



## Review

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

Megan L. Blatchford<sup>a,\*</sup>, Chris M. Mannaerts<sup>a</sup>, Yijian Zeng<sup>a</sup>, Hamideh Nouri<sup>b</sup>, Poolad Karimi<sup>c</sup>

<sup>a</sup> ITC-UTWENTE, Hengelostraat 99, 7514 AE Enschede, Netherlands

<sup>b</sup> University of Göttingen, Division of Agronomy, Göttingen 37075, Germany

<sup>c</sup> IHE Institute for Water Education, Westvest 7, 2611 AX Delft, Netherlands

## ARTICLE INFO

Edited by Marie Weiss

## Keywords:

Crop water productivity

Crop yield

Evapotranspiration

Remote sensing

In-situ

## 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 (ET<sub>a</sub>), 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, is reviewed. Then, the accuracy of current remote sensing products is 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 ET<sub>a</sub> 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 < 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 Organization (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 (ET<sub>a</sub>), above ground

\* Corresponding author.

E-mail addresses: [m.l.blatchford@utwente.nl](mailto:m.l.blatchford@utwente.nl) (M.L. Blatchford), [c.m.m.mannaerts@utwente.nl](mailto:c.m.m.mannaerts@utwente.nl) (C.M. Mannaerts), [y.zeng@utwente.nl](mailto:y.zeng@utwente.nl) (Y. Zeng), [hamideh.nouri@uni-goettingen.de](mailto:hamideh.nouri@uni-goettingen.de) (H. Nouri), [p.karimi@un-ihe.org](mailto:p.karimi@un-ihe.org) (P. Karimi).

<https://doi.org/10.1016/j.rse.2019.111413>

Received 20 September 2018; Received in revised form 3 September 2019; Accepted 4 September 2019

Available online 09 October 2019

0034-4257/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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<sup>-1</sup> year<sup>-1</sup>, 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, 1990; 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):

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

The crop yield is defined as the seasonal crop yield and the  $ET_a$  is taken as the accumulated crop  $ET_a$ , from start of season (SOS) to end of season (EOS). The conversion factor, 10<sup>-1</sup>, converts  $ET_a$  from mm to m<sup>3</sup> ha<sup>-1</sup>. By using  $ET_a$  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 (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):

$$GPP \text{ (gC m}^{-2} \text{ day}^{-1}\text{)} \\ = \epsilon \text{ LUE}_{\text{max}} \text{ (gC MJ}^{-1}\text{)} \sum \text{PAR (MJ m}^{-2} \text{ day}^{-1}\text{)} \times \text{fAPAR (-)} \quad (2)$$

where  $\epsilon$  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<sub>2</sub> that is fixed by the plant in photosynthesis. The maximum LUE ( $LUE_{\text{max}}$ ) 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) among 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:

$$NPP \text{ (gC m}^{-2} \text{ day}^{-1}\text{)} = GPP \text{ (gC m}^{-2} \text{ day}^{-1}\text{)} - AR \text{ (gC m}^{-2} \text{ day}^{-1}\text{)} \quad (3)$$

The distinction between the GPP, NPP, DMP, AGBP and yield are shown in Fig. 1. The  $LUE_{\text{max}}$  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  $LUE_{\text{max}}$  for different vegetation types, including cropland. In agricultural

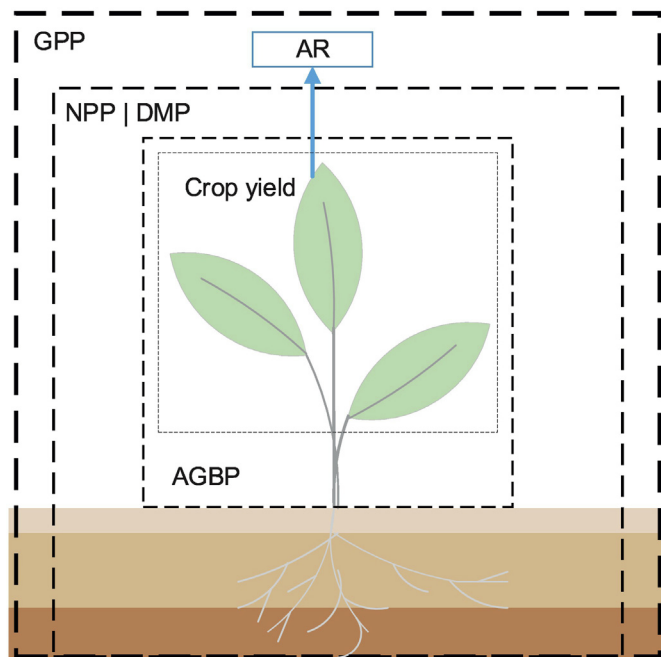


Fig. 1. Distinction between GPP, NPP, DMP, AGBP and crop yield products, where each box compares the plant parts associated with each product.

applications, the  $LUE_{max}$  is not solely defined for cropland, but for specific crop type.

In agricultural applications the NPP is then converted to DMP (3), typically through static conversion factors, before being converted to crop yield (4):

$$DMP (kg ha^{-1}) = \frac{1}{0.045} NPP (gC m^{-2}) \quad (4)$$

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 ( $\theta$ ) of the harvestable product (Prince et al., 2001):

$$Crop\ yield (kg m^{-2}) = f \cdot HI (-) \sum_{SOS}^{EOS} DMP (kg ha^{-1}) / (1 - \theta) \quad (5)$$

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  $\theta$  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  $LUE_{max}$ , HI and  $\theta$  (Zwart et al., 2010).

### 2.3. Evapotranspiration

$ET_a$  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  $ET_a$  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:

$$R_n (W m^{-2}) = LE + H + G \quad (6)$$

where,  $LE (W m^{-2})$  is the latent heat flux,  $R_n$  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  $ET_a$  by  $LE/\lambda$ , where  $\lambda$  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  $ET_a$  based on the ETLook model (Pelgrum et al., 2012) and is defined as:

$$LE = \frac{\Delta (R_n - G) + \rho_{air} \times C_p (e_{sat} - e_a) / r_a}{\Delta + \gamma \left(1 + \frac{r_s}{r_a}\right)} \quad (7)$$

where  $\Delta = d(e_{sat})/dT (kPa ^\circ C^{-1})$  is the slope of the curve relating saturated water vapour pressure to air temperature ( $T^\circ C$ ).  $\rho_{air} (kg m^{-3})$  is the density of air,  $C_p (MJ kg^{-1} ^\circ C^{-1})$  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  $ET_a$ , and  $\gamma (kPa ^\circ C^{-1})$  is the psychrometric constant. This approach further partitions  $ET_a$  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:

$$Relative\ Error\ (%) = \frac{|absolute\ error|}{accepted\ value} \times 100 \quad (8)$$

The absolute error is defined as:

$$absolute\ error = experimental\ value - accepted\ value \quad (9)$$

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 s.Eqs (8)-(9). 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:

$$\text{absolute error} = \Delta x \quad (10)$$

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

$$R = \frac{X}{Y} \quad (11)$$

$$\delta R \approx |R| \cdot \sqrt{\left(\frac{\delta X}{X}\right)^2 + \left(\frac{\delta Y}{Y}\right)^2} \quad (12)$$

where, in this case,  $R$  represents the CWP,  $\delta R$  represents the uncertainty of CWP,  $|R|$  represents the absolute value of the mean, and  $\delta R/|R|$  represents the relative uncertainty or percent error. Similarly,  $X$  in this case represents the crop yield and  $Y$  represents the  $ET_a$ .

When possible, the error associated with different methods to estimate yield,  $ET_a$ , 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 Eq. (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  $ET_a$ . These methods have historically been accepted as standards in estimating crop yield and  $ET_a$  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 < 0.5 ha, 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<sup>2</sup> and 15.7–23.4% for triangular plots of 2.7 m<sup>2</sup> when compared to a whole-plot harvest estimate on a 44 m<sup>2</sup> 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 > 40 m<sup>2</sup>, yet bias of 14% with 60 m<sup>2</sup> 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<sup>2</sup> for mixed cropping. Bias of 28% for sorghum and 17% for yam was found in plot sizes of 50 m<sup>2</sup> and 100 m<sup>2</sup>. The bias was not reduced until plot sizes increased to 200 m<sup>2</sup> (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), > 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<sup>2</sup>) 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 are first estimated and then can be converted to yield estimates through crop and location specific conversion factors, as per s.Eqs. (3)-(5). 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 et al., 2013a](#)). 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 measurements to the pattern of flux measurements by Bowen ratio energy balance (see [Section 3.2](#)) 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.

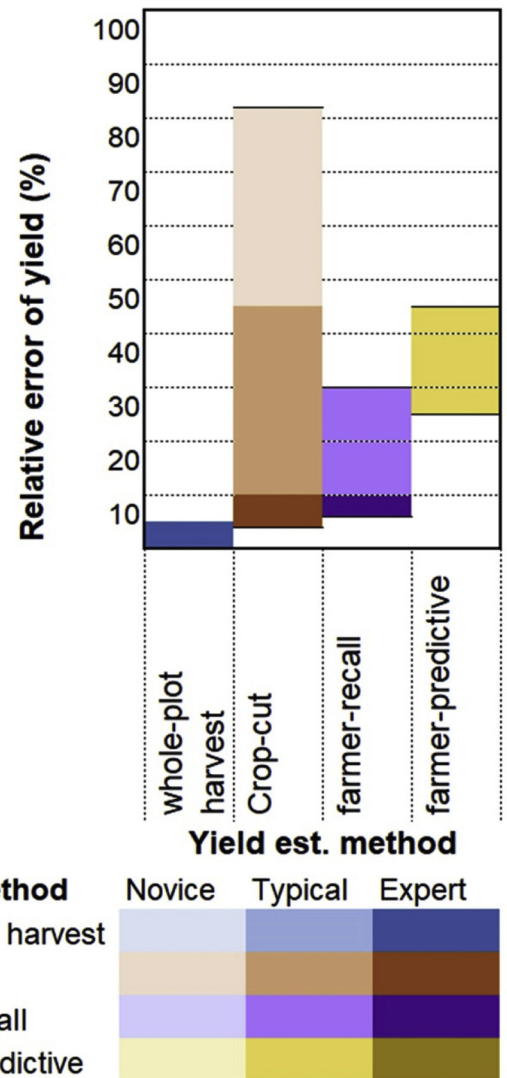


Fig. 2. Relative error associated with in-situ methods of crop yield estimation. All methods provide estimates for at field scale for cropping season.

The reported agreement in measurements between the two methods in non-agricultural lands varies significantly, from 8 to 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 [Fig. 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 Fig. 2, the y-axis is the suggested relative error range, as defined in Eq. (8) 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  $ET_a$ .

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 < 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 Fig. 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, is 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.

### 3.2. Evapotranspiration

Several in-situ measurement systems exist to determine  $ET_a$ . 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 Fig. 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  $ET_a$  error reported post-2011 is considered expert error, as the literature cited here was undertaken by scientists.

Lysimetry 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

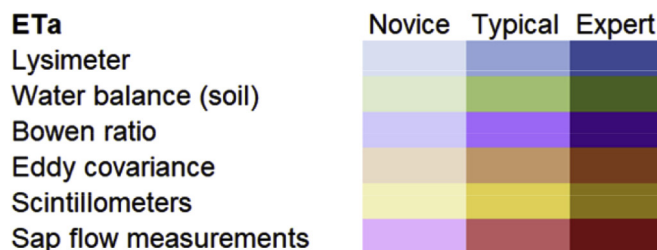
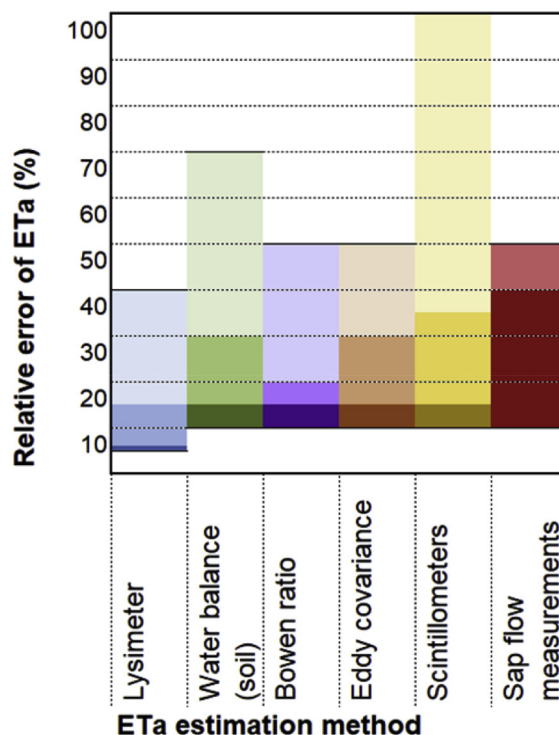


Fig. 3. Relative error associated with in-situ methods of  $ET_a$  estimation used for irrigation performance, adapted from Allen et al. (2011).

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. Lysimetry, 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. Lysimetry is often used for the validation of other in-situ  $ET_a$  methods as it is generally accepted to be the most accurate method to estimate  $ET_a$ .

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  $\pm 0.3$ – $0.5 \text{ mm day}^{-1}$  with resulting error ranging from 24 to 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  $ET_a$ . 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  $ET_a$  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  $ET_a$  measurement methods. Results varied considerably. On an annual scale in a lentil field, BREB overestimated  $ET_a$  by 10–43% as compared to lysimeter  $ET_a$  (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  $ET_a$  estimates. These discrepancies suggest that estimates of the scalar turbulent fluxes of H and LE are underestimated and/or that  $R_n$  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 and 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 and 88.8  $W m^{-2}$ , equivalent to an  $ET_a$  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  $ET_a$  were not reported. The footprint consisted of > 90% agricultural fields.

Sap flow  $ET_a$  measurement uncertainty in cotton was estimated to be 0.03–0.5  $mm h^{-1}$ , 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 et al., 2017b). The difficulty in using sap-flow measurements as a stand-alone method to estimate  $ET_a$  is that it actually measures transpiration, not  $ET_a$ . 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  $ET_a$  measurements.

It is also worth noting that the crop coefficient (Kc) is a widely accepted method to estimate  $ET_a$  from reference evapotranspiration ( $ET_o$ ) 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  $ET_o$  multiplied by a Kc. To obtain  $ET_a$  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  $ET_a$  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  $ET_a$  estimated by Kc varied considerably between years as compared to lysimeters. Liu and Luo (2010) found that the Kc approach showed reasonable seasonal  $ET_a$  with 10% relative error for winter wheat and summer maize as compared to lysimetry. However, peak  $ET_a$  was underestimated and the mean relative error of  $ET_a$  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  $ET_a$  of cherry trees, with relative error of < 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  $ET_a$  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 (lysimetry), plot scale (soil water balance and sap flow measurements), field scale (Bowen ratio and EC), and several hectares (scintillometers).

### 3.3. 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  $ET_a$  through simple error propagation, using Eqs. (11)–(12). 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  $ET_a$  in-situ measurement. These derived errors, however, do not take into account spatial scale differences between the crop yield and  $ET_a$  measurements. Fig. 4 shows the CWP relative error for each combination of the previously described crop yield and  $ET_a$  in-situ techniques. The relative error is plotted on the y-axis, the  $ET_a$  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 < 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 and 11% when combined with lysimeter, between 10 and 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 and 42% for the combination of lysimeter and farmer estimates (recall) to > 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%.

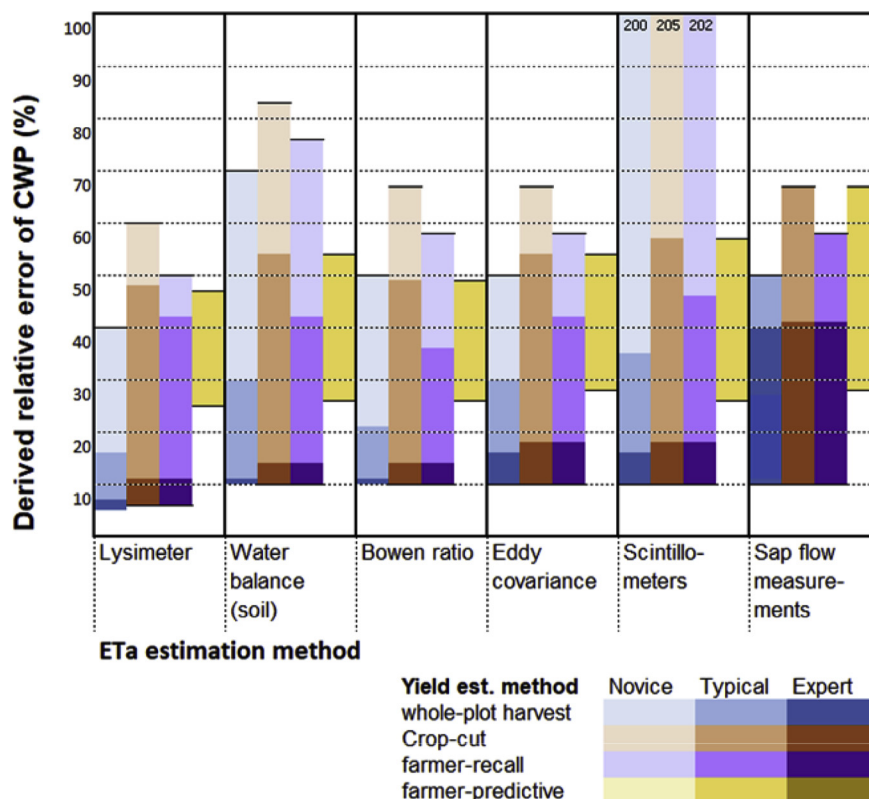


Fig. 4. Relative error associated with CWP derived from in-situ methods of estimating  $ET_a$  and crop yield. When numbers are located at the top of y-axis, they indicate value of relative error (when it goes) beyond 100%.

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  $ET_a$  method), the target relative error is therefore in the range of 2% (lysimetry 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.

#### 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  $ET_a$  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  $ET_a$  was used to assess the evaporative fraction (Bastiaanssen et al., 1998; Su, 2002), spatial distribution represented through the coefficient of variation (CV) of  $ET_a$  (Bastiaanssen et al., 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,  $ET_a$  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  $ET_a$  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 synthesised (Table 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 Eq. (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 (s.Eqs. (3)-(5)). Though these corrections are simple, they can impose significant errors. The implications of validating crop yield intermediates are discussed in Section 4.1.

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-



**Table 1**  
Stated accuracy of remote sensing derived crop yield, ordered by publication year.

Author	Location	Study size	Year/s	Main crops	Variable	Sensor (Spatial/temporal resolution)	Accuracy of study <sup>a,b</sup>	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 <sup>-1</sup> RMSE (maize yield) = 0.81; 1.06 ton ha <sup>-1</sup> RE = 10% R <sup>2</sup> = 0.709 R <sup>2</sup> = 0.34 RMSE = 94% RE = 33%	Crop-cut Farmer reported yields - recall EC
Löw et al. (2017)	Central Asia - Fergana Valley Global model, tested globally	363,000 ha	2010–2014	Mixed	Crop yield	Landsat-5 TM (30 m 16-day)	R <sup>2</sup> = 0.709 R <sup>2</sup> = 0.34 RMSE = 94% RE = 33%	Farmer reported yields - recall EC
Wang et al. (2017a)	Global	Global	2004–2014	not specified	GPP	MODIS (500 m 8-day)	Bias = -0.22 ton ha <sup>-1</sup> day <sup>-1</sup> RE = 5.8–6.2% RE = 8.7–14.7% r = 0.75 R <sup>2</sup> = 0.37–0.57	EC Crop-cut and Farmer reported yield - recall
Madugundu et al. (2017) Yilma (2017)	Saudi Arabia Awash, Ethiopia	50 ha 14,000 ha	2015 2014–2016	Maize Sugarcane	GPP ABGP	Landsat 8 (30 m 16-days) Landsat 8 (30 m 16-days)	Yield: R <sup>2</sup> = 0.61; RE = 30–61% GPP: R <sup>2</sup> = 0.90; RMSE = 0.02–0.11 ton ha <sup>-1</sup> day <sup>-1</sup> ; bias = 0–0.07 ton ha <sup>-1</sup> day <sup>-1</sup> ; RE = 0.5–88% (median = 11.9%) bias < 49.8%	EC Crop-cut and Farmer reported yield - recall
Yuan et al. (2016)	Global model, tested globally		2001–2011	GPP: 36 cropped sites Yield: 12 cropped sites	GPP, Yield	MODIS (1 km 8-day)	Yield: R <sup>2</sup> = 0.61; RE = 30–61% GPP: R <sup>2</sup> = 0.90; RMSE = 0.02–0.11 ton ha <sup>-1</sup> day <sup>-1</sup> ; bias = 0–0.07 ton ha <sup>-1</sup> day <sup>-1</sup> ; RE = 0.5–88% (median = 11.9%) bias < 49.8%	EC
Tang et al. (2015)	Global model, tested in USA	Global	2004–2005	Maize/soybean	GPP	MODIS (1 km 8-day)	MODIS GPP: r = 0.86; RMSE = 0.06 ton ha <sup>-1</sup> day <sup>-1</sup> ; bias = -0.00 ton ha <sup>-1</sup> day <sup>-1</sup>	EC
Yan et al. (2015)	Global model, tested globally	Global	2000–2010	Mixed	GPP	MODIS (1 km 8-day)	TEC GPP model: r = 0.77; RMSE = 0.08 ton ha <sup>-1</sup> day <sup>-1</sup> ; bias = -0.02 ton ha <sup>-1</sup> day <sup>-1</sup>	EC
Yuan et al. (2015)	Global model, tested in USA	Global	2001–2005	Soybean	GPP	MODIS (1 km 16-day & 1 km 8-day)	MODIS GPP: R <sup>2</sup> = 0.64–0.67; RMSE = 0.09–0.10 ton ha <sup>-1</sup> day <sup>-1</sup> ; bias = negligible VPM: R <sup>2</sup> = 0.5–0.79; RMSE = 0.08–0.10 ton ha <sup>-1</sup> day <sup>-1</sup> ; bias = 0.02–0.04	EC
Sibley et al. (2013)	USA - Nebraska		2007–2010	Maize	Crop yield	MODIS (250 m daily)	R <sup>2</sup> irrigated = 0.22 R <sup>2</sup> rainfed = 0.09 R <sup>2</sup> all = 0.52 r = 0.71 and 0.8	Farmer reported yields - recall EC
Sjöström et al. (2013)	Global model, tested in Africa	Global	2005–2006	Millet/grassland	GPP	MODIS (1 km 8-day)	RE (maize) = -69.2% to -78.4% RE (orchard) = -74.1%	EC
Wang et al. (2013b)	Global model, tested globally	Global	2008–2009	Maize, orchard	GPP	MODIS (1 km 8-day)	Bias (yield) = +0.5 t ha <sup>-1</sup> RE = 9.1%	Farmer reported yields - recall; crop cuts; and ministry statistics EC
Zwart and Bastiaanssen (2007)	Mexico - Yaqui River coastal plain	225,000 ha	1999–2000	Wheat	AGBP, crop yield	NOAA-AVHRR (1 km Monthly) Landsat TM	RMSE = 2 ton ha <sup>-1</sup> year <sup>-1</sup> RE = 20%	Farmer reported yields - recall; crop cuts; and ministry statistics EC
Turner et al. (2005)	Global model, tested in USA	Global	2000	Corn/soy	NPP	MODIS (1 km 8-day)	RE (state) = 5% R <sup>2</sup> (county) = 0.01–0.46	Ministry statistics based on farmer reported yields - recall
Reeves et al. (2005)	Global model, tested in USA	> 12,000 ha/county	2001–2002	Wheat	Crop yield	MODIS (1 km 8-day)	R <sup>2</sup> (climate) = 0.33–0.67 RE (regional) = 6% RE (field) = not reported	Regional scale: Regional statistics Field scale: crop-cut Regional statistics
Bastiaanssen et al. (2003)	India - Siroa			Wheat	Crop yield	Landsat ETM (30 m 16-day)   NOAA (1.1 km daily)	RE (field) = not reported	Regional scale: Regional statistics Field scale: crop-cut Regional statistics
Bastiaanssen and Ali (2003)	Pakistan – Indus Basin		1993–1994	Wheat, rice, cotton, sugar-cane	AGBP/Crop yield	AVHRR (1.1 km monthly)	RE = 22%–42%	Regional statistics

(continued on next page)

Table 1 (continued)

Author	Location	Study size	Year/s	Main crops	Variable	Sensor (Spatial temporal resolution)	Accuracy of study <sup>a,b</sup>	Method of validation
Lobell et al. (2003)	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
Samarasinghe (2003)	Sri-Lanka	1,752,100 ha	1999–2000	Tea, coconut, rice, rubber	AGBP/Crop yield	NOAA-AVHRR images (1.1 km 10-day)	R <sup>2</sup> = 0.47	Regional statistics
Lobell et al. (2002)	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 <sup>2</sup> = 0.78 RMSE = 0.37 ton ha <sup>-1</sup> RE = 5.9%	Farmer reported yield

<sup>a</sup> Abbreviations accuracy metrics used in this table: R<sup>2</sup> – coefficient of determination; r - correlation coefficient; RE – Relative Error (or percentage error); RMSE – root mean square error.  
<sup>b</sup> GPP (gC m<sup>-1</sup>) units are converted to ton ha<sup>-1</sup> using Eq. (2) to ease comparison between GPP and AGBP and yield errors.

GPP typically underestimated GPP by  $-0.06$  to  $-0.41$  gC m<sup>-2</sup> day<sup>-1</sup>, the EC-LUE had a positive bias of  $0.16$ – $0.37$  gC m<sup>-2</sup> day<sup>-1</sup>, and the VPM had a positive bias of  $1.02$ – $1.70$  gC m<sup>-2</sup> day<sup>-1</sup>. 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<sup>2</sup> of 0.9 and a RMSE and bias ranging between 1.75 and 5 gC m<sup>-2</sup> day<sup>-1</sup> at EC sites and 0.03–3.34 gC m<sup>-2</sup> day<sup>-1</sup> 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<sup>-2</sup> day<sup>-1</sup> and 1.18 gC m<sup>-2</sup> day<sup>-1</sup>, respectively. The r, RMSE and bias at the cropped site for 2005 and 2006 was 0.71 and 0.8, 0.97 gC m<sup>-2</sup> day<sup>-1</sup> and 0.73 gC m<sup>-2</sup> day<sup>-1</sup>, and  $-0.59$  gC m<sup>-2</sup> day<sup>-1</sup> and  $-0.32$  gC m<sup>-2</sup> day<sup>-1</sup>, 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<sup>-2</sup> day<sup>-1</sup>, and  $-0.16$  gC m<sup>-2</sup> day<sup>-1</sup>, respectively, across cropped sites. The MODIS products had an r, RMSE and bias of 0.77, 3.38 gC m<sup>-2</sup> day<sup>-1</sup>, and  $-0.76$  gC m<sup>-2</sup> day<sup>-1</sup>, 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<sup>-2</sup> year<sup>-1</sup> and 105 gC m<sup>-2</sup> year<sup>-1</sup> for soybean and corn respectively, corresponding to over 2 ton ha<sup>-1</sup> year<sup>-1</sup> 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<sup>-2</sup> year<sup>-1</sup> and 34 gC m<sup>-2</sup> year<sup>-1</sup> for the cropped sites and grassland site respectively. The EC GPP and NPP were scaled to 5 km × 5 km 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 et al., 2017a). The R<sup>2</sup>, RMSE and bias at the cropped sites was 0.34, 94%, and  $-10$  gC m<sup>-2</sup> day<sup>-1</sup>, 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<sub>max</sub>. 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<sup>2</sup> for all sites increased to 0.79).

Similarly, a study in China found that without calibration of LUE<sub>max</sub> the performance of MODIS GPP performed much worse in croplands compared to other vegetation (Wang et al., 2013b). MOD17 was compared to 10 EC sites, including four maize sites and an orchard. The RMSE over the maize sites ranged between 59.7 and 89.4 gC m<sup>-2</sup> 8-day<sup>-1</sup>. The relative errors ranged between -69.2% to -78.4%. The RMSE at the orchard site was 51.2 gC m<sup>-2</sup> 8-day<sup>-1</sup> 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<sub>max</sub> was adjusted for, the results improved considerably for all sites. The maize sites RMSE reduced to 14.6–17.8 gC m<sup>-2</sup> 8-day<sup>-1</sup> and the relative error reduced to 3.1–11.5% (Wang et al., 2013b).

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<sup>2</sup> 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<sup>2</sup> 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 normalized 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<sup>2</sup> of 0.98. The RMSE values for different fields ranged between 1.39 and 2.18 ton ha<sup>-1</sup> for each field over the growing season. WUE and CWP based approaches showed similar results for R<sup>2</sup>. The CWP model had the lowest RMSE values (1.07–1.58 ton ha<sup>-1</sup>). 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<sup>2</sup> 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<sup>-1</sup>, while the Landsat crop-model based regression had RMSE values of just over 2 ton ha<sup>-1</sup>.

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 < 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<sup>2</sup> of 0.78 and a RMSE of 0.37 ton ha<sup>-1</sup>.

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<sup>-1</sup> 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<sup>-1</sup>, 0.62 ton ha<sup>-1</sup>, 0.55 ton ha<sup>-1</sup>, and 13.5 ton ha<sup>-1</sup>, 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<sup>2</sup> 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<sup>2</sup> was 0.47 and the RMSE was 0.43 ton ha<sup>-1</sup>. 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<sup>2</sup> 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<sub>max</sub> adjustment (Wang et al., 2013b) and up 70% and even 90% (Wang et al., 2017a). 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 LUE<sub>max</sub> 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.

Fig. 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 1. The in-situ values are taken from Fig. 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 < 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  $\theta$ . 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.

Integration of remote sensing into crop models through data assimilation methods is becoming more prevalent, including models such

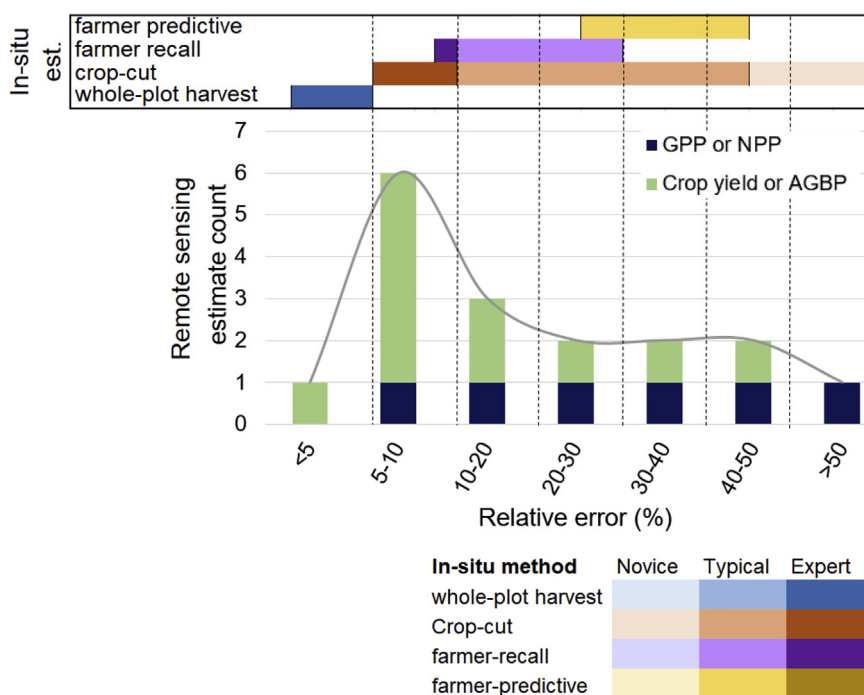


Fig. 5. Count of relative error ranges of remote sensing-based GPP, NPP, AGBP and yield reported in, or derived from, literature compared to in-situ relative errors.

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 ( $\theta$ ) 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 and 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 and 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 (Feres 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  $\theta$  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  $\theta$  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  $LUE_{max}$  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  $LUE_{max}$  values from literature, which varied significantly between crops, particularly between C3 and C4 crops. The importance of distinguishing  $LUE_{max}$  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  $LUE_{max}$ , 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,  $f$ ,  $\theta$ , and  $LUE_{max}$ , accurate land classification is important to ensure that the appropriate crop specific conversion factors or look-up tables for the AGBP fraction, HI and  $LUE_{max}$  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  $ET_a$  estimation on an annual or seasonal scale. They demonstrated that the absolute error of  $ET_a$  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  $ET_a$  estimates over cropped areas, with 60% of the studies achieving an error of  $\leq 5\%$ . 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 difference between SEBAL estimated  $ET_a$  and lysimeter  $ET_a$  of 9.3 and 15.4% for onion and potato fields respectively. The range of errors was 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  $ET_a$  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., 2007, 2011) and EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA-SAF MSG ET) (Trigo et al., 2011). MOD16 (1 km, daily resolution) is based on the PM model and LSA-SAF MET (3 km at the nominal position at 0° longitude, 30-min) 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<sup>-1</sup> 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 to 0.48 mm day<sup>-1</sup> or 9–30% with a mode error of 0.3 mm day<sup>-1</sup> or 20%. The MOD16 errors in croplands ranged from 36 to 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<sup>-1</sup> for both sites, and bias of -0.10 and 0.27 mm day<sup>-1</sup>. MOD16 had  $R^2$  of 0.90 and 0.91, RMSE of 0.72 and 0.47 mm day<sup>-1</sup>, and bias of -0.39 and 0.26 mm day<sup>-1</sup>, 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<sup>-1</sup> and < 0.1 mm day<sup>-1</sup> when  $ET_a$  is < 0.4 mm day<sup>-1</sup>. 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 errors 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<sup>-1</sup>) and SSEBop had the highest RMSE (1.67 mm day<sup>-1</sup>). 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 and 12.98 mm 8-day<sup>-1</sup> 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  $ET_a$  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  $ET_a$  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 > 30 years (Bausch and Neale, 1987; Neale et al., 1989). The Kcb remote sensing approach estimates  $ET_a$  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., 2010, 2013; 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 Guadalquivir 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  $0.75 \text{ mm day}^{-1}$ . 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  $ET_a$  estimates show a large range of reported errors. Locally parameterised and calibrated  $ET_a$  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  $ET_a$  varies widely, between location and models. Karimi and Bastiaanssen (2015) suggest remote sensing-based  $ET_a$  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 vary considerably. Some models have reported errors of < 15% (Bhattarai et al., 2016; Singh and Senay, 2016) while other models have reported errors of > 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 allow 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, 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  $ET_a$  and crop yield. The only cited literature using this method for  $ET_a$  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 1 km. 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 1 km 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 et al. 2013b). 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 (Fig. 4). However, the typical errors have large ranges from 7 to 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  $ET_a$  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  $ET_a$  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  $ET_a$ , are. The accuracy of CWP was therefore derived by propagating the reported accuracy of both remote sensing and in-situ  $ET_a$  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  $ET_a$  it ranges from 5 to 15% (lysimeter) to 50% (sap flow measurements). This propagates to CWP errors from field measurements that range from 7 to 67%. Based on remote sensing reported accuracy of  $ET_a$  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  $ET_a$ , 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.

## Acknowledgements

This work was partly funded by the United Nations Food and Agricultural Organization (FAO) as part of the Water Productivity Open Access Portal (WaPOR) project, in which we undertook the validation efforts. The WaPOR validation efforts were undertaken under contract number 2016/HQR/NRL-CPA320260 (FRAME project).

## Declaration of competing interest

None.

## References

- 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](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











- ATBD\_FAPAR1km-V2.11.41.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>.
- 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, K., Lui, C., Zheng, X., Pihlatie, M., Li, B., Haapanala, S., Vesala, T., Lui, H., Wang, Y., Liu, G., Hu, F., 2013a. 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., 2013b. 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., 2017a. 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., 2017b. 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>.
- 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>.
- 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](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](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. Model.* 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.
- 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., Ganelle, 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>.
- 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. Model.* 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>.
- 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>.