occasions where the range in CI increases with an increasing step size; the ETIa entropy range is greater when Ni = 3000 than when Ni = 1000 and the NDVI entropy range is greater at Ni = 1000 than when Ni = 300. The minimum entropy range $\Delta H(x)j$ occurs at Ni = 3,000,000 and is less than 0.003, for all datasets. This is followed by Ni = 1,000,000, where the range is entropy is less than

0.006 for all datasets. At Ni = 300,000, Ni = 100,000, Ni = 30,000 and Ni = 10,000 the $\Delta H(x)j$ is less than 0.04, 0.05, 0.06 and 0.08 respectively. For sample sizes less than Ni = 3000, the entropy ranges are much larger, ranging from ETo $\Delta H(x)j = 0.05$ when Ni = 3000, to NPP $\Delta H(x)j = 0.68$ when Ni = 100. The $\Delta H(x)i \leq 0.05$ performance indicator is met at Ni = 3,000 for ETo, Ni = 10,000 for SR and SM and Ni = 30,000 for NDVI, NPP and ETIa. The $\Delta H(x)j \leq 0.05$ performance indicator is met at Ni = 3,000 for ETo, Ni = 10,000 for SM, SR and ETo, at Ni = 30,000 for NPP and ETIa. Therefore for $\Delta H(x)i,j$ condition is met at Ni =3,000 for ETo, Ni = 10,000 for SR and SM, at Ni = 30,000 for NPP, ETIa and NDVI.

The CI and entropy precision increases for both the CI and the differential entropy with increasing sample size. This is reflected in the increasing density of the cluster for increasing sample sizes. The CI values are converging to a single, central value. Comparatively, the differential entropy values are converging to a higher value. Differential entropy increasing coupled with decreasing variation for increasing sample size conforms to expectations (Figure 6.1).

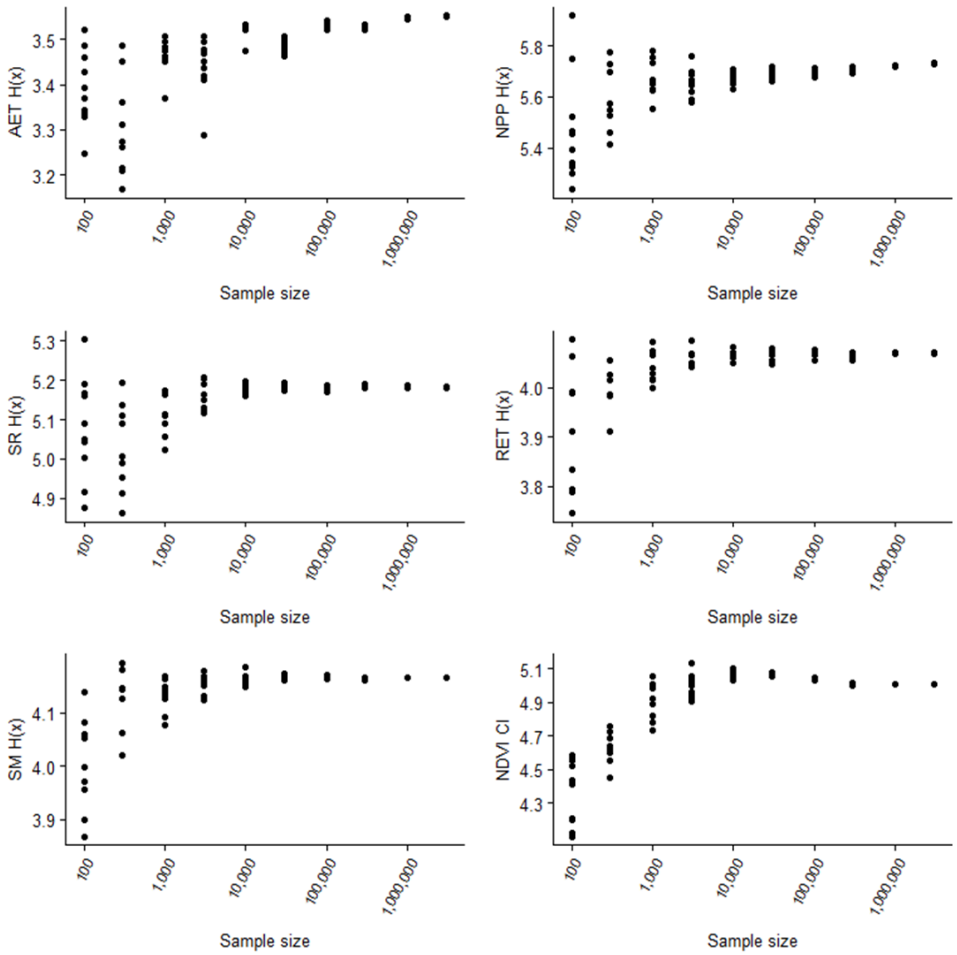

Figure 6.4: Entropy, H(x), plotted against sample size for the sampling schedule for each iteration (x10). Note that each plot has a different scale for Y-axis. AET is ETIa and RET is ETo.

The ETIa dataset was used as example to show the performance of the CI indicator when using for climate classes. The CI range for each climate class and each sample size is shown in Figure 6.5. The classes with the largest CI interval ranges have the deepest saturation of red and the classes with the lowest ranges have the deepest saturation of green. The classes where the sample size was not large enough for any of the samples in the sample set to provide information on the CI are grey. The smallest sample size, Ni = 100, shows the greatest CI variation for all classes, with the CI range frequently exceeding 2 (or 200%). This includes classes with the largest representation, such as the arid desert hot class (BWh). As the sample size increases the CI range decreases for most classes. Exceptions

occur for some sample size increments where the class has a faction of total area of less than 1%. When the sample size reached Ni = 300,000, the CI is below the previous sample size, Ni = 100,000, for all classes. At Ni = 100,000, one class has greater CI than the class before (greater than CIj >0.05). This is for the class with the smallest area, temperate dry warm summer (Csb), and the number of sample points representing this class is still <100. At Ni=3000, all classes had samples to estimate the CI for the entire sample set.

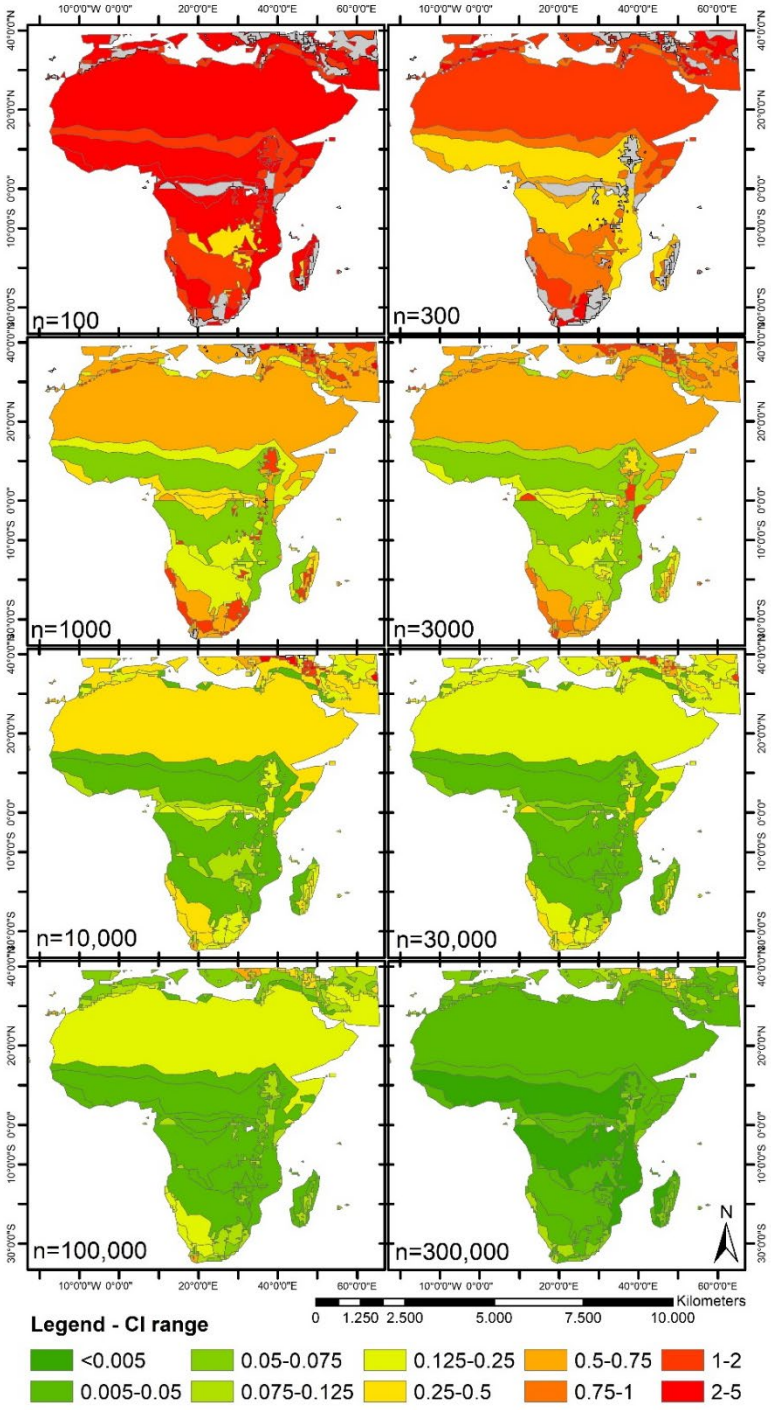

Figure 6.5: The CI range for each sample set for ETIa using Köppen-Geiger climate classes for all sample sizes (Ni).

When the sample size is Ni = 3000 the CI range is ΔCIj <2 for all classes and at 1000, the CI range is ΔCIj >2 on one occasion, cold (continental) dry warm summer (Dsb). When the sample size is Ni = 30,000, one class has a ΔCIj = 1 and all other classes have a range in CI of less than 0.8. Four classes have a range in CI with ΔCIj <0.05 and seven classes with ΔCIj <0.15. When the sample size is Ni=100,000 the maximum CI range is ΔCI=0.51. Four classes have a range in CI with ΔCI <0.05 and 12 have a ΔCI<0.15. When the sample size is Ni=300,000 the maximum CI range is ΔCI = 0.34. Eight classes have a range in CI with ΔCIj<0.05 and 14 (of 16) have a ΔCIj<0.15. The samples that exceed ΔCI<0.15 are from the humid continental classes (cold (continental) dry hot summer (Dsa) and Dsb).

When applying the test to only the major classes - arid, equatorial temperate and humid continental – the same overall CI trend is observed, a decreasing ΔCIj for all major classes with increasing sample size. When the sample size is Ni = 3000, one class has a range in CI of ΔCIj <0.05 (equatorial) and two others have a ΔCIj <0.15 (equatorial and arid). When the sample size is Ni=300,000 the maximum CI range is ΔCIj = 0.34. When the sample size is Ni = 10,000 up to and including Ni=100,000, one class has a range in CI of ΔCI<0.05 (equatorial) and three have a ΔCIj <0.15 (equatorial, arid and warm temperate). When the sample size reaches Ni=300,000, three classes have a range in CI of ΔCIj <0.05 (equatorial, arid and warm temperate). The minimum ΔCIj in humid continental class is ΔCIj =0.19 and occurs at Ni=300,000. The humid continental has the smallest area, <1% of the total area, and therefore has the smallest number of representative sample points.

Figure 6.6 shows the correlation between the ETIa and NPP with increasing sample size. Similar to the CI for each dataset, the correlation shows the most variation at low sample sizes and converges as the sample size increases. This shows that, like CI and H(x), negligible further insight into the relationship is gained beyond a certain sample size. In this case depending on the users preferred margin of error, this is likely to fall between a sample size of 30,000, where correlation ranges from 0.41-.45, and 300,000, where correlation ranges from 0.41-0.43.

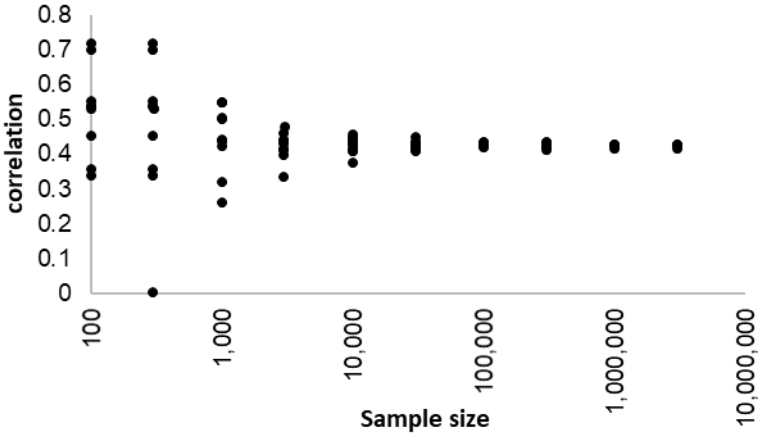

Figure 6.6: Correlation between ETIa and NPP plotted against sample size for the sampling schedule for each iteration (x10).

4. Discussion

In this research, a new and operative methodology was proposed to define a representative sample for arbitrarily large, continuous datasets, using a progressive sampling approach combined with two performance indicators. The purpose being to increase efficiency of validation and quality assessment tasks where the entire dataset has previously been used. The results showed that the assessed datasets in the continent of Africa and the Middle East, without classification of zones, can suitably be represented by a small fraction dataset as the performance condition of both indicators, CI and entropy, was met at Ni = 10,000 for ETo, SM and SR, Ni = 30,000 for ETIa and NDVI and at Ni = 100,000 for NPP. This represents 0.01% for the of the total dataset size of the 250m resolution datasets, and 0.41% of the total dataset size 25km resolution dataset (Table 6.1).

Though no directly comparable study exists, several studies in other fields that use progressive sampling for to increase training efficiency of discrete datasets exist. Six studies that look at a combined 22 different datasets, including land cover type (Lazarevic and Obradovic, 2001; Peng et al. 2004), traffic data (Umarani and Punithavalli, 2011), waveform (Lazarevic and Obradovic, 2001; Ng and Dash, 2006; Peng et al. 2004), simulated data (ElRafey and Wojtusiak, 2017; Umarani and Punithavalli, 2011), wine quality data (ElRafey and Wojtusiak, 2017), with varying number of categories or attributes. The effective sample size was determined by each author and is not related to the indicators selected in this study. Irrelevant of the parameter, the sampling schedule or the number of categories, we found a power relationship between the effective sample size per category among

these studies (Figure 6.7). Although a small pool of data, the power relationship between the total dataset size and the sample size required for effective algorithm training has a good coefficient of determination (R^2=0.76). If this power relation is applied to the dataset size used in this research, as if there is one category, an effective sample size is extrapolated to be 0.2% and 3% of total dataset size for the 250m and 25km resolution datasets respectively, which is similar to our results. However, none of these datasets are arbitrarily large, or continuous and in theory this should the sample size would stop increasing beyond a certain point (Cherkassky et al. 1999). The entropy performance measure accounts for this, as the negligible marginal gain in new information with increasing sample size is reflected in a negligible marginal gain in entropy.

While, the proposed approach should remain effective regardless of different state variable with different spatial-temporal heterogeneity, the determined sample size may vary. This was seen between the six observed datasets. The required sample size should be particularly influenced by the dataset complexity rather than specifically dataset resolution (unless the resolution is increasing dataset complexity). For example, in non-discrete applications of progressive sampling, samples with increasing number of attributes is associated with increasing sample size, as the effective sample size was frequently determined per category.

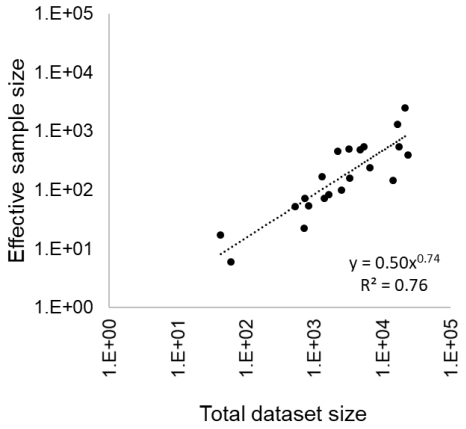

Figure 6.7: Effective sample size per category taken from literature. Note both axis' use log scale.

While the CI indicator can be easily determined based on user performance requirements, differential maximum entropy is less obvious. This study used a nominal requirement of $\Delta H(x) < 0.05$. If the acceptable value decreased to $\Delta H(x) < 0.02$ the required sample size required increases significantly. For example, the ETIa sample size

requirement becomes Ni = 300,000 as compared to Ni = 30,000. Alternatively, if it is increased to $\Delta H(x) < 0.10$ the acceptable sample size is Ni = 1000 for this indicator (although it would still need to meet the CI indicator requirement). This not only highlights the importance of selecting an acceptable entropy increment suitable to user application, but also highlights the sensitivity of the entropy indicator.

The sample size required to reach the acceptable performance increased once climate classification was introduced. When the sample size reached 300,000 only 8 reached the acceptable ΔCI. A sample size of 30,000, achieved the acceptable performance for only 4 and 7 classes respectively. This suggests that the sample size included in the sampling schedule does not meet the PCE criterion when considering all classes. Similarly, the PCE criterion was not met when only the major climate classes were considered. The smallest classes, humid continental classes, did not meet the acceptable error for the entire sampling schedule. This was a result of the small representation of humid continental class. Rather than increase the sample size, it may be more appropriate to use methods such as a stratified random sample, or a progressive boosting to optimize sample size and account for imbalanced data (Lazarevic and Obradovic, 2001; Soleymani et al. 2018). This would align with approaches used in land cover classification where a minimum sample size per class is often defined (EFTAS and FAO, 2015). This highlights the importance of selecting a sampling approach that suits the user needs. For studies focused on the Middle East and Africa, a lower confidence in the humid continental class, which is located predominately in Europe, may be acceptable.

While the approach has only been tested on the region of Africa and the Middle East to climate data, it is possible to extend the approach to other regions or for other categories. While the convergence point for both indicators is expected to change based on data distribution and on how the data is categorized (e.g, climate of land cover type), the progressive sampling approach should still be valid, and is expected though the sampling size may increase based on data complexity resulting from increased environment, land cover climate, topography types etc. In the case of added complexity or categories a complexity measure, as applied with entropy to soil moisture retrievals (Kumar et al. 2018), may also by useful.

This study used spatially continuous datasets (no data gaps). In cases where data gaps exist the sampling schedule should be applied to pixels where data exists (exclude no data), or the sample size will increase, however, dependent on dataset, this will result in areas with large data gaps missing from the evaluation.

The two indicators selected cover both the mean and standard deviations, through the CI of the dataset, and the distribution of the dataset through the differential entropy. They are useful indicators in

the absence of true values, which is relevant for large spatial datasets with little ground-truth information available. These indicators are therefore not intended to define the accuracy of the dataset itself, but how accurately the sample represents the dataset. The condition of both indicators should be met when defining the suitable sample size. This was seen in this example when the CI acceptable criterion was met before the differential entropy acceptable was met. The usefulness of the approach is that the performance criterion (i.e. CI and entropy) can be defined dependent on the application and therefore accuracy requirement of the user.

The selected indicators selected are not suitable for all continuous datasets, as not all data behaves as a continuous dataset. For example, although precipitation is a continuous variable, the number of zero rain points will cause the dataset to behave as a discrete-continuous distribution (Friederichs and Hense, 2007). For datasets that behave discretely, it may be useful to use the two-step approach that is commonly used in precipitation validation, whereby the product is first validated categorically and then quantitatively (Wilks, 2006). Though, this shows limitations as for precipitation this will no longer represent dry or arid regions well.

Finally, a similar approach could, in theory, be applied to determining sample size for comparison against ground-truth measurements. Many limitations are imposed on ground-truth data collection, such as sample design (due to access constraints) and resource limitations. However, it could guide a best practice or an or objective for a community, for example fluxnet, where data collection is less intensive and can be collected without experts, for example citizen science.

5. Conclusion

The progressive sampling approach combined with CI and differential maximum entropy performance measures, can be applied to determine a suitable sample size for physical consistency and cross-comparison tests of a continuous, arbitrarily large remote sensing-based datasets. The approach showed that the amount of data required to represent large datasets (1.22E+09 pixels) datasets is comparatively small (10,000-100,000 pixels) and therefore can significantly reduce computing time and resources. This can be used to run initial tests or product analysis. It is suggested that using a representative sample, rather than the whole dataset, can effectively balance insight to the quality of the dataset and reduce processing efforts required in validation procedures, such as continental cross-comparisons, that are computationally exhaustive. This will become even more useful as dataset resolution increases.

Chapter 7

A reflection on the validity of the crop water productivity indicator concept in context of the SDGs

1. Relevant SDGs and their intentions

The United Nations (UN) Sustainable Development Goals (SDG) are a global agenda for sustainable development (UN, 2016). They contain both 'Outcome' (circumstances to be attained) and 'Means of Implementation' (MoI) targets (Bartram 2018). MoI targets are to provide a way on how the SDGs can be attained. Each goal has at least one accompanying indicator used to monitor, report and measure the performance. SDG6 specifically sets goals for sustainable water consumption:

SDG6 – 'Ensure availability and sustainable management of water and sanitation' – and it contains six outcome targets (SDG6.1-SDG6.6) and 2 MoI targets SDG6A-SDG6B).

SDG6.4 is defined as "By 2030, substantially increase water use efficiency across all sectors and ensure sustainable withdrawals and supply of freshwater to address water scarcity and substantially reduce the number of people suffering from water scarcity". SDG6.4 is accompanied by two indicators:

- 6.4.1: "Change in water-use efficiency over time"
- 6.4.2: "Level of water stress: freshwater withdrawal as a proportion of available freshwater resources"

SDG6.4 may contradict with other SDG Outcome targets and therefore the goal of SDG6 itself, by focusing on efficiency (SDG6.4.1) rather than total freshwater volume. Namely, it may jeopardise: SDG6.1 – By 2030, achieve equitable access to safe and affordable drinking water for all; SDG6.6 – By 2020, protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers and lake; as well as SDG6.4 itself.

Further to consider is SDG2 – 'End hunger, achieve food security and improved nutrition and promote sustainable agriculture'. Specific to CWP, are SDG2.3 and SDG2.4 which are focused on doubling agricultural productivity and ensuring sustainable food production systems respectively.

It may therefore be necessary to strike a balance between water consumption and food production. This chapter aims to discuss the value and limitations of water productivity in the context of SDGs. It aims to consider how valid is the water productivity concept and how it might or not jeopardise/conflict with reaching other SDG goals.

2. Why water productivity? – intentions, benefits and limitations of the indicator

In agriculture, the SDG indicator SDG6.4.1 term 'water use efficiency' is frequently interpreted as referring to CWP. It is generally adopted as a necessary part of the paradigm of sustainable intensification of agriculture (Rockström et al. 2017; Brauman, Siebert, and Foley 2013). This is evident in the UN FAO WaPOR project - the FAO portal to monitor WAter Productivity through Open access of Remotely sensed derived data. The project aims to monitor CWP and water and food security in Africa and the Middle East and also to identify and reduce CWP gapsand contribute to the sustainable increase of agricultural production. The definition given by FAO on water productivity in agriculture (i.e., CWP) is the value of output in relation to the quantity of water beneficially consumed to produce this output (FAO, 2020). There are several benefits to this definition:

- It provides a clear and tangible definition – similar to the 'crop per drop' approach however, more open to variation in the definition of beneficial output (compared to yield only).
- It focuses purely on the physically attainable optimum, though crop production is governed only by transpiration, evaporation, and is related to crop specific growth and management practices (Molden et al. 2010).
- It acknowledges that water that does not leave the system through evapotranspiration can have other beneficial uses (Molden and Sakthivadivel 1999).
- It can be monitored and measured through remote sensing techniques which covers more expansive range and capabilities as compared to traditional in-situ methods of measuring irrigated efficiencies Blatchford et al. (2019).

Several authors (Booker and Trees 2020; Wilchens et al. 2015; Molden et al. 2010; Zoebl et al. 2006) have discussed the limitations and potential dangers of CWP. Though this indicator can be a powerful tool, the caution other authors suggest should be heeded, and these limitations include:

- The transpiration ratio determines the physical attainability and therefore limits in scope for improvement (Molden et al. , 2010).
- Social, economic limitations which inhibit certain growers and farmers ability to improve CWP (Wichelns 2015) as well as the incentive for farmers to improve water use efficiency rather than increase yield.

- Potential to increase actual water consumption due to marginal water productivity decreases with increasing water supply (Zoebl 2006).
- High CWP even though very low yield/production (due to a non-linear relation between transpiration and carbon assimilation at very low T) and gain decrease with increasing productivity or yield (Blatchford et al. 2018).

3. Potential contradictions in the CWP indicator in the context of SDGs

Large studies suggest WP is not necessarily higher or lower in irrigated or rainfed croplands. It has been shown that typically average CWP values will be higher for certain crops under irrigated and others under rainfed conditions (Brauman, Siebert, and Foley 2013; Mekonnen and Hoekstra 2011). Figure 7.1 shows the average 10-year [2009-2018] water productivity, as a function of dry matter productivity (DMP) for irrigated and rainfed crops in Africa and the Middle East. The lower (orange) and upper (blue) boundary functions were estimated using Blatchford et al. (2018) and are taken as functions of the 5th percentile WP and the 95th percentile WP for each 5 ton ha^{-1} DMP increment increase. It should be noted that the DMP-CWP does not account for yield conversion factors or LUE conversion factors. The rainfed and irrigated DMP and WP have similar ranges, however, they have different distributions. At lower DMP, the rainfed croplands are achieving higher upper thresholds of WP. The rainfed croplands also has a steeper lower boundary function, suggesting irrigated croplands are more frequently underperforming, or crops that are irrigated frequently have lower WP. The rainfed croplands WP has two distribution frequency peaks at 1 kg m^{-3} and 2.2 kg m^{-3}, while the irrigated croplands have two at 0.5 kg m^{-3} and at 1.7 kg m^{-3}. However, conversely, for DMP, irrigated croplands have high DMP productivity with the peak distribution frequency at 7-8 ton ha^{-1} compared to 2-3 ton ha^{-1} in rainfed croplands. This highlights that despite much lower dry matter production in rainfed croplands as compared to irrigated croplands, the WP is frequently higher in rainfed croplands, suggesting a trade-off is required between water productivity and land productivity – or SDG6.4.1 and SDG2.

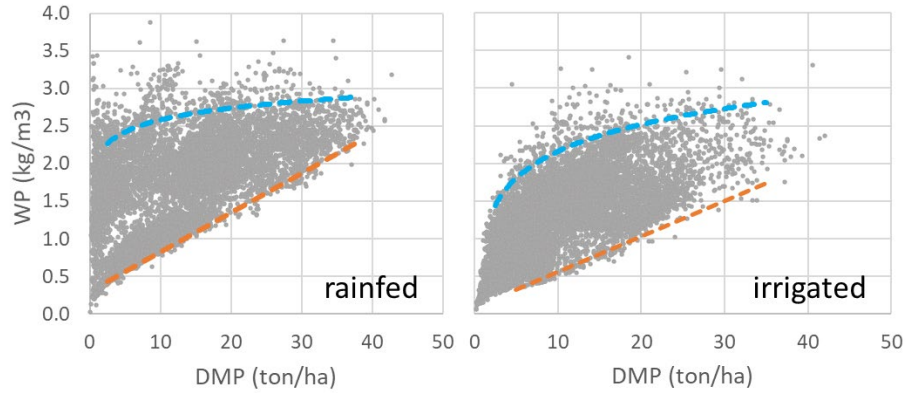

Figure 7.1: WP plotted against DMP in Africa and the Middle East for rainfed and irrigated crops. The WP [WP (kgDMP m^{-3}) = DMP (kg ha^{-1} $year^{-1}$) /ETIa (mm $year^{-1}$) /10] was derived from the NPP and ETIa datasets on the WaPOR portal. The data was extracted for a random stratified sample of 30,000 points for each of the irrigated [class 41] and rainfed cropland classes [class 42] as per the WaPOR land classification map.

Brauman et al. (2013) stipulated that food production can be increased for a given amount of water consumption, or that food production can be maintained while decreasing water consumption. The latter is used as a base for defining yield zones to improve CWP by Bastiaanssen and Steduto (2017) and Blatchford et al. (2018). The purpose of this approach is to target increases in yield while maintaining ETa. Practices that reduce the evaporation component, or target irrigation timing (i.e. precision irrigation), rather than increasing water consumed time, so it is consumed at times to maximise production – flowering etc., or on planting date (e.g. Blatchford et al. 2018). However, due to the near linear relationship between yield and transpiration for a healthy crop, increasing food production is likely to actually increase water consumption (not maintain) for a given crop, which may therefore have indicators SDG6.4.1 and SDG6.4.2 at odds.

Brauman et al. (2013) showed that variation in CWP was not solely a function of climate, however, did show that each crop had its highest CWP values in a specific climate zones (and as much variability from non-climate drivers as from climate). They state that increase in CWP would result in areduction of irrigation water requirement in more arid climates, but not necessarily in water savings.

"*Though the full impact of increasing water productivity of irrigated crops would not be gained in irrigation savings, it would result*

in a lower irrigation water requirement, particularly in more arid climates." Brauman, Siebert, and Foley (2013)

Based on this, we considered WP by aridity for irrigated and rainfed croplands in Africa and the Middle East. Figure 7.2 shows the 95th percentile DMP and DMP WP for aridity increments of 0.1 (i.e. for bands 0-0.1, 0.1-0.2 etc.) for rainfed and irrigated croplands. The 95th percentile would be considered a target or bright spot – a high performing area (Cai et al. 2011). The rainfed croplands have high WP for all aridity increments with only small gains as aridity decreases. However, despite decreasing marginal gains in 95th percentile DMP after the aridity index gets >0.65 (where climate is classed as humid), the DMP shows clear gains with decreasing aridity until available energy is equal to available water (PET/PCP=1). This suggests that for a given aridity, a high WP, may be a result of a water stressed crop with low DMP as water use efficiency is higher when canopies are water-stressed (Fuchs 1975 – in Zoebl 2006). Therefore, these areas of high WP in highly arid areas may not reflect desirable management practices or goals given very low land productivity and therefore contradict SDG2.4.

Irrigation in drylands can decrease marginal gains in water use efficiencies (i.e. higher under rainfed) (Zoebl 2006). The irrigated croplands have lower 95th percentile WP at low aridity then rainfed areas, however, they have higher DMP or land productivity. In the irrigated croplands the steepness of the curve of increased DMP and WP are similar when aridity is <0.5. Aridity index between 0.5-0.8 shows a 95th percentile DMP increase, while 95th percentile WP is decreasing. The 95th percentile DMP begins decreasing when the aridity index >0.9. This may be the point of diminishing returns, or the point when gains in yield comes at the expense of an efficient input use (Zoebl 2006).

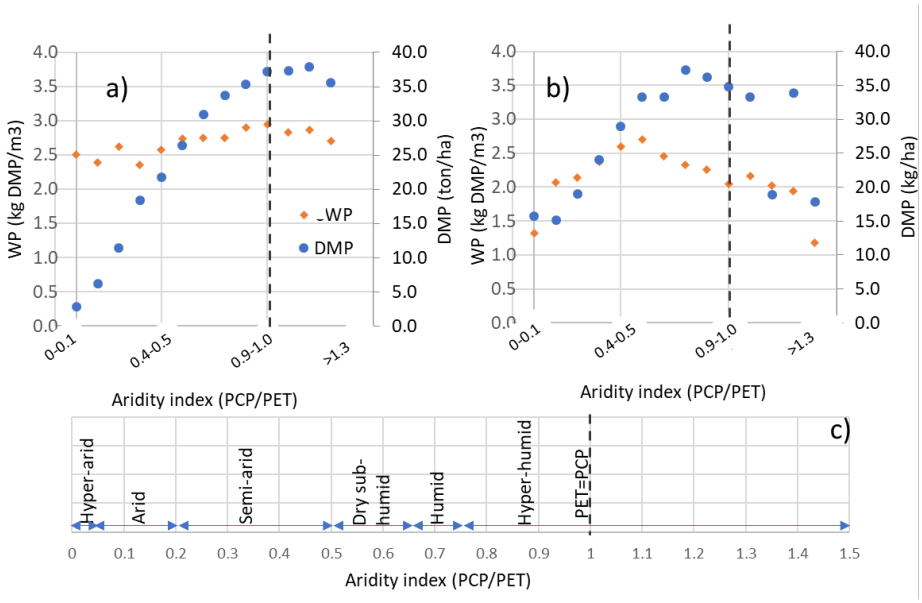

Figure 7.2: 95th percentile WP (same dataset as Fig. 1) and 95th percentile DMP plotted against aridity zones for a) rainfed crop lands and b) irrigated crop lands. c) shows the climate classification for a given aridity index range as defined by UNEP (1992). The dashed line represents where PET=PCP. The aridity index is taken from the "Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2"[5]

Bastiaanssen and Steduto (2017) and Blatchford et al. (2018) attempted to apply a correction factor for climate as a means to set global CWP targets. The CWP is then scaled based on the ratio of actual reference evapotranspiration to a global average evapotranspiration. This linear adjustment may not be suitable considering Figure 7.2. Reference evapotranspiration should decrease with decreasing aridity, and, particularly in irrigated agriculture, aridity is showing a clear optimum performance at aridity in the range or 0.3-0.7 and applying a linear adjustment factor would not create a 95th percentile slope = 0 in irrigated areas (though it may in the rainfed croplands given the 95th percentile WP trend in rainfed areas). Further, this approach does not account for very low DMP. Rather, it attempts to equalise WP across aridity zones and essentially lifts the WP in these zones despite very low DMP, which may not be a suitable target if food production and irrigation water are available in that area. Figure 7.2 would benefit by

[5]https://figshare.com/articles/Global_Aridity_Index_and_Potential_Evapotranspiration_ET0_Climate_Database_v2/7504448/3

further considering CWP by aridity and to distinguish in very arid areas are associated with lower yielding crops.

4. Conclusion

Increasing CWP, or setting targets of CWP bright spots, or upper percentiles may result in unintended consequences. Selecting bright spots in rainfed arid zones may be setting targets at odds or in contradiction to be able to achieve increased yields or land productivity (SDG2.3). Setting targets or bright spots to maximise CWP in less arid zones in irrigated croplands may also be at a cost of higher yields. Further, increasing CWP may also lead to an increase in water consumption, in contradiction to the indicator – reduce water stress (SGD6.4.2).

The CWP concept can be a useful one, however, it can also be misleading. On both a local, regional and global scale, users must reflect on the goals and select or prioritise indicators, improving CWP may not always be the most useful indicator. We need to focus on the sustainable improvement in CWP rather than the increase in CWP for sustainable agriculture (SDG2.4).

Chapter 8
Concluding remarks and prospects

Monitoring CWP can be a useful indicator and way to better understand and optimize water use consumption in agriculture. Remote sensing of CWP provides a promising way to monitor and evaluate CWP over extensive areas, currently achievable at appropriate spatial and temporal resolutions. Global and continental evapotranspiration and primary productivity remote sensing-based datasets are becoming increasingly available at higher resolutions. This provides opportunity to monitor CWP in regions previously difficult to access and assess. However, mischaracterization of the quality of the dataset can be unsafe as it may result in misguided application or analysis. It is important that users understand the quality of the data, and proceed with caution when using these datasets in decision making.

1. Conclusions and implications

1.1 Summary of conclusions

There are three components to understanding the quality of a dataset: (i) understanding the target or benchmark, (ii) defining a process or strategy of evaluation and (iii) executing the evaluation strategy and determining if the dataset meets the target.

The key to understanding the suitability of CWP datasets is to first understand the benchmark – or goal accuracy. Only then the user can relate the actual accuracy to a target. A tangible target is to meet the standards of in-situ measurements of crop yield and water consumption. In Chapter 3 the quality of large CWP datasets were compared to the quality of in-situ measurement. The review revealed that remote sensing can estimate CWP within the error range of in-situ methods. However, the review also revealed a great deal of heterogeneity in the reporting of both errors and uncertainty. The characterisation of the errors, e.g. random error or systematic bias, will largely define if the data products are suitable or not for different applications of CWP.

Chapters 4, 5, and 6, define and execute an evaluation strategy for an example database – the UN FAO WaPOR database. Chapter 4 presents a multi-faceted approach to evaluate a large evapotranspiration dataset with little in-situ data for verification. The approach considered the multiple available resolutions and scales of the database to define quality at multiple scales. The database showed reasonable accuracy with errors ranging from 26.3%-40.4%. This fell within the typical range of evapotranspiration estimates from large remote sensing studies (25-60%) and within the upper threshold of accuracy gained with in-situ measurements (5-30%).

The implication of dataset spatial resolution on CWP application (use in irrigation performance) was considered in Chapter 5. A simple study comparing 30m, 100m and 250m resolution datasets showed that while each dataset had consistent temporal trends, the magnitude of the trend in both space and time is smoothed by the 100m and 250m resolution datasets. This is frequently resulting in large differences in the irrigation performance assessment criteria for inter-plot comparisons. This is particularly noticed in irrigation schemes with the smallest field sizes. This highlights the importance of selecting the spatial resolution appropriate and related to irrigation scheme characteristics when undertaking irrigation performance assessment with (using) remote sensing input. It also shows that even within one database the quality of database varies for the same product across resolutions depending the application.

Analysing such large databases can face many challenges, including processing challenges. To improve the efficiency of the evaluation approach, a method to estimate a representative sample size, dependent on evaluation characteristics, was developed. Chapter 6, highlighted that a simple progressive sampling approach, combined with an accuracy focused indicator –confidence interval, and a distribution indicator – maximum probability entropy, was able to significantly reduce the sample size required for certain validation procedures. Relevant procedures in the approach used in Chapter 4 include cross-comparison and spatial and temporal consistency. A sample size of 10,000-30,000 (representing <0.01% of the total database size) was determined suitable to represent quality of a continental dataset.

Finally, while reflecting on the validity of the concept of CWP, we note that the concept can be a useful one, however, it can also be misleading. On both a local, regional and global scale, users must reflect on the goals and select or prioritise indicators, improving CWP may not always be the most useful indicator. We need to focus on the sustainable improvement in CWP rather than the increase in CWP for sustainable agriculture.

1.2 Implications

The research in this thesis is expected to be useful and play an significant role for the evaluation and interpretation of remote sensing-based dataset quality, particularly for applications in agriculture. The results can guide both academics and real-life agricultural application end-users in using remote sensing-based evapotranspiration, yield and CWP datasets.

By reviewing the current literature on evapotranspiration, yield and CWP accuracy by in-situ and remote sensing methods, users can

easily associate the quality of remote sensing-based products with best ground-truth observations. As ground-truth is the standard in agriculture, but has quite a range of uncertainty, Chapter 3, was able to show that large scale remote sensing-based observations, can provide similar user confidence. However, it was also able to show that the error range of both in-situ and remote sensing-based estimates is significant and may still not yet be suitable for several applications. This can help academic users when incorporating such datasets in analysis and models, and can also help professional end-users from the decision-maker group, i.e. irrigation scheme managers or policy makers.

Chapter 4 provides an exhaustive validation approach for evaluating evapotranspiration over a large and in-situ data scarce continent, Africa. This chapter provides a methodology that be applied to other products in regions with limited data availability. The approach in Chapter 4 has also been applied to NPP and CWP products in the UN FAO WaPOR validation V2.0 report (FAO 2020c). This will be used to inform users of product quality and strengths and weaknesses of data dependent on climate and land cover attributes. Further, it is one of the first scientific publications with a focus on validation of remote sensing-based evapotranspiration for the full extent of Africa and it is currently still the highest resolution dataset available for the full extent of Africa. Understanding the quality of this dataset will better inform FAO policy, project activities and research and may have positive influence at country, basin and scheme level decision-making on agricultural water use and crop production.

A representative sample was used for some of the validation activities in Chapter 4. In Chapter 5, the impact of varying resolutions on the interpretation of CWP and other irrigation performance indicators were compared. This highlighted the influence of spatial scale when interpreting CWP data and outcomes. This not only provides users with a guide to the resolution needed based on farm plot size, but informs academics on the threats of performing and drawing conclusions from spatial analysis with a dataset that is too coarse to identify e.g. plot-to-plot variation.

Chapter 6 described the determination of representative sample size. The proposed approach can significantly reduce the processing time while still providing a statistically valid representation of a large remote sensing dataset. This can be useful with any very large dataset and may become particularly useful as more high-resolution remote sensing data becomes available.

Finally, Chapter 7 reflects on the validity of the concept of CWP itself and its limitations. This is important because although CWP can be a useful indicator, it may prove misleading if used as a stand-alone indicator.

2. Further challenges on the way ahead

Understanding the interaction between water, energy and mass provides interesting parallels with the concept of CWP, which is de facto a function of mass production in terms of carbon gains and water consumed through transpiration. The importance of water and carbon interactions on the field have long been recognised in agriculture. This is seen through significant research into yield and transpiration relationships (Sinclair et al. 1984), which is the relationship that founds the WP concept. Linking these concepts to energy has also been long recognised in either the biomass production or evapotranspiration studies. (Monteith 1977; 1972) related the plant biomass proportionally to the photosynthetically active radiation (PAR) over the course of the growing cycle (Lobell 2013; Ahmed et al. 2010). Both field measurements and remote sensing link evapotranspiration or water consumed to the energy balance, where the latent energy is equivalent to the energy consumed in thermodynamic phase change (liquid to vapor) or evapotranspiration. However, in both remote sensing and agriculture, the link between all three has not been explored in depth. Understanding the physical interactions and partitioning of energy into mass (carbon) and water, may offer the development of useful empirical relationships to observe the spatial and temporal consistency in agricultural CWP component products.

This relationship could be explored at different scales for different agro-ecological zones, e.g. catchment or field scale, as well as determining if such consistency can be used in the validation of CWP datasets. Also, it does not yet include the complexity of considering irrigated areas, where the evapotranspiration is often larger than the precipitation for a given field. So yet to be explored is this energy and mass balance partitioning in agriculture, particularly over irrigated lands.

Another interesting tool and way forward for CWP validation may be by utilizing vegetation indices. In a given vegetation index or VI range, both the NPP and evapotranspiration can be estimated through VI methods (Nagler et al. 2013; Glenn et al. 2011; Mu et al., 2007). Currently each of these datasets are estimated independently, but to assure consistency between the datasets the two can be empirically related, given that biomass production has been described as having a positive and near linear relationship with transpiration. While these relationships are empirical and often specific to a local climate, they could be useful validation tools in places where the relationship is known, as for example in the Lower Colorado River Basin (Nagler et al. 2009). They can also be used as a tool, where ground

data exists to develop the relationship and extrapolate to a larger region for comparison. In theory, this would allow ground observations of NPP or evapotranspiration (rather than both) to validate CWP.

In conclusion, while it is suitable to apply large remote sensing-based datasets for the monitoring and evaluating of CWP, there is more still to be explored with regard to carbon-water-energy links, that may enhance both validation and physical understanding of CWP. With further work, this link may be able to better utilize existing ground data and further increase data quality of agricultural water productivity datasets.

Bibliography

Abernethy, C.L., 1990. Indicators of the performance of irrigation water distribution systems, in: Symposium on the Performance Evaluation of Irrigation Systems, 23rd Nov. International Irrigation Management Institute (IIMI), Colombo, pp. 22. http://publications.iwmi.org/pdf/h_7823.pdf

Ahmed, B.M., Tanakamaru, H., Tada, A., 2010. Application of remote sensing for estimating crop water requirements, yield and water productivity of wheat in the Gezira Scheme. Int. J. Remote Sens. 31, 4281–4294. https://doi.org/10.1080/01431160903246733

Al Zayed, I.S., Elagib, N.A., Ribbe, L., Heinrich, J., 2016. Satellite-based evapotranspiration over Gezira Irrigation Scheme, Sudan: A comparative study. Agric. Water Manag. 177, 66–76. https://doi.org/10.1016/j.agwat.2016.06.027

Albarakat, R., Lakshmi, V. 2019. Comparison of normalized difference vegetation index derived from landsat, MODIS, and AVHRR for the mesopotamian marshes between 2002 and 2018. Remote Sens. 11. doi:10.3390/rs11101245.

Alexandratos, N., Bruinsma, J. 2013. World Agriculture towards 2015/2030: The 2012 Revision. ESA Working Paper No. 12-03. Vol. 20. ROME. https://doi.org/10.1016/S0264-8377(03)00047-4.

Alexandridis, T., Asif, S., Ali, S. 1999. Water performance indicators using satellite imagery for the Fordwah Eastern Sadiqia (South). Irrigation and Drainage Project.

Ali, M. H., Talukder, M.S.U. 2008. Increasing Water Productivity in Crop Production-A Synthesis. Agric. Water Manag. 95 (11): 1201–13. https://doi.org/10.1016/j.agwat.2008.06.008.

Allen, R.G., Smith, M., Pruitt, W.O., Pereira, L.S., 1996. Modifications to the FAO crop coefficient approach, in: Proceedings of the ASAE International Conference on Evapotranspiration and Irrigation Scheduling 3–6 Nov. San Antonio, pp. 124-132.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - guidelines for computing crop water requirements. FAO Irrigation and Drain Paper 56. FAO, Rome.

Allen, R.G., Tasumi, M., Trezza, R., 2007. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model. J. Irrig. Drain. Eng. 133, 380-394. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380)

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agric. Water Manag. 98, 899–920. https://doi.org/10.1016/j.agwat.2010.12.015

Anderson, M.C., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., D'Urso, G., Pimstein, A., Gao, F., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. Hydrol. Earth Syst. Sci. 15, 223–239. https://doi.org/10.5194/hess-15-223-2011

Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P. 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. Remote Sens. Environ. 122, 50–65. https://doi.org/10.1016/j.rse.2011.08.025Anderson, R.G., Alfieri, J.G., Tirado-Corbalá, R., Gartung, J., McKee, L.G., Prueger, J.H., Wang, D., Ayars, J.E., Kustas, W.P., 2017. Assessing FAO-56 dual crop coefficients using eddy covariance flux partitioning. Agric. Water Manag. 179, 92–102. https://doi.org/10.1016/j.agwat.2016.07.027

Ardö, J., Mölder, M., El-Tahir, B. A, Elkhidir, H. A. M. 2008. Seasonal variation of carbon fluxes in a sparse savanna in semi arid Sudan. Carbon Balance and Manag. 3 (7): 1–18. https://doi.org/10.1186/1750-0680-3-7

Ardö, J. 2015. Comparison between Remote Sensing and a Dynamic Vegetation Model for Estimating Terrestrial Primary Production of Africa. Car Bal. and Manag. 10 (1): 8. https://doi.org/10.1186/s13021-015-0018-5.

Arora, V. K., Singh, C.B., Sidhu, A.S., and Thind, S.S. 2011. Irrigation, Tillage and Mulching Effects on Soybean Yield and Water Productivity in Relation to Soil Texture. Agric. Water Manag. 98 (4): 563–68. https://doi.org/10.1016/j.agwat.2010.10.004.

Asrar, G., Myneni, R.B., Choudhury, B.J., 1992. Spatial heterogeneity in vegetation canopies and remote sensing of absorbed photosynthetically active radiation: A modeling study. Remote Sens. Environ. 41, 85–103. https://doi.org/10.1016/0034-4257(92)90070-Z

Atta, Y., Gaafar, I., Hassan, W., El, A. 2015. Validation of Accurate Determination of Maize Water Requirements in Nile Delta. In International Commission on Irrigation and Drainage 26th Euro-Mediterranean Regional Conference and Workshops, Innovate to Improve Irrigation Performances, (pp. 4). Montpellier, France. Available at: https://icid2015.sciencesconf.org/65471/document

Awulachew, S.B., Ayana, M., 2011. Performance of irrigation: An assessment at different scales in Ethiopia. Exp. Agric. 47, 57–69. https://doi.org/10.1017/S0014479710000955

Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. Remote Sens. Environ. 202, 129–141. https://doi.org/10.1016/j.rse.2017.04.014

Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Glob. Chang. Biol. 9, 479–492. https://doi.org/10.1046/j.1365-2486.2003.00629.x

Bartram, J. 2018. Policy Review of the Means of Implementation Targets and Indicators for the Sustainable Development Goal for Water and Sanitation. npj Clean Water, no. December 2017: 1–5. https://doi.org/10.1038/s41545-018-0003-0.

Bastiaanssen, W.G.M., 1998. Remote Sensing in Water Resources Management: The State of the Art. International Water Management Institute (IWMI), Colombo. http://publications.iwmi.org/pdf/H022865.pdf

Bastiaanssen, W.G.M., Bos, M.G. 1999. Irrigation performance indicators based on remotely sensed data: a review of literature. Irrig. Drain. Syst. 13, 291–311. https://doi.org/10.1023/A:1006355315251

Bastiaanssen, W.G.M., Bandara, K. 2001. Evaporative depletion assessments for irrigated watersheds in Sri Lanka. Irrig. Sci. 21, 1-5. https://doi.org/10.1007/s002710100046

Bastiaanssen, W.G.M., Ali, S., 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. Agric. Ecosyst. Environ. 94, 321–340. https://doi.org/10.1016/S0167-8809(02)00034-8

Bastiaanssen, W.G.M., Steduto, P., 2017. The water productivity score (WPS) at global and regional level: Methodology and first results from remote sensing measurements of wheat, rice and maize. Sci. Total Environ. 575, 595–611. https://doi.org/10.1016/j.scitotenv.2016.09.032

Bastiaanssen, W.G.M., Van Der Wal, T., Visser, T.N.M. 1996. Diagnosis of regional evaporation by remote sensing to support irrigation performance assessment. Irrig. Drain. Syst. 10, 1–23. https://doi.org/10.1007/BF01102762

Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1.Formulation. J. Hydrol. 212–213, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4

Bastiaanssen, W.G.M., Thiruvengadachari, S., Sakthivadivel, R., Molden, D.J., 1999. Satellite remote sensing for estimating productivities of land and water. Int. J. Water Resour. Dev. 15, 181–194. https://doi.org/10.1080/07900629949005

Bastiaanssen, W.G.M., Brito, R.A.L., Bos, M.G., Souza, R.A., Cavalcanti, E.B., Bakker, M.M. 2001. Low cost satellite data for monthly irrigation performance monitoring: Benchmarks from Nilo Coelho, Brazil. Irrig. Drain. Syst. 15, 53-79. https://doi.org/10.1023/A:1017967021198

Bastiaanssen, W.G.M., Zwart, S.J., Pelgrum, H., 2003. Remote sensing analysis, in: Van Dam, J.C., Malik, R.S. (Eds.), Water Productivity of Irrigated Crops in Sirsa District, India. Integration of Remote Sensing, Crop and Soil Models and Geographical Information Systems. AGRIS, Wageningen, pp. 85–96.

Bastiaanssen, W. G. M., Cheema, M. J M., Immerzeel, W. W., Miltenburg, I. J., Pelgrum, H., 2011. Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model, Water Resource. Res. 48: W11512. https://doi.org/10.1029/2011WR010482

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. Remote Sens. Environ. 184, 668–681. https://doi.org/10.1016/j.rse.2016.07.030

Bausch, W.C., Neale, C.M.U., 1987. Crop Coefficients Derived From Reflected Canopy Radiation: a Concept. Trans. ASAE. Eng. 30, 0703–0709. https://doi.org/10.13031/2013.30463

Beck, H. E., De Roo, A., Van Dijk A. I. J. M. 2015. Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*. J. Hydrometeo. 16: 1478-1501. https://doi.org/10.1175/JHM-D-14-0155.1

Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens. Environ. 114, 1312–1323. https://doi.org/10.1016/j.rse.2010.01.010

Beyrich, F., Bange, J., Hartogensis, O.K., Raasch, S., Braam, M., Van Dinther, D., Gräf, D., Van Kesteren, B., Van den Kroonenberg, A.C., Maronga, B., Martin, S., Moene, A.F., 2012. Towards a Validation of Scintillometer Measurements: The LITFASS-2009 Experiment. Boundary-Layer Meteorol. 144, 83–112. https://doi.org/10.1007/s10546-012-9715-8

Bhattarai, N., Shaw, S.B., Quackenbush, L.J., Im, J., Niraula, R., 2016. Evaluating five remote sensing based single-source surface energy balance models for estimating daily evapotranspiration in a humid subtropical climate. Int. J. Appl. Earth Obs. Geoinf. 49, 75–86. https://doi.org/10.1016/j.jag.2016.01.010

Bierhuizen, J., and Slayer. R. 1965. Effect of Atmospheric Concentration of Water Vapour and CO2 in Determining Transpiration-Photosynthesis Relationships of Cotton Leaves. Agric. Meteorol. 2 (4): 259–70.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, 2008. Evaluation of measurement data — Guide to the expression of uncertainty in Measurement JCGM 100:2008, GUM 1995 with minor corrections. First ed. JCGM. https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

Blatchford, M.L., Karimi, P., Bastiaanssen, W.G.M., Nouri, H. 2018. From Global Goals to Local Gains — A Framework for Crop Water Productivity. ISPRS Int. J. Geo-Inf. 7 (414): 414. https://doi.org/10.3390/ijgi7110414.

Blatchford , M. L., Mannaerts, C. M., Zeng, Y., Nouri, H., Karimi, P. 2019. Status of accuracy in remotely sensed and in-situ agricultural water productivity estimates: A review. Remote Sens. Environ. 234 (1 December 2019): 111413. https://doi.org/10.1016/j.rse.2019.111413

Blatchford, M.L., Mannaerts, C.M., Njuki, S.M., Nouri, H., Zeng, Y., Pelgrum, H., Wonink, S., Karimi, P. 2020. Evaluation of WaPOR V2 evapotranspiration products across Africa. Hydrol. Process. doi:10.1002/hyp.13791.

Bloom, A.J., Chapin, F.S., Mooney, H.A., 1985. Resource Limitation in Plants-An Economic Analogy. Annu. Rev. Ecol. Syst. 16, 363–392. https://doi.org/10.1146/annurev.es.16.110185.002051

Booker, J.F., Trees, W.S., 2020, Implications of Water Scarcity for Water Productivity and Farm Labor, Water, 12, 308; doi:10.3390/w12010308.

Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential Uses and Limitations of Crop Models. Agron. J. 88, 704–716. https://doi.org/10.2134/agronj1996.00021962008800050005x

Bos, M.G., Nugteren, J. 1990. On irrigation efficiencies. 4. Ed. International Institute for Land Reclamation and Improvement. Publication 19. ILRI, Wageningen. http://edepot.wur.nl/71061

Bossio, D., Noble, A., Molden, D.J., Nangia, V., 2008. Land Degradation and Water Productivity in Agricultural Landscapes, in: Bossio, D., Geheb, K. (Eds.), Conserving Land, Protecting Water. CABI International, London, pp. 20–32. doi: 10.1079/9781845933876.0020

Boulain, N., Cappelaere, B., Séguis, L., Favreau, G., Gignoux, J. 2009. Water balance and vegetation change in the Sahel: A case study at the watershed scale with an eco-hydrological model. J. Arid Environ. 73 (12): 1125–1135. https://doi.org/10.1016/j.jaridenv.2009.05.008

Brauman, K., Siebert, S., Foley, J. 2013. Improvements in Crop Water Productivity Increase Water Sustainability and Food Security—a

Global Analysis. Environ. Res. Letters 8 (2): 024030. https://doi.org/10.1088/1748-9326/8/2/024030.

Briggs, L.J., Shantz, H.L. 1914. Relative Water Requirements of Plants. J. of Agric. Research 5: 116–32.

Brisson, N., Gary, C., Justes, E., Roch, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J., Hénault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model STICS. Eur. J. Agron. 18, 309–332. https://doi.org/10.1016/S1161-0301(02)00110-7

Brouwer, C., Prins, K., Heiblom, M., 1989. Irrigation Water Management: Irrigation Scheduling, FAO, Rome.

Burt, C.M., Clemmens, A.J., Strelkoff, T.S., Solomon, K.H., Bliesner, R.D., Hardy, L.A., Howell, T.A., Eisenhauer, D.E., 1997. Irrigation Performance Measures: Efficiency and Uniformity. J. Irrig. Drain. Eng. 123, 423–442. https://doi.org/10.1061/(ASCE)0733-9437(1997)

Buysse, P., Bodson, B., Debacq, A., De Ligne, A., Heinesch, B., Manise, T., Moureaux, C., Aubinet, M., 2017. Carbon budget measurement over 12 years at a crop production site in the silty-loam region in Belgium. Agric. For. Meteorol. 246, 241–255. https://doi.org/10.1016/j.agrformet.2017.07.004

Cai, X., Molden, D., Mainuddin, M., Sharma, B., Ahmad, M., Karimi, K. 2011. Producing More Food with Less Water in a Changing World: Assessment of Water Productivity in 10 Major River Basins. Water Int. 36 (1): 42–62. https://doi.org/10.1080/02508060.2011.542403.

Calera, A., Campos, I., Osann, A., D'Urso, G., Menenti, M., 2017. Remote sensing for crop water management: From ET modelling to services for the end users. Sensors. 17, 1104. https://doi.org/10.3390/s17051104

Campos, I., Neale, C.M.U., Calera, A., Balbontín, C., González-Piqueras, J., 2010. Assessing satellite-based basal crop coefficients for irrigated grapes (Vitis vinifera L.). Agric. Water Manag. 98, 45–54. https://doi.org/10.1016/j.agwat.2010.07.011

Campos, I., Villodre, J., Carrara, A., Calera, A., 2013. Remote sensing-based soil water balance to estimate Mediterranean holm oak savanna (dehesa) evapotranspiration under water stress conditions. J. Hydrol. 494, 1–9. https://doi.org/10.1016/j.jhydrol.2013.04.033

Campos, I., González-Gómez, L., Villodre, J., González-Piqueras, J., Suyker, A.E., Calera, A., 2018a. Remote sensing-based crop biomass with water or light-driven crop growth models in wheat commercial fields. Field Crops Res. 216, 175-188. https://doi.org/10.1016/j.fcr.2017.11.025

Campos, I., Neale, C.M.U., Arkebauer, T.J., Suyker, A.E., Gonçalves, I.Z., 2018b. Water productivity and crop yield: A simplified remote sensing driven operational approach. Agric. For. Meteorol. 249, 501–511. https://doi.org/10.1016/j.agrformet.2017.07.018

Casley, D., Kumar, K., 1988. The collection, analysis and use of monitoring and evaluation data. The John Hopkins University Press, Baltimore.
http://documents.worldbank.org/curated/en/534751468763750439/pdf/multi-page.pdf

Cellier, P., Olioso, A., 1993. A simple system for automated long-term Bowen ratio measurement. Agric. For. Meteorol. 66, 81–92. https://doi.org/10.1016/0168-1923(93)90083-T

Chen, C., Li, Y., 2013. A Robust Multiquadric Method for Digital Elevation Model Construction. Math. Geosci. 45, 297–319. https://doi.org/10.1007/s11004-013-9451-8

Cherkassky, V., Xuhui, S., Mulier, F.M., Vapnik, V.N., 1999. Model complexity control for regression using VC generalization bounds, in: IEEE Transactions on Neural Networks. IEEE, pp. 1075–1089. https://doi.org/10.1109/72.788648

Chishugi, J. B., Alemaw, B. F. 2009. The Hydrology of the Congo River Basin: A GIS-Based Hydrological Water Balance Model. World Environmental and Water Resources Congress 2009: Great Rivers. (pp. 1–16). Kansas City, USA: ASCE.

Chiti, T., Certini, G., Grieco, E., Valentini, R. 2010. The role of soil in storing carbon in tropical rainforests: The case of Ankasa Park, Ghana. Plant and Soil 331 (1–2): 453–461. https://doi.org/10.1007/s11104-009-0265-x

Choudhury, B.J., Ahmed, N.U., Idso, S.B., Reginato, R.J., Daughtry, C.S.T., 1994. Relations between evaporation coefficients

and vegetation indices studied by model simulations. Remote Sens. Environ. 50, 1–17. https://doi.org/10.1016/0034-4257(94)90090-6

Clark, D.A., Brown, S., Kicklighter, D.W., Chambers, J.Q., Thomlinson, J.R., Ni, J., 2001. Measuring net primary production in forests: concepts and field methods. Ecol. Appl. 11, 356–370. https://doi.org/10.1890/1051-0761(2001)011[0356:MNPPIF]2.0.CO;2

Clemmens, A.J., Burt, C.M., 1997. Accuracy of irrigation efficiency estimates. J. Irrig. Drain. Eng. 123, 443–453. https://doi.org/10.1061/(ASCE)0733-9437(1997)123:6(443)

Congalton, R.G., Green, K., 2009. Assessing the Accuracy of Remotely Sensed Data - Principles and Practices. Int. J. Appl. Earth Obs. Geoinf. 11, 183. https://doi.org/10.1016/j.jag.2009.07.002

Conijn, J.G., Bindraban, P.S., Schröder, J.J., Jongschaap, R.E.E. 2018. Can Our Global Food System Meet Food Demand within Planetary Boundaries? Agric., Ecosys. and Environ. 251 (June 2017): 244–56. https://doi.org/10.1016/j.agee.2017.06.001.

Conrad, C., Dech, S.W., Hafeez, M., Lamers, J.P.A., Tischbein, B., 2013. Remote sensing and hydrological measurement based irrigation performance assessments in the upper Amu Darya Delta, Central Asia. Phys. Chem. Earth, Parts A/B/C, 61–62, 52–62. https://doi.org/10.1016/j.pce.2013.05.002

Cleverly, J., Chen, C., Boulain, N., Villalobos-Vega, R., Faux, R., Grant, N., Yu, Q., Eamus, D. 2013. Aerodynamic Resistance and Penman–Monteith Evapotranspiration over a Seasonally Two-Layered Canopy in Semiarid Central Australia. J. Hydrometeo. 14 (5): 1562–1570. https://doi.org/10.1175/jhm-d-13-080.1

Córdova, M., Carrillo-Rojas, G., Crespo, P., Wilcox, B., Célleri, R. 2015. Evaluation of the Penman-Monteith (FAO 56 PM) Method for Calculating Reference Evapotranspiration Using Limited Data. Mount. Resear. and Dev. 35 (3): 230–239. https://doi.org/10.1659/mrd-journal-d-14-0024.1

Courault, D., Seguin, B., and Olioso, A. 2005. Review on Estimation of Evapotranspiration from Remote Sensing Data: From Empirical to Numerical Modeling Approaches. Irrig. and Drain. Syst. 19 (3–4): 223–49. https://doi.org/10.1007/s10795-005-5186-0.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Elem. Inf. Theory 1–748. https://doi.org/10.3390/e7040253

David, I., 1978. Non-sampling Errors in Agricultural Surveys. Review, Current Findings, and Suggestions for Future Research, in: Philippine Statistical Association Annual Conference, June 19. Manila.

De Bruin, H., Trigo, I., Bosveld, F., Meirink, J. 2016. Thermodynamically based model for actual evapotranspiration of an extensive grass field close to FAO reference, suitable for remote sensing application. J. Hydrometeo. 17 (5): 1373–1382. https://doi.org/10.1175/JHM-D-15-0006.1

De Bruin, H. A. R., Trigo, I.F. 2019. A new method to estimate reference crop evapotranspiration from geostationary satellite imagery: Practical considerations. Water 11 (2): 382. https://doi.org/10.3390/w11020382

Degefu, D. M., Weijun, H., Zaiyi, L., Liang, Y., Zhengwei, H., Min, A. 2018. Mapping Monthly Water Scarcity in Global Transboundary Basins at Country-Basin Mesh Based Spatial Resolution. Scientific Reports 8 (1): 1–10. https://doi.org/10.1038/s41598-018-20032-w

Delécolle, R., Maas, S.J., Guérif, M., Baret, F., 1992. Remote sensing and crop production models: present trends. ISPRS J. Photogramm. Remote Sens. 47, 145–161. https://doi.org/10.1016/0924-2716(92)90030-D

Dembélé, M., Zwart, S.J. 2016. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. Int. J. Remote Sens. 37 (17): 3995–4014. https://doi.org/10.1080/01431161.2016.1207258

Dhungel, R., Allen ,R. G., Trezza, R., Robison, C. W. 2014. Comparison of latent heat flux using aerodynamic methods and using the penman-monteith method with satellite-based surface energy balance. Remote Sens. 6 (9): 8844–8877. https://doi.org/10.3390/rs6098844

Diskin, P., 1999. Agricultural Productivity Indicators Measurement Guide, Arlington, Va: Food Security and Nutrition Monitoring (IMPACT) Project, ISTI, Methods. U.S. Agency for International Development, Arlington. https://reliefweb.int/sites/reliefweb.int/files/resources/842682301AA

98504C1256F070044D507-
USAID_Agricultural_indicators_December_1997.pdf

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., Ceccato, P. 2018. Validation of the CHIRPS satellite rainfall estimates over eastern Africa. Quarterly J. Royal Meteorol. Society 144 (April): 292–312. https://doi.org/10.1002/qj.3244

Donald, C.M., 1962. In search of yield. Aust. Inst. Agric. Sci. 28, 171–178.

Doorenbos, J., Pruitt, W.O., 1977. Guidelines for predicting crop water requirements. FAO Irrigation and Drainage Paper No. 24. FAO, Rome. http://www.fao.org/3/a-f2430e.pdf

Dore, S., Hymus, G.J., Johnson, D.P., Hinkle, C.R., Valentini, R., Drake, B.G., 2003. Cross validation of open-top chamber and eddy covariance measurements of ecosystem CO2 exchange in a Florida scrub-oak ecosystem. Glob. Chang. Biol. 9, 84–95. https://doi.org/10.1046/j.1365-2486.2003.00561.x

Duchemin, B., Hadria, R., Erraki, S., Boulet, G., Maisongrande, P., Chehbouni, A., Escadafal, R., Ezzahar, J., Hoedjes, J.C.B., Kharrou, M.H., Khabba, S., Mougenot, B., Olioso, A., Rodriguez, J.C., Simonneaux, V., 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. Agric. Water Manag. 79, 1–27. https://doi.org/10.1016/j.agwat.2005.02.013

Duchemin, B., Maisongrande, P., Boulet, G., Benhadj, I., 2008. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. Environ. Model. Softw. 23, 876–892. https://doi.org/10.1016/j.envsoft.2007.10.003

Dugas, W.A., Bland, W.L., 1989. The accuracy of evaporation measurements from small lysimeters. Agric. For. Meteorol. 46, 119–129. https://doi.org/10.1016/0168-1923(89)90116-0

Dugas, W.A., Reicosky, D.C., Kiniry, J.R., 1997. Chamber and micrometeorological measurements of CO2 and H2O fluxes for three C4 grasses. Agric. For. Meteorol. 83, 113–133. https://doi.org/10.1016/S0168-1923(96)02346-5

Durgam, O.Y., Gobin, A., Duveiller, G., Tychon, B. 2020. A study on trade-offs between spatial resolution and temporal sampling density for wheat yield estimation using both thermal and calendar time. Int J Appl Earth Obs Geoinf. 86, 101988. https://doi.org/10.1016/j.jag.2019.101988

Duveiller, G. 2012. Caveats in calculating crop specific pixel purity for agricultural monitoring using MODIS time series. In Proceedings of the SPIE; Vol. 8531, pp. 1–10. https://doi.org/10.1117/12.974625

Duveiller, G., Baret, F., Defourny, P., 2011. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. Remote Sens. Environ. 115, 2686–2701. https://doi.org/10.1016/j.rse.2011.05.026

Duveiller, G., Lopez-Lozano, R., Cescatti, A., 2015. Exploiting the multi-angularity of the MODIS temporal signal to identify spatially homogeneous vegetation cover: A demonstration for agricultural monitoring applications. Remote Sens. Environ. 166, 61–77. https://doi.org/10.1016/j.rse.2015.06.001

EFTAS, FAO, 2015. Protocol for land cover validation. Rome, Italy.

El-Marsafawy, S.M., Swelam, A., Ghanem, A., 2018. Evolution of crop water productivity in the Nile Delta over three decades (1985-2015). Water. 10, 1168. https://doi.org/10.3390/w10091168

El-Rafey, A., Wojtusiak, J., 2017. Recent advances in scaling-down sampling methods in machine learning. Wiley Interdiscip. Rev. Comput. Stat. 9. https://doi.org/10.1002/wics.1414

Elnmer, A., Khadr, M., Allam, A., Kanae, S., Tawfik, A. 2018. Assessment of irrigation water performance in the Nile Delta using remotely sensed data. Water. 10, 1375. https://doi.org/10.3390/w10101375

Ershadi, A., Mccabe, M.F., Evans, J.P., Walker, J.P. 2013. Effects of spatial aggregation on the multi-scale estimation of evapotranspiration. Remote Sens. Env. 131, 51–62. https://doi.org/10.1016/j.rse.2012.12.007

Ershadi, A., McCabe, M.F., Evans, J.P., Chaney, N.W., Wood, E.F., 2014. Multi-site evaluation of terrestrial evaporation models using

FLUXNET data. Agric. For. Meteorol. 187, 46–61. https://doi.org/10.1016/j.agrformet.2013.11.008

Estrada, A., Morales, E.F., 2004. NSC: A New Progressive Sampling Algorithm, in: Proceedings of the Workshop: Machine Learning Learning for Scientific Data Analysis (Iberamia). pp. 335–344.

Evett, S.R., Schwartz, R.C., Howell, T.A., Louis Baumhardt, R., Copeland, K.S., 2012. Can weighing lysimeter ET represent surrounding field ET well enough to test flux station measurements of daily and sub-daily ET? Adv. Water Resour. 50, 79–90. https://doi.org/10.1016/j.advwatres.2012.07.023

Falkenmark, M., Rockström, J., Karlberg, L.. 2009. Present and Future Water Requirements for Feeding Humanity. Food Security 1 (1): 59–69. https://doi.org/10.1007/s12571-008-0003-x.

Fanadzo, M., Chiduza, C., Mnkeni, P.N.S. 2010. Overview of smallholder irrigation schemes in south Africa: Relationship between farmer crop management practices and performance. Afr. J. Agric. Res. 5, 3514–3523. doi:10.5897/AJAR10.001.

FAO. 1997. Irrigation potential in Africa: A basin approach. FAO. Rome, Italy. Available at: http://www.fao.org/3/W4347E/w4347e00.htm

FAO, 1982. Estimation of crop areas and yields in agricultural statistics. FAO economic and social development paper No. 22. FAO, Rome.

FAO, 2017. Using Remote Sensing in support of solutions to reduce agricultural water productivity gaps. Database Methodology: Level 1 Data. WaPOR beta release.

FAO, 2018. WaPOR Database Methodology: Level 1. Remote Sensing for Water Productivity Technical Report: Methodology Series. FAO, Licence: CC BY-NC-SA 3.0 IGO, Rome.

FAO, 2019. WaPOR - The FAO portal to monitor Water Productivity through Open access or Remotely sensed derived data. https://wapor.apps.fao.org/home/1 (accessed 26 March 20).

FAO. 2020a. WaPOR Database methodology, V2 release (not yet published). Rome, Italy.

FAO. 2020b. WaPOR - The FAO portal to monitor Water Productivity through Open access or Remotely sensed derived data, FAO, Rome, Italy, https://wapor.apps.fao.org/home/1, date accessed: 26-10-2019.

FAO. 2020c. WaPOR V2 quality assessment – Technical Report on the Data Quality of the WaPOR FAO Database version 2. Rome. https://doi.org/10.4060/
cb2208en

FAO and IHE Delft. 2019. WaPOR quality assessment Technical report on the data quality of the WaPOR FAO database version 1.0, 134 pp. Rome, Italy. Available at: http://www.fao.org/3/ca4895en/ca4895en.pdf

Fereres, E., Soriano, M.A., 2007. Deficit irrigation for reducing agricultural water use. J. Exp. Bot. 58, 147–159. https://doi.org/10.1093/jxb/erl165

Fermont, A., Benson, T., 2011. Estimating yield of food crops grown by smallholder farmers: A Review in the Uganda Context, IFPRI Discussion Paper 01097. International Food Policy Research Institute (IFPRI). Washington. http://www.ifpri.org/publication/estimating-yield-food-crops-grown-smallholder-farmers

Fielding, W.J., Riley, J., 1997. How big should on-farm trials be and how many plots should be measured? in: PLA Notes. IIED, London, pp. 19–22.

Figueroa, R.L., Zeng-Treitler, Q., Kandula, S., Ngo, L.H., 2012. Predicting sample size required for classification performance. BMC Med. Inform. Decis. Mak. 12, 8. https://doi.org/10.1186/1472-6947-12-8

Foken, T., 2008. The energy balance closure problem: An overview. Ecol. Appl. 18, 1351–1367. https://doi.org/10.1890/06-0922.1

Foken, T., Leclerc, M.Y., 2004. Methods and limitations in validation of footprint models. Agric. For. Meteorol. 127, 223–234. https://doi.org/10.1016/j.agrformet.2004.07.015

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Fox, A.M., Huntley, B., Lloyd, C.R., Williams, M., Baxter, R., 2008. Net ecosystem exchange over heterogeneous Arctic tundra: Scaling between chamber and eddy covariance measurements. Global Biogeochem. Cycles 22, 1–15. https://doi.org/10.1029/2007GB003027

Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., Sobrino, J.A., 2015. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. Remote Sens. Environ. 161, 131–148. https://doi.org/10.1016/j.rse.2015.02.014

Frenken, K. 2005. Irrigation in Africa in figures, FAO WATER REPORTS 29, AQUASTAT Survey, FAO Land and Water Development Division, ISBN 9251054142.

Friederichs, P., Hense. A. 2007. Statistical Downscaling of Extreme Precipitation Events Using Censored Quantile Regression. Monthly Weather Review June 2007: 2365–78. https://doi.org/https://doi.org/10.1175/MWR3403.1.

Fukunaga, K., Hayes, R.R., 1989. Effects of Sample Size in Classifier Design. IEEE Trans. Pattern Anal. Mach. Intell. 11, 873–885. https://doi.org/10.1109/34.31448

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Harrison, L., Hoell, A., Michaelsen, J. 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data 2: 150066. https://doi.org/10.1038/sdata.2015.66

Gao F., Masek J., Schwaller, M., Hall, F. 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Transactions on Geoscience and Remote Sens. 44 (8): 2207–2218. https://doi.org/10.1109/TGRS.2006.872081

Gebler, S., Hendricks Franssen, H.J., Pütz, T., Post, H., Schmidt, M., Vereecken, H., 2015. Actual evapotranspiration and precipitation measured by lysimeters: A comparison with eddy covariance and tipping bucket. Hydrol. Earth Syst. Sci. 19, 2145–2161. https://doi.org/10.5194/hess-19-2145-2015

Geerts, S., Raes, D., 2009. Deficit irrigation as an on-farm strategy to maximize crop water productivity in dry areas. Agric. Water Manag. 96, 1275–1284. https://doi.org/10.1016/j.agwat.2009.04.009

Gebrechorkos, S. H., Hülsmann, S., Bernhofer, C. 2018. Evaluation of multiple climate data sources for managing environmental resources in East Africa. Hydrol. Earth Syst. Sci. 22 (8): 4547–4564. https://doi.org/10.5194/hess-22-4547-2018

Ghilain, N., Arboleda, A., Gellens-Meulenberghs, F. 2011. Evapotranspiration modelling at large scale using near-real time MSG SEVIRI derived data. Hydrol. Earth Syst. Sci. 15: 771–786. https://doi.org/10.5194/hess-15-771-2011

Glenn, E.P., Huete, A.R., Nagler, P.L., Hirschboeck, K.K., Brown, P., 2007. Integrating Remote Sensing and Ground Methods to Estimate Evapotranspiration. Crit. Rev. Plant Sci. 26, 139–168. https://doi.org/10.1080/07352680701402503

Glenn, E. P., Huete, A. R., Nagler, P. L., Nelson, S. G. 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. Sensors 8 (4): 2136–2160. https://doi.org/10.3390/s8042136

Glenn, E.P., Nagler, P.L., Huete, A.R., 2010. Vegetation Index Methods for Estimating Evapotranspiration by Remote Sensing. Surv. Geophys. 31, 531–555. https://doi.org/10.1007/s10712-010-9102-2

Glenn, E.P., Neale, C.M.U., Hunsaker, D.J., Nagler, P.L., 2011. Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems. Hydrol. Process. 25, 4050–4062. https://doi.org/10.1002/hyp.8392

González-Dugo, M.P., Escuin, S., Cano, F., Cifuentes, V., Padilla, F.L.M., Tirado, J.L., Oyonarte, N., Fernández, P., Mateos, L., 2013. Monitoring evapotranspiration of irrigated crops using crop coefficients derived from time series of satellite images. II. Application on basin scale. Agric. Water Manag. 125, 92-104. http://dx.doi.org/10.1016/j.agwat.2013.03.024

Gobron, N., Pinty, B., Verstraete, M.M., Widlowski, J-L., 2000. Advanced Vegetation Indices Optimized for Up-Coming Sensors:

Design, Performance, and Applications. IEEE Trans. Geosci. Remote Sens. 38, 2489-2505. doi: 10.1109/36.885197

Gu, B., Liu, B., Hu, F., Liu, H., 2001. Efficiently Determining the Starting Sample Size for Progressive Sampling. Mach. Learn. ECML 2001 192–202. https://doi.org/10.1007/3-540-44795-4_17

Guerra, E., Ventura, F., Snyder, R.L., 2015. Crop Coefficients : A Literature Review. J. Irrig. Drain Eng. 142, 1–5. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000983

Guillevic, P. C., Olioso, A., Hook, S. J., Fisher, J. B., Lagouarde, J. P., Vermote, E. F. 2019. Impact of the revisit of thermal infrared remote sensing observations on evapotranspiration uncertainty-A sensitivity study using AmeriFlux Data. Remote Sens. 11 (5): 573. https://doi.org/10.3390/rs11050573

Guodong, T., Honglu, L., Fahu, L., 2016. Evaluation of dual crop coefficient approach on evapotranspiration calculation of cherry trees. Int J Agric Biol Eng. 9, 29–39. https://doi.org/10.3965/j.ijabe.20160903.1886

Gupta, K.K., Attri, J.P., Singh, A., Kaur, H., Kaur, G., 2016. Basic concepts for sample size calculation: Critical step for any clinical trials! Saudi J. Anaesth. 10, 328–331. https://doi.org/10.4103/1658-354X.174918

Hanks, B.J., and Tanner, C.B. 1952. Water Consumption by Plants as Influenced by Soil Fertility. Agron. J. 44: 98–100.

Hanks, R.J., Gardner H.R, and Florian R.L. 1969. Plant Growth-Evapotranspiration Relations for Several Crops in the Central Great Plains. Agron. J. 61 (Jan-Feb): 30–34.

Hatfield, J.L., G. Asrar, G., Kanemasu, E. T. 1984. Intercepted Photosynthetically Active Radiation Estimated by Spectral Reflectance. Remote Sens. Environ. 14 (1–3): 65–75. https://doi.org/10.1016/0034-4257(84)90008-7.

Hatfield, J.L., Prueger, J.H., Kustas, W.P., Anderson, M.C., Alfieri, J.G. 2016. Evapotranspiration: Evolution of Methods to Increase Spatial and Temporal Resolution. In Improving Modeling Tools to Assess Climate Change Effects on Crop Response, Volume 7https://doi.org/10.2134/advagricsystmodel7.2015.0076

Hay, R.K.M., 1995. Harvest index: a review of its use in plant breeding and crop physiology. Ann. Appl. Biol. 126, 197–216. https://doi.org/10.1111/j.1744-7348.1995.tb05015.x

Hellegers, P.J.G.J., Soppe, R., Perry, C.J., Bastiaanssen, W.G.M., 2009. Combining remote sensing and economic analysis to support decisions that affect water productivity. Irrig. Sci. 27, 243–251. https://doi.org/10.1007/s00271-008-0139-7

Hellegers, P.J.G.J., Soppe, R., Perry, C.J., Bastiaanssen, W.G.M. 2010. Remote Sensing and Economic Indicators for Supporting Water Resources Management Decisions. Water Resour. Manag. 24, 2419-2436. https://doi.org/10.1007/s11269-009-9559-2

Heydari, S.S., Mountrakis, G., 2018. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. Remote Sens. Environ. 204, 648–658. https://doi.org/10.1016/j.rse.2017.09.035

Hills, R. C. 1978. The Structure of the Inter-Tropical Convergence Zone in Equatorial Africa and Its Relationship to East African Rainfall. Transactions of the Institute of British Geographers. 4 (3): 329–352. https://doi.org/10.2307/622055

Hirschi, M., Michel, D., Lehner, I., Seneviratne, S.I., 2017. A site-level comparison of lysimeter and eddy covariance flux measurements of evapotranspiration. Hydrol. Earth Syst. Sci. 21, 1809–1825. https://doi.org/10.5194/hess-21-1809-2017

Hobbins, M. T., Ramírez, J. A, Brown, T. C. 2001. The complementary relationship in regional evapotranspiration: An enhanced advection-aridity model. Water Resour. Res. 37 (5): 1389–1403. https://doi.org/10.1029/2000WR900359

Hoogeveen, J., Faurès, J.M., Peiser, L., Burke, J., Van de Giesen, N. 2015. GlobWat – a Global Water Balance Model to Assess Water Use in Irrigated Agriculture. Hydrol. Earth Syst. Sci. 19 (9): 3829–44. https://doi.org/10.5194/hess-19-3829-2015.

Howell, T.A., Schneider, A.D., Dusek, D.A., Marek, T.H., Steiner, J.L., 1995. Calibration and scale performance of bushland weighing lysimeters. Trans. ASAE 38, 1019–1024. doi:10.13031/2013.27918

Howell, T.A., Evett, S.R., Tolk, J.A., Copeland, K.S., Marek, T.H., 2015. Evapotranspiration, water productivity and crop coefficients for irrigated sunflower in the U.S. Southern High Plains. Agric. Water Manag. 162, 33–46. https://doi.org/10.1016/j.agwat.2015.08.008

Hu, G., Jia, L., Menenti, M., 2015. Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011. Remote Sens. Environ. 156, 510–526. https://doi.org/10.1016/j.rse.2014.10.017

Hunsaker, D.J., Fitzgerald, G.J., French, A.N., Clarke, T.R., Ottman, M.J., Pinter, J.P.J., 2007. Wheat Irrigation Management Using Multispectral Crop Coefficients: I. Crop Evapotranspiration Prediction. Trans. ASABE 50, 2017–2033. https://doi.org/10.13031/2013.24105

Idso, S.B., Jackson, R.D., Reginato, R.J., 1977. Remote-sensing of crop yields. Science 196, 19–25. https://doi.org/10.1126/science.196.4285.19

Imukova, K., Ingwersen, J., Hevart, M., Streck, T., 2016. Energy balance closure on a winter wheat stand: Comparing the eddy covariance technique with the soil water balance method. Biogeosciences 13, 63–75. https://doi.org/10.5194/bg-13-63-2016

Irmak, S., Skaggs, K.E., Chatterjee, S., 2014. A Review of the Bowen Ratio Surface Energy Balance Method for Quantifying Evapotranspiration and other Energy Fluxes. Trans. ASABE 57, 1657–1674. https://doi.org/10.13031/trans.57.10686

Ismail, A.M.A., 1993. A critical analysis of "harvest index." Qatar Univ. Sci. J. 13, 253–263. http://hdl.handle.net/10576/9624

Jackson, R.D., Slater, P.N., Pinter, J.P.J., 1983. Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres. Remote Sens. Environ. 13, 187–208. https://doi.org/10.1016/0034-4257(83)90039-1

Jarchow, C.J., Didan, K., Barreto-Muñoz, A., Nagler, P.L., Glenn, E.P. 2018. Application and comparison of the MODIS-derived enhanced vegetation index to VIIRS, landsat 5 TM and landsat 8 OLI platforms: A case study in the arid colorado river delta, Mexico. Sensors. 18. doi:10.3390/s18051546.

Jaynes, E.T., 1957. Information Theory and Statistical Mechanis. Phys. Rev. 106, 620–630.

Jaynes, E.T., 2003. Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK.

Jiang, L., Islam, S., Carlson, T.N., 2004. Uncertainties in latent heat flux measurement and estimation: Implications for using a simplified approach with remote sensing data. Can. J. Remote Sens. 30, 769–787. https://doi.org/10.5589/m04-038

Jiang, Z., Chen, Y., Li, J., Dou, W. 2005. The impact of spatial resolution on NDVI over heterogeneous surface. In Proceedings of the IEEE; Bejing, pp. 1310–1313. https://doi.org/10.1109/IGARSS.2005.1525361

Jiang, Y., Xu, X., Huang, Q., Huo, Z., Huang, G., 2015. Assessment of irrigation performance and water productivity in irrigated areas of the middle Heihe River basin using a distributed agro-hydrological model. Agric. Water Manag. 147, 67–81. https://doi.org/10.1016/j.agwat.2014.08.003

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. Eur. J. Agron. 92, 141–152. https://doi.org/10.1016/j.eja.2017.11.002

John, G., Langley, P., John, H., 1996. Static Versus Dynamic Sampling for Data Mining. Kdd 367–370. https://doi.org/10.1007/s00221-011-2539-9

Kalma, J.D., McVicar, T.R., McCabe, M.F., 2008. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. Surv. Geophys. 29, 421–469. https://doi.org/10.1007/s10712-008-9037-z

Karatas, B.S., Akkuzu, E., Unal, H.B., Asik, S., Avci, M. 2009. Using satellite remote sensing to assess irrigation performance in Water User Associations in the Lower Gediz Basin, Turkey. Agric. Water Manag. 96, 982–990. https://doi.org/10.1016/j.agwat.2009.01.010

Karimi, P., Molden, D., Notenbaert, A., Peden, D. 2012. Nile basin farming systems and productivity. In The Nile River Basin: Water, Agriculture, Governance and Livelihoods; Awulachew, S.B., Smakhtin,

V., Molden, D., Peden, D., Eds.; Routledge - Earthscan: Abingdon, UK, pp. 133–153.

Karimi, P., Bastiaanssen, W.G.M., 2015. Spatial evapotranspiration, rainfall and land use data in water accounting - Part 1: Review of the accuracy of the remote sensing data. Hydrol. Earth Syst. Sci. 19, 507–532. https://doi.org/10.5194/hess-19-507-2015

Karimi, P., Bhembe, B., Blatchford, M.L., De Fraiture, C., 2019. Global Satellite-Based ET Products for the Local Level Irrigation Management: An Application of Irrigation Performance Assessment in the Sugarbelt of Swaziland. Remote Sens. 11, 705. https://doi.org/10.3390/rs11060705

Kasampalis, D., Alexandridis, T., Deva, C., Challinor, A., Moshou, D., Zalidis, G., 2018. Contribution of Remote Sensing on Crop Models: A Review. J. Imaging 4, 52. https://doi.org/10.3390/jimaging4040052

Kemanian, A.R., Stöckle, C.O., Huggins, D.R., Viega, L.M., 2007. A simple method to estimate harvest index in grain crops. F. Crop. Res. 103, 208–216. https://doi.org/10.1016/j.fcr.2007.06.007

Khan, M.S., Liaqat, U.W., Baik, J., Choi, M., 2018. Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. Agric. For. Meteorol. 252, 256–268. https://doi.org/10.1016/j.agrformet.2018.01.022

Kharrou, M.H., Le Page, M., Chehbouni, A., Simonneaux, V., Er-Raki, S., Jarlan, L., Ouzine, L., Khabba, S., Chehbouni, G., 2013. Assessment of Equity and Adequacy of Water Delivery in Irrigation Systems Using Remote Sensing-Based Indicators in Semi-Arid Region, Morocco. Water Resour. Manag. 27, 4697–4714. https://doi.org/10.1007/s11269-013-0438-5

Kijne, J.W., 2003. Why agricultural water productivity is important for the global water challenge, in: Kijne, J.W., FAO. (Eds.), Unlocking the Water Potential of Agriculture. FAO, Rome. pp. 23-24. http://www.fao.org/3/y4525e/y4525e00.htm

Kijne, J., Barron, J., Hoff, H., Rockström, J. 2009. Opportunities to Increase Water Productivity in Agriculture with Special Reference to Africa and South Asia. A Report Prepared by Stockholm Environment

Institute, for the Swedish Ministry of Environment Presentation at CSD 16: 48. http://indiaenvironmentportal.org.in/files/opportunitiestoincreasewaterprod2009.pdf.

Kim, S.Y., 2009. Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. BMC Bioinformatics 10, 4–7. https://doi.org/10.1186/1471-2105-10-147

Kirby, A., Gebski, V., Keech, A.C., 2002. Determining the sample size in a clinical trial. Med. J. Aust. 177, 256–257. https://doi.org/10.5694/j.1326-5377.2003.tb05241.x

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kumar, S. V., Dirmeyer, P.A., Peters-Lidard, C.D., Bindlish, R., Bolten, J., 2018. Information theoretic evaluation of satellite soil moisture retrievals. Remote Sens. Environ. 204, 392–400. https://doi.org/10.1016/j.rse.2017.10.016

Kutzbach, L., Schneider, J., Sachs, T., Giebels, M., Nykänen, H., Shurpali, N.J., Martikainen, P.J., Alm, J., Wilmking, M., 2007. CO2 flux determination by closed-chamber methods can be seriously biased by inappropriate application of linear regression. Biogeosciences 4, 1005–1025. https://doi.org/10.5194/bgd-4-2279-2007

Lachin, J.M., 1981. Introduction to sample size determination and power analysis for clinical trials. Control. Clin. Trials 2, 93–113. https://doi.org/10.1016/0197-2456(81)90001-5

Last, M., 2009. Improving data mining utility with projective sampling. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 487–495. https://doi.org/10.1145/1557019.1557076

Langensiepen, M., Kupisch, M., Van Wijk, M.T., Ewert, F., 2012. Analyzing transient closed chamber effects on canopy gas exchange for optimizing flux calculation timing. Agric. For. Meteorol. 164, 61–70. https://doi.org/10.1016/j.agrformet.2012.05.006

Lazarevic, A.O., Obradovic, Z., 2001. Data reduction using multiple models integration, in: de Raedt, L., Siebes, A. (Ed.),

Principles of Data Mining and Knowledge Discovery: 5th European Conference, PKDD 2001. Springer, Berlin, pp. 301–313.

LBPTC. 2010. Joint Limpopo River Basin Study Scoping Phase Final Report. Federal Ministry for Economic Cooperation and Development and DFID. Available at: http://www.limpopo.riverawarenesskit.org/LIMPOPORAK_COM/_SYSTEM/DMSSTORAGE/3451EN/LIMCOM2010_SCOPINGSTUDY_ENG.PDF

Leca, A., Parisi, L., Lacointe, A., Saudreau, M. 2011. Comparison of Penman–Monteith and non-linear energy balance approaches for estimating leaf wetness duration and apple scab infection, Agric. and For. Meteorol. 151 (2011): 1158–1162. https://doi.org/10.1016/j.agrformet.2011.04.010

Lehner, B., Grill, G. 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol. Process. 27 (15): 2171–2186. https://doi.org/10.1002/hyp.9740

Li, K. Y., Coe, M. T., Ramankutty, N. 2005. Investigation of hydrological variability in West Africa using land surface models. J. Climate 18 (16): 3173–3188. https://doi.org/10.1175/JCLI3452.1

Li, Z.L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., Yan, G., Zhang, X., 2009. A review of current methodologies for regional Evapotranspiration estimation from remotely sensed data. Sensors. 9, 3801–3853. https://doi.org/10.3390/s90503801

Liu, Y., Luo, Y., 2010. A consolidated evaluation of the FAO-56 dual crop coefficient approach using the lysimeter data in the North China Plain. Agric. Water Manag. 97, 31–40. https://doi.org/10.1016/j.agwat.2009.07.003

Liu, Z., Wang, L., Wang, S. 2014. Comparison of Different GPP Models in China Using MODIS Image and ChinaFLUX Data, Remote Sens. 6: 10215–31. https://doi.org/10.3390/rs61010215.

Liu, S., Xu, Z., Song, L., Zhao, Q., Ge, Y., Xu, T., Ma, Y., Zhu, Z., Jia, Z., Zhang, F., 2016. Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. Agric. For. Meteorol. 230–231, 97–113. https://doi.org/10.1016/j.agrformet.2016.04.008

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. F. Crop. Res. 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008

Lobell, D.B., Ortiz-Monasterio, J.I., Addams, C.L., Asner, G.P., 2002. Soil, climate, and management impacts on regional wheat productivity in Mexico from remote sensing. Agric. For. Meteorol. 114, 31–43. https://doi.org/10.1016/S0168-1923(02)00138-7

Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. Agric. Ecosyst. Environ. 94, 205–220. https://doi.org/10.1016/S0167-8809(02)00021-X

Loew, A., Bell, W., Brocca, L., Bulgin, C.E., Burdanowitz, J., Calbet, X., Donner, R. V, Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J., Schaepman-Strub, G., Schröder, M., 2017. Validation practices for satellite based earth observation data across communities. Rev. Geophys. https://doi.org/10.1002/2017RG000562

Lorite, I.J., Santos, C., Testi, L., Fereres, E., 2012. Design and construction of a large weighing lysimeter in an almond orchard. Spanish J. Agric. Res. 10, 238-250. https://doi.org/10.5424/sjar/2012101-243-11

Löw, F., Biradar, C., Fliemann, E., Lamers, J.P.A., Conrad, C., 2017. Assessing gaps in irrigated agricultural productivity through satellite earth observations—A case study of the Fergana Valley, Central Asia. Int. J. Appl. Earth Obs. Geoinf. 59, 118–134. https://doi.org/10.1016/j.jag.2017.02.014

Luo, G., 2016. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Netw. Model. Anal. Heal. Informatics Bioinforma. 5, 1–16. https://doi.org/10.1007/s13721-016-0125-6

Madugundu, R., Al-Gaadi, Tola, E., K.A., Kayad, A.G., Jha, C.S., 2017. Estimation of gross primary production of irrigated maize using Landsat-8 imagery and Eddy Covariance data. Saudi J. Biol. Sci. 24, 410–420. https://doi.org/10.1016/j.sjbs.2016.10.003

Mahalanobis, P.C., Sengupta, J.M., 1951. On the Size of Sample Costs In Crop-Cut Experiments. Bull. Int. Stat. Inst. 33, 359–404. http://hdl.handle.net/10263/2010

Majozi, N. P., Mannaerts, C .M., Ramoelo, A., Mathieu, R., Mudau, A. E., Verhoef, W. 2017a. An intercomparison of satellite-based daily evapotranspiration estimates under different eco-climatic regions in South Africa. Remote Sens. 9 (4): 307. https://doi.org/10.3390/rs9040307

Majozi, N. P., Mannaerts, C. M., Ramoelo, A., Mathieu, R., Nickless, A., Verhoef, W. 2017b. Analysing surface energy balance closure and partitioning over a semi-arid savanna FLUXNET site in Skukuza, Kruger National Park, South Africa. Hydrol. Earth Syst. Sci. 21 (7): 3401–3415. https://doi.org/10.5194/hess-21-3401-2017

Makarovic, B., 1973. Progressive Sampling for Digital Terrain Models. ITC J. 3, 145–153.

Malik, S.J., 1993. Farm Production Data in Rural Household Surveys, in: Puetz, D., von Braun, J. (Eds.), Data Needs for Food Policy in Developing Countries: New Directions for Household Surveys. International Food Policy Research Institute, Washington.

Maljanen, M., Martikainen, P.J., Walden, J., Silvola, J., 2001. CO2 exchange in an organic field growing barley or grass in eastern Finland. Glob. Chang. Biol. 7, 679–692. https://doi.org/10.1111/j.1365-2486.2001.00437.x

Mamadou, O., Cohard, J. M., Galle, S., Awanou, C. N., Diedhiou, A., Kounouhewa, B., Peugeot, C. 2014. Energy fluxes and surface characteristics over a cultivated area in Benin: daily and seasonal dynamics. Hydrol. Earth Syst. Sci. 18: 893–914. https://doi.org/10.5194/hess-18-893-2014

Mateos, L., Gonzalez-Dugo, M. P., Testi, L., Villalobos, F. J., 2013. Monitoring evapotranspiration of irrigated crops using crop coefficients derived from time series of satellite images. I. Method validation, Agric. Water Manag. 125, 81-91. http://dx.doi.org/10.1016/j.agwat.2012.11.005

Matondo, J. I., Mortensen, P. 1998. Water resource assessment for the Zambezi River Basin. Water Int. 23 (4): 256–262. https://doi.org/10.1080/02508069808686780

Mauder, M., Genzel, S., Fu, J., Kiese, R., Soltani, M., Steinbrecher, R., Zeeman, M., Banerjee, T., De Roo, F., Kunstmann, H., 2018. Evaluation of energy balance closure adjustment methods by independent evapotranspiration estimates from lysimeters and

hydrological simulations. Hydrol. Process. 32, 39–50. https://doi.org/10.1002/hyp.11397

McDonald, J. E. 1961. On the Ratio of Evaporation to Precipitation. Bulletin of the American Meteorol. Society 42 (3): 185–189. https://doi.org/10.1175/1520-0477-42.3.185

Meek, C., Thiesson, B., Heckerman, D., 2002. The learning-curve sampling method applied to model-based clustering. J. Mach. Learn. Res. 2, 397–418. https://doi.org/10.1162/153244302760200678

Mekonnen, M.M., Hoekstra, A.Y. 2011. The Green, Blue and Grey Water Footprint of Crops and Derived Crop Products. Hydrol. Earth Syst. Sci. 15 (5): 1577–1600. https://doi.org/10.5194/hess-15-1577-2011.

Melton, F.S., Johnson, L.F., Lund, C.P., Pierce, L.L., Michaelis, A.R., Hiatt, S.H., Guzman, A., Adhikari, D.D., Purdy, A.J., Rosevelt, C., Votava, P., Trout, T.J., Temesgen, B., Frame, K., Sheffner, E.J., Nemani, R.R., 2012. Satellite irrigation management support with the terrestrial observation and prediction system: A framework for integration of satellite and surface observations to support improvements in agricultural water resource management. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 1709–1721. https://doi.org/10.1109/JSTARS.2012.2214474

Menenti, M., Choudhury, B.J., 1993. Parameterization of land surface evaporation by means of location dependent potential evaporation and surface temperature range, in: Exchange Processes at the Land Surface for a Range of Space and Time Scales, IAHS, Yokohama, Japan (1993), pp. 561-568. http://hydrologie.org/redbooks/a212/iahs_212_0561.pdf

Menenti, M., Visser, T., Morabito, J.A., Drovandi, A., 1989. Appraisal of irrigation performance with satellite data and georeferenced information, in: Rydzewski, J.R. and Ward, C.F. (Eds.) Irrigation, Theory and Practice, Proc. of the Int. Conf., Institute of Irrigation Studies, Southhampton, 12-15 September 1989. Pentech Press, London., pp. 785–801.

Menenti, M., Chambouleyron, J., Morábito, J., Fornero, L., Stefanini, L., 1992. Appraisal and Optimization of Agricultural Water Use in Large Irrigation Schemes: II. Applications. Water Resour. Manag. 6, 201–221. https://doi.org/10.1007/BF00872356

Merbold, L., Ardö, J., Arneth, A., Scholes, R. J., Nouvellon, Y., Grandcourt, A., Archibald, S., Bonnefond, J.M., Boulain, N., Brueggemann, N., Bruemmer, C., Cappelaere, B., Ceschia, E., El-Khidir, H.A.M., El-Tahir, B.A., Falk, U., Lloyd, J., Kergoat, L., Le Dantec, V., Mougin, E., Muchinda, M., Mukelabai, M.M, Ramier, D., Roupsard, O., Timouk, F., Veenendaal, F.M., Kutsch, W.L., Kergo, W.L. 2009. Precipitation as driver of carbon fluxes in 11 African ecosystems, Biogeosciences 6 (6): 1027–1041. https://doi.org/10.5194/bg-6-1027-2009

Mira, M., Weiss, M., Baret, F., Courault, D., Hagolle, O., Gallego-Elvira, B., Olioso, A., 2015. The MODIS (collection V006) BRDF/albedo product MCD43D: Temporal course evaluated over agricultural landscape. Remote Sens. Environ. 170, 216–228. https://doi.org/10.1016/j.rse.2015.09.021

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., Dolman, A. J. 2011. Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci. 15: 453–469. https://doi.org/10.5194/hess-15-453-2011

Miyata, A., Leuning, R., Denmead, O.T., Kim, J., Harazono, Y., 2000. Carbon dioxide and methane fluxes from an intermittently flooded paddy field. Agric. For. Meteorol. 102, 287–303. https://doi.org/10.1016/S0168-1923(00)00092-7

Molden, D., Sakthivadivel, R. 1999. Water Accounting to Assess Use and Productivity of Water. Int. J. of Water Resour. Dev. 15 (1–2): 55–71. https://doi.org/10.1080/07900629948934.

Molden, D., Oweis, T., Steduto, P., Bindraban, P., Hanjra, M.A., Kijne, J. 2010. Improving Agricultural Water Productivity: Between Optimism and Caution. Agric. Water Manag. 97 (4): 528–35. https://doi.org/10.1016/j.agwat.2009.03.023.

Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9, 747–766. https://doi.org/10.2307/2401901

Monteith, J.L., 1977. Climate and the efficiency of crop production in Britain. Phil. Trans. Roy. Soc. L. 281, 277–294. https://doi.org/10.1098/rstb.1977.0140

Moorhead, J.E., 2015. Lysimetric Evaluation of Eddy Covariancec and Scintillometer Systems for the Texas High Plains (Doctoral dissertation). Texas Tech University, Lubbock.

Moorhead, J.E., Marek, G.W., Colaizzi, P.D., Gowda, P.H., Evett, S.R., Brauer, D.K., Marek, T.H., Porter, D.O., 2017. Evaluation of sensible heat flux and evapotranspiration estimates using a surface layer scintillometer and a large weighing lysimeter. Sensors. 17, 2350. https://doi.org/10.3390/s17102350

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 Sens. Environ. 49, 246-263. https://doi.org/10.1016/0034-4257(94)90020-5

Morisette, J. T., Privette, J. L., Justice C. O., Running S. W. 1998. MODIS Land Validation Plan. (September). Available at: https://modis.gsfc.nasa.gov/data/atbd/land_val.pdf

Morisette, J.T., Privette, J.L., Justice, C.O., 2002. A framework for the validation of MODIS Land products. Remote Sens. Environ. 83, 77–96. https://doi.org/10.1016/S0034-4257(02)00088-3

Moureaux, C., Ceschia, E., Arriga, N., Béziat, P., Eugster, W., Kutsch, W.L., Pattey, E., 2012. Eddy covariance measurements over crops, in: Aubinet, M., Vesala, T., Papale, D. (Eds.), Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer, New York, pp. 319–331.

Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 111, 519-536. https://doi.org/10.1016/j.rse.2007.04.015

Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019

Mu, Q., Zhao, M., Running, S. W. 2013. Algorithm Theoretical Basis Document: MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3) Collection 5. NASA Headquarters. Available at: https://modis-land.gsfc.nasa.gov/pdf/MOD16ATBD.pdf

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J.B., Guo, Z., Jung, M., Maignan, F., McCabe, M.F., Reichle, R., Reichstein, M., Rodell, M.,

Sheffield, J., Teuling, A.J., Wang, K., Wood, E.F., Zhang, Y. 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. Geophys. Resear. Letters. 38 (6): 1–7. https://doi.org/10.1029/2010GL046230

Murphy, J., Casley, D.J., Curry, J.J., 1991. Farmers' Estimations as a Source of Production Data: Methodological Guidelines for Cereals in Africa. World Bank technical paper; no. WTP 132., Africa Technical Department series, The World Bank, Washington. http://documents.worldbank.org/curated/en/313721468002132955/Farmers-estimations-as-a-source-of-production-data-methodological-guidelines-for-cereals-in-Africa

Muthoni, F. K., Odongo, V. O., Ochieng, J., Mugalavai, E. M., Mourice, S. K., Hoesche-Zeledon, I., Mwila, M., Bekunda, M. 2019. Long-term spatial-temporal trends and variability of rainfall over Eastern and Southern Africa. Theor. and Appl. Climatol. 137: 1869–1882. https://doi.org/10.1007/s00704-018-2712-1

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ. 83, 214-231. https://doi.org/10.1016/S0034-4257(02)00074-3

Nagler, P.L., Glenn, E.P., Nguyen, U., Scott, R.L., Doody, T., 2013. Estimating riparian and agricultural actual evapotranspiration by reference evapotranspiration and MODIS enhanced vegetation index. Remote Sens. 5, 3849–3871. https://doi.org/10.3390/rs5083849

Neale, C.M.U., Bausch, C.W., Heermann, D.F., 1989. Development of Reflectance-Based Crop Coefficients for Corn. Trans. ASAE. 32, 1891-1900. https://doi.org/10.13031/2013.31240

Ng, W., Dash, M., 2006. An evaluation of progressive sampling for imbalanced data sets. Proc. - IEEE Int. Conf. Data Mining, ICDM 657–661. https://doi.org/10.1109/ICDMW.2006.28

Noellemeyer, E., Fernández, R., Quiroga, A. 2013. Crop and Tillage Effects on Water Productivity of Dryland Agriculture in Argentina. Agric. 3 (1): 1–11. https://doi.org/10.3390/agriculture3010001.

Norman, D.W., Worman, F.D., Siebert, J.D., Modiakgotla, E., 1995. The farming systems approach to development and appropriate

technology generation. FAO, Rome. http://www.fao.org/3/v5330e/V5330e00.htm

Nouri, H., Glenn, E.P., Beecham, S., Boroujeni, S.C., Sutton, P., Alaghmand, S., Noori, B., Nagler, P., 2016. Comparing three approaches of evapotranspiration estimation in mixed urban vegetation: Field-based, remote sensing-based and observational-based methods. Remote Sens. 8, 492. https://doi.org/10.3390/rs8060492

Nouri, H., Stokvis, B., Borujeni, S.C., Galindo, A., Brugnach, M., Blatchford, M.L., Alaghmand, S., Hoekstra, A.Y. 2020a. Reduce blue water scarcity and increase nutritional and economic water productivity through changing the cropping pattern in a catchment. J. Hydrol. 588, 125086. https://doi.org/10.1016/j.jhydrol.2020.125086

Nouri, H., Nagler, P., Chavoshi, S., Noori, B., Barreto, A., Sina, M., Galindo, A. 2020b. Effect of spatial resolution of satellite images on estimating the greenness and evapotranspiration of urban green spaces. Hydrol. Process., 3183–3199. https://doi.org/10.1002/hyp.13790

Odi-Lara, M., Campos, I., Neale, C.M.U., Ortega-Farías, S., Poblete-Echeverría, C., Balbontín, C., Calera, A., 2016. Estimating evapotranspiration of an apple orchard using a remote sensing-based soil water balance. Remote Sens. 8, 253. https://doi.org/10.3390/rs8030253

Odongo, V. O., Van der Tol., C., Becht, R., Hoedjes, J. C. B., Ghimire, C. P., Su, Z. 2016. Energy partitioning and its controls over a heterogeneous semi-arid shrubland ecosystem in the Lake Naivasha Basin, Kenya. Ecohydrology. 9 (7): 1358–1375. https://doi.org/10.1002/eco.1732

Oweis, T., Hachum, A. 2003. Improving Water Productivity in the Dry Areas of West Asia and North Africa. In: Water productivity in agriculture: limits and opportunities for improvement. CABI Publishing. USA.

Pax-Lenney, M., Woodcock, C.E. 1997 The effect of spatial resolution on the ability to monitor the status of agricultural lands. Remote Sens. Environ. 1997, 61, 210–220. https://doi.org/10.1016/S0034-4257(97)00003-5

Paw, K. T. 1992. A discussion of the Penman form equations and comparisons of some equations to estimate latent energy flux density. Agric. For. Meteorol. 57 (4): 297–304. https://doi.org/10.1016/0168-1923(92)90125-N

Paw, U. K. T., Gao, W. 1988. Applications of solutions to non-linear energy budget equations. Agric. For. Meteorol. 43: 121–145. https://doi.org/10.1016/0168-1923(88)90087-1

Payton, M.E., 1996. Confidence Intervals for the Coefficient of Variation. Conf. Appl. Stat. Agric. https://doi.org/10.4148/2475-7772.1320

Pelgrum, H., Miltenburg, I.J., Cheema, M.J.M., Klaasse, A., Bastiaanssen, W.G.M., 2012. ET Look: A novel continental evapotranspiration algorithm, in: Proceedings of a symposium held at Jackson Hole, Wyoming, USA, September 2010, IAHS-AISH Publ. 352, 120–123.

Peng, K., Obradovic, Z., Vucetic, S., 2004. Towards efficient learning of neural network ensembles from arbitrarily large datasets. Front. Artif. Intell. Appl. 110, 623–627.

Pinter, P.J.J., Hatfield, J.L.L., Schepers, J.S.S., Barnes, E.M.M, Moran, M.S.S., Daughtry, C.S.T, Upchurch, R. 2003. Remote Sensing for Crop Management. Photogramm. Eng. Remote Sens. 69 (6): 647–64. https://doi.org/10.1109/IGARSS.2007.4423218.

Pinty, B., Lavergne, T., Dickinson, R.E., Widlowski, J.L., Gobron, N., Verstraete, M.M., 2006. Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate models. J Geophys. Res. Atmos. 111, 1-20, https://doi.org/10.1029/2005JD005952

Poate, D., 1988. A review of methods for measuring crop production from smallholder procedures. Exp. Agric. 24, 1–14. https://doi.org/10.1017/S0014479700015659

Poate, D., Casley, D.J., 1985. Estimating Crop Production in Development Projects: Methods and Their Limitations (English). The World Bank, Washington. http://documents.worldbank.org/curated/en/271251468759322233/Estimating-crop-production-in-development-projects-methods-and-their-limitations

Povey, A.C., Grainger, R.G., 2015. Known and unknown unknowns: Uncertainty estimation in satellite remote sensing. Atmos. Meas. Tech. 8, 4699–4718. https://doi.org/10.5194/amt-8-4699-2015

Prince, S.D., Haskett, J., Steininger, M., Strand, H., Wright, R., 2001. Net Primary Production of U.S. Midwest Croplands from Agricultural Harvest Yield Data. Ecol. Appl. 11, 1194–1205. https://doi.org/10.1890/1051-0761(2001)011[1194:NPPOUS]2.0.CO;2

Provost, F., Jensen, D., Oates, T., 1999. Efficient progressive sampling. Proc. fifth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '99 23–32. https://doi.org/10.1145/312129.312188

Provost, F., Jensen, D., Oates, T., 2001. Progressive sampling, in: Liu, H., Motodo, H. (Eds.), Instance Selection and Construction for Data Mining. Springer Science+Business Media, Dordrecht, p. 448.

Prueger, J.H., Hatfield, J.L., Aase, J.K., Pikul Jr., J.L., 1997. Bowen-Ratio Comparison with Lysimeter Evapotranspiration. Agron. J. 89, 730–736. https://doi.org/10.2134/agronj1997.00021962008900050004x

Pumpanen, J., Ilvesniemi, H., Perämäki, M., Hari, P., 2003. Seasonal patterns of soil CO2 efflux and soil air CO2 concentration in a Scots pine forest: comparison of two chamber techniques. Glob. Chang. Biol. 9, 371–382. https://doi.org/10.1046/j.1365-2486.2003.00588.x

Raes, D., 2017. AquaCrop training handbooks. Book I. Understanding AquaCrop (Manual). FAO, Rome. http://www.fao.org/3/a-i6051e.pdf

Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2018. Reference Manual, Chapter 3 – AquaCrop, Version 6.0 – 6.1 May 2018. FAO, Rome, pp 141. http://www.fao.org/3/a-br248e.pdf

Rajan, A., Kuang, Y.C., Ooi, M.P.-L., Demidenko, S.N., 2017. Moments and Maximum Entropy Method for Expanded Uncertainty Estimation in Measurements. IEEE, pp. 3–8.

Ramier, D., Boulain, N., Cappelaere, B., Timouk, F., Rabanit, M., Lloyd, C. R., Boubkraoui, S., Métayer, F., Descriox, L., Wawrzyniak, V. 2009. Towards an understanding of coupled physical and biological

processes in the cultivated Sahel - 1. Energy and water. J. Hydrol. 375 (1–2): 204–216. https://doi.org/10.1016/j.jhydrol.2008.12.002

Ramírez-Cuesta, J.M., Allen, R.G., Zarco-Tejada, P.J., Kilic, A., Santos, C., Lorite, I.J. 2019. Impact of the spatial resolution on the energy balance components on an open-canopy olive orchard. International Journal of Applied Earth Observation and Geoinformation, 74, 88–102. https://doi.org/10.1016/j.jag.2018.09.001

Ran, Y., Li, X., Sun, R., Kljun, N., Zhang, L., Wang, X., Zhu, G., 2016. Spatial representativeness and uncertainty of eddy covariance carbon flux measurements for upscaling net ecosystem productivity to the grid scale. Agric. For. Meteorol. 230–231, 114–127. https://doi.org/10.1016/j.agrformet.2016.05.008

Rana, G., Katerji, N. 1998. A Measurement Based Sensitivity Analysis of the Penman-Monteith Actual Evapotranspiration Model for Crops of Different Height and in Contrasting Water Status. Theor. Appl. Climatol. 60: 141–149. https://doi.org/10.1007/s007040050039

Rana, G., Katerji, N., 2000. Measurement and estimation of actual evapotranspiration in the field under Mediterranean climate: A review. Eur. J. Agron. 13, 125–153. https://doi.org/10.1016/S1161-0301(00)00070-8

Rauhala, U.A., 1989. Compiler positioning system: an array algebra formulation of digital photogrammetry. Photogramm. Eng. Remote Sens. 55, 317–326.

Raupach, M.R., 1998. Influences of local feedbacks on land-air exchanges of energy and carbon. Glob. Chang. Biol. 4, 477–494. https://doi.org/10.1046/j.1365-2486.1998.t01-1-00155.x

Reed, G.F., Lynn, F., Meade, B.D., 2002. Quantitative Assays 9, 1235–1239. https://doi.org/10.1128/CDLI.9.6.1235

Rees, D., Farrell, G., Orchard, J., 2012. Potatoes, in: Crop Post-Harvest: Science and Technology, Volume 3: Perishables. Blackwell Publishing Ltd, West Sussex, pp. 334–359.

Reeves, M.C., Zhao, M., Running, S.W., 2005. Usefulness and limits of MODIS GPP for estimating wheat yield. Int. J. Remote Sens. 26, 1403–1421. https://doi.org/10.1080/01431160512331326567

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L.T.,

Kim, G., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R., Ruddick, A.G., Sienkiewicz, M., Woollen, J. 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. J. Climate 24: 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Riederer, M., Serafimovich, A., Foken, T., 2014. Net ecosystem CO2 exchange measurements by the closed chamber method and the eddy covariance technique and their dependence on atmospheric conditions. Atmos. Meas. Tech. 7, 1057–1064. https://doi.org/10.5194/amt-7-1057-2014

Rochette, P., Hutchinson, G.L., 2005. Measurement of Soil Respiration in situ: Chamber Techniques, in: Hatfield, J.L., Baker, J.M. (Eds.), Micrometeorology in Agricultural Systems. ASA Monograph. No. 47, American Society of Agronomy, Madison, pp. 247–286.

Rockström, J., Barron, J. 2007. Water Productivity in Rainfed Systems: Overview of Challenges and Analysis of Opportunities in Water Scarcity Prone Savannahs. Irrig. Sci. 25 (3): 299–311. https://doi.org/10.1007/s00271-007-0062-3.

Rockstrom, J., Lannerstad, M., Falkenmark, M. 2007. Assessing the Water Challenge of a New Green Revolution in Developing Countries. Proceedings of the National Academy of Sciences of the United States of America. 104 (15): 6253–60. https://doi.org/10.1073/pnas.0605737104.

Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L., Wetterstrand, H. DeClerck, F., Shah, M., Steduto, P., de Fraiture, C., Hatibu, N., Unver, O., Bird, J., Sibanda, L., Smith, J. 2017. Sustainable Intensification of Agriculture for Human Prosperity and Global Sustainability. Ambio 46 (1): 4–17. https://doi.org/10.1007/s13280-016-0793-6.

Rodell, M., Famiglietti J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., Lo, M. H. 2018. Emerging trends in global freshwater availability. Nature. 557: 651–659. https://doi.org/10.1038/s41586-018-0123-1

Roerink, G.J., Bastiaanssen, W.G.M., Chambouleyron, J., Menenti, M., 1997. Relating Crop Water Consumption to Irrigation Water Supply by Remote Sensing. Water Resour. Manag. 11, 445–465. https://doi.org/10.1023/A:1007982411718

Roerink, G.J., Su, Z., Menenti, M., 2000. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. Phys. Chem. Earth, Part B Hydrol. Ocean. Atmos. 25, 147–157. https://doi.org/10.1016/S1464-1909(99)00128-8

Rockström, J., Falkenmark, M., Lannerstad, M., Karlberg, L. 2012. The planetary water drama: Dual task of feeding humanity and curbing climate change. Geophys. Resear. Letters. 39 (15): 1–8. https://doi.org/10.1029/2012GL051688

Rozelle, S., 1991. Rural household data collection in developing countries: Designing instruments and methods for collecting farm production data (Working paper). Food and Nutrition Policy Program, Department of Agriculture Economics and Cornell, Washington.

Rubel, F., Kottek, M. 2010. Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. Meteorol. Zeitschrift 19 (2): 135–141. https://doi.org/10.1127/0941-2948/2010/0430

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. Bioscience 54, 547-560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Sadras, V.O., Connor, D.J., 1991. Physiological basis of the response of harvest index to the fraction of water transpired after anthesis: A simple model to estimate harvest index for determinate species. F. Crop. Res. 26, 227–239. https://doi.org/10.1016/0378-4290(91)90001-C

Sadras, V.O., Cassman, K.G., Grassini, P., Hall, A.J., Bastiaanssen, W.G.M., Laborte, A.G., Milne, A.E., Sileshi, G., Steduto, P., 2015. Yield gap analysis of field crops, Methods and case studies (FAO Water Reports 41). FAO, Rome. http://www.fao.org/3/a-i4695e.pdf

Samarasinghe, G.B., 2003. Growth and yields of Sri Lanka's major crops interpreted from public domain satellites. Agric. Water Manag. 58, 145–157. https://doi.org/10.1016/S0378-3774(02)00130-0

Sánchez, J.M., López-Urrea, R., Doña, C., Montoro, A., Caselles, V., Galve, J.M., 2016. Discrepancies between eddy covariance and

lysimeter measurements in the assessment of energy balance modeling in vineyards, in: Proceedings Volume 9998, Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII; 99980R, 25 Oct. SPIE Remote Sensing, Edinburgh. pp1-9. https://doi.org/10.1117/12.2241436

Santamaría-Bonfil, G., Fernández, N., Gershenson, C., 2016. Measuring the complexity of continuous distributions. Entropy 18. https://doi.org/10.3390/e18030072

Sarkar, A., Guo, J., Siegmund, N., Apel, S., Czarnecki, K., 2016. Cost-efficient sampling for performance prediction of configurable systems. Proc. - 2015 30th IEEE/ACM Int. Conf. Autom. Softw. Eng. ASE 2015 342–352. https://doi.org/10.1109/ASE.2015.45

Schectman, O., 2013. Methods of Clinical Epidemiology. Methods Clin. Epidemiol. 33–49. https://doi.org/10.1007/978-3-642-37131-8

Schmid, H.P., 2002. Footprint modeling for vegetation atmosphere exchange studies: A review and perspective. Agric. For. Meteorol. 113, 159–183. https://doi.org/10.1016/S0168-1923(02)00107-7

Schneider, P., Ghent, D., Prata, F., Corlett, G.K., Remedios, J.J., 2012. AATSR Validation: LST Validation Protocol. ESA Contract Number: 19054/05/NL/FF, Report for the European Space Agency.

Sebhat, M. Y., Wenninger, J. 2014. Water balance of the Juba and Shabelle Rivers basins the Horn of Africa. Int. J. Agric. Policy and Res. 2 (6): 238–255. http:/dx.doi.org/10.14303/irjas.2014.027

Seckler, D., Sampath, R.K., Raheja, S.K., 1988. An index for measuring the performance of irrigation management systems with an application. Water Resour. Bull. 24, 855–860. https://doi.org/10.1111/j.1752-1688.1988.tb00937.x

Senay, G.B., Budde, M., Verdin, J.P., Melesse, A.M., 2007. A coupled remote sensing and simplified surface energy balance approach to estimate actual evapotranspiration from irrigated fields. Sensors. 6, 979-1000. https://doi.org/10.3390/s7060979

Senay, G.B., Leake, S., Nagler, P.L., Artan, G., Dickinson, J., Cordova, J.T., Glenn, E.P., 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing

methods. Hydrol. Process. 25, 4037–4049. https://doi.org/10.1002/hyp.8379

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. J. Am. Water Resour. Assoc. 49, 577–591. https://doi.org/10.1111/jawr.12057

Senay, G.B., Friedrichs, M., Singh, R.K., Velpuri, N.M., 2016. Evaluating Landsat 8 evapotranspiration for water use mapping in the Colorado River Basin. Remote Sens. Environ. 185, 171–185. https://doi.org/10.1016/j.rse.2015.12.043

Shanahan, J., Schepers, J.S., Francis, D.D., Varvel, G.E., Wilhelm, W.W., Tringe, J.M., Schlemmer, M.R., Major, D.J., 2001. Use of Remote-Sensing Imagery to Estimate Corn Grain. Agronomy & Horticulture -- Faculty Publications. 9, 583–589. http://digitalcommons.unl.edu/agronomyfacpub/9

Shannon, C.E.C.., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/10.1145/584091.584093

Sharma V., Kilic A., Irmak S. 2016. Impact of scale/resolution on evapotranspiration from Landsat and MODIS images. Water Resour. Res. 52 (3): 1800–1819. https://doi.org/10.1002/2015WR017772

Sibley, A.M., Grassini, P., Thomas, N.E., Cassman, K.G., Lobell, D.B., 2013. Testing remote sensing approaches for assessing yield variability among maize fields. Agron. J. 106, 24–32. https://doi.org/10.2134/agronj2013.0314

Sinclair, T.R., Tanner, C.B., Bennett, J.M. 1984. Water-Use Efficiency in Crop Production. BioScience 34 (1): 36–40. https://doi.org/10.2307/1309424.

Sim, J., Reid, N., 1999. Statistical Inference by Confidence Intervals: Issues of Interpretation and Utilization. Phys. Ther. 79, 186–195. https://doi.org/10.1093/ptj/79.2.186

Sinclair, T.R., 1998. Historical Changes in Harvest Index and Crop Nitrogen Accumulation. Crop Sci. 38, 638–643. https://doi.org/10.2135/cropsci1998.0011183X003800030002x

Singh, R., 2003. Use of Satellite Data and Farmers Eye Estimate for Crop Yield Modelling. Jour. Ind. Soc. Ag. Stat. 56, 166–176. http://agris.fao.org/agris-search/search.do?recordID=IN2004000941

Singh, R.K., Senay, G.B., 2016. Comparison of four different energy balance models for estimating EvapoTranspiration in the Midwestern United States. Water. 8, 9. https://doi.org/10.3390/w8010009

Singh, R., Van Dam, J.C., Feddes, R.A., 2006. Water productivity analysis of irrigated crops in Sirsa district, India. Agric. Water Manag. 82, 253–278. https://doi.org/10.1016/j.agwat.2005.07.027

Sjöström, M., Zhao, M., Archibald, S., Arneth, A., Cappelaere, B., Falk, U., De Grandcourt, A., Hanan, N., Kergoat, L., Kutsch, W., Merbold, L., Mougin, E., Nickless, A., Nouvellon, Y., Scholes, R.J., Veenendaal, E.M., Ardö, J., 2013. Evaluation of MODIS gross primary productivity for Africa using eddy covariance data. Remote Sens. Environ. 131, 275–286. https://doi.org/10.1016/j.rse.2012.12.023

Smets, B., Lacaze, R., Freitas, S.C., Jann, A., Calvet, J.C., Camacho, F., Baret, F., Paulik, C., d'Andrimont, R., Tansey, K., others, 2013. Operating The Copernicus Global Land Service. ESA Spec. Publ. 722, 66.

Soleymani, R., Granger, E., Fumera, G., 2018. Progressive Boosting for Class Imbalance and Its Application to Face Re-Identificatio. Expert Syst. Appl. 101, 271–291. https://doi.org/10.1016/j.eswa.2018.01.023

Song, C., Dannenberg, M.P., Hwang, T., 2013. Optical remote sensing of terrestrial ecosystem primary productivity. Prog. Phys. Geogr. 37, 834–854. https://doi.org/10.1177/0309133313507944

Steduto, P., Todorovic, M., 2001. The Agro-Ecological Characterisation of Apulia Region Methodology and Experience 34, 143–164.

Steduto, P., Çetinkökü, Ö., Albrizio, R., Kanber, R., 2002. Automated closed-system canopy-chamber for continuous field-crop monitoring of CO2 and H2O fluxes. Agric. For. Meteorol. 111, 171–186. https://doi.org/10.1016/S0168-1923(02)00023-0

Steduto, P., Todorovic, M., Caliandro, A., Rubino, P. 2003. Daily reference evapotranspiration estimates by the Penman-Monteith

equation in Southern Italy. Constant vs. variable canopy resistance. Theor. and Appl. Climatol. 74 (3–4): 217–225. https://doi.org/10.1007/s00704-002-0720-6

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., Ludwig C. 2015. The Trajectory of the Anthropocene: The Great Acceleration. Anthrop. Rev. 2 (1): 81–98. https://doi.org/10.1177/2053019614564785.

Stein, A., Ettema, C., 2003. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. Agric. Ecosyst. Environ. 94, 31–47. https://doi.org/10.1016/S0167-8809(02)00013-0

Stoy, P.C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M.A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., Van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, H., Merbold, L., Montagnani, L., Papale, D., Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., Vaccari, F., Varlagin, A., 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. Agric. For. Meteorol. 171–172, 137–152. https://doi.org/10.1016/j.agrformet.2012.11.004

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. Discuss. 6, 85–100. https://doi.org/10.5194/hess-6-85-2002

Su, C.H., Ryu, D., Crow, W.T., Western, A.W., 2014. Beyond triple collocation: Applications to soil moisture monitoring. J. Geophys. Res. Atmos. 119, 6419–6439. https://doi.org/10.1002/2013JD021043

Su, Z., Timmermans, W., Zeng, Y., Schulz, J., John, V.O., Roebeling, R.A., Poli, P., Tan, D., Kaspar, F., Kaiser-Weiss, A.K., Swinnen, E., Toté, C., Gregow, H., Manninen, T., Rijhelä, A., Calvet, J.-C., Ma, Y., Wen, J., 2018. An overview of European efforts in generating climate data records. Bull. Am. Meteorol. Soc. 99, 349–359. https://doi.org/10.1175/BAMS-D-16-0074.1

Sud, U.C., Ahmad, T., Gupta, V.K., Chandra, H., Sahoo, P.M., Aditya, K., Singh, M., Biswas, A., 2016. Synthesis of Literature and Framework: Research on Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping (Working Paper No. 5). ICAR-Indian Agricultural Statistics Research Institute, New Delhi.

Sugita, M., Matsuno, A., El-Kilani, R. M. M, Abdel-Fattah, A., Mahmoud, M. A. 2017. Crop evapotranspiration in the Nile Delta under different irrigation methods. Hydrol. Sciences J. 62 (10): 1618–1635. https://doi.org/10.1080/02626667.2017.1341631

Sukhatme, P. V., 1947. The Problem of Plot Size in Large-Scale Yield Surveys. J. Am. Stat. Assoc. 42, 297–310. https://doi.org/10.1080/01621459.1947.10501928

Suyker, A.E., Verma, S.B., 2010. Coupling of carbon dioxide and water vapor exchanges of irrigated and rainfed maize-soybean cropping systems and water productivity. Agric. For. Meteorol. 150, 553–563. https://doi.org/10.1016/j.agrformet.2010.01.020

Swinnen, E., Van Hoolst, R., 2018. Gio Global Land Component - Lot I "Operation of the Global Land Component"- Framework Service Contract N° 388533 (JRC). Algorithm Theoretical Basis Document, Dry Matter Productivity. Collection 300 m, Version 1. VITO, EC Copernicus Global Land, Brussels. https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/GIOGL1_ATBD_DMP300m-V1_I1.11.pdf

Swinnen, E., Van Hoolst, R., Toté, C., 2015. Gio Global Land Component - Lot I "Operation of the Global Land Component". Framework Service Contract N° 388533 (JRC). Quality Assessment Report. Dry Matter Productivity (DMP) - PROBA-V., Algorithm Theoretical Basis Document, Issue I1.01. VITO, EC Copernicus Global Land, Brussels. https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/GIOGL1_QAR_DMP300m_V1_I1.12.pdf

Tagesson, T., Fensholt, R., Guiro, I., Rasmussen, M. O., Huber, S, Mbow, C., Garcia, M., Horion, S., Sandholt, I., Holm-Rasmussen, B., Göttsche, F.M., Ridler, M., Olén , N., Olsen, J.L., Ehammer, A.E., Madsen, M., Olesen, F.S. Ardö, J. 2015. Ecosystem properties of semiarid savanna grassland in West Africa and its relationship with environmental variability. Glob. Chang. Biol. 21 (1): 250–264. https://doi.org/10.1111/gcb.12734

Taghvaeian, S., Neale, C.M.U., Osterberg, J.C., Sritharan, S.I., Watts, D.R. 2018. Remote Sensing and GIS Techniques for Assessing Irrigation Performance: Case Study in Southern California. J. Irrig. Drain. Eng. 144, 05018002. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001306

Tang, X., Ding, Z., Li, H., Li, X., Luo, J., Xie, J., Chen, D., 2015. Characterizing ecosystem water-use efficiency of croplands with eddy covariance measurements and MODIS products. Ecol. Eng. 85, 212–217. https://doi.org/10.1016/j.ecoleng.2015.09.078

Tanner, C.B., Sinclair, R. 1983. Efficient Water Use in Crop Production: Research or Re-Search? Limitations to Efficient Water Use in Crop Production, 1–27. https://doi.org/10.2134/1983.limitationstoefficientwateruse.c1.

Taylor, J.R., 1997. An Introduction to Error Analysis. University Science Books, Sausalito.

Teillet, P.M., Staenz, K., Williams, D.J. 1997. Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. Remote Sens. Environ. 1997, 61, 139–149. https://doi.org/10.1016/S0034-4257(96)00248-9

Teixeira, A.H. de C., Bastiaanssen, W.G.M., Ahmad, M.D., Bos, M.G. 2009. Reviewing SEBAL input parameters for assessing evapotranspiration and water productivity for the Low-Middle San Francisco River basin, Brazil. Part B: Application to the regional scale. Agric. For. Meteorol. 149, 477–490. https://doi.org/10.1016/j.agrformet.2008.09.014

The Nile Basin Initiative Secretariat. 2014. Nile Waters Technical Bulletin from The Nile Basin Initiative Secretariat, Understanding Nile Basin Hydrology Mapping Actual Evapotranspiration over the Nile Basin. Entebbe, Ethiopia. Available at: www.nilebasin.org

Thiruvengadachari, S., Sakthivadivel, R., 1997. Satellite remote sensing for assessment of irrigation system performance: A case study in India. Research Report 9, International Irrigation Management Institute (IIMI), Colombo, Sri Lanka. http://dx.doi.org/10.3910/2009.017

Todd, R., Evett, S.R., Howell, T.A., 2000. The Bowen ratio-energy balance method for estimating latent heat flux of irrigated alfalfa evaluated in a semi-arid, advective environment. Agric. For. Meteorol. 103, 335–348. https://doi.org/10.1016/S0168-1923(00)00139-8

Trigo, I.F., Dacamara, C.C., Viterbo, P., Roujean, J.-L., Olesen, F., Barroso, C., Camacho-de-Coca, F., Carrer, D., Freitas, S.C., García-

Haro, J., Geiger, B., Gellens-Meulenberghs, F., Ghilain, N., Meliá, J., Pessanha, L., Siljamo, N., Arboleda, A., 2011. The Satellite Application Facility on Land Surface. Int. J. Remote Sens. 32, 2725–2744. https://doi.org/10.1080/01431161003743199

Tucker, C.J., Sellers, P.J., 1986. Satellite remote sensing of primary production. Int. J. Remote Sens. 7, 1395–1416. https://doi.org/10.1080/01431168608948944

Tucker, C.J., Holben, B.N., Elgin, J.H., McMurtrey, J.E., 1980. Relationship of spectral data to grain yield variation. Photogram. Eng. Remote Sens. 46, 657-666. https://www.cabdirect.org/cabdirect/abstract/19810714985

Turner, D.P., Ritts, W.D., Cohen, W.B., Maeirsperger, T.K., Gower, S.T., Kirschbaum, A.A., Running, S.W., Zhao, M., Wofsy, S.C., Dunn, A.L., Law, B.E., Campbell, J.L., Oechel, W.C., Kwon, H.J., Meyers, T.P., Small, E.E., Kurc, S.A., Gamon, J.A., 2005. Site-level evaluation of satellite-based global terrestrial gross primary production and net primary production monitoring. Glob. Chang. Biol. 11, 666–684. https://doi.org/10.1111/j.1365-2486.2005.00936.x

Uddin, J., Hancock, N.H., Smith, R.J., Foley, J.P., 2013. Measurement of evapotranspiration during sprinkler irrigation using a precision energy budget (Bowen ratio, eddy covariance) methodology. Agric. Water Manag. 116, 89–100. https://doi.org/10.1016/j.agwat.2012.10.008

Umarani, V., Punithavalli, M., 2011. Analysis of the progressive sampling-based approach using real life datasets. Open Comput. Sci. 1, 221–242. https://doi.org/10.2478/s13537-011-0016-y

UN, 2016. Sustainable Development Goals. https://sustainabledevelopment.un.org/sdgs (accessed 30 August 2020).

Unkovich, M., Baldock, J., Forbes, M., 2010. Chapter 5 - Variability in Harvest Index of Grain Crops and Potential Significance for Carbon Accounting, Adv. Agron. 105, 173-219. https://doi.org/10.1016/S0065-2113(10)05005-4

USAID. 2016. Southern Africa Drought - Fact Sheet #4 - Fiscal Year (FY) 2016. 6 May 2016 (pp 1–9). Available at: https://www.usaid.gov/sites/default/files/documents/1866/southern_africa_dr_fs02_05-06-2016.pdf

Valiant, L.G., 1984. A theory of the learnable. Commun. ACM 27, 1134–1142. https://doi.org/10.1145/1968.1972

Van Dam, J.C., Singh, R., Bessembinder, J.J.E., Leffelaar, P.A., Bastiaanssen, W.G.M., Jhorar, R.K., Kroes, J.G., Droogers, P., 2006. Assessing Options to Increase Water Productivity in Irrigated River Basins Using Remote Sensing and Modelling Tools. Water Resour. Dev. 22, 115–133. https://doi.org/10.1080/07900620500405734

Van der Meer, F., 2012. Remote-sensing image analysis and geostatistics. Int. J. Remote Sens. 33, 5644–5676. https://doi.org/10.1080/01431161.2012.666363

Van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6, 3109–3129. https://doi.org/10.5194/bg-6-3109-2009

Van Wart, J., van Bussel, L.G.J., Wolf, J., Licker, R., Grassini, P., Nelson, A., Boogaard, H., Gerber, J., Mueller, N.D., Claessens, L., et al. 2013. Use of agro-climatic zones to upscale simulated crop yield potential. F. Crop. Res. 143, 44–55. https://doi.org/10.1016/j.fcr.2012.11.023

Vanikiotis, T.; Stagakis, S.; Kyparissis, A. Effects of satellite spatial resolution on gross primary productivity estimation through light use efficiency modeling. In Proceedings of the Proc. SPIE 10773, Sixth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2018) (6 August 2018); 2018; p. 107731R.

Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P. 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139: 35-39. https://doi.org/10.1016/j.rse.2013.07.013

Verger, A., Baret, F., Weiss, M., 2017. Algorithm Theoretical Basis Document: Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Fraction of green Vegetation Cover (FCover), Collection 1km, Version 2, Issue 1.41. https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_FAPAR1km-V2_I1.41.pdf

Verma, V., Marchant, T., Scott, C., 1988. Evaluation of crop-cut methods and farmer reports for estimating crop production. Results of a methodological study in five African countries. Longacre, London. https://ageconsearch.umn.edu/record/257587

Verstraeten, W.W., Veroustraete, F., Feyen, J., 2008. Assessment of Evapotranspiration and Soil Moisture Content Across Different Scales of Observation. Sensors. 8, 70–117. https://doi.org/10.3390/s8010070

Viets, F.G. Jr. 1962. Fertiliser and the Efficient Use of Water. Adv. Agron. 14: 223–64.

Voisin, N., Wood, A. W., Lettenmaier, D. P. 2008. Evaluation of Precipitation Products for Global Hydrological Prediction. J. Hydrometeorol. 9: 388–407. https://doi.org/10.1175/2007JHM938.1

Wagner, S.W., Reicosky, D.C., 1992. Closed-Chamber Effects on Leaf Temperature, Canopy Photosynthesis, and Evapotranspiration. Agron. J. 84, 731-738. https://doi.org/10.2134/agronj1992.00021962008400040035x

Wang, D., Alimohammadi, N. 2012. Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale. Water Resourc. Res. 48: W05546. https://doi.org/10.1029/2011WR011444

Wang, K., Lui, C., Zheng, X., Pihlatie, M., Li, B., Haapanala, S., Vesala, T., Lui, H., Wang, Y., Liu, G., Hu, F., 2013. Comparison between eddy covariance and automatic chamber techniques for measuring net ecosystem exchange of carbon dioxide in cotton and wheat fields. Biogeosciences 10, 6865–6877. https://doi.org/10.5194/bg-10-6865-2013

Wang, X., Ma, M., Li, X., Yi, S., Tan, J., Huang, G., Zhang, Z., Zhao, T., Feng, J., Ma, Z., Wei, W., Bai, Y., 2013. Validation of MODIS-GPP product at 10 flux sites in northern China. Int. J. Remote Sens. 34, 587–599. https://doi.org/10.1080/01431161.2012.715774

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., Du, Q., 2017. Evaluation of the latest MODIS GPP products across multiple biomes using global eddy covariance flux data. Remote Sens. 9, 418. https://doi.org/10.3390/rs9050418

Wang, Y., Zhang, X., Xiao, X., Heitman, J., Horton, R., Ren, T., 2017. An empirical calibration for heat-balance sap-flow sensors in

maize. Agron. J. 109, 1122–1128. https://doi.org/10.2134/agronj2016.10.0611

Wang, J., Ma, J., Clarke-Sather, A., Qu, J., 2018. Estimating changes in the green water productivity of cropping systems in Northern Shaanxi Province in China's Loess Plateau. Water 10, 1198. https://doi.org/10.3390/w10091198

Warmuth, M.K., Liao, J., Ratsch, G., Mathieson, M., Putta, S., Lemmen, C., 2003. Active learning with support vector machines in the drug discovery process. J Chem Inf Comput Sci 43, 667-673.

Wiegand, C.L., Richardson, A.J., 1990. Use of spectral vegetation indexes to infer leaf-area, evapotranspiration and yield. 2. Results. Agron. J. 82, 630–636. https://www.cabdirect.org/cabdirect/abstract/19900737878

Wiegand, C.L., Richardson, A.J., Escobar, D.E., Gerbermann, A.H., 1991. Vegetation indices in crop assessments. Remote Sens. Environ. 35, 105–119. https://doi.org/10.1016/0034-4257(91)90004-P

Wichelns, D. 2015. Virtual Water and Water Footprints Do Not Provide Helpful Insight Regarding International Trade or Water Scarcity. Ecol. Indic. 52: 277–83. https://doi.org/10.1016/j.ecolind.2014.12.013.

Wilks, D., 2006. Forecast Verification, in: Wilks, D. (Ed.), Statistical Methods in the Atmospheric Sciences. Elsevier, pp. 255–335.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D.D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B.E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Verma, S., Valentini, R., 2002. Energy balance closure at FLUXNET sites. Agric. For. Meteorol. 113, 223–243. https://doi.org/10.1016/S0168-1923(02)00109-0

WMO, 2011. Systematic Observation Requirements For Satellite-Based Data Products for Climate - 2011 Update. Supplemental details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)", GCOS- No. 154. World Meteorological Organization, Geneva. https://library.wmo.int/doc_num.php?explnum_id=3710

Wohlfahrt, G., Klumpp, K., Soussana, J.-F., 2012. Eddy Covariance Measurements over Grasslands, in: Aubinet, M., Vesala, T., Papale, D. (Eds.), Eddy Covariance : A Practical Guide to Measurement and Data Analysis. Springer, New York, pp. 333–344.

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B., Ojima, D., 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. Remote Sens. Environ. 91, 256–270. https://doi.org/10.1016/j.rse.2004.03.010

Xin, Q., Broich, M., Suyker, A.E., Yu, L., Gong, P., 2015. Multi-scale evaluation of light use efficiency in MODIS gross primary productivity for croplands in the Midwestern United States. Agric. For. Meteorol. 201, 111–119. https://doi.org/10.1016/j.agrformet.2014.11.004

Yan, H., Wang, S., Billesbach, D., Oechel, W., Bohrer, G., Meyers, T., Martin, T.A., Matamala, R., Phillips, R.P., Rahman, F., Yu, Q., Shugart, H.H., 2015. Improved global simulations of gross primary product based on a new definition of water stress factor and a separate treatment of C3 and C4 plants. Ecol. Modell. 297, 42–59. https://doi.org/10.1016/j.ecolmodel.2014.11.002

Yee, M.S., Pauwels, V.R.N., Daly, E., Beringer, J., Rüdiger, C., McCabe, M.F., Walker, J.P., 2015. A comparison of optical and microwave scintillometers with eddy covariance derived surface heat fluxes. Agric. For. Meteorol. 213, 226–239. https://doi.org/10.1016/j.agrformet.2015.07.004

Yilma, W.A., 2017. Computation and Spatial Observation of Water Productivity in Awash River Basin (Thesis WSE-HELWD.17.08). IHE Delft, Delft.

Young, K.D., Lewis, R.J., 1997. What Is Confidence? Part 1: The Use and Interpretation of Confidence Intervals. Ann. Emerg. Med. 30, 307–310. https://doi.org/10.1016/S0196-0644(97)70166-5

Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L.L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A.H., Goulden, M.L., Hollinger, D.Y., Hu, Y., Law, B.E., Stoy, P.C., Vesala, T., Wofsy, S.C., 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. Agric. For. Meteorol. 143, 189–207. https://doi.org/10.1016/j.agrformet.2006.12.001

Yuan, W., Liu, S., Yu, G., Bonnefond, J., Chen, J., Davis, K., Desai, A.R., Goldstein, A.H., Gianelle, D., Rossi, F., Suyker, A.E., Verma, S.B., 2010. Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. Remote Sens. Environ. 114, 1416–1431. https://doi.org/10.1016/j.rse.2010.01.022

Yuan, W., Cai, W., Nguy-Robertson, A.L., Fang, H., Suyker, A.E., Chen, Y., Dong, W., Liu, S., Zhang, H., 2015. Uncertainty in simulating gross primary production of cropland ecosystem from satellite-based models. Agric. For. Meteorol. 207, 48–57. https://doi.org/10.1016/j.agrformet.2015.03.016

Yuan, W., Chen, Y., Xia, J., Dong, W., Magliulo, V., Moors, E., Olesen, J.E., Zhang, H., 2016. Estimating crop yield using a satellite-based light use efficiency model. Ecol. Indic. 60, 702–709. https://doi.org/10.1016/j.ecolind.2015.08.013

Zanotelli, D., Montagnani, L., Manca, G., Tagliavini, M., 2013. Net primary productivity, allocation pattern and carbon use efficiency in an apple orchard assessed by integrating eddy covariance, biometric and continuous soil chamber measurements. Biogeosciences 10, 3089–3108. https://doi.org/10.5194/bg-10-3089-2013

Zeng, Y., Su, Z., Calvet, J.-C., Manninen, T., Swinnen, E., Schulz, J., Roebeling, R., Poli, P., Tan, D., Riihelä, A., Tanis, C.-M., Arslan, A.-N., Obregon, A., Kaiser-Weiss, A., John, V.O., Timmermans, W., Timmermans, J., Kaspar, F., Gregow, H., Barbu, A.-L., Fairbairn, D., Gelati, E., Meurey, C., 2015. Analysis of current validation practices in Europe for space-based climate data records of essential climate variables. Int. J. Appl. Earth Obs. Geoinf. 42, 150–161. https://doi.org/10.1016/j.jag.2015.06.006

Zeng, Y., Su, Z., Van Der Velde, R., Wang, L., Xu, K., Wang, X., Wen, J. 2016. Blending satellite observed, model simulated, and in situ measured soil moisture over Tibetan Plateau. Remote Sens. 8, 1–22. doi:10.3390/rs8030268.

Zeng, X., Luo, G., 2017. Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. Heal. Inf. Sci. Syst. 5, 2. https://doi.org/10.1007/s13755-017-0023-z

Zeng, Y.., Su, Z., Barmpadimos, I., Perrels, A., Poli, P., Boersma, K.F., Frey, A., Ma, X., de Bruin, K., Goosen, H., John, V.O., Roebeling, R., Schulz, J., Timmermans, W., 2019. Towards a Traceable Climate Service: Assessment of Quality and Usability of Essential Climate Variables. Remote Sens. 11, 1186. https://doi.org/https://doi.org/10.3390/rs11101186

Zhang, Y., Leuning, R., Chiew, F. H. S., Wang, E., Zhang, L., Liu, C., Sun, F., Peel, M.C., Shen, Y., Jung, M. 2012. Decadal Trends in Evaporation from Global Energy and Water Balances. J. Hydrometeorol. 13 (1): 379–391. https://doi.org/10.1175/JHM-D-11-012.1

Zhang, Z., Tian, F., Hu, H., Yang, P., 2014. A comparison of methods for determining field evapotranspiration: Photosynthesis system, sap flow, and eddy covariance. Hydrol. Earth Syst. Sci. 18, 1053–1072. https://doi.org/10.5194/hess-18-1053-2014

Zhang, L.X., Zhou, D.C., Fan, J.W., Hu, Z.M., 2015. Comparison of four light use efficiency models for estimating terrestrial gross primary production. Ecol. Modell. 300, 30–39. https://doi.org/10.1016/j.ecolmodel.2015.01.001

Zhao, M., Heinsch, F.A., Nemani, R.R., Running, S.W., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. Remote Sens. Environ. 95, 164–176. https://doi.org/10.1016/j.rse.2004.12.011

Zhuang, R., Zeng, Y., Manfreda, S., Su, Z. 2020. Quantifying long-term land surface and root zone soil moisture over Tibetan plateau. Remote Sens. 12, 1–20. doi:10.3390/rs12030509.

Zoebl, D. 2006. Is Water Productivity a Useful Concept in Agricultural Water Management? Agric. Water Manag. 84 (3), 265–73. https://doi.org/10.1016/j.agwat.2006.03.002.

Zwart, S.J., Bastiaanssen, W.G.M., 2004. Review of measured crop water productivity values for irrigated wheat, rice, cotton and maize. Agric. Water Manag. 69, 115–133. https://doi.org/10.1016/j.agwat.2004.04.007

Zwart, S.J., Bastiaanssen, W.G.M., 2007. SEBAL for detecting spatial variation of water productivity and scope for improvement in eight irrigated wheat systems. Agric. Water Manag. 89, 287–296. https://doi.org/10.1016/j.agwat.2007.02.002

Zwart, S.J., Leclert, L.M.C., 2010. A remote sensing-based irrigation performance assessment: A case study of the Office du Niger in Mali. Irrig. Sci. 28, 371–385. https://doi.org/10.1007/s00271-009-0199-3

Zwart, S.J., Bastiaanssen, W.G.M., De Fraiture, C., Molden, D.J., 2010. A global benchmark map of water productivity for rainfed and irrigated wheat. Agric. Water Manag. 97, 1617–1627. https://doi.org/10.1016/j.agwat.2010.05.018

Summary

The role of water in agriculture is essential and plays an important role in food security. We need to better understand how to optimise water use in agriculture to meet both global food security and water management efficiency goals. Earth observation with satellite offers the opportunity to map, monitor and better understand the dynamics of agricultural water. However, the applicability and accuracy of these observations need to be better understood. This dissertation aims to better understand the suitability of large remote sensing-based datasets for the monitoring of crop water productivity (CWP). CWP being an indicator of agricultural water efficiency.

The dissertation is composed of seven chapters.

Chapter 1 is introductory and describes; the importance of CWP in agriculture, how CWP can be monitored and the importance of understanding the accuracy of remote sensing based CWP estimates.

Chapter 2 assesses the accuracy of remotely sensed CWP against the accuracy of estimated in-situ CWP. The accuracy of CWP based on in-situ methods, which are assumed to be the user's benchmark for CWP accuracy, and the CWP derived from remote sensing methods are reviewed. CWP from non-locally parametrised studies can achieve accuracy within the error bounds of in-situ methods. However, it was shown that Uncertainty varies significantly dependent on method and application.

Chapter 3 evaluates the accuracy of a large remote sensing-based evapotranspiration dataset, the denominator in the CWP indicator, over Africa. The evaluation uses multiple methods to compensate for sparse ground-truth data. The dataset showed mixed results at point, daily scale to annual, basin scale, which were similar to those reported by other literature in Chapter 2. The research highlighted the risks in using these large datasets in CWP monitoring, due to the high variation in accuracy of a large dataset.

Chapter 4 quantifies the suitability of varying remote sensing-based resolutions for application in agricultural productivity. Three performance indicators, adequacy, equity and productivity are tested in five irrigation schemes for three spatial resolutions, 250m, 100m and 30m. This is frequently resulting in large differences in the irrigation performance assessment criteria for inter-plot comparisons. This is particularly noticed in irrigation schemes with the smallest field

sizes. This highlights the importance of selecting the spatial resolution appropriate for to scheme characteristics when undertaking irrigation performance assessment using remote sensing input.

Chapter 5 quantifies a sample size to improve processing time of validation activities of large continental CWP datasets. A progressive sampling approach, as typically applied in machine learning to train algorithms, combined with two performance measures, was applied to estimate the required sample size. The proposed approach can significantly reduce the processing time while still providing a statistically valid representation of a large remote sensing dataset. This can be useful as more high-resolution remote sensing data becomes available.

Chapter 6 reflects on the implications of the CWP concept and how the concept itself may at times contradict the global food security and water management efficiency goals. The CWP concept can be a useful one, however, it can also be misleading. On both a local and global scale, users must reflect on the goals and select or prioritise indicators, improving CWP may not always be the most useful indicator. We need to focus on the sustainable improvement in CWP rather than the increase in CWP for sustainable agriculture.

Future studies should show how the interactions between carbon and transpiration can better be utilised as a tool in understanding the quality of agricultural water monitoring datasets.

Samenvatting

De rol van water in de landbouw is essentieel en speelt een belangrijke rol in de voedselzekerheid. We moeten beter begrijpen hoe we het watergebruik in de landbouw kunnen optimaliseren om te voldoen aan de wereldwijde doelstellingen voor voedselzekerheid. Naast voedselproductie is efficiëntie in watergebruik in de landbouw ook belangrijk en essentieel om te voldoen aan int'le doelstellingen (i.e. "SDG") voor de globale waterzekerheid. Aardobservatie met satellieten biedt de mogelijkheid om de dynamiek van het watergebruik in de landbouw beter in kaart te brengen, in ruimte en tijd. Hierdoor kan ook beter toezicht gehouden worden op watergebruik in bijvoorbeeld de geirrigeerde landbouw. De toepasbaarheid en nauwkeurigheid van deze op satelliet-gebaseerde schattingen moeten echter beter worden begrepen. Dit proefschrift heeft tot doel de geschiktheid van grote op teledetectie gebaseerde datasets voor de monitoring van

gewaswaterproductiviteit (CWP) beter te begrijpen. CWP is een indicator voor de efficiëntie van het gebruik van water in de landbouw.

Het proefschrift bestaat uit zeven hoofdstukken.

Hoofdstuk 1 is inleidend en beschrijft het belang van CWP in de landbouw, hoe CWP kan worden gemonitord en het groot belang om een beter inzicht te verkrijgen in de nauwkeurigheid van op teledetectie gebaseerde CWP schattingen.

Hoofdstuk 2 vergelijkt de nauwkeurigheid van op satelliet-gebaseerde CWP waarden met de nauwkeurigheid van de in-situ geschatte gewasopbrengsten en waterverbruik (CWP). De nauwkeurigheid van CWP op basis van in-situ methoden, waarvan meestal wordt aangenomen dat ze de "benchmark" zijn van de gebruiker voor CWP nauwkeurigheid, en de CWP afgeleid van methoden voor teledetectie moeten worden herzien. Het blijkt dat satelliet gebaseerde CWP waarden, zelfs verkregen uit niet-lokaal gecalibreerde studies de nauwkeurigheid bereiken binnen de foutgrenzen van in-situ methoden. Er wordt ook aangetoond dat onzekerheid aanzienlijk varieert afhankelijk van de toegepaste in-situ schattingsmethodes.

Hoofdstuk 3 evalueert de nauwkeurigheid van een continentale op teledetectie gebaseerde verdampingsgegevensset, de noemer in de CWP-indicator. De evaluatie maakt gebruik van verschillende methoden om de schaars beschikbare grondwaarnemingen in Afrika en Midden-Oosten te compenseren. De analyze leidde tot gemengde resultaten, al naar gelang men vergelijkt op punt locatie of op stroombekken niveau, en op verschillende tijdschalen i.e. dagelijks tot jaarlijks. De nauwkeurigheden van de CWP bepalingen zijn echter vergelijkbaar met die gerapporteerd door andere literatuur in hoofdstuk 2. Het onderzoek wees ook op de risico's van het gebruik van grote datasets in CWP monitoring, vanwege de grote interne variatie in nauwkeurigheid, meestal aanwezig in zeer grote continentale en globale datasets.

Hoofdstuk 4 kwantificeert de geschiktheid van verschillende ruimtelijke waarnemingsresoluties (op basis van teledetectie) voor toepassing in de landbouwproductiviteit, en met name in irrigatiegebieden. Drie prestatie-indicatoren, toereikendheid, billijkheid en productiviteit werden getest in vijf irrigatieschema's voor drie ruimtelijke resoluties, 250m, 100m en 30m. Toepassing van verschillende resoluties leidt vaak tot grote verschillen in de bovennoemde beoordelingscriteria van irrigatieprestaties e.g. bij het vergelijken tussen irrigatieplots. Dit wordt vooral opgemerkt in irrigatieschema's met de kleinste

veldgroottes. Dit benadrukt het grote belang van het selecteren van de gepaste ruimtelijke resolutie voor irrigatieschema's bij het uitvoeren van irrigatieprestatie analyses met behulp van remote sensing data.

Hoofdstuk 5 kwantificeert de steekproefgrootte van validatietesten van grote continentale CWP-datasets. Doel was de om de digitale verwerkingstijd en gemak van de statistische analyses te verbeteren. Een progressieve bemonsteringsbenadering, zoals meestal toegepast in "machine learning" om algoritmen te trainen, in combinatie met twee prestatiemetingen, werd toegepast om de vereiste steekproefgrootte te schatten. De voorgestelde aanpak kan de verwerkingstijd aanzienlijk verkorten en tegelijkertijd een statistisch geldige weergave van een grote remote sensing dataset bieden. Dit kan handig zijn omdat er vandaag steeds meer "big data" en remote sensing gegevens met hoge resolutie beschikbaar komen.

Hoofdstuk 6 gaat in op de implicaties van het CWP-concept en hoe dit concept soms zelf tot tegenspraak kan leiden met de wereldwijde doelstellingen op het gebied van voedselzekerheid en waterzekerheid. Het CWP-concept kan zeer nuttig zijn, maar het kan ook misleidend zijn. Op zowel lokale als regionale, continentale schaal moeten gebruikers nadenken over de doelen en indicatoren selecteren of prioriteren. Enkel verbetering van CWP nastreven is misschien niet altijd de meest nuttige indicator om tegerlijkertijd voedsel- en water zekerheid te bevorderen. We moeten ons richten op de duurzame verbetering van CWP in plaats van enkel de toename nastreven van CWP in de duurzame landbouw.

Toekomstige studies moeten ook nog beter aantonen hoe de fysische en biochemische interacties tussen de koolstof cyclus, nutrienten en gewasverdamping (i.e. de water cyclus) nog beter kunnen worden gebruikt als instrument om nog meer inzicht te krijgen in de kwaliteit van op satelliet-gebaseerde gegevenssets voor de monitoring van waterproductiviteit in de landbouw.

Author's biography and publications

Megan Leigh Blatchford was born in May 1989 in Nambour, Australia. She received her BEng degree in Environmental Engineering from the University of Queensland in Brisbane. After a few years in government and industry she went on to further pursue her education. She awarded her MSc degree in Water Science and Engineering – Hydraulic Engineering – Land and Water Development in 2016 from UNESCO-IHE in Delft. Based on her research in her MSc., she was awarded a four-year PhD candidacy at the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente. Her research interests include remote sensing of vegetation, biomass, coupled to evapotranspiration and water use of crops, aiming to improve understanding and management of agricultural and natural ecosystem systems.

Blatchford, M.L., Mannaerts, C.M., Njuki, S.M., Nouri, H. Zeng, Y., Pelgrum, H., Wonick, S., & Karimi, P. (2020) Influence of spatial resolution on remote sensing-based irrigation performance assessment, case study: WaPOR V2, Remote sensing (2020), 12, 2949: https://doi.org/10.3390/rs12182949

Blatchford, M.L., Mannaerts, C.M., Zeng, Y. (2020) Determining representative sample size for validation of continuous, endless continental remote sensing data, International Journal of Applied Earth Observation and Geoinformation. JAG (2020) 94, 102235: https://doi.org/10.1016/j.jag.2020.102235

Blatchford, M.L., Mannaerts, C.M., Njuki, S.M., Nouri, H. Zeng, Y., Pelgrum, H., Wonick, S., & Karimi, P. (2020) Evaluation of WaPOR V2 evapotranspiration products across Africa. Hydrological Processes (2020) 34:3200-3221 Special Issue. https://doi.org/10.1002/hyp.13791

Blatchford, M.L., Mannaerts, C.M., Zeng, Y., Nouri, H. & Karimi, P. (2019). *Status of accuracy in remotely sensed and in-situ agricultural water productivity estimates: A review*. Remote Sensing of Environment, 234 (1 Dec. 2019), 111413. https://doi.org/10.1016/j.rse.2019.111413

Blatchford, M.L, Karimi, P., Bastiaanssen, W. G. M., & Nouri, H. (2018). *From Global Goals to Local Gains—A Framework for Crop Water Productivity*. ISPRS International Journal of Geo-Information, 7(11), 414. https://doi.org/10.3390/ijgi7110414

Nouri, H., Stokvis, B., Borujeni, S.C., Galindo, A., Brugnach, M., **Blatchford, M.L.**, Alaghmand, S., Hoekstra, A.Y. (2020) Reduce blue water scarcity and increase nutritional and economic water productivity through changing the cropping pattern in a catchment, Journal of Hydrology. 588, 125086. https://doi.org/10.1016/j.jhydrol.2020.125086

Su, Z., Zeng, Y., Romano N., Manfreda, S., Francés, F., Dor, E.D., Szabó, B., Vico, G., Nasta, P., Zhuang, R., Francos, N., Mészáros, J., Dal Sasso, S.F., Bassiouni, M., Zhang, L., Rwasoka, D.T., Retsios, B., Yu, L., **Blatchford, M.L**., & Mannaerts, C.M. (2020) An Integrative Information Aqueduct to Close the Gaps between Satellite Observation of Water Cycle and Local Sustainable Management of Water Resources, Water 12 (5), 1495. https://doi.org/10.3390/w12051495

Karimi, P., Bongani, B., **Blatchford M.L**, & de Fraiture, C. (2019). *Global Satellite-Based ET Products for the Local Level Irrigation Management: An Application of Irrigation Performance Assessment in the Sugarbelt of Swaziland.* Remote Sensing 11 (6), 705. https://doi.org/10.3390/rs11060705

Nouri, H., Stokvis, B., Galindo, A., **Blatchford, M.,** & Hoekstra, A. Y. (2019). *Water scarcity alleviation through water footprint reduction in agriculture: The effect of soil mulching and drip irrigation*. Science of the Total Environment, 653, 241–252. https://doi.org/10.1016/j.scitotenv.2018.10.311

FAO**,** 2020. WaPOR V2 Quality Assessment - Technical Report on the Data Quality of the WaPOR database Version 2 (2020), FAO, Rome. http://www.fao.org/3/cb2208en/cb2208en.pdf (First authors C.M.Mannaerts & **M.L.Blatchford** contributed equally to this work).