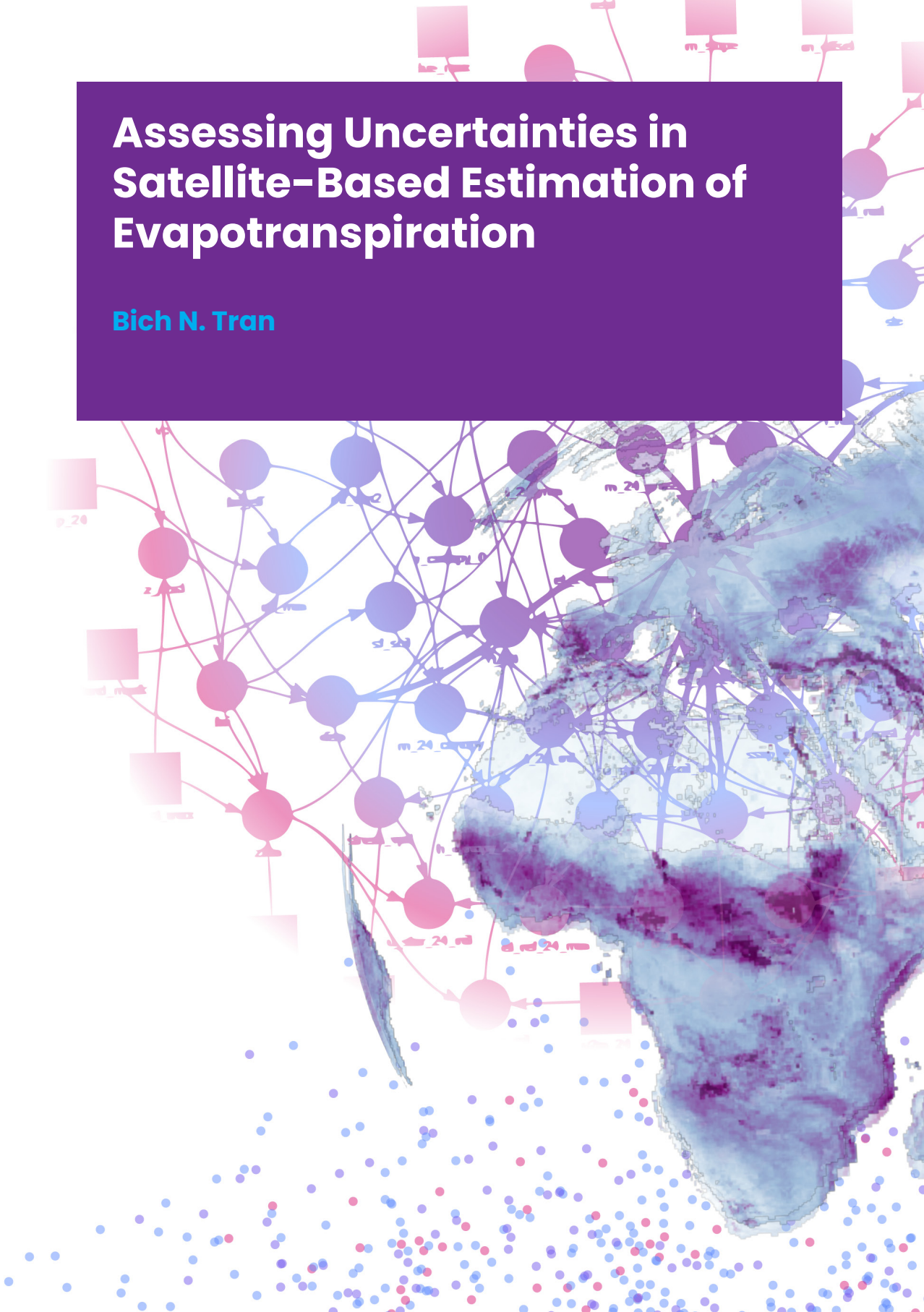
Assessing Uncertainties in Satellite-Based Estimation of Evapotranspiration

Bich N. Tran



ASSESSING UNCERTAINTIES IN SATELLITE-BASED ESTIMATION OF EVAPOTRANSPIRATION

Bich N. Tran

ASSESSING UNCERTAINTIES IN SATELLITE-BASED ESTIMATION OF EVAPOTRANSPIRATION

DISSERTATION

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. H. Bijl,
chair of the Board for Doctorates
and
in fulfilment of the requirement of the Vice Rector of IHE Delft
Institute for Water Education, Prof.dr. G.P.W. Jewitt,
to be defended in public on
Thursday, 16 April 2026 at 17:30 hours

by

Ngoc Bich TRAN

This dissertation has been approved by the (co)promotors.

Composition of the doctoral committee:

Rector Magnificus TU Delft	chairperson
Vice Rector IHE Delft	vice-chairperson
Prof.dr. G.P.W. Jewitt	TU Delft / IHE Delft, promotor
Prof.dr.ir. R. Uijlenhoet	TU Delft, promotor
Dr.ir. M.L. Mul	IHE Delft, copromotor

Independent members:

Prof.dr.ir. M. Hrachowitz	TU Delft
Dr.ir. A.M.J. Coenders	TU Delft
Dr. L.A. Melsen	Wageningen University & Research
Prof.dr. A. Ruhoff	UFRGS, Brazil
Prof.dr.ir. P. van der Zaag	TU Delft / IHE Delft, reserve member

Dr. S.D. Seyoum and Dr. J. van der Kwast (IHE Delft) and L. Peiser (FAO) have contributed greatly to the supervision of this dissertation.

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

Published by IHE Delft Institute for Water Education
www.un-ihe.org
ISBN 978-90-73445-80-2

In loving memory of my late grandmother, who taught me to cherish the ordinary.

ACKNOWLEDGEMENTS

If my PhD journey is the process of evaporation, then, Marloes, you are the incoming solar radiation, providing the daily energy that lifts my fluid, ever-evolving work into meaningful outputs. Thank you for your daily support and pragmatic advice, which has been essential to shaping and lifting this research toward completion. To Graham and Remko, thank you for being the vapour pressure deficit to my evaporation. Your guidance, both demanding and deeply encouraging, has carried me through some of my most resistant moments. Thank you, Solomon and Hans, for being the wind that propels my research at stagnant times. I am deeply grateful to have you all in my supervisory team and for our monthly meetings, which help keep me on track with my PhD trajectory. With every question and suggestion, you stir new ideas and turn my understanding into concrete steps forward.

I feel sincerely grateful for the generosity of many collaborators, researchers, and reviewers. I asked and received help from a lot of them, who are specifically acknowledged in each of my articles. My special thanks go to my co-authors, Suzan and Kawa, whose dedication and collaboration were invaluable. Sincere thanks to Livia and Bert for always supporting my research and offering practical perspectives. I am also thankful to Annemarie, Henk, Pasquale, and Rado for their openness in sharing their insights. My gratitude extends to all teachers and trainers in my doctoral education for organizing such excellent learning experiences, with particular appreciation for Prof. Gerard Heuvelink, Prof. Sytze de Bruin, and Prof. Hans-Peter Schmid.

Even a well-rested field can exhibit greater evaporation through advection from nearby fields. Likewise, my intellectual growth comes not only from my field of research but also from the unexpected exchange of ideas and knowledge across boundaries. Thank you, my learning partners, Roos, Jonatan, and Rossella, for thinking and writing with me on modelling and transition. I would like to thank Margreet, Pieter, Tobias, Jean-Phillipe, Bruno, Francois, and Marcel for supporting our CAT group and always welcoming my questions with engaging and constructive discussions. I am thankful for the knowledge gained from the Summer School of Situated Modelling, the Decolonizing Science course, and the IAHS working group on Co-Creating Water Knowledge. Special thanks to Amit, Sören, and Edwin for their writing (and thinking) advice, which has helped me overcome writer's block.

There can be no evaporation without moisture reserves on the surface, just as there can be no fruitful work without rest, replenishment, and, at times, a full reset. It is in those pauses that my mind forms new connections and finds my way through setbacks. I am, therefore, sincerely grateful to my dear friends and colleagues at IHE Delft and TU Delft for helping me hold onto perspective, replenish joy, and keep my spirit from running dry. Thank you, Adele, Milk, Omar, Yared, Claire, Abebe, and Seleshi, for your advice on

navigating the PhD journey. Learning from your experiences has helped me a lot in finding my way. Thank you, Adele, Binny, Celia, Ain, Alyssa, Enya, Gaby, Eliana, and Sumi for chill-working with me at TU library or café, and on PhD Writing Days. Thank you, Binny, Adam, and Adele, for also being my constant helplines and making Mina feel more like home. I would like to thank my colleagues, Amani, Claire, Marloes, Elga, Amit, Solomon, Hans, Ahmed, Vineet, Sajid, and Neha for their support and for sharing experiences during lunch talks, coffee/snack breaks, after-lunch walks, and after-work drinks.

I also wish to thank the IHE Delft community and friends I met in the Netherlands for making my PhD journey feel less solitary: chị Hà, chị Thảo Nguyên, Coti, Khin, Ingrid, Bota, Henry, Nafn, Tumaini, Saidee, Haris, Santiago, Mercy, Cristiane, Deborah, Vanida, Yiman, Jiaqi, Gofran, Mutaz, Aftab, Sebrian, Ifan, Eliza, Selam, Lina, Janvière, Tanya, Thaine, Asia, Roya, Leila, Zoubida, Quico, Enrico, Pau, Annida, Rachel, Noor, Krishna, Mila, Mariam, Saber, Alejandro, Marie, Heidi, Carmen, Johnathan, Meijun, Saïd, Sara, Beatriz, Giovanni, Lorenzo, Belen, Emanuelle, Desa, Italo, Maia, Rodrigo, Alvaro,...

To my family and friends beyond the Netherlands, you are the rain that arrives from across the sea. Thank you for cheering me on from afar, for being both my winter haven and the roots that ground my being wherever I am, and for repeatedly asking, “When will you finish your PhD?” almost every time we spoke over the last years. Last but not least, Manu, thank you for irrigating my life with steady streams of positivity and possibility. È nel tuo essere che ho trovato il mio divenire. Grazie di cuore per tutti.

SUMMARY

Satellites are increasingly used to acquire data and derive information in water sciences and management. This research focuses on satellite-based estimation of evapotranspiration (ET), a prominent water flux in terrestrial water balance, and also the key link between water and surface energy balances. Quantification of ET is required for many applications in water management, but it is challenging to gauge in situ. Many models have been developed to estimate actual ET (ET_a) using data from satellite remote sensing. Some of these satellite-based ET_a models are applied to generate data products covering a global extent, which are increasingly being used in water management at various scales, from agricultural fields to river basins. While previous studies have evaluated satellite-based ET_a models, it remains uncertain whether they can provide reliable estimates in all conditions.

This research aims to enhance the understanding of uncertainties in satellite-based estimation of ET_a and further explore the implications of these uncertainties for applications in water management. Uncertainties in satellite-based ET_a estimates arise from both methodological and technical factors. Technical uncertainties arise from input data sources, model structure, parameters, and data processing, which can be assessed through methods from probability and statistics. Methodological uncertainties arise from modelling approaches and subjective modelling choices, which are often not quantifiable or even recognised. Several methods to assess technical uncertainties in satellite-based ET_a estimates have been developed, which could be applied before (ex-ante) and after (ex-post) the generation of data products using satellite-based ET_a models. In this study, uncertainties in satellite-based estimation of ET_a were assessed from two perspectives: one informed by academic literature (Part I) and the other by direct engagement with the quality assessment and application of a satellite-based ET_a product (Part II).

Part I aims to identify how uncertainty has been evaluated in academic literature on satellite-based ET_a models and data products, and the status of uncertainty reported. A systematic quantitative literature review was conducted in Chapter 2 to identify and appraise the advances and caveats of uncertainty assessments for satellite-based ET_a estimates. This review shows the common uncertainty assessment approaches, but also their diverse approaches, and constraints due to the availability and quality of the reference. In addition, a meta-analysis of studies that employed in-situ validation against Eddy Covariance (EC) measurements was conducted in Chapter 3, which presents the status of uncertainty in terms of two commonly reported performance metrics: Root Mean Square Error (RMSE) and percentage bias (PBIAS). This analysis also shows that PBIAS in satellite-based ET_a estimates varies significantly across studies utilizing different model types, but without a significant improvement over the decade.

Part II focuses on the satellite-based ET_a model and data that were developed and applied in the WaPOR project, which exemplifies the challenges in uncertainty assessment of satellite-based ET_a products at the continental and global scale. The objectives of Part II are: (1) to evaluate the technical uncertainties of a satellite-based ET_a data through both ex-ante and ex-post uncertainty assessments and (2) to explore the implications of uncertainties for the applications in water management.

Meteorological forcings are required in ET_a modelling and often sourced from weather stations or climate reanalysis. In Chapter 4, the ex-ante uncertainty assessment of meteorological forcings from reanalysis data (GEOS5, ERA5, AgERA5) was investigated in Africa and Southwest Asia by direct comparison with in-situ weather stations and spatio-temporal intercomparison. The differences between these datasets are non-stationary and most pronounced in Central and Southern Africa and Southwest Asia. Uncertainty propagations were applied to analyse how uncertainties in ERA5 impact the estimation of ET for a reference crop. The results using two error propagation methods (Monte Carlo and Taylor expansion) were compared, illustrating the trade-off between accuracy and computational cost in estimating uncertainty.

In Chapter 5, the compound and relative uncertainty were evaluated for the WaPOR version 3 Level 1 (WaPORv3L1) ET_a data product, which was derived using the ETLook model. In-situ validation against EC measurements indicates comparable compound uncertainty to that of other satellite-based ET_a products. However, in-situ validation still faces challenges in dry and tropical climates, where the performance metrics were poor. Triple collocation analysis (TCA) was employed to estimate relative uncertainty in terms of RMSE and correlation coefficient. Although TCA is valuable in revealing the spatial variation of uncertainty, the results highly depend on the choices of the datasets, and underestimating uncertainties compared to in-situ validation.

To explore the implications for applications of satellite-based ET_a estimation, Chapter 6 examined the WaPOR project in terms of problem framing, model choices, and configurations, and the consequences of uncertainties. The context of the WaPOR project and its emphasis on agricultural water use and water productivity led to a certain choice of model and configuration. While uncertainties are mainly perceived and managed as predominantly technical problems, reflecting on methodological uncertainties encourages exploring alternative data and models and involving participation from data users.

From the experience and knowledge gained from this research, additional data quality information is recommended for satellite-based data products, including outputs from ex-ante and ex-post uncertainty assessments. Uncertainty assessments should be seen as a modelling practice contingent upon the specific methods, references, and methodological choices. Therefore, future work should incorporate additional information about known issues and users' feedback, while improving both ex-ante and ex-post assessment methods.

SAMENVATTING

Waterwetenschappen en -beheer gebruiken steeds vaker satellieten om gegevens te verzamelen en informatie af te leiden. Dit onderzoek richt zich op satellietgebaseerde schattingen van evapotranspiratie (ET), een belangrijke waterstroom in de terrestrische waterbalans en tevens de belangrijkste schakel tussen de water- en oppervlakte-energiebalansen. Veel waterbeheertoepassingen vereisen het bepalen van de hoeveelheid van ET, maar dit is lastig ter plaatse te meten. Er zijn veel modellen ontwikkeld om de werkelijke ET (ET_a) te schatten met behulp van satellietobservatiegegevens. Sommige van deze op satellieten gebaseerde ET_a-modellen worden toegepast om producten te genereren die een wereldwijde dekking hebben. Deze producten worden steeds vaker gebruikt in waterbeheer op verschillende schaalniveaus, van landbouwvelden tot rivierbekkens. Hoewel eerdere studies op satellieten gebaseerde ET_a-modellen hebben geëvalueerd, blijft het onzeker of ze onder alle omstandigheden betrouwbare schattingen kunnen leveren.

Dit onderzoek heeft tot doel het inzicht in de onzekerheden bij satellietgebaseerde schattingen van ET_a te vergroten en de implicaties van deze onzekerheden voor waterbeheertoepassingen verder te onderzoeken. Onzekerheden in satellietgebaseerde ET_a-schattingen vloeien voort uit zowel methodologische als technische factoren. Technische onzekerheden vloeien voort uit onzekerheden in de invoergegevens, modelstructuur, parameters en gegevensverwerking. Deze onzekerheden kunnen worden bepaald met behulp van methoden uit de kansrekening en statistiek. Methodologische onzekerheden vloeien voort uit modelleringsbenaderingen en subjectieve model keuzes, die vaak niet kwantificeerbaar zijn of zelfs niet worden onderkend. Er zijn verschillende methoden ontwikkeld om technische onzekerheden in satellietgebaseerde ET_a-schattingen vast te stellen. Deze kunnen worden toegepast vóór (ex ante) en na (ex post) het genereren van de dataproducten met behulp van satellietgebaseerde ET_a-modellen. In deze studie werden onzekerheden in satellietgebaseerde schattingen van ET_a beoordeeld vanuit twee perspectieven: het ene gebaseerd op academische literatuur (deel I) en het andere op directe betrokkenheid bij de kwaliteitsbeoordeling en toepassing van een satellietgebaseerd ET_a-product (deel II).

Deel I heeft tot doel vast te stellen hoe onzekerheden van satellieten gebaseerde ET_a-modellen en gegevensproducten worden geëvalueerd in de academische literatuur, en wat de status is van de gerapporteerde onzekerheid. In hoofdstuk 2 is een systematisch kwantitatief literatuuronderzoek uitgevoerd om de vorderingen en kanttekeningen bij onzekerheidsbeoordelingen voor op satellieten gebaseerde ET_a-schattingen in kaart te brengen en te beoordelen. Dit onderzoek toont de gangbare benaderingen voor onzekerheidsbeoordeling, maar ook de uiteenlopende benaderingen en beperkingen als gevolg van de beschikbaarheid en kwaliteit van de referentiedata. Daarnaast is in

hoofdstuk 3 een meta-analyse uitgevoerd van studies waarin gebruik is gemaakt van in-situ validatie ten opzichte van Eddy Covariance (EC)-metingen. Deze analyse geeft de status van onzekerheid weer in termen van twee veelgebruikte prestatiemaatstaven: Root Mean Square Error (RMSE) en percentage bias (PBIAS). Deze analyse laat ook zien dat PBIAS in satellietgebaseerde ET_a-schattingen aanzienlijk varieert tussen studies die gebruikmaken van verschillende modeltypes, maar dat er in het afgelopen decennium geen significante verbetering is opgetreden.

Deel II richt zich op het satellietgebaseerde ET_a-model en de gegevens die zijn ontwikkeld en toegepast in het WaPOR-project, dat een voorbeeld is van de uitdagingen bij de onzekerheidsbeoordeling van satellietgebaseerde ET_a-producten op continentale en mondiale schaal. De doelstellingen van deel II zijn: (1) het evalueren van de technische onzekerheden van satellietgebaseerde ET_a-gegevens door middel van zowel ex-ante als ex-post onzekerheidsbeoordelingen en (2) het onderzoeken van de implicaties van onzekerheden voor waterbeheertoepassingen.

ET_a-modellen maken gebruik van meteorologische forceringen die vaak verkregen worden van weerstations of klimaatheranalyse. In hoofdstuk 4 werd de ex-ante onzekerheidsbeoordeling van meteorologische forceringen uit heranalysegegevens (GEOS5, ERA5, AgERA5) onderzocht in Afrika en Zuidwest-Azië door middel van een directe vergelijking met in-situ weerstations en een onderlinge ruimtelijk-temporele vergelijking. De verschillen tussen deze datasets zijn niet-stationair en het meest uitgesproken in Centraal- en Zuidelijk Afrika en Zuidwest-Azië. Er werden foutvoortplanting toegepast om te analyseren hoe onzekerheden in ERA5 van invloed zijn op de schatting van ET voor een referentiegewas. De resultaten van twee methoden (Monte Carlo en Taylorontwikkeling) werden vergeleken, waarmee de afweging tussen nauwkeurigheid en rekenkosten bij het schatten van onzekerheid werd geïllustreerd.

In hoofdstuk 5 werden de samengestelde en relatieve onzekerheid geëvalueerd voor het WaPOR versie 3 Level 1 (WaPORv3L1) ET_a gegevensproduct, dat werd afgeleid met behulp van het ETLook-model. In-situ validatie ten opzichte van EC-metingen wijst op een samengestelde onzekerheid die vergelijkbaar is met die van andere satellietgebaseerde ET_a-producten. In-situ validatie kent echter nog steeds uitdagingen in droge en tropische klimaten, waar de prestatiestatistieken slecht waren. Triple collocation analysis (TCA) werd gebruikt om de relatieve onzekerheid te schatten in termen van RMSE en correlatiecoëfficiënt. Hoewel TCA waardevol is voor het blootleggen van de ruimtelijke variatie van onzekerheid, zijn de resultaten sterk afhankelijk van de keuze van de datasets en worden onzekerheden onderschat in vergelijking met in-situ validatie.

Om de implicaties voor toepassingen van satellietgebaseerde ET_a-schattingen te onderzoeken, werd in hoofdstuk 6 het WaPOR-project bekeken in termen van probleemformulering, modelkeuzes en configuraties, en de gevolgen van onzekerheden. De context van het WaPOR-project en de nadruk die daarin wordt gelegd op watergebruik in de landbouw en waterproductiviteit hebben geleid tot een bepaalde keuze van model

en configuratie. Hoewel onzekerheden voornamelijk worden gezien en beheerd als overwegend technische problemen, stimuleert het nadenken over methodologische onzekerheden het onderzoeken van alternatieve gegevens en modellen en het betrekken van gegevensgebruikers.

Op basis van de ervaring en kennis die uit dit onderzoek is opgedaan, wordt aanbevolen om aanvullende informatie over de gegevenskwaliteit, zoals de ex-ante en ex-post onzekerheidsbeoordelingen, op te nemen in de satellietgebaseerde gegevensproducten. Onzekerheidsbeoordelingen zijn een onderdel van het modelleringsproces die afhankelijk is van de specifieke methoden, referenties en methodologische keuzes. Daarom moet in toekomstig werk aanvullende informatie over bekende problemen en feedback van gebruikers worden opgenomen, terwijl zowel de ex-ante als de ex-post beoordelingsmethoden moeten worden verbeterd.

CONTENTS

1
INTRODUCTION

Satellites are increasingly used to acquire data and derive information in water sciences and management. The research leading to this dissertation is motivated by a question of several researchers and water managers: “How accurate and precise are estimates based on satellite-based data?”. Specifically, this research focuses on satellite-based estimation of evapotranspiration. This first chapter introduces readers to the definition, importance, quantification and satellite-based estimation of evapotranspiration, which sets the context for the research problem. To address the research problem, a conceptual framework based on probability theory was applied to understand and examine uncertainty. Based on that, research objectives and questions are formulated. This chapter outlines the overall research approach and design, concluding with an overview of thesis structure.

This chapter is partly based on:

Tran, B.N., van der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G. and Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. Hydrology and Earth System Sciences, 27(24), pp.4505-4528. https://doi.org/10.5194/hess-27-4505-2023

1.1 BACKGROUND

1.1.1 Evapotranspiration in water sciences and management

Evaporation is defined broadly as the vaporization of water from a liquid to a gaseous state. Terrestrial evaporation includes water vaporization from the soil and open water (soil and open water evaporation), from precipitation intercepted by vegetation foliage (interception loss), and water taken from the soil by plants (transpiration). Transpiration and evaporation from soil and water surfaces (sometimes including canopy interception) are often combined into the term 'evapotranspiration' (ET), which is commonly used in agriculture, hydrology, and remote sensing.

The choice between the term 'evapotranspiration' and 'evaporation' reflects disciplinary training and perspective, and often causes confusion (Ahmad et al., 2025). Hydrologists such as Savenije (2004) dismissed the term 'evapotranspiration' as a failure to conceptualise different 'evaporation' processes separately. In presenting the bulk flux of water vapor from the land surface, Miralles et al. (2020) advocated the simpler term 'evaporation' as a more physically accurate and consistent than 'evapotranspiration'. However, the term 'evapotranspiration' has been used more widely (Miralles et al., 2020). Ahmad et al. (2025) contended that 'evapotranspiration' better illustrates the complexity of coupled physical and physiological process, fostering multidisciplinary collaboration, and maintaining relevance across scientific fields. For multidisciplinary relevance, this dissertation uses the abbreviation ET as equivalent to 'total evaporation'. For clarity, ET here includes interception loss as a component of abiotic evaporation (Oliveira et al., 2024).

ET is the key link between water and energy balances of the land surface. When water evaporates, it absorbs energy from the surface as latent heat. The latent heat flux (LE) [$W \cdot m^{-2}$] is the energy used to convert liquid water into water vapor during evapotranspiration: $LE = \lambda \times ET$, where ET [$m \cdot s^{-1}$] is the rate of water evaporated and λ [$J \cdot m^{-3}$] is the specific latent heat of vaporisation of water. LE is a major flux in the surface energy balance (Brutsaert, 1982; Rijtema, 1965):

$$R_n = G + H + LE \tag{1.1}$$

where R_n [$W \cdot m^{-2}$] is net radiation, G [$W \cdot m^{-2}$] is soil heat flux, H [$W \cdot m^{-2}$] is sensible heat flux. The typical value of λ is 2.45 $MJ \cdot kg^{-1}$ or 2.45×10^3 $MJ \cdot m^{-3}$ (assuming the density of water is 1000 $kg \cdot m^{-3}$ at 20°C).

In the terrestrial water balance, ET is recognised as the second-largest water flux after precipitation, and estimated to be more than 60 % of precipitation on land (Korzoun et al., 1978; Oki and Kanae, 2006). The partitioning of precipitation over a catchment into

actual ET (ET_a) and runoff is mainly controlled by the relationship between the total precipitation and the potential ET (ET_p) (Budyko, 1974). The latter, a concept introduced by Thornthwaite (1948), denotes the potential amount of water that the atmosphere could remove from the surface if water were abundantly available. Both actual ET_a and ET_p are fundamental in studying and understanding water cycles and climatic conditions, which consequently impact ecosystems, water resources, and food production (Fisher et al., 2017, 2011).

In hydrology, ET is often viewed as "water loss" since once the water is evaporated, it is no longer immediately accessible to plants and animals at a location. Higher ET means less water available for streamflow and groundwater recharge. However, evaporated water is not lost; it continues its journey in the water cycle, returning as precipitation elsewhere through moisture recycling (Eltahir and Bras, 1996; Savenije, 1995; van der Ent et al., 2010). Jackson and Head (2022) argued that the conceptualization of ET_a as the largest loss in the water balance of a territory has influenced how atmospheric water is managed and who controls it, especially in the dry and hot regions. Treating ET_a as 'green water' flow to be controlled and managed—like 'blue water' (rivers and groundwater)—also necessitates scientific and technical tools to monitor and influence water resources management and planning (Falkenmark and Rockström, 2004; Jewitt, 2006).

In agricultural water management, estimating ET is crucial for designing and operating water supply as ET_a equals the crop water consumption or 'consumptive use' (ASCE, 1930). Water consumption by plants from the soil moisture in the root zone enables carbon uptake through photosynthesis, prevents heat stress, and supports nutrient transport within plant stems (Katul et al., 2012). Therefore, crop growth requires an appropriate amount of water (Jensen et al., 2016), which is governed by numerous factors including climate, soil properties, crop type, growth stages, and management practices (Steduto et al., 2012). To determine crop water requirements, several standardised methodologies have been developed, mainly based on the potential ET for a reference crop such as grass or alfalfa, also called reference ET (ET_o) (Allen et al., 1998; ASCE, 2005; Doorenbos and Pruitt, 1977).

1.1.2 Quantification of ET

The process of ET_a depends on numerous factors, including the atmospheric and vegetation conditions, the water availability in the soil, water bodies, and the canopy, as well as the surface resistances to ET_o (Monteith, 1965; Penman, 1956; Shuttleworth and Wallace, 1985). As a result, a great variety of ET_a quantification methods have been developed over centuries (Brutsaert, 1982). These methods differ on the spatial and

temporal scales[1] of application, theoretical basis, sources of input data (**Table 1-1**). Diverse combinations of theoretical frameworks, spatiotemporal scales, and input data sources result in a wide range of estimation methods, consequently yielding varying results. For example, at the field scale, water balance-based methods such as lysimeters and evaporation pans give different results than turbulence-based methods such as Eddy Covariance (EC) flux tower (Brutsaert and Parlange, 1998; Ding et al., 2010; Hirschi et al., 2017; Rowshon et al., 2014). At the continental and global scales, different ET_a models, parameters, input data sources and processing techniques result in a wide range of ET_a estimates (Chen et al., 2014; Long et al., 2014).

Table 1-1. The choices of theoretical basis, spatial and temporal support, and sources of input data in ET_a estimation.

Theoretical basis	water balance/budget, surface energy balance, aerodynamics, turbulent eddies
Spatial support	pixel/grid size[2], field, basin/catchment, region, continent
Temporal support	instantaneous, sub-daily (day/night), daily, sub-monthly (e.g., 5-day, 8-day, 16-day), monthly, yearly
Sources of input data	measured by in-situ instruments, measured by ground-based remote sensor, measured by sensors on remote platforms (e.g. drones, aeroplanes, satellites)

The complexity of on-site ET_a measurement makes it both difficult and expensive to routinely measure and monitor ET_a temporally and spatially, as this requires a dense network of in-situ gauging stations. Given its importance for many ET-based science and applications, many have advocated for advancing capabilities for ET_a observation, specifically regarding accuracy, spatial and temporal resolution, spatial coverage, and long-term monitoring (Fisher et al., 2017; Huang et al., 2025). For this purpose, satellite remote sensing (RS) currently offers the most feasible means to regularly monitor ET_a spatially over large surface areas (e.g., river basin, irrigation schemes) up to global coverage.

[1] Scale (both spatial and temporal) is best described as a triplet of support (or grain), spacing and extent. Support is the volume, shape, size, and orientation that a measurement represents (Blöschl and Sivapalan, 1995).

[2] In the realm of remote sensing, the spatial support of pixel values can be equivalent to the resolution of remote sensing images, which is the (average) size of its constituent pixels (Bierkens et al., 2000).

1.1.3 Satellite-based ET estimation

Satellite remote sensing (RS) is the process of acquiring data about an object or phenomenon using instruments mounted on satellites. Satellite sensors measure the radiation reflected or emitted from the earth's surface or atmosphere in different regions of the electromagnetic spectrum. These instruments generate electrical current corresponding to the intensity of reflected or emitted radiation and convert electrical signals to discrete values in digital imagery (Campbell and Wynne, 2011, p.103). Raw digital imagery acquired by sensors on satellites (*satellite observations* in **Figure 1-1**) must undergo a chain of processing and analysis to derive useful information for applications, such as water resources management and hydrological sciences.

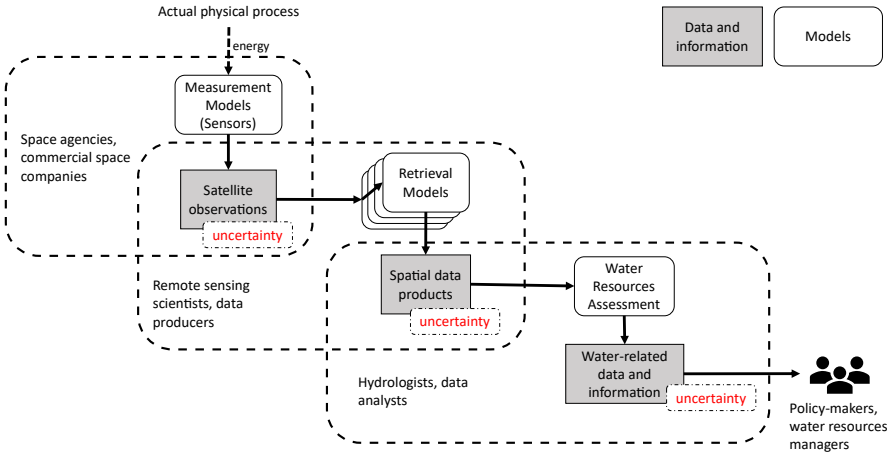


Figure 1-1. Generation of uncertainty in applications of remote sensing in water resources management. Uncertainty is an inherent attribute of each data and information component.

Drawing from the work on information theory and the physics of information (Boisot and Canals, 2004), Bennett (2025) defines 'satellite data' as "that which contains information about the Earth", 'satellite information' as "what which can modify understandings of the Earth", and 'satellite knowledge' as "that which enables its producer to act in adaptive ways in and upon". For example, satellite data can be a flat binary file containing an array of numbers, which contains information about variability and quantity of a geophysical variable, such as maps of ET_a. From this information, one can produce hydrological knowledge such as about variation and anomaly of ET_a over time, which suggests potential actions regarding water use and demand.

Because analysis-ready satellite imagery involves multiple processing steps that use diverse methods and ancillary data, it aligns more closely with satellite information than

with satellite data (Bennett, 2025). It both contains information and reshapes understanding, since its form is shaped by processing and calculation methods. Such outputs are often called 'satellite data products' to distinguish them from raw data.

ET_a and other hydrological variables are not directly measured by sensors onboard satellites; instead, they are derived through retrieval models, the outputs of which are often referred to as "data products". Some data products, including those for ET_a, are generated through several levels of data processing (ESA, 2021; NASA, 2021), a process illustrated by the multiple layers of retrieval models shown in **Figure 1-1**. It should also be noted that 'directly measured' also involves a measurement model (i.e. the way we relate the numerical output of a measuring instrument with the underlying physical state of the object that we want to measure).

Many satellite-based ET_a models have been developed and applied to derive estimation from RS sensor signals. These models estimate ET_a from visible and/or thermal infrared RS data and include widely recognized models such as SEBAL (Surface Energy Balance Algorithm for Land) (Bastiaanssen et al., 1998), TSEB (Two-source Surface Energy Balance) (Kustas and Norman, 1999), SEBS (Surface Energy Balance System) (Su, 2002), METRIC (Mapping Evapotranspiration at high Resolution with Internalised Calibration) (Allen et al., 2007), ALEXI (Atmosphere-Land Exchange Inverse) (Anderson et al., 2011), PT-JPL (Priestley-Taylor from Jet Propulsion Lab) (Fisher et al., 2008), and GLEAM (Global Land Evaporation Amsterdam Model) (Miralles et al., 2011b).

The diversity of models (Li et al., 2009; Mohan et al., 2020; Wang and Dickinson, 2012; K. Zhang et al., 2016), input data sources, and processing techniques result in a wide range of satellite-based global and regional ET_a estimates (Chen et al., 2014; Jiménez et al., 2011; Long et al., 2014). Retrieval models for ET_a estimates require access to the data, software or source code, and expertise in these models. Several projects have provided platforms to increase public access to various data products which are generated by these satellite-based ET_a models. These projects and outputs include MODIS16 (Mu et al., 2011), SSEBop (Senay et al., 2013), GLEAM (Miralles et al., 2025, 2011b), WaPOR (FAO, 2020a, 2018a, 2018b), ECOSTRESS (Fisher et al., 2020), and OpenET (Melton et al., 2022).

1.2 PROBLEM STATEMENT

Remote sensing has a potential to provide water-related data and information more economically and efficiently than ground-based data collection methods alone (Hughes et al., 2015; van Dijk and Renzullo, 2011). There is now an unprecedented opportunity to estimate ET globally thanks to the proliferation of sensors and satellites and advances in computational power. However, knowledge gaps remain regarding the uncertainty of spatial data derived from satellite remote sensing. These uncertainties can cascade into

misinformation and misinterpretation when applying these products in water management at various scales (**Figure 1-1**). For example, uncertainty in ET_a estimation introduce uncertainty in the estimation of irrigation water use and demand (McDermid et al., 2023; Rhenals and Bras, 1981), which might lead to welfare losses for farmers and hinder efforts to manage the impacts of irrigation withdrawals (Foster et al., 2020).

Understanding the sources and impact of uncertainty helps data users know what level of confidence they can have in ET_a estimates and the inferred information about water resources (e.g., crop water consumption, water depletion). Given that more satellite-based ET_a data products are becoming available, information about the uncertainties in satellite-based ET_a estimates is important for data users (i.e., water managers and policymakers) to apply them properly. Characterising these uncertainties is needed to better communicate derived water-related information to decision-makers.

While many studies have evaluated the performance of satellite-based ET_a models, none of them has concluded that a single model performs best in all situations (e.g., Ferguson et al., 2010; Vinukollu et al., 2011b). The strengths and limitations of satellite-based ET_a models are mostly related to the application and region each was developed for or the assumptions used in the evaluation. For example, some satellite-based ET_a models have been developed for agricultural application and, thus, tested more in homogeneous areas (e.g. single crop) than in heterogeneous areas (e.g. a mixture of different vegetations types and bare soil). As a result, when certain models are applied to generate global data products, it remains uncertain whether they can provide reliable estimates across all intended environments (Fuentes et al., 2024; Shufen Pan et al., 2020).

1.3 CONCEPTUAL FRAMEWORK

1.3.1 Concepts of uncertainty

Uncertainty can be generally defined as the state of being not completely confident or sure of something. The terms ‘error’, ‘accuracy’, ‘bias’, and ‘precision’ are sometimes used to characterize uncertainty. All these terms indicate quantifiable information about what is certain or uncertain. However, they are different from ‘uncertainty’ by definition (Foody and Atkinson, 2003; Heuvelink, 1998; Loew et al., 2017). The ‘error’ represents the difference between what is measured and its true value (JCGM, 2012). The true value is the (exact) value according to the theoretical definition of the variable being measured or estimated. If we perfectly know the true value, we have no measurement error, which eliminates uncertainty. In essence, uncertainty stems from unknown true values and errors of a measurement or estimation.

When a measurement can be repeated, its uncertainty can be described using probability distributions of the measured values or measurement errors compared to a reference

(Foody and Atkinson, 2003; Montanari, 2007; Povey and Grainger, 2015). **Figure 1-2** illustrates the relationship between uncertainty and other related terms when uncertainty is described by the probability distribution of a measured value or error. 'Accuracy' is defined as "the expectation (i.e., expected value) of overall error' (e.g., Foody and Atkinson, 2003). 'Bias' (i.e., the difference between the estimated value and the true value) is considered a measure of inaccuracy. Likewise, 'precision' (i.e., a measure of the scatter of a number of estimated values) can be described using standard deviation and variance of the probability distribution of measured values since they both denote the errors spread around the mean.

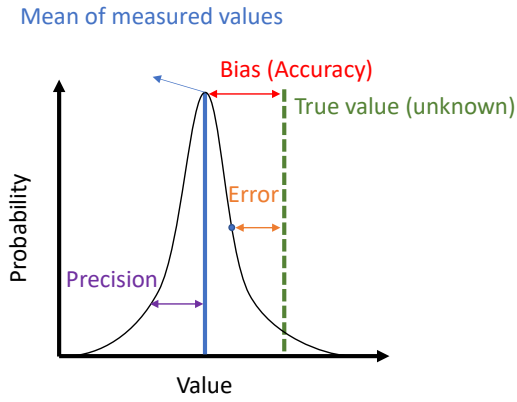


Figure 1-2. Uncertainty as described by the probability distribution of measured values. Adapted from JCGM (2012) and Povey and Grainger (2015).

However, some uncertainties cannot be described using a probability distribution function in modelling or measurement. These are called the 'known, unquantifiable unknowns' (i.e., what we know to exist but cannot quantify) and the 'unknown unknowns' (i.e., what we do not know to exist because we cannot observe) (Povey and Grainger, 2015). The suitability of probability theory for quantifying uncertainty is widely debated in hydrological science (e.g., Beven, 2016; Nearing et al., 2016). Nearing et al. (2016) argue that there is epistemological uncertainty that must be recognized before selecting probability theory as the framework to estimate epistemic uncertainty (i.e., what we do not know for certain). Their uncertainty classification includes philosophical and linguistic aspects that are not quantifiable. Uncertainty assessment of satellite remote sensing data products typically reports quantifiable errors but not 'unknown' and 'unquantifiable' errors (Povey and Grainger, 2015).

1.3.2 Sources of uncertainties in satellite-based estimation of ET

ET_a is derived from models rather than being directly measured by sensors; thus, ET_a data products are characterized by a high level of processing by data providers (ESA, 2021; NASA, 2021). Raw satellite imagery undergoes a chain of pre- and post-processing and modelling to generate useful data and information for applications (**Figure 1-3**). The retrieval models of data with low level of processing (e.g., radiance, vegetation indices) share common formulas and usually require only raw satellite images. High-level data products like satellite-based ET_a relies on various models with different concepts, assumptions, and data sources. Raw satellite images undergo processes (e.g., radiometric, geometric, and atmospheric corrections) to generate higher-level data products such as surface reflectance, land surface temperature, emissivity. These processes introduce uncertainty into the input data required for satellite-based ET_a estimation, such as vegetation indices (Hadjimitsis et al., 2010) and land surface temperature (Li et al., 2013). Therefore, the uncertainty of satellite-based ET data products is related to uncertainty in model structure and parameterization, and input data.

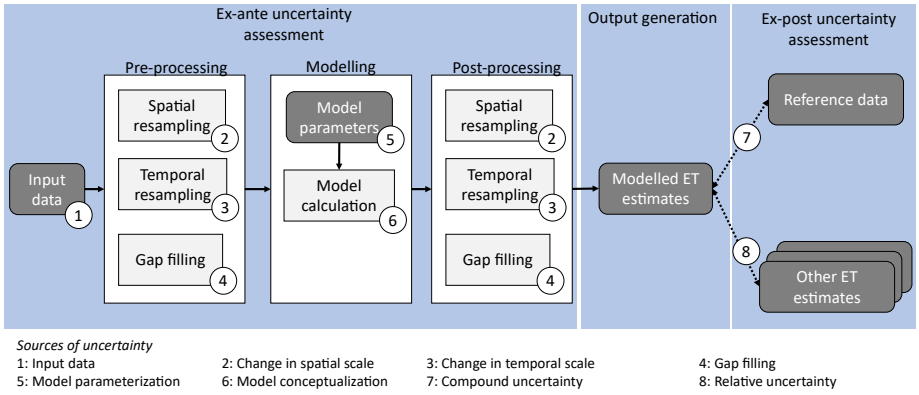


Figure 1-3. The sources of uncertainty in ET_a estimates from the typical workflow in remote sensing-based models. Compound uncertainty is the aggregation of all uncertainties from input data, change of temporal and spatial scale, gap filling, model parameterization, and model conceptualization.

Satellite-based data are typically acquired when satellite passes over specific areas of interest, resulting in essentially instantaneous ET_a estimates. Because many operational applications necessitate ET_a estimates over longer time intervals, such as daily, 8-day, or monthly totals, various methods have been developed to upscale these instantaneous satellite-based ET_a estimates to daily values (Jiang et al., 2021). Moreover, the designated resolution and the return interval of satellites might not be suitable for operational

applications. Thus, gap-filling and spatial downscaling steps are common in many retrieval models of satellite-based ET_a data products.

Changes in spatial and temporal scales and gap filling during data pre- and post-processing steps also introduce more uncertainty. Modelled estimates are typically validated against a reference that is considered more accurate. In earth sciences, model validation is always partial and can only demonstrate the agreement between modelled results and observation (Oreskes et al., 1994). For satellite remote sensing, the reference is often the estimates from in-situ measurements because they are considered to provide a more accurate representation of actual ground conditions than satellite retrievals. Woodhouse (2021) argued that in-situ measurements should not be considered the 'ground truth' for satellite-based estimates; instead, they should be treated as estimates that have their own associated uncertainty. In this research, the errors compared to the reference dataset are termed 'compound uncertainty' because they aggregate all sources of uncertainty. Meanwhile, comparison with other equivalent ET_a estimates yields 'relative uncertainty', as these uncertainty estimates are determined in relation to the chosen estimates. Since reference data also carries uncertainty, 'compound uncertainty' can be seen as a form of 'relative uncertainty' specially determined against the reference data.

1.3.3 Uncertainty assessment

Uncertainty assessment refers to the process of identifying, characterizing, and quantifying the various sources of uncertainties in each measurement, estimate, model result, or analysis. Its goal is to understand how these uncertainties might affect the results, conclusions, or decisions, and to communicate the level of confidence in those results. Uncertainty in satellite-based data products can be assessed using various methods and approaches at different stages of production. In this study, uncertainty assessments are categorized as ex-ante and ex-post (**Figure 1-3**). **Ex-ante assessments** are performed prior to output data generation to analyse and estimate potential errors and uncertainties, which are crucial steps in model development and calibration. **Ex-post assessments** are conducted after output data generation, using independent datasets (e.g., a reference dataset or other estimates) to evaluate the compound and relative uncertainty of the satellite-based data product.

The sources of uncertainty mentioned previously (**Figure 1-3**) stem from the technical factors of a given satellite-based ET_a model such as its structure, parameters, or input data. These "**technical uncertainties**" can be assessed through standard methods like probability and statistics. For example, the uncertainty from parameters and input data can be quantified with uncertainty analysis, also called uncertainty propagation or error propagation (Crosetto et al., 2001; Heuvelink, 1998).

There are many techniques for **uncertainty propagation** and their suitability depends on several factors, including the number of uncertain inputs, uncertainty distribution and correlation of input variables, and model linearity (Mohammadi and Cremaschi, 2022). For satellite-based models, analytical techniques which are based on propagation of moments formulas (Taylor, 1997) are not always suitable because these models have complex relationships and input uncertainty is not always normally distributed. Numerical techniques are generally applicable to outputs from satellite-based models (Crosetto et al., 2001; Heuvelink, 1998).

The contribution of each input factor to the total uncertainty in the model output can be determined by **sensitivity analysis** (Crosetto et al., 2001; Rakovec et al., 2014; Saltelli et al., 2021). Such analysis is primarily used to identify the factors that contribute most to the model uncertainty (Saltelli et al., 2019). In general, there are two main approaches to sensitivity analysis: local and global sensitivity analysis. Local sensitivity analysis defines the model's sensitivity to an input factor (e.g., parameter or variable) as the first-order partial derivative of the model with respect to this input factor (Saltelli et al., 2019). In contrast, global sensitivity analysis explores the whole variation range of input factors (Razavi and Gupta, 2015).

Validation is often applied to confirm a data product's fit-for-purpose instead of sensitivity analysis and uncertainty analysis of its retrieval model (Crosetto et al., 2001). The definition of validation in modelling is context-dependent and has become more well-defined over time (Bellocchi et al., 2011). Model validation does not prove that the model is true but rather proves that it is empirically adequate (Oreskes et al., 1994). A valid model is one that does not contain known or detectable flaws and is internally consistent, rather than being an assertion of the reality. Meanwhile, validation of model results involves quantifying the accuracy compared to a reference (often in-situ datasets), which proves the validity of the data for its intended application.

In RS, validation often only refers to the data itself and not the model (Bayat et al., 2021; Loew et al., 2017; Wu et al., 2019b). Because RS-derived data products are model results, their validation depends on the quality and quantity of input parameters and the accuracy of auxiliary hypotheses that were used to derive them (Oreskes et al., 1994). Therefore, validating a satellite-based ET_a model does not imply that the model can be applied with any forcing data or settings to produce accurate output.

Uncertainty also arises from differences in modelling approaches and subjective choices when establishing the workflow illustrated in **Figure 1-3**. These "**methodological uncertainties**" are sometimes not consciously recognized by modelers and often not easily captured by statistics (Melsen, 2022; Melsen et al., 2019). Methodological uncertainty is present throughout the entire workflow in **Figure 1-3**, including the choice of uncertainty assessment method and reference data. It arises from the choices made by scientists and engineers, such as the specific algorithms used to retrieve data, the

assumptions embedded in physical models, techniques for handling missing data or fusing different data sources, and even the selection of uncertainty assessment methods.

1.4 RESEARCH OBJECTIVES

This research aims to enhance the understanding of both technical and methodological uncertainties in satellite-based ET_a estimation and further explore the implications of these uncertainties for water management. Specifically, the study aims to:

1. Identify how uncertainty has been evaluated in academic literature on satellite-based ET_a models and data products, and the status of uncertainty reported.
2. Evaluate the technical uncertainties of a satellite-based ET_a data product through both ex-ante and ex-post uncertainty assessments.
3. Explore the implications of uncertainties in a satellite-based ET_a data product for its applications.

1.5 RESEARCH APPROACH AND DESIGN

This research approached the uncertainties in satellite-based estimation of ET_a from two perspectives: one informed by academic literature and the other by direct engagement with the development and application of a satellite-based ET_a model (**Figure 1-4**). From the first perspective (Part I), a systematic quantitative literature review was conducted to examine how uncertainties in satellite-based ET_a are assessed in academic literature published between 2011 and 2021, a decade before the inception of this research. This approach examines the evidence base for claims regarding the accuracy and precision of satellite-based ET_a estimates. Quantified uncertainties from the reviewed literature were then surveyed and analysed using a meta-analysis approach to determine the overall status of uncertainty in satellite-based ET_a within academic literature.

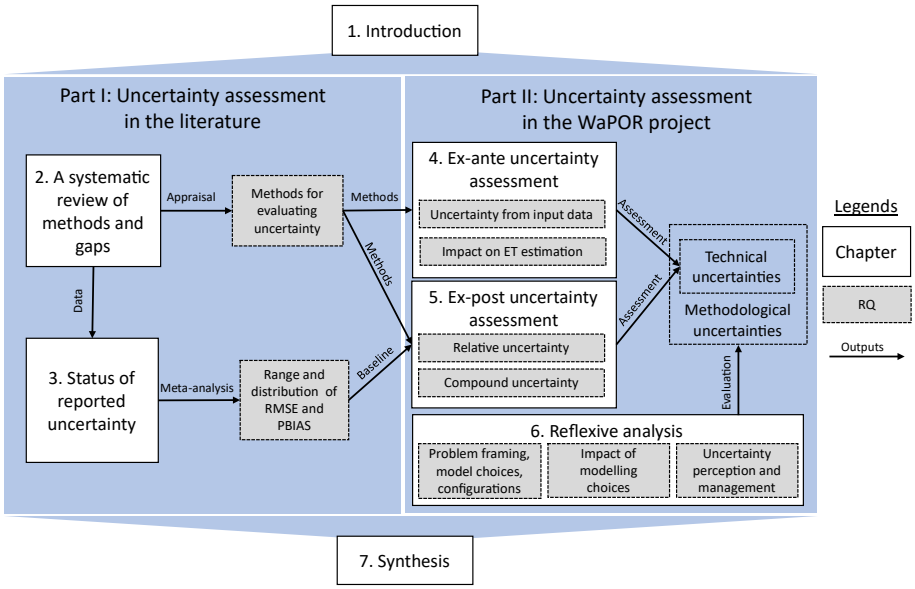


Figure 1-4. The research approach and outline of this thesis.

The second perspective (Part II) focuses on a specific satellite-based ET_a model for two primary reasons. First, the implications of uncertainties in satellite-based ET estimation are case-specific and depend on the actors and decisions involved in the development and application of a model (FAO, 2023a). Second, this research was funded by and undertaken as part of the WaPOR project[3], which aims to monitor water productivity using remote sensing data (FAO, 2020a, 2018a, 2018b). This allowed the research to benefit from the project's resources and acquired data. To some extent, the funding source pre-determined the satellite-based ET_a model to be evaluated. This work contributes to the WaPOR project's technical work package focused on the quality assessment of its ET_a data, independent from the data provider.

The research design both shaped and was shaped by the quality assessment activities within the WaPOR project. Based on the appraisal of uncertainty assessment methods in Part I, a set of suitable methods was selected to evaluate technical uncertainties in both the input data and the model (ex-ante assessments), as well as in the derived data product (ex-post assessments). The range and distribution of uncertainty metrics from Part I served as the baseline for ex-post uncertainty assessment conducted in Part II. For the ex-ante uncertainty assessment, the uncertainties in meteorological forcings and their impact

[3] https://www.fao.org/in-action/remote-sensing-for-water-productivity/en

on the estimation of ET for a reference crop in Africa and Southwest Asia were assessed (**Figure 1-5**). This assessment was conducted as part of the WaPOR project's first quality assessment report, focusing on the period from 2018 to 2022 (FAO and IHE Delft, 2024). For the ex-post uncertainty assessment, the relative and compound uncertainties of the WaPOR version 3 global data product was assessed across various locations with in-situ references worldwide, also for the same period. This assessment was conducted as part of the WaPOR project's second quality assessment report. Additionally, the compound uncertainty from direct validation was compared with the relative uncertainty obtained from alternative methods to investigate the applicability of these approaches for future quality assessments.

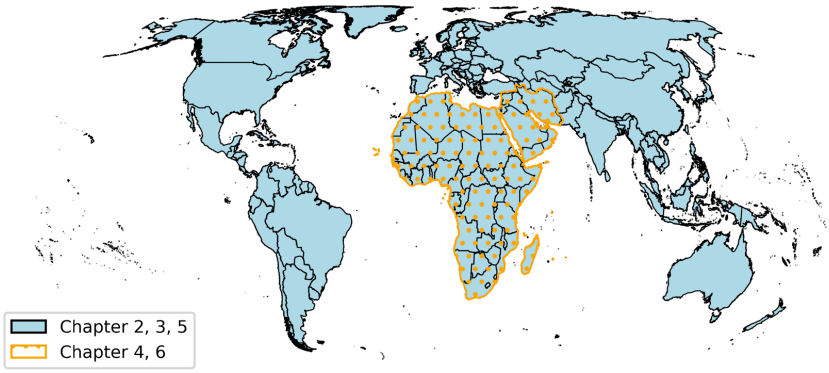


Figure 1-5. Geographical regions corresponding to the areas of interest in each chapter.

A qualitative approach was adopted to explore the implications of uncertainties in satellite-based estimation of ET_a for its applications. This involved a reflexive analysis of the WaPOR project, specifically examining problem framings, model choices, and configurations, as well as the consequences of uncertainties. Drawing on Holland's (1999) definition of reflexivity as the process of "turning back upon, or taking account of, itself" this analysis focused on the WaPOR project's entire process of ET modelling, not just its final ET_a estimates, to understand how decisions and assumptions influenced its outcomes. Using a reflexive lens (Walsh, 2003), lessons were drawn from multiple studies and secondary sources, including water accounting studies in the Litani and Niger Basins, and irrigation advisory services in Egypt and Tunisia.

1.6 RESEARCH QUESTIONS

The research process for this dissertation involved the iterative formulation and refinement of specific research questions as the research approach and design were implemented. **Table 1-2** presents research questions addressed in this dissertation.

Table 1-2. Research questions and their links to research objectives and thesis chapters.

Research objective	Research questions (RQs)	Addressed in chapter
Objective 1	RQ 1.1: What are the common and emerging methods used to assess uncertainty in satellite-based ET_a estimates? RQ 1.2: What are the advances and caveats of these uncertainty assessment methods? RQ 1.3: In which contexts are the uncertainties of satellite-based ET_a assessed with these methods?	Chapter 2
	RQ 1.4: Which metrics are used to report uncertainty in satellite-based ET_a estimates? RQ 1.5: What is the typical range of uncertainty in satellite-based ET_a estimates globally based on performance metrics reported from previous studies?	Chapter 3
Objective 2	RQ 2.1: What is the uncertainty of meteorological forcing from the input datasets for modelling ET? RQ 2.2: What is the impact of uncertainty in meteorological variables on the estimated ET_o?	Chapter 4
	RQ 2.4: What is the compound uncertainty of the WaPOR project's ET_a data product? RQ 2.5: How well can the relative uncertainty of the WaPOR's ET_a data product predict the compound uncertainty from direct validation?	Chapter 5
Objective 3	RQ 3.1: What are the problem framings that influence the model choices and configurations in the WaPOR project? RQ 3.2: What the impacts of subjective modelling choices observed in the WaPOR project? RQ 3.3: How are uncertainties stemming from the development and application of the ETLook model specifically perceived and managed within the WaPOR project?	Chapter 6

1.7 THESIS OUTLINE

The rest of this thesis includes two parts and a synthesis chapter (**Figure 1-4**). Each chapter presents a different set of methods; therefore, the detailed methods are presented in each chapter instead of in a separate chapter.

Part I focuses on the assessment of uncertainty on satellite-based ET_a estimation in the academic literature, which comprises Chapter 2 and 3. **Chapter 2** examines the methods and existing gaps in uncertainty assessment through a systematic quantitative literature review. **Chapter 3** presents the status of uncertainty in satellite-based ET_a estimates reported in the literature through a meta-analysis of studies that employed the most common uncertainty assessment method.

Part II of this dissertation focuses on the satellite-based ET_a model that was developed and applied in the WaPOR project, known as ETLook. Part II begins with a brief introduction to the WaPOR project, the ETLook model, and the global data product generated using this model. **Chapter 4** presents an ex-ante uncertainty assessment of the ETLook model, investigating the uncertainties in the model's meteorological forcings, and analysing how these uncertainties impact the estimation of ET for a reference crop. **Chapter 5** provides an ex-post uncertainty assessment, evaluating the compound and relative uncertainty in outputs of the ETLook model, specifically the WaPOR version 3 data product. **Chapter 6** contains an analysis on how uncertainties in the ETLook model are influenced by the problem framings, model choices and configurations, and how uncertainty in its outputs was perceived and managed within the WaPOR project.

Finally, the key outputs of this research and the answers to the research questions are synthesized in **Chapter 7**. This synthesis aims to provide recommendations for water managers, policymakers, and RS data providers regarding the quality assessment of satellite-based ET data products. **Chapter 7** also includes the author's personal reflection on the research processes that led to this dissertation.

PART I: UNCERTAINTY ASSESSMENT IN THE LITERATURE

In Part I, the uncertainty assessments of ET_a estimate derived from satellite-based models from previous studies are appraised through a systematic quantitative review of their methods (Chapter 2) and a meta-analysis of their results (Chapter 3).

2

A SYSTEMATIC REVIEW OF METHODS AND GAPS

This chapter is based on:

Tran, B.N., van der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G. and Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. Hydrology and Earth System Sciences, 27(24), pp.4505-4528. https://doi.org/10.5194/hess-27-4505-2023

2.1 INTRODUCTION

Previous literature reviews have discussed satellite-based ET_a estimates and uncertainty (**Figure 2-1**). Many of these reviews focused on outlining the methods to estimate ET_a using satellite-based models (e.g., Courault et al., 2005; Kustas and Norman, 1996; Wang and Dickinson, 2012; K. Zhang et al., 2016) and sometimes discussed the uncertainties in the estimation (Glenn et al., 2011; Kalma et al., 2008; Karimi and Bastiaanssen, 2015). However, none of these explored how uncertainties of satellite-based ET_a estimates are currently being assessed, which is an important issue in remote sensing and the production of spatial data (Bielecka and Burek, 2019; Mayr et al., 2019; Wu et al., 2019b). In an overview of Essential Climate Variables, Bayat et al. (2021) concluded that satellite-based ET_a data products lack a good practice protocol for operational validation, compared to other variables. Meanwhile, in-situ measurements of ET_a also suffer from errors and uncertainty (Allen et al., 2011a) and, thus, require complete documentation that provides sufficient information to ascertain the expected accuracy and representativeness of the reported ET_a estimates (Allen et al., 2011b).

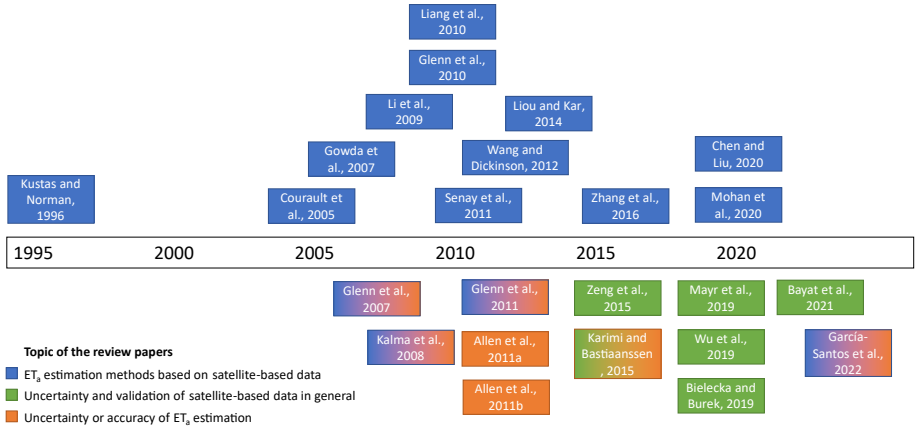


Figure 2-1. Previous literature reviews on satellite-based ET_a estimation, uncertainty, and validation of RS-derived data.

These reviews highlight the need to better advance the uncertainty assessment of satellite-based ET_a, leading to the following research questions:

- What are the common and emerging methods used to assess uncertainty in satellite-based ET_a estimates?
- What are the advances and caveats of these uncertainty assessment methods?
- In which contexts are the uncertainties of satellite-based ET_a assessed with these methods?

To answer these questions and build on existing literature, this chapter provides a literature review of previous studies that assessed the uncertainty or accuracy of satellite-based ET_a models or the output data products of these models. Given that many literature reviews on the uncertainty or accuracy of ET_a estimation have been published prior to 2011 (**Figure 2-1**), we focus on the period from 2011 onward to evaluate whether the studies in this period adopted the valuable contributions and recommendations from these previous reviews. Given the growing volume of literature published in the field, we followed a systematic quantitative review approach to avoid subjectivity or bias towards particular products, authors, or approaches. We identified research articles with a set of predetermined criteria and categorized these articles based on the methods used to assess uncertainties and the context where uncertainties were evaluated. We then quantified the number of articles per category to identify any trends or gaps in literature. Furthermore, we appraised the advances and caveats of the existing methods and provided recommendations for future studies.

The rest of this chapter is organized as follows: Section 2.2 describes the methods used to analyse literature. The results of the literature analysis concerning assessment methods and the context when these methods are used are discussed in Section 2.3 and 2.4. Finally, Section 2.5 summarizes the key points and recommendations for future research.

2.2 SYSTEMATIC QUANTITATIVE LITERATURE REVIEW METHOD

In this literature review, we specifically focus on how quantifiable uncertainty in satellite-based ET_a estimates has been assessed in recent years (2011-2021). We employed Pickering and Byrne's (2014) systematic quantitative literature review method, which includes systematic search, categorization, and quantification of literature. We chose this approach to objectively highlight trends and gaps in current satellite-based ET_a uncertainty assessment methods through quantitative results. The literature search is systematic, but undeniably not exhaustive; thus, certain papers may be omitted if they do not meet the specified inclusion criteria.

2.2.1 Identification and database search

The academic electronic databases Web of Science and Scopus were searched (last access: 24/07/2023) using the combination of the three search terms: "evapotranspiration", "remote sensing", and "uncertainty", or their variants (**Table 2-1**). The term "transpiration" and "interception" were not used since they only represent components of ET_a. Since different terms for satellite remote sensing, evaporation, and uncertainty can be used in the title and abstract, the variants of search terms were identified based on a set of 34 prior articles (Badgley et al., 2015; Bhattarai et al., 2019; Byun et al., 2014; Chen et al., 2015, 2014; Guillevic et al., 2019; He et al., 2020; Jiménez et al., 2011; H. C.

Jung et al., 2019; Khan et al., 2018; Kiptala et al., 2013; Li et al., 2015; W. Liu et al., 2016; Long et al., 2014; McCabe et al., 2016; B. Mueller et al., 2013; Mueller et al., 2011; Shufen Pan et al., 2020; Rajib et al., 2018; Ramoelo et al., 2014; Rodell et al., 2011; Ruhoff et al., 2013; Senkondo et al., 2019; Sörensson and Ruscica, 2018; Vinukollu et al., 2011a, 2011b; S. Wang et al., 2015; Westerhoff, 2015; Xu et al., 2019; Yang et al., 2017, 2015b; Yilmaz et al., 2014; Yuan et al., 2012; Zeng and Cai, 2018). These articles were identified in a preliminary literature search prior to this study.

Table 2-1. Search terms and variants. Search terms were combined using AND operator and variants were combined using OR operator. The asterisk * was used to include similar terms.

	Search terms combined by < AND >		
Variants combined by < OR >	Evaporation	Remote sensing	Uncertainty
	Evapotranspiration	Remotely-sensed	Accuracy
	Latent heat	Remotely sensed	Data quality
		Earth observation	Variability
		Satellite*	Reliability
		Global ** product	Evaluat*
		Global ** data*	Validat*
			Performance

The search result was limited to a publication date from 2011 to 2021, and duplicates were removed. Only English articles (> 99 % of results) that reported original research and were published in scientific peer-reviewed journals were considered. Review papers, conference proceedings and grey literature were not included because they have different formats and provide limited details of the methods used for uncertainty assessment.

2.2.2 Relevance and eligibility screening

From the search result, we identified papers that attempted to assess the accuracy or uncertainty of one or more satellite remote sensing-based estimations of terrestrial ET_a, either from model simulations or analysis-ready data products. The models of interest were diagnostic satellite-based ET_a models[4], such as the models that were reviewed by Chen and Liu (2020), Courault et al., (2005), and Zhang et al. (2016). To identify relevant papers, we screened the title and abstract using the ASReview software, a semi-automated screening system that incorporates an active learning classifier to rank the order of papers based on their relevance to the articles that were included previously (van de Schoot et al., 2021) [Website: https://asreview.nl/]. ASReview assists by identifying 95 % of the eligible studies after screening between 8 to 33 % of the sampled studies (van de Schoot

[4] Diagnostic satellite-based ET models are static models that estimate ET at single period of time (snapshot) using satellite data as the primary inputs for independent variables in the models.

et al., 2021). Based on the number of articles and the efficiency of ASReview, we established criteria to stop screening when 100 irrelevant records had been found consecutively (3 % of the total records) and at least 10 % of the total records had been screened. After screening titles and abstracts, we assessed the eligibility of each paper by reading the full-text articles and finally included 676 articles in our review (**Figure 2-2**).

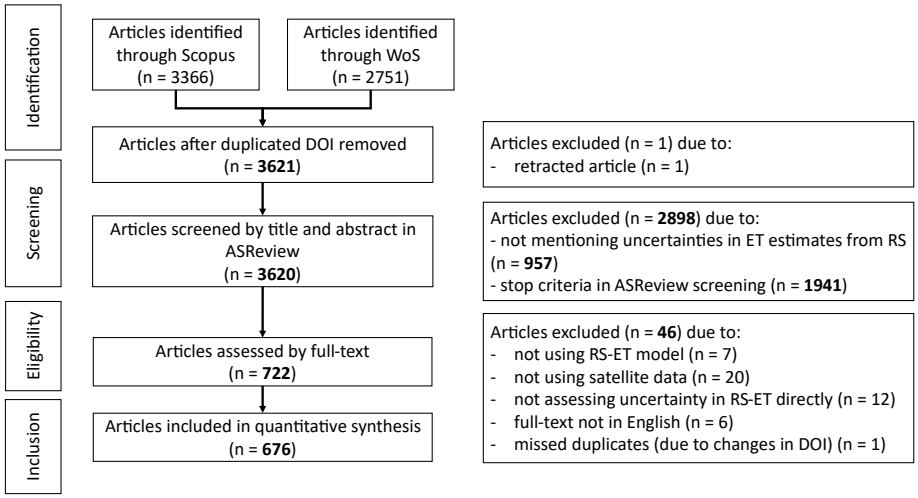


Figure 2-2. Results of article selection from database search (identification), title and abstract screening (screening), and full-text assessment (eligibility).

2.2.3 Article organization and analysis

Each included article was classified into categories based on methods, objectives of the study, and results (**Table 2-2**). The total number and percentage of research papers per category were then synthesized from the literature database, and the patterns and trends in assessing the uncertainty of satellite-based ET_a were discerned.

Table 2-2. Categories and subcategories used to organize the included papers.

Category group	Categories
Objective of the study	Model development, Model improvement, Model implementation, Product evaluation, Model evaluation
Sources of uncertainty	Compound uncertainty, Relative uncertainty, Change of spatial scale, Change of temporal scale, Model parameterization, Input data, Gap filling
Types of approach	Sensitivity analysis, Uncertainty propagation, Validation, Inter-comparison, Others
Uncertainty metrics	RMSE, bias, variance, …
Types of reference	In-situ measurement (EC, lysimeter…) Catchment water balance
Temporal support	Sub-daily, daily, from 5 to 16 days, monthly, season, annual
Spatial support	Less than 100 m, from 100 m to 500 m, from 500 m to 5km, from 5 km to 1º, more than 1º, basin, continent, global.
Spatial coverage	Field, region, continent, global

2.3 REVIEW OF THE METHODS FOR UNCERTAINTY ASSESSMENT

The selected articles assess uncertainty in satellite-based ET_a using mainly 8 approaches: (1) validation, (2) intercomparison, (3) sensitivity analysis, (4) evaluation of input data, (5) uncertainty propagation, (6) three-cornered hat and triple collocation (TCH/TC), (7) physical consistency, (8) ensemble of estimates. **Figure 2-3** shows the upset plot (Lex et al., 2014) of all reviewed articles by the approach of uncertainty assessment and the intersections of more than one approach. Most articles (532 out of 601) used a validation approach. There are a few other approaches that were less frequently used and often in combination with validation, as shown by the number of intersections with 'Validation' (**Figure 2-3**).

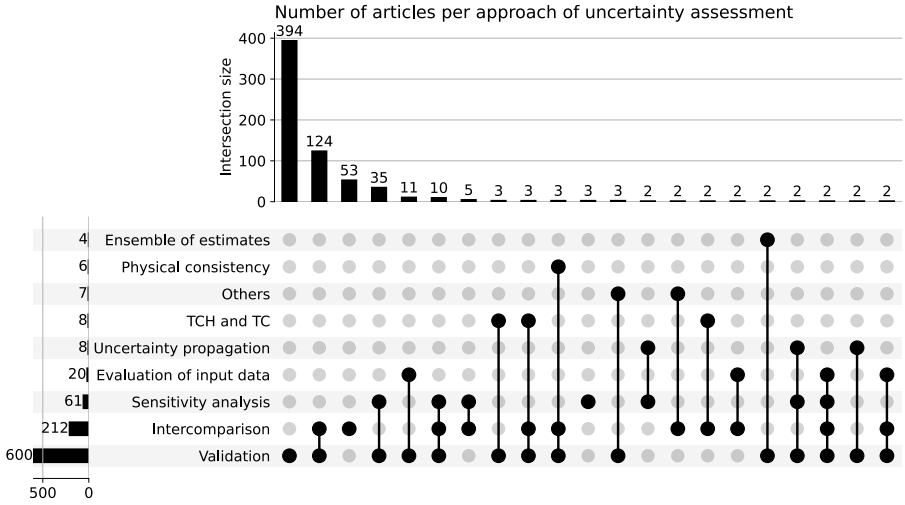


Figure 2-3. Uncertainty assessment approaches used in the reviewed articles ($N = 676$). The horizontal bar chart displays the number of articles using specific approaches (categories), while the vertical bar chart represents article counts within the intersections of multiple categories. Each vertical bar corresponds to an intersection in the column beneath it. Black circles denote the categories on the respective rows present in the intersection, while grey circles signify categories absent from the intersection. Intersections with less than 2 articles were excluded from the graph for improved presentation. TCH/TC stands for 'Three-Cornered Hat/Triple Collocation'. 'Others' are approaches that are used only once, which are recorded in (Tran, 2023).

Except for the validation and intercomparison approach, other approaches showed no increasing or a decreasing proportion in selected literature from 2011 to 2021 (**Figure 2-4**). Approaches other than validation and intercomparison have only been used by a small group of researchers and not applied widely or increasingly. Thus, this section discusses the application of the most common uncertainty assessment approaches.

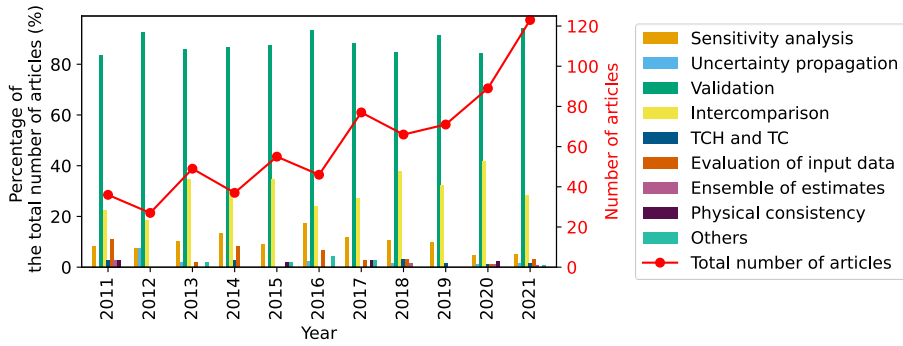


Figure 2-4. The proportion of reviewed articles per year for each approach to assessing satellite-based ET_a uncertainties.

2.3.1 Validation

In validation, satellite-based ET_a model results are compared to a 'reference' method that is considered by the researcher as the 'best' or most valid measure. The choice of the 'reference' method introduces subjectivity into the model evaluation (Melsen et al., 2019). In the case of satellite-based ET, three types of 'reference' are typically used: (1) in-situ measurements (N = 572), (2) catchment water balance (N = 83) and (3) output from models run with ground-based input data (N = 9). Almost all articles that used the validation approach considered an in-situ measurement as their reference (**Figure 2-5**), while other types of reference data were much less considered.

a. Using in-situ measurements as the validation reference

Several in-situ methods have been developed to estimate ET_a on the ground, including Eddy Covariance (EC), lysimeters, the Bowen ratio energy balance (BREB), etc. (**Table A-1**). These measurements are often considered the 'observation' or 'reference' to validate satellite-based ET_a. Among these, EC is the predominant method for validation and was considered in 424 out of 600 articles (**Figure 2-5**). Four factors explain the popularity of the EC method: (1) its relatively large network of stations, (2) long-term temporal coverage of flux towers, (3) open access of data (e.g., FLUXNET[5]) and (4) direct measurement of water vapor concentration and vertical wind speed of the air parcels to calculate latent heat flux.

[5] An international network of hundreds of EC sites (and researchers) mainly in the Global North., which aims to provides long-term observations of ecosystem through monitoring ET and carbon fluxes. See https://fluxnet.org/.

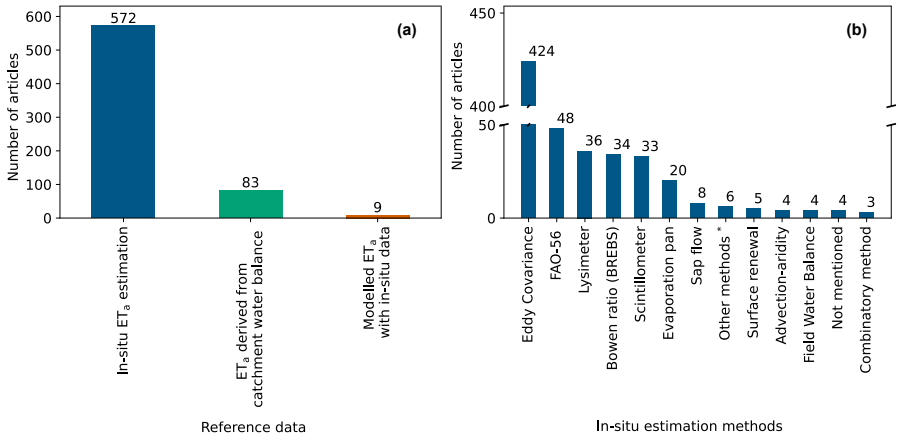


Figure 2-5. Different reference data (a) and in-situ estimation methods (b) used for satellite-based ET_a validation in reviewed articles ($N = 600$). *Other methods for in-situ ET_a estimation include volumetric soil water content difference ($N = 1$), canopy temperature and meteorology monitoring system ($N = 1$), portable chamber ($N = 1$), atmometer ($N = 1$), Open Top Chamber ($N = 1$), and crop coefficient method using ET_o equations other than FAO-56 ($N = 1$).

Using in-situ methods for validation faces three main challenges: (1) the cost to set up and maintain measuring stations; (2) the mismatch between the source area of measurement and the spatial resolution of an RS-based estimate, and (3) errors in measurements and assumptions. For example, the cost of a complete EC system is about ten times the cost of a weather station with basic meteorological instruments. Although the EC method can be used to monitor other fluxes (e.g., carbon dioxide and nitrogen oxide), the high cost of the EC system still limits the number of sampling points and regions (FLUXNET, 2017; Oliphant, 2012). The low sampling density can be compensated for with low-cost systems (Markwitz and Siebickeor, 2019) but at the expense of lower accuracy. In order to obtain validation data at a global scale, EC networks need to be expanded in many regions (e.g., Africa, South Asia, Southwest Asia, and South America).

Spatial support of in-situ measurements often does not overlap with the pixel footprint of the satellite images (i.e., the area the pixel value represents). The spatial support of in-situ measurements varies among methods, averaging from 1 m^2 (micro-lysimetry) to a few km^2 depending on wind speed and wind direction (eddy covariance and scintillometry). Certain methods for measuring components of ET_a have more limited spatial support (e.g., sap-flow measurement for transpiration). For homogeneous pixels (with the same geophysical and ecological characteristics), in-situ measurements can be representative of an entire pixel. However, when the pixel covers a large area, satellite-

based ET_a validation frequently involves heterogeneous pixels. Therefore, multiple sites and upscaling methods are required to best aggregate from site-specific to pixel-scale information (e.g., Li et al., 2018; Liu et al., 2016).

Every in-situ measurement technique is subject to uncertainty and error. Even the most widely used technique, the EC flux tower, has limitations in terms of measurement (10–20 % error) and spatial support (Glenn et al., 2011; J. Wang et al., 2015). All methods have common sources of error and uncertainty, such as sensor response (detection limit), calibration error (sensor drift over time), noise (spurious random spikes in the signal from the sensor), and poor installation and maintenance (Allen et al., 2011a). Additionally, each method has specific sources of error and uncertainty due to its theoretical assumptions. For example, the EC method requires fully developed turbulent fluxes to ensure that the net vertical transfer of water vapor is caused by eddies, and the area must be horizontal and uniform. Moreover, the lack of energy balance closure in EC measurements needs particular attention, since the gap can be up to 30 % of available energy (Allen et al., 2011b; Bambach et al., 2022; Vendrame et al., 2020; Wilson et al., 2002). The problem arises due to a scale mismatch of energy balance components and unaccounted exchange fluxes on heterogenous landscapes (Foken, 2008).

Dealing with scale mismatches and uncertainty of reference in-situ measurements is challenging, and there is no consistent method in the reviewed literature. Some studies only mentioned these issues when discussing the validation result. Information about the spatial support and uncertainty of in-situ measurements is not always available to researchers if they acquire reference data from other sources. However, without reporting the spatial support and uncertainty of measurements, we might easily draw biased conclusions: when the validation results are good, we might conclude that the model is good without questioning the quality of the reference, but when the results are not so good, we could argue that it is because of the imperfect reference measurements and conclude that the model still is good. Hence, it is important to accompany validation results with the best knowledge about the uncertainty and scale mismatches of reference datasets.

b. Using the residual of the water balance as the validation reference

ET_a of an area can be estimated as the residual of the water balance (WB) when the inflow (e.g., precipitation, irrigation supply), change in storage, and outflows of water (e.g., runoff, water conveyance) of that area are known. This approach is mainly used for assessment at a river basin scale. It assumes that the residual from the basin WB should be the total ET_a of the basin: $ET_a = P - Q - \frac{dS}{dt}$, where P is precipitation, Q is river discharge, and $\frac{dS}{dt}$ is the total change in basin water storage over time. This water balance approach assumes that there is no other water inflow or outflow across the catchment boundary. In some studies, $\frac{dS}{dt}$ is assumed to be negligible over a long period of time (a year or longer), which results in a more simplified water balance $ET_a = P - Q$.

For long-term periods (e.g., years, decades), total water storage change (TWSC) over time ($\frac{\mathrm{d}S}{\mathrm{d}t}$) is assumed to be zero, such that ET_a estimates are then validated with only $P - Q$ (e.g., Vinukollu et al., 2011). However, this assumption does not hold true in many regions of the world where groundwater is overexploited at an accelerated rate. For short-term periods (i.e., months), TWSC is often estimated from GRACE RS-based total water storage anomaly (TWSA) products. However, the TWSA products only cover the period from 2002 with a gap of 11 months from 2017 to 2018 between the GRACE and GRACE-FO missions. Some techniques have been developed to reconstruct this gap in the GRACE time series (e.g., Yang et al., 2021). However, the uncertainties in gap-filled $\frac{\mathrm{d}S}{\mathrm{d}t}$ estimates are still less known than uncertainties in the initial estimates from GRACE and GRACE-FO (Boergens et al., 2022).

The uncertainty in ET_a estimated by this approach depends on the choice and data quality of other variables (e.g., precipitation and river discharge) in the WB (Senay et al., 2011). Lehmann et al. (2022) have compared the residual calculated from 1694 combinations of P, Q, and ET_a datasets with $\frac{\mathrm{d}S}{\mathrm{d}t}$ derived from GRACE and found that none of these combinations can close the WB in all tested basins. They also suggested that using some combinations of P, Q, and ET_a datasets cancels their errors in the GRACE-based WB. Because of the errors in the P, Q, and $\frac{\mathrm{d}S}{\mathrm{d}t}$ components, studies that use WB-derived ET_a as a reference to validate ET_a without accounting for uncertainties in these components risk biased conclusions.

To account for errors in P, Q, and $\frac{\mathrm{d}S}{\mathrm{d}t}$, some researchers have tried to use multiple datasets (e.g., Weerasinghe et al., 2020). Recently, Schoups and Nasseri (2021) proposed treating uncertainties in datasets as unknown random variables. Instead of using the WB to determine these uncertainties, they estimated ET_a (and other water fluxes) by combining WB constraints and uncertainty estimation into a comprehensive probabilistic model. Although only applicable for river basins where GRACE resolution is suitable, this could be a good direction for future research on these water fluxes.

2.3.2 Intercomparison

Intercomparison is the second most widely used method (212 out of 676 studies). In intercomparison, the satellite-based ET_a estimates from multiple models are compared without assuming a superior one. This approach is mainly used to evaluate the relative uncertainty of a model compared to others (170 out of 212 studies). Intercomparison has also been used to evaluate other sources of uncertainty. For example, uncertainty from a change of spatial support can be evaluated by comparing model outputs using different input upscaling methods (e.g., Ershadi et al., 2013; Sharma et al., 2016). Intercomparison

has also been used to evaluate uncertainty due to the choice of input datasets (e.g., Badgley et al., 2015; Long et al., 2011; Wang et al., 2016).

Since the satellite-based ET_a datasets have both temporal and spatial dimensions, comparing satellite-based ET_a models or products is usually done by aggregating over one or two dimensions (i.e., resampling to a lower resolution). The simplest method of intercomparison involves aggregating ET_a estimates both temporally and spatially into one value (e.g., global annually averaged ET_a) and then comparing this value from different models or products (e.g., Mueller et al., 2013; Pan et al., 2020). Other methods of intercomparison involve comparing time series of spatially aggregated ET_a (e.g., monthly basin-scale ET_a). Aggregating over one of the two spatial dimensions is sometimes applied (e.g., Chen et al., 2019; Pan et al., 2020). The time series can also be aggregated by land cover classes (e.g., Weerasinghe et al., 2020) or climate zones (e.g., Trambauer et al., 2014), describing how satellite-based ET_a uncertainty varies in different conditions. For spatial intercomparison, temporally aggregated satellite-based ET_a maps can be compared visually (e.g., Weerasinghe et al., 2020) or by using simple map algebra (e.g., M. Jung et al., 2019). Only a few studies have applied metrics to evaluate the spatial similarity between two datasets, such as the Spatial Efficiency metric (SPAEF) (H. C. Jung et al., 2019; Stisen et al., 2021) and the degree correlation measure of spherical harmonic coefficients (López et al., 2017). None of these methods can characterize uncertainty in satellite-based ET_a fully, thus, combining them would provide a more comprehensive intercomparison.

2.3.3 Uncertainty propagation

Only 8 out of 676 articles applied the uncertainty propagation approach, mainly the Monte Carlo (MC) methods, to evaluate uncertainty in satellite-based ET_a. In MC methods, the model inputs are randomly sampled from their distributions and fed into the model to generate outputs repeatedly. The variance of the output distribution will then be considered as the uncertainty in the model output (i.e., ET_a estimate) associated with the input variables. The limited application of uncertainty propagation can be attributed to its complexity and computational demand. Sensitivity analysis and uncertainty propagation are ideally carried out in tandem (Crosetto et al., 2001; Saltelli et al., 2019), but only 5 out of 8 articles combined these approaches. The uncertainty propagation approach was also used for investigations beyond uncertainty quantification. For example, Talsma et al. (2018) used MC methods to determine the uncertainties in ET_a partitioning (i.e., soil evaporation, interception, and transpiration) in three satellite-based ET_a models (MOD16, PT-JPL, and GLEAM) due to the relative uncertainty in the key variables.

In the reviewed studies, uncertainty propagation was done only at one or a few fixed locations by assuming the probability distribution of the input variables, then simulating a range of ET_a values at these locations. This approach is computationally inexpensive

but does not fully characterize uncertainties in a spatial field of ET_a. To fully quantify uncertainty in a scene, Cawse-Nicholson et al. (2020) introduced a method based on MC methods and spatial-statistical models (Cressie, 1993). With this method, the probability distribution of ET_a per pixel in a satellite scene can be quantified and presented as percentile maps. This distribution was almost always non-Gaussian for all pixels in ET_a scenes, which means simple linear error propagation is not possible (Cawse-Nicholson et al., 2020). Future studies of satellite-based ET_a would benefit from the development of new methods to quantify uncertainty spatially.

2.3.4 Sensitivity analysis

Sensitivity analysis is the third most used approach in the reviewed literature but is only applied in a small proportion of the reviewed studies (61 out of 676 articles). Out of these, only 7 studies applied global sensitivity analysis. Sobol′s (2001) method was applied to the parameters of the MODIS16 algorithm (K. Zhang et al., 2019), the TSEB model (Burchard-Levine et al., 2020), and three satellite-based ET_a models (PT-DTsR, MODIS16 algorithm, and PML) (Cao et al., 2021). This method was also applied to input variables of satellite-based ET_a models alone (e.g., Gomis-Cebolla et al., 2019). Elhag (2016) applied a similar variance-based sensitivity measure for the SEBS model but did not refer to Sobol's method. The Extended Fourier Amplitude Sensitivity Test has also been applied for global sensitivity analysis (García et al., 2013). This limited number of studies shows the application of global sensitivity analysis onto satellite-based ET_a models has been under-researched during the last decade, despite the importance of global sensitivity analysis in environmental modelling (Saltelli et al., 2021).

Most articles that applied sensitivity analysis (54 out of 61) did not mention or apply a global sensitivity analysis method and thus, were considered local sensitivity analysis. In most of these studies, local sensitivity analysis was done by changing one parameter at a time (One-at-A-Time) and calculating the ratio of change in ET_a over change in parameter (e.g., Long et al., 2011). In the reviewed articles, the One-at-A-Time method has been implemented differently in terms of three factors: (1) the selection of parameters for local sensitivity analysis according to their importance judged by the researchers, (2) the range of values over which parameters are allowed to vary, and (3) the calculation of sensitivity for specific land covers. This suggests that local sensitivity analysis is influenced by the subjectivity of the researchers.

2.3.5 Evaluation of input data

The uncertainties of key input datasets are sometimes evaluated by researchers in studies that assess uncertainty in satellite-based ET_a without explicitly being propagated to model outputs. This approach ranked fourth in the number of articles with 20 out of 676. The key input datasets considered by researchers include air temperature, incoming shortwave

radiation, incoming longwave radiation, wind speed, and land surface temperature (e.g., Li et al., 2017; Pardo et al., 2014; Peng et al., 2016; Vinukollu et al., 2011a). Input datasets were evaluated through validation with their in-situ counterpart. Although other input datasets like Vegetation Indices are also important in satellite-based ET_a models, the in-situ measurements of these are often not available for evaluation (Vinukollu et al., 2011a). Some of the forcing datasets of satellite-based ET_a models are not remotely sensed data but are products from atmospheric data assimilation systems (e.g., Global Land Data Assimilation System (GLDAS) and ECMWF reanalysis (ERA)), which are sometimes provided with uncertainty estimates from data providers. Evaluating the input data provides crucial *a priori* information for propagating uncertainty to ET_a estimates. Furthermore, even if uncertainty propagation is not conducted, these assessments can help to identify sources of uncertainty in satellite-based ET_a; as the saying goes, "garbage in, garbage out".

2.3.6 Triple collocation and three-cornered hat method

The Three-Cornered Hat method (TCH) (Premoli and Tavella, 1993) and Triple Collocation (TC) (McColl et al., 2014; Stoffelen, 1998) are related to the intercomparison approach in the sense that these techniques assess the relative uncertainty of three datasets without assuming one is the best. Therefore, these techniques are useful when there is a lack of high-quality reference datasets. Both TC and TCH methods require a set of three datasets with the assumption that their errors are independent (Sjoberg et al., 2021). The difference between TCH and TC is that TC can only be used to assess uncertainties of uncorrelated datasets, while TCH can be used when there are correlations with proper constraints (Sjoberg et al., 2021; Xu et al., 2019). However, to date few studies have evaluated uncertainties in satellite-based ET_a using TC (Barraza Bernadas et al., 2018a; Khan et al., 2018; Kibria et al., 2021; Miralles et al., 2011a) and TCH (He et al., 2020; Long et al., 2014; Xu et al., 2019). The proportion of studies that used these methods is less than 2 % of the total reviewed articles and is not increasing (**Figure 2-4**). This low adoption might be attributed to the limitations of these methods: (1) the lack of information about biases and only estimation of random errors (e.g., RMSE, standard deviation, or variances), (2) the required conditions to achieve reliable error estimates (large samples, similar scales and magnitudes of errors between datasets) (Sjoberg et al., 2021), and (3) the reliability of TCH as an alternative to direct validation (Wu et al., 2019a).

2.3.7 Physical consistency

Physical consistency can be understood as the plausibility that an ET_a estimate is consistent with the physical conditions or characteristics of the area it represents. Consistency check or physical validation was proposed by Zeng et al. (2015) as the final

step in a general validation process for big remote sensing datasets. When there is limited reference data and ground-based measurements, physical validation is critical to assess the quality of data products (Blatchford et al., 2020b). Although physical validation does not quantify uncertainty using metrics, it provides an evaluation of the data quality. This is useful to identify the regions and conditions in which RS estimates are more uncertain and where more effort in direct validation approaches is required.

Only 6 studies in the selected literature have attempted to quantify this plausibility (**Figure 2-3**), but they defined physical consistency differently. For example, Rwasoka et al. (2011) used ET_o estimates as a threshold to decide whether ET_a estimates from the SEBS model were physically inconsistent. Blatchford et al. (2020) used the ET_a/P ratio and water availability ($P - Q$) to evaluate the physical consistency of the WaPOR ET_a product. López et al. (2017) developed a technique to assess the hydrological consistency of ET_a by transforming both ET_a and P data into spherical harmonics and then using spherical harmonic coefficients to calculate the degree correlation. These studies are not the same as validating satellite-based ET_a with $P - Q$ or $P - Q - \frac{dS}{dt}$ as discussed previously, since these residuals were not considered the best reference of ET_a.

Another method to assess physical plausibility without explicit water balance is through the Budyko curve. The Budyko curve describes the semi-empirical relationship between long-term ET_a and its limiting factors, i.e., precipitation and potential ET (ET_p), for river basins (Budyko, 1974). Koppa and Gebremichael (2017) validated the physical consistency of ET_a by calculating the RMSE of the Euclidean distance between the data points and the Budyko curve in ET_a/P and ET_p/P space. Weerasinghe et al. (2020) simply calculated the mean difference (bias) between satellite-based ET_a and Budyko-derived ET_a to evaluate which satellite-based ET_a product exceeds the energy and water limit defined by the Budyko curve. They also noticed that if a data point does not align with the Budyko curve, it might also mean that the ET_a of the basin exceeds the water or energy limit, for example, due to human activities. Therefore, the interpretation of physical plausibility needs to consider the actual knowledge about water resources in the basin, instead of focusing only on model generated numbers.

2.3.8 Using ensemble of estimates

Intercomparison studies sometimes lead to ensemble-mean products of all available products, based on the assumption that no model performs best, so an ensemble of them would be preferable (Bhattarai et al., 2019; Elnashar et al., 2021). This approach has been used the least in the reviewed articles (**Figure 2-4**). Some researchers have evaluated the uncertainty in an ensemble (a set) of satellite-based ET_a estimates from different models by calculating the average and range of all members in the ensemble (Elnashar et al., 2021; Guo et al., 2020; Vinukollu et al., 2011a). This approach is the same as the multi-model ensembles in climate modelling. The model structural uncertainty can only be quantified

if independent models are sampled from the entire possible model space and avoid the over-representation of one model structure (Abramowitz and Gupta, 2008). For example, Vinukollu et al. (2011a) selected three satellite-based ET_a models, namely SEBS (Su, 2002), PM-Mu or MODIS16 (Mu et al., 2007), PT-Fi or PT-JPL (Fisher et al., 2008), which are based on distinct equations used to estimate ET.

Using ensembles of satellite-based ET_a estimates provides uncertainties of the ensemble but not each individual member of the ensemble. Thus, some studies went further by merging the datasets of the ensemble and calculating the difference between this merged dataset with each ensemble member (Baik et al., 2018; Elnashar et al., 2021). If simply averaging all the ET_a products, the bias of different models can be cancelled in regions where they perform differently but accumulated in regions where they perform in the same manner. Hence, the ensemble products may arguably produce better estimation in some areas (Yao et al., 2017), but not a better understanding of the physical processes and drivers needed to improve satellite-based ET_a (Yao et al., 2017; K. Zhang et al., 2016). Therefore, it is considered more useful to use the range of the ensemble to identify the outlier data products or the uncertainty of all data products.

2.4 CONTEXT OF UNCERTAINTY ASSESSMENT

The context in which the uncertainty of satellite-based ET_a is assessed determines which method is selected and how it is applied. This context includes the objective of the satellite-based ET_a estimates, the spatial and temporal support at which ET_a is assessed, geographic location, and the availability of reference datasets. This section describes the context in which 676 reviewed articles assessed uncertainties in satellite-based estimation of ET_a.

2.4.1 Objectives of reviewed articles

The review shows that uncertainties in satellite-based ET_a estimates were assessed at all stages, from developing a new model to evaluating its data product. Uncertainty in satellite-based ET_a was assessed in the context of model implementation (34 % of reviewed articles), model development (13 % of all reviewed articles), model improvement (17 %), model evaluation (19 %), and product evaluation (16 %) (**Figure 2-6**). Here, model implementation means that a pre-existing model was applied to new case studies or to achieve some specific research objective without considerable modification or further development of the model.

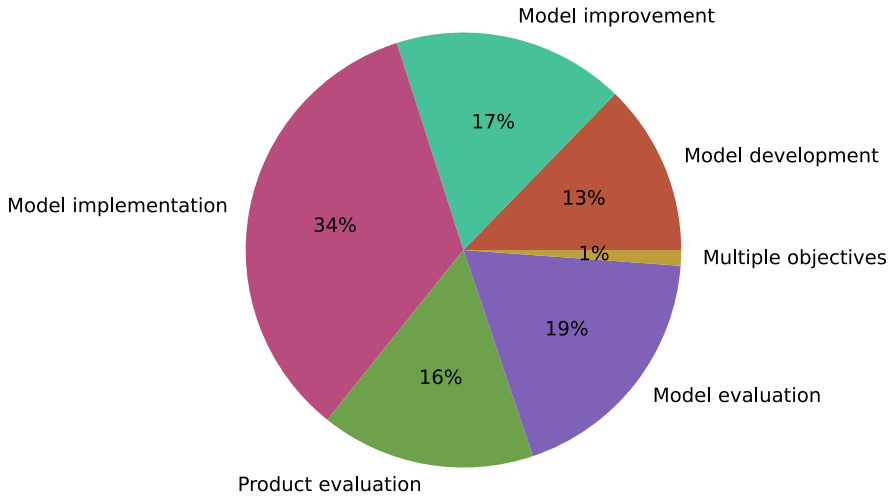


Figure 2-6. Research objective of the reviewed articles ($N = 676$).

The prominence of model implementation as the main objective in the reviewed articles could be due to a perceived need to assess the uncertainty of satellite-based ET_a estimates for each application despite previous validation. This is an important attitude in the research community since it helps to provide feedback on appropriate applications, and improvement of satellite-based ET_a models. Therefore, studies in the context of model implementation should not be overlooked.

2.4.2 Sources of uncertainty evaluated

The reviewed articles evaluated all sources of uncertainty as categorized in the theoretical framework (**Figure 1-3**), with a strong focus on compound uncertainty. **Figure 2-7** shows that the majority (406 out of 676) of reviewed articles assess only compound uncertainty without disaggregating into other sources. The second largest set of articles assessed both compound uncertainty and the relative uncertainty of satellite-based ET_a estimates. Other sources of uncertainty are remarkably less evaluated in the selected literature. According to the number of articles in each set (**Figure 2-7**), the level of interest in different sources of uncertainty can be ranked as follows: compound, relative, input data, model parameterization, change of spatial support, change of temporal support, and finally gap filling. This does not necessarily show the ranking of importance of the uncertainty sources, but rather the availability of methods and data needed to assess them.

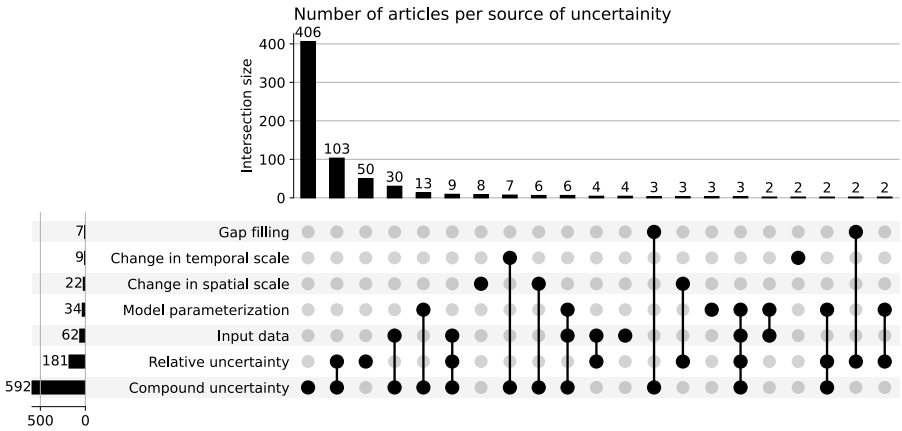


Figure 2-7. The source of uncertainty assessed in reviewed articles ($N = 676$). The horizontal bar chart displays the number of articles assessing specific sources of uncertainty (categories), while the vertical bar chart represents article counts within the intersections of multiple categories. Each vertical bar corresponds to an intersection in the column beneath it. Black circles denote the categories on the respective rows present in the intersection, while grey circles signify categories absent from the intersection. Intersections with less than 2 articles were excluded for improved presentation.

The uncertainties due to temporal upscaling are affected by several factors related to location (Jiang et al., 2021). These factors includes vegetation cover, soil moisture (Gentine et al., 2007; Hoedjes et al., 2008), cloud coverage (as discussed in research by Van Niel et al., 2012), cloud frequency (as explored in studies by Xu et al., 2015), air pollution effects (as indicated in research by Zhang et al., 2013), the return interval of the satellite (Alfieri et al., 2017), the satellite overpass time (Jiang et al., 2021), and the number of instantaneous values used for upscaling (Liu, 2021). Consequently, applying a single temporal upscaling method for the entire globe results in spatially varying uncertainties in satellite-based ET_a estimates.

2.4.3 Spatial and temporal support of assessments

Uncertainties in satellite-based ET_a estimates are specific for different spatial and temporal supports. The reviewed studies evaluated satellite-based ET_a uncertainties at spatial supports ranging from less than 100 m up to global, and temporal support ranging from sub-daily to annual (**Figure 2-8**). Most studies evaluated uncertainties in satellite-based ET_a estimates at spatial supports of 500 m to 5 km (268 out of 676) and less than 100 m (191 out of 676). This can be attributed to the availability of RS datasets that are widely used to estimate ET_a, such as MODIS (250 m to 1 km) and Landsat (30 m to

100 m). In the case of validation, the spatial support of uncertainty assessment was determined by the spatial support of the ground truth reference. For temporal support, uncertainty was mostly evaluated by daily ET_a (365 out of 676), although RS datasets provide observations at the satellite overpass time with a temporal resolution of 5-16 days. This shows that the temporal support of uncertainty assessment is driven more by practical needs and less by the availability of datasets.

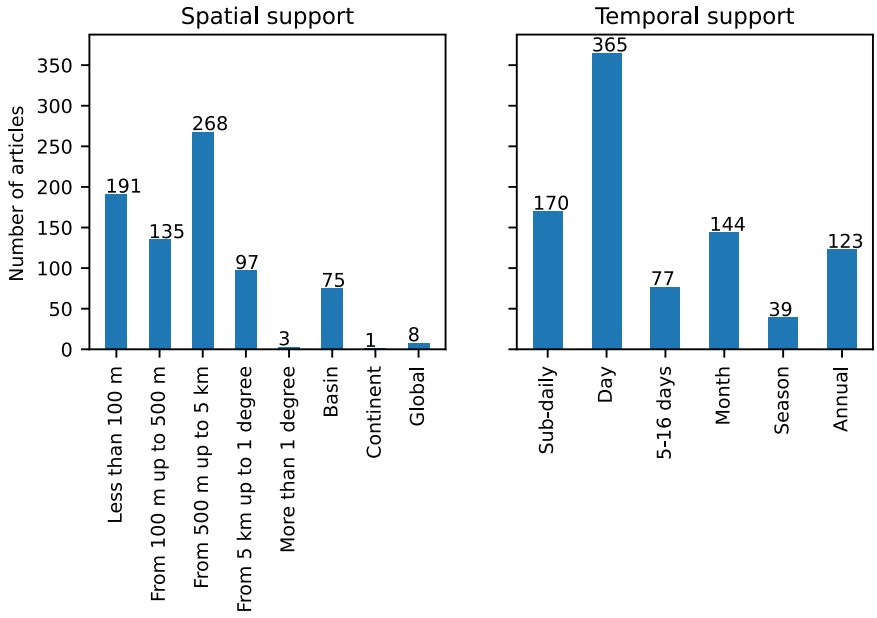


Figure 2-8. Number of articles per range of spatial and temporal support at which uncertainty in satellite-based ET_a was assessed (total number of articles $N = 676$).

2.4.4 Geographical distribution

Assessment of satellite-based ET_a uncertainties is not evenly distributed across the globe. The number of articles per country where uncertainties in satellite-based ET_a were assessed is shown in **Figure 2-9**. Each article was tagged by the country where the sites of study are located. The highest number of articles assessed ET_a in China. Because the most common approach is validation and the most common reference used is EC measurements, ET_a was mainly assessed where there are EC stations (i.e., AmeriFlux, AsiaFlux, ChinaFlux, OzFlux, EuroFlux, FLUXNET). Even when the studies aimed to validate satellite-based ET_a globally, the estimated uncertainty is not universal since these networks do not cover many regions. These studies were also included in **Figure 2-9**.

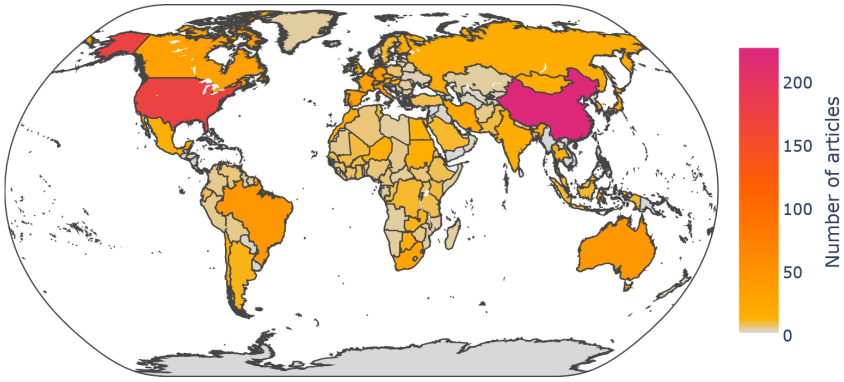


Figure 2-9. Number of articles per country where uncertainties in satellite-based ET_a were assessed, as identified using the study area or locations of in-situ reference sites in the articles' full-text.

Based on its popularity, EC can be considered the de facto standard ET_a estimation approach for validation of satellite-based ET. However, this popularity is mainly driven by the number of publications in countries where EC towers are more densely distributed (e.g., China and the United States of America). In countries where there are very few or no EC towers available, the most common reference used for validation of satellite-based ET_a is the water balance method (**Figure 2-10**). In a few countries in North Africa and Southwest Asia, the most common method is to use the FAO-56 method (Allen et al., 1998) in combination with crop coefficients to estimate ground-based references for validation (e.g., Egypt and Iran).

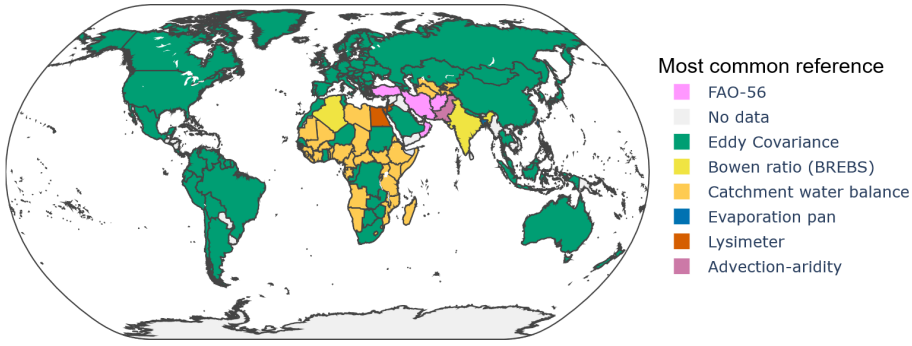


Figure 2-10. The most common reference used for validation of satellite-based ET_a per country, as identified in the articles' full-text.

2.5 CONCLUSIONS

This chapter identifies and appraises methods for uncertainty assessment of satellite-based ET_a estimates by applying a systematic quantitative literature review approach. The majority of reviewed articles assess uncertainty in satellite-based ET_a estimates by validation against EC measurements. In regions where in-situ measurements are limited, most studies used the residual in the water balance as a reference for validation. It is important to make use of existing EC networks for global validation of satellite-based ET_a estimates. However, there is still a gap in the availability of in-situ data for global validation, as most are concentrated in North America, East Asia, and Europe. Moreover, the challenges in energy balance closure and scale mismatch persist through the reviewed studies.

Comparing performance of satellite-based ET_a models and investigating the sources and geographical distribution of uncertainty in their ET_a estimates remains important for many applications. Global assessments provide a broad perspective on satellite-based ET_a uncertainties by considering factors that affect data quality on a large scale, such as satellite sensor characteristics, model characteristics, geographical and climatic factors. Local assessments, on the other hand, focus on specific study areas, which may have unique conditions and sources of uncertainty that are overlooked in global assessments. Therefore, future research should combine local and global evaluation efforts.

For validation of satellite-based ET_a estimates with in-situ methods, we provide specific recommendations:

- The uncertainty of the reference datasets, including correction for surface energy balance closure, should be evaluated and reported.
- Satellite-based ET_a estimates should be converted to values at the temporal and spatial scale of reference datasets.
- The common metrics (RMSE, bias/mean error, correlation coefficient, coefficient of determination), their formulations, and mean ET_a or scale-independent metrics should be reported in validation studies.
- The statistical significance of validation metrics should be tested, and the number of data points used should be reported.
- In addition, uncertainties in satellite-based estimates should be characterized using multiple metrics that are scale-independent to facilitate comparison of studies across regions with different ET_a ranges.

- Validation of satellite-based ET_a models and data products should be reported at different levels of spatial and temporal scales, covering multiple locations.

This chapter highlights the importance of applying multiple and diverse approaches for assessing uncertainty in spatiotemporal satellite-based ET_a data, particularly when validation datasets are limited. Key methods include intercomparison, sensitivity analysis, uncertainty propagation, physical consistency check, input data evaluation, triple collocation, and the ensemble of estimates. Among these, sensitivity analysis and uncertainty propagation are especially valuable for advancing satellite-based ET_a techniques, as they help identify and quantify sources of uncertainty. However, this chapter highlights a notable gap in the literature: few studies have applied these techniques, with most relying on less computationally intensive options. This limits the ability to capture the detailed spatiotemporal distribution of satellite-based ET_a uncertainty.

Since uncertainty is an inherent characteristic of any satellite-based data products, a challenge remains in effectively characterizing it across space and time. This involves not only estimating overall errors but also pinpointing where and when uncertainty is likely to be highest. Several studies have aimed to offer spatially explicit uncertainty in thematic classification, such as land cover and soil type. These studies, like the ones mentioned by Woodcock (2002), have primarily focused on qualitative mapping techniques. However, for quantitative remote sensing, which involves mapping continuous variables like ET_a, there is a need for methods that can effectively characterize spatially explicit uncertainty.

3

STATUS OF REPORTED UNCERTAINTY

This chapter is partly based on:

Tran, B.N., Van Der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G. and Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. Hydrology and Earth System Sciences, 27(24), pp.4505-4528. https://doi.org/10.5194/hess-27-4505-2023

and

Tran, B. N., Hakzi, K., and Mul, M.: A meta-analysis of percentage bias in remote-sensing-based evapotranspiration estimates (in preparation).

3.1 INTRODUCTION

Despite the extensive body of literature assessing uncertainty of satellite-based estimation of ET_a (Chapter 2), only a few studies attempted to synthesize their results. For example, a review by Karimi and Bastiaanssen (2015) employed statistical methods to synthesize the results of 32 studies that validated satellite-based ET_a estimates. This synthesis included estimating the probability density function of mean absolute percentage error (MAPE) from 46 estimations of seasonal ET_a. Kalma et al. (2008) summarized the relative error and RMSE of satellite-based ET_a estimates reported in 30 studies. These syntheses are limited in number of studies considered ($N < 50$) and the selection of studies were not systematic. Another limitation of synthesizing these results is that the selected studies used different validation data and field instruments, which do not have equivalent spatial support and accuracy. Therefore, this chapter aims to synthesize and analyse the results of the reviewed articles in Chapter 2, in order to provide an updated status of uncertainty in satellite-based ET_a estimates in terms of accuracy and precision. The outcomes of this analysis would be a useful reference for future studies to evaluate the results of satellite-based ET_a uncertainty assessment. The main research questions of this chapter are:

- Which metrics are used to report uncertainty in satellite-based ET_a estimates?
- What is the typical range of uncertainty in satellite-based ET_a estimates globally based on performance metrics reported from previous studies?

3.2 ON THE USE OF METRICS TO REPORT UNCERTAINTY

The 676 articles included in the systematic review in Chapter 2 mainly reported uncertainty (RMSE), accuracy (bias or mean error), and the goodness-of-fit with a reference dataset (R^2) (**Figure 3-1**). Although quantifiable uncertainty in measurement is theoretically represented as a probability distribution, this has rarely been done in the literature. The reviewed studies used a wide range of metrics to report their uncertainty assessment (33 metrics). Most studies used three metrics, while some used up to twelve. Larger number of metrics provide more description of uncertainty, but some metrics might be challenging to interpret.

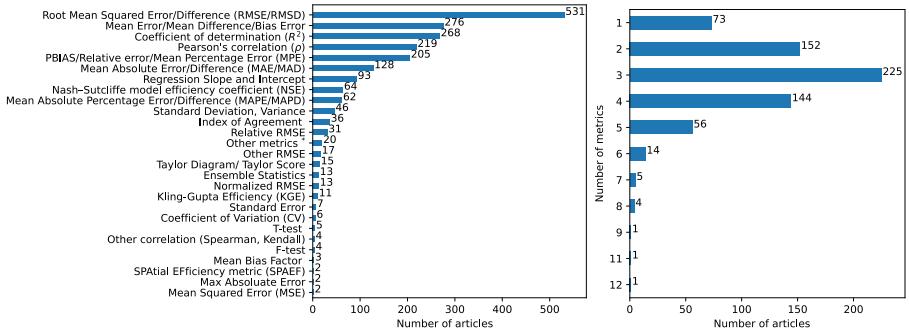


Figure 3-1. Number of studies per choice of metric to report uncertainty and the number of metrics used.

Root Mean Square Error (RMSE) is the most widely used metric in the reviewed articles (531 out of 676 articles). Metrics related to RMSE include normalized RMSE (normalized by standard deviation) and relative RMSE (as a percentage of mean ET_a). Very few studies (17 articles) used modified RMSE to report more robust results and few consider random error and systematic error, such as robust RMSE (Bisquert et al., 2016), systematic and unsystematic RMSE (Yebra et al., 2013), biased and unbiased RMSE (Martens et al., 2017).

Other commonly used metrics are mean error (ME), coefficient of determination (R^2), Pearson's correlation (ρ), percentage bias (PBIAS), which are used in more than 200 articles out of 676 articles (**Figure 3-1**). These metrics complement RMSE by diagnosing different facets of uncertainty (Section 1.3). ME and PBIAS reveal systematic over- or under-estimation (accuracy), R^2 quantifies the proportion of variance in reference data explained by satellite-based estimates, and ρ captures the strength of the linear relationship between satellite-based estimates and reference. However, many studies only reported RMSE, without other metrics or with only one or two (most studies used less than three metrics).

We also observed inconsistent use of metrics such as R^2, which might cause misinterpretation of results, especially when comparing studies. For example, the second most used evaluation metric was referred to using many names including mean error, mean difference, bias error, or bias. Meanwhile, the coefficient of determination (R^2) has the opposite issue, in which the same term was used with different formulas. R^2 is a measure of goodness-of-fit for regression models. There are at least 8 formulas for R^2 in the literature (Kvålseth, 1985), but only one formula can be used for any type of model fitting (i.e., R_1^2 in Kvålseth, 1985). Since many studies did not report which formula they used, we did not distinguish between different R^2 formulas in **Figure 3-1**. Nevertheless,

we observed that at least four different formulas of R^2 were used in the reviewed articles including the squared coefficient of correlation (**Table B-1**).

3.3 META-ANALYSIS METHOD

To quantify the overall range of uncertainty in satellite-based ET_a estimates reported in the literature, we conducted a meta-analysis of reported performance metrics. Meta-analysis is a quantitative method for synthesizing results from multiple studies to derive more robust conclusions than a single study alone (Gurevitch et al., 2018). Meta-analysis offers a way to accumulate knowledge and interpret divergent results from primary studies in a meaningful, overarching way (Evaristo and McDonnell, 2017). The following sections outline the research steps in this study.

3.3.1 Study selection

For a meta-analysis of uncertainty in satellite-based ET_a estimates, we included only studies that applied the same assessment method and reported the same metric. We selected the most commonly used uncertainty assessment method in order to include the largest number of data points. From the systematic quantitative literature review (Chapter 2), the majority of studies used EC flux towers as reference for in-situ validation. We selected these studies for meta-analysis of reported satellite-based ET_a uncertainty (N = 372) (**Figure 3-2**). Next, we screened the identified studies based on the performance metrics mentioned. We selected articles that used RMSE, the most commonly used metric. From 372 articles, 348 articles that reported RMSE of satellite-based ET_a estimates from the validation with EC flux tower were included. RMSE cannot always be interpreted as the same effect size across studies with different underlying magnitudes of ET_a (e.g., humid vs. arid climates). To complement results of RMSE and analyse systematic bias in satellite-based ET_a estimates, we also included studies that reported PBIAS (including ones that reported ME together with mean ET_a to derive PBIAS). Bias caused by the model structure or systematic errors in input data is a critical issue in ET_a estimates (Chen and Liu, 2020). We evaluated the results section of each study to extract the reported performance metrics from in-situ validation against EC measurements (Section 3.3.2).

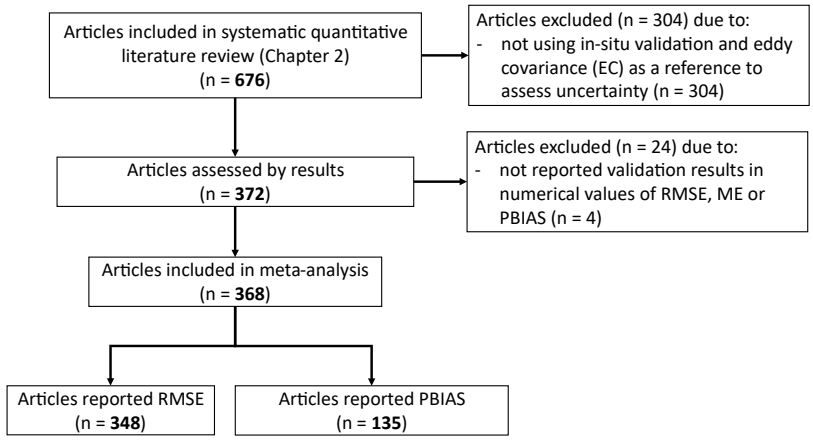


Figure 3-2. Number of articles included in meta-analysis after further screening systematic quantitative literature review dataset in Chapter 2.

3.3.2 Data extraction

For the studies identified for the meta-analysis, the reported performance metrics of in-situ validation (i.e., effect sizes) are recorded for each EC site or a group of EC sites in the studies. Heterogeneity in effect sizes can be caused by many factors, including model performance, input data, EC site characteristics, and reference quality. Therefore, the study characteristics that could explain the heterogeneity (i.e., moderators) are also investigated. The process of extracting effect sizes and moderators involved reviewing the full text of each article and identifying tables, figures, or key sentences that reported mean error, bias, or PBIAS of satellite-based ET_a estimates when validated against EC data.

Effect sizes

In meta-analysis, effect size is a standardized measure of the outcome that can be compared across studies (Nakagawa et al., 2023). For satellite-based model's performance or accuracy, effect size could be the goodness-of-fit statistics of model estimates against in-situ data, such as coefficient of determination (R^2) (Bahrami et al., 2022), Pearson correlation coefficient (ρ) (Hinge et al., 2021) and residual standard error (Zolkos et al., 2013). The effect size chosen must be comparable or standardized across studies, meaning that all included studies should be reporting essentially the same quantity or one that can be converted to a common metric.

RMSE has the unit of the estimates, so it can be expressed in $mm \cdot d^{-1}$ for ET_a or $W \cdot m^{-2}$ for latent heat flux. RMSE is scale-dependent, which means RMSE values are not inherently standardized. Therefore, to quantify the overall RMSE from across different

studies, unit conversion is needed. Recorded RMSE values were thus converted to a common unit and temporal scale (for example, daily or monthly ET_a in $mm \cdot d^{-1}$) to facilitate pooling of reported values into expected error magnitude. RMSE values in units other than $mm \cdot d^{-1}$ were converted to $mm \cdot d^{-1}$ assuming constant rate of ET_a over the temporal support. For example, 365 $mm \cdot year^{-1}$ was converted to 1 $mm \cdot d^{-1}$ and 0.1 $mm \cdot h^{-1}$ was converted to 2.4 $mm \cdot d^{-1}$.

We also used PBIAS as the effect size for cross-model and cross-site comparison. We further selected 135 studies that reported PBIAS or provided sufficient data to derive PBIAS (i.e., mean error and mean ET_a).

Moderators

A moderator in a meta-analysis is any study-level characteristic that is examined to explain why effect sizes differ across studies (Li et al., 2020). The examined model-specific characteristics include types of model and input data. Based on the model classification schemes commonly used in various literature reviews of satellite-based ET_a estimation methods (**Figure 2-1**), we grouped reported PBIAS based on types of models:

- "Factor": models that computes a reduction factor to scale reference/potential ET_a or available energy ($R_n - G$),
- "Empirical": statistical or machine-learning methods that derive ET_a from direct relationship with RS observables,
- "Ts-VI": models that use the triangular/trapezoidal relationship between ET_a, surface temperature and vegetation indices,
- "1SEB" and "2SEB": models that calculates sensible heat flux and then derive latent heat flux as a residual from either one-source ("1SEB") or two-source ("2SEB") surface energy balance,
- "PM" and "PT": models that relies on combined formulas such as Penman-Monteith ("PM") or Priestley-Taylor ("PT").

Models that are not identified as one of the mentioned types considered are grouped into the "Others" type since they have a limited sample size (i.e., only one study).

It is important to note that EC flux measurements do not necessarily provide the true ET_a values but simply represent ground-based conditions. One challenge of EC measurements is that they often fail to close the surface energy balance (1.1), meaning that LE measured by EC is not equal to $R_n - G - H$ measured at the same EC tower. Typically, only about 70-90 % of the available energy is captured by sensible and latent heat flux, resulting in a 10-30 % gap (Wilson et al., 2002). Several advancements in instruments and data post-processing have improved the energy imbalance of EC techniques, however, there is still at least 15 % gap is irreducible imbalance (Mauder et al., 2024). Therefore, when

validating satellite-based estimates, the energy balance gap is an important factor that could affect the resulting accuracy.

3.3.3 Data analysis

Summary statistics of RMSE and PBIAS in satellite-based ET_a estimates were calculated based on the data extracted from selected literature. To mitigate the disproportionate influence of outliers on statistical analyses, we removed data points that fell outside of the conventional ± 2 standard deviations from the mean of PBIAS distribution (less than 1 % of the original data). This approach, which captures roughly the 95 % confidence interval, was applied prior to the analysis (Helsel et al., 2020). The Mann-Kendall test (Kendall, 1955; Mann, 1945) was used to detect whether there is a yearly monotonic trend in effect sizes. In order to choose an appropriate statistical test for PBIAS cross-model and cross-site studies, we first determine if the PBIAS distribution of each model type is normally distributed using the Shapiro-Wilk test (Shapiro and Wilk, 1965).

When the PBIAS results for the different moderator types (i.e., model type) are not normally distributed or have imbalanced sample size, we applied a two-stage non-parametric hypothesis test for significant differences between groups (Helsel et al., 2020; Strobl et al., 2020). First, the Kruskal-Wallis H-test (Kruskal and Wallis, 1952) was applied to assess the null hypothesis (significant level $\alpha = 0.05$) that the distributions of PBIAS are identical across moderator types. Kruskal-Wallis tests the equality of median ranks and ignores distribution shape, which is robust for skewed distributions and unequal variances. When the Kruskal-Wallis test was significant, we performed post-hoc pairwise comparisons of mean rank sums using Dunn's test with Holm's α-adjustment to determine which pair of model types differed significantly (Dunn, 1964). When there was a significant difference between a pair, descriptive statistics were compared to identify which model type had a lower mean and standard deviation of PBIAS across included studies.

3.4 GLOBAL UNCERTAINTY FROM REPORTED RMSE

The reported RMSE values for daily ET_a ($N = 3{,}167$) range from 0.01 to 6.65 $mm \cdot d^{-1}$ with the mean value of 1.18 $mm \cdot d^{-1}$ (**Table 3-1**), which is comparable with RMSE previously reported by Kalma et al. (2008). When converting RMSE values from the reported unit to a common unit of $mm \cdot d^{-1}$, the mean RMSE is the highest for validation of instantaneous ET_a estimates (2.81 $mm \cdot d^{-1}$) and the lowest for monthly (0.78 $mm \cdot d^{-1}$). In general, studies with larger temporal support for validation have lower mean RMSE in $mm \cdot d^{-1}$. For the validation at temporal support of 3-hour, 10-day, and week, definitive conclusions are hindered by the small number of studies and records. The decrease of RMSE with increasing temporal support is due to the averaging and corrective effect of temporal

upscaling. Therefore, improving temporal upscaling and gap-filling methods are crucial for reducing uncertainty in satellite-based estimation of ET_a.

Table 3-1. Descriptive statistics of reported RMSE values (in $mm \cdot d^{-1}$) in reviewed articles ($N = 348$) with validation of satellite-based ET_a estimates with EC flux towers. STD is standard deviation, pct is percentile.

Temporal support	Number of records (N)	median	mean	STD	min	25th pct	75th pct	max
instantaneous	703	0.93	2.81	1.46	0.20	1.67	3.61	8.63
30-min	130	1.59	1.66	0.80	0.16	0.10	2.16	4.13
hour	135	0.62	1.03	1.03	0.16	0.40	1.29	6.00
3-hour	18	2.48	2.58	0.80	1.44	1.92	3.11	4.54
day	3,167	0.93	1.18	0.82	0.01	0.70	1.30	6.65
week	237	0.76	0.82	0.35	0.02	0.58	1.00	2.61
8-day	528	0.75	0.87	0.47	0.02	0.56	1.10	3.40
10-day	22	0.93	1.04	0.43	0.4	0.80	1.18	2.20
16-day	53	0.49	0.50	0.14	0.22	0.40	0.62	0.89
month	499	0.62	0.82	0.35	0.02	0.58	1.00	2.6
year	71	0.83	0.80	0.31	0.15	0.61	1.02	1.49
overall	5,563	0.95	1.31	1.06	0.01	0.67	1.49	8.63

Figure 3-3 shows that very high RMSE values were mainly from validation results that used a single EC site. Validation using data from a greater number of EC sites tends to yield lower RMSE values. This might be attributed to the fact that when papers only report average RMSE values across multiple EC sites, the average RMSE is lower than the individual site with highest RMSE . Moreover, the random errors at each site are likely to be uncorrelated or partially cancel each other out when averaged, which further reduces the overall RMSE. As RMSE is inherently dependent on the scale of ET, sites with lower ET_a values or the practice of averaging ET_a across multiple sites are more likely to exhibit lower average RMSE values. Unfortunately, only a limited number of studies provided information on relative RMSE or the average ET_a corresponding to RMSE values, which hindered the derivation of scale-independent RMSE values across all studies.

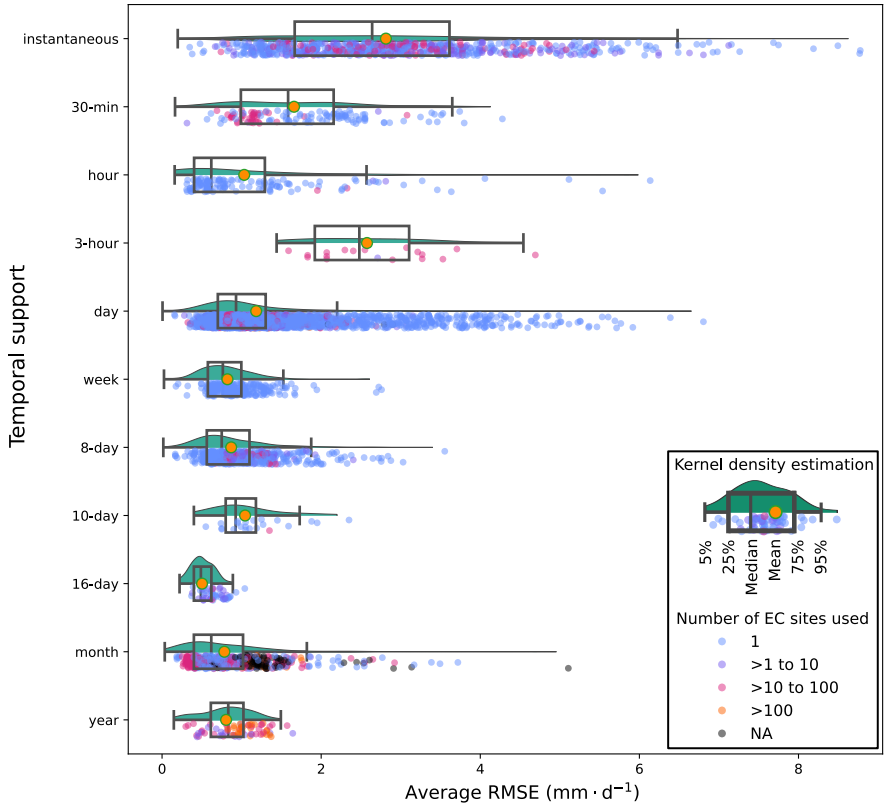


Figure 3-3. RMSE (mm·d^{-1}) of satellite-based ET_a estimates from the validation with Eddy Covariance (EC) observations in reviewed articles ($N = 348$). The scattered dots represent RMSE values reported in articles. The colour of the dots shows the number of EC sites used in validation. The green area under the curve represents the kernel density estimation of the underlying probability distribution. The box-and-whiskers plot represents 5th, 25th, 50th (median), 75th, and 95th percentiles of the distribution. The orange circle inside the box-and-whisker plot represents mean value.

The large range of RMSE obtained from the meta-analysis can be explained by the diversity of reviewed studies in terms of models, resampling methods, and validation context (e.g., temporal scales, land cover, climate, amount of data). For example, some studies validate satellite-based estimates of ET_a from global products, while others validate estimates from models that were calibrated to reduce RMSE. Moreover, many studies reported RMSE of latent heat flux (in W·m^{-2} or MJ·m^{-2}·d^{-1}) averaged from estimates at the satellite overpass time. The reported accuracy varies at different times of the day due to weather conditions, and is, thus, not representative of the entire day. We

converted these values to mm·d^{-1} (**Table 3-1**) only for comparison between different temporal supports. The range of RMSE presented in **Figure 3-3** and **Table 3-1** should only be considered as a baseline for typical error in satellite-based estimation of ET_a. Using only RMSE to compare ET_a model performance across different studies or validation sites is not recommended. The following section analyses PBIAS, which is a scale-independent performance metric commonly reported in studies.

3.5 Accuracy from reported PBIAS

3.5.1 PBIAS distribution

The reported PBIAS values for ET_a estimates (N = 2,063) from 135 studies range from -164 to 230 %, excluding the outliers beyond two standard deviations (**Figure 3-4**). The mean PBIAS of all records is 6.8 %, and standard deviation is 33 %. The probability distribution of PBIAS in ET_a estimates is unlikely to be normal and it is skewed in the positive range of PBIAS (**Figure 3-4**b). This suggests a slight overestimation of ET_a over the included studies. Karimi and Bastiaanssen (2015) analysed the mean absolute percentage error (MAPE) of satellite-based ET_a estimates and also found positively skewed distribution, with a mean MAPE of 5.4 % and a standard deviation of 5 %.

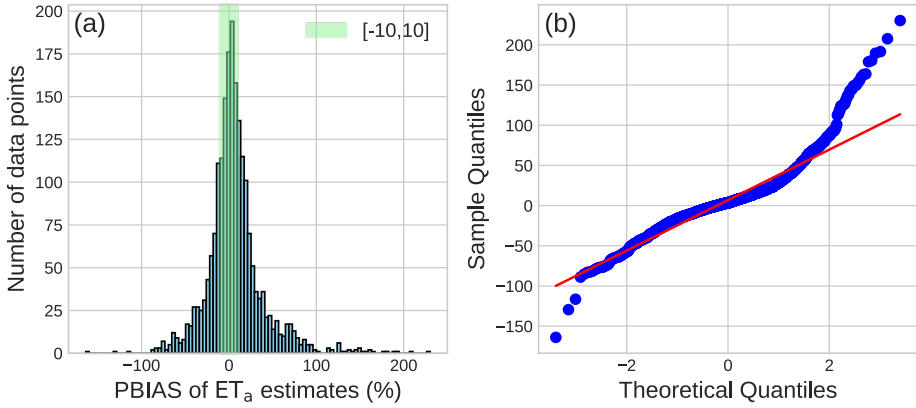


Figure 3-4. Distribution of PBIAS in satellite-based ET_a estimates from 135 studies that validated against EC flux tower data: (a) histogram and (b) Q-Q plot

The range of PBIAS is much wider than the range of mean absolute percentage error (MAPE) in a study by Karimi and Bastiaanssen (2015) (up to 20 %). This is likely due to the difference in validation period and number of studies. Karimi and Bastiaanssen (2015) only considered validation results from 31 studies (N = 46) where at least a season (6 months) or longer period is considered. In this study, we considered a larger number of studies (135) with any validation periods (**Figure 3-5**). Overall, a longer validation period

results in narrower distribution of PBIAS, with smaller standard deviation and less outliers. However, there was no significant difference between medians of different validation periods (Kruskal-Wallis p-value = 0.52).

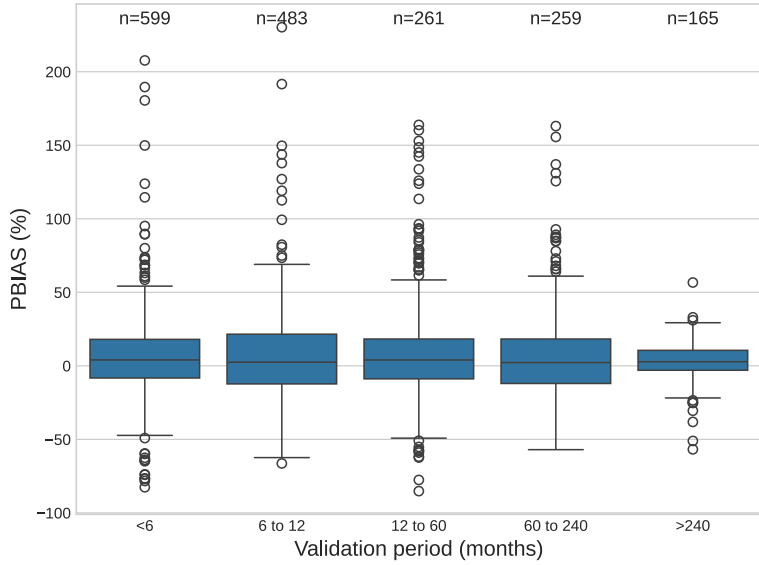


Figure 3-5. PBIAS (%) in satellite-based ET_a estimates categorized by the length of validation period.

The distribution of PBIAS over the years of study publication is shown in **Figure 3-6**. There is no statistically significant trend in both the mean and median PBIAS over the study period from 2011 to 2021, as indicated by the Mann-Kendall test (p = 0.76 for mean PBIAS and p = 0.27 for median). This suggests that the reported accuracy of satellite-based ET_a estimates has remained relatively stable over the decade. The number of validation records per year increased substantially from only 14 in 2011 to 573 in 2021. This reflects both a growing number of publications (as shown in Chapter 2) and validation efforts in the academic literature on satellite-based estimation of ET_a. Additionally, the 2021 publications show a narrower PBIAS range, due to the larger number of data points compared to earlier years, such as 2012 and 2018, which had similar mean PBIAS values but with larger standard error due to limited data size.

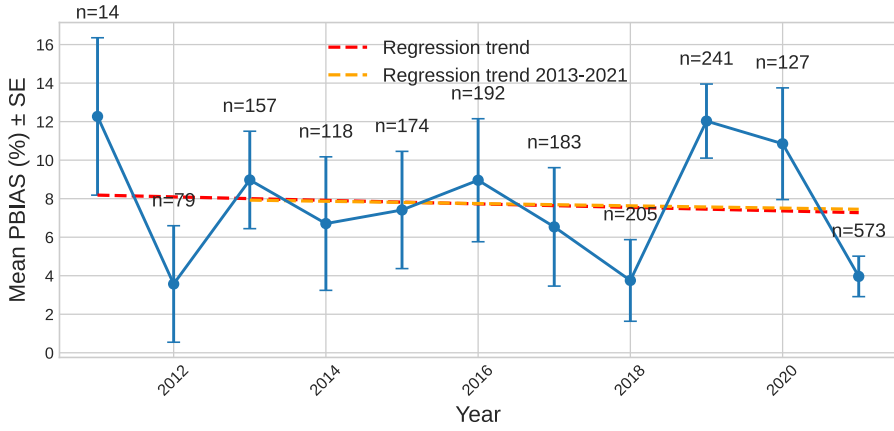


Figure 3-6. Mean and standard deviation of PBIAS in satellite-based ET_a estimates by the year of publication.

3.5.2 PBIAS by model type

The distribution of PBIAS demonstrates the overall expected performance of each model type across included studies. The mean (± standard deviation) of PBIAS ranges from 0.1 ± 32.3 % for PM type to 16.8 ± 65.2 % for Ts-VI type (**Table 3-2**). Empirical model type demonstrates slightly higher average but lower variation of PBIAS than PM model type (1.0 ± 12.5 %). Empirical models, which range from simple regression to machine learning algorithms, estimate ET_a by calibrating the relationship between in-situ ET_a data with satellite-based input data. Therefore, their performance is likely more precise than other process-based models (**Table 3-2**). The mean PBIAS is generally lower for complex model types (which require more inputs), such as PM (0.1 ± 32.3 %) and 2SEB (7.5 ± 23.2 %), than for simpler models like Ts-VI (16.8 ± 65.2 %), PT (11.3 ± 34.9 %), and 1SEB (11.7 ± 39.3 %).

Table 3-2. Descriptive statistics of reported PBIAS values (in %) by model type in reviewed articles ($N = 135$) with validation of satellite-based ET_a estimates with EC flux towers. STD is standard deviation, pct is percentile.

Model Type	count	mean	STD	min	25th pct	median	75th pct	max
1SEB	556	11.7	39.3	-89.0	-9.0	7.3	24.6	230.3
2SEB	407	7.5	23.2	-61.8	-7.0	3.3	18.8	127.0
Empirical	113	1.0	12.5	-62.4	-1.5	1.1	6.0	33.0
Factor	106	9.3	29.2	-64.3	-6.1	3.8	15.7	137.0
PM	573	0.1	32.3	-164.0	-15.8	0.5	13.8	191.6
PT	250	11.3	34.9	-85.2	-8.7	8.6	28.9	163.9
Ts-VI	30	16.8	65.2	-82.6	-6.3	5.7	25.0	189.6
Others	28	6.8	14.4	-30.0	2.2	7.5	15.2	38.6

The probability distribution of PBIAS varies moderately by model type (**Figure 3-7**a). Studies using PM, 1SEB and 2SEB types of models accounted for the largest number of studies, suggesting their popularity. Dunn's post-hoc test indicated that median PBIAS were significantly different between PM and three other models: 1SEB, 2SEB, and PT (**Figure 3-7**b). For other pairs, no significant difference was detected, which could also be due to small sample sizes of some model types such as Ts-VI and Others. It could be inferred that mean PBIAS in PM model types (0.1 ± 32.3 %) is generally lower than 1SEB, 2SEB, and PT model types. This confirms the observation by Chen and Liu (2020) that two-leaf conductance-based models—primarily of the PM type—apparently exhibit the lowest bias when compared to flux measurements.

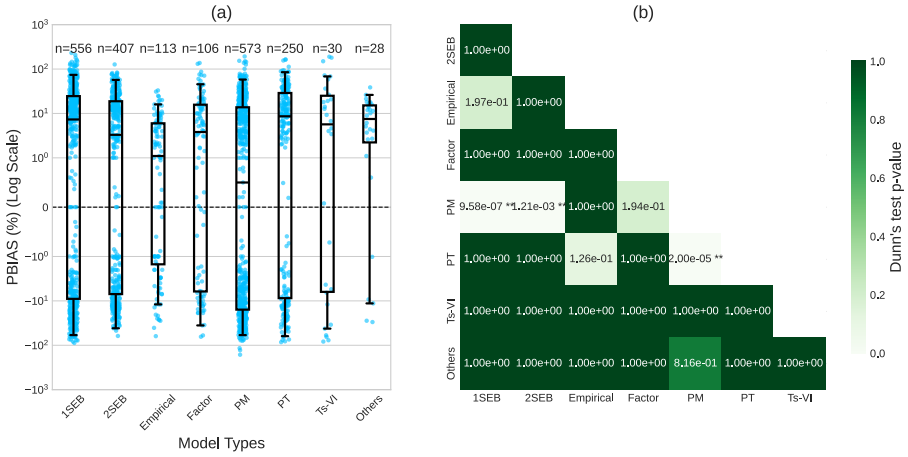


Figure 3-7. PBIAS (%) in satellite-based ET_a estimates categorized by model type. (a) The box-and-whisker plot represents the 5th, 25th, 50th (median), 75th, and 95th percentiles of the distribution. (b) Dunn's post-hoc test results at a significant level of $\alpha = 0.05$ (significant *p*-values marked with **). Note: y-axis is displayed on a log scale. **Figure B-1** shows the same figure on linear scale.

3.6 Impact of correcting energy balance in Eddy Covariance estimation

Multiple strategies are developed to address the EC energy imbalance. For validation of satellite-based ET_a estimates using post-processed EC data, the most common approach is to force closure, adjusting H and LE proportionally so that their sum equals available energy while preserving the Bowen ratio ($\beta = \frac{H}{\mathrm{LE}}$). The Bowen ratio (BR)-based correction assumes the imbalance affects both H and LE similarly (Twine et al., 2000). Studies using airborne and lysimeter comparison show that this assumption is reasonable in many cases (Mauder et al., 2020). Alternatively, some studies assign the entire energy residual (ER) of the balance to either H or LE, which improved the agreement of EC with other methods at specific sites (Mauder et al., 2020).

Either using BR, ER-based, or other partitioning schemes, the choice of correction has significant effects on validation results, leading to up to 50 % uncertainty in daily ET_a or 30 % of cumulative crop water use in a growing season (Bambach et al., 2022). To evaluate the impact of surface energy balance correction on the PBIAS of satellite-based ET_a estimates across studies, we further selected 10 (out of 135) studies that explicitly described the treatment methods and applied more than one energy balance closure treatment (**Table B-2**).

There is a significant difference in median PBIAS between unclosed and closed treatments (Kruskal-Wallis p-value < 0.01). **Figure 3-8** shows that correction methods such as ER and BR significantly reduced PBIAS in ET_a estimates. These findings suggest that correcting EC flux measurements for energy balance closure could potentially improve the resulting accuracy of ET_a estimates. Since most studies did not explicitly state how energy balance closure was addressed, it is unclear whether the reported PBIAS values are overestimated.

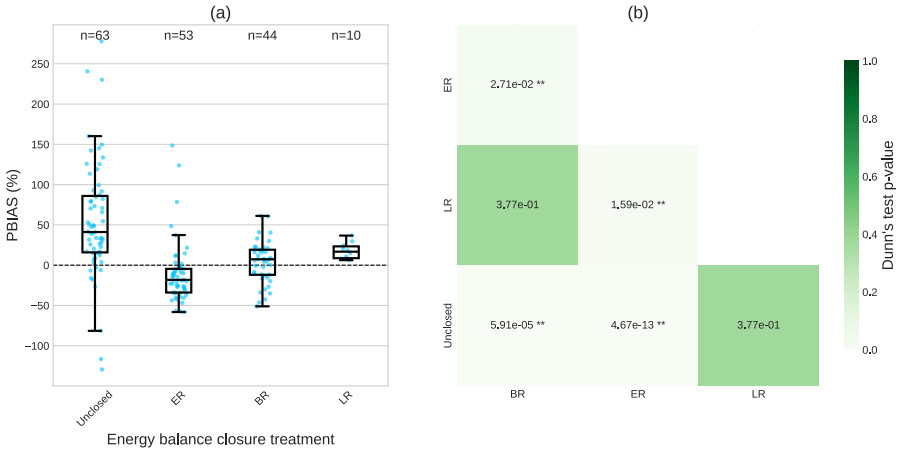


Figure 3-8. PBIAS (%) in satellite-based ET_a estimates categorized by treatment of energy balance closure. (a) The box-and-whisker plot represents the 5th, 25th, 50th (median), 75th, and 95th percentiles of the distribution. (b) Dunn's post-hoc test results at a significant level of $\alpha = 0.05$ (significant p-values marked with **). ER: Energy residual method. BR: Bowen ratio conservation method. LR: Linear regression.

3.7 LIMITATIONS OF META-ANALYSIS

The control factors of ET_a and uncertainty in their estimates are not the same globally (K. Zhang et al., 2016). As the distribution of validation sites is concentrated in regions where EC flux towers are available (**Figure 2-10**), the results of the validation, thus, are not necessarily transferable to other areas. Therefore, when interpreting the uncertainty of satellite-based ET_a estimates for a specific application, we should consider the validation metrics at each site individually and the variation of these metrics among all locations. Moreover, the study-level factors, including model-specific characteristics, influence the effect size of the meta-analysis. They can also attenuate or amplify each other's effect. In this study, the interaction effects between two moderators (e.g., model type and land cover type) were neglected because the number of records across their joint categories was too sparse and uneven.

No matter which metrics are used, the validation metrics that compare estimate with reference only represent actual error if the reference is the absolute truth. This is never the case because in-situ measurements and upscaling methods are never perfect. Wu et al. (2019b) suggested that validation should be performed in conjunction with uncertainty associated with in-situ measurements and the statistical significance of performance metrics. However, uncertainty in in-situ measurements has not been reported in the selected literature. Therefore, this study only analysed studies that reported results from different treatments of errors in energy balance from EC flux measurements.

The results of this meta-analysis do not necessarily invalidate site-specific conclusions drawn from individual primary studies. Rather, the primary utility of meta-analysis lies in evaluating overall model performance across a broad range of study variations (Evaristo and McDonnell, 2017). These variations include, for example, model parameterization, input data, site characteristics, and even the modeler's expertise. For instance, a single study may compare one representative model from PM, PT, 1SEB, and 2SEB types and find that one model outperforms the others. However, such a result is contingent on the specific conditions of that study, including the chosen model configuration, input data, and the modeler's expertise. Meta-analysis does not confirm or refute the findings of individual studies but instead offers a synthesized comparison across a larger body of evidence. Nonetheless, its conclusions are limited by the quality, consistency, and reporting standards of the underlying primary studies.

3.8 CONCLUSIONS

This chapter discusses the use of performance metrics in reporting uncertainty of satellite-based ET_a estimates and analyses the reported RMSE and PBIAS from studies validated with EC flux measurements. Most studies use a set of three performance metrics or less,

with RMSE, ME, and R^2 being the most used. Each performance metric only reflects an aspect of model performance and does not reveal the variation of uncertainty in satellite-based ET_a estimates due to various factors, including both model-specific and site-specific factors. While RMSE stands as the most employed metric in literature, it is unsuitable for comparing uncertainties across different studies due to its inherent scale dependency. Therefore, meta-analyses were conducted on both the reported RMSE and PBIAS from studies that evaluated satellite-based ET_a estimates by validating against EC flux measurements.

The meta-analysis of 348 studies reveals the distribution and RMSE across studies. RMSE varies among studies due to different ranges of ET_a, models, resampling methods, and site conditions, ranging from 0.01 to 6.65 $mm \cdot d^{-1}$, with a mean of 1.18 $mm \cdot d^{-1}$. Validation with multiple sites reported a lower average and smaller variation of RMSE than validation at single sites. In general, studies with larger temporal support of validation have lower values of mean RMSE in $mm \cdot d^{-1}$.

The meta-analysis of PBIAS values reveals that the mean PBIAS across 135 studies was 6.8 ± 33 %, with a range from -164 to 230 %. Over a decade, studies have not reported a significant improvement in accuracy. Studies reported significantly higher accuracy for the PM model type across studies than for the 1SEB, 2SEB, and PT model types. Other model types are less reported, thus resulting in insufficient sample size in literature. The correction of energy balance closure in EC flux measurements could improve PBIAS of satellite-based ET_a estimates significantly.

Reported metrics from a single validation exercise only reflect the uncertainty of satellite-based ET_a estimates at specific locations and for specific model implementation. Therefore, continuous and consistent validation practice remains important to control the quality of ET_a models and data products, especially when new input data or model modifications are used. The synthesized distribution of RMSE and PBIAS reported in this chapter offers a baseline for the uncertainty in satellite-based estimation of ET_a for consequent chapters and future studies.

PART II: UNCERTAINTY ASSESSMENT IN THE WAPOR PROJECT

Part II comprises three chapters focused on assessing the technical uncertainties inherent in the WaPOR project's AETI data products through ex-ante (Chapter 4) and ex-post approaches (Chapter 5) and the methodological uncertainties of the overall project through reflexive analysis (Chapter 6). Before addressing the research questions in the following chapters, the introduction of Part II presents the WaPOR project, its data products and satellite-based model: the ETLook model.

1. WAPOR PROJECT

WaPOR stands for "**Wa**ter **P**roductivity through **O**pen access of **R**emotely sensed derived data". The WaPOR project is led by the Food and Agriculture Organization of the United Nations (FAO) and funded by the Ministry of Foreign Affairs of the Netherlands. The first phase of the WaPOR project (2016 – 2020) focused on the development and improvement of the WaPOR portal's database. The second phase of the project (2021-2026) continues the improvement and expansion of the WaPOR database to global coverage. At the same time, it places a greater focus on applying and using the data through collaboration with key stakeholders in partner countries[6].

2. WAPOR DATA PORTAL

The WaPOR project developed a database that comprises several variables related to land and water, including ET_a (coded in the WaPOR database as Actual Evapotranspiration and Interception, or AETI) and reference ET (RET). The WaPOR data are openly accessed through the WaPOR web portal (**Figure II-1**). The WaPOR portal is also

[6] Algeria, Mali, Tunisia, Iraq, Jordan, Palestine, Egypt, Sudan, Ethiopia, Kenya, Mozambique, Pakistan and Colombia.

complemented with an Application Programming Interface (API)[7]. The API connects the WaPOR database with web portals[8] and other application software (e.g., Python package, QGIS plugin, mobile apps), enabling direct access to data on several variables. Users interact with the web portal and software to extract information about land and water for specific applications, such as estimating crop water consumption and evaluating irrigation performance.

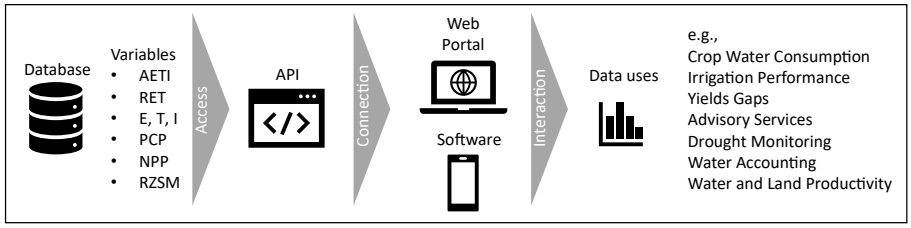


Figure II-1. Diagram of the WaPOR project's data flow. AETI: Actual Evapotranspiration and Interception. RET: Reference Evapotranspiration. E: Evaporation. T: Transpiration. I: Interception. PCP: Precipitation. NPP: Net Primary Production. TBP: Total Biomass Production. GBWP: Gross Biomass Water Productivity. RSM: Relative root-zone Soil Moisture.

2.1. WaPOR-ET data products

The WaPOR-ET data products, which include Actual Evapotranspiration and Interception (AETI), and components of ET_a (i.e., Transpiration, Evaporation, and Interception) are available at three distinct levels of spatial resolutions. These levels are the global level, with a spatial resolution of 250-300 m (Level 1); the national level, 100 m (Level 2); and the sub-national level, 20–30 m (Level 3). The highest temporal resolution is dekad[9] (**Table II-1**). This makes WaPOR-ET one of the highest-resolution global satellite-based ET products (Huang et al., 2025). Moreover, the WaPOR-ET data products are operational with a latency of 5 days, which enables near real-time[10] applications such as

[7] API is a digital tool that allows streamlined connection between different computer systems. An API simplifies interaction by hiding a system's complexity, exposing only useful parts while keeping the front end consistent, even if the system changes.

[8] https://data.apps.fao.org/wapor/ or http://wapor.apps.fao.org/ (deprecated since December 2023)

[9] A dekad is defined as a period of roughly 10 days. A calendar month is divided into 3 dekads: dekad 1 (from the 1st to 10th day), dekad 2 (from 11th to 20th day) and dekad 3 (from 21st to the end of the month). Dekad 1 and 2 consists of 10 days, while the duration of dekad 3 ranges from 8 to 11 days (FAO, 2020a).

[10] Near real-time refers to data that are available almost immediately after it is collected, but with a small delay.

forecasting and monitoring. Since its first version, the WaPOR-ET data products have been updated, incorporating improvements in methodology, input data, and spatial coverage. **Table II-1** also describes the different versions of the WaPOR-ET data products. These versions progressively revised the database methodology, prioritising improvements that are based on the findings of quality assessments carried out by IHE Delft and ITC-Twente (FAO, 2021; FAO and IHE Delft, 2019).

Table II-1. Overview of different versions of WaPOR-ET data. D: dekadal, A: annual, S: seasonal, M: monthly. Source: https://www.fao.org/in-action/remote-sensing-for-water-productivity/wapor-data/en (last access: 10/12/2024).

WaPOR version	Level	Spatial resolution (m)	Temporal resolution	Spatial coverage	Temporal coverage	Release
v1	L1	250	D, A	Africa and the Near East[11]	2009 - 2018	June 2018
	L2	100	D, S, A	Selected basins and countries		
	L3	30	D, S, A	Selected pilot areas		
v2	L1	250	D, M, A	Africa and the Near East	2009 - 2023	June 2019
	L2	100	D, M, A	Selected basins and countries		
	L3	30	D, M, A	Selected pilot areas		
v3	L1	300	D, M, A	Global	2018 - Present	October 2023
	L2	100	D, M, A	Africa, the Near East, Colombia, Sri Lanka		
	L3	20	D, M, A	Selected pilot areas		

2.2. Uses of WaPOR-ET data

The WaPOR-ET data have been used in several studies for various purposes. At the river basin scale, it has been used in quantification of water consumptive use (Al-Bakri et al., 2022; Al-Omoush et al., 2025) and its trend (Barideh and Nasimi, 2022), calculation of groundwater balance (Kivi et al., 2022) and basin water accounting (Al-Omoush et al., 2025; Amdar et al., 2024; Kivi et al., 2022). At the field scale, it has been used to quantify crop water consumption (Mebrie et al., 2023), crop and economic water productivity (Hazimeh and Jaafar, 2024; Jaafar et al., 2024; Pitoro et al., 2025; Sequeira et al., 2025), evaluate irrigation performance (Blatchford et al., 2020a; Chukalla et al., 2022), predict

[11] The term "Near East" is a location referenced in relation to Europe, whereas "Southwest Asia" offers a more objective geographic reference to the position of the same region on the Asian continent.

user crop biomass (Servia et al., 2022), estimate applied irrigation volume (Corbari et al., 2025), and monitor land and water productivity gaps (Chukalla et al., 2024).

As the second phase of the WaPOR project focuses on demand-driven data applications, it develops a compendium of user-focused solutions and practical tools based on information from the WaPOR database. Most uses and applications presented on the WaPOR project website are by user organizations based in the Netherlands (e.g., IHE Delft, Wageningen University and Research), Europe (e.g., European Space Agency, FAO), Sri Lanka (e.g., International Water Management Institute) (**Figure II-2**). Some applications are by user organizations within the project's partner countries, such as Egypt, Tunisia, Iran, Jordan, Lebanon, Ethiopia, Rwanda, and Kenya. The WaPOR database is used and applied mostly African countries (e.g., Ethiopia, Kenya, Egypt…), which is mostly due to the coverage of WaPOR version 2. This demonstrates the project's strong focus on applications in the regions of Africa and Southwest Asia, while the tools and solutions are mainly developed by project partners in the Global North.

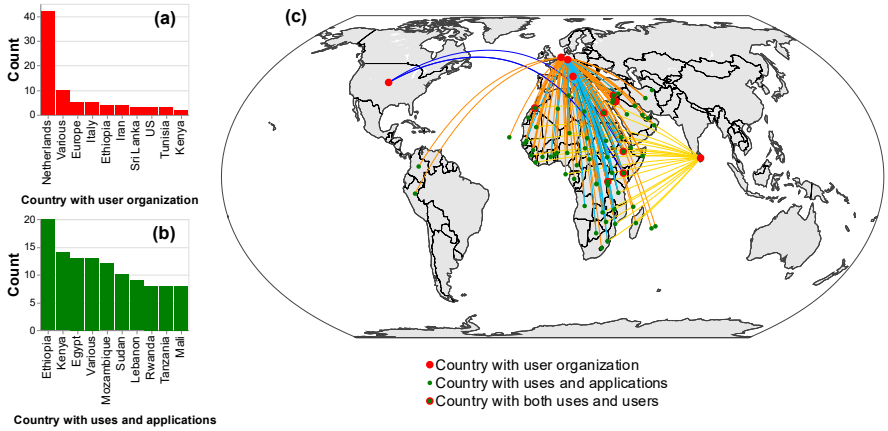


Figure II-2. Users and uses of the WaPOR-ET data products. (a) Countries with user organization (top 10 with most applications). (b) Countries with most uses and applications (top 10). (c) Links between country with user organization and country with uses, coloured by country. Data source: https://www.fao.org/in-action/wapor-uses-applications-catalogue/en/ (last access: 10/12/2024).

3. ETLOOK MODEL

ETLook is a model developed by Pelgrum et al. (2010) and Bastiaanssen et al. (2012) to estimate ET under all weather and surface conditions by combining satellite-based data from optical and passive microwave sensors. The WaPOR project adopted the ETLook

model as the core algorithm for generating the WaPOR-ET data at the three levels (**Table II-2**).

3.1. Model formulation

The ETLook model uses the Penman-Monteith equation (Monteith, 1965), which has been modified to separately calculate evaporation from soil (II.1) and transpiration from canopy (II.2) (Bastiaanssen et al., 2012). It estimates interception using the model by von Hoyningen-Hune (1983) with satellite-based vegetation indices and precipitation data (II.3).

$$\lambda E_{\mathrm{e}} = \frac{\Delta\left(R_{\mathrm{n,\,soil}} - G\right) + \rho c_{\mathrm{p}}\left(\frac{\Delta_{\mathrm{e}}}{r_{\mathrm{a,soil}}}\right)}{\Delta + \gamma\left(1 + \frac{r_{\mathrm{soil}}}{r_{\mathrm{a,soil}}}\right)} \tag{II.1}$$

$$\lambda E_{\mathrm{t}} = \frac{\Delta\left(R_{\mathrm{n,\,canopy}}\right) + \rho c_{\mathrm{p}}\left(\frac{\Delta_{\mathrm{e}}}{r_{\mathrm{a,canopy}}}\right)}{\Delta + \gamma\left(1 + \frac{r_{\mathrm{canopy}}}{r_{\mathrm{a,canopy}}}\right)} \tag{II.2}$$

$$\frac{E_{\mathrm{i}}}{86400} = 0.2\mathrm{LAI}\left[1 - \left(\frac{1}{1 + \frac{\left(c_{\mathrm{veg}}\right)P}{0.2\mathrm{LAI}}}\right)\right] \tag{II.3}$$

where E_{e}, E_{t}, and E_{I} [kg·m^{-2}·s^{-1}] are the soil evaporation, transpiration, and interception, respectively; λ [J·kg^{-1}] is the latent heat of vaporisation of water at specific air temperature; $R_{\mathrm{n,soil}}$ and $R_{\mathrm{n,canopy}}$ [W·m^{-2}] are the net radiation partitioned for the soil and canopy; G [W·m^{-2}] is the ground heat flux; ρ [kg·m^{-3}] is the air density; c_{p} [J·kg^{-1}·K^{-1}] is the specific heat of dry air; Δ_{e} [Pa] is the vapour pressure deficit (VPD), which is the difference between the saturated and actual vapour pressure ($e_{\mathrm{s}} - e_{\mathrm{a}}$); Δ [Pa·K^{-1}] is the slope of the saturation vapour pressure curve; γ [Pa·K^{-1}] is the psychrometric constant; r_{soil} and r_{canopy} [s·m^{-2}] are the surface resistances of soil and canopy; $r_{\mathrm{a,soil}}$ and $r_{\mathrm{a,canopy}}$ [s·m^{-2}] are the aerodynamic resistances of soil and canopy; LAI [m^2·m^{-2}] is the Leaf Area Index; c_{veg} [-] is the vegetation cover; and P [mm·d^{-2}] is the precipitation.

In the ETLook model, R_n, G, Δ, c_p, γ, and Δ_e are calculated from meteorological data following the standard procedures described in the FAO Irrigation and Drainage paper 56 (FAO56), by Allen et al. (1998). The ETLook model integrates several other models to derive the variables required for the equations (Bastiaanssen et al., 2012), including net radiation partitioning (Shuttleworth and Wallace, 1985), soil surface resistance (Camillo and Gurney, 1986; Dolman, 1993; Shu Fen Shun, 1982), canopy surface resistance (Choudhury et al., 1986; Holtslag, 1984), aerodynamic resistance (Jiang et al., 2006),

vegetation cover (Carlson and Ripley, 1997), and effective LAI (Allen et al., 2006; Ben Mehrez et al., 1992).

3.2. Model implementation

The ETLook model was developed and operationalized on a global scale by eLEAF[12], a partner of the WaPOR project. This implementation, referred to here as WaPOR-ETLook, generates the WaPOR-ET data products. FAO has developed an open-source Python implementation of the ETLook model, called pyWaPOR (FAO, 2025a). The pyWaPOR model can be implemented at any location, although its spatial extent can be limited by computational power and the availability of cloud-free satellite images. The pyWaPOR model also allows model users to customize input data sources and model parameterization.

The ETLook model requires several input data for its implementation, such as meteorological data, soil moisture, surface reflectance (for vegetation index and albedo), a digital elevation model (DEM) and land cover classification (LCC). Meteorological data includes air temperature, air pressure, wind speed, humidity, precipitation, and solar radiation. These are derived from either in-situ weather stations (Bastiaanssen et al., 2012) or a combination of climate reanalysis and meteorological satellites (FAO, 2020a, 2018a, 2018b).

The WaPOR-ETLook model differs from the first version of ETLook model mainly in the derivation of soil moisture information (Blatchford et al., 2020b). In the WaPOR-ETLook model, the relative soil moisture content is estimated using a modified trapezoid model proposed by Yang et al. (2015a). This model uses optical and thermal imagery rather than passive microwave remote sensing as applied in the original model (FAO, 2020a, 2018a, 2018b). Therefore, land surface temperature (LST) is also a key input in the WaPOR-ETLook model (**Table C-1**). The sources of input data for the WaPOR-ET data production vary by level and version (**Table C-2**).

4. PREVIOUS UNCERTAINTY ASSESSMENTS

Prior to the WaPOR-ETLook implementation, Bastiaanssen et al. (2012) conducted the first local sensitivity analysis of the ETLook model in the Indus basin. This analysis focused on two representative land uses (bare soil and irrigated rice-wheat) and five selected input parameters: relative and saturated soil moisture, normalized difference vegetation index (NDVI), soil surface resistance parameters, and minimum stomatal

[12] https://eleaf.com/

resistance. Among these, surface soil moisture was found to be the most influential for describing ET variability at the selected locations. The uncertainty analysis of the selected input parameters indicated only a 3.4 % error in the model's output, which demonstrate the promising applicability of the ETLook model for agricultural land (Bastiaanssen et al. (2012). However, this uncertainty assessment was limited to specific locations, a selective set of parameters, and the assumed uncertainty of those input parameters.

Several ex-post uncertainty assessments of the ETLook model have been conducted through the quality assessment activities within the WaPOR project (Blatchford et al., 2020b; FAO, 2021; FAO and IHE Delft, 2019). These assessments inform the applications of the WaPOR-ET data products and the development of subsequent versions. The quality assessment of WaPOR version 1 focused on the consistency of the WAPOR-ET data with various sources, such as other data products, hydrological model outputs, water balance for large river basins, and Eddy Covariance (EC) flux towers in Africa (FAO and IHE Delft, 2019). The report suggested higher errors in locations with higher aridity and highlighted the need to improve the spatial resolution of input data, which was later implemented in the version 3 (**Appendix C**). While WaPOR version 2 primarily improved other variables, a key change to the ET data was an update to the land cover. Evaluations of the version 2 WaPOR-ET data products found them to perform well, though with some overestimation in irrigated fields in Egypt (Blatchford et al., 2020b). The report also recommended verifying the behaviour of meteorological and thermal data, noting lower performance at sites with lower quality of NDVI and LST inputs, or at sites that are energy-limited (FAO, 2021).

Outside of the model development context, some studies have evaluated the WaPOR-ET data products using different methods, focusing on a specific study area or application. Some studies have suggested that the uncertainties in the WaPOR-ET data might influence their results (Al-Bakri et al., 2022; Al-Omoush et al., 2025; Chukalla et al., 2022; Hazimeh and Jaafar, 2024; Kivi et al., 2022; Mebrie et al., 2023). For example, Al-Omoush et al. (2025) found that the estimates from WaPORv2L2 ET data were approximately 50 % lower compared to other five ET data products in the Amman-Zarqa Basin, Jordan. This underestimation can lead to an exaggeration of recoverable water in their ground water model if the WaPOR-ET data was not correct. Beyond the uncertainties inherent to the WaPOR-ET data, methodological uncertainties stemming from the choice of ET models (Hazimeh and Jaafar, 2024) and spatial resolution (Blatchford et al., 2020a) also impact analyses such as crop water productivity and irrigation performance (Seijger et al., 2023).

5. SIGNIFICANCE AND RELEVANCE

The WaPOR project presents an interesting case study for understanding the implication of uncertainties in satellite-based estimation for water resources assessment and management. By stitching several empirical and semi-empirical models and data sources together, the ETLook model simplifies and enables the efficient calculation of ET for various surface and climate conditions on a global scale. As Van Stan II and Simons (2025) argued, this form of "patchwork empiricism" risks creating a mismatch between global models and local realities. For example, the vegetation cover model was tested by Jiang et al. (2006) using ground-based surface reflectance measurements from cotton experiments conducted by Huete et al. (1985) in Arizona. Meanwhile, the effective LAI model used by Ben Mehrez et al. (1992) was based on experimental studies of different corn strata in Canada (Rochette et al., 1991). Consequently, the choice of these sub-models introduces methodological uncertainties when the ETLook model is applied to different locations across the world.

The influence of the ETLook model extends beyond scientific research on ET, as it could affect water management practices or tools that rely on WaPOR-ET data products. As one of the most easy-to-access and high-resolution ET data in many regions, particularly Africa and Southwest Asia (**Table II-1**), the WaPOR-ET data products are used in analyses of water resources at multiple scales. Furthermore, the WaPOR-ET data products are supported by FAO and various governments for being used in tools and applications that inform farmers, water managers, and policy makers.

From the above, it is important for the quality assessment of the WaPOR-ET data to include uncertainty assessments, which are presented in this part of the thesis. Chapter 4 focuses on ex-ante uncertainty assessment by examining the input meteorological data. This is particularly relevant because the major updates in version 3 of the WaPOR-ET data is a change of input datasets (**Appendix C**). Chapter 5 focuses on the ex-post uncertainty assessment of the WaPOR-ET data version 3 Level 1. Its global coverage makes it possible to validate the data in regions where Eddy Covariance data are available (**Figure 2-10**). These analyses contribute directly to the quality assessment of the latest version of the WaPOR-ET data products. Chapter 6 analyses the influence of problem framings, model choices and configurations on the WaPOR-ETLook model, as well as the perception and management of output uncertainties in the WaPOR project.

4

EX-ANTE UNCERTAINTY ASSESSMENT

This chapter is partly based on:

Tran, B. N., Dehati, S., Seyoum, S., van der Kwast, J., Jewitt, G., Uijlenhoet, R., and Mul, M.L., 2025. Evaluating reanalysis datasets as meteorological input for estimating reference evapotranspiration in Africa and Southwest Asia. Hydrological Sciences Journal. https://doi.org/10.1080/02626667.2025.2600682

4.1 INTRODUCTION

Almost all the methods to calculate reference ET (ET_o) and actual ET (ET_a), including the FAO56 method, require data of at least some meteorological variables that are often measured with standard weather stations, such as air temperature, pressure, humidity, windspeed, and solar radiation. Since some variables are not always available, the FAO56 guideline provides alternative equations to estimate missing meteorological variables (Allen et al., 1998, chap. 3). Still, many regions lack weather stations, especially Africa and Southwest Asia (Dinku, 2019; van de Giesen et al., 2014). Therefore, mapping ET_o over large regions depends on either interpolated gridded weather datasets or climate reanalysis data (Abatzoglou et al., 2018; Singer et al., 2021). At the same time, the estimation of ET_a aided by satellite observation are also dependent on models and forcing inputs (McCabe et al., 2017; Tran et al., 2023). Meteorological data are therefore essential forcing inputs for both ET_o and ET_a calculations. Despite the increasing development and use of spatial ET_o data, there is limited understanding of the accuracy and precision of these data, and how they are related to the uncertainties of the meteorological input data.

Reanalysis is a compelling alternative to interpolated gridded weather datasets for mapping ET_o. Reanalysis are data generated using data assimilation techniques to couple numerical weather prediction (NWP) models with past observations (Eyre et al., 2022; Slivinski, 2018). Recent developments in atmospheric reanalysis have greatly improved spatial resolution, notably the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5) and its derived dataset AgERA5 (up to about 11 km resolution) (Copernicus Climate Change Service, 2020; Hersbach et al., 2020). In addition, the short latency of some NWP models and data assimilation systems facilitates operational near real-time monitoring of ET_o and ET_a. For example, the FAO's portal to monitor Water Productivity through Open access of Remotely sensed derived data (WaPOR) provides global daily ET_o updated within 3 days calculated using the Goddard Earth Observing System version 5 (GEOS5) dataset, which has a latency of less than a day (FAO, 2024).

Parker (2016) argued that the reliability of reanalyses is unclear due to partial understanding of the errors and uncertainties in NWP models and thus advocated both quantitative and qualitative assessment of uncertainties in reanalysis data. The explicit and standardized quantification of uncertainties in NWP models and input parameters is often overlooked. For example, Wang et al. (2024) emphasized that uncertainties in cloud optical thickness, aerosol optical depth, and ozone significantly impact solar radiation estimates, a key input for ET_o and ET_a calculation. Lang et al. (2024) demonstrated that coarse-resolution reanalysis data can introduce substantial errors in solar radiation estimates, due to mixed-pixel effects, especially under cloudy conditions (L. Wang et al., 2024a).

Since the sensitivity of ET_o models to errors in meteorological forcing varies with different models, space, and time (Fisher et al., 2017), it is important to analyse uncertainty in each ET_o model when using reanalysis data. Furthermore, climate reanalysis and reanalysis-based evaporation estimates are increasingly used to study hydrological processes, which may lead to errors in these estimates being amplified and misinterpreted in hydrological studies. This necessitates a comprehensive review and description of both evaporation retrieval models and their forcing components (McCabe et al., 2017). The reliability of ET_o calculated using particular reanalysis datasets has been investigated, mainly in southern Europe and China (e.g., Ippolito et al., 2024; Martins et al., 2017; Xu et al., 2024) where past weather observations for reanalysis are more available than in other parts of the world (Brönnimann et al., 2018; Soci et al., 2024). For instance, in Africa and Southwest Asia, where weather observations are scarce, the reanalysis data quality is largely unknown as well as the impact of the meteorological uncertainty on ET_o calculations.

The objective of this chapter is therefore to assess the uncertainty of meteorological forcing from reanalysis products (namely GEOS5, ERA5, and AgERA5) and the resulting uncertainty in ET_o over Africa and Southwest Asia. In this study, we focused on ET_o since the definition and calculation of ET_o depends only on meteorological forcing. However, we also extend our discussion to the impact of uncertainty in meteorological forcing on ET_a estimation where relevant.

4.2 MATERIALS AND METHODS

We assessed the uncertainty of three reanalysis data products (GEOS5, ERA5, AgERA5) for five meteorological inputs in the FAO56 ET_o calculation, namely air temperature, atmospheric pressure, windspeed, vapour pressure, and solar radiation for a five-year period (from 2018 to 2022). Our assessment entails three components: uncertainty between products, nominal accuracy, and quantitative impact of uncertainty in inputs on ET_o (**Figure 4-1**). The uncertainty between products was assessed by spatial and temporal pair-wise comparison. The nominal accuracy[13] was assessed by comparison with time-series data from in-situ measurements. Finally, the impact of uncertainty in inputs on ET_o was assessed by two error propagation methods (Monte Carlo simulations and Taylor expansion).

[13] The results from comparison with in-situ measurement is considered nominal accuracy since in-situ measurements also have errors and not necessarily present true values of the grid cells due to spatial scale mismatch.

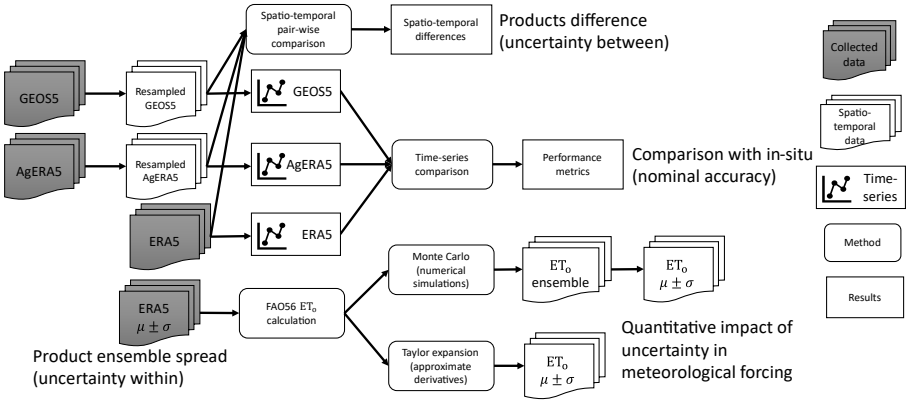


Figure 4-1. Schematization of the methodological framework. Collected reanalysis data (GEOS5, AgERA5, and ERA5) were resampled and processed for spatio-temporal pair-wise comparison to calculate difference between products, which represents uncertainty between products. Time-series were extracted at grid cells for comparison with in-situ measurements, to calculate performance metrics, which represent nominal accuracy. The ensemble spread of ERA5, which represents uncertainty within ERA5 product, was used to propagate errors in FAO56 reference evapotranspiration (ET_o) calculation using Monte Carlo and Taylor expansion methods.

4.2.1 Study areas and in-situ data

This study covers the land mass of Africa and Southwest Asia (30ºS – 60ºN, 40ºW – 40ºE) with a wide range of climates, from arid to tropical (**Figure 4-2**). About half of the study area is arid desert in the North Africa and Southwest Asia regions. There are a few climate monitoring networks that cover only fractions of the study area. The Trans-African Hydro-Meteorological Observatory (TAHMO) is an initiative that has successfully extended a network of meteorological and hydrological stations in sub-Saharan Africa (van de Giesen et al., 2014). Currently, TAHMO provides hydro-meteorological measurements from the largest number of stations in the region of interest. However, most stations are distributed in specific regions: western, eastern and southern Africa (**Figure 4-2**).

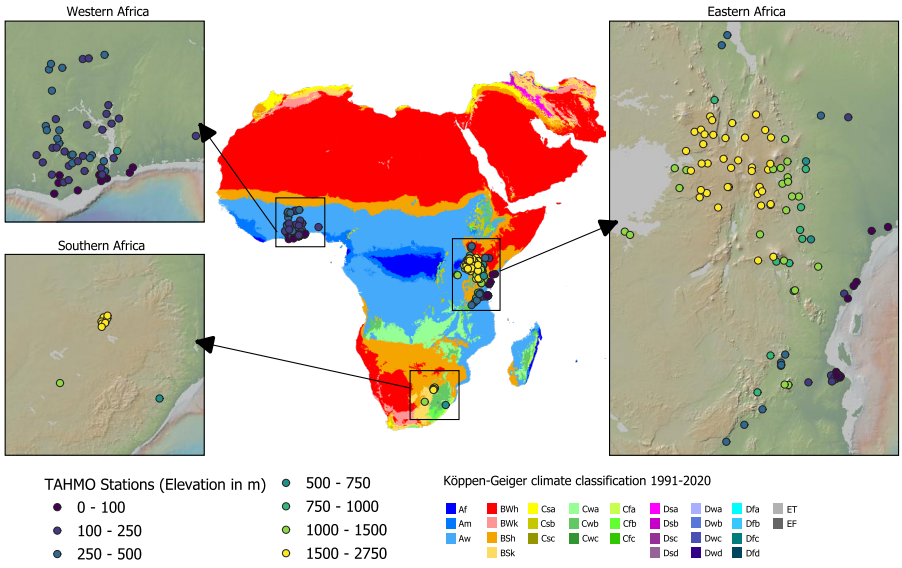


Figure 4-2. Climate classification map of study area and the locations of in-situ observations. Data source: TAHMO, Köppen-Geiger map (Beck et al., 2023). Base map: Natural Earth NE1_50M_SR_W.

In-situ climate data were collected from 174 TAHMO stations (**Figure 4-2**). **Table 4-1** shows the distribution of TAHMO stations by climate classes. The dataset includes hourly measurement of standard meteorological variables (i.e., air temperature, relative humidity, wind speed and direction, solar radiation and atmospheric pressure) from ATMOS 41 sensors (METER, 2023). Daily mean air temperature [^{o}C], relative humidity [%], wind speed at 2 m [$m \cdot s^{-1}$], air pressure [mbar], and solar radiation [$W \cdot m^{-2}$] were computed by averaging hourly data (TAHMO, 2023).

Table 4-1. Number of stations per each of major climate classes in study area. The classes with <0.5% of total area and no stations are omitted from the table.

Code	Description	Area [% of total]	Number of stations
BW	Arid, desert	51.0	3
Aw	Tropical, savannah	22.1	114
BS	Arid, steppe, hot	14.6	9
Cw	Temperate, dry winter	4.5	14
Am	Tropical, monsoon	2.5	4
Cs	Temperate, dry summer	2.1	13
Af	Tropical, rainforest	1.8	4
Cf	Temperate, no dry season	0.5	13
Total area		**99.1**	**174**

The quality of in-situ data varies and depends on the accuracy specification of sensors, sensor performance and operation continuity. We acquired data from stations with the best quality flags provided by the TAHMO quality control procedure (van de Giesen et al., 2014). Quality control procedure for TAHMO stations is extended from the procedures in the Oklahoma Mesonetwork (Shafer et al., 2000). TAHMO employs both automated and manual methods to ensure data quality, including tests for range, sensor accuracy, climate condition, temporal changes, dips and spikes, and changes in variance (Annor, 2023, p. 125). In addition, we evaluated the quality and integrity of the acquired dataset by visually checking the time series of each climatic variable at each station.

4.2.2 Reanalysis data

a. Datasets description

We selected and retrieved data for a study period of five years, from 1/1/2018 to 31/12/2022, which we considered sufficient sample of daily values for analysing spatio-temporal pattern of uncertainty (1826 data points per time-series). This is also the period for which we have access to TAHMO climate data. The hourly GEOS5 and ERA5 data and daily AgERA5 data were retrieved from the sources that are provided in **Table 4-2**.

Table 4-2. Spatial and temporal resolution of the three reanalysis datasets used. The specific data products were acquired from the data source given in parenthesis.

Dataset name	Spatial resolution	Temporal resolution	Product name (data source)
GEOS5	0.31º × 0.25º (~30 km at the equator)	hourly	GEOS-FP tavg1_2d_slv_Nx (https://opendap.nccs.nasa.gov/dods/GEOS-5/fp/0.25_deg/assim/tavg1_2d_slv_Nx.info) GEOS-FP tavg1_2d_rad_Nx (https://opendap.nccs.nasa.gov/dods/GEOS-5/fp/0.25_deg/assim/tavg1_2d_rad_Nx.info)
ERA5	0.25º × 0.25º (~28 km at the equator)	hourly	ERA5 hourly data on single levels from 1940 to present (https://doi.org/10.24381/cds.adbb2d47) (Hersbach et al., 2023)
AgERA5	0.1º × 0.1º (~11 km at the equator)	daily	Agrometeorological indicators from 1979 to present derived from reanalysis (https://doi.org/10.24381/cds.6c68c9bb) (Boogaard et al., 2020)

GEOS5 stands for the Goddard Earth Observing System version 5, a global atmospheric model, developed by the NASA Global Modelling and Assimilation Office (GMAO) (Rienecker et al., 2008). The GEOS Data Assimilation System is the integration of GEOS5 Atmospheric General Circulation Model and the Grid-point Statistical

Interpolation Analysis (Rienecker et al., 2008). GMAO runs the GEOS Forward Processing stream, which generates both forecasts and assimilation products. The meteorological variables from the time-average 1-hourly data, 2-dimensional, single-level [14] atmospheric state variables (tavg1_2d_slv_Nx) and radiative fluxes (tavg1_2d_rad_Nx) data products were retrieved. The GEOS5 dataset includes the following variables: hourly air temperature at 2 m [K], northward and eastward components of wind at 10 m [m s^{-1}], specific humidity at 2 m [kg kg^{-1}], sea level pressure [Pa], surface pressure [Pa], and surface incoming shortwave flux [J m^{-2} h^{-1}]. The GEOS5 system derives air pressure at surface level from mean sea level pressure using the United States Geological Survey's 1-km Global Elevation (GTOPO30) raster data (Rienecker et al., 2008).

ERA5 is the fifth generation of the ECMWF atmospheric reanalysis of the global climate. ERA5 is generated by combining the model forecasts from the Integrated Forecasting System Cy41r2 with vast amounts of historical observations using the 4D-Var assimilation scheme (Hersbach et al., 2020). ERA5 provides data products for several climate variables at 137 pressure levels from the surface up to 80 km. In this study, the ERA5 hourly data on single levels (Hersbach et al., 2023) was retrieved from ECMWF's Climate Data Store (CDS). The ERA5 dataset includes the following variables: hourly 2 m temperature [K], 2-m dewpoint temperature [K], 10-m *v*-component and *u*-component of wind speed [$m \cdot s^{-1}$], mean sea level pressure [Pa], surface pressure [Pa], and surface solar radiation downwards [$J \cdot m^{-2} \cdot h^{-1}$]. ERA5 surface pressure was computed using surface elevation data interpolated from the Shuttle Radar Topography Mission Digital Elevation 30m data (SRTM30) combined with other surface elevation datasets (ECMWF, 2024a).

AgERA5 comprises agrometeorological indicators derived from reanalysis, providing input needed for most crop growth models. Daily AgERA5 data is produced by aggregating ERA5 hourly data to daily at the local time zone and downscaling towards a finer topography at a 0.1° spatial resolution (Boogaard et al., 2023). The ERA5 data was corrected using regression equations that were calibrated with the ECMWF's operational high-resolution atmospheric model (HRES) for each variable and grid (Boogaard et al., 2020). AgERA5 provides data products for 12 meteorological variables at the surface level and daily timestep. The AgERA5 dataset was collected from ECMWF's CDS, which includes the following variables: daily 2-m temperature [K], 2-m dewpoint temperature [K], 10-m wind speed [$m \cdot s^{-1}$], and solar radiation flux [$J \cdot m^{-2} \cdot d^{-1}$].

[14] In the context of climate reanalysis, single level data includes variables measured or modelled close to the surface.

b. Reanalysis data pre-processing

Prior to our analysis, we applied a pre-processing procedure on the retrieved reanalysis data products to ensure consistency among variables and units for accurate comparisons. Simple linear temporal and spatial aggregation were applied to achieve the same resolution for comparison of different input datasets. For spatial and temporal comparison, we analysed meteorological variables at the coarser resolution of the data products to avoid introducing errors due to spatial downscaling and elevation correction. Reanalysis data was converted to the same unit as in-situ data.

The average windspeed at 10 m is derived from the windspeed components collected from ERA5 and GEOS5:

$$u_{10} = \sqrt{{u_{10x}}^2 + {u_{10y}}^2} \tag{4.1}$$

where u_{10} [m·s^{-1}] is wind speed at 10 m, u_{10x} [m·s^{-1}] is 10-m eastward wind or u-component, and u_{10y} [m·s^{-1}] is 10 m northward wind or v-component.

Notably, the reanalysis datasets do not have relative humidity except for AgERA5. Therefore, we compared vapour pressure derived from ERA5 and AgERA5 dew-point temperature and from GEOS5 specific humidity, which is required for the calculation of ET_o. The calculation of vapour pressure from specific humidity is:

$$e_a = \frac{q_v \cdot P}{\varepsilon} \tag{4.2}$$

where e_a [kPa] is vapour pressure, q_v [kg·kg^{-1}] is specific humidity, P [kPa] is air pressure, and ε [-] is the ratio of molecular weight of water to dry air (ε = 0.622).

The saturation vapour pressure at actual temperature was calculated following Allen *et al.* (1998, Eq. 11):

$$e_s(T) = 0.6108 \exp\left(\frac{17.27\,T}{T + 237.3}\right) \tag{4.3}$$

where $e_s(T)$ is saturated vapour pressure [kPa] at the actual air temperature T [°C].

The actual vapour pressure e_a [kPa] equals the saturated vapour pressure e_s at the dewpoint temperature T_d [°C]: $e_a = e_s(T_d)$. Therefore, the saturated vapour pressure was calculated by substituting T_d from reanalysis data for T in Equation (4.3), following Allen *et al.* (1998).

For in-situ dataset, vapour pressure was derived from minimum and maximum relative humidity following Allen *et al.* (1998, Eq. 17) since dew-point temperature was not available:

$$e_a = \frac{e_s(T_{\min}) \times \frac{\mathrm{RH}_{\min}}{100} + e_s(T_{\max}) \times \frac{\mathrm{RH}_{\max}}{100}}{2} \quad (4.4)$$

where $\mathrm{RH}_{\min}$ is daily minimum relative humidity [%], $\mathrm{RH}_{\max}$ is daily maximum relative humidity [%], and $e_s(T)$ is the saturation vapour pressure [kPa] at the same temperature.

4.2.3 Uncertainty assessment methods

a. Spatial and temporal pair-wise comparison

The uncertainty between products was assessed by pair-wise comparison. Before that, we aggregated hourly reanalysis data to daily, daily to monthly, and monthly to yearly by arithmetic averaging. The higher-resolution datasets (GEOS5 and AgERA5) were resampled to the spatial resolution of ERA5 to ensure that all datasets represent the same level of detail, allowing for an unbiased comparison. For spatial comparison, we computed the yearly average of the differences between each pair for mean air temperature at 2 m, wind speed at 10 m, vapour pressure, and solar radiation. For atmospheric pressure, we compared the pressure at sea level and the pressure at surface originally retrieved from GEOS5 and ERA5. Since the AgERA5 dataset does not include data of air pressure, it was excluded from air pressure comparison. For spatio-temporal comparison, we used the latitude-time Hovmöller diagrams of monthly average maps (Hovmöller, 1949), which helps visually detect seasonal anomalies or any dynamics of the discrepancy between datasets.

b. Comparison with in-situ data and performance metrics

The nominal accuracy of reanalysis data was evaluated by comparison with in-situ measurements. Daily time-series of air temperature, air pressure, windspeed, vapour pressure, and solar radiation were extracted from the reanalysis datasets at the grids containing observation stations. The performance metrics we used to validate reanalysis data against in-situ data include the coefficient of determination (R^2); the root mean square error (RMSE), the bias, and the relative bias (PBIAS) (**Table B-1**). The R^2 metric is the square of Pearson correlation coefficient, which measures how well the variables from in-situ data are correlated to the temporal variation of variables derived from reanalysis products. The bias measures mean residuals, while the RMSE measures the root mean square difference between reanalysis and in-situ data. These metrics are widely used in in-situ validation of earth observation data (Mayr et al., 2019; Tran et al., 2023).

4.2.4 Error propagation methods

We analysed the impact of uncertainty in ET_o propagated from the meteorological inputs from reanalysis on ET_o by applying error propagation methods in the FAO56 calculation of ET_o. We applied and compared the Monte Carlo (MC) method with the Taylor method. The MC method is a statistical approach to estimate the uncertainty in a complex mathematical model $g(\cdot)$ by performing random sampling (Heuvelink, 1998). It entails randomly simulating inputs based on known or assumed probability distributions, applying these inputs to the model, and deriving the uncertainty and variability from the resulting outputs (Kroese and Rubinstein, 2012). The MC method is well-suited for non-linear functions and models with multiple variables. However, it can be computationally expensive and time-consuming when a large number of simulations are needed to obtain accurate estimates of the probability distribution (Heuvelink, 1998; Kroese et al., 2014).

The Taylor method is based on the theory of error propagation, which applies the Taylor expansion for linear approximation of non-linear functions (Taylor, 1997). The main advantages of using Taylor method are efficient computation and the analytical form of the variance of the output error. However, when the operation $g(\cdot)$ is strongly non-linear or involves many inputs, like the ET_o calculation, the approximation error may increase and computational efficiency may decrease (Heuvelink, 1998, p. 43). Therefore, comparing the Taylor method and the MC method can provide more insights to guide future applications of ET_o error propagation. Our intention was to evaluate whether Taylor method can be an alternative to MC method for operational ET_o uncertainty estimation.

a. Calculation of FAO56 reference evapotranspiration

The daily ET_o [mm·d^{-1}] was calculated following the FAO56 Penman-Monteith equation for reference crop following the procedure described by Allen *et al.* (1998):

$$\mathrm{ET_o} = \frac{0.408\Delta(R_\mathrm{n} - G) + \gamma \frac{900}{T_\mathrm{mean} + 273} u_2 (e_\mathrm{s} - e_\mathrm{a})}{\Delta + \gamma(1 + 0.34u_2)} \tag{4.5}$$

where Δ [kPa·°C^{-1}] is the slope of saturation vapour pressure curve, R_n [MJ·m^{-2}·d^{-1}] is the net radiation at the reference crop surface, G [MJ·m^{-2}·d^{-1}] is the soil heat flux density assumed to be zero for day period, γ [kPa·°C^{-1}] is the psychrometric constant, T_mean [°C] is the daily mean air temperature, u_2 [m·s^{-1}] is daily average wind speed at 2 m, e_s [kPa] is saturation vapour pressure, e_a [kPa] is actual vapour pressure, T_mean is the average of minimum and maximum air temperature (T_min and T_max).

The slope of saturation vapour pressure curve (Δ) was calculated following Allen *et al.* (1998, Eq. 13). The psychrometric constant (γ) was calculated following Allen *et al.* (1998, eq. 8).

The net radiation (R_{n}) was calculated by subtracting the net longwave radiation (R_{nl}) from the net shortwave solar radiation (R_{ns}):

$$R_{\mathrm{n}} = R_{\mathrm{ns}} - R_{\mathrm{nl}} = (1-\alpha)R_{\mathrm{s}} - R_{\mathrm{nl}} \tag{4.6}$$

where $\alpha = 0.23$ is surface albedo for the hypothetical grass reference and R_{s} [MJ·m^{-2}·d^{-1}] is the incoming solar radiation from reanalysis data. R_{nl} [MJ·m^{-2}·d^{-1}] was calculated following Allen et al. (1998, Eq. 39).

$$R_{\mathrm{nl}} = \sigma \left[\frac{(T_{\min}+273.16)^4 + (T_{\max}+273.16)^4}{2}\right] \times \left(0.34 - 0.14\sqrt{e_{\mathrm{a}}}\right) \times \left(1.35\frac{R_{\mathrm{s}}}{R_{\mathrm{so}}} - 0.35\right) \tag{4.7}$$

where $\sigma = 4.903 \cdot 10^{-9}$ MJ·K^{-4}·m^{-2}·d^{-1} is Stefan-Boltzmann constant, $\left(0.34 - 0.14\sqrt{e_{\mathrm{a}}}\right)$ and $\left(1.35\frac{R_{\mathrm{s}}}{R_{\mathrm{so}}} - 0.35\right)$ are correction terms for air humidity and cloudiness respectively, and R_{so} [MJ·m^{-2}·d^{-1}] is the clear-sky radiation calculated from the extraterrestrial radiation R_{a}, which was, in turn, calculated following Allen *et al.* (1998, Eq. 21).

The wind speed at 2 m was estimated from reanalysis windspeed at 10 m using the logarithmic wind speed profile (Allen et al., 1998, Eq. 47):

$$u_2 = u_{10}\frac{4.87}{\ln(67.8 \times z - 5.42)} \tag{4.8}$$

where u_{10} [m·s^{-1}] is the wind speed at 10 m, u_2 [m·s^{-1}] is the wind speed at 2 m, and z = 10 m is the height at which wind speed is calculated.

b. Uncertainty estimation of reanalysis datasets

Since only the ERA5 data product comes with uncertainty estimation, we focused on the propagation of ERA5 data uncertainty estimates in the FAO56 ET_o calculation. The uncertainty estimates of the ERA5 data product was produced by ECMWF by sampling from 10 underlying ensemble members every 3 hours (ECMWF, 2023b). This uncertainty estimation addresses mostly random errors in the observations and sea surface temperature model parameterization. The estimates are closely related to the uncertainty of the assimilated observations, which have evolved considerably over time. The ERA5 uncertainty estimates vary in different zones due to the uneven distribution of measurements that are used to correct the forecasting model through data assimilation. However, the systematic errors were not addressed and not correlated to the computed uncertainty estimates (ECMWF, 2023b).

The mean and spread (standard deviation) of the ERA5 ensemble were retrieved from ECMWF's CDS (Hersbach et al., 2023) (**Figure S1**, **Appendix D**). We considered each meteorological variable as a quantitative spatial attribute $A(\cdot) = \{A(x) \mid x \in D\}$ where D is the domain of interest in a 3-dimensional space (longitude-latitude-time). Then, $A(x)$ is the value of $A(\cdot)$ at a certain point in time and space $x \in D$. An error model of that variable was assumed: $A(x) = b(x) + V(x)$, where $b(x)$ is a deterministic function of x, and $A(x)$ and $V(x)$ are random variables (Heuvelink, 1998, p. 10). We assumed that $V(x)$ were the random errors that follow a normal distribution $N(\mu, \sigma^2)$, where μ is the mean (equal to zero) and σ^2 is the variance. Instead of assuming an error model for $\sigma(x)$, we utilized the hourly ensemble spread $\sigma(x)$ at each ERA5 grid, which presents the temporal and spatial variability of errors, the covariance of meteorological variables, and autocorrelation.

c. Monte Carlo method

The Monte Carlo (MC) method was applied to propagate random errors from the five meteorological variables to ET_o. Since each daily output ET_o map is computed from the five input maps $A_i(\cdot)$ $(i = 1, \dots, 5)$, the output ET_o maps are also random functions: $U(\cdot) = g(A_1(\cdot), \dots, A_5(\cdot))$, where $g(\cdot)$ is the FAO56 ET_o calculation. The MC method randomly samples N sets of realizations $a_{i,j}(x)(j = 1, \dots, N)$ from the distribution of $A_i(x)$ described by the daily ERA5 ensemble mean and standard deviation maps. For each set of realizations, an ET_o map $u_j(x) = g(a_{1,j}(x), \dots, a_{5,j}(x))$ was calculated. The error of ET_o was estimated by calculating the statistics of N outputs $u_j(x)$, including standard deviation normalized with mean values (σ_{norm}), and percentiles.

Although the MC method can generate the entire distribution of $u_j(x)$, the level of accuracy may be arbitrary depending on the method of the random generator and number of simulations. Since the efficiency of MC method is proportional to $\sqrt{N}$ (Heuvelink, 1998, p. 106), we experimented with N between 100 to 10,000 (10 times less efficient) to assess the impact of N on the results. Due to the computational burden of random sampling, we employed the Latin-Hypercube sampling (LHS) (Lee et al., 2014; Stein, 1987) to improve computational efficiency.

d. Taylor method

For implementing the Taylor method, we utilized the Python package *uncertainties* (Lebigot, 2017) to facilitate the calculation of derivatives. For that, we defined the daily ensemble mean and ensemble spread of ERA5 reanalysis products as the nominal mean and standard deviation of uncertain arrays (*uncertainties.unumpy.uarray*) for each meteorological variable. The *uarray* objects of each meteorological variable were used directly as inputs in each equation in the FAO56 ET_o calculation procedure to estimate the uncertain arrays of daily ET_o over the study area. The advantage of this stepwise

calculation is that we can analyse also the propagated errors of intermediate variables, such as Δ, γ, and R_{n}.

4.3 RESULTS AND DISCUSSION

4.3.1 On the uncertainty of meteorological forcings from climate reanalysis

a. Spatial and temporal pair-wise comparison

The scatterplots of meteorological variables show that all products are well correlated grid by grid (**Figure 4-3**). The correlation between paired datasets for air pressure, air temperature and vapour pressure are very high (0.97 to >0.99). Meanwhile, windspeed (0.85 to 0.94) and solar radiation (0.82 to 0.98) show slightly lower correlation, especially between GEOS5 and other two products. The data points in the scatterplot exhibit the largest spread for solar radiation, followed by windspeed, vapour pressure, air temperature, and air pressure. The highest correlation of the three pairs is between AgERA5 and ERA5 (**Figure 4-3** a, d, g, and j), which is expected since AgERA5 is derived from ERA5. However, the comparison between AgERA5 and ERA5 shows the largest mean bias of air temperature (-0.41 $^{\circ}$C) and windspeed (0.14 m·s^{-1}) among three pairs (**Figure 4-3** a & d).

The difference between datasets varies with location (**Figure 4-4)**. In case of air temperature, the largest difference between AgERA5 and ERA5 can be seen in areas with variable elevation (e.g., the Great Rift Valley and the Iranian highlands) (**Figure 4-4**a). This can be explained by the fact that an elevation correction using vertical lapse rate is applied to derive AgERA5 from ERA5 (Boogaard et al., 2023). GEOS5 air temperature is lower near the equator, while it is higher at mid-latitude when compared to both ERA5 and AgERA5 (**Figure 4-4** b & c).

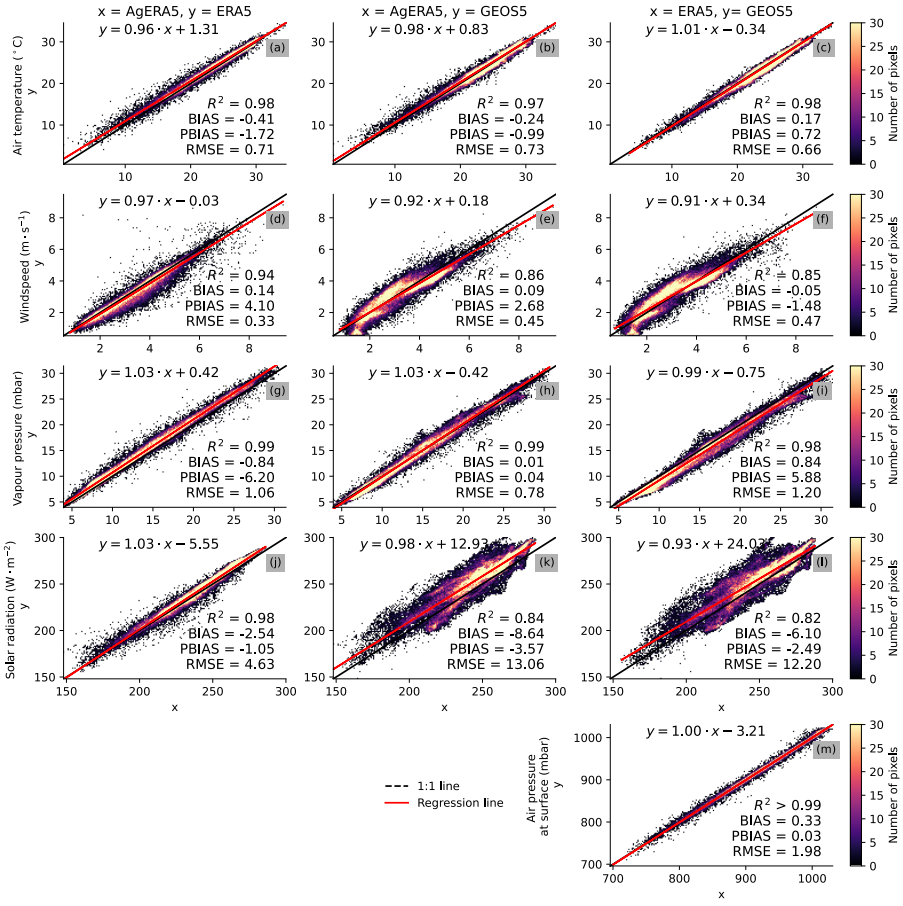


Figure 4-3. Comparison of mean annual air temperature, windspeed, vapour pressure, solar radiation, and air pressure at the surface from ERA5, AgERA5, and GEOS5 for the period 2018-2022. The scatterplots show the correlation between each pair of two datasets (column) for each meteorological variable (row). The performance metrics (R^2, RMSE, BIAS, PBIAS) and linear regression coefficients were calculated grid-wise by reshaping mean annual 2-dimensional arrays into 1-dimensional series, showing the spatial correlation between two products.

The difference between AgERA5 and ERA5 windspeed is close to zero in the majority of the grid cells, and slightly negative near the coastline and mountainous areas (**Figure 4-4** d). This is likely due to the topographical correction algorithm of AgERA5. Compared to GEOS5, both AgERA5 and ERA5 show greater windspeed below the Sahel, in Iran, and in the Southern Africa and lower windspeed in the Sahara and Congo rainforest, with differences up to 3 $m \cdot s^{-1}$ **(Figure 4-4** e & f). Moreover, the difference between the products over the Congo rainforest increases below the equator. Given that GEOS5 is prone to a larger error (Rienecker et al., 2008, p. 25) and ERA5 exhibits lower uncertainty in the Southern hemisphere (**Figure S1**, **Appendix D**), the larger difference between the two products is likely due to the errors in GEOS5.

In case of vapour pressure, the difference between products is greater in the middle range (between 20 and 25 mbar) (**Figure 4-3** g-i). The difference between AgERA5 and ERA5 is slightly positive over the majority of the area (lower than 2 mbar) and greater in Western Africa and the coastline of the Red Sea (**Figure 4-4**g). The difference in vapour pressure between GEOS5 and ERA5 is highest among the three pairs. GEOS5 vapour pressure is found to be higher than ERA5 and AgERA5 in the Sahel and Congo rainforest, but lower in the Ethiopian highlands (**Figure 4-4** h & i).

GEOS5 estimates higher solar radiation values than ERA5 and AgERA5 in the humid tropical parts of Western and Central Africa (10ºS – 20ºN, 0º – 30ºE), and lower over the rest of the continent (**Figure 4-4** k & l). Although AgERA5 is derived from ERA5, it shows slightly lower solar radiation values than ERA5 in most areas, except for the mountainous areas around Lake Victoria and Ethiopia (**Figure 4-4** j). This is likely due to the topographic correction used for AgERA5.

When comparing air pressure at the surface, ERA5 and GEOS5 exhibit greater similarity than for other variables (**Figure 4-3** m). Differences between the two products are primarily observed in regions with high elevation and coastlines (**Figure 4-4** m). However, these differences appear to be random and only present in a small number of grid cells, which does not affect the overall correlation. The difference between the two datasets for air pressure at the surface is much smaller than the difference in air pressure at sea level (**Figure S2**, **Appendix D**), which underlines the effect of topographical correction of air pressure for both datasets.

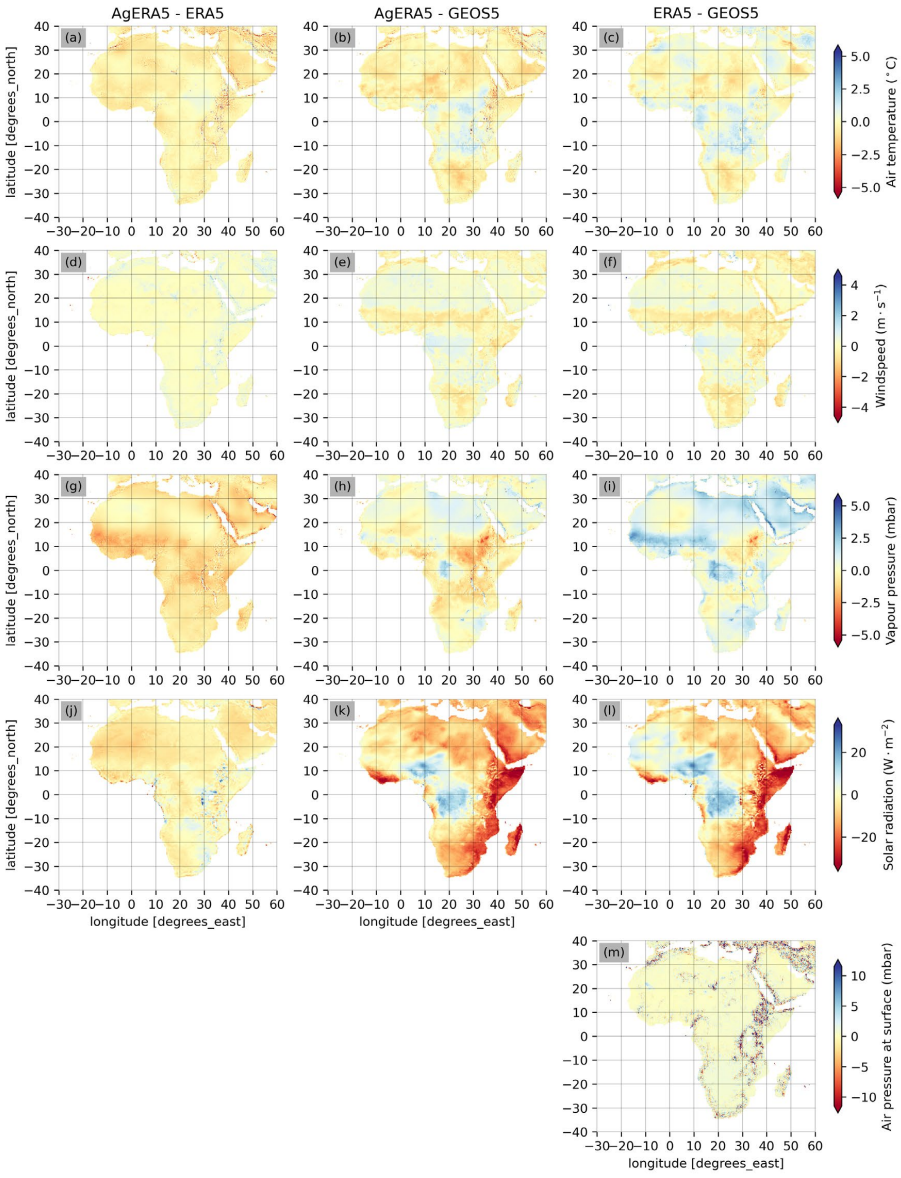


Figure 4-4. Mean annual difference of air temperature, windspeed, vapour pressure, solar radiation, and air pressure at the surface from ERA5, AgERA5, and GEOS5 for the period 2018-2022.

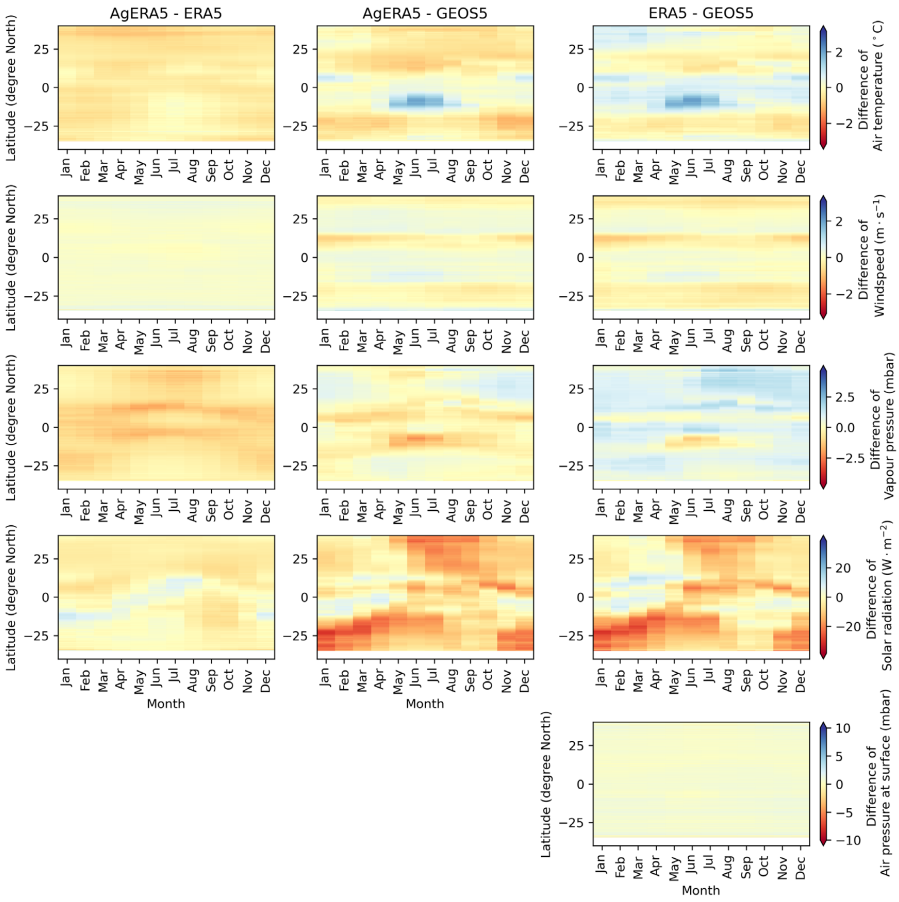


Figure 4-5. Hovmöller diagrams showing the monthly and latitudinal variation of mean difference in air temperature, windspeed, vapour pressure, solar radiation, and air pressure at the surface between GEOS5, AgERA5, and ERA5 for the period 2018-2022.

The differences between the datasets vary seasonally, especially for air temperature, vapour pressure, and solar radiation (**Figure 4-5**). The highest discrepancy in air temperature between the GEOS5 and ERA5 or AgERA5 products occurs between May and July (up to 2 °C). Moreover, during these months, the differences of air temperature near the equator are larger than the other months. In case of vapour pressure, the largest difference between GEOS5 and the other two products occurs between May and July below the equator and between July and September above the equator. The vapour pressure difference shows a seasonal pattern aligning with air temperature differences, but it has an inverse relationship – positive air temperature differences correspond to

negative vapour pressure differences, and vice versa. The difference in solar radiation between GEOS5 and the other two products also varies seasonally. From June to October, GEOS5 provides higher solar radiation values, whereas for the rest of the year, GEOS5 values of solar radiation are lower below the equator. In case of windspeed, the latitudinal pattern of differences between the products is mostly consistent throughout the year. The difference in air pressure at the surface is also very low (<5 mbar) and not seasonally variable.

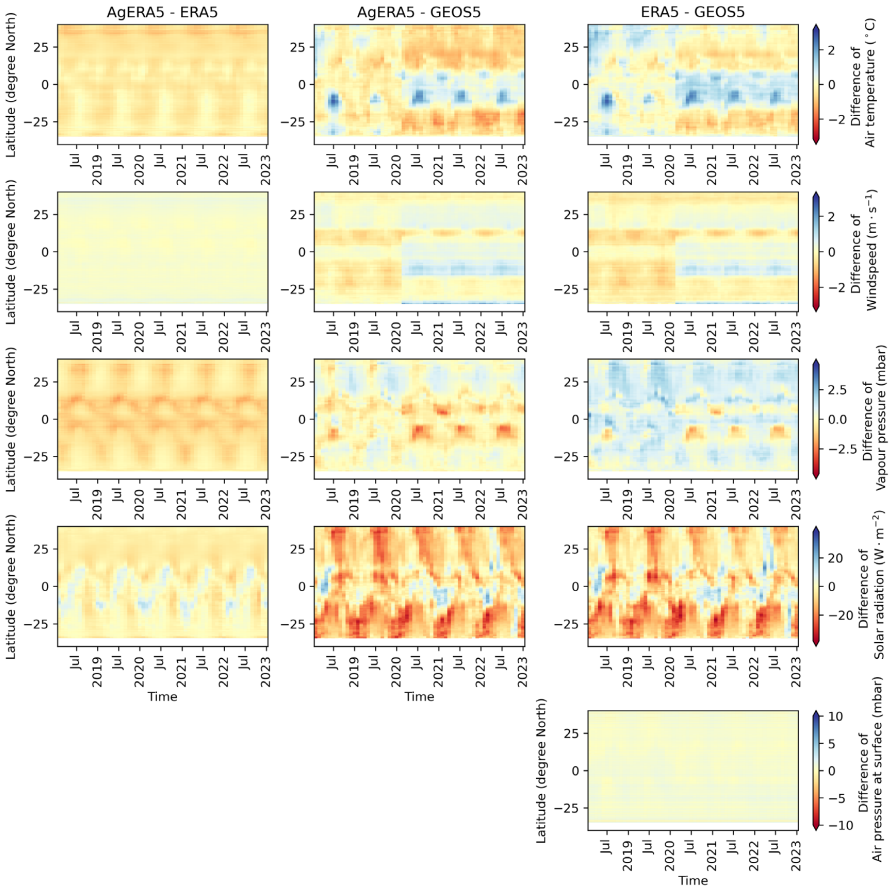


Figure 4-6. Hovmöller diagrams showing the temporal and latitudinal variation of mean difference in air temperature, windspeed, vapour pressure, solar radiation, and air pressure at the surface between GEOS5, AgERA5, and ERA5 for the period 2018-2022.

The time-series of air temperature, windspeed and vapour pressure difference between GEOS5 and the other two datasets show a shift in pattern from March 2020 (**Figure 4-6**).

This shift also coincides with an increase in the spatial difference between the data products, where GEOS5 air temperature becomes even lower near the equator and higher at mid-latitude compared to ERA5 and AgERA5. After the shift, the difference in windspeed between the data products becomes more positive in the tropics (between 20°N and 20°S), where the values of GEOS5 become smaller than ERA5 and AgERA5. In case of vapour pressure, this shift causes GEOS5 values to become lower than ERA5 and AgERA5 between the 10°S and 10°N. For other latitudes, the shift has less impact on vapour pressure differences.

According to the notices from Global Modelling and Assimilation Office (GMAO), the developer of GEOS5, the GEOS-FP model was updated on April 7, 2020. This upgrade introduced two changes that addressed (1) a bias in the heating tendency within the stratosphere and (2) errors in the diagnostics of convective mass flux (GMAO, 2020). Although it was not clearly mentioned how this upgrade affected each variable in the final data product, the shift observed in **Figure 4-6** suggests that these changes might have caused the temporal inconsistency of the reanalysis datasets. The Hovmöller plots for solar radiation and air pressure at the surface do not show any discernible shift between GEOS5 and the other two products. Since GEOS5 derives solar radiation data from the tavg1_2d_rad_Nx product and other variables originate from tavg1_2d_slv_Nx, the discrepancy suggests a systematic change within the GEOS5 model that specifically affects the tavg1_2d_slv_Nx product.

b. Comparison with in-situ data and performance metrics

In terms of nominal accuracy, the comparison with TAHMO stations shows that all datasets exhibit comparable performance (**Figure 4-7**). ERA5 and AgERA5 show slightly better performance metrics than GEOS5. AgERA5 performs slightly better than ERA5 for all variables, except for vapour pressure. Among the five variables, reanalysis datasets perform best for air pressure at surface and air temperature, showing bias close to zero and high correlation. Meanwhile, the nominal accuracy for windspeed, vapour pressure, and solar radiation is much lower. In general, all three datasets overestimate windspeed and solar radiation, and underestimate vapour pressure. In case of windspeed and solar radiation, the reanalysis datasets show very low correlation ($R^2 < 0.5$) at almost all stations. Some outlier stations have windspeed PBIAS up to –200 %, while solar radiation PBIAS is up to +100 %. For vapour pressure, the correlation between reanalysis and in-situ data is similar to air temperature ($R^2 > 0.6$), although PBIAS is generally from –5 % to –10 %. The overestimation of windspeed and solar radiation, and the underestimation of vapour pressure derived from dew-point temperature is also observed in validation studies of other reanalysis data in Iran (Radmanesh et al., 2023), Spain and Portugal (Martins et al., 2017) and China (Xu et al., 2024).

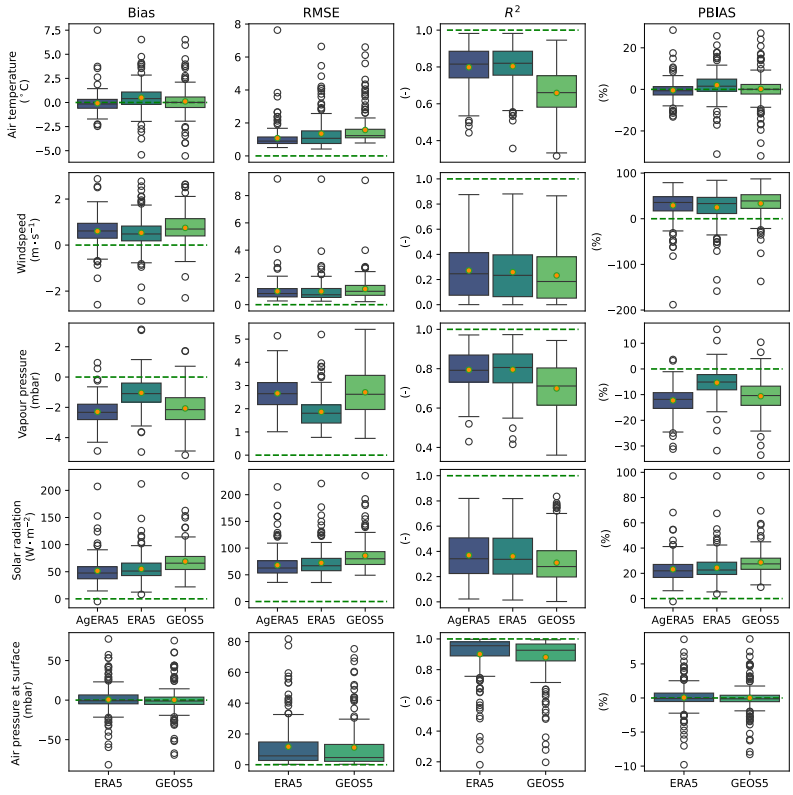


Figure 4-7. Performance metrics of meteorological variables from reanalysis datasets (indicated by different colours) compared to measurements at 174 TAHMO sites. The box-and-whiskers plots represent the 25th (Q1), 50th (median), 75th (Q3) percentiles of the probability distribution. The orange circles inside the box-and-whisker plots represent the mean value. The white circles represent outliers, which exceed the range [Q1 – 1.5 × IQR, Q3 + 1.5 × IQR], where IQR = Q3 – Q1 is the interquartile range.

Figure 4-8 shows the spatial variability of the average R^2 at the TAHMO sites, averaged over the three reanalysis datasets. The maps for RMSE, BIAS, and PBIAS are shown in **Figure S4-6** (**Appendix D**). Overall, the area with the best performance is Southern Africa, with less stations covering a smaller surface area. The other two areas have a lower performance, due to their complex topography (Eastern Africa) and proximity to the coast (Western Africa). These observations are aligned with findings from validation of reanalysis datasets in other regions (Pelosi et al., 2020; Pelosi and Chirico, 2021). In case of air temperature and vapour pressure, the average R^2 of the reanalysis datasets shows a decreasing pattern towards the coastline, which is not observed for solar radiation, air pressure, and windspeed.

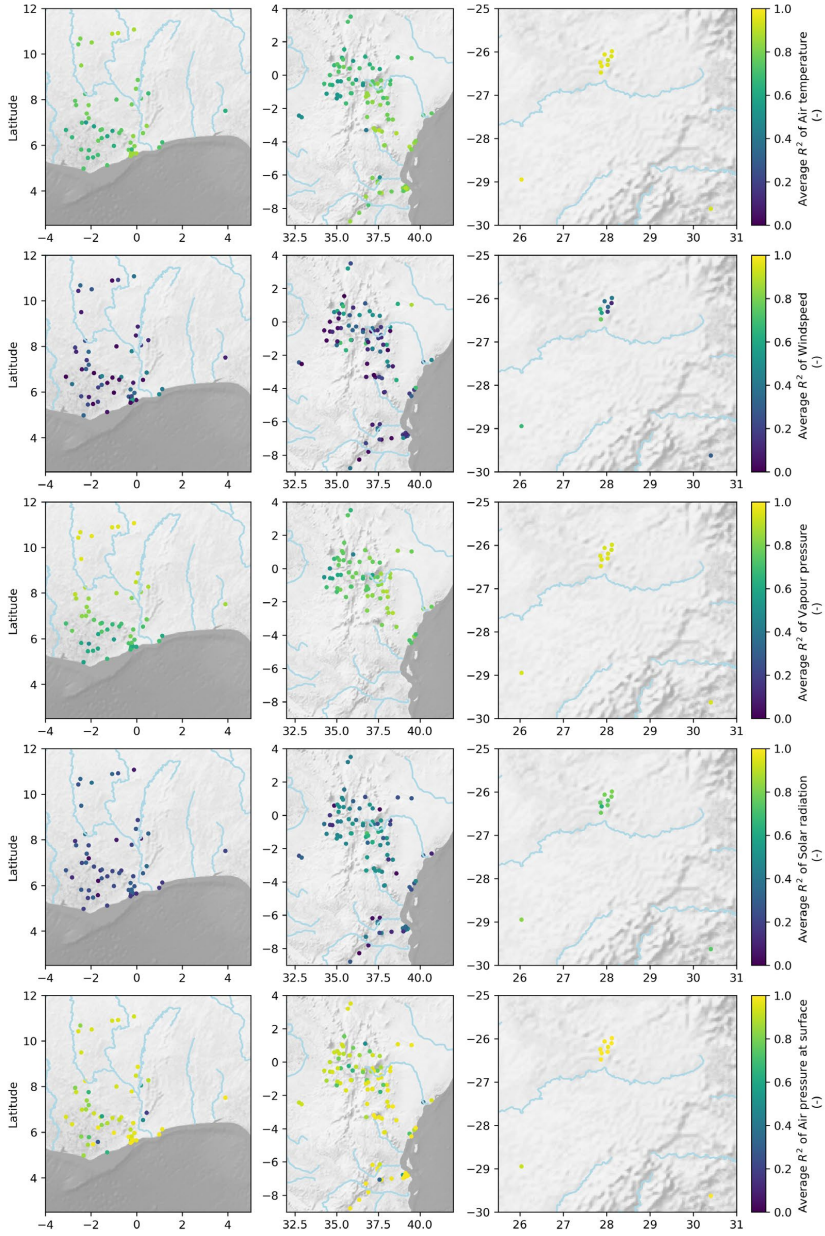


Figure 4-8. Spatial variability of the average R^2 computed at the TAHMO sites, averaged over the three reanalysis datasets, AgERA5, ERA5, and GEOS5. Base map: Natural Earth Shaded Relief and Rivers.

c. Comparing the uncertainty of reanalysis from multiple assessments

The uncertainties associated with the meteorological variables from the reanalysis datasets were assessed by different approaches. The uncertainty estimates from the ERA5 ensemble represent the random errors within ECMWF's NWP model and its data assimilation system. Meanwhile, the discrepancy between different reanalysis datasets indicates the relative errors between the different NWP models. The performance metrics calculated using in-situ measurements as reference indicate the nominal accuracy of the reanalysis data. Here, we compare and discuss the findings from the three approaches to identify commonalities.

The uncertainty estimates for the ERA5 reanalysis show that all variables have the largest ensemble spread between May and September, especially around 20°N (**Figure S1**, **Appendix D**). The seasonal variability of the discrepancy between reanalysis datasets is also higher between May and September, but only for air temperature and between the equator and 20°S (**Figure 4-5**). While ERA5 has a larger ensemble spread in the Northern hemisphere than in the Southern hemisphere in general, the differences between the reanalysis products do not show the same pattern. This spatial variability between products is more aligned with the reported errors from the GEOS5 NWP initial estimates and assimilated observations for wind and humidity, which is larger in the Southern hemisphere than in the Northern hemisphere (Rienecker et al., 2008, p. 25). Unfortunately, the GEOS5 data products do not include uncertainty estimates or quality indicators for quantitative analysis.

The spatial variability of the ERA5 ensemble spread is different for each variable (**Figure S1**, **Appendix D**). Air temperature has the largest yearly average uncertainty estimates in the tropics. Dew point temperature and wind component speed are most uncertain in the Sahara and semi-arid Southern Africa and air pressure in Central Africa (**Figure S1**, **Appendix D**). The discrepancy between the datasets also tends to be higher in the tropics for air temperature (**Figure 4-4**). However, for vapour pressure and windspeed, the discrepancy is higher in the Congo rainforest and below the Sahel.

The spatial difference between ERA5 and AgERA5 is mainly observed in mountainous and coastal areas. This is likely because AgERA5 is derived by calibrating ERA5 data with ECMWF's operational high-resolution atmospheric model (HRES). According to Boogaard et al. (2023), the greatest enhancements of AgERA5 compared to HRES are observed at grid points situated in mountainous regions or along coasts and lakes, improved particularly for the variables of temperature, humidity, and windspeed. However, AgERA5 data is still limited by the accuracy of the HRES operational model, since it is assumed that HRES represents actual conditions the best. The performance metrics at coastal stations (**Figure S7**, **Appendix D**) show AgERA5 performs better for windspeed, but worse for air temperature, and the same for solar radiation and vapour pressure in comparison with all stations (**Figure 4-7**). In case of stations at elevations

higher than 1000 m (**Figure S8**, **Appendix D**), the performance metrics of AgERA5 are only marginally better than ERA5 for all stations.

4.3.2 Impact of meteorological input uncertainties from reanalysis data

a. Propagation of reanalysis error in reference evapotranspiration calculation

The ET_o uncertainty estimated using the Taylor method shows the exact same spatial and temporal pattern compared to the MC method (**Figure 4-9**). However, it overestimates the normalized standard deviation up to about 20 % in the Sahara, especially in the summer months (**Figure 4-9**). In areas with a high uncertainty of ET_o, such as the Western coastline, the Taylor method shows better agreement with the MC method ($\frac{\sigma_{\mathrm{norm,MC}}}{\sigma_{\mathrm{norm,T}}} \approx 1$).

The MC simulation result with $N = 500$ realizations were found to be optimal, as they provided similar outcomes to simulations with larger sample size, while significantly reducing computational time, which is slightly more than using the Taylor method (**Figure S9**, **Appendix D**).

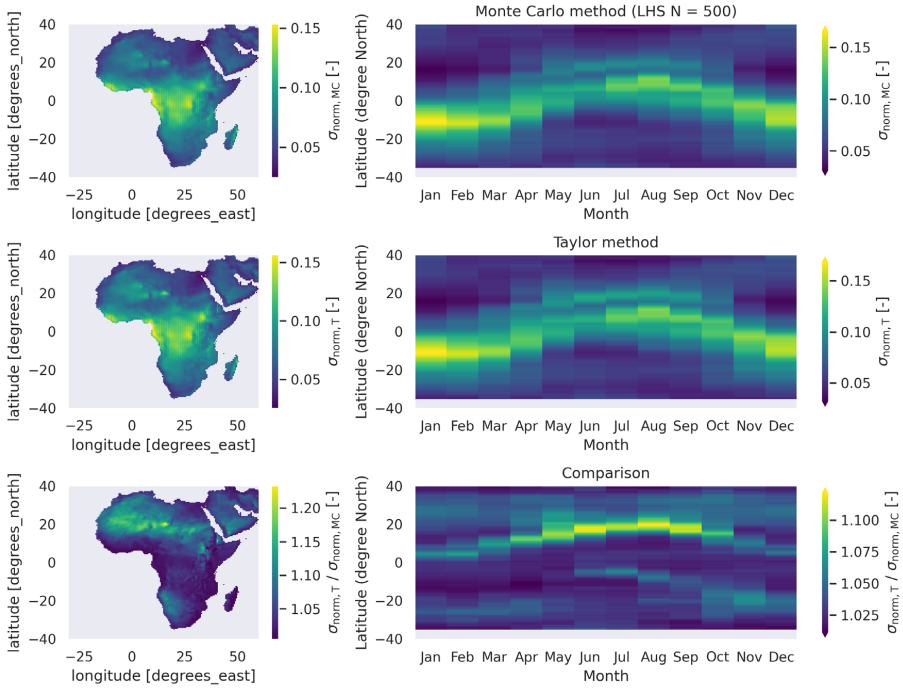


Figure 4-9. Yearly average and monthly-latitudinal variation of normalized standard deviation σ_{norm} of daily reference evapotranspiration estimated by Monte Carlo method (first row), Taylor method (second row), and their ratio (third row).

The advantage of the MC method is that a probability density function (PDF) and confidence interval can be estimated from the sampled realizations. Daily ET_o was calculated for 500 MC simulations of ERA5 meteorological inputs by Latin-Hypercube sampling (LHS). From the ensemble of ET_o, the standard deviation (σ) and 90 % confidence interval of the ET_o values were estimated for each grid cell. **Figure 4-10** shows the spatial and temporal variation of the estimated ET_o uncertainty. The Sahara shows the highest standard deviation [mm·d^{-1}] but a relatively lower normalized standard deviation (σ_{norm}) compared to other regions. This is likely due to higher ET_o values in the Sahara (**Figure 4-10** Area [1]). The tropics shows a high σ_{norm}, especially at the Western coastline (**Figure 4-10** Area [3]) and the Congo rainforest (**Figure 4-10** Area [4]). The time-series of 90 % confidence intervals shows that this high uncertainty is consistent throughout the years for these regions (**Figure 4-10**). Meanwhile, for other regions, the estimated uncertainty varies with season (**Figure 4-10** Area [1] and Area [2]).

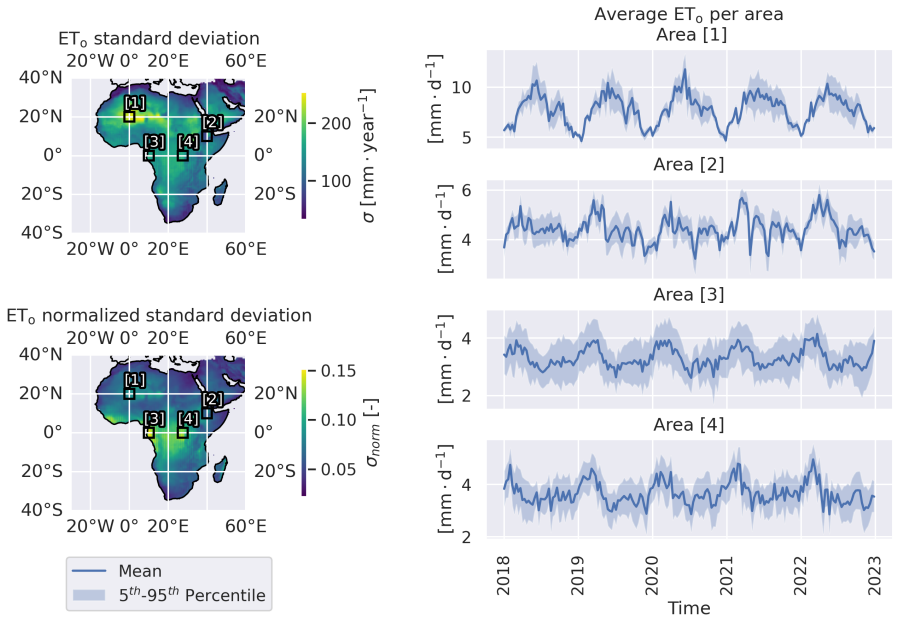


Figure 4-10. Standard deviation of daily reference evapotranspiration (ET_o) resulted from 500 Monte Carlo simulations, averaged for 5 years (top left) and normalized with daily mean ET_o (bottom left), and the dekadal ET_o in four specific areas indicated in the maps, with 90 % confidence interval (right column).

Since the month of June is where the MC method and the Taylor method show the largest difference in estimated σ_{norm} (**Figure 4-9**), we compared the PDF of the MC samples to a normal distribution $\mathcal{N}(\mu, \sigma^2)$ characterized by the μ and σ obtained from the Taylor

method on one day of June (**Figure 4-11**). In general, the PDF obtained from the MC method fits quite well with the normal distribution obtained from the Taylor method, even when the ET_o model involves a non-linear combination of input variables. At locations where the normal distribution is rejected by the Kolmogorov-Smirnov test, the *Q-Q* plot and histogram show that the largest discrepancy is observed in the tail quantiles, where ET_o is extremely low (**Figure 4-11** location [1] and [3]). This suggests that while the Taylor method provides a reasonable estimate of the PDF for the central part of ET_o values range, it overestimates the occurrence of extremely low ET_o values. Therefore, in cases where accurate estimation of tail quantiles is critical, the MC method is preferred over the Taylor method. Since the MC results are not in analytical form, it is not feasible to analyse the effect of changing the input error on the output without running the entire simulation again (Heuvelink, 1998, pp. 44–45). For this reason, where the Taylor method matches well with the MC result in the central quantiles, it can be applied to update uncertainty estimates when more information about input errors becomes available.

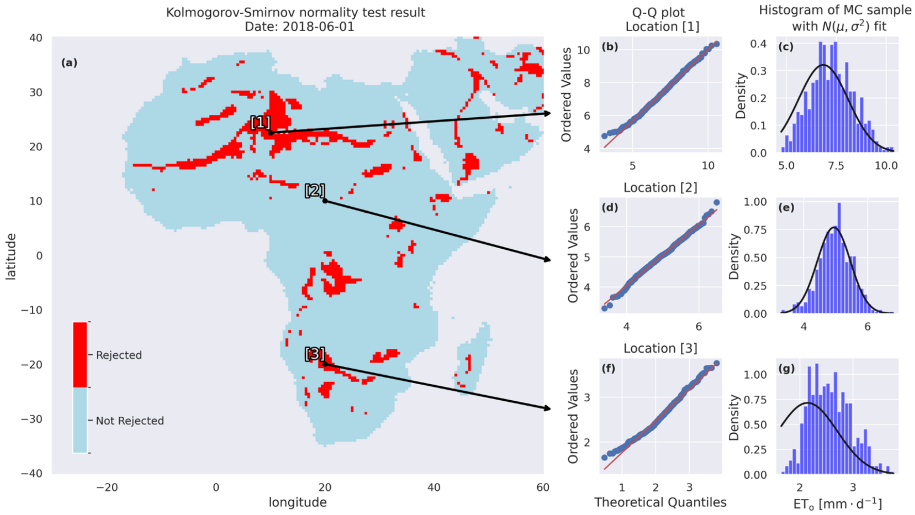


Figure 4-11. Kolmogorov-Smirnov normality test comparing Monte Carlo sample of calculated ET_o ($N = 500$) to the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with μ and σ obtained from the Taylor method (significance level of $\alpha = 0.05$). The Q-Q plot and histogram at location [1] and [3] (normal distribution is rejected) and location [2] (not rejected) are on the right panel.

b. Impact of uncertainties in meteorological data from reanalysis

The uncertainties associated with meteorological variables from reanalysis datasets can affect the accuracy and precision of ET_o estimation. The propagation of ERA5 errors in ET_o estimates shows that random errors are particularly high in the tropics and the Sahel

(**Figure 4-9**). The estimated uncertainty in ET_o also varies seasonally, following similar patterns to the errors in input variables (**Figure S1**, **Appendix D**). This indicates that non-stationary uncertainty in the reanalysis data also leads to non-stationary uncertainty in ET_o calculations.

Since ET_o estimates are used to calculate crop water requirement, high uncertainty in ET_o can affect irrigation recommendations. Given that ET_a in arid and semi-arid climates is predicted to be primarily controlled by water supply (rainfall or irrigation) (K. Zhang et al., 2016), effective irrigation management is essential to limit agricultural water consumption. Therefore, precise and accurate ET_o estimates are crucial in these regions. The estimated uncertainty of ET_o propagated from input reanalysis data is generally less than 15 % of ET_o, based on σ_{norm} (**Figure 4-9**). In terms of mm·year^{-1}, this uncertainty is highest in the Sahel (**Figure 4-10**). Therefore, application of ET_o for irrigation in this region could be improved considerably with more accurate input meteorological data.

Furthermore, the differences between reanalysis datasets indicate that the choice of input dataset may influence the estimated ET_o, particularly for solar radiation, vapour pressure, and wind speed (**Figure 4-3**). Since the nominal accuracy of ERA5 and AgERA5 at TAHMO sites is generally better than that of GEOS5 (**Figure 4-7**), these datasets are more recommended for field-scale estimation. If GEOS5 is needed for high-latency applications, the error propagation methods employed in this study can be applied to GEOS5, assuming its uncertainty quantification is feasible. This will provide more robust results with a clearer understanding of the confidence or reliability of the estimates.

4.3.3 Limitations of ex-ante uncertainty assessment methods

The multiple uncertainty assessment methods used in this study have certain limitations. Firstly, the inter-comparison approach requires all datasets to have equivalent spatial resolution (grid size), temporal resolution, and physical quantity. We resampled higher-resolution dataset to coarser resolution to avoid bias caused by the loss of detail in coarser-resolution dataset. However, it is difficult to separate the impact of spatial resampling from the differences between datasets. For example, AgERA5 data is derived from ERA5 by applying both spatial downscaling and bias correction for topographical condition and coastal areas (Boogaard et al., 2024). As a result, the differences between AgERA5 and ERA5 represent both these corrections and the errors introduced by downscaling and subsequently upscaling processes.

Secondly, comparing reanalysis data with in-situ measurements only provides nominal accuracy at a few clusters of TAHMO sites, mainly in tropical savannah climate. These clusters do not cover the tropics and the Sahara, which are the areas with the largest uncertainties indicated by spatio-temporal comparison and ERA5 ensemble spread. Furthermore, the accuracy of in-situ measurements is also a factor influencing the resulting performance metrics. For example, in-situ windspeed measurement at 2 m can

be affected by local terrain effects at some sites (Pelosi and Chirico, 2021). For equivalent comparison, the reanalysis windspeed at 10 m was used to derive windspeed at 2 m using a logarithmic windspeed profile, however, this cannot account for local terrain effects.

Thirdly, the error propagation method requires an assumption for the error models of input variables. In this study, we assumed that the errors in meteorological variables from reanalysis data follow a normal distribution. For ERA5, we characterized the error distribution using the mean and standard deviation provided by its ensemble uncertainty quantification. However, in the case of AgERA5 and GEOS5, where such uncertainty quantification is not available, making assumptions about the mean and standard deviation of errors would be inherently incomplete, as they vary for every grid cell and every time step. Additionally, none of these approaches account for systematic error and uncertainty due to different NWP models.

The challenge of uncertainty propagation is that one needs to define the prior uncertainty of inputs, the parameter space, and the models. Obtaining that information at every pixel is challenging. We tested uncertainty propagation for one day (with the largest uncertainty) to investigate the magnitude and spatial distribution of uncertainty. In fact, the most challenging part is estimating prior uncertainty realistically. Even when all prior uncertainties are known, significant computational power is needed to estimate uncertainty at every pixel. This is also a challenge: the more Monte Carlo simulations, the more time and computational power required. Therefore, if we want to use uncertainty propagation to quantify uncertainty at every pixel and every date in the spatio-temporal datasets, the computational demand will be immense. However, if we accept that some uncertainty is irreducible, we can move away from increasingly complex and detailed models and instead use uncertainty propagation as a tool to assess the validity of claims based on data.

Lastly, it is important to note that the methodological uncertainties associated with the FAO56 ET_o equation were not evaluated in this study. Given that the FAO56 definition of ET_o is based on idealized conditions, ambiguity regarding the consideration of local advection effects may introduce biases in FAO56 ET_o estimates when applied to arid and semiarid regions with non-ideal or smaller fields (de Bruin et al., 2016; Pereira et al., 2021). Therefore, updates on the FAO56 ET_o equation still need to be considered for its use in estimating water demand.

4.4 CONCLUSIONS

This chapter evaluates the uncertainty of meteorological data inputs for FAO56 ET_o calculation from GEOS5, ERA5, and AgERA5 reanalysis data products through spatiotemporal inter-comparison (between-product uncertainty), comparison with in-situ measurements (nominal accuracy), and ensemble spread (within-product uncertainty).

The spatio-temporal inter-comparison of all data products shows that the differences between products are non-stationary. The major differences between climatic input datasets are in Central and Southern Africa, and Southwest Asia. This uncertainty between reanalysis datasets is due to the model uncertainty of the employed numerical weather prediction model and estimated errors of observations in data assimilation systems. All reanalysis datasets predict air temperature and air pressure well, but overestimate windspeed and solar radiation and underestimate vapour pressure at the reference sites in TAHMO network. Comparison with in-situ data shows that all reanalysis datasets have comparable performance, but ERA5 and AgERA5 perform slightly better. Although having better latency, GEOS5 has lower nominal accuracy and some temporal inconsistency due to changes in the data assimilation system. Therefore, near-real time applications that depend on the GEOS5 dataset are subject to more errors and not recommended to be used for trend analysis.

The study also analyses the error propagation in the FAO56 ET_o calculation using reanalysis as meteorological forcing. This is part of the ex-ante uncertainty assessment in the development of WaPOR version 3. The error propagation results show that the uncertainty in ET_o estimates propagated from the estimated uncertainty in the ERA5 reanalysis dataset is consistently higher in the tropics. The Taylor method showed a consistent spatial and temporal pattern of uncertainty and adequate accuracy compared to the Monte Carlo method. Since every uncertainty assessment method has its limitations, applying multiple approaches and comparing their results could help identify the limitation in reanalysis data and better inform the application of reanalysis data, especially in data-scarce regions like Africa and Southwest Asia.

5

EX-POST UNCERTAINTY ASSESSMENT

This chapter is based on:

Tran, B. N., Seyoum, S., van der Kwast, J., Jewitt, G., Uijlenhoet, R., and Mul, M.L.: Assessing uncertainty in WaPOR version 3 global evapotranspiration data: insights from in-situ validation and triple collocation. (in preparation)

5.1 INTRODUCTION

Ex-post uncertainty assessments are instrumental in understanding the reliability of satellite-based data products. These assessments often involve comparison with a reference or other equivalent data products (Chapter 2) to quantify uncertainty in terms of performance metrics (Chapter 3). The most widely used ex-post uncertainty assessment method for satellite-based ET_a estimation is validation against an in-situ reference dataset, such as Eddy Covariance (EC) flux measurements **(Figure 2-5**). This method relies on the existing international network of ET_a measurement sites. The major challenges for these in-situ measurements are insufficient spatial coverage and energy balance closure (Yi et al., 2024). Due to the limited spatial coverage of in-situ measurements, the validation of satellite-based ET_a estimation in many regions often relies on residuals from the basin water balance as a reference (**Figure 2-10**). While this method is useful for evaluating ET_a at the basin scale, it is affected by uncertainties in other components of the water balance and does not reveal the spatial and temporal pattern of uncertainty within the basin.

These challenges are also observed in the quality assessments of the WaPOR-ET data products. Previous validation of the WaPOR-ET v2 data products was conducted at multiple scales, from pixel to river basin, across the African continent. Due to the sparse distribution of EC sites and other ground-based measurements in Africa, the evaluation of WaPOR v2 relied on in-situ validation against 14 EC sites and basin water balance (Blatchford et al., 2020b). As in-situ validation alone is insufficient for a comprehensive assessment of the WaPOR-ET product across continental Africa, Blatchford et al. (2020b) further evaluated the consistency between WaPOR data levels and their plausibility based on vegetation and climatic conditions. This provides qualitative information about the data quality and physical plausibility, but not uncertainty estimates.

The evaluation of the WaPOR-ET v2 data product provided important recommendations for the improvement in the WaPOR-ET v3 data product, which was released in 2023 with global coverage (Table II-1). This version incorporates several improvements, including the downscaling of thermal data and the use of higher-resolution meteorological and other input data. To continue the uncertainty assessment of the WaPOR-ET data product, in-situ validation needs to be extended to other regions and updated with more reference datasets now available within the spatial coverage of the WaPOR-ET v3 data product. Since the WaPOR-ET v3 L1 product, hereinafter WaPORv3L1, covers a global extent, it presents more opportunities for validation with EC data from international and regional EC networks (e.g., FLUXNET, ICOS, OzFlux, AmeriFlux).

Other approaches for ex-post uncertainty assessment have been developed and applied in many studies, such as intercomparison, Triple Collocation (TC), Three-Cornered Hat (TCH), physical consistency, and ensemble of estimates (Section 2.3). Among these

methods, the TC and TCH methods are potentially useful for ex-post uncertainty assessment because of its ability to estimate uncertainty (error variances) in three or more datasets, without assuming one "true" reference. This is especially valuable for ET_a estimation because, while numerous methods exist including satellite-based models , high-quality reference data are available only in limited regions. Moreover, several studies have used TCH and TC methods to quantify uncertainty spatially (Khan et al., 2018; Miralles et al., 2011a). By spatially and temporally collocating three datasets, these methods can quantify uncertainty at every pixel in the dataset. However, only a few studies have utilized the TC (Barraza Bernadas et al., 2018a; Khan et al., 2018; Kibria et al., 2021; Miralles et al., 2011a) and TCH (He et al., 2020; Long et al., 2014; Xu et al., 2019; Zhu et al., 2022) methods to evaluate uncertainty in satellite-based ET_a estimates (Chapter 2). Given the advantages but underutilization of the TCH and TC methods, its reliability for uncertainty assessment needs to be evaluated and compared with the conventional in-situ validation approach.

This chapter aims to explore the applicability of the TC and TCH methods for ex-post uncertainty assessment, using the WaPORv3L1 data product as an example. We compared its performance metrics with the results from in-situ validation. Given that the choice of the ET_a data can influence the outcomes of the TC and TCH methods (Shao et al., 2022; Wu et al., 2019a), we also evaluated the impact of data selection on their outputs using multiple satellite-based ET_a data products. This chapter will address the following research questions:

- What is the compound uncertainty of the WaPOR-ET data product based on in-situ validation?
- How well can the relative uncertainty of the WaPOR-ET data product predict the compound uncertainty from in-situ validation?

The remainder of this chapter is structured as follows: First, a brief review of the TC and TCH methods is provided, leading to the selection of a methodological procedure. Second, the materials and methods used for compound uncertainty and relative uncertainty are described. The results of this analysis are presented in Section 5.4, followed by a discussion of the implications and recommendations in Section 5.5. Finally, the chapter concludes with a summary of the key findings.

5.2 TRIPLE COLLOCATION AND THREE-CORNERED HAT APPROACH

The TC and TCH methods were independently developed (Sjoberg et al., 2021). Both methods estimate error variances of three collocated datasets, assuming they estimate the same physical quantity at approximately same time and location (Premoli and Tavella, 1993; Stoffelen, 1998). The key difference between the two methods is that the TCH

method assumes that the three datasets have been calibrated with only additive bias and random error (Sjoberg et al., 2021):

$$X_i = a_i + \Theta + \epsilon_i \tag{5.1}$$

where X_i (for $i \in \{1,2,3\}$) represents the i-th dataset; Θ is the unknown true value; a_i is the additive bias; and ϵ_i is the random error from the i-th dataset.

Meanwhile, the TC method assumes uncalibrated datasets with calibration coefficient or multiplicative bias (b_i) (Stoffelen, 1998):

$$X_i = a_i + b_i\Theta + \epsilon_i \tag{5.2}$$

where b_i is the multiplicative bias or calibration coefficient, which is different from one for uncalibrated dataset.

Both TC and TCH assume that linear calibration of the datasets is sufficient. For this linear system to be solvable, both methods rely on two assumptions regarding the random errors ϵ_i. First, it is assumed that the random errors have a zero mean ($\overline{\epsilon_i} = 0$), which means that the estimated error component is purely random and not a systematic error. Second, the random errors of all datasets are uncorrelated to each other (zero error cross-correlation), which implies the error covariance between two different datasets is zero ($Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$). Based on the variances of pairwise differences between datasets, the TCH method derives the variance of random errors in each dataset as (Premoli and Tavella, 1993; Sjoberg et al., 2021):

$$\sigma_{\epsilon_i}^2 = \frac{1}{2}\left(\sigma_{x_i-x_j}^2 + \sigma_{x_i-x_k}^2 - \sigma_{x_j-x_k}^2\right) \tag{5.3}$$

where ${\sigma_{\epsilon_i}}^2$ is the random error variance of the i-th dataset; $\sigma_{x_i-x_j}^2$ is the variance of the difference between the i-th and j-th datasets; $\sigma_{x_i-x_k}^2$ is the variance of the difference between the i-th and k-th datasets; $\sigma_{x_j-x_k}^2$ is the variance of the difference between the j-th and k-th datasets.

In addition, the TC method assumes the random errors to be uncorrelated with the true value (error orthogonality), which means the covariance of the errors with the true values is zero ($Cov(\epsilon_i, \Theta) = 0$). By substituting the linear error model (5.2) in covariance equations and applying the three assumptions, the TC method leads to the reduced equations (Gruber et al., 2016b; Stoffelen, 1998):

$$Q_{ij} = Cov(X_i, X_j) = \begin{cases} b_i^2\sigma_\Theta^2 + \sigma_{\epsilon_i}^2, & \text{for } i = j \\ b_i b_j \sigma_\Theta^2, & \text{for } i \neq j \end{cases} \tag{5.4}$$

where $i, j \in \{1,2,3\}$; Q_{ij} is the covariance of the i-th and j-th datasets; σ_θ^2 is the variance of the true values; and $\sigma_{\epsilon_i}^2$ is the random error variance of the i-th dataset.

From a linear system of six equations based on (5.5), we can solve for six unknowns ($\sigma_{\epsilon_i}^2$ and $b_i^2\sigma_\theta^2$ for $i \in \{1,2,3\}$). Given the three variances (Q_{ii} for $i \in \{1,2,3\}$) and covariances (Q_{ij} for $i, j \in \{1,2,3\}$ and $i \neq j$) are known from sufficient collocated data points of the three datasets, the TC method estimates the random error variances ($\sigma_{\epsilon_i}^2$) and the signal variances ($b_i^2\sigma_\theta^2$) as (Gruber et al., 2016b; Stoffelen, 1998):

$$b_i^2\sigma_\theta^2 = \frac{Q_{ij}Q_{ik}}{Q_{jk}} \tag{5.5}$$

$$\sigma_{\epsilon_i}^2 = Q_{ii} - \frac{Q_{ij}Q_{ik}}{Q_{jk}}$$

where $i, j, k \in \{1,2,3\}$ and $i \neq j \neq k$.

The TCH method yields in the same estimates as the TC method when the three datasets are similarly (un)calibrated to each other (similar b_i), except it does not assume error orthogonality (Sjoberg et al., 2021). The impact of violating the error orthogonality assumption is generally negligible for error variance estimation (ϵ_i), but considerable for the calibration coefficient (b_i) (Gruber et al., 2016b; Vogelzang et al., 2022). Meanwhile, violating the assumption of calibrated datasets has a large impact on the estimation of the error variances when three datasets are uncalibrated and present largely different multiplicative biases (Sjoberg et al., 2021). Hence, for geophysical variables like ET_a, for which satellite-based models are usually not extensively calibrated, the TC method is preferred to the TCH method.

The key challenge of both TC and TCH is that the assumption of zero error cross-correlation is not easily upheld or verified in practice. For satellite-based ET_a estimation, many datasets are derived from the same sensors or satellite platforms (e.g., MODIS, Landsat), utilize the same sources of meteorological input (e.g., reanalysis, gridded meteorological dataset), or employ the same modelling scheme (e.g., PM, SEB) or data processing (e.g., gap-filling). Studies that applied the TC method for satellite-based ET_a data often used a combination with datasets from in-situ measurements, reanalysis, or land surface models (Barraza Bernadas et al., 2018b; Khan et al., 2018; Miralles et al., 2010). These studies assumed that different sources of ET_a estimation have zero error cross-correlation. This assumption is typically based on the premise that the datasets originate from different sources are independent. However, non-zero error cross-correlation has been observed even in datasets where error independence is commonly assumed (Gruber et al., 2016a; Yilmaz and Crow, 2014). For example, He et al. (2025) tested the reliability of TCH-estimated relative uncertainty and found it is sensitive to the

choice of dataset, even in combinations of three ET_a datasets from independent models and input data sources.

Using datasets of largely different spatial support might cause representativeness errors in the estimation of error variances (Gruber et al., 2016b). Representativeness errors occur when the three datasets represent the same phenomena, but not the same quantity due to different spatial and temporal supports (Vogelzang and Stoffelen, 2022). For example, satellite-based data usually has spatial resolution higher or close to the spatial support of in-situ measurements, (e.g., 100 to 1000 m for EC flux towers), but lower than that of land surface models and reanalysis products (10 km to 100 km). For such a combination, the TC method will penalize both in-situ and coarse-resolution datasets, but the uncertainty estimation of the intermediate-resolution dataset will remain unbiased (Gruber et al., 2016b). This combination, however, still requires the availability of in-situ data, similar to the conventional in-situ validation approach. Wu et al. (2019) noted that the TC technique is less affected by representativeness errors than in-situ validation, which is hampered by the scale mismatch between satellite-based and ground-based estimates. In case a combination of satellite-based ET_a data with two coarse-resolution land surface models or reanalysis datasets was used for TC, the higher-resolution satellite-based data will be disproportionately penalized (Gruber et al., 2016b).

There have been also several developments based on the classical TC method by Stoffelen (1998), primarily focusing on extending its applicability to a number of datasets other than three (Pan et al., 2015; Su et al., 2014; Vogel and Ménard, 2023; Vogelzang and Stoffelen, 2022, 2021), addressing non-zero error cross-correlations (González-Gambau et al., 2020; Gruber et al., 2016a), and deriving correlation coefficients (McColl et al., 2014). Notably, the Extended Triple Collocation (ETC) method proposed by McColl et al. (2014) allows the estimation of the correlation coefficient with the unknown true signal only from the covariance of three datasets:

$$\rho^2_{\Theta, X_i} = \frac{Q_{ij} Q_{ik}}{Q_{ii} Q_{jk}} \tag{5.6}$$

where ρ^2_{Θ, X_i} is the squared correlation coefficient (coefficient of determination) of the i-th dataset (X_i) with the unknown true signal (Θ), for $i, j, k \in \{1,2,3\}$ and $i \neq j \neq k$.

This method provides more insights from the same combination of three datasets without altering the core assumptions or increasing the computational burden since ρ^2_{Θ, X_i} is estimated from the same covariances as $\sigma^2_{\epsilon_i}$. Similar to TCH and TC, the estimation of random error variance and correlation coefficient from the ETC method also depends on the choice of three datasets and only represents the relative uncertainty of the selected dataset in reference to one another (Wu et al., 2019a). Therefore, caution should be taken when ranking datasets by uncertainty estimated using any TCH and TC methods.

5.3 MATERIALS AND METHODS

The uncertainty metrics were estimated using the conventional in-situ validation with Eddy Covariance (EC) flux measurements, which is the most used method in literature (Chapter 2). The conventional in-situ validation metrics indicate the compound uncertainty of the WaPORv3L1 data product. The ETC procedure was selected to estimate its relative uncertainties, indicated by the random error variance and correlation coefficient in comparison with other datasets. To address the second research question, the results of the two approaches were compared and analysed to understand the applicability of the TC method as an alternative for in-situ validation.

5.3.1 Eddy Covariance flux dataset

Eddy Covariance flux data was used as reference for in-situ validation. The most comprehensive EC flux dataset is FLUXNET2015 with 212 sites (Pastorello et al., 2020), including gap-filled time-series, estimated uncertainties, and measurement metadata. However, this dataset covers only the period until 2014. Therefore, in order to evaluate the WaPORv3L1 data product (available from 2018), we acquired FLUXNET data products available from regional networks, including AmeriFlux, ICOS (Integrated Carbon Observation System), and OzFlux (Table 5-1).

Table 5-1. Summary of FLUXNET data products collected from regional networks.

Dataset	Sources	Number of sites
AmeriFlux	https://ameriflux.lbl.gov/ (last access: 09/10/2024)	131
ICOS	Warm Winter 2020 Team and ICOS Ecosystem Thematic Centre (2022), and Drought 2018 Team and ICOS Ecosystem Thematic Centre (2020)	77
OzFlux	https://data.ozflux.org.au/ (Isaac et al., 2017) (last access: 04/11/2024)	14

The FLUXNET data products are EC flux measurements that have been post-processed using ONEFlux (Open Network-Enabled Flux processing pipeline), the standard data processing procedure used for the FLUXNET2015. The ONEFlux pipeline includes the quality control of half-hourly turbulent fluxes and the gap-filling of missing and quality-flagged hourly data, energy balance closure correction, and aggregation to daily intervals (Pastorello et al., 2020). The post-processed data product using ONEFlux was used so that the data quality of all flux sites is controlled following a standardized flux data post-processing procedure that is widely accepted within the flux measurement community.

For the study period from 2018 to 2022, FLUXNET data products from **222** EC flux sites were used (**Figure 5-1**).

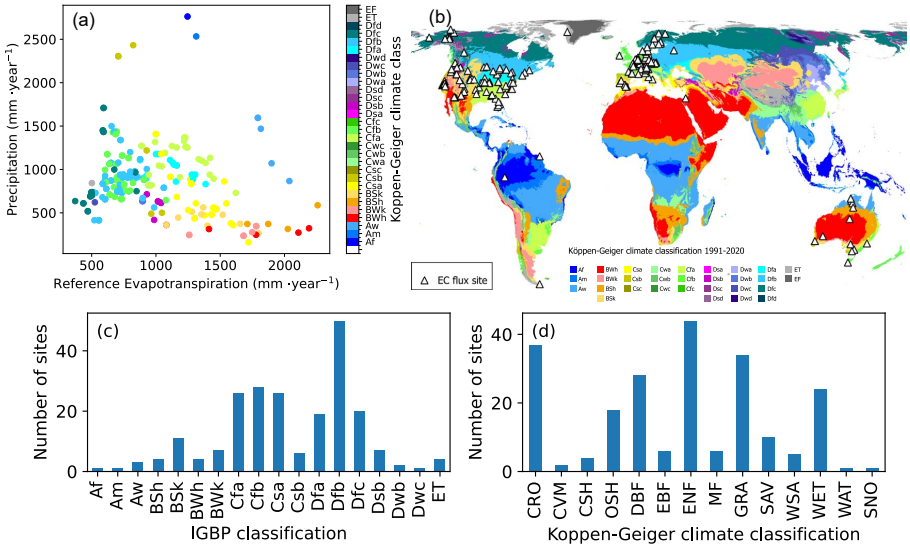


Figure 5-1. Distribution of EC flux sites used in this study across (a) the range of the annual precipitation and reference evapotranspiration (Data: AQUAMAPS-AgERA5), (b) geographical regions, (c) International Geosphere-Biosphere Programme (IGBP) land classification, and (d) Köppen-Geiger climate classification. Basemap: Köppen-Geiger climate classification 1991-2020 (Beck et al., 2023).

5.3.2 Satellite-based global data products

To estimate the relative uncertainty at a spatial support similar to that of WaPORv3L1 (300 m, dekadal), global satellite-based datasets with comparable spatial and temporal support were selected (**Table 5-2**). Time-series data were acquired from the selected products for the five-year study period (2018–2022), which is the overlap period of these products.

Table 5-2. Description of satellite-based global data products. List of abbreviations: ERA5 (European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5), AgERA5 (Agrometeorological indicators derived from ERA5), GMAO (Global Modelling and Assimilation Office), MERRA (Modern-Era Retrospective analysis for Research and Applications), VIIRS (Visible Infrared Imaging Radiometer Suite), MODIS (Moderate Resolution Imaging Spectroradiometer), GLDAS (Global Land Data Assimilation System), FEWS NET (The Famine Early Warning Systems Network).

Dataset name	Resolution		Methodology		Data source
	Spat. (m)	Temp. (days)	Model (type)	Main data inputs	
WaPORv3L1	300	10	WaPOR-ETLook (PM)	**Meteorological:** ERA5, AgERA5, GEOS5. **Satellite:** VIIRS (TIR and VNIR)	Actual evapotranspiration and interception (Global - Dekadal - 300m) - WaPOR version 3 (FAO, 2025b)
MOD16A2GF	500	8	MODIS16 (Mu et al., 2013, 2011) (PM)	**Meteorological:** GMAO, MERRA. **Satellite:** MODIS (TIR and VNIR)	MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500m SIN Grid V061 (Running et al., 2021)
PMLv2	500	8	PML_V2 (Gan et al., 2018; Y. Zhang et al., 2019, 2016) (PM)	**Meteorological:** GLDAS. **Satellite:** MODIS (TIR and VNIR)	Penman-Monteith-Leuning Evapotranspiration Version 2 (PML_V2) v0.1.8
SSEBopv61	1000	10	(Senay et al., 2023) (1SEB)	**Meteorological:** TerraClimate. **Satellite:** VIIRS (TIR)	SSEBop Actual Evapotranspiration Products (Version 6.1) (FEWS NET, 2024)

5.3.3 OpenET dataset

The ensemble of satellite-based ET_a models in the OpenET project was utilized to explore the impact of combining different model structures on the results of TC. The OpenET project provides an operational data generation system to provide high-resolution and low-latency satellite-based ET_a data from an ensemble of models for water managers in the Western United States (Melton et al., 2022). The benchmark validation dataset of the OpenET ensemble includes post-processed daily (and monthly) ET_a data for 194 locations with in-situ ET_a measurements within the contiguous United States (CONUS) (Volk et al., 2023c). For the study period from 2018 to 2022, the OpenET dataset was available for **53** sites, spreading over 10 land cover types (**Figure 5-2**). The post-processing methods for the EC data is detailed in Volk et al. (2023).

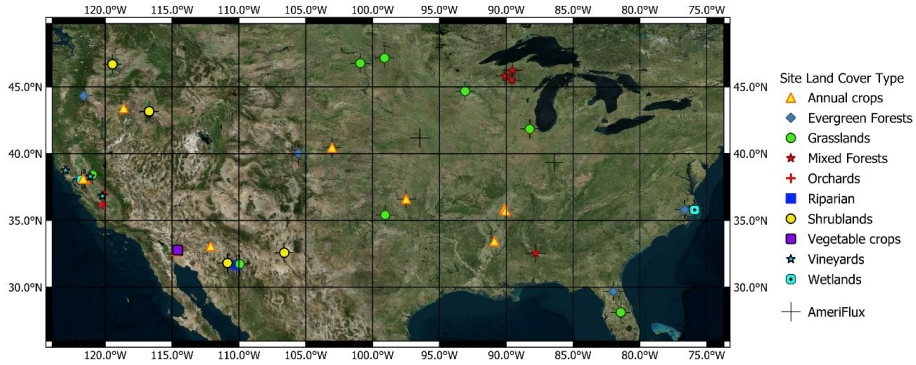


Figure 5-2. Locations of the selected sites from the OpenET benchmark dataset. Data source: Volk et al. (2023b).

In addition, the satellite-based ET_a daily time-series for these locations, derived from multiple models within the OpenET ensemble, were provided by Volk et al. (2023c). **Table 5-3** describes the five satellite-based ET_a models that constitute the OpenET ensemble. The SIMS (Satellite Irrigation Management Support) model is also an ensemble member and available from Volk et al. (2023d). However, it was not included in this study due to a large gap in data for the study period (47 % of the time-series).

Table 5-3. ET_a datasets from the OpenET ensemble of models. Data source: Volk et al., (2023d). Adapted from Melton et al. (2022, Table 1 and 2). List of abbreviations: PT (Priestley-Taylor), 1SEB (One-source Surface Energy Balance), 2SEB (Two-source Surface Energy Balance), CIMIS (California Irrigation Management Information System), NLDAS (North American Land Data Assimilation System), gridMET (Gridded Surface Meteorological), CFRS (Climate Forecast System Reanalysis).

Dataset name	Resolution		Methodology	
	Spat. (m)	Temp. (days)	Model (type)	Main inputs
PT-JPL	30	1	Priestley-Taylor Jet Propulsion Laboratory, ver. 0.2.1 (PT)	**Meteorological:** CIMIS, NLDAS. **Satellite:** Landsat (TIR and VNIR), MODIS (fAPAR).
eeMETRIC	30	1	Google Earth Engine implementation of the METRIC model, ver. 0.20.15 (1SEB)	**Meteorological:** CIMIS, NLDAS, gridMET. **Satellite:** Landsat (TIR, VNIR).
geeSEBAL	30	1	Surface Energy Balance Algorithm for Land using Google Earth Engine, ver. 0.2.1 (1SEB)	**Meteorological:** NLDAS, gridMET. **Satellite:** Landsat (TIR, VNIR).
DisALEXI	30	1	Disaggregation of the Atmosphere-Land Exchange Inverse, ver. 0.0.27 (2SEB)	**Meteorological:** CFRS. **Satellite:** Landsat (TIR and VNIR), MODIS (VNIR).
SSEBop	30	1	Operational Simplified Surface Energy Balance, version 0.1.5 (1SEB)	**Meteorological:** CIMIS, gridMET. **Satellite:** Landsat (TIR and VNIR).

5.3.4 In-situ validation

The objective of in-situ validation is to evaluate the compound uncertainty of the WaPORv3L1 dataset through direct comparison with in-situ estimates from the EC flux measurements. For that, dekadal ET_a time-series were derived from both datasets. The satellite-based datasets were sampled at the coordinates of the flux sites. Since the horizontal wind direction information was missing in the obtained EC flux dataset, it was

not possible to conduct dynamic flux footprint analysis for all sites. Instead, we applied weighted average sampling using a buffer area centred by the selected flux sites (a synthetic pixel) to sample dekadal ET_a time-series. This sampling method helps reduce bias caused by spatial heterogeneity (Zhou et al., 2025) especially when flux sites are located at the corner of the original pixels. The selected buffer distance was 100 m based on the length of synthetic pixel (200 m) that closely matches with the representative static flux footprint (Chu et al., 2021; Volk et al., 2023a).

The final gap-filled latent heat flux (LE) product from ONEFlux was used, which consisted only of measured or high-quality gap-filled records (LE_F_MDS). The daily time-series of LE from the EC flux datasets were aggregated to dekadal ET_a by averaging daily values. Daily LE values were converted to daily ET_a by dividing by the latent heat of vaporization. The latent heat of vaporization (λ) can vary with air temperature (Stull, 1988):

$$\lambda = (2.501 - 0.0023 \times T_{\text{air}}) \times 10^6 \tag{5.7}$$

where T_{air} [^{o}C] is the daily average air temperature.

However, since air temperature data was not available at all flux sites, a constant latent heat was used: $\lambda = 2.45 \times 10^6$ J·kg^{-1} at $T_{\text{air}} = 20$ ^{o}C (Allen et al., 1998). For 75 stations without gaps in the air temperature data, we compared the difference between daily ET_a calculated using constant and temperature-dependent latent heat (Equation (5.7)) was minimal (**Figure E-1**). Therefore, we considered the assumed constant λ for all stations to be sufficient. Finally, daily ET_a data was aggregated to dekads by arithmetic averaging.

The energy balance closure is the largest source of uncertainty in the EC flux measurements (Mauder et al., 2024). However, there is currently no one-site-fit-all method to correct for the energy balance closure (T. Wang et al., 2024) since each flux site presents a unique set of biometeorological factors that contribute to the lack of energy balance closure (Cui and Chui, 2019). Therefore, the energy balance closure correction was considered only as an indicator of uncertainty in in-situ daily ET_a. The LE values corrected for the energy balance closure (EBC) factor was calculated at 169 sites following the ONEFlux method (**Figure E-2**), which preserves the half-hourly Bowen ratio (Pastorello et al., 2020). For the other 53 sites, the EBC-corrected LE was missing due to lack of net radiation and ground heat flux measurements at flux sites.

The compound uncertainty of satellite-based data products was evaluated using four performance metrics from in-situ validation: bias, percent bias (PBIAS), root mean square error (RMSE), and coefficient of correlation (ρ) (**Table B-1**). The results are presented in Section 5.4. For comparison with the random error variance estimates from the ETC method, centred or unbiased RMSE (uRMSE) was also calculated (**Table B-1**).

5.3.5 Triple collocation analysis (TCA)

To estimate error variance and correlation coefficient, ET_a values from the three datasets were first collocated for approximately the same spatial and temporal support. For the estimation of uncertainty in the WaPORv3L1, all datasets were resampled to dekadal temporal resolution. The 8-day datasets were resampled using linear interpolation, where the 8-day value was uniformly assigned to each day within the period, and then these daily values were averaged to obtain dekadal values. The daily datasets were resampled by averaging all values within each dekad. The sample size of collocated data (i.e., the number of triplets) should be at least 100, with an optimal size exceeding 500 to achieve highest precision (Scipal et al., 2008; Tsamalis, 2022). Thus, for each triple combination, a minimum threshold of 100 triplets was used. For locations with less than 100 triplets, the results were considered invalid.

The covariance matrix for each combination of WaPORv3L1 and two other datasets other from either **Table 5-2** or **Table 5-3** was computed using their collocated dekadal time-series. Applying the ETC solution by McColl et al. (2014), we calculated the error variance ($\sigma^2_{\epsilon_{\mathrm{WaPOR}}}$) and coefficient of determination ($\rho^2_{\Theta,\, X_{\mathrm{WaPOR}}}$) for the WaPORv3L1 data with Equation (5.5) and (5.6) respectively. The results of the TCA for the WaPORv3L1 provide uncertainty information are expressed as uRMSE and correlation coefficient:

$$\mathrm{uRMSE}_{\mathrm{TC}} = \sqrt{\sigma^2_{\epsilon_{\mathrm{WaPOR}}}} \tag{5.8}$$

where $\mathrm{uRMSE}_{\mathrm{TC}}$ is the unbiased root mean squared error based on the TC approach, which is estimated from TC analysis result for $\sigma^2_{\epsilon_{\mathrm{WaPOR}}}$.

$$\rho_{\mathrm{TC}} = \sqrt{\rho^2_{\Theta,\, X_{\mathrm{WaPOR}}}} \tag{5.9}$$

where ρ_{TC} is the correlation coefficient based on the TC approach, which is estimated from TC analysis result for $\rho^2_{\Theta,\, X_{\mathrm{WaPOR}}}$. These metrics were compared with the $\mathrm{uRMSE}_{\mathrm{val}}$ and ρ_{val} resulting from in-situ validation (Section 5.3.4). In case of negative solutions for $\sigma^2_{\epsilon_{\mathrm{WaPOR}}}$ and $\rho^2_{\Theta,\, X_{\mathrm{WaPOR}}}$, indicating some assumptions of the TC method were violated, the results were also considered invalid. The results of triple collocation represent relative uncertainty (Section 5.4.2).

a. Triple collocation with global datasets

To estimate the uncertainty of the WaPORv3L1 dataset spatially, we used global datasets described in **Table 5-2**. The collocated time-series of the study period (2018 – 2022)

consists of 180 data points (5 years × 36 dekads) at each pixel. The required conditions for achieving reliable uncertainty estimates using TCA include large sample sizes and similar scales and magnitudes of errors among datasets (Sjoberg et al., 2021). Therefore, we resampled all global datasets to the same grid size as the coarsest dataset in a combination to address the difference in the spatial resolution and avoid bias towards higher-resolution datasets. We tested the assumption that linear calibration is sufficient by inspecting scatterplots of each pair of datasets. Since the scatter points are distributed evenly along a straight line (**Figure 5-3**), the assumption is likely to hold (Vogelzang, 2024). The assumption of non-zero error correlation might not hold when some datasets are not completely independent, e.g., due to common source of input data. Thus, the impact of the dataset selection was inspected by comparing TCA results derived from different combinations of three datasets selected from the four global datasets. This analysis reflects the uncertainty associated with uncertainty estimates using TCA.

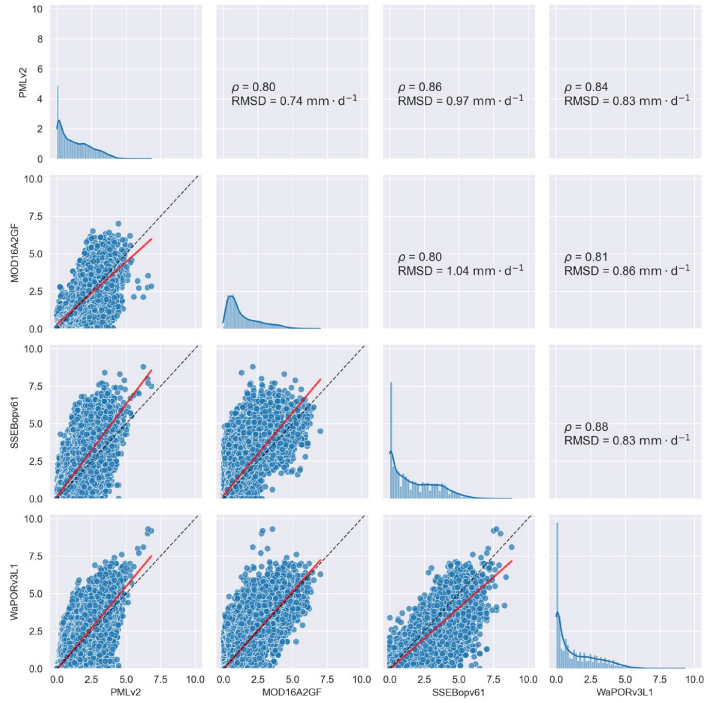


Figure 5-3. Cross-correlation of dekadal ET_a from satellite-based global datasets (WaPORv3L1, PMLv2, MOD16A2GF, and SSEBopv61) at the locations of 222 EC flux sites. The diagonal sub-plots show histogram of dekadal ET_a for each data product. The off-diagonal sub-plots show the scatter plots of each pair of data products. RMSD and ρ were calculated using formulas in **Table B-1**.

b. Triple collocation with the OpenET dataset

In addition to the global datasets, we conducted TCA using the OpenET benchmark dataset since it provides ET_a estimates from a diversity of model structures. The spatial support of the OpenET flux sites was bounded by a static flux footprint (210 m or 7×7 pixels of 30 m), which was determined based on optimal upwind direction and speed, and representativeness of the hourly footprint model (Volk et al., 2023b, 2023a). This is close to spatial resolution of the WaPORv3L1 (300 m) so we considered the OpenET model datasets can be collocated with the dekadal timeseries from WaPORv3L1. The OpenET dataset provides post-processed daily ET_a data from EC measurement and ensemble member models at the spatial support of flux sites. The daily ET_a values were aggregated to dekadal intervals by averaging. Similar to **Figure 5-3**, we also tested the assumption that linear calibration is sufficient for the OpenET benchmark datasets in **Figure E-3**.

5.4 RESULTS

5.4.1 Compound uncertainty from in-situ validation

a. Performance of the WaPORv3L1 data product

The performance metrics of the WaPORv3L1 dekadal ET_a across the selected EC flux sites are presented in **Figure 5-5**. The site-average metrics shows good correlation (ρ = 0.69) and low bias (bias = 0.084 mm·d^{-1}, PBIAS = 9.15 %) **(Figure 5-5)**. The site-average RMSE is 0.80 mm·d^{-1}, relatively low compared to the typical RMSE for satellite-based dekadal ET_a estimates in the literature (0.8 to 1.18 mm·d^{-1}) (**Table 3-1**).

Figure 5-4 also shows that there is a great variation of performance metrics across all flux sites. For the majority of the flux sites, the correlation of WaPORv3L1 with in-situ estimate is high (ρ > 0.8) and RMSE is low (RMSE < 1 mm·d^{-1}). The site-specific values of average bias and PBIAS for WaPORv3L1 are symmetrically distributed with respect zero, ranging from -2 to 2 mm·d^{-1} for bias (mostly within 100 % for PBIAS). The scatter plot of WaPORv3L1 against in-situ dekadal ET_a values for all site-intervals also shows a symmetrical distribution with respect to the 1:1 line. This means the WaPORv3L1 dataset might overestimate and underestimate dekadal ET_a at specific sites, but there is no notable systematic bias across all the included flux sites.

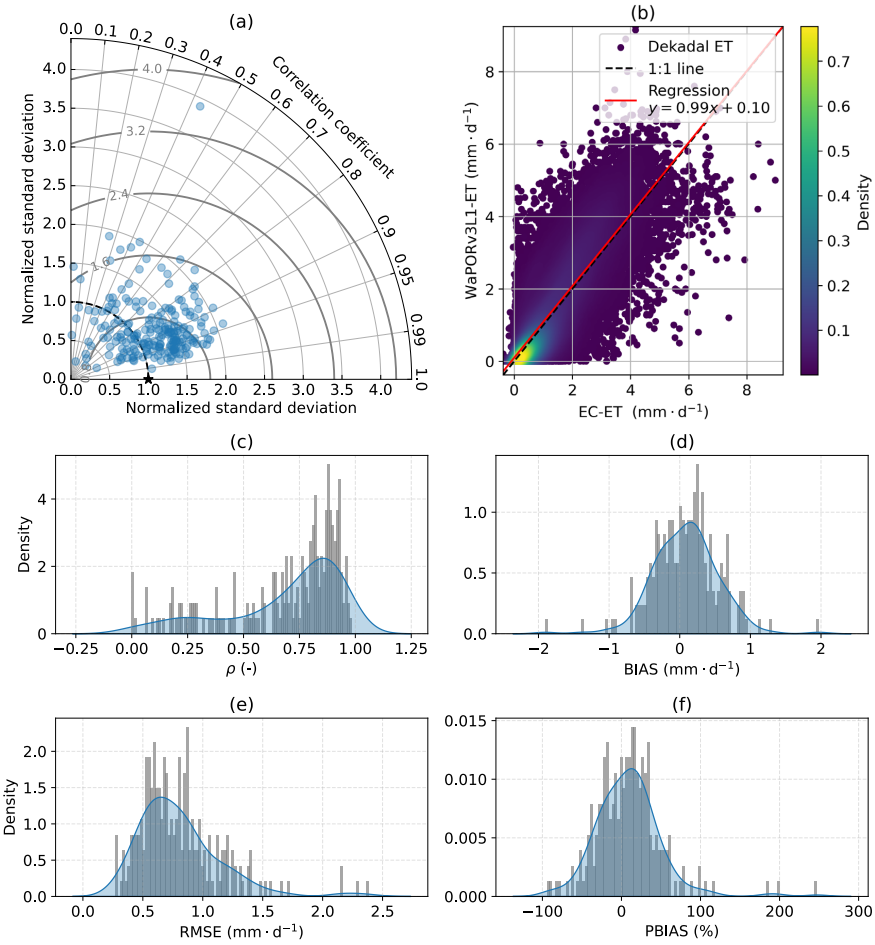


Figure 5-4. Performance of WaPOR-ET v3L1 product at 222 EC sites between 2018–2022. (a) Taylor diagram, (b) regression analysis of all data points, and kernel density estimation (KDE) of the probability density function for four performance metrics: (c) coefficient of correlation ρ, (d) bias, (e) RMSE, and (f) PBIAS.

The variation of performance metrics across the land cover types and climates of the flux sites provides guidance on the suitability of WaPORv3L1 in different conditions. The performance of the WaPORv3L1 dataset for cropland, shrubland and forest is relatively better than for savannah, water, and snow. The results also show lower bias and higher RMSE for homogeneous cropland (CRO) sites than for mixed mosaic of cropland (CVM) sites (**Figure 5-5**). For sites located in temperate, continental, and polar climates, WaPORv3L1 demonstrates an overall good correlation ($\rho > 0.5$). Meanwhile,

WaPORv3L1 exhibits low correlation ($\rho < 0.5$) and high relative RMSE and bias for sites located in tropical and dry climates (**Figure 5-5**). However, the number of sites per land cover class and climate class is not always comparable (**Figure 5-1**), thus limiting the robustness of the comparison. For example, the number of CRO sites (37) is much higher than CVM sites (2).

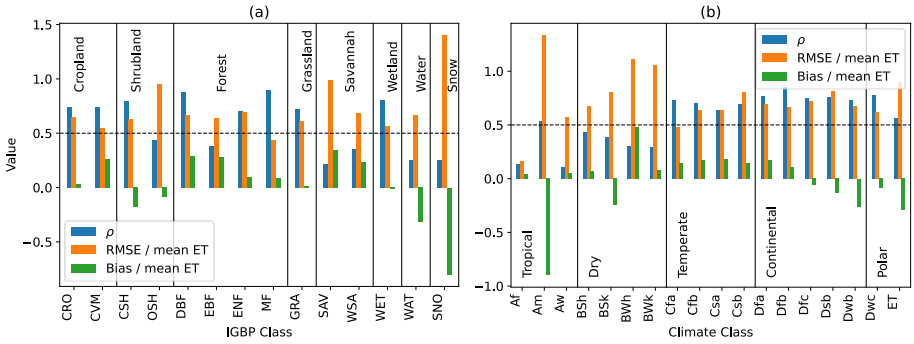


Figure 5-5. Average normalized performance metrics of WaPORv3L1 product by (a) International Geosphere-Biosphere Programme (IGBP) land cover classification and (b) Köppen–Geiger climate classification.

The validation results in **Figure 5-4** and **Figure 5-5** using the EBC-corrected data records at 169 EC flux sites are shown in **Figure E-4** and **Figure E-5**. The average performance metrics across these sites shows lower correlation ($\rho = 0.66$), higher bias (-0.12 mm·d^{-1}) and RMSE (0.82 mm·d^{-1}). The scatter plot between WaPORv3L1 and EBC-corrected dekadal ET_a values for all site-intervals also shows less agreement with respect to the 1:1 line, with more underestimation for high values of ET_a (**Figure E-4**). **Figure E-5** shows the average performance metrics of WaPORv3L1 for difference land cover classes and climate classes when applying EBC correction to the EC flux data. The correlation coefficients are not affected by the EBC correction for most sites, compared to **Figure 5-5**. However, the bias and RMSE decrease for cropland (CRO) and forest (DBF, EBF, ENF) but increase for shrubland (CSH). For other land cover classes, the changes are less discernible. The EBC-corrected LE flux was unavailable for comparison over water bodies (WAT). This constraint also applied to the sites located in the tropical climate zones (Af and Am). For sites located in other climates, the average bias becomes more negative and the average RMSE increases for most cases (**Figure E-5**).

b. Comparison with other global data products

Compared to the global datasets in **Table 5-2**, the WaPORv3L1 data has performance metrics that are within the same range of values across all flux sites (**Figure 5-6**). This finding aligns with expectations, given that the four datasets also exhibit a strong

correlation, as shown in **Figure 5-3**. WaPORv3L1 shows slightly higher median correlation than SSEBopv61 and MOD16A2GF and closely approaches that of PMLv2. The median bias and PBIAS of WaPORv3L1 are also closer to zero and comparable to PMLv2 than other products. However, the range of PBIAS is the largest for WaPORv3L1. Notably, SSEBopv61 shows the highest median bias and PBIAS and systematically overestimates ET_a, which could be partially because SSEBopv61 has the lowest spatial resolution (1 km) among these data products.

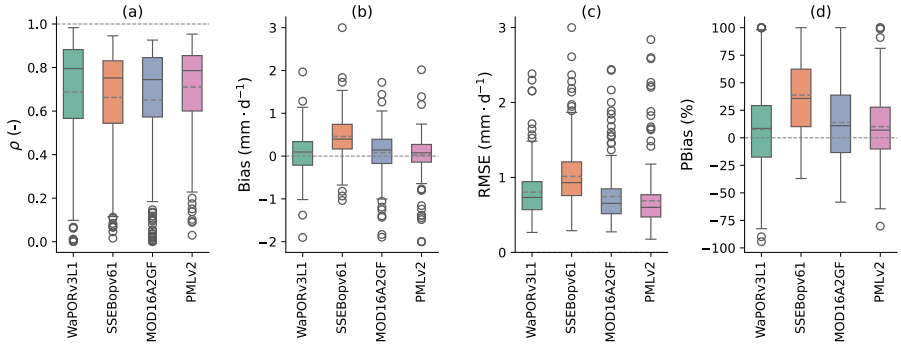


Figure 5-6. The averaged performance metrics of dekadal ET_a from WaPORv3L1, SSEBopv61, MOD16A2GF, and PMLv2 across EC flux 222 sites.

Based on the statistical summary of performance metrics, WaPORv3L1 showed performance similar to that of WaPORv2L1, based on 14 EC flux sites in Africa (ρ = 0.71, RMSE = 1.2 mm·d^{-1}, BIAS = 0.5 mm·d^{-1}) (Blatchford et al., 2020b). However, the dataset previously used to validate WaPORv2 was excluded from the WaPORv3L1 validation due to a temporal coverage mismatch and the lack of the ONEFlux post-processed data products. To evaluate the improvement of WaPORv3L1, we specifically compared the performance metrics of WaPORv2L1 and WaPORv3L1 at 9 flux sites and for 24 river basins in Africa for their overlapping period (2018 – 2022) (**Figure E-7** and **Figure E-10**). The results of these analyses also demonstrate that both versions of WaPOR-ET data perform similarly across 9 flux sites. However, an improvement in WaPORv3L1 performance is observed at 3 sites, with a slightly lower performance at the remaining sites (**Figure E-7**). At basin scale, WaPORv3L1 exhibits better agreement than WaPORv2L1 with the basin water balance residual, which was calculated using an ensemble of multiple datasets of precipitation, storage change, and runoff (**Figure E-10**). However, both versions of the WaPOR-ET data insufficiently reproduced the interannual variability in the water balance residual, a challenging issue also observed in other leading satellite-based products (Zhao et al., 2025).

5.4.2 Relative uncertainty from triple collocation analysis

a. Spatial distribution of relative uncertainty

Figure 5-7 shows the spatial distribution of relative uncertainty estimated by the TCA of WaPORv3L1 with two combinations of satellite-based global data products. In many regions, both combinations show high correlation with the underlying signal of ET_a (ρ_{TC} close to 1) and low uRMSE, which is indicated by green colour. Regions with high uRMSE and low correlation, coefficient indicated by red colour, are found in Western and Central Africa, Indian Subcontinent, Greater Mekong Subregion, Central-West Brazil and Northern Australia. These are regions dominated by tropical savanna climates (Aw). In the dry regions, such as the Sahara, Southwest Asia, and Central Australia, the TCA using WaPORv3L1-SSEBopv61-PMLv2 showed low correlation coefficient and low uRMSE in WaPORv3L1. This could be due to low ET_a ranges and the challenges to detect the variation of ET_a at low range (< 1 $mm \cdot d^{-1}$). Meanwhile, the TCA using WaPORv3L1-MOD16A2GF-PMLv2 yielded invalid results in these regions due to data masks within MOD16A2GF dataset.

One of the advantages of TCA is its capability to estimate both ρ_{TC} and uRMSE at the pixel level. This is essential for spatially investigating the relative uncertainty of WaPORv3L1 within a region. For example, in the Nile delta, satellite-based global datasets exhibit variation in ET_a estimates in terms of trends, magnitude, and spatial distribution (**Figure 5-8**). As a result, the relative uncertainty estimated by TCA also shows spatially varied results, even within a region with similar climate and land cover. Spatially variable uncertainty (both correlation coefficient and random error variance) can unintentionally lead to a misinterpretation of ET_a spatial and/or temporal variations. This risk is of particular concern when the ET_a estimates are used in comparative analyses such as land and water productivity gap analyses (e.g. Chukalla et al., 2024). Therefore, spatially distributed estimation of relative uncertainty could provide complementary information for the applications of satellite-based ET_a data products.

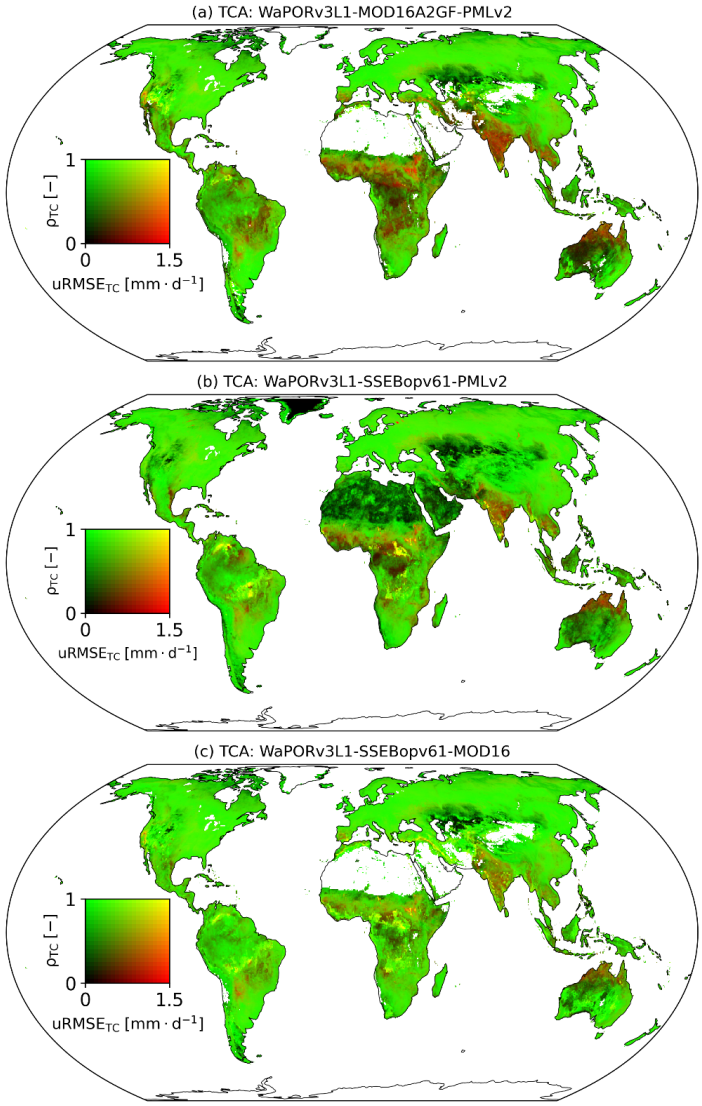


Figure 5-7. Temporal average of error standard deviation ($uRMSE_{TC}$) and correlation coefficient (ρ_{TC}) of dekadal ET_a from WaPORv3L1 calculated by triple collocation analysis (TCA) with (a) MOD16A2GF and PMLv2, (b) SSEBopv61 and PMLv2, and (c) SSEBopv61 and MOD16A2GF. TCA was conducted at pixel level: 500m for (a) and 1000 m for (b), then resampled to 30km resolution for visualization. Results are not shown for areas where TCA yielded invalid estimates.

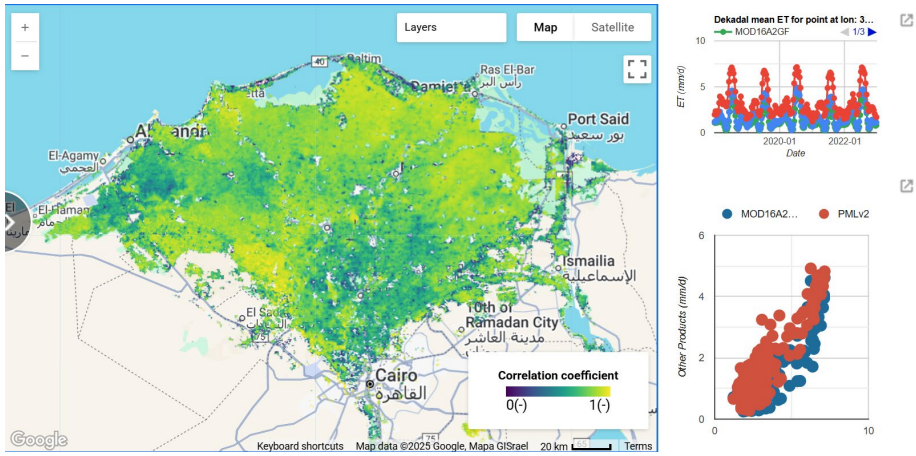


Figure 5-8. Temporal average of correlation coefficient of dekadal ET_a from WaPORv3L1 calculated by triple collocation analysis (TCA) with MOD16A2GF and PMLv2. High-resolution map zooming into the Nile delta, Egypt (left panel). Comparison of the three satellite-based datasets at pixel-scale (right panel). The high-resolution maps can be viewed at: https://ee-wapor.projects.earthengine.app/view/tca

b. Impact of dataset selection on triple collocation analysis

The selection of datasets influences the estimated uncertainties derived from TCA (Wu et al., 2019a). **Figure 5-7** also indicates that TCA using SSEBopv61 underestimated the $uRMSE_{TC}$ of WaPORv3L1 relative to the results obtained from the TCA performed with PMLv2 and MOD16A2GF. This difference is likely attributed to the presence of non-zero error cross-correlation among the selected datasets. All four satellite-based global datasets were generated using distinct meteorological data sources and algorithms. However, because WaPORv3L1 and SSEBopv61 were both derived using satellite data from the same sensor (i.e., VIIRS) (**Table 5-2**), it is likely that their random errors are cross-correlated, which violates a major assumption of TCA. The results presented here confirm that this is violation may lead to underestimation of error variances of those datasets (González-Gambau et al., 2020). Conversely, TCA using the WaPORv3L1-MOD16A2GF-PMLv2 combination is expected to penalize uncertainty of WaPORv3L1, because the three data products rely on the same satellite data sources (i.e., MODIS).

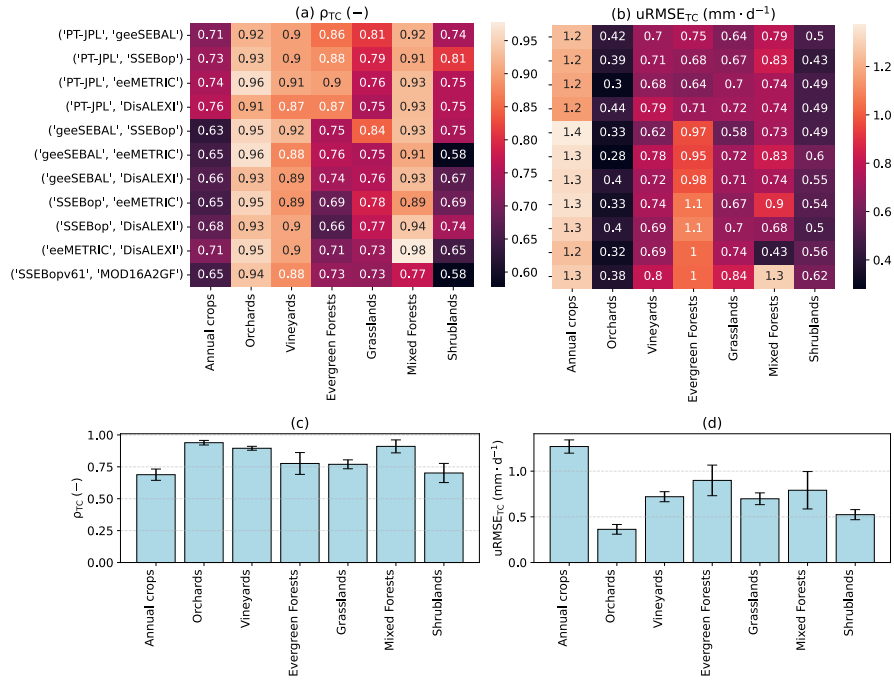


Figure 5-9. Correlation coefficient (ρ_{TC}) and random error (uRMSE$_{TC}$) of WaPORv3L1-ET calculated by triple collocation with OpenET's ensemble models at 53 benchmark sites (a-b), and average values of all combinations (c-d). 'SSEBop' refers to an ensemble member model in the OpenET dataset, while 'SSEBopv61' refers to the global data product. 'Annual crops' include all sites with field crops (e.g., maize, corn, wheat, alfalfa…). The TCA at the remaining land cover types did not yield valid results due to insufficient sample size (N < 100).

While using the same input data is a common cause of error correlation (as two datasets share errors arising from the sensors, calibration, or processing chain), other suggested causes include shared algorithms or models (Pan et al., 2015), shared ancillary data, such as observed data assimilated into reanalysis products to derive meteorological inputs (Gruber et al., 2020), and violations of other TC assumptions, such as error non-orthogonality and random errors (Gruber et al., 2016a; Vogelzang and Stoffelen, 2022). While identifying all the causes for inaccurate estimation of uncertainty using TCA is beyond the scope of this study, we utilized the OpenET ensemble-member models to investigate the influence of shared models on TCA results. These datasets were all derived using VNIR and TIR data from Landsat, while using different model structure. Therefore, we would expect the estimation of uncertainty in WaPORv3L1 to be conservative (or overestimated) when collocating with two other datasets that are more correlated to each

other than to WaPORv3L1 (**Figure E-3**). At the same time, the variation of TCA results when using ET_a estimates from different models would indicate the impact of modelling schemes. **Figure 5-9** shows the variation of TCA-estimated $uRMSE_{TC}$ and ρ_{TC} using various combinations of WaPORv3L1 and two other satellite-based modelling schemes from the OpenET suite. Most combinations showed TCA results with high correlation of WaPORv3L1 in orchards, vineyards and mixed forests ($\rho > 0.8$). Most land cover types showed low average $uRMSE_{TC}$ (< 1 $mm \cdot d^{-1}$), except annual crops. This indicates that WaPORv3L1 differs greatly from other datasets in annual crops.

The results of TCA varied both due to different combinations of datasets and land cover classes. The uncertainty range, however, appeared wider across different land cover types than across different data combinations. For example, the TCA results (average of all combinations) showed the $uRMSE_{TC}$ of WaPORv3L1 ranges from 0.4 (orchards) to 1.3 $mm \cdot d^{-1}$ (annual crops), and the correlation coefficient ranges from 0.7 (annual crops and shrublands) to 0.9 (orchards). Meanwhile, the largest uncertainty range across different combinations was $uRMSE_{TC}$ from 0.43 to 1.3 $mm \cdot d^{-1}$ in mixed forest, and from 0.66 to 0.92 for correlation coefficient in evergreen forest. It is also noteworthy that the results exhibit significant differences across the three distinct crop types as defined by OpenET, particularly since these are categorized under a single class (CRO) in global classifications (**Figure 5-5**).

The disagreement in TCA results between different combinations is most notable in evergreen forest and mixed forest. For evergreen forest, the TCA using a combination of a SEB model (e.g., geeSEBAL, SSEBop, eeMETRIC, and disALEXI) and a PT-type model (i.e., PT-JPL) yielded lower $uRMSE_{TC}$ (0.64 to 0.75 $mm \cdot d^{-1}$) and higher correlation coefficient (0.86 to 0.90) compared to combinations using two SEB models (0.95 to 1.1 $mm \cdot d^{-1}$; 0.66 to 0.76). Njuki et al. (2023) observed that one-source surface energy balance (1SEB) models tend to have biases when estimating ET_a in tall, dense forests due to the treatment of the land surface as a single homogeneous layer and inadequate parameterization of canopy turbulence in heat transfer processes. Rashid and Tian (2024) reported larger bias by the PT-JPL model in denser vegetation. In their study, Blatchford et al. (2020) also drew attention to the lowest performance of the WaPORv2 data product at an evergreen forest site. In light of these studies, the higher range of uncertainty estimates in TCA at forest sites suggests that the underlying biases of different models in specific biomes influence the error correlation between datasets, and thus, variability of TCA results.

5.4.3 Comparison of relative and compound uncertainty

The relationship between the compound uncertainty, indicated by $uRMSE_{val}$ and ρ_{val} from validation against EC flux measurement, and the relative uncertainty, indicated by $uRMSE_{TC}$ and ρ_{TC} estimated by the TCA with global ET_a products, was investigated. Their scatter plots for $uRMSE_{TC}$ and ρ_{TC} and linear regression relationship at 215 EC flux sites are shown in **Figure 5-10.** The TCA results for WaPORv3L1 included three cases: WaPORv3L1-MOD16A2GF-PMLv2, WaPORv3L1-SSEBopv61-PMLv2, and WaPORv3L1-SSEBopv61- MOD16A2GF. Since the direct comparison with in-situ ET_a estimates from the EC flux measurements is considered the de facto standard for uncertainty assessment of satellite-based ET_a estimates, the relationship between TCA and validation against EC was regarded here as a reference for evaluating the reliability of dataset choice for TCA.

The $uRMSE_{TC}$ are weakly correlated with $uRMSE_{val}$ for all cases (ρ = 0.40 to 0.51) (**Figure 5-10**). The ρ_{val} and ρ_{TC} are slightly more correlated than uRMSE (ρ = 0.6 to 0.64). All TCA combinations underestimate random error (uRMSE) and overestimate correlation coefficient (ρ), compared to results from in-situ validation. The TCA with WaPORv3L1-MOD16A2GF-PMLv2 resulted in highest $uRMSE_{TC}$ and lowest ρ_{TC}. As discussed in Section 5.4.2, this was expected since the TCA using MOD16A2GF-PMLv2 penalizes WaPORv3L1 more than when using SSEBopv61-PMLv2 and SSEBopv61-MOD16A2GF due to different satellite data input (**Table 5-2**). However, TCA using MOD16A2GF-PMLv2 results were slightly closer to the compound uncertainty indicators than using SSEBopv61-PMLv2 and SSEBopv61-MOD16A2GF. Since MOD16A2GF and PMLv2 are both based on PM-type models like WaPORv3L1, but using different satellite input, the TCA results suggests that the independence of input data for ET_a models leads to more agreement between TCA and in-situ validation.

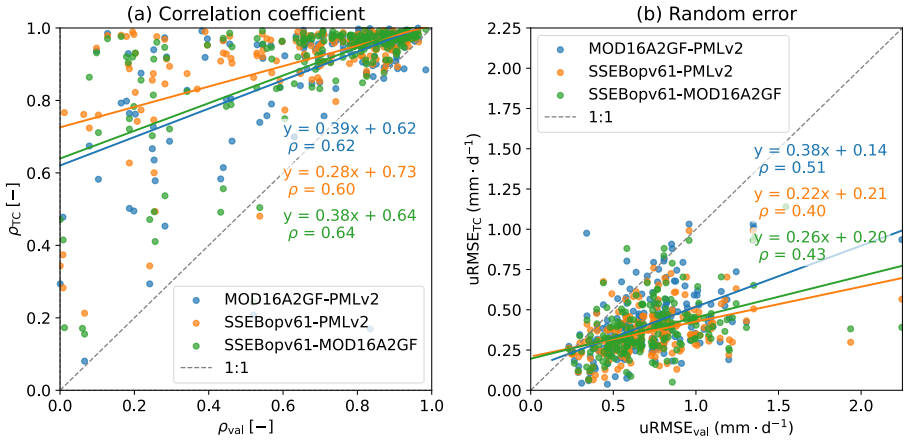


Figure 5-10. The correlation coefficient (ρ) (a) and random error (uRMSE) (b) of WaPORv3L1 at 222 EC flux sites, derived from in-situ validation against EC flux dataset and from TCA with MOD16A2GF-PMLv2 and SSEBopv61-PMLv2.

To further investigate the impact of underlying independent models on the relationship between relative uncertainty estimated and compound uncertainty, the Euclidean distances between $uRMSE_{TC}$ and $uRMSE_{val}$, and between ρ_{val} and ρ_{TC} for each combination of OpenET datasets were calculated across different land cover classes (**Figure 5-11**). As previously shown in Section b, the TCA using PT-JPL with one SEB-type model resulted in higher $uRMSE_{TC}$ and lower ρ_{TC} than other combinations, especially for forest sites. Here, the Euclidean distances showed better agreement between the compound and relative uncertainty when the TCA using PT-JPL with one SEB-type model. In case of ρ_{TC}, the three combinations with the lowest Euclidean distance to ρ_{val} were PT-JPL-geeSEBAL, PT-JPL-eeMETRIC, and PT-JPL-DisALEXI. Meanwhile, PT-JPL-eeMETRIC, PT-JPL-SSEBop, and PT-JPL-DisALEXI were the three combinations with highest similarity between $uRMSE_{TC}$ and $uRMSE_{val}$. Given the expectation that TCA using WaPORv3L1, PT-JPL and SEB-type model would exhibit greater independence in modelling schemes, the results in **Figure 5-11** also supports that TCA results using datasets derived from more independent modelling schemes agree more with in-situ validation results.

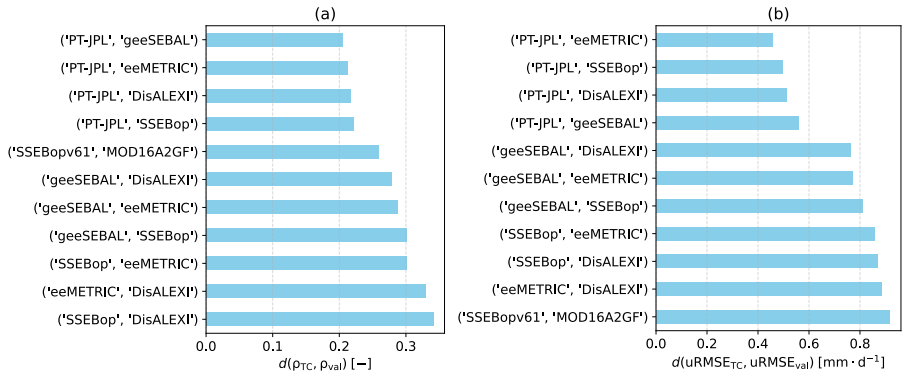


Figure 5-11. Euclidean distance (d) between ρ_{TC} and ρ_{val} (a) and between $uRMSE_{TC}$ and $uRMSE_{val}$ (b).

Figure 5-11 also shows the results from the TCA with global data products, SSEBopv61, PMLv2, and MOD16A2GF. These products have much larger spatial support (300 to 1000 m) than the OpenET dataset (210 m or 7×7 pixels of 30 m), thus, their ET_a estimation will contain signals from larger area. Therefore, we would expect that the relative uncertainty derived from TCA using global data products would exhibit lower similarity to the compound uncertainty from in-situ validation (higher d) than the TCA using the OpenET models, which is the case for $uRMSE_{TC}$ (**Figure 5-11**). However, the ρ_{TC} by the TCA with SSEBopv61 shows relatively close to in-situ validation compared to the TCA with only SEB-type models. This suggests that underlying dependent models affect the ρ_{TC} from TCA more than representativeness errors.

5.5 Discussion

5.5.1 Limitations of ex-post uncertainty assessment methods

Uncertainty assessment methods have limitations and gaps, which are previously reviewed and discussed in Chapter 2. This chapter applied and discussed the utility of TCA for assessing correlation and random errors, which enables uncertainty quantification across the entire spatial extent of satellite-based datasets even in the absence of in-situ data. The non-zero error cross-correlation and representativeness errors indicate the uncertainty of the TCA results. The spread of uRMSE in **Figure 5-10** could be partially explained by the suboptimal TCA sample size, 180 triplets (5 years × 36 dekads) per location, which causes uncertainty in estimated error variances (Scipal et al., 2008; Tsamalis, 2022). On the other hand, in-situ validation with EC flux measurements, despite having its own inherent uncertainties due to data quality and scale mismatch, is still dominantly used to assess uncertainty in satellite-based ET_a. The limited data sample

also affects in-situ validation, and even more so since long-term EC data records are also not available in many locations. Furthermore, some assumptions of TC and TCH methods are also required for in-situ validation (Gruber et al., 2016b). Since both in-situ validation and TCA have uncertainty, the purpose of TCA was to evaluate the relative uncertainty of WaPORv3L1, not ranking datasets by lowest error variance and highest correlation.

The inherent energy imbalance observed in EC flux measurements affects the quality of in-situ reference at many flux sites. Post-processed EC flux data at 169 sites demonstrates that applying energy balance correction can alter the daily ET_a value by a substantial range, from -40 to 80 % (**Figure E-2**). Furthermore, the spatial scale mismatch, driven by the dynamic flux footprint variation and the heterogeneity of the surrounding surface, may introduce errors in the sampled ET_a value when compared to a larger sampling area. For example, the source of latent heat flux measured at a site might originate mainly from a pixel not directly collocated with the site's location. This potential representativeness error, however, is considered manageable in the current study because the sampling area is defined as within 100 m around the site. While one could employ a dynamic EC flux footprint model (Kljun et al., 2015) to account for high heterogeneity, the impact of footprint variation is considered minimal for the current EC datasets, as they exhibit considerable surface homogeneity.

In this study, the assumptions of TCA have been investigated to some extent by using scatter plots (system biases between datasets are minimal), comparing different datasets (the impact of dependent errors), using datasets with similar spatial and temporal resolution (limit representativeness errors). Two of the main concerns about triple collocation were investigated: the choice of datasets and reliability compared to in-situ validation. The choice of datasets was limited to global datasets that have similar spatial and temporal resolution as WaPORv3L1, which allows for applying TCA at WaPORv3L1's resolution. This leads to inevitable non-zero error correlation, which impact was evaluated by comparing different combinations of datasets. Although TCA metrics are useful for indicating correlation and random errors, they do not account for bias.

5.5.2 Future research

This chapter has shown the implementations and comparison of in-situ validation and TCA as ex-post uncertainty assessments for WaPORv3L1. This section discusses potential future research agendas to improve ex-post uncertainty assessment for satellite-based ET_a data production at global and continental scales.

To improve the robustness of TCA, several refinements should be considered. First, implementing TCA at a coarser resolution is recommended to focus on the investigation of spatial pattern rather than absolute results. This approach allows for the inclusion of other coarse-resolution and reanalysis datasets, such as GLEAM and GLDAS. Second, to

improve the precision of the results and increase the statistical sample size, the analysis period should be extended (aiming for more than 500 dekads, which equates to at least 14 years of data). Currently, WaPORv3L1 data is only available from 2018 to the present, which limits the current sample size. Finally, future studies could employ more specialized techniques like Extended Collocation (Gruber et al., 2016a) for more than three datasets and Correlated Triple Collocation (González-Gambau et al., 2020) for combinations with two datasets exhibiting strong errors correlation. These methods explicitly account for error correlation of one pair of datasets, while still relying on non-zero error correlation for other pairs. Finally, the presentation and interpretation of TCA can be enhanced with recent visualization methods, such as Triple Collocation Diagram (Siu et al., 2024) and Comparison Map Profile (He et al., 2025).

This study also underscores the importance of the existing publicly available datasets from regional flux networks for evaluation and quality assessment of satellite-based ET_a data products. These datasets serve a purpose extending beyond the validation of satellite-based data, including other applications, such as monitoring greenhouse gas emission. Ongoing initiatives are important to improve the access and accuracy of EC flux data, such as the FLUXNET Data System (The FLUXNET Coordination Team et al., 2025) and improving the ONEFLUX post-processing pipeline (Zhang et al., 2024), and development of benchmark dataset development for validation of ensemble models (Volk et al., 2023a). Regional networks that are not yet part of FLUXNET could provide a source of in-situ data despite not employing FLUXNET standards yet, such as the NENA-ET and SAEON networks. Strategic in-situ validation could focus on designing campaigns to collect in-situ ET_a measurements over a short period in critical areas, including areas with high heterogeneity, less observed types of climates and land cover (Yi et al., 2024, fig. 2).

Over the past two decades, significant advancements have been made in improving EC measurements and interpretation, particularly in reducing the systematic energy imbalance (Mauder et al., 2024). In addition, recent efforts to address the scale mismatch such as investigating spatial sampling methods (Zhou et al., 2025), the impact of heterogeneity (Burchard-Levine et al., 2021; Wu et al., 2023), and upscaling methods of in-situ measurements (Du et al., 2023; Li et al., 2021), should be considered for improving future in-situ validation of satellite-based ET_a data products.

The strategy of combining in-situ measurements and multiple satellite-based ET_a models could help address the lack of long-term observation, while opportunistically providing new information and insights for improving global satellite-based ET_a modelling and data production. New methods for reconciling and utilizing existing data are instrumental for this purpose. Constraining $\mathrm{uRMSE}_{\mathrm{TC}}$ and ρ_{TC} using in-situ validation at certain quality control flux site. Improvements in mapping $\mathrm{uRMSE}_{\mathrm{TC}}$ and ρ_{TC} could be achieved by

employing specialized TC and TCH methods and global datasets that are derived using independent models and input datasets.

Without robust and consistent methods for uncertainty assessment, claims about the improvement in accuracy seem elusive, most likely based on the increase in spatial resolution of the thermal sensor (e.g., ECOSTRESS, VIIRS). However, this only reduces the uncertainty due to spatial scale mismatch. In other words, more detailed data is not necessarily more accurate and more useful for all applications (Braun, 2021), for example, improving hydrological understanding (McCabe et al., 2017). This is also true for temporal variability, however, since TCA estimates temporal average uncertainty, it could not indicate this. For this, temporal trends comparison, such as with water balance residual, can be used instead. As Zhao et al. (2025) argued, a critical limitation of satellite-based ET_a products (in their analysis, GLEAM and OpenET) is the interannual variability. This undermines the uses of satellite-based estimation for detecting trends of ET_a (water use, hydrological responses) over time.

5.6 Conclusions

This chapter assessed the compound uncertainty of WaPORv3L1 through in-situ validation against 222 Eddy Covariance flux sites. Furthermore, the relative uncertainty was evaluated using extended triple collocation with global data products and OpenET ensemble models. The compound uncertainty of WaPORv3L1 falls within the typical range of performance metrics reported in the existing literature. Overall, the performance of WaPORv3L1 is comparable to other data products with similar spatial resolution and coverage. Most validation sites exhibit strong performance, characterized by a correlation coefficient exceeding 0.8 and a RMSE of less than 1 $mm \cdot d^{-1}$. However, the performance metrics demonstrate significant variability across different climate and land cover types. WaPORv3L1 achieves good performance in continental and temperate climates but shows reduced skill in dry and tropical regions, which is potentially exacerbated by the limited number of available flux sites in those climatic zones. Performance is notably low over certain land cover types: evergreen broadleaf forests, savannas, water bodies, and snow. Furthermore, the current validation lacks representation from urban sites. It is also observed that sites presenting very low correlation, high RMSE, and substantial bias often exhibit high land surface heterogeneity and a large residual in energy balance closure.

The relative uncertainty exhibited only a weak correlation with the compound uncertainty. Both the TCA using global data products and the analysis employing the OpenET ensemble models consistently underestimated RMSE and overestimated the correlation coefficient. Furthermore, the TCA using global data products indicated that uncertainty is higher in tropical regions than in temperate regions, which is consistent with the in-situ validation results. Our investigation into the impact of dataset selection on TCA results

suggests that the use of shared input data has a greater influence on the reliability of the TCA than the use of shared underlying models. While the RMSE derived from TCA is more affected by representativeness errors, the correlation coefficient is more influenced by the shared underlying models. Moreover, the influence of shared underlying models on TCA results was particularly prominent in evergreen and mixed forests.

In conclusion, TCA represents a valuable additional assessment to in-situ validation, as it reveals the spatial variation of uncertainty estimates, which is particularly useful in areas that lack in-situ data. However, the absolute RMSE and correlation coefficient derived from TCA are often overly optimistic, and its reliability is highly dependent on data selection, representativeness errors, and land cover types. Therefore, TCA should be used with caution, especially for ranking the performance of models or data products. In this study, we selected data products with similar temporal and spatial resolutions as for TCA, in order to estimate relative uncertainty at WaPORv3L1 resolution. However, when uncertainty estimation at the native spatial resolution of the data product is not required, priority should be given to selecting data products that feature a greater diversity of input data and underlying models.

6

REFLEXIVE ANALYSIS OF METHODOLOGICAL UNCERTAINTIES

This chapter is based on:

Tran, B. N. and Mul, M.L.: Blurry eyes from space: A reflexive analysis of uncertainty research on WaPOR's satellite-based model and data. (submitted to Water Alternatives, under revision)

6.1 INTRODUCTION

Statistical methods, particularly probability theory, have been widely used to represent uncertainty, as repeatable measurements and simulations allow for quantification in probabilistic terms (Montanari, 2007; Povey and Grainger, 2015). However, as Nearing et al. (2016) point out, uncertainty is not solely a technical issue; epistemological uncertainty, which arises from the choice of probability frameworks, as well as philosophical and linguistic uncertainties, resist quantification. Melsen (2022) refers to this distinction by categorizing statistical approaches as addressing only 'technical uncertainty'. While probability-based approaches capture an important subset of uncertainty, they do not fully encompass the broader epistemological and methodological challenges inherent in environmental modelling.

Beyond technical uncertainty, methodological uncertainties arise from the subjective choices embedded in modelling practices, including model selection, calibration strategies, and conceptual framing (Foody and Atkinson, 2003; Melsen et al., 2019). Methodological uncertainties emerge from differences in approaches and interpretations among researchers, reflecting the epistemological underpinnings of the study rather than just the mechanics of the model. A similar typology of uncertainties is also discussed in satellite remote sensing (RS). While RS practitioners typically focus on quantifying errors that are known, Povey and Grainger (2015) argue that there are other types: "known, unquantifiable unknowns", factors that are recognized but cannot be measured, and "unknown unknowns", factors that remain entirely unobserved or unrecognized. Since the methodological uncertainties are often not quantified or quantifiable, explicit reflection and account for these uncertainties are important to avoid rhetorical use of models (Saltelli and Funtowicz, 2014).

Krueger and Alba (2022) contend that the hydrology community's long-standing debate on uncertainty fosters reflexive engagement with ontological and epistemological commitments in interdisciplinary water research (e.g., Beven, 2002; Nearing et al., 2016). Reflexivity, widely discussed in qualitative research, has been defined and applied in various ways. For example, Walsh (2003) observed four distinct practices: personal reflexivity, interpersonal reflexivity, methodological reflexivity, and contextual reflexivity. Methodological reflexivity "acknowledges the theoretical commitments that surround each research project" while contextual reflexivity is "concerned with situating a given study in its cultural and historical milieu" (Walsh, 2003, p. 56).

Nevertheless, reflexivity seems less prominent in quantitative research like water modelling (ter Horst et al., 2024). Sensitivity auditing builds on sensitivity analysis, a well-established approach in modelling communities (Saltelli, 2004), thus, introduces reflexivity without requiring a radical departure from current modelling workflows. Reflexivity in sensitivity auditing is operationalized through tasks such as identifying and

scrutinizing assumptions, evaluating uncertainties, and ensuring transparency in model design and use (Lo Piano et al., 2022).

Inspired by methodological and contextual reflexivity (Walsh, 2003) and sensitivity auditing (Saltelli et al., 2014), we open up our reflection on the uncertainty assessment of the WaPOR-ET data and the ETLook model to investigate beyond the technical uncertainties. Specifically, we assess how the methodological uncertainties can be ascribed to the institutional, social or political context of commissioning, developing, and using the outputs of the ETLook model (i.e., the WaPOR-ET data). We ask, 'What are the conditions, assumptions, and problem framings that influence how the WaPOR project was developed and the ETLook model was developed and applied?', 'How are the uncertainties in the WaPOR data due to the choices in developing and applying ETLook model addressed in the project?', and 'What are potential consequences of methodological uncertainties?'. By explicitly discussing these questions, we demonstrate how sensitivity auditing could enhance methodological and contextual reflexivity, before using the WaPOR data indiscriminately.

Our engagement with the WaPOR project over seven years (2019-2025) was predominantly through the application of the ETLook model outputs for various applications, mainly water accounting, and the quality assessment [15] of the WaPOR database. Therefore, we position ourselves more as data and model users than modelers [16], although we were involved in a few meetings with modelers to provide feedback on model development. As model users, we have made ourselves familiar with the ETLook model through self-studying: reviewing project and model documents (FAO, 2020b, 2020a, 2018a, 2018b), running an open-source [17] version of the model in small areas (FAO, 2025a), and developing educational resources on the concepts and tools used in the model (IHE Delft, 2023).

In addition to our direct experience, we are also informed by project documents and reports, meeting recordings, and interpersonal exchanges with professionals in the WaPOR project and those who are utilising the WaPOR-ET data. Our major limitation is

[15] Data quality assessment could entail evaluating various criteria such as accuracy, coherence, comparability, accessibility, timeliness and relevance (FAO, 2023b). It typically involves comparing outputs against reference data, verifying adherence to best practices, and ensuring methodological soundness (Barsi et al., 2019).

[16] Within this chapter, we heuristically define modeler as the specialist who creates and develops a model, model user as the analyser who runs and configures the model for specific cases, and data user as end-user who consumes and interprets output data from the model.

[17] Open source refers to a type of software or model that is made publicly available (through online repository) for anyone to use, modify, and distribute freely.

our lack of interaction with different actors within and around the WaPOR project, who potentially use and are influenced by the use of the ETLook model outputs. Therefore, we learn about their experiences regarding the use of the WaPOR-ET data through secondary sources, mainly from fieldwork activities of colleagues in Lebanon, Tunisia, Jordan, and Egypt (Contractor et al., 2023; El-Wattar, 2022; Matar, 2019).

6.2 WaPOR's satellite-based model and its ecosystem

We first unpack the social and technical context of the WaPOR project, which is situating the WaPOR-ET data and the ETLook model. Since 2015, the Dutch Government has supported FAO to establish the WaPOR project and database. The origins of this initiative can be traced to a pivotal encounter described in a blog post by a former diplomat specializing in water affairs at the Dutch Ministry of Foreign Affairs (MoFA) and First Secretary of the Netherlands Embassy in Sana'a, Yemen, in which he recounts the visit by a professor and their exchange about the use of satellite images for monitoring agricultural water use in Yemen (**Appendix F**). This account highlights the role of the professor's visit in shaping the early conceptualization of the WaPOR project. By linking a funding request for domestic water supply in Yemen and a simplified inference about the cause of water shortage[18] by this professor, the diplomat was motivated to advocate for the use of "satellite information" to manage water resources.

Once the idea of developing satellite data to map water use in agriculture was conceived by the ministry, many organizations in the network of experts in water management and geospatial technology also got involved, mainly Dutch institutes (Box 1).

[18] Yemen has been undeniably suffering from water crisis, and its agricultural production consumes a large proportion of its water. However, this apolitical explanation about the cause of domestic water shortage abstracts the issue from its local context, neglecting the external economic forces and political ramification that shifted Yemeni rain-fed agriculture towards irrigated cash cropping (Varisco, 2019).

Box 1. Key partners in the WaPOR project[19]

The Dutch Ministry of Foreign Affairs serves as the main funder for the WaPOR project, contributing to the initiative's goal of using remote sensing for monitoring and enhancing water productivity. FAO serves as the executing agency for the project, leading its technical and methodological coordination. Within FAO, the Land and Water Division is responsible for budget management and overall project oversight, ensuring alignment with Sustainable Development Goal (SDG) 6 (clean water and sanitation) and SDG 2 (zero hunger).

IHE Delft is responsible for developing water accounting reports using remotely sensed data (WA+), database quality assessment, and delivering specialized training curricula targeting stakeholders such as government agencies, academia, and technical staff. The International Water Management Institute (IWMI) is also a key partner for implementing capacity-building activities and piloting field-level applications of WaPOR data in selected countries.

The FRAME Consortium, who won an open tender to develop and operationalize the WaPOR database version 1 and 2 in phase 1 (2016 – 2020), comprises the following key partners: eLEAF, University of Twente's Faculty of Geo-Information Science and Earth Observation (ITC-Twente), the Flemish Institute for Technological Research (VITO), and WaterWatch. Later in phase 2 (2021 – 2026), eLEAF, a Dutch private company specializing in satellite-based monitoring technologies, was primarily responsible for producing the WaPOR database.

Through synergy with another project[20], Wageningen University & Research (WUR) also contributes to the WaPOR project with its expertise in agricultural research and resource management, providing a scientific foundation for WaPOR's tools and methodologies. In the same project, MetaMeta, a social enterprise in the Netherlands, focuses on end-user needs and engagement.

The WaPOR project capitalized on the advancements in remote sensing technology pioneered by Dutch geospatial technology institutions (eLEAF and ITC-Twente) and the Netherlands' expertise in water management (IHE Delft and WUR). These capabilities

[19] It is not difficult to recognize that the key partners in the project are the same "old epistemic friends who share a long history of developing and actively promoting and circulating a particular body of water knowledge" in the Water Accounting initiative (Zwarteveen et al., 2019, p. 235).

[20] Water Productivity Improvement in Practice. See https://www.wur.nl/en/project/Water-Productivity-Improvement-in-Practice-Water-PIP.htm (accessed on 31 January 2025)

aligned with the Dutch Ministry of Foreign Affairs' international development objectives, particularly regarding sustainable agriculture and water resource management in the Global South. Meanwhile, FAO brought its longstanding institutional platform for capacity building in the Global South to design the project for practical applicability. By funding the WaPOR project, the Dutch government reinforces its international influence in water knowledge and sustainable development, particularly in target countries in the West Asia, Northern Africa, Sahelian and Horn of Africa regions (FAO, 2020b). The WaPOR data portal and API are instrumental in expanding WaPOR's technical ecosystem with development of ICT-based applications in target countries where the project aims to increase national capacity and make the WaPOR data utilized at field level more.

6.2.1 Counting 'crop per drop'

The MoFA funds the WaPOR project to monitor and improve agricultural water productivity (WP) gaps (briefly defined in Box 2), aligning with the global objective, particularly SDG Indicator 6.4.1 "Change in water use efficiency (WUE) over time" (United Nations, 2021). This initiative is embedded in a broader institutional shift from traditional irrigation efficiency (IE) metrics toward WP.

Box 2. Water Productivity (WP) definition and the use of WaPOR data for WP indicator

> The WaPOR project defines WP as the net quantity of output per unit of water used to produce that output. Since there are many ways to define output for agricultural water management, there can be many indicators for WP, such as biophysical, nutritional, economic, and social WP (van Steenbergen et al., 2022). Among those, only biophysical WP can be monitored with satellite RS. The project focuses on the indicator of biophysical WP, which is specifically calculated as the amount of crop (biomass or yield) produced per unit of water consumed by the crop through evapotranspiration (ET), expressed in $kg \cdot m^{-3}$.
>
> The definition of WP gaps as "the difference between the maximum attainable WP and the currently achieved WP at the field scale" (Zheng et al., 2018) was adopted in the project. The maximum attainable WP could be determined by, for example, the highest levels of productivity that a farm has already reached under real-world conditions (Chukalla et al., 2024).

As early as 1958, the agricultural WUE was expressed by researchers at WUR in terms of kg crop production per m^3 water transpired, indicating the physiological performance of crops in optimizing water for plant growth (Bessembinder et al., 2005). IWMI later broadened this concept to "the physical mass of production or the economic value of production measured against gross inflow, net inflow, depleted water, process depleted

water, or available water" (Molden, 1997, p. v). Molden contends that irrigation efficiency (IE) is a limited measure for agricultural water use, which ignores crop yield and water reuse. This critique underpinned the shift from IE to WP, positioning it as a more comprehensive framework for assessing benefits from water use at the basin scale.

However, both WUE and WP have been contested for their reductionist focus on water as a single input for agricultural production (Giordano et al., 2021; Zoebl, 2006). Drawing on dry-farming studies and historical agronomic perspectives, Zoebl (2006) demonstrates that high WP values do not inherently indicate better agricultural practices, forewarning that the shift toward WP, mainly promoted by IWMI and FAO, risks oversimplifying the complexities of water and agricultural management and sidelining broader sustainable farming concerns.

The topic of IE and productivity encompasses diverse perspectives and ongoing debates (Boelens and Vos, 2012; Knox et al., 2012; Seijger et al., 2025; van Halsema and Vincent, 2012), which cannot be settled through scientific and methodological approaches. As Lankford (2012) suggests, these discussions must be linked with political and societal concerns regarding resource governance. In the light of this academic debate, the concept of monitoring 'crop per drop' to enhance agricultural water management adopted by the WaPOR project is not merely a technical intervention but also a politically charged standpoint.

6.2.2 Turning the eyes from space onto water

The institutional shift from IE to WP and its operationalization coincided with the rapid proliferation of satellite-based models in the late 1990s. Traditionally, WUE or WP data were collected at the field scale, where most variables were readily available to practitioners—except, perhaps, for ET_a (Bessembinder et al., 2005). The diversity of satellite-based ET_a models, the growing number of satellites, and advancements in sensor technology have made obtaining ET_a data more accessible to researchers, yet not fully providing sufficient capabilities for global ET-based applications (Fisher et al., 2017).

Among all the data layers offered in the WaPOR database, ET_a data seems to be appreciated and adopted the most, since it has more potential use for water management beyond WP calculation, such as basin water assessment, drought impact study, and crop water use[21]. While FAO has several modelling tools to determine ET_a and crop yield at field level (e.g., FAO-56, AquaCrop), "none of those tools can be used in an operational

[21] Almost three fourths of the applications listed on FAO's catalogue of WaPOR uses and applications use ET data layer, more than other data layers https://www.fao.org/in-action/wapor-uses-applications-catalogue/en/ (accessed on 17 December 2024)

way to assess and monitor water productivity rapidly, routinely and consistently in time and space at affordable costs" (FAO, 2020b, p. 7). Therefore, the need for the WaPOR database has been justified with this technological gap "to develop an improved methodology to monitor water productivity at low costs[22], in near real-time, that can be applied at different spatial scales" (FAO, 2020b, p. 7). This approach attempts to mitigate these challenges by centralizing WP data through an open-access portal, thereby lowering technical entry barriers for (potential) WP practitioners, allowing them to bypass the complexity of modelling[23] and focus solely on interpreting the data (i.e., the model outputs).

From the above, we could infer that the WaPOR project, which frames the use of satellite-based models, has socio-political and technical commitments to operationalize the WP concept and capitalize on satellite RS for water management. In the next two sections, we describe how these commitments are affected by methodological uncertainties of the ETLook model, through two main themes: the choices in and about modelling[24] of ET_a and the perception and management of uncertainties.

6.3 A PARTIAL ACCOUNT OF CHOICES

Modelling entails a myriad of choices and junctures[25], which would be impossible to unpack completely in the limited space of this chapter. Here, we discuss the choice of modelers and models and provide some examples of modelling choices that are made to meet the requirements of the WaPOR database.

6.3.1 The modelers and the model

As part of the FRAME Consortium, eLEAF was appointed through a public tender process to develop and implement the production of the WaPOR data. The Netherlands-based high-tech company capitalizes on its proprietary in-house developed ETLook

[22] The WaPOR data is free of charge and publicly available. The cost here most likely refers to the cost for database production, dissemination, maintenance, and other project cost, which was roughly five million euros.

[23] This complexity also includes choosing a model out of the various models in Table 1-1.

[24] Choices within the modelling process itself and choices regarding the adoption and application of models.

[25] Junctures are "where things are done in a way that could be different" and other choices could be made that change the socio-political outcomes of model (Krueger and Alba, 2022, p. 12).

model, which forms the core of its geospatial consultancy services[26]. Compared to other satellite-based ET_a models, ETLook is a rather new model, first introduced by Pelgrum et al. (2010) at the Remote Sensing and Hydrology symposium, presenting some examples of field-level applications in China, Australia, and India. ETLook is "the successor of SEBAL" (Brombacher et al., 2022), which is one of the most academically used models for estimation of ET_a. The ETLook model uses the Penman-Monteith (PM) equation, a longstanding method which is also used by FAO in its guideline for crop water requirement (Allen et al., 1998). The PM equation relies on the quantification of surface and aerodynamic resistance, both of which are contingent upon the varying biophysical conditions of the canopy and soil surfaces. In order to quantify canopy and soil parameters required in the PM equation, the modelers at eLEAF predefine parameter values for each land cover class. Consequently, the accuracy of the ETLook model outputs rely on the accuracy of land cover classification data (Brombacher et al., 2022) and the underlying assumption that a single parameter value can be generalized for each land cover class regardless of location.

Additionally, the access to the ETLook model is not completely open, limiting independent verification or modification by an external party. On FAO's request, eLEAF provides description of the model in methodology reports and a set of computer code in Python programming language which contains mathematical formulas for calculating soil and canopy parameters, meteorological variables and spectral indices[27] to solve the PM equation. However, the complete procedure for data processing and gap-filling[28] remains proprietary since it is, as an expert at eLEAF explained to us, their business to provide data processing services. Moreover, eLEAF relies on paid cloud-computing platforms[29] to handle the vast computational demands of running the ETLook model at continental

[26] eLEAF delivers satellite-derived data for crop monitoring, irrigation planning, yield forecasting, and specialized solutions for sugarcane cultivation, supporting the entire agri-business value chain. See https://eleaf.com/science/algorithms/ and https://eleaf.com/services/agriculture/ (accessed on 30 November 2024)

[27] Spectral indices are mathematical combinations of different bands of light captured by satellite sensors, reducing the complex surface conditions to numerical values that indicate various surface characteristics from vegetation greenness to soil variations.

[28] Satellite data are collected at specific moments, much like a snapshot. As satellites orbit the Earth, they revisit the same location only after several days, leading to potential gaps, which deteriorates the ET model accuracy (Wang et al., 2023). These gaps are typically addressed using various statistical techniques to ensure continuity in data. However, gap-filled data cannot fully assure the same quality as actually captured data (Song et al., 2021).

[29] A cloud-computing platform, Amazon Web Services in this case, provides pay-as-you-go access to Amazon's computing servers over the internet.

and global scale. This underscores an economic barrier to the computational infrastructure required to replicate the output of the ETLook model. All these factors make the model highly specialized and exclusive[30].

As discussed, the ETLook model is not merely a technical tool. It plays a role in circulating certain methodologies of water quantification and centralizing the management and distribution of water data within organizations and countries that have technical capacity and infrastructure. Although WaPOR data products are generated using a common methodology that aims to support consistent cross-region and cross-period comparisons of ET_a, their accuracy varies because model performance differs under varying regional conditions (see Chapter 5). Arguably, the dissemination of outputs from the ETLook model through WaPOR portal balance the unequal access and capacity to satellite RS technology by providing data free-of-charge and readily available through the internet.

6.3.2 The database

The specific assignment of developing a low-cost, near real-time, and multi-scale WP database requires specific configurations of ETLook model for each spatial level. The configuration choices are related to spatial resolution, data input, and data preprocessing. For example, the spatial resolution of thermal data[31] used in ETLook is coarser than the desired resolution of the continental and global levels. Instead of running the model at the original resolution of thermal data, eLEAF modelers decided to apply resampling and downscaling techniques[32] to enhance the spatial resolution of thermal data to meet the technical requirement by the project. This choice could be explained technically in that the computer model requires all input data to be in the same grid size, which allows performance of calculation pixel by pixel. However, an alternative could be running the model at native resolution of thermal data, to reduce uncertainties due to resampling and

[30] In contrast to this exclusive reliance on a single model, a project in the western USA with a similar focus on addressing ET data gaps through satellite RS has adopted multiple models developed by various research groups (https://etdata.org/).

[31] WaPOR version 2 was produced using MODIS thermal data (1 km resolution). WaPOR version 2 uses VIIRS data (375 m resolution). In ETLook, thermal data is used in combination with vegetation information to infer the soil moisture, which has effect on the transpiration of plant.

[32] Resampling refers to simple averaging the numbers at the large pixel into smaller pixel size, generalizing the value for smaller pixels. Downscaling involves another model to relate the coarse-resolution data to other sources of data with finer resolution, using their statistical relationship to estimate the desired data at finer resolution.

downscaling[33]. However, the ETLook model might not be equivalently valid at all levels of resolutions. Moreover, this alternative is seen as less desirable to the project as it would not result in a good enough resolution for quantification of ET at field level, which would attract more potential uses of the WaPOR database (FAO, 2020b, p. 20).

Another crucial configuration choice concerns the temporal extent of the model. Since RS data are only available for the period during which a satellite remains in its orbit, data availability presents a fundamental challenge for modelling. This limitation, often perceived as a technical issue, has led modelers, in the case of WaPOR version 2 (2009–2022), to integrate two different sources of multispectral data[34]: one providing finer resolution from 2014 onward and another covering the period from 2009 to 2014 with coarser resolution (FAO, 2018b). The impact of this decision on the temporal patterns of ET analysis became evident to us only through our water accounting study of the Niger River Basin (FAO and IHE Delft, 2020, p. 38). We could have exercised greater caution regarding the limited temporal availability of satellite data and avoided modelling for the desired period continuously, which creates the illusion of long-term ET data suitable for historical analysis. However, the modelers appear to have prioritized their responsibility for the database requirements over these concerns.

From the discussion above, we have demonstrated that while model configuration choices may be technically justified, they can have immense impacts on water assessments based on model outputs. These examples further highlight the importance of evaluating modelling decisions alongside case-specific assessments, rather than uncritically treating WaPOR data as an inherently reliable and continuous source for interpreting water fluxes. Moreover, we recognize that many consequences cannot be fully anticipated during the modelling process. This reflection reminds us that use of model results should not be presented as representing absolute truth but rather as contingent upon the assumptions, limitations, and context in which they were produced.

6.4 PERCEPTION AND MANAGEMENT OF UNCERTAINTY

Since uncertainties in modelling are inevitably present and, particularly in the case of satellite-based models, often exceedingly challenging to assess, sensitivity auditing encourages a critical examination of whether these uncertainties have been understated to

[33] The uncertainties due to spatial resampling and downscaling in satellite ET retrieval are especially challenging to evaluate and have only been studied in a few cases (see Chapter 2).

[34] This was a particular issue for the national level data. For continental level data, only the lower resolution was used, which does not show the same issue.

make results appear more robust. To reflect on this question, we draw on our own experience using WaPOR data for water accounting, as well as insights from other users, to explore how uncertainty is perceived and managed within the WaPOR project.

6.4.1 Data users

As a WaPOR data user in the Netherlands, we conducted water accounting studies for five river basins (Litani, Jordan, Awash, Niger, and Nile) using WaPOR data as prescribed in the project's terms of reference[35]. A simple basin water balance[36] is the first step to check if the total amount of water entering the basin (precipitation) equals the amount leaving basin (ET, river discharge to the sea), accounting for any changes in storage. Our calculation for Litani using WaPOR data revealed an imbalance of approximately 20 % of precipitation.

In response to this issue, an expert in the team, who had been previously a co-developer of ETLook (Pelgrum et al., 2010), explained that this was mostly due to bias in ET data, suggesting multiplying the basin ET value with a correction factor to close the water balance. The expert's attributing water imbalance to ET inaccuracy suggests less confidence in this data than other components (precipitation and change in water storage). However, at that time, we had no clear understanding of this presumption and no possible explanation for the shortfalls of ETLook in this area.

Later, thanks to the data acquired from the Litani River Basin Authority for another study (Matar, 2019), we found that by accounting for the huge amount of water diversion from the basin for hydropower generation and for domestic uses in Beirut (which cannot be observed with RS data), we could resolve this discrepancy without 'correcting' ET data. If this adjustment had been applied, we fear that it would have led to us making a substantial overestimation of agricultural water consumption in the Bekaa Valley, Litani, thereby misinterpreting the potential of reducing agricultural water use to improve WP at basin level.

A similar pattern of scepticism and adjustment of ET values is also observed in the Egyptian Ministry of Water Resources and Irrigation (MWRI) in utilizing WaPOR data

[35] IHE Delft is responsible for implementing the project output 'Assessment of the consequences and sustainability of possible increases in water productivity by means of water accounting'. This output has been delayed due to staff turnover at IHE. As we started our positions at IHE Delft in the middle of the WaPOR first phase, we took over these studies with very limited time for implementing them.

[36] Water balance assumes the principle of mass conservation, a common assumption in hydrology. Uncertainties arise due to measurement errors, unobserved water fluxes (such as deep percolation or water diversion by human). Discrepancies in water balance signal the need to check for these issues.

for a project to improve the management and use of water. El-Wattar (2022)'s interviews with MWRI officials revealed that data users exhibited a dual attitude toward WaPOR data, simultaneously appreciating its ET data while expressing scepticism regarding its accuracy when findings deviated from their expectations or estimation by other methods. Notably, one MWRI data user, tasked with designing a national monitoring system for water allocation, addressed uncertainty in the ET values from WaPOR by applying a similar approach suggested to us by the expert in the water accounting study of the Litani Basin: multiplying the ET values with different correction factors until the output aligned with her common-sense judgment (El-Wattar, 2022, p. 30).

From the above, it can be inferred that the perception of uncertainty in the modelled ET values creates a tendency among data users to modify the data. Since users have neither access to the underlying ETLook model nor the technical capacity to re-run the model, the application of correction factors serves as an informal calibration process[37]. However, as shown in the case of Litani water accounting, this approach could introduce arbitrary bias, particularly when assumptions remain implicit and unverified, and potential alternative explanations are overlooked. In both cases, we contend that field data and local insights are crucial in explaining the uncertainties of using WaPOR data. Therefore, WaPOR data should be used in a complementary manner with field data to enhance the accuracy and contextual relevance of analyses.

However, some applications of WaPOR data implicitly assume that satellite RS alone provides a superior or sufficient basis for decision-making at field level. For example, there have been some attempts to use WaPOR data to provide advisory services to farmers to irrigate their fields. The pioneering example is the PlantVillage mobile app, which was developed and introduced in Kenya by Penn State University (USA), "because they consider WaPOR a powerful instrument to provide open-access and reliable data for their extension/advisory work focusing on good practices of soil and moisture conservation" (FAO, 2020b, p. 37). The inclusion of access to the WaPOR database in this mobile app happened even before the first WaPOR data quality assessment had been published in 2019. This early adoption gives a sense of trust in the reliability of the data by the institute and, at the same time, demonstrates their technical advantage in leveraging the, at that time, relatively underutilized API provided by FAO to enhance access to the database.

In contrast, other cases of the field-level mobile app using WaPOR data faced significant challenges in adoption and trust. National institutions, such as the National Agricultural Research Centre in Jordan and the Institut National des Grandes Cultures (INGC) in

[37] Model calibration refers to the process of adjusting a model's parameters to make its outcomes more accurate when compared to observation data (Oreskes et al., 1994). In this case, the data user introduced a new parameter (i.e., the correction factor) and the observation data is replaced with their own perception of the right ET amount.

Tunisia, expressed low trust in WaPOR's ET_a data, due to the lack of validation at field level, so they used the ET_o data[38] instead of ET_a for the calculation of crop water requirement in their irrigation advisory mobile apps (Contractor et al., 2023). The ET-based irrigation advice given by mobile app was not trusted and used by the farmers in Tunisia "as it asked them to irrigate when they did not have the water to do so" (Contractor et al., 2023, p. 45). Although using ET_a could have provided a more realistic estimation of crop water needs, it still would not address broader constraints such as limited access to water and other essential agricultural inputs. This habitual reliance on model outputs, repackaged and disseminated as part of the WaPOR database, often occurs without "doing the right sums"[39], highlighting that the consideration of methodological uncertainty is as important as technical uncertainty.

6.4.2 Modelers and model users

In the WaPOR project, both eLEAF modelers and model users have engaged in the quality assessment of the WaPOR database. During the second phase of the project, our team was responsible for evaluating the quality of version 3 data. This assessment encompasses all WaPOR data components, including biomass production, precipitation, and ET_o. However, given that this chapter focuses on the ETLook model, we limit our discussion to our experience with ET_a data.

One of the main challenges in assessing data quality is acquiring reliable reference sources and field data for comparison with WaPOR data. As outlined in Chapter 2, the standard practice in evaluating remotely sensed data involves validation against ground measurements. In the case of satellite-based ET_a estimation, the eddy covariance (EC) technique[40] is considered the most reliable reference, as it is widely used for validation in the majority of studies (Chapter 2). However, in WaPOR's target regions (Africa and Southwest Asia) the availability of such data is severely limited, both in spatial distribution and temporal coverage. For instance, FLUXNET includes only a few stations in these regions, and the temporal extent of the available datasets is often restricted to specific historical periods. While we acknowledge that no measurement technique is entirely free of bias and error, the lack of field data significantly hampers our ability to

[38] WaPOR's ET_o is calculated using the PM equation parameterized for a hypothetical reference crop in FAO-56 report (Allen et al., 1998).

[39] Rule 6 in sensitivity auditing is about asking whether the use of quantification is not foreclosing important alternative ways of problem framing while focusing on a narrow-defined problem (Saltelli et al., 2020)

[40] Eddy Covariance techniques estimate ET flux by measuring vertical wind speed and moisture concentration in the air mass at the sensor placed above the vegetation canopy.

evaluate satellite-based data quality. To mitigate this limitation, we supplemented our comparisons with flux site data by incorporating alternative data sources, including other models and methodologies. Our assessments are partial and inherently constrained by the extent of our knowledge about other available references and actual field conditions.

During the development phase of version 3, we had several meetings with experts from eLEAF, as well as the FAO project manager and technical officer, to discuss potential adjustments to the database methodology based on data quality assessments. Based on these discussions, the eLEAF team conducted several runs of the model to test adjusted configurations and parameterizations in selected areas. The choice of these test areas was informed both by concerns over uncertainty in specific conditions (such as irrigation in semi-arid regions) and by the availability of reference data. Each team conducted independent evaluations of preliminary model outputs using a range of methods and data sources.

The eLEAF team concentrated on comparisons with data from a limited number of EC sites, to which they had access through professional contacts. In discussions with a manager at eLEAF, she emphasized their primary objective: to determine whether the model's results were 'good enough' and to identify areas for improvement. By testing the model at a few selected locations, they were able to evaluate its performance in specific contexts. However, a key unresolved issue remains: the extent to which a model calibrated for a handful of locations can be reliably applied across entire continents.

Meanwhile, the FAO team focused on assessing the water balance of major river basins and comparing WaPOR data with outputs from other models. Through collaboration with a project funded by European Space Agency, the FAO team also facilitated the comparison of ETLook with another widely used model, TSEB, in case studies conducted in Tunisia, Lebanon, and Spain (Guzinski et al., 2021). This collaboration has led to methodological advancements, particularly in thermal data downscaling, which contributed to improvements in ETLook's ability to model ET_a at the field level. As part of this effort, the team developed an open-source version of ETLook, called pyWaPOR, which incorporates data processing and gap-filling functionalities (FAO, 2025a).

While the quality assessment activities only provide a partial understanding of uncertainties, they demonstrate the significant effort and commitment of both modelers and data users to refining model outputs. As illustrated here, the primary focus of quality assessment has been on technical uncertainties, specifically inconsistencies between WaPOR data and reference datasets. This approach assumes that ET data accuracy can be statistically inferred from comparisons with field data. However, given the limited representativeness of available validation sites, even perfect alignment with field-level data at selected locations does not necessarily imply broader accuracy across diverse landscapes. Thus, while ongoing quality assessments contribute to improving the model's robustness, uncertainty remains an inherent challenge.

The development of pyWaPOR presents a valuable opportunity to explore technical uncertainties in the modelling process by enabling quantitative experiments with different configurations, as well as uncertainty and sensitivity analyses. However, the computational demands of running ETLook over large geographic areas present a significant challenge for implementing these analyses across the entire WaPOR domain. Nevertheless, as an open-access model, ETLook invites users to implement and refine it according to their specific applications. To support this process, we have incorporated pyWaPOR into our training curriculum as part of the project. Our goal is to familiarize WaPOR data users with the concepts and methodologies involved in the modelling process, enhancing transparency and allowing earthly eyes to scrutinize the assumptions underpinning the model.

6.5 CONCLUSIONS

The emphasis on agricultural water use and the rising interest in water productivity among powerful nations and institutions, coupled with advancements in their capacity for satellite RS and ET modelling, has legitimized the use of satellite-based ET_a models in developing a global database for monitoring WP. Inspired by the call for greater reflexivity in quantitative modelling, we have examined the socio-political context of utilizing the WaPOR data considering the uncertainties of employing the ETLook model. These include associated methodological uncertainties, problem framing, model selection, and configuration choices. Although alternative satellite-based ET_a models exist, the socio-political commitments of the WaPOR project have led to the exclusive use of the ETLook model in specific configurations to meet prescribed requirements for the database. Furthermore, we examined how the uncertainties associated with WaPOR data are perceived and managed by various users and modelers. In applications such as water accounting and irrigation advisory, as well as in quality assessments conducted by modelers and model users, diverse approaches are employed to understand and assess uncertainty, with a predominant focus on technical dimensions.

WaPOR represents a coordinated effort by governments and institutions to democratize access to RS by providing global satellite-based data for enhancing WP. However, neither improving WP nor simplifying access to data inherently result in more just and sustainable water management. Modelers, model users, and data users must critically evaluate uncertainties and, when necessary, adapt methodologies to local and context-specific conditions, ensuring that the ‘eyes from space’ vision is effectively anchored in observations on the ground. This imperative arises not only because all methods inevitably contain errors that influence spatial and temporal WP evaluation, but also due to the broader political and societal implications of using WP concepts and satellite RS.

Recognizing that data derivation from satellite-based models involves subjective decisions and framing makes purely technical solutions less straightforward than it seems. Uncertainty research encourages exploring alternative data, model structures, and parameters in modelling. By explicitly admitting and addressing uncertainties, modelling can be more open and involve participation from data users, who would have more insights and knowledge about the actual condition in the area of interest. The retrieved ET data could have supported a wider array of applications beyond WP. Through contextual and methodological reflexivity in using the satellite-based model, we can create opportunities for alternative pathways toward a more just and sustainable approach to water management, both through and beyond increasing overall WP.

7
EPILOGUE

This chapter summaries the main conclusions from the previous chapters and addresses the research questions in Chapter 1. Following that, this chapter provides the author's personal reflection on the research leading to this dissertation. Finally, this chapter concludes with an outlook for future research.

7.1 MAIN CONCLUSIONS

This dissertation assesses uncertainties in satellite-based estimation of ET from two perspectives, through both the literature (Part I) and the case study of the WaPOR-ET data product (Part II).

First, methods for assessing uncertainty in satellite-based ET_a estimates were identified and appraised through a systematic quantitative literature review (Chapter 2). This review indicates that efforts to assess uncertainties vary in methodology and are constrained by the availability of reference data. The most common method found in the literature is in-situ validation against Eddy Covariance (EC) measurements. However, this primary approach faces a significant gap as EC data availability is heavily concentrated in North America and Europe. In regions where such in-situ measurements are limited, studies must rely on the residual in the water balance as a reference for validation. Even where EC data exists, persistent challenges remain across reviewed studies, including non-closure of the energy balance and a space-time scale mismatch. Given the limits of in-situ validation, this research highlights the importance of applying diverse approaches, especially when validation datasets are limited.

In addition, a meta-analysis of studies that validated satellite-based ET_a estimates against EC measurements reveals the status of reported compound uncertainty in terms of the most employed metric, Root Mean Square Error (RMSE), and percent bias (PBIAS) (Chapter 3). While RMSE is the most employed metric, it is a measure of error variance and scale dependent. PBIAS represents systematic errors and is scale-independent, which would be more suitable for comparing uncertainties across studies. The meta-analysis found RMSE ranged from 0.01 to 6.65 $mm \cdot d^{-1}$, with a mean of 1.18 $mm \cdot d^{-1}$. For PBIAS, the overall mean is 6.8 ± 33 %, without a significant improvement over the past decade. This analysis also shows that bias in satellite-based ET_a estimation can vary significantly across studies utilizing different model types. Therefore, continuous and consistent uncertainty assessment practice remains important for ET models, especially when they are applied to generate global data products.

The general challenges of uncertainty assessment are exemplified in the WaPOR project (Part II), which presents an interesting case study for exploring the implications of uncertainties for applications in water management. The ETLook model, the core algorithm for generating the WaPOR-ET data, is the interface between satellite observation and derived ET data. The modelling choices in the ETLook model introduce methodological uncertainties when this model is applied worldwide. This is not merely a scientific issue as WaPOR-ET data products are among the most accessible high-resolution ET_a data in many regions and are supported by the FAO and various governments for use in tools that inform farmers and decision-makers, particularly Africa and Southwest Asia. To address these shortcomings, more diverse approaches were

proposed and later explored in Part II, including uncertainty propagation (Chapter 4) and triple collocation (Chapter 5), which offer the ability to quantify the spatiotemporal distribution of uncertainties.

The ex-ante uncertainty assessment of input meteorological data (Chapter 4) is particularly relevant as WaPOR version 3 considers a change of input datasets. This assessment of reanalysis data (GEOS5, ERA5, and AgERA5) for Africa and Southwest Asia revealed that the largest differences between products occur in Central and Southern Africa and Southwest Asia, the very regions where in-situ validation data is scarcest. It also found that, while air temperature and pressure align well with in-situ reference, the datasets tend to overestimate windspeed and solar radiation, and underestimate vapour pressure. The ex-post uncertainty assessment (Chapter 5) then validated the WaPORv3L1 data product against 222 EC flux sites globally. The results showed the WaPORv3L1's performance is generally comparable to other global products, with strong performance in continental and temperate climates. However, the performance was significantly reduced in dry and tropical regions and over challenging land covers like evergreen broadleaf forests and savannas. This finding confirms that compound uncertainty in satellite-based ET_a estimates is highest in data-scarce regions, where they are more relied on. Relative uncertainty was evaluated using triple collocation analysis (TCA), which is valuable for revealing the spatial variation of uncertainty, where reference data is missing. However, the TCA results were overly optimistic, consistently underestimating RMSE and overestimating correlation coefficient compared to in-situ validation. The reliability of TCA was found to be highly dependent on data selection, with common input data having a greater influence than common model structure.

Finally, Chapter 6 explored the implications of uncertainties in a satellite-based ET_a data product for its applications by investigating the problem framings that influence the model choices and configurations, the observed impacts of these choices, and how uncertainties were perceived and managed within the WaPOR project. It shows how these subjective factors are bound by the socio-techno-political context of the WaPOR project. The emphasis on agricultural water use and water productivity by powerful nations and institutions legitimized the development of the WaPOR database and choices that configure the ETLook model to meet the prescribed requirements of the WaPOR database. Uncertainties are mainly perceived and managed by users and modelers predominantly as a technical problem, often overlooking the politically charged standpoint of governing water as a resource intended for optimized productivity.

While some uncertainty arising from technical sources can be quantified partially, other sources of uncertainties stem from subjective factors that cannot be quantified. Therefore, uncertainty quantification is partial and subjective and should only be used to indicate areas for improvement and caution. Moreover, uncertainty assessments are contingent on the specific methods, reference data, and methodological choices employed. Therefore,

modelers and data users must critically evaluate uncertainties, adapt methodologies to the specific context, and reflect on the implications for the specific application of the data.

7.2 OUTLOOK

7.2.1 Development of data quality information

The Committee on Earth Observation Satellites (CEOS)'s Quality Assurance Framework for Earth Observation (QA4EO; http://qa4eo.org/; last access: 19/11/2025) defines five validation stages, from Stage 0 (Product accuracy has not been assessed) to Stage 4 (Uncertainty quantification and characterization are systematically updated when new product version is released and as the time-series expands) (Gruber et al., 2020). This framework should be used to guide the data quality assessment efforts for satellite-based ET_a data products, including WaPOR and those similar to it. According to this framework, the quality assessment of WaPORv2 has reached Stage 2, which implies that product accuracy is estimated over a considerable set of locations and time periods, and the spatial, temporal, and product-to-product consistency with similar data products has been evaluated (Blatchford et al., 2020b; FAO, 2021).

This dissertation contributes to the validation Stage 3 of WaPORv3 by: (1) quantifying uncertainties in the product through in-situ validation; (2) propagating uncertainties from meteorological analysis; and (3) characterizing uncertainties over multiple locations and time periods through triple collocation analysis with similar products. Further development of the WaPOR database should incorporate these estimated uncertainties as data quality layers. For instance, these data quality layers could include the uncertainties estimated by propagating uncertainties in meteorological forcings (Chapter 4) and those derived from triple collocation analysis with other data products (Chapter 5). Such complementary data quality layers would guide users on the reliability of data products within their regions of interest.

The WaPORv3 data quality layers report data gaps in the Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) inputs as an indicator of the data quality. The assumption is that high data quality when these input gaps are minimal or absent. As shown in this research, this assumption is often invalid when the primary source of uncertainty originates elsewhere, such as in meteorological data or the ETLook model itself. Therefore, data quality information should be expanded to include insights from this research's ex-ante and ex-post uncertainty assessments. More importantly, qualitative information, such as known issues and user feedback on specific applications, should be made publicly available on the data repository. This transparency would complement the quality assessment reports and highlight areas for continuous improvement, especially for areas lacking in-situ validation.

The Copernicus Land Monitoring Service (CLMS) has recently adopted the ETLook model, as well as the TSEB-PT model, for the generation of a global ET data product (Guzinski et al., 2025). The CLMS ET data product continues the effort of the WaPOR project in providing global ET data product at 300 m and dekadal resolution. The validation of the CLMS prototype data product was limited in tropical climates, where the number of cloud-free satellite images is the lowest (Guzinski et al., 2025). The uncertainty assessment methods applied in this dissertation could be useful for quality assessment of the CLMS ET product. First, the techniques detailed in Chapter 4 and Chapter 5 can provide uncertainty estimates spatially even in regions lacking in-situ data. Second, these methods can be implemented over continental and global extents using relatively low computational resources, making them suitable for operational data products.

7.2.2 Improving ex-ante uncertainty assessment

Sources of uncertainty that are outside the scope of previous chapters include the calibration and validation of raw satellite imagery. Addressing these uncertainties would greatly increase the scope of research beyond ET estimation to include the satellite observations themselves. Including uncertainties from raw satellite imagery poses challenges in estimating prior uncertainty, as this often requires assumptions regarding their probability distribution. One approach is to make assumptions about the uncertainty distribution of NDVI (e.g., normally distributed with zero mean and constant variance across temporal and spatial domains) and propagate these through the ET_a model. However, these uncertainties can be spatially variable (e.g., LST uncertainty), similar to meteorological forcings (Chapter 4).

Uncertainties in satellite observation, such as reflectance and brightness temperature, depend on the inverse solution of Radiative Transfer Models (RTMs), which simulates how radiation propagates through and interacts with different media, such as the atmosphere, vegetation, or soil (Gallucci et al., 2024). Uncertainties in RTMs are also dynamic, stemming from a combination of input data variability, model approximations, and external factors that change over time and space (Gallucci et al., 2024; Gorroño et al., 2024). Therefore, further ex-ante uncertainty assessments should investigate other critical inputs to ET_a models, including satellite-based LST and vegetation indices.

7.2.3 Improving ex-post uncertainty assessment

To advance ex-post uncertainty assessment methods, future research should prioritize several areas, focusing on advancing triple collocation techniques. This includes applying Multiple Collocation or Extended Collocation for satellite-based estimation of ET_a, for example, building on frameworks developed in the soil moisture community (Gruber et al., 2020). Future work should also use Extended Collocation (Gruber et al., 2016a) to

quantify the impact of non-zero error cross-correlation between datasets. Concurrently, efforts to improve validation references should be continued, especially through the development of benchmark datasets at field/pixel scale (similar to the OpenET benchmark). Such datasets should be operationalized and continuously updated for use in the uncertainty assessment of extended and updated data products.

Recent research and development efforts in remote sensing are focused on improving spatiotemporal resolution, driven by advances in missions, including ECOSTRESS (Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station), SBG (Surface Biology and Geology) (Hook, 2025), LSTM (The Copernicus Land Surface Temperature Monitoring) (Courtier et al., 2025), TRISHNA (Thermal infrared Imaging Satellite for High-resolution Natural resources Assessment) (Hagolle et al., 2025), and Landsat Next (Roy et al., 2026). In addition, small satellite constellations, such as Hydrosat (Hydrosat, 2023) and Copernicus Contributing Missions (Copernicus Data Space Ecosystem, 2025), are increasingly used for field-scale applications. Therefore, further improvements in ex-ante and ex-post uncertainty assessment methods are needed to also address the challenges of evaluating satellite-based ET_a estimates at high resolution (less than 100 m).

7.3 REFLECTIONS

7.3.1 Ontological and epistemological commitments

Knowledge creation is fundamentally shaped by ontology and epistemology. Ontology defines what exists and what is recognized as a meaningful subject of knowledge. Epistemology defines what can be known, how it can be known, and what counts as valid knowledge. To critically reflect on my ontological and epistemological commitments, I initially formulated two questions:

- How do I know uncertainty exists in satellite remote sensing?
- How do I conceptualize and acquire knowledge about uncertainty?

As a practitioner of water accounting (defined here as the systematic acquisition of information about water fluxes, flows, and storage), I used satellite remote sensing data products for three years prior to this PhD research. Through this practical experience, I was exposed to the inconsistencies between various satellite-based data products and other sources of data. These discrepancies lead to scepticisms and doubts among data users (including myself) regarding the information these data products represent. I interpreted these observed inconsistencies and doubts as ‘uncertainty’. This heuristic interpretation serves as my ontological commitment to the existence of uncertainties in satellite-based data.

Regarding the conceptualization of uncertainty and the nature of knowledge about it, I applied my prior training in probability, statistics, and hydrology. I conceptualized uncertainty as the probability distribution of errors derived from repeated measurements. Consequently, the uncertainty of a dependent variable can be propagated from the uncertainties in independent variables. However, the applicability of probability theory is limited in the field of remote sensing for several reasons. First, every satellite retrieval constitutes a single, one-time measurement. There are no practical repeated measurements of surface reflectance or radiometric temperature at the specific time of satellite overpass. While pre-mission calibration can estimate uncertainty in satellite retrieval, this process does not fully account for conditions onboard satellites or throughout the mission lifespan. Second, space agencies still face challenges in providing rigorous uncertainty estimates for derived data products that extend beyond the sensor specification (e.g., the accuracy of a Thermal Infrared (TIR) sensor calibrated in lab conditions). Without reliable uncertainty estimates for satellite retrieval (e.g., the accuracy of a TIR sensor in mission conditions), probability theory would propagate uncertainty estimates that are not grounded in measured or validated statistics.

Acknowledging these limitations, my adoption of probability theory represents an epistemological commitment to this dominant approach to uncertainty within the natural sciences and technology. Within the probability framework, uncertainty assessment is intrinsically a modelling practice, which entails conceptualization (of uncertainty), formalization (selection of appropriate models and parameters), implementation, and evaluation of the estimated uncertainties. Therefore, the outputs of uncertainty assessment are also contingent upon these modelling choices. For that reason, I chose to explore and compare several uncertainty assessment methods (Chapter 4 and 5), while also analysing the methods and results detailed in the literature (Chapter 2 and 3).

The value of probability theory lies in its ability to offer a quantitative approach for assessing how substantial or negligible uncertainty is, typically by examining standard errors or the spread of the probability distribution. However, this type of uncertainty information is not directly actionable and requires a decision-support model to relate the uncertainty information to optimal actions and the probability of consequences (Gruber et al., 2025). Furthermore, there are facets of uncertainty that cannot be described by probability theory, such as complete ignorance (total lack of information), ambiguity about which model to use, and subjective modelling choices (Beven, 2016; Melsen et al., 2019). This study applied this epistemology of uncertainty to improve understanding and support its effective operationalization in practice. However, uncertainty quantification does not aim to reduce or eliminate uncertainties; instead, it allows for reflection on the limits of knowledge.

7.3.2 Interdisciplinary research

During the final two years of my PhD research, I participated in several interdisciplinary programs, including the Constructive Advanced Thinking program, the Situated Hydrological Modelling summer school at IRI THESys (Integrative Research Institute on Transformations of Human-Environment Systems), and the Decolonising Science course at IHE Delft. These opportunities facilitated interdisciplinary conversations regarding the practices and impacts of remote sensing and quantitative modelling. The outcomes of this engagement were twofold. First, the disagreements and points of confusion encountered in these conversations helped to situate my research within the specific context of the WaPOR project and its broader social and political dimensions. Second, this process led to a reflection on my own previously discussed ontological and epistemological commitments.

My collaboration with two human geographers and a political scientist within the Constructive Advanced Thinking program focused on investigating the practices of modelling in water governance (Alba et al., 2025a). Through this interdisciplinary engagement, I was introduced to concepts such as reflexivity (Krueger and Alba, 2022), power sensitivity (ter Horst et al., 2023), and situated modelling (Klein et al., 2024). The primary outcome of this collaboration was our proposal to engage with the power of models by situating modelling practices (Alba et al., 2025b). We argued that situating modelling within its broader context entails examining where and how models are produced, who develops and applies them, for whom they are intended, and how their results shape reality.

In Chapter 6, I operationalized reflexivity to investigate methodological uncertainties inherent in satellite-based estimation of ET, specifically those uncertainties that cannot be quantified. These reflections shifted the focus from the limitations of technical uncertainty assessment methods discussed in Chapter 1 to the methodological uncertainties that remain unaddressed in literature. This approach also helps to interrelate quantified uncertainty with unquantifiable uncertainty, without integrating them into a single unified framework.

REFERENCES

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci. Data 5, 170191. https://doi.org/10.1038/sdata.2017.191

Abramowitz, G., Gupta, H., 2008. Toward a model space and model independence metric. Geophys. Res. Lett. 35. https://doi.org/10.1029/2007GL032834

Ahmad, S., McElwain, J., Nair, R., Gill, L., 2025. Comment on "On the Use of the Term 'Evapotranspiration'" by Miralles et al. (2020). Water Resour. Res. 61, e2024WR038835. https://doi.org/10.1029/2024WR038835

Alba, R., Horst, R. ter, Madrigal, J.G., Tran, B., 2025a. Modelling practices in water governance [WWW Document]. CAT - Water Models. URL https://cat-water-models.github.io/ (accessed 11.19.25).

Alba, R., ter Horst, R., Tran, B.N., Klein, A., Unverzagt, K., Godinez-Madrigal, J., Verzijl, A., Rusca, M., Vos, J., Venot, J.-P., Zwarteveen, M., Krueger, T., 2025b. Situating Hydrological Modeling: A Proposal for Engaging With the Power of Models. WIREs Water 12, e70030. https://doi.org/10.1002/wat2.70030

Al-Bakri, J.T., D'Urso, G., Batchelor, C., Abukhalaf, M., Alobeiaat, A., Al-Khreisat, A., Vallee, D., 2022. Remote Sensing-Based Agricultural Water Accounting for the North Jordan Valley. Water 14, 1198. https://doi.org/10.3390/w14081198

Alfieri, J.G., Anderson, M.C., Kustas, W.P., Cammalleri, C., 2017. Effect of the revisit interval and temporal upscaling methods on the accuracy of remotely sensed evapotranspiration estimates. Hydrol. Earth Syst. Sci. 21, 83–98. https://doi.org/10.5194/hess-21-83-2017

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

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011b. Evapotranspiration information reporting: II. Recommended documentation. Agric. Water Manag. 98, 921–929. https://doi.org/10.1016/j.agwat.2010.12.016

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - Guidelines for computing crop water requirements, FAO Irrigation and drainage paper. FAO - Food and Agriculture Organization of the United Nations, Rome.

Allen, R.G., Pruitt, W.O., Wright, J.L., Howell, T.A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Yrisarry, J.B., Smith, M., Pereira, L.S., Raes, D., Perrier, A., Alves, I., Walter, I., Elliott, R., 2006. A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO56 Penman-Monteith method. Agric. Water Manag. 81, 1–22. https://doi.org/10.1016/j.agwat.2005.03.007

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

Al-Omoush, R.A., Al-Bakri, J.T., Abdelal, Q., Al-Kilani, M.R., Hamdan, I., Aljarrah, A., 2025. Developing a Remote Sensing-Based Approach for Agriculture Water Accounting in the Amman–Zarqa Basin. Water 17, 2106. https://doi.org/10.3390/w17142106

Amdar, N., Seyoum, S., Al-Bakri, J., Rutten, M., Jewitt, G., Mul, M., 2024. Developing a water budget for the Amman-Zarqa basin using water accounting plus and the pixel-based soil water balance model. Model. Earth Syst. Environ. https://doi.org/10.1007/s40808-024-02159-0

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

Annor, F.O., 2023. Small Reservoirs in Northern Ghana: Monitoring, Physical Processes, and Management. Delft University of Technology, Delft, The Netherlands. https://doi.org/10.4233/UUID:81E5E8A8-2BEE-4AF5-B1BC-C7B210F9CB55

ASCE, 2005. The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil Engineers, Reston, VA.

ASCE, 1930. Consumptive Use of Water in Irrigation. Trans. Am. Soc. Civ. Eng. 94, 1349–1377. https://doi.org/10.1061/TACEAT.0004220

Badgley, G., Fisher, J.B., Jiménez, C., Tu, K.P., Vinukollu, R., 2015. On Uncertainty in Global Terrestrial Evapotranspiration Estimates from Choice of Input Forcing Datasets. J. Hydrometeorol. 16, 1449–1455. https://doi.org/10.1175/JHM-D-14-0040.1

Bahrami, H., McNairn, H., Mahdianpari, M., Homayouni, S., 2022. A Meta-Analysis of Remote Sensing Technologies and Methodologies for Crop Characterization. Remote Sens. 14, 5633. https://doi.org/10.3390/rs14225633

Baik, J., Liaqat, U.W., Choi, M., 2018. Assessment of satellite- and reanalysis-based evapotranspiration products with two blending approaches over the complex landscapes and climates of Australia. Agric. For. Meteorol. 263, 388–398. https://doi.org/10.1016/j.agrformet.2018.09.007

Bambach, N., Kustas, W., Alfieri, J., Prueger, J., Hipps, L., McKee, L., Castro, S.J., Volk, J., Alsina, M.M., McElrone, A.J., 2022. Evapotranspiration uncertainty at micrometeorological scales: the impact of the eddy covariance energy imbalance and correction methods. Irrig. Sci. https://doi.org/10.1007/s00271-022-00783-1

Barideh, R., Nasimi, F., 2022. Investigating the changes in agricultural land use and actual evapotranspiration of the Urmia Lake basin based on FAO's WaPOR database. Agric. Water Manag. 264, 107509. https://doi.org/10.1016/j.agwat.2022.107509

Barraza Bernadas, V., Grings, F., Restrepo-Coupe, N., Huete, A., 2018a. Comparison of the performance of latent heat flux products over southern hemisphere forest ecosystems: estimating latent heat flux error structure using in situ measurements and the triple collocation method. Int. J. Remote Sens. 39, 6300–6315. https://doi.org/10.1080/01431161.2018.1458348

Barraza Bernadas, V., Grings, F., Restrepo-Coupe, N., Huete, A., 2018b. Comparison of the performance of latent heat flux products over southern hemisphere forest ecosystems: estimating latent heat flux error structure using in situ measurements and the triple collocation method. Int. J. Remote Sens. 39, 6300–6315. https://doi.org/10.1080/01431161.2018.1458348

Barsi, Á., Kugler, Z., Juhász, A., Szabó, G., Batini, C., Abdulmuttalib, H., Huang, G., Shen, H., 2019. Remote sensing data quality model: from data sources to lifecycle

phases. Int. J. Image Data Fusion 10, 280–299. https://doi.org/10.1080/19479832.2019.1625977

Bastiaanssen, W.G.M., Cheema, M.J.M., Immerzeel, W.W., Miltenburg, I.J., Pelgrum, H., 2012. Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model. Water Resour. Res. 48. https://doi.org/10.1029/2011WR010482

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

Bayat, B., Camacho, F., Nickeson, J., Cosh, M., Bolten, J., Vereecken, H., Montzka, C., 2021. Toward operational validation systems for global satellite-based terrestrial essential climate variables. Int. J. Appl. Earth Obs. Geoinformation 95, 102240. https://doi.org/10.1016/j.jag.2020.102240

Bellocchi, G., Rivington, M., Donatelli, M., Matthews, K., 2011. Validation of Biophysical Models: Issues and Methodologies, in: Lichtfouse, E., Hamelin, M., Navarrete, M., Debaeke, P. (Eds.), Sustainable Agriculture Volume 2. Springer Netherlands, Dordrecht, pp. 577–603. https://doi.org/10.1007/978-94-007-0394-0_26

Ben Mehrez, M., Taconet, O., Vidal-Madjar, D., Valencogne, C., 1992. Estimation of stomatal resistance and canopy evaporation during the HAPEX-MOBILHY experiment. Agric. For. Meteorol. 58, 285–313. https://doi.org/10.1016/0168-1923(92)90066-D

Bennett, M.M., 2025. Satellite data, information, or knowledge? Critiquing how Arctic environmental NGOs derive meaning and power from imagery. Digit. Geogr. Soc. 8, 100116. https://doi.org/10.1016/j.diggeo.2025.100116

Bessembinder, J.J.E., Leffelaar, P.A., Dhindwal, A.S., Ponsioen, T.C., 2005. Which crop and which drop, and the scope for improvement of water productivity. Agric. Water Manag. 73, 113–130. https://doi.org/10.1016/j.agwat.2004.10.004

Beven, K., 2016. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol. Sci. J. 61, 1652–1665. https://doi.org/10.1080/02626667.2015.1031761

Beven, K., 2002. Towards a coherent philosophy for modelling the environment. Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci. 458, 2465–2484. https://doi.org/10.1098/rspa.2002.0986

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16, 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Beyrich, F., Hartogensis, O.K., de Bruin, H.A.R., Ward, H.C., 2021. Scintillometers, in: Foken, T. (Ed.), Springer Handbook of Atmospheric Measurements. Springer International Publishing, Cham, pp. 969–997. https://doi.org/10.1007/978-3-030-52171-4_34

Bhattarai, N., Mallick, K., Stuart, J., Vishwakarma, B.D., Niraula, R., Sen, S., Jain, M., 2019. An automated multi-model evapotranspiration mapping framework using remotely sensed and reanalysis data. Remote Sens. Environ. 229, 69–92. https://doi.org/10.1016/j.rse.2019.04.026

Bielecka, E., Burek, E., 2019. Spatial data quality and uncertainty publication patterns and trends by bibliometric analysis. Open Geosci. 11, 219–235. https://doi.org/10.1515/geo-2019-0018

Bierkens, M., Finke, P., De Willigen, P., 2000. Upscaling and downscaling methods for environmental research. Kluwer Academic.

Bisquert, M., Sánchez, J.M., López-Urrea, R., Caselles, V., 2016. Estimating high resolution evapotranspiration from disaggregated thermal images. Remote Sens. Environ. 187, 423–433. https://doi.org/10.1016/j.rse.2016.10.049

Blatchford, M., M. Mannaerts, C., Zeng, Y., Nouri, H., Karimi, P., 2020a. Influence of Spatial Resolution on Remote Sensing-Based Irrigation Performance Assessment Using WaPOR Data. Remote Sens. 12, 2949. https://doi.org/10.3390/rs12182949

Blatchford, M., Mannaerts, C.M., Njuki, S.M., Nouri, H., Zeng, Y., Pelgrum, H., Wonink, S., Karimi, P., 2020b. Evaluation of WaPOR V2 evapotranspiration products across Africa. Hydrol. Process. 34, 3200–3221. https://doi.org/10.1002/hyp.13791

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: A review. Hydrol. Process. 9, 251–290. https://doi.org/10.1002/hyp.3360090305

Boelens, R., Vos, J., 2012. The danger of naturalizing water policy concepts: Water productivity and efficiency discourses from field irrigation to virtual water trade. Agric. Water Manag., Irrigation efficiency and productivity: scales, systems and science 108, 16–26. https://doi.org/10.1016/j.agwat.2011.06.013

Boergens, E., Kvas, A., Eicker, A., Dobslaw, H., Schawohl, L., Dahle, C., Murböck, M., Flechtner, F., 2022. Uncertainties of GRACE-Based Terrestrial Water Storage Anomalies for Arbitrary Averaging Regions. J. Geophys. Res. Solid Earth 127, e2021JB022081. https://doi.org/10.1029/2021JB022081

Boisot, M., Canals, A., 2004. Data, information and knowledge: have we got it right? J. Evol. Econ. 14, 43–67. https://doi.org/10.1007/s00191-003-0181-9

Boogaard, H., de Wit, A., Lazebnik, J., Schubert, J., Grijn, G. van der, 2024. Data Stream 2: AgERA5 historic and near real time forcing data: Algorithm Theoretical Basis (ATBD) [WWW Document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. URL https://confluence.ecmwf.int/pages/viewpage.action?pageId=278550984 (accessed 8.12.24).

Boogaard, H., Schubert, J., de Wit, A., Lazebnik, J., Hutjes, R., Van der Grijn, G., 2020. Agrometeorological indicators from 1979 to present derived from reanalysis. https://doi.org/10.24381/cds.6c68c9bb

Boogaard, H., Van der Grijn, G., Schubert, J., 2023. Global Agriculture Downscaling and bias correction [WWW Document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. URL https://confluence.ecmwf.int/display/CKB/Global+Agriculture+Downscaling+and+bias+correction (accessed 8.6.24).

Braun, A.C., 2021. More accurate less meaningful? A critical physical geographer's reflection on interpreting remote sensing land-use analyses. Prog. Phys. Geogr. Earth Environ. 45, 706–735. https://doi.org/10.1177/0309133321991814

Brombacher, J., Silva, I.R. de O., Degen, J., Pelgrum, H., 2022. A novel evapotranspiration based irrigation quantification method using the hydrological similar pixels algorithm. Agric. Water Manag. 267, 107602. https://doi.org/10.1016/j.agwat.2022.107602

Brönnimann, S., Allan, R., Atkinson, C., Buizza, R., Bulygina, O., Dahlgren, P., Dee, D., Dunn, R., Gomes, P., John, V.O., Jourdain, S., Haimberger, L., Hersbach, H., Kennedy, J., Poli, P., Pulliainen, J., Rayner, N., Saunders, R., Schulz, J., Sterin, A.,

Stickler, A., Titchner, H., Valente, M.A., Ventura, C., Wilkinson, C., 2018. Observations for Reanalyses. https://doi.org/10.1175/BAMS-D-17-0229.1

Brutsaert, W., 1982. Evaporation into the Atmosphere. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-017-1497-6

Brutsaert, W., Parlange, M.B., 1998. Hydrologic cycle explains the evaporation paradox. Nature 396, 30–30. https://doi.org/10.1038/23845

Brutsaert, W., Stricker, H., 1979. An advection-aridity approach to estimate actual regional evapotranspiration. Water Resour. Res. 15, 443–450. https://doi.org/10.1029/WR015i002p00443

Budyko, M.I., 1974. Climate and Life. Academic Press.

Burchard-Levine, V., Nieto, H., Riaño, D., Migliavacca, M., El-Madany, T.S., Guzinski, R., Carrara, A., Martín, M.P., 2021. The effect of pixel heterogeneity for remote sensing based retrievals of evapotranspiration in a semi-arid tree-grass ecosystem. Remote Sens. Environ. 260, 112440. https://doi.org/10.1016/j.rse.2021.112440

Burchard-Levine, V., Nieto, H., Riaño, D., Migliavacca, M., El-Madany, T.S., Perez-Priego, O., Carrara, A., Martín, M.P., 2020. Seasonal adaptation of the thermal-based two-source energy balance model for estimating evapotranspiration in a semiarid tree-grass ecosystem. Remote Sens. 12. https://doi.org/10.3390/rs12060904

Byun, K., Liaqat, U.W., Choi, M., 2014. Dual-model approaches for evapotranspiration analyses over homo- and heterogeneous land surface conditions. Agric. For. Meteorol. 197, 169–187. https://doi.org/10.1016/j.agrformet.2014.07.001

Camillo, P.J., Gurney, R.J., 1986. A resistance parameter for bare-soil evaporation models. Soil Sci. 141, 95.

Campbell, J.B., Wynne, R.H., 2011. Introduction to Remote Sensing, Fifth Edition. ed. The Guilford Press, New York, NY.

Cao, M., Wang, W., Xing, W., Wei, J., Chen, X., Li, J., Shao, Q., 2021. Multiple sources of uncertainties in satellite retrieval of terrestrial actual evapotranspiration. J. Hydrol. 601. https://doi.org/10.1016/j.jhydrol.2021.126642

Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens. Environ. 62, 241–252. https://doi.org/10.1016/S0034-4257(97)00104-1

Cawse-Nicholson, K., Braverman, A., Kang, E.L., Li, M., Johnson, M., Halverson, G., Anderson, M., Hain, C., Gunson, M., Hook, S., 2020. Sensitivity and uncertainty quantification for the ECOSTRESS evapotranspiration algorithm – DisALEXI. Int. J. Appl. Earth Obs. Geoinformation 89. https://doi.org/10.1016/j.jag.2020.102088

Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. Remote Sens. Environ. 237, 111594. https://doi.org/10.1016/j.rse.2019.111594

Chen, X., Su, Z., Ma, Y., Middleton, E.M., 2019. Optimization of a remote sensing energy balance method over different canopy applied at global scale. Agric. For. Meteorol. 279. https://doi.org/10.1016/j.agrformet.2019.107633

Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J.B., Li, Xin, Li, Xianglan, Liu, S., Ma, Z., Miyata, A., Mu, Q., Sun, L., Tang, J., Wang, K., Wen, J., Xue, Y., Yu, G., Zha, T., Zhang, L., Zhang, Q., Zhao, T., Zhao, L., Yuan, W., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens. Environ. 140, 279–293. https://doi.org/10.1016/j.rse.2013.08.045

Chen, Y., Yuan, W., Xia, J., Fisher, J.B., Dong, W., Zhang, X., Liang, S., Ye, A., Cai, W., Feng, J., 2015. Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. J. Hydrol. 528, 537–549. https://doi.org/10.1016/j.jhydrol.2015.06.059

Choudhury, B.J., Reginato, R.J., Idso, S.B., 1986. An analysis of infrared temperature observations over wheat and calculation of latent heat flux. Agric. For. Meteorol. 37, 75–88. https://doi.org/10.1016/0168-1923(86)90029-8

Chu, H., Luo, X., Ouyang, Z., Chan, W.S., Dengel, S., Biraud, S.C., Torn, M.S., Metzger, S., Kumar, J., Arain, M.A., Arkebauer, T.J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N.A., Chen, J., Chen, X., Clark, K., Desai, A.R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J.A., Gough, C.M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J.F., Knox, S.H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J.W., Noormets, A., Novick, K., Oberbauer, S.F., Oechel, W., Oikawa, P., Papuga, S.A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W.L., Richardson, A.D., Russell, E.S., Scott, R.L., Starr, G., Staebler, R., Stoy, P.C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R.C., Suyker, A., Ueyama, M., Vargas, R., Wood, J.D., Zona, D., 2021. Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. Agric. For. Meteorol. 301–302, 108350. https://doi.org/10.1016/j.agrformet.2021.108350

Chukalla, A.D., Mul, M.L., Karimi, P., 2024. Establishing the water resources implications for closing the land and water productivity gaps using remote sensing – A case study of sugarcane. Field Crops Res. 318, 109589. https://doi.org/10.1016/j.fcr.2024.109589

Chukalla, A.D., Mul, M.L., van der Zaag, P., van Halsema, G., Mubaya, E., Muchanga, E., den Besten, N., Karimi, P., 2022. A framework for irrigation performance assessment using WaPOR data: the case of a sugarcane estate in Mozambique. Hydrol. Earth Syst. Sci. 26, 2759–2778. https://doi.org/10.5194/hess-26-2759-2022

Contractor, Q.A., El-Wattar, S., Joshi, D., Schmitter, P., Verzjil, A., Mul, M., 2023. WaPOR Gender Equality and Social Inclusion Strategy [unpublished report].

Copernicus Climate Change Service, 2020. Agrometeorological indicators from 1979 to present derived from reanalysis.

Copernicus Data Space Ecosystem, 2025. Copernicus Contributing Missions | Copernicus Data Space Ecosystem [WWW Document]. URL https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions (accessed 11.19.25).

Corbari, C., Paciolla, N., Sheffield, J., Labbassi, K., Dos Santos Araujo, D.C., Berendsen, S., Szantoi, Z., 2025. Estimates of Irrigation Water Volume by Assimilation of Satellite Land Surface Temperature or Soil Moisture Into a Water-Energy Balance Model in Morocco. Water Resour. Res. 61, e2024WR038926. https://doi.org/10.1029/2024WR038926

Courault, D., Seguin, B., Olioso, A., 2005. Review on estimation of evapotranspiration from remote sensing data: From empirical to numerical modeling approaches. Irrig. Drain. Syst. 19, 223–249. https://doi.org/10.1007/s10795-005-5186-0

Courtier, B., Perry, M., Soszynska, A., Ghent, D., 2025. The Land Surface Temperature Retrieval Algorithmfor LSTM: An Overview and Results.

Cressie, N.A.C. (Ed.), 1993. Statistics for spatial data, Revised edition. ed, Wiley series in probability and mathematical statistics Applied probability and statistics. Wiley, New York.

Crosetto, M., Moreno Ruiz, J.A., Crippa, B., 2001. Uncertainty propagation in models driven by remotely sensed data. Remote Sens. Environ. 76, 373–385. https://doi.org/10.1016/S0034-4257(01)00184-5

Cui, W., Chui, T.F.M., 2019. Temporal and spatial variations of energy balance closure across FLUXNET research sites. Agric. For. Meteorol. 271, 12–21. https://doi.org/10.1016/j.agrformet.2019.02.026

de Bruin, H.A.R., Trigo, I.F., Bosveld, F.C., Meirink, J.F., 2016. A Thermodynamically Based Model for Actual Evapotranspiration of an Extensive Grass Field Close to FAO Reference, Suitable for Remote Sensing Application. https://doi.org/10.1175/JHM-D-15-0006.1

Ding, R., Kang, S., Li, F., Zhang, Y., Tong, L., Sun, Q., 2010. Evaluating eddy covariance method by large-scale weighing lysimeter in a maize field of northwest China. Agric. Water Manag. 98, 87–95. https://doi.org/10.1016/j.agwat.2010.08.001

Dinku, T., 2019. Chapter 7 - Challenges with availability and quality of climate data in Africa, in: Melesse, A.M., Abtew, W., Senay, G. (Eds.), Extreme Hydrology and Climate Variability. Elsevier, pp. 71–80. https://doi.org/10.1016/B978-0-12-815998-9.00007-5

Dolman, A.J., 1993. A multiple-source land surface energy balance model for use in general circulation models. Agric. For. Meteorol. 65, 21–45. https://doi.org/10.1016/0168-1923(93)90036-H

Doorenbos, J., Pruitt, W.O., 1977. Guidelines for predicting crop water requirements, Rev. ed, FAO irrigation and drainage paper ; 24. Food and Agriculture Organization of the United Nations, Rome.

Drought 2018 Team, ICOS Ecosystem Thematic Centre, 2020. Drought-2018 ecosystem eddy covariance flux product for 52 stations in FLUXNET-Archive format. https://doi.org/10.18160/YVR0-4898

Du, X., Wu, X., Tang, R., Zeng, Q., Li, Z., Wang, J., Jiang, Z., Wang, K., You, D., Wen, J., Xiao, Q., 2023. An Improved Upscaling Method of In Situ Measurements With Consideration of Their Uncertainty for the Spatial Scale Match Between Satellite and In Situ Measurements. IEEE Trans. Geosci. Remote Sens. 61, 1–17. https://doi.org/10.1109/TGRS.2023.3291883

Dunn, O.J., 1964. Multiple Comparisons Using Rank Sums. Technometrics 6, 241–252. https://doi.org/10.1080/00401706.1964.10490181

Elhag, M., 2016. Inconsistencies of SEBS Model Output Based on the Model Inputs: Global Sensitivity Contemplations. J. Indian Soc. Remote Sens. 44, 435–442. https://doi.org/10.1007/s12524-015-0502-0

Elnashar, A., Wang, L., Wu, B., Zhu, W., Zeng, H., 2021. Synthesis of global actual evapotranspiration from 1982 to 2019. Earth Syst. Sci. Data 13, 447–480. https://doi.org/10.5194/essd-13-447-2021

Eltahir, E.A.B., Bras, R.L., 1996. Precipitation recycling. Rev. Geophys. 34, 367–378. https://doi.org/10.1029/96RG01927

El-Wattar, S., 2022. Entangled logics of water and land productivity: conversations among Egyptian farmers and WaPOR's remote sensing data (MSc thesis). IHE Delft Institute for Water Education.

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

ESA, 2021. User Guides - Sentinel-2 MSI - Processing Levels [WWW Document]. URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/processing-levels (accessed 9.7.21).

European Centre for Medium-Range Weather Forecasts (ECMWF), 2023a. ERA5: data documentation [WWW Document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. URL https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation (accessed 7.29.24).

European Centre for Medium-Range Weather Forecasts (ECMWF), 2023b. ERA5: uncertainty estimation [WWW Document]. Copernic. Knowl. Base - ECMWF Conflu. Wiki. URL https://confluence.ecmwf.int/display/CKB/ERA5%3A+uncertainty+estimation (accessed 7.29.24).

Evaristo, J., McDonnell, J.J., 2017. A role for meta-analysis in hydrology. Hydrol. Process. 31, 3588–3591. https://doi.org/10.1002/hyp.11253

Eyre, J.R., Bell, W., Cotton, J., English, S.J., Forsythe, M., Healy, S.B., Pavelin, E.G., 2022. Assimilation of satellite data in numerical weather prediction. Part II: Recent years. Q. J. R. Meteorol. Soc. 148, 521–556. https://doi.org/10.1002/qj.4228

Falkenmark, M., Rockström, J., 2004. Balancing Water for Humans and Nature: The New Approach in Ecohydrology. Earthscan.

FAO, 2025a. pyWaPOR – A Python implementation of ETLook algorithm (v.3.6.0).

FAO, 2025b. Actual evapotranspiration and interception (Global - Dekadal - 300 m) - WaPOR v3 Description.

FAO, 2024. Reference Evapotranspiration (Global - Daily - Approximately 30km) - WaPOR v3 - “FAO catalog” [WWW Document]. URL https://data.apps.fao.org/catalog//iso/6f487301-9731-4317-b620-faeccc976443 (accessed 7.18.24).

FAO, 2023a. Remote sensing determination of evapotranspiration – Algorithms, strengths, weaknesses, uncertainty and best fit-forpurpose. FAO, Cairo. https://doi.org/10.4060/cc8150en

FAO, 2023b. FAO Statistics and Data Quality Assurance Framework.

FAO, 2021. WaPOR V2 Quality Assessment. Technical report on the data quality of the WaPOR FAO database version 2. FOOD & AGRICULTURE ORG, S.l.

FAO, 2020a. WaPOR database methodology.

FAO, 2020b. Mid-term evaluation of the project "Monitoring water productivity by remote sensing as a tool to assess possibilities to reduce water productivity gaps.

FAO, 2018a. WaPOR Database Methodology: Level 1. Remote Sensing for Water Productivity. Food and Agriculture Organization of the United Nations, Rome.

FAO, 2018b. WaPOR Database Methodology: Level 2. Remote Sensing for Water Productivity. Rome.

FAO, IHE Delft, 2024. WaPOR Quality Assessment: Technical report evaluating WaPOR v3. IHE Delft Institute for Water Education, Delft, The Netherlands.

FAO, IHE Delft, 2020. Water Accounting in the Niger River Basin. FAO, Rome. https://doi.org/10.4060/cb1274en

FAO, IHE Delft, 2019. WaPOR quality assessment. Technical report on the data quality of the WaPOR FAO database version 1.0. Rome.

Feng, J., Wang, Z., 2013. A satellite-based energy balance algorithm with reference dry and wet limits. Int. J. Remote Sens. 34, 2925–2946. https://doi.org/10.1080/01431161.2012.748990

Ferguson, C.R., Sheffield, J., Wood, E.F., Gao, H., 2010. Quantifying uncertainty in a remote sensing-based estimate of evapotranspiration over continental USA. Int. J. Remote Sens. 31, 3821–3865. https://doi.org/10.1080/01431161.2010.483490

FEWS NET, 2024. SSEBop Dekadal Actual Evapotranspiration Products (Version 6.1).

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J.-H., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resour. Res. 56, e2019WR026058. https://doi.org/10.1029/2019WR026058

Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M.F., Hook, S., Baldocchi, D., Townsend, P.A., Kilic, A., Tu, K., Miralles, D.D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A.J., French, A., Schimel, D., Famiglietti, J.S., Stephens, G., Wood, E.F., 2017. The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resour. Res. 53, 2618–2626. https://doi.org/10.1002/2016WR020175

Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112, 901–919. https://doi.org/10.1016/j.rse.2007.06.025

Fisher, J.B., Whittaker, R.J., Malhi, Y., 2011. ET come home: potential evapotranspiration in geographical ecology. Glob. Ecol. Biogeogr. 20, 1–18. https://doi.org/10.1111/j.1466-8238.2010.00578.x

FLUXNET, 2017. FLUXNET - Site Summary [WWW Document]. FLUXNET. URL https://fluxnet.org/sites/site-summary/ (accessed 1.20.23).

Foken, T., 2008. The Energy Balance Closure Problem: An Overview. Ecol. Appl. 18, 1351–1367.

Foken, T., Aubinet, M., Leuning, R., 2012. The Eddy Covariance Method, in: Aubinet, M., Vesala, T., Papale, D. (Eds.), Eddy Covariance: A Practical Guide to Measurement and Data Analysis, Springer Atmospheric Sciences. Springer Netherlands, Dordrecht, pp. 1–19. https://doi.org/10.1007/978-94-007-2351-1_1

Foody, G.M., Atkinson, P.M., 2003. Uncertainty in Remote Sensing and GIS. John Wiley & Sons.

Foster, T., Mieno, T., Brozović, N., 2020. Satellite-Based Monitoring of Irrigation Water Use: Assessing Measurement Errors and Their Implications for Agricultural Water

Management Policy. Water Resour. Res. 56, e2020WR028378. https://doi.org/10.1029/2020WR028378

Frehlich, R., 1992. Laser Scintillation Measurements of the Temperature Spectrum in the Atmospheric Surface Layer. J. Atmospheric Sci. 49, 1494–1509. https://doi.org/10.1175/1520-0469(1992)049%253C1494:LSMOTT%253E2.0.CO;2

Fuentes, I., Vervoort, R.W., McPhee, J., 2024. Global evapotranspiration models and their performance at different spatial scales: Contrasting a latitudinal gradient against global catchments. J. Hydrol. 628, 130477. https://doi.org/10.1016/j.jhydrol.2023.130477

Gallucci, D., Cimini, D., Turner, E., Fox, S., Rosenkranz, P.W., Tretyakov, M.Y., Mattioli, V., Larosa, S., Romano, F., 2024. Uncertainty in simulated brightness temperature due to sensitivity to atmospheric gas spectroscopic parameters from the centimeter- to submillimeter-wave range. Atmospheric Chem. Phys. 24, 7283–7308. https://doi.org/10.5194/acp-24-7283-2024

Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F.H.S., Yu, Q., 2018. Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. Ecohydrology 11, e1974. https://doi.org/10.1002/eco.1974

García, M., Sandholt, I., Ceccato, P., Ridler, M., Mougin, E., Kergoat, L., Morillas, L., Timouk, F., Fensholt, R., Domingo, F., 2013. Actual evapotranspiration in drylands derived from in-situ and satellite data: Assessing biophysical constraints. Remote Sens. Environ. 131, 103–118. https://doi.org/10.1016/j.rse.2012.12.016

Gentine, P., Entekhabi, D., Chehbouni, A., Boulet, G., Duchemin, B., 2007. Analysis of evaporative fraction diurnal behaviour. Agric. For. Meteorol. 143, 13–29. https://doi.org/10.1016/j.agrformet.2006.11.002

Giordano, M., Scheierling, S.M., Tréguer, D.O., Turral, H., McCornick, P.G., 2021. Moving beyond 'more crop per drop': insights from two decades of research on agricultural water productivity. Int. J. Water Resour. Dev.

Glenn, E.P., Doody, T.M., Guerschman, J.P., Huete, A.R., King, E.A., McVicar, T.R., Dijk, A.I.J.M.V., Niel, T.G.V., Yebra, M., Zhang, Y., 2011. Actual evapotranspiration estimation by ground and remote sensing methods: the Australian experience. Hydrol. Process. 25, 4103–4116. https://doi.org/10.1002/hyp.8391

Global Modeling and Assimilation Office (GMAO), 2020. GEOS FP to be Updated on April 7, 2020 [WWW Document]. GMAO - Glob. Model. Assim. Off. Res. Site. URL https://gmao.gsfc.nasa.gov/news/geos_system_news/2020/GEOS_FP_upgrade_5_25_1p5.php (accessed 8.6.24).

Gomis-Cebolla, J., Jimenez, J.C., Sobrino, J.A., Corbari, C., Mancini, M., 2019. Intercomparison of remote-sensing based evapotranspiration algorithms over amazonian forests. Int. J. Appl. Earth Obs. Geoinformation 80, 280–294. https://doi.org/10.1016/j.jag.2019.04.009

González-Gambau, V., Turiel, A., González-Haro, C., Martínez, J., Olmedo, E., Oliva, R., Martín-Neira, M., 2020. Triple Collocation Analysis for Two Error-Correlated Datasets: Application to L-Band Brightness Temperatures over Land. Remote Sens. 12, 3381. https://doi.org/10.3390/rs12203381

Gorroño, J., Guanter, L., Valentin Graf, L., Gascon, F., 2024. A Framework for the Estimation of Uncertainties and Spectral Error Correlation in Sentinel-2 Level-2A

Data Products. IEEE Trans. Geosci. Remote Sens. 62, 1–13. https://doi.org/10.1109/TGRS.2024.3435021

Gruber, A., Bulgin, C.E., Dorigo, W., Embury, O., Formanek, M., Merchant, C., Mittaz, J., Muñoz-Sabater, J., Pöppl, F., Povey, A., Wagner, W., 2025. Making Sense of Uncertainties: Ask the Right Question. Surv. Geophys. https://doi.org/10.1007/s10712-025-09889-5

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.-C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., van der Schalie, R., Wigneron, J.-P., Wagner, W., 2020. Validation practices for satellite soil moisture retrievals: What are (the) errors? Remote Sens. Environ. 244, 111806. https://doi.org/10.1016/j.rse.2020.111806

Gruber, A., Su, C.-H., Crow, W.T., Zwieback, S., Dorigo, W.A., Wagner, W., 2016a. Estimating error cross-correlations in soil moisture data sets using extended collocation analysis. J. Geophys. Res. Atmospheres 121, 1208–1219. https://doi.org/10.1002/2015JD024027

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016b. Recent advances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs. Geoinformation, Advances in the Validation and Application of Remotely Sensed Soil Moisture - Part 1 45, 200–211. https://doi.org/10.1016/j.jag.2015.09.002

Guillevic, P.C., Olioso, A., Hook, S.J., Fisher, J.B., Lagouarde, J.-P., Vermote, E.F., 2019. Impact of the Revisit of Thermal Infrared Remote Sensing Observations on Evapotranspiration Uncertainty—A Sensitivity Study Using AmeriFlux Data. Remote Sens. 11, 573. https://doi.org/10.3390/rs11050573

Guo, X., Yao, Y., Zhang, Y., Lin, Y., Jiang, B., Jia, K., Zhang, X., Xie, X., Zhang, L., Shang, K., Yang, J., Bei, X., 2020. Discrepancies in the Simulated Global Terrestrial Latent Heat Flux from GLASS and MERRA-2 Surface Net Radiation Products. Remote Sens. 12, 2763. https://doi.org/10.3390/rs12172763

Gurevitch, J., Koricheva, J., Nakagawa, S., Stewart, G., 2018. Meta-analysis and the science of research synthesis. Nature 555, 175–182. https://doi.org/10.1038/nature25753

Guzinski, R., Nieto, H., Barrios, J.M., Ghariani, W., Gellens-Meulenberghs, F., De Pue, J., Lacaze, R., 2025. Towards a global actual evapotranspiration product for the Copernicus Land Monitoring Service. EGUsphere 1–63. https://doi.org/10.5194/egusphere-2025-4342

Guzinski, R., Nieto, H., Sánchez, J.M., López-Urrea, R., Boujnah, D.M., Boulet, G., 2021. Utility of Copernicus-Based Inputs for Actual Evapotranspiration Modeling in Support of Sustainable Water Use in Agriculture. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 11466–11484. https://doi.org/10.1109/JSTARS.2021.3122573

Guzinski, R., Nieto, H., Sandholt, I., Karamitilios, G., 2020. Modelling High-Resolution Actual Evapotranspiration through Sentinel-2 and Sentinel-3 Data Fusion. Remote Sens. 12, 1433. https://doi.org/10.3390/rs12091433

Hadjimitsis, D.G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M.G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toulios, L., Clayton, C.R.I., 2010. Atmospheric correction for satellite remotely sensed data intended for agricultural

applications: impact on vegetation indices. Nat. Hazards Earth Syst. Sci. 10, 89–95. https://doi.org/10.5194/nhess-10-89-2010

Hagolle, O., Gamet, P., Bhattacharya, B.K., Roujean, J.-L., Boulet, G., Autret, E., Picard, G., Roupioz, L., Irvine, M., Marcq, S., Salcedo, C., Maisongrande, P., 2025. TRISHNA: an Indo-French Missionto Study the Thermography of the Earth at FineSpatio-Temporal resolution.

Hazimeh, R., Jaafar, H., 2024. Impact of ET and biomass model choices on economic irrigation water productivity in water-scarce basins. Agric. Water Manag. 292, 108651. https://doi.org/10.1016/j.agwat.2023.108651

He, S., Yang, Q., Zhang, L., Shi, Z., Wang, X., Lv, B., 2025. Consistency Assessment and Uncertainty Analysis of Spatial-temporal Characteristics of Evaporation Data in the Greater Mekong Subregion. https://doi.org/10.1175/JHM-D-24-0014.1

He, X., Xu, T., Xia, Y., Bateni, S.M., Guo, Z., Liu, S., Mao, K., Zhang, Y., Feng, H., Zhao, J., 2020. A Bayesian Three-Cornered Hat (BTCH) Method: Improving the Terrestrial Evapotranspiration Estimation. Remote Sens. 12, 878. https://doi.org/10.3390/rs12050878

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2020. Statistical methods in water resources, version 1.1. ed, Techniques and Methods, U.S. Geological Survey Techniques and Methods. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/tm4A3

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., 2023. ERA5 hourly data on single levels from 1940 to present. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803

Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modelling with GIS. CRC Press, London. https://doi.org/10.4324/9780203016114

Hill, R.J., 1997. Algorithms for Obtaining Atmospheric Surface-Layer Fluxes from Scintillation Measurements. J. Atmospheric Ocean. Technol. 14, 456–467. https://doi.org/10.1175/1520-0426(1997)014%253C0456:AFOASL%253E2.0.CO;2

Hinge, G., Mohamed, M.M., Long, D., Hamouda, M.A., 2021. Meta-Analysis in Using Satellite Precipitation Products for Drought Monitoring: Lessons Learnt and Way Forward. Remote Sens. 13, 4353. https://doi.org/10.3390/rs13214353

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

Hoedjes, J.C.B., Chehbouni, A., Jacob, F., Ezzahar, J., Boulet, G., 2008. Deriving daily evapotranspiration from remotely sensed instantaneous evaporative fraction over olive orchard in semi-arid Morocco. J. Hydrol. 354, 53–64. https://doi.org/10.1016/j.jhydrol.2008.02.016

Holland, R., 1999. Reflexivity. Hum. Relat. 52, 463–484. https://doi.org/10.1177/001872679905200403

Holtslag, A.A.M., 1984. Estimates of diabatic wind speed profiles from near-surface weather observations. Bound.-Layer Meteorol. 29, 225–250. https://doi.org/10.1007/BF00119790

Hook, S.J., 2025. ECOSTRESS, SBG-TIR andHyTES - Status and Results.

Hovmöller, E., 1949. The Trough-and-Ridge diagram. Tellus 1, 62–66. https://doi.org/10.1111/j.2153-3490.1949.tb01260.x

Huang, J., Sehgal, V., Alvarez, L.V., Brocca, L., Cai, S., Cheng, R., Cheng, X., Du, J., El Masri, B., Endsley, K.A., Fang, Y., Hu, J., Jampani, M., Kibria, M.G., Koren, G., Li, L., Liu, L., Mao, J., Moreno, H.A., Rigden, A., Shi, M., Shi, X., Wang, Y., Zhang, X., Fisher, J.B., 2025. Remotely Sensed High-Resolution Soil Moisture and Evapotranspiration: Bridging the Gap Between Science and Society. Water Resour. Res. 61, e2024WR037929. https://doi.org/10.1029/2024WR037929

Huete, A.R., Jackson, R.D., Post, D.F., 1985. Spectral response of a plant canopy with different soil backgrounds. Remote Sens. Environ. 17, 37–53. https://doi.org/10.1016/0034-4257(85)90111-7

Hughes, D.A., Jewitt, G., Mahé, G., Mazvimavi, D., Stisen, S., 2015. A review of aspects of hydrological sciences research in Africa over the past decade. Hydrol. Sci. J. 60, 1865–1879. https://doi.org/10.1080/02626667.2015.1072276

Hwang, K., Choi, M., 2013. Seasonal trends of satellite-based evapotranspiration algorithms over a complex ecosystem in East Asia. Remote Sens. Environ. 137, 244–263. https://doi.org/10.1016/j.rse.2013.06.006

Hydrosat, 2023. Satellite Technology - Hydrosat. URL https://hydrosat.com/ (accessed 11.19.25).

IHE Delft, 2023. WaPOR Concepts and Validation | OCW IHE DELFT [WWW Document]. URL https://ocw.un-ihe.org/enrol/index.php?id=214 (accessed 1.27.25).

Ippolito, M., De Caro, D., Cannarozzo, M., Provenzano, G., Ciraolo, G., 2024. Evaluation of daily crop reference evapotranspiration and sensitivity analysis of FAO Penman-Monteith equation using ERA5-Land reanalysis database in Sicily, Italy. Agric. Water Manag. 295, 108732. https://doi.org/10.1016/j.agwat.2024.108732

Isaac, P., Cleverly, J., McHugh, I., van Gorsel, E., Ewenz, C., Beringer, J., 2017. OzFlux data: network integration from collection to curation. Biogeosciences 14, 2903–2928. https://doi.org/10.5194/bg-14-2903-2017

Jaafar, H., Karimi, P., Borgomeo, E., 2024. Economic irrigation water productivity of wheat and potato: An earth observation perspective on policy implications in the Litani Basin, Lebanon. Agric. Water Manag. 306, 109180. https://doi.org/10.1016/j.agwat.2024.109180

Jackson, S., Head, L., 2022. The politics of evaporation and the making of atmospheric territory in Australia's Murray-Darling Basin. Environ. Plan. E Nat. Space 5, 1273–1295. https://doi.org/10.1177/25148486211038392

Jensen, M.E., Burman, R.D., Allen, R.G., 2016. Evaporation, Evapotranspiration, and Irrigation Water Requirements, ASCE Manuals and Reports on Enginerring Practice. American Society of Civil Engineers, Virginia. https://doi.org/10.1061/9780784414057

Jewitt, G., 2006. Integrating blue and green water flows for water resources management and planning. Phys. Chem. Earth Parts ABC, Water for Sustainable Socio-Economic

Development, Good Health for All and Gender Equity 31, 753–762. https://doi.org/10.1016/j.pce.2006.08.033

Jiang, L., Zhang, B., Han, S., Chen, H., Wei, Z., 2021. Upscaling evapotranspiration from the instantaneous to the daily time scale: Assessing six methods including an optimized coefficient based on worldwide eddy covariance flux network. J. Hydrol. 596, 126135. https://doi.org/10.1016/j.jhydrol.2021.126135

Jiang, Z., Huete, A.R., Chen, J., Chen, Y., Li, J., Yan, G., Zhang, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. Remote Sens. Environ. 101, 366–378. https://doi.org/10.1016/j.rse.2006.01.003

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S.I., McCabe, M.F., Wood, E.F., Rossow, W.B., Balsamo, G., Betts, A.K., Dirmeyer, P.A., Fisher, J.B., Jung, M., Kanamitsu, M., Reichle, R.H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., Wang, K., 2011. Global intercomparison of 12 land surface heat flux estimates. J. Geophys. Res. Atmospheres 116. https://doi.org/10.1029/2010JD014545

Joint Committee for Guides in Metrology (JCGM), 2012. International vocabulary of metrology—Basic and general concepts and associated terms, BIPM. Sèvres, France.

Jung, H.C., Getirana, A., Arsenault, K.R., Holmes, T.R.H., McNally, A., 2019. Uncertainties in Evapotranspiration Estimates over West Africa. Remote Sens. 11, 892. https://doi.org/10.3390/rs11080892

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., Reichstein, M., 2019. The FLUXCOM ensemble of global land-atmosphere energy fluxes. Sci. Data 6, 74. https://doi.org/10.1038/s41597-019-0076-8

Kalma, J.D., McVicar, T.R., McCabe, M.F., 2008. Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. Surv. Geophys. 29, 421–469. https://doi.org/10.1007/s10712-008-9037-z

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

Katul, G.G., Oren, R., Manzoni, S., Higgins, C., Parlange, M.B., 2012. Evapotranspiration: A process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system. Rev. Geophys. 50. https://doi.org/10.1029/2011RG000366

Kendall, M.G., 1955. Further Contributions to the Theory of Paired Comparisons. Biometrics 11, 43–62. https://doi.org/10.2307/3001479

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

Kibria, S., Masia, S., Sušnik, J., Hessels, T.M., 2021. Critical comparison of actual evapotranspiration estimates using ground based, remotely sensed, and simulated data in the USA. Agric. Water Manag. 248, 106753. https://doi.org/10.1016/j.agwat.2021.106753

Kiptala, J.K., Mohamed, Y., Mul, M.L., Van der Zaag, P., 2013. Mapping evapotranspiration trends using MODIS and SEBAL model in a data scarce and heterogeneous landscape in Eastern Africa. Water Resour. Res. 49, 8495–8510. https://doi.org/10.1002/2013WR014240

Kivi, Z.R., Javadi, S., Karimi, N., Shahdany, S.M.H., Moghaddam, H.K., 2022. Performance evaluation and verification of groundwater balance using WA + as a new water accounting system. Environ. Monit. Assess. 194, 580. https://doi.org/10.1007/s10661-022-10193-7

Kleijn, J., 2024. WaPOR Born Out of Necessity! Inter-Reg. Tech. Platf. Water Scarcity IRTP-WS. URL https://www.fao.org/platforms/water-scarcity/Outreach/blog-on-water-scarcity/blog-detail/pasquale-steduto/2024/07/04/wapor-born-out-of-necessity/en (accessed 1.7.25).

Klein, A., Unverzagt, K., Alba, R., Donges, J.F., Hertz, T., Krueger, T., Lindkvist, E., Martin, R., Niewöhner, J., Prawitz, H., Schlüter, M., Schwarz, L., Wijermans, N., 2024. From situated knowledges to situated modelling: a relational framework for simulation modelling. Ecosyst. People 20, 2361706. https://doi.org/10.1080/26395916.2024.2361706

Kljun, N., Calanca, P., Rotach, M.W., Schmid, H.P., 2015. A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). Geosci. Model Dev. 8, 3695–3713. https://doi.org/10.5194/gmd-8-3695-2015

Knipper, K.R., Kustas, W.P., Anderson, M.C., Alfieri, J.G., Prueger, J.H., Hain, C.R., Gao, F., Yang, Y., McKee, L.G., Nieto, H., Hipps, L.E., Alsina, M.M., Sanchez, L., 2019. Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. Irrig. Sci. 37, 431–449. https://doi.org/10.1007/s00271-018-0591-y

Knipper, K.R., Kustas, W.P., Anderson, M.C., Nieto, H., Alfieri, J.G., Prueger, J.H., Hain, C.R., Gao, F., McKee, L.G., Alsina, M.M., Sanchez, L., 2020. Using high-spatiotemporal thermal satellite ET retrievals to monitor water use over California vineyards of different climate, vine variety and trellis design. Agric. Water Manag. 241, 106361. https://doi.org/10.1016/j.agwat.2020.106361

Knox, J.W., Kay, M.G., Weatherhead, E.K., 2012. Water regulation, crop production, and agricultural water management—Understanding farmer perspectives on irrigation efficiency. Agric. Water Manag., Irrigation efficiency and productivity: scales, systems and science 108, 3–8. https://doi.org/10.1016/j.agwat.2011.06.007

Koppa, A., Gebremichael, M., 2017. A Framework for Validation of Remotely Sensed Precipitation and Evapotranspiration Based on the Budyko Hypothesis. Water Resour. Res. 53, 8487–8499. https://doi.org/10.1002/2017WR020593

Korzoun, V.I., Sokolov, A.A., Budyko, M.I., Voskresensky, K.P., Kalinin, G.P., Konoplyantsev, A.A., Korotkevich, E.S., Kuzin, P.S., Lvovich, M.I., 1978. World water balance and water resources of the earth. Stud. Rep. Hydrol. UNESCO.

Kroese, D.P., Brereton, T., Taimre, T., Botev, Z.I., 2014. Why the Monte Carlo method is so important today. WIREs Comput. Stat. 6, 386–392. https://doi.org/10.1002/wics.1314

Kroese, D.P., Rubinstein, R.Y., 2012. Monte Carlo methods. WIREs Comput. Stat. 4, 48–58. https://doi.org/10.1002/wics.194

Krueger, T., Alba, R., 2022. Ontological and epistemological commitments in interdisciplinary water research: Uncertainty as an entry point for reflexion. Front. Water 4. https://doi.org/10.3389/frwa.2022.1038322

Kruskal, W.H., and Wallis, W.A., 1952. Use of Ranks in One-Criterion Variance Analysis. J. Am. Stat. Assoc. 47, 583–621. https://doi.org/10.1080/01621459.1952.10483441

Kustas, W.P., Norman, J.M., 1999. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agric. For. Meteorol. 94, 13–29. https://doi.org/10.1016/S0168-1923(99)00005-2

Kustas, W.P., Norman, J.M., 1996. Use of remote sensing for evapotranspiration monitoring over land surfaces. Hydrol. Sci. J. 41, 495–516. https://doi.org/10.1080/02626669609491522

Kvålseth, T.O., 1985. Cautionary Note about R 2. Am. Stat. 39, 279–285. https://doi.org/10.1080/00031305.1985.10479448

Lang, Q., Wang, L., Qin, W., Wang, Z., Su, X., Zhang, M., 2024. Effects of increasing spatial resolution on the spatial information content and accuracy of downward surface shortwave radiation. Int. J. Appl. Earth Obs. Geoinformation 133, 104128. https://doi.org/10.1016/j.jag.2024.104128

Lankford, B., 2012. Towards a political ecology of irrigation efficiency and productivity. Agric. Water Manag., Irrigation efficiency and productivity: scales, systems and science 108, 1–2. https://doi.org/10.1016/j.agwat.2012.03.005

Lebigot, E.O., 2017. Uncertainties: a Python package for calculations with uncertainties. [WWW Document]. Uncertainties Python Package 301. URL https://pythonhosted.org/uncertainties/ (accessed 7.30.24).

Lee, A., Baudin, M., Collette, Y., Martinez, J.-M., 2014. Randomized Designs — pyDOE 0.3.6 documentation [WWW Document]. PyDOE Exp. Des. Package Python. URL https://pythonhosted.org/pyDOE/randomized.html#latin-hypercube (accessed 7.30.24).

Lehmann, F., Vishwakarma, B.D., Bamber, J., 2022. How well are we able to close the water budget at the global scale? Hydrol. Earth Syst. Sci. 26, 35–54. https://doi.org/10.5194/hess-26-35-2022

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., Pfister, H., 2014. UpSet: Visualization of Intersecting Sets. IEEE Trans. Vis. Comput. Graph. 20, 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

Li, X., Dusseldorp, E., Su, X., Meulman, J.J., 2020. Multiple moderator meta-analysis using the R-package Meta-CART. Behav. Res. Methods 52, 2657–2673. https://doi.org/10.3758/s13428-020-01360-0

Li, X., Liu, S., Li, H., Ma, Y., Wang, J., Zhang, Y., Xu, Z., Xu, T., Song, L., Yang, X., Lu, Z., Wang, Z., Guo, Z., 2018. Intercomparison of Six Upscaling Evapotranspiration Methods: From Site to the Satellite Pixel. J. Geophys. Res. Atmospheres 123, 6777–6803. https://doi.org/10.1029/2018JD028422

Li, X., Liu, S., Yang, X., Ma, Y., He, X., Xu, Z., Xu, T., Song, L., Zhang, Y., Hu, X., Ju, Q., Zhang, X., 2021. Upscaling Evapotranspiration from a Single-Site to Satellite Pixel Scale. Remote Sens. 13, 4072. https://doi.org/10.3390/rs13204072

Li, X., Xin, X., Jiao, J., Peng, Z., Zhang, H., Shao, S., Liu, Q., 2017. Estimating subpixel surface heat fluxes through applying temperature-sharpening methods to MODIS data. Remote Sens. 9. https://doi.org/10.3390/rs9080836

Li, Z., Jia, L., Lu, J., 2015. On Uncertainties of the Priestley-Taylor/LST-Fc Feature Space Method to Estimate Evapotranspiration: Case Study in an Arid/Semiarid Region in Northwest China. Remote Sens. 7, 447–466. https://doi.org/10.3390/rs70100447

Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013. Satellite-derived land surface temperature: Current status and perspectives. Remote Sens. Environ. 131, 14–37. https://doi.org/10.1016/j.rse.2012.12.008

Li, Z.-L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., Yan, G., Zhang, X., 2009. A Review of Current Methodologies for Regional Evapotranspiration Estimation from Remotely Sensed Data. Sensors 9, 3801–3853. https://doi.org/10.3390/s90503801

Liaqat, U.W., Choi, M., 2017. Accuracy comparison of remotely sensed evapotranspiration products and their associated water stress footprints under different land cover types in Korean peninsula. J. Clean. Prod., Sustainable Development of Energy, Water and Environmental Systems 155, 93–104. https://doi.org/10.1016/j.jclepro.2016.09.022

Liaqat, U.W., Choi, M., 2015. Surface energy fluxes in the Northeast Asia ecosystem: SEBS and METRIC models using Landsat satellite images. Agric. For. Meteorol. 214–215, 60–79. https://doi.org/10.1016/j.agrformet.2015.08.245

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

Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., Li, X., Sang, Y.-F., 2016. A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. J. Hydrol. 538, 82–95. https://doi.org/10.1016/j.jhydrol.2016.04.006

Liu, Z., 2021. The accuracy of temporal upscaling of instantaneous evapotranspiration to daily values with seven upscaling methods. Hydrol. Earth Syst. Sci. 25, 4417–4433. https://doi.org/10.5194/hess-25-4417-2021

Lo Piano, S., Sheikholeslami, R., Puy, A., Saltelli, A., 2022. Unpacking the modelling process via sensitivity auditing. Futures 144, 103041. https://doi.org/10.1016/j.futures.2022.103041

Loew, A., Bell, W., Brocca, L., Bulgin, C.E., Burdanowitz, J., Calbet, X., Donner, R.V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J.-C., Schaepman-Strub, G., Schröder, M., Verhoelst, T., 2017. Validation practices for satellite-based Earth observation data across communities. Rev. Geophys. 55, 779–817. https://doi.org/10.1002/2017RG000562

Long, D., Longuevergne, L., Scanlon, B.R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour. Res. 50, 1131–1151. https://doi.org/10.1002/2013WR014581

Long, D., Singh, V.P., Li, Z.-L., 2011. How sensitive is SEBAL to changes in input variables, domain size and satellite sensor? J. Geophys. Res. Atmospheres 116. https://doi.org/10.1029/2011JD016542

López, O., Houborg, R., McCabe, M.F., 2017. Evaluating the hydrological consistency of evaporation products using satellite-based gravity and rainfall data. Hydrol. Earth Syst. Sci. 21, 323–343. https://doi.org/10.5194/hess-21-323-2017

Lu, P., Urban, L., Zhao, P., 2004. Granier's thermal dissipation probe (TDP) method for measuring sap flow in trees: Theory and practice. Acta Bot. Sin. 46, 631–646.

Mann, H.B., 1945. Nonparametric Tests Against Trend. Econometrica 13, 245–259. https://doi.org/10.2307/1907187

Martens, B., Miralles, D.G., Lievens, H., Van Der Schalie, R., De Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3:

Satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Martins, D.S., Paredes, P., Raziei, T., Pires, C., Cadima, J., Pereira, L.S., 2017. Assessing reference evapotranspiration estimation from reanalysis weather products. An application to the Iberian Peninsula. Int. J. Climatol. 37, 2378–2397. https://doi.org/10.1002/joc.4852

Matar, A., 2019. Evaluating criteria for water allocation in closed or closing basins : Case study of Litani River Basin - Lebanon. IHE Delft Institute for Water Education.

Mauder, M., Foken, T., Cuxart, J., 2020. Surface-Energy-Balance Closure over Land: A Review. Bound.-Layer Meteorol. 177, 395–426. https://doi.org/10.1007/s10546-020-00529-6

Mauder, M., Jung, M., Stoy, P., Nelson, J., Wanner, L., 2024. Energy balance closure at FLUXNET sites revisited. Agric. For. Meteorol. 358, 110235. https://doi.org/10.1016/j.agrformet.2024.110235

Mayr, S., Kuenzer, C., Gessner, U., Klein, I., Rutzinger, M., 2019. Validation of Earth Observation Time-Series: A Review for Large-Area and Temporally Dense Land Surface Products. Remote Sens. 11, 2616. https://doi.org/10.3390/rs11222616

McCabe, M.F., Ershadi, A., Jimenez, C., Miralles, D.G., Michel, D., Wood, E.F., 2016. The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data. Geosci. Model Dev. 9, 283–305. https://doi.org/10.5194/gmd-9-283-2016

McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E.C., Franz, T.E., Shi, J., Gao, H., Wood, E.F., 2017. The future of Earth observation in hydrology. Hydrol. Earth Syst. Sci. 21, 3879–3914. https://doi.org/10.5194/hess-21-3879-2017

McColl, K.A., Vogelzang, J., Konings, A.G., Entekhabi, D., Piles, M., Stoffelen, A., 2014. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. Geophys. Res. Lett. 41, 6229–6236. https://doi.org/10.1002/2014GL061322

McDermid, S., Nocco, M., Lawston-Parker, P., Keune, J., Pokhrel, Y., Jain, M., Jägermeyr, J., Brocca, L., Massari, C., Jones, A.D., Vahmani, P., Thiery, W., Yao, Y., Bell, A., Chen, L., Dorigo, W., Hanasaki, N., Jasechko, S., Lo, M.-H., Mahmood, R., Mishra, V., Mueller, N.D., Niyogi, D., Rabin, S.S., Sloat, L., Wada, Y., Zappa, L., Chen, F., Cook, B.I., Kim, H., Lombardozzi, D., Polcher, J., Ryu, D., Santanello, J., Satoh, Y., Seneviratne, S., Singh, D., Yokohata, T., 2023. Irrigation in the Earth system. Nat. Rev. Earth Environ. 4, 435–453. https://doi.org/10.1038/s43017-023-00438-5

Mebrie, D.W., Assefa, T.T., Yimam, A.Y., Belay, S.A., 2023. A remote sensing approach to estimate variable crop coefficient and evapotranspiration for improved water productivity in the Ethiopian highlands. Appl. Water Sci. 13, 168. https://doi.org/10.1007/s13201-023-01968-5

Melsen, L.A., 2022. It Takes a Village to Run a Model—The Social Practices of Hydrological Modeling. Water Resour. Res. 58, e2021WR030600. https://doi.org/10.1029/2021WR030600

Melsen, L.A., Teuling, A.J., Torfs, P.J.J.F., Zappa, M., Mizukami, N., Mendoza, P.A., Clark, M.P., Uijlenhoet, R., 2019. Subjective modeling decisions can significantly

impact the simulation of flood and drought events. J. Hydrol. 568, 1093–1104. https://doi.org/10.1016/j.jhydrol.2018.11.046

Melton, F.S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Senay, G.B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., Daudert, B., Doherty, C., Dunkerly, C., Friedrichs, M., Guzman, A., Halverson, G., Hansen, J., Harding, J., Kang, Y., Ketchum, D., Minor, B., Morton, C., Ortega-Salazar, S., Ott, T., Ozdogan, M., ReVelle, P.M., Schull, M., Wang, C., Yang, Y., Anderson, R.G., 2022. OpenET: Filling a Critical Data Gap in Water Management for the Western United States. JAWRA J. Am. Water Resour. Assoc. 58, 971–994. https://doi.org/10.1111/1752-1688.12956

METER, 2023. ATMOS41 Manual [WWW Document]. URL https://library.metergroup.com/Manuals/20635_ATMOS41_Manual_Web.pdf (accessed 7.29.24).

Michel, D., Jiménez, C., Miralles, D.G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M.F., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. Hydrol. Earth Syst. Sci. 20, 803–822. https://doi.org/10.5194/hess-20-803-2016

Miralles, D.G., Bonte, O., Koppa, A., Baez-Villanueva, O.M., Tronquo, E., Zhong, F., Beck, H.E., Hulsman, P., Dorigo, W., Verhoest, N.E.C., Haghdoost, S., 2025. GLEAM4: global land evaporation and soil moisture dataset at 0.1 resolution from 1980 to near present. Sci. Data 12, 416. https://doi.org/10.1038/s41597-025-04610-y

Miralles, D.G., Brutsaert, W., Dolman, A.J., Gash, J.H., 2020. On the Use of the Term "Evapotranspiration." Water Resour. Res. 56, e2020WR028055. https://doi.org/10.1029/2020WR028055

Miralles, D.G., Crow, W.T., Cosh, M.H., 2010. Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations. https://doi.org/10.1175/2010JHM1285.1

Miralles, D.G., De Jeu, R. a. M., Gash, J.H., Holmes, T.R.H., Dolman, A.J., 2011a. Magnitude and variability of land evaporation and its components at the global scale. Hydrol. Earth Syst. Sci. 15, 967–981. https://doi.org/10.5194/hess-15-967-2011

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

Mohammadi, S., Cremaschi, S., 2022. Efficiency of uncertainty propagation methods for moment estimation of uncertain model outputs. Comput. Chem. Eng. 166, 107954. https://doi.org/10.1016/j.compchemeng.2022.107954

Mohan, M.M.P., Kanchirapuzha, R., Varma, M.R.R., 2020. Review of approaches for the estimation of sensible heat flux in remote sensing-based evapotranspiration models. J. Appl. Remote Sens. 14, 041501. https://doi.org/10.1117/1.JRS.14.041501

Molden, D., 1997. Accounting for Water Use and Productivity. IWMI.

Montanari, A., 2007. What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. Hydrol. Process. 21, 841–845. https://doi.org/10.1002/hyp.6623

Monteith, J.L., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205–234.

Montgomery, R.B., 1948. Vertical Eddy Flux of Heat in the Atmosphere. J. Atmospheric Sci. 5, 265–274. https://doi.org/10.1175/1520-0469(1948)005%253C0265:VEFOHI%253E2.0.CO;2

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

Mu, Q., Zhao, M., Running, S.W., 2013. MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3) Algorithm Theoretical Basis Document Collection 5. NASA Headquarters, Washington, DC.

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

Mueller, B, Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P., Dolman, A., Fisher, J., Jung, M., Ludwig, F., Maignan, F., Miralles, D., McCabe, M., Reichstein, M., Sheffield, J., Wang, K., Wood, E., Zhang, Y., Seneviratne, S., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. Hydrol. EARTH Syst. Sci. 17, 3707–3720. https://doi.org/10.5194/hess-17-3707-2013

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P.A., Dolman, A.J., Fisher, J.B., Jung, M., Ludwig, F., Maignan, F., Miralles, D.G., McCabe, M.F., Reichstein, M., Sheffield, J., Wang, K., Wood, E.F., Zhang, Y., Seneviratne, S.I., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. Hydrol. Earth Syst. Sci. 17, 3707–3720. https://doi.org/10.5194/hess-17-3707-2013

Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J.B., Guo, Z., Jung, M., Maignan, F., McCabe, M.F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A.J., Wang, K., Wood, E.F., Zhang, Y., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. Geophys. Res. Lett. 38. https://doi.org/10.1029/2010GL046230

Nakagawa, S., Yang, Y., Macartney, E.L., Spake, R., Lagisz, M., 2023. Quantitative evidence synthesis: a practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. Environ. Evid. 12, 8. https://doi.org/10.1186/s13750-023-00301-6

NASA, 2021. Data Processing Levels | Earthdata [WWW Document]. URL https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels/ (accessed 9.7.21).

Nearing, G.S., Tian, Y., Gupta, H.V., Clark, M.P., Harrison, K.W., Weijs, S.V., 2016. A philosophical basis for hydrological uncertainty. Hydrol. Sci. J. 61, 1666–1678. https://doi.org/10.1080/02626667.2016.1183009

Njuki, S.M., Mannaerts, C.M., Su, Z., 2023. Accounting for Turbulence-Induced Canopy Heat Transfer in the Simulation of Sensible Heat Flux in SEBS Model. Remote Sens. 15, 1578. https://doi.org/10.3390/rs15061578

Oki, T., Kanae, S., 2006. Global Hydrological Cycles and World Water Resources. Science 313, 1068–1072. https://doi.org/10.1126/science.1128845

Oliphant, A.J., 2012. Terrestrial Ecosystem-Atmosphere Exchange of CO2, Water and Energy from FLUXNET; Review and Meta-Analysis of a Global in-situ Observatory. Geogr. Compass 6, 689–705. https://doi.org/10.1111/gec3.12009

Oliveira, S., Cunha, J., Nóbrega, R.L.B., Gash, J.H., Valente, F., 2024. Enhancing global rainfall interception loss estimation through vegetation structure modeling. J. Hydrol. 631, 130672. https://doi.org/10.1016/j.jhydrol.2024.130672

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science 263, 641–646. https://doi.org/10.1126/science.263.5147.641

Pan, M., Fisher, C.K., Chaney, N.W., Zhan, W., Crow, W.T., Aires, F., Entekhabi, D., Wood, E.F., 2015. Triple collocation: Beyond three estimates and separation of structural/non-structural errors. Remote Sens. Environ. 171, 299–310. https://doi.org/10.1016/j.rse.2015.10.028

Pan, Shufen, Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J.E.M.S., Ottlé, C., Poulter, B., Zaehle, S., Running, S.W., 2020. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. Hydrol. Earth Syst. Sci. 24, 1485–1509. https://doi.org/10.5194/hess-24-1485-2020

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J.E.M.S., Ottlé, C., Poulter, B., Zaehle, S., Running, S.W., 2020. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. Hydrol. Earth Syst. Sci. 24, 1485–1509. https://doi.org/10.5194/hess-24-1485-2020

Pardo, N., Sanchez, M., Timmermans, J., Su, Z., Perez, I., Garcia, M., 2014. SEBS validation in a Spanish rotating crop. Agric. For. Meteorol. 195, 132–142. https://doi.org/10.1016/j.agrformet.2014.05.007

Parker, W.S., 2016. Reanalyses and Observations: What's the Difference? https://doi.org/10.1175/BAMS-D-14-00226.1

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M.A., Ardö, J., Arkebauer, T., Arndt, S.K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L.B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Boike, J., Bolstad, P.V., Bonal, D., Bonnefond, J.-M., Bowling, D.R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S.P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T.R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B.D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P.S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K.J., Cinti, B.D., Grandcourt, A. de, Ligne, A.D., De Oliveira, R.C., Delpierre, N., Desai, A.R., Di Bella, C.M., Tommasi, P. di, Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H.A.M., Eugster, W., Ewenz, C.M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A.H., Gough, C.M., Goulden, M.L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B.U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-

Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W.L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Yingnian, Liddell, M., Limousin, J.-M., Lion, M., Liska, A.J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H.A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J.H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A.M.S., Merbold, L., Meyer, W., Meyers, T., Miller, S.D., Minerbi, S., Moderow, U., Monson, R.K., Montagnani, L., Moore, C.E., Moors, E., Moreaux, V., Moureaux, C., Munger, J.W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J.E., Ourcival, J.-M., Papuga, S.A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R.P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S.M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V.R. de, Restrepo-Coupe, N., Reverter, B.R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.R., Sánchez-Cañete, E.P., Sanchez-Mejia, Z.M., Schmid, H.P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R.L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R.M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J.P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Yuelin, Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., Papale, D., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3

Paw U, K.T., Qiu, J., Su, H.-B., Watanabe, T., Brunet, Y., 1995. Surface renewal analysis: a new method to obtain scalar fluxes. Agric. For. Meteorol. 74, 119–137. https://doi.org/10.1016/0168-1923(94)02182-J

Pelgrum, H., Miltenburg, I.J., Cheema, M.J.M., Klaasse, A., 2010. ETLook: a novel continental evapotranspiration algorithm 5.

Pelosi, A., Chirico, G.B., 2021. Regional assessment of daily reference evapotranspiration: Can ground observations be replaced by blending ERA5-Land meteorological reanalysis and CM-SAF satellite-based radiation data? Agric. Water Manag. 258, 107169. https://doi.org/10.1016/j.agwat.2021.107169

Pelosi, A., Terribile, F., D'Urso, G., Chirico, G., 2020. Comparison of ERA5-Land and UERRA MESCAN-SURFEX Reanalysis Data with Spatially Interpolated Weather Observations for the Regional Assessment of Reference Evapotranspiration. Water 12, 1669. https://doi.org/10.3390/w12061669

Peng, Z.Q., Xin, X., Jiao, J.J., Zhou, T., Liu, Q., 2016. Remote sensing algorithm for surface evapotranspiration considering landscape and statistical effects on mixed pixels. Hydrol. Earth Syst. Sci. 20, 4409–4438. https://doi.org/10.5194/hess-20-4409-2016

Penman, H.L., 1956. Evaporation: an introductory survey. Neth. J. Agric. Sci. 4, 9–29. https://doi.org/10.18174/njas.v4i1.17768

Pereira, L.S., Paredes, P., Hunsaker, D.J., López-Urrea, R., Jovanovic, N., 2021. Updates and advances to the FAO56 crop water requirements method. Agric. Water Manag. 248, 106697. https://doi.org/10.1016/j.agwat.2020.106697

Pickering, C., Byrne, J., 2014. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. High. Educ. Res. Dev. 33, 534–548. https://doi.org/10.1080/07294360.2013.841651

Pitoro, V.S.J., Franco, J.R., de Souza Correa, L.R., Manjate, M.J., Román, R.M.S., 2025. Evaluating Water Productivity and Efficiency in Irrigated Fields Using Remote Sensing: A Case Study in Mozambique's Sugarcane Cultivation. Water Conserv. Sci. Eng. 10, 14. https://doi.org/10.1007/s41101-025-00340-9

Povey, A.C., Grainger, R.G., 2015. Known and unknown unknowns: uncertainty estimation in satellite remote sensing. Atmospheric Meas. Tech. 8, 4699–4718. https://doi.org/10.5194/amt-8-4699-2015

Premoli, A., Tavella, P., 1993. A revisited three-cornered hat method for estimating frequency standard instability. IEEE Trans. Instrum. Meas. 42, 7–13. https://doi.org/10.1109/19.206671

Radmanesh, Y., Tabrizi, M.S., Etedali, H.R., Azizian, A., Babazadeh, H., 2023. Comparative evaluation of the accuracy of re-analysed and gauge-based climatic data in Iran. J. Earth Syst. Sci. 132, 190. https://doi.org/10.1007/s12040-023-02202-1

Rajib, A., Merwade, V., Yu, Z., 2018. Rationale and Efficacy of Assimilating Remotely Sensed Potential Evapotranspiration for Reduced Uncertainty of Hydrologic Models. Water Resour. Res. 54, 4615–4637. https://doi.org/10.1029/2017WR021147

Rakovec, O., Hill, M.C., Clark, M.P., Weerts, A.H., Teuling, A.J., Uijlenhoet, R., 2014. Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models. Water Resour. Res. 50, 409–426. https://doi.org/10.1002/2013WR014063

Ramoelo, A., Majozi, N., Mathieu, R., Jovanovic, N., Nickless, A., Dzikiti, S., 2014. Validation of Global Evapotranspiration Product (MOD16) using Flux Tower Data in the African Savanna, South Africa. Remote Sens. 6, 7406–7423. https://doi.org/10.3390/rs6087406

Rashid, T., Tian, D., 2024. Improved 30-m Evapotranspiration Estimates Over 145 Eddy Covariance Sites in the Contiguous United States: The Role of ECOSTRESS, Harmonized Landsat Sentinel-2 Imagery, Climate Reanalysis, and Deep Neural Network Postprocessing. Water Resour. Res. 60, e2023WR036313. https://doi.org/10.1029/2023WR036313

Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resour. Res. 51, 3070–3092. https://doi.org/10.1002/2014WR016527

Rhenals, A.E., Bras, R.L., 1981. The irrigation scheduling problem and evapotranspiration uncertainty. Water Resour. Res. 17, 1328–1338. https://doi.org/10.1029/WR017i005p01328

Rienecker, M.M., Suarez, M.J., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R.D., Gelaro, R., Stajner, I., Nielsen, J.E., 2008. The GEOS-5 Data Assimilation System-Documentation of Versions 5.0.1, 5.1.0, and 5.2.0 (No. NASA/TM-2008-104606-VOL-27).

Rijtema, P.E., 1965. An analysis of actual evapotranspiration (Doctor of Philosophy). Wageningen University, Wageningen.

Rochette, P., Pattey, E., Desjardins, R.L., Dwyer, L.M., Stewart, D.W., Dubé, P.A., 1991. Estimation of maize (*Zea mays* L.) canopy conductance by scaling up leaf stomatal conductance. Agric. For. Meteorol. 54, 241–261. https://doi.org/10.1016/0168-1923(91)90008-E

Rodell, M., McWilliams, E.B., Famiglietti, J.S., Beaudoing, H.K., Nigro, J., 2011. Estimating evapotranspiration using an observation based terrestrial water budget. Hydrol. Process. 25, 4082–4092. https://doi.org/10.1002/hyp.8369

Rowshon, M.K., Amin, M.S.M., Mojid, M.A., Yaji, M., 2014. Estimated evapotranspiration of rice based on pan evaporation as a surrogate to lysimeter measurement. Paddy Water Environ. 12, 35–41. https://doi.org/10.1007/s10333-013-0356-4

Roy, D.P., Wulder, M.A., Gorelick, N., Hansen, M., Healey, S., Hostert, P., Huntington, J., Radeloff, V.C., Scambos, T., Schaaf, C., Woodcock, C.E., Zhu, Z., 2026. The next Landsat: Mission turning point? Remote Sens. Environ. 332, 115087. https://doi.org/10.1016/j.rse.2025.115087

Ruhoff, A.L., Paz, A.R., Aragao, L.E.O.C., Mu, Q., Malhi, Y., Collischonn, W., Rocha, H.R., Running, S.W., 2013. Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin. Hydrol. Sci. J. 58, 1658–1676. https://doi.org/10.1080/02626667.2013.837578

Running, S., Mu, Q., Zhao, M., Moreno, A., 2021. MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500m SIN Grid V061 | NASA Earthdata.

Rwasoka, D.T., Gumindoga, W., Gwenzi, J., 2011. Estimation of actual evapotranspiration using the Surface Energy Balance System (SEBS) algorithm in the Upper Manyame catchment in Zimbabwe. Phys. Chem. Earth 36, 736–746. https://doi.org/10.1016/j.pce.2011.07.035

Saltelli, A. (Ed.), 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Wiley, Hoboken, NJ.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environ. Model. Softw. 114, 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

Saltelli, A., Funtowicz, S., 2014. When All Models Are Wrong. Issues Sci. Technol. 30, 79–85.

Saltelli, A., Jakeman, A., Razavi, S., Wu, Q., 2021. Sensitivity analysis: A discipline coming of age. Environ. Model. Softw. 146, 105226. https://doi.org/10.1016/j.envsoft.2021.105226

Saltelli, A., Pereira, Â.G., Sluijs, J.P.V. der, Funtowicz, S., 2014. What do I make of your latinorum? Sensitivity auditing of mathematical modelling. Int. J. Foresight Innov. Policy.

Savenije, H.H.G., 2004. The importance of interception and why we should delete the term evapotranspiration from our vocabulary. Hydrol. Process. 18, 1507–1511. https://doi.org/10.1002/hyp.5563

Savenije, H.H.G., 1995. New definitions for moisture recycling and the relationship with land-use changes in the Sahel. J. Hydrol. 167, 57–78. https://doi.org/10.1016/0022-1694(94)02632-L

Schoups, G., Nasseri, M., 2021. GRACEfully Closing the Water Balance: A Data-Driven Probabilistic Approach Applied to River Basins in Iran. Water Resour. Res. 57, e2020WR029071. https://doi.org/10.1029/2020WR029071

Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., Wagner, W., 2008. A possible solution for the problem of estimating the error structure of global soil moisture data sets. Geophys. Res. Lett. 35. https://doi.org/10.1029/2008GL035599

Seijger, C., Chukalla, A., Bremer, K., Borghuis, G., Christoforidou, M., Mul, M., Hellegers, P., van Halsema, G., 2023. Agronomic analysis of WaPOR applications: Confirming conservative biomass water productivity in inherent and climatological variance of WaPOR data outputs. Agric. Syst. 211, 103712. https://doi.org/10.1016/j.agsy.2023.103712

Seijger, C., Urfels, A., Christoforidou, M., Hellegers, P., Borghuis, G., Langan, S., van Halsema, G., 2025. More food, but less land and water for nature: Why agricultural productivity gains did not materialize. Agric. Water Manag. 307, 109229. https://doi.org/10.1016/j.agwat.2024.109229

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. JAWRA J. Am. Water Resour. Assoc. 49, 577–591. https://doi.org/10.1111/jawr.12057

Senay, G.B., Leake, S., Nagler, P.L., Artan, G., Dickinson, J., Cordova, J.T., Glenn, E.P., 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. Hydrol. Process. 25, 4037–4049. https://doi.org/10.1002/hyp.8379

Senay, G.B., Parrish, G.E.L., Schauer, M., Friedrichs, M., Khand, K., Boiko, O., Kagone, S., Dittmeier, R., Arab, S., Ji, L., 2023. Improving the Operational Simplified Surface Energy Balance Evapotranspiration Model Using the Forcing and Normalizing Operation. Remote Sens. 15, 260. https://doi.org/10.3390/rs15010260

Senkondo, W., Munishi, S.E., Tumbo, M., Nobert, J., Lyon, S.W., 2019. Comparing Remotely-Sensed Surface Energy Balance Evapotranspiration Estimates in Heterogeneous and Data-Limited Regions: A Case Study of Tanzania's Kilombero Valley. Remote Sens. 11, 1289. https://doi.org/10.3390/rs11111289

Sequeira, E., Leão de Sousa, P., Correia, A.M., Rolim, J., 2025. The Current Status of Irrigated Agriculture in Cape Verde and Its Link to Water Scarcity. Agronomy 15, 1625. https://doi.org/10.3390/agronomy15071625

Servia, H., Pareeth, S., Michailovsky, C.I., de Fraiture, C., Karimi, P., 2022. Operational framework to predict field level crop biomass using remote sensing and data driven models. Int. J. Appl. Earth Obs. Geoinformation 108, 102725. https://doi.org/10.1016/j.jag.2022.102725

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., Fredrickson, S.E., Hughes, T.W., 2000. Quality Assurance Procedures in the Oklahoma Mesonetwork.

Shao, X., Zhang, Y., Liu, C., Chiew, F.H.S., Tian, J., Ma, N., Zhang, X., 2022. Can Indirect Evaluation Methods and Their Fusion Products Reduce Uncertainty in Actual Evapotranspiration Estimates? Water Resour. Res. 58, e2021WR031069. https://doi.org/10.1029/2021WR031069

Shapiro, S.S., Wilk, M.B., 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52, 591–611. https://doi.org/10.2307/2333709

Sharma, V., Kilic, A., Irmak, S., 2016. Impact of scale/resolution on evapotranspiration from Landsat and MODIS images. Water Resour. Res. 52, 1800–1819. https://doi.org/10.1002/2015WR017772

Shu Fen Shun, 1982. Moisture and heat transport in a soil layer forced by atmospheric conditions (M.Sc. thesis). University of Connecticut.

Shuttleworth, W.J., Wallace, J.S., 1985. Evaporation from sparse crops-an energy combination theory. Q. J. R. Meteorol. Soc. 111, 839–855. https://doi.org/10.1002/qj.49711146910

Singer, M.B., Asfaw, D.T., Rosolem, R., Cuthbert, M.O., Miralles, D.G., MacLeod, D., Quichimbo, E.A., Michaelides, K., 2021. Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present. Sci. Data 8, 224. https://doi.org/10.1038/s41597-021-01003-9

Siu, L.W., Zeng, X., Sorooshian, A., Cairns, B., Ferrare, R.A., Hair, J.W., Hostetler, C.A., Painemal, D., Schlosser, J.S., 2024. Summarizing multiple aspects of triple collocation analysis in a single diagram. Front. Remote Sens. 5. https://doi.org/10.3389/frsen.2024.1395442

Sjoberg, J.P., Anthes, R.A., Rieckh, T., 2021. The Three-Cornered Hat Method for Estimating Error Variances of Three or More Atmospheric Datasets. Part I: Overview and Evaluation. J. Atmospheric Ocean. Technol. 38, 555–572. https://doi.org/10.1175/JTECH-D-19-0217.1

Slivinski, L.C., 2018. Historical Reanalysis: What, How, and Why? J. Adv. Model. Earth Syst. 10, 1736–1739. https://doi.org/10.1029/2018MS001434

Sobol′, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul., The Second IMACS Seminar on Monte Carlo Methods 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Villaume, S., Haimberger, L., Woollen, J., Buontempo, C., Thépaut, J.-N., 2024. The ERA5 global reanalysis from 1940 to 2022. Q. J. R. Meteorol. Soc. 150, 4014–4048. https://doi.org/10.1002/qj.4803

Song, L., Bateni, S.M., Xu, Y., Xu, T., He, X., Ki, S.J., Liu, S., Ma, M., Yang, Y., 2021. Reconstruction of remotely sensed daily evapotranspiration data in cloudy-sky conditions. Agric. Water Manag. 255, 107000. https://doi.org/10.1016/j.agwat.2021.107000

Sörensson, A.A., Ruscica, R.C., 2018. Intercomparison and Uncertainty Assessment of Nine Evapotranspiration Estimates Over South America. Water Resour. Res. 54, 2891–2908. https://doi.org/10.1002/2017WR021682

Steduto, P., Hsiao, T.C., Fereres, E., Raes, D., 2012. Crop yield response to water, FAO irrigation and drainage paper. Food and Agriculture Organization of the United Nations, Rome.

Stein, M., 1987. Large Sample Properties of Simulations Using Latin Hypercube Sampling. Technometrics 29, 143–151. https://doi.org/10.1080/00401706.1987.10488205

Stisen, S., Soltani, M., Mendiguren, G., Langkilde, H., Garcia, M., Koch, J., 2021. Spatial patterns in actual evapotranspiration climatologies for europe. Remote Sens. 13. https://doi.org/10.3390/rs13122410

Stoffelen, A., 1998. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. J. Geophys. Res. Oceans 103, 7755–7766. https://doi.org/10.1029/97JC03180

Strobl, B., Etter ,Simon, van Meerveld ,Ilja, and Seibert, J., 2020. Accuracy of crowdsourced streamflow and stream level class estimates. Hydrol. Sci. J. 65, 823–841. https://doi.org/10.1080/02626667.2019.1578966

Stull, R.B. (Ed.), 1988. An Introduction to Boundary Layer Meteorology. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-009-3027-8

Su, C.-H., Ryu, D., Crow, W.T., Western, A.W., 2014. Beyond triple collocation: Applications to soil moisture monitoring. J. Geophys. Res. Atmospheres 119, 6419–6439. https://doi.org/10.1002/2013JD021043

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–100. https://doi.org/10.5194/hess-6-85-2002

Sur, C., Kang, S., Kim, J., Choi, M., 2015. Remote sensing-based evapotranspiration algorithm: a case study of all sky conditions on a regional scale. GIScience Remote Sens. 52, 627–642. https://doi.org/10.1080/15481603.2015.1056288

TAHMO, 2023. TAHMO Weather Station Manual [WWW Document]. URL https://usercontent.one/wp/tahmo.org/wp-content/uploads/2023/06/TAHMOSTATIONmanual2023.pdf (accessed 7.29.24).

Talsma, C.J., Good, S.P., Miralles, D.G., Fisher, J.B., Martens, B., Jimenez, C., Purdy, A.J., 2018. Sensitivity of Evapotranspiration Components in Remote Sensing-Based Models. Remote Sens. 10, 1601. https://doi.org/10.3390/rs10101601

Taylor, J., 1997. Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements, 2nd Edition, Published by University Science Books.

ter Horst, R., Alba, R., Vos, J., Rusca, M., Godinez-Madrigal, J., Babel, L.V., Veldwisch, G.J., Venot, J.-P., Bonté, B., Walker, D.W., Krueger, T., 2024. Making a case for power-sensitive water modelling: a literature review. Hydrol. Earth Syst. Sci. 28, 4157–4186. https://doi.org/10.5194/hess-28-4157-2024

ter Horst, R., Alba, R., Vos, J., Rusca, M., Godinez-Madrigal, J., Babel, L.V., Veldwisch, G.J., Venot, J.-P., Bonté, B., Walker, D.W., Krueger, T., 2023. Making a case for power-sensitive water modelling: a literature review. Hydrol. Earth Syst. Sci. Discuss. 1–31. https://doi.org/10.5194/hess-2023-164

Teuling, A.J., 2018. A Forest Evapotranspiration Paradox Investigated Using Lysimeter Data. Vadose Zone J. 17, 170031. https://doi.org/10.2136/vzj2017.01.0031

The FLUXNET Coordination Team, Torn, M., Pastorello, G., Papale, D., Sabbatini, S., Isaac, P., Arndt, S., Keenan, T., Moore, D., Novick, K., 2025. Original FLUXNET Data System announcement. FLUXNET. URL https://fluxnet.org/original-fluxnet-data-system-announcement/ (accessed 11.6.25).

Thom, A.S., Stewart, J.B., Oliver, H.R., Gash, J.H.C., 1975. Comparison of aerodynamic and energy budget estimates of fluxes over a pine forest. Q. J. R. Meteorol. Soc. 101, 93–105. https://doi.org/10.1002/qj.49710142708

Thornthwaite, C.W., 1948. An Approach toward a Rational Classification of Climate. Geogr. Rev. 38, 55–94. https://doi.org/10.2307/210739

Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L., Uhlenbrook, S., 2014. Comparison of different evaporation estimates over the African continent. Hydrol. EARTH Syst. Sci. 18, 193–212. https://doi.org/10.5194/hess-18-193-2014

Tran, B., 2023. Systematic Quantitative Literature Review - Uncertainty assessment of Evapotranspiration Remote Sensing. https://doi.org/10.4121/797dcaff-56e3-45ae-a931-f6f4a3135d26

Tran, B.N., van der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G., Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. Hydrol. Earth Syst. Sci. 27, 4505–4528. https://doi.org/10.5194/hess-27-4505-2023

Tsamalis, C., 2022. Clarifications on the equations and the sample number in triple collocation analysis using SST observations. Remote Sens. Environ. 272, 112936. https://doi.org/10.1016/j.rse.2022.112936

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol. 103, 279–300. https://doi.org/10.1016/S0168-1923(00)00123-4

United Nations, 2021. Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators: note / by the Secretary-General. UN,.

van de Giesen, N., Hut, R., Selker, J., 2014. The Trans-African Hydro-Meteorological Observatory (TAHMO). WIREs Water 1, 341–348. https://doi.org/10.1002/wat2.1034

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., Oberski, D.L., 2021. An open source machine learning framework for efficient and transparent systematic reviews. Nat. Mach. Intell. 3, 125–133. https://doi.org/10.1038/s42256-020-00287-7

van der Ent, R.J., Savenije, H.H.G., Schaefli, B., Steele-Dunne, S.C., 2010. Origin and fate of atmospheric moisture over continents. Water Resour. Res. 46. https://doi.org/10.1029/2010WR009127

van Dijk, A.I.J.M., Renzullo, L.J., 2011. Water resource monitoring systems and the role of satellite observations. Hydrol. Earth Syst. Sci. 15, 39–55. https://doi.org/10.5194/hess-15-39-2011

van Halsema, G.E., Vincent, L., 2012. Efficiency and productivity terms for water management: A matter of contextual relativism versus general absolutism. Agric. Water Manag. 108, 9–15. https://doi.org/10.1016/j.agwat.2011.05.016

Van Niel, T.G., McVicar, T.R., Roderick, M.L., van Dijk, A.I.J.M., Beringer, J., Hutley, L.B., van Gorsel, E., 2012. Upscaling latent heat flux for thermal remote sensing studies: Comparison of alternative approaches and correction of bias. J. Hydrol. 468–469, 35–46. https://doi.org/10.1016/j.jhydrol.2012.08.005

Van Stan II, J.T., Simons, J., 2025. Water Models as Geographical Chimera: Precipitation Interception Routines as an Example of “Patchwork Empiricism” 18.

van Steenbergen, F., Mulder, E., Bremer, K., Mul, M., Chukalla, A., Chevalking, S., Deligianni, A., Pluijm, L., Deribe, M., 2022. Compendium of Approaches to Improve Water Productivity.

Varisco, D., 2019. Pumping Yemen Dry: A History of Yemen’s Water Crisis. Hum. Ecol. 47, 317–329. https://doi.org/10.1007/s10745-019-0070-y

Vendrame, N., Tezza, L., Pitacco, A., 2020. Comparison of sensible heat fluxes by large aperture scintillometry and eddy covariance over two contrasting−climate vineyards. Agric. For. Meteorol. 288–289, 108002. https://doi.org/10.1016/j.agrformet.2020.108002

Vinukollu, R.K., Meynadier, R., Sheffield, J., Wood, E.F., 2011a. Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends. Hydrol. Process. 25, 3993–4010. https://doi.org/10.1002/hyp.8393

Vinukollu, R.K., Wood, E.F., Ferguson, C.R., Fisher, J.B., 2011b. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. Remote Sens. Environ. 115, 801–823. https://doi.org/10.1016/j.rse.2010.11.006

Vogel, A., Ménard, R., 2023. How far can the statistical error estimation problem be closed by collocated data? Nonlinear Process. Geophys. 30, 375–398. https://doi.org/10.5194/npg-30-375-2023

Vogelzang, J., 2024. Triple collocation manual v2.0.

Vogelzang, J., Stoffelen, A., 2022. On the Accuracy and Consistency of Quintuple Collocation Analysis of In Situ, Scatterometer, and NWP Winds. Remote Sens. 14, 4552. https://doi.org/10.3390/rs14184552

Vogelzang, J., Stoffelen, A., 2021. Quadruple Collocation Analysis of In-Situ, Scatterometer, and NWP Winds. J. Geophys. Res. Oceans 126, e2021JC017189. https://doi.org/10.1029/2021JC017189

Vogelzang, J., Stoffelen, A., Verhoef, A., 2022. The Effect of Error Non-Orthogonality on Triple Collocation Analyses. Remote Sens. 14, 4268. https://doi.org/10.3390/rs14174268

Volk, J.M., Huntington, J., Melton, F.S., Allen, R., Anderson, M.C., Fisher, J.B., Kilic, A., Senay, G., Halverson, G., Knipper, K., Minor, B., Pearson, C., Wang, T., Yang, Y., Evett, S., French, A.N., Jasoni, R., Kustas, W., 2023a. Development of a Benchmark Eddy Flux Evapotranspiration Dataset for Evaluation of Satellite-Driven Evapotranspiration Models Over the CONUS. Agric. For. Meteorol. 331, 109307. https://doi.org/10.1016/j.agrformet.2023.109307

Volk, J.M., Huntington, J.L., Melton, F., Minor, B., Wang, T., Anapalli, S., Anderson, R.G., Evett, S., French, A., Jasoni, R., Bambach, N., Kustas, W.P., Alfieri, J., Prueger, J., Hipps, L., McKee, L., Castro, S.J., Alsina, M.M., McElrone, A.J., Reba, M., Runkle, B., Saber, M., Sanchez, C., Tajfar, E., Allen, R., Anderson, M., 2023b. Post-processed data and graphical tools for a CONUS-wide eddy flux evapotranspiration dataset. Data Brief 48, 109274. https://doi.org/10.1016/j.dib.2023.109274

Volk, J.M., Huntington, J.L., Melton, F., Minor, B., Wang, T., Anapalli, S.S., Anderson, R.G., Evett, S.R., French, A.N., Jasoni, R., Bambach, N., Kustas, W.P., Alfieri, J.G., Prueger, J., Hipps, L., McKee, L., Bustamante, S.J.C., Alsina, M. del M., McElrone, A., Reba, M., Runkle, B., Saber, M., Sanchez, C.A., Tajfar, E., Allen, R.G., Anderson, M., 2023c. Post-processed data and graphical tools for a CONUS-wide eddy flux evapotranspiration dataset. https://doi.org/10.5281/zenodo.7636781

Volk, J.M., Huntington, J.L., Melton, F.S., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Ruhoff, A., Senay, G.B., Minor, B., Morton, C., Ott, T., Johnson, L., Andrade, B.C.D., Carrara, W., Doherty, C.T., Dunkerly, C., Friedrichs, M., Guzman, A., Hain, C., Halverson, G., Kang, Y., Knipper, K., Laipelt, L., Ortega-Salazar, S., Pearson, C., Parrish, G.E.L., Purdy, A., ReVelle, P., Wang, T., Yang, Y., 2023d. OpenET model data for assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. https://doi.org/10.5281/zenodo.10119477

von Hoyningen-Hune, J., 1983. Die Interzeption des Niederschlages in landwirtschaftlichen Pflanzenbeständen. Interzeption Niederschlages Landwirtsch. Pflanzenbeständen 1–53.

Walsh, R., 2003. The methods of reflexivity. Humanist. Psychol. 31, 51–66. https://doi.org/10.1080/08873267.2003.9986934

Wang, J., Zhuang, J., Wang, W., Liu, S., Xu, Z., 2015. Assessment of Uncertainties in Eddy Covariance Flux Measurement Based on Intensive Flux Matrix of HiWATER-MUSOEXE. IEEE Geosci. Remote Sens. Lett. 12, 259–263. https://doi.org/10.1109/LGRS.2014.2334703

Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Rev. Geophys. 50. https://doi.org/10.1029/2011RG000373

Wang, L., Lang, Q., Wang, Z., Feng, L., Zhang, M., Qin, W., 2024a. Quantifying and Mitigating Errors in Estimating Downward Surface Shortwave Radiation Caused by Cloud Mask Data. IEEE Trans. Geosci. Remote Sens. 62, 1–15. https://doi.org/10.1109/TGRS.2024.3438879

Wang, L., Lu, Y., Wang, Z., Li, H., Zhang, M., 2024b. Hourly solar radiation estimation and uncertainty quantification using hybrid models. Renew. Sustain. Energy Rev. 202, 114727. https://doi.org/10.1016/j.rser.2024.114727

Wang, S., Pan, M., Mu, Q., Shi, X., Mao, J., Brümmer, C., Jassal, R.S., Krishnan, P., Li, J., Black, T.A., 2015. Comparing Evapotranspiration from Eddy Covariance Measurements, Water Budgets, Remote Sensing, and Land Surface Models over Canadaa,b. J. Hydrometeorol. 16, 1540–1560. https://doi.org/10.1175/JHM-D-14-0189.1

Wang, T., Alfieri, J., Mallick, K., Arias-Ortiz, A., Anderson, M., Fisher, J.B., Girotto, M., Szutu, D., Verfaillie, J., Baldocchi, D., 2024. How advection affects the surface energy balance and its closure at an irrigated alfalfa field. Agric. For. Meteorol. 357, 110196. https://doi.org/10.1016/j.agrformet.2024.110196

Wang, Y., Hu, J., Li, R., Song, B., Hailemariam, M., Fu, Y., Duan, J., 2023. Increasing Cloud Coverage Deteriorates Evapotranspiration Estimating Accuracy From Satellite, Reanalysis and Land Surface Models Over East Asia. Geophys. Res. Lett. 50, e2022GL102706. https://doi.org/10.1029/2022GL102706

Wang, Y.Q., Xiong, Y.J., Qiu, G.Y., Zhang, Q.T., 2016. Is scale really a challenge in evapotranspiration estimation? A multi-scale study in the Heihe oasis using thermal remote sensing and the three-temperature model. Agric. For. Meteorol., Oasis-desert system 230–231, 128–141. https://doi.org/10.1016/j.agrformet.2016.03.012

Warm Winter 2020 Team, ICOS Ecosystem Thematic Centre, 2022. Warm Winter 2020 ecosystem eddy covariance flux product for 73 stations in FLUXNET-Archive format—release 2022-1 (Version 1.0). https://doi.org/10.18160/2G60-ZHAK

Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L., van Griensven, A., 2020. Can we trust remote sensing evapotranspiration products over Africa? Hydrol. Earth Syst. Sci. 24, 1565–1586. https://doi.org/10.5194/hess-24-1565-2020

Westerhoff, R.S., 2015. Using uncertainty of Penman and Penman–Monteith methods in combined satellite and ground-based evapotranspiration estimates. Remote Sens. Environ. 169, 102–112. https://doi.org/10.1016/j.rse.2015.07.021

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, 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., Valentini, R., Verma, S., 2002. Energy balance closure at FLUXNET sites. Agric. For. Meteorol., FLUXNET 2000 Synthesis 113, 223–243. https://doi.org/10.1016/S0168-1923(02)00109-0

Woodhouse, I.H., 2021. On 'ground' truth and why we should abandon the term. J. Appl. Remote Sens. 15, 041501. https://doi.org/10.1117/1.JRS.15.041501

Wu, J., Feng, Y., Zheng, C., Zeng, Z., 2023. Dense flux observations reveal the incapability of evapotranspiration products to capture the heterogeneity of evapotranspiration. J. Hydrol. 622, 129743. https://doi.org/10.1016/j.jhydrol.2023.129743

Wu, X., Xiao, Q., Wen, J., You, D., 2019a. Direct Comparison and Triple Collocation: Which Is More Reliable in the Validation of Coarse-Scale Satellite Surface Albedo Products. J. Geophys. Res. Atmospheres 124, 5198–5213. https://doi.org/10.1029/2018JD029937

Wu, X., Xiao, Q., Wen, J., You, D., Hueni, A., 2019b. Advances in quantitative remote sensing product validation: Overview and current status. Earth-Sci. Rev. 196, 102875. https://doi.org/10.1016/j.earscirev.2019.102875

Xu, C., Wang, W., Hu, Y., Liu, Y., 2024. Evaluation of ERA5, ERA5-Land, GLDAS-2.1, and GLEAM potential evapotranspiration data over mainland China. J. Hydrol. Reg. Stud. 51, 101651. https://doi.org/10.1016/j.ejrh.2023.101651

Xu, T., Guo, Z., Xia, Y., Ferreira, V.G., Liu, S., Wang, K., Yao, Y., Zhang, X., Zhao, C., 2019. Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. J. Hydrol. 578, 124105. https://doi.org/10.1016/j.jhydrol.2019.124105

Xu, T., Liu, S., Xu, L., Chen, Y., Jia, Z., Xu, Z., Nielson, J., 2015. Temporal Upscaling and Reconstruction of Thermal Remotely Sensed Instantaneous Evapotranspiration. Remote Sens. 7, 3400–3425. https://doi.org/10.3390/rs70303400

Yang, X., Tian, S., You, W., Jiang, Z., 2021. Reconstruction of continuous GRACE/GRACE-FO terrestrial water storage anomalies based on time series decomposition. J. Hydrol. 603, 127018. https://doi.org/10.1016/j.jhydrol.2021.127018

Yang, X., Yong, B., Ren, L., Zhang, Y., Long, D., 2017. Multi-scale validation of GLEAM evapotranspiration products over China via ChinaFLUX ET measurements. Int. J. Remote Sens. 38, 5688–5709. https://doi.org/10.1080/01431161.2017.1346400

Yang, Y., Guan, H., Long, D., Liu, B., Qin, G., Qin, J., Batelaan, O., 2015a. Estimation of Surface Soil Moisture from Thermal Infrared Remote Sensing Using an Improved Trapezoid Method. Remote Sens. 7, 8250–8270. https://doi.org/10.3390/rs70708250

Yang, Y., Long, D., Guan, H., Liang, W., Simmons, C., Batelaan, O., 2015b. Comparison of three dual-source remote sensing evapotranspiration models during the MUSOEXE-12 campaign: Revisit of model physics: Two-source remote sensing ET model comparision. Water Resour. Res. 51, 3145–3165. https://doi.org/10.1002/2014WR015619

Yao, Y., Liang, S., Li, X., Zhang, Y., Chen, J., Jia, K., Zhang, X., Fisher, J.B., Wang, X., Zhang, L., Xu, J., Shao, C., Posse, G., Li, Y., Magliulo, V., Varlagin, A., Moors, E.J., Boike, J., Macfarlane, C., Kato, T., Buchmann, N., Billesbach, D.P., Beringer, J., Wolf, S., Papuga, S.A., Wohlfahrt, G., Montagnani, L., Ardö, J., Paul-Limoges, E., Emmel, C., Hörtnagl, L., Sachs, T., Gruening, C., Gioli, B., López-Ballesteros, A., Steinbrecher, R., Gielen, B., 2017. Estimation of high-resolution terrestrial

evapotranspiration from Landsat data using a simple Taylor skill fusion method. J. Hydrol. 553, 508–526. https://doi.org/10.1016/j.jhydrol.2017.08.013

Yebra, M., Van Dijk, A., Leuning, R., Huete, A., Guerschman, J.P., 2013. Evaluation of optical remote sensing to estimate actual evapotranspiration and canopy conductance. Remote Sens. Environ. 129, 250–261. https://doi.org/10.1016/j.rse.2012.11.004

Yi, K., Senay, G.B., Fisher, J.B., Wang, L., Suvočarev, K., Chu, H., Moore, G.W., Novick, K.A., Barnes, M.L., Keenan, T.F., Mallick, K., Luo, X., Missik, J.E.C., Delwiche, K.B., Nelson, J.A., Good, S.P., Xiao, X., Kannenberg, S.A., Ahmadi, A., Wang, T., Bohrer, G., Litvak, M.E., Reed, D.E., Oishi, A.C., Torn, M.S., Baldocchi, D., 2024. Challenges and Future Directions in Quantifying Terrestrial Evapotranspiration. Water Resour. Res. 60, e2024WR037622. https://doi.org/10.1029/2024WR037622

Yilmaz, M.T., Anderson, M.C., Zaitchik, B., Hain, C.R., Crow, W.T., Ozdogan, M., Chun, J.A., Evans, J., 2014. Comparison of prognostic and diagnostic surface flux modeling approaches over the Nile River basin. Water Resour. Res. 50, 386–408. https://doi.org/10.1002/2013WR014194

Yilmaz, M.T., Crow, W.T., 2014. Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis. https://doi.org/10.1175/JHM-D-13-0158.1

Yuan, W., Liu, S., Liang, S., Tan, Z., Liu, H., Young, C., 2012. Estimations of Evapotranspiration and Water Balance with Uncertainty over the Yukon River Basin. Water Resour. Manag. 26, 2147–2157. https://doi.org/10.1007/s11269-012-0007-3

Zeng, R., Cai, X., 2018. Hydrologic Observation, Model, and Theory Congruence on Evapotranspiration Variance: Diagnosis of Multiple Observations and Land Surface Models. Water Resour. Res. 54, 9074–9095. https://doi.org/10.1029/2018WR022723

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. Geoinformation 42, 150–161. https://doi.org/10.1016/j.jag.2015.06.006

Zhang, K., Kimball, J.S., Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation. Wiley Interdiscip. Rev. Water 3, 834–853. https://doi.org/10.1002/wat2.1168

Zhang, K., Zhu, G., Ma, J., Yang, Y., Shang, S., Gu, C., 2019. Parameter Analysis and Estimates for the MODIS Evapotranspiration Algorithm and Multiscale Verification. Water Resour. Res. 55, 2211–2231. https://doi.org/10.1029/2018WR023485

Zhang, W., Nelson, J.A., Miralles, D.G., Mauder, M., Migliavacca, M., Poyatos, R., Reichstein, M., Jung, M., 2024. A New Post-Hoc Method to Reduce the Energy Imbalance in Eddy Covariance Measurements. Geophys. Res. Lett. 51, e2023GL107084. https://doi.org/10.1029/2023GL107084

Zhang, X., Wu, J., Wu, H., Chen, H., Zhang, T., 2013. Improving temporal extrapolation for daily evapotranspiration using radiation measurements. J. Appl. Remote Sens. 7, 073538. https://doi.org/10.1117/1.JRS.7.073538

Zhang, Y., Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sens. Environ. 222, 165–182. https://doi.org/10.1016/j.rse.2018.12.031

Zhang, Y., Peña-Arancibia, J.L., McVicar, T.R., Chiew, F.H.S., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y.Y., Miralles, D.G., Pan, M., 2016. Multi-decadal trends in global terrestrial evapotranspiration and its components. Sci. Rep. 6, 19124. https://doi.org/10.1038/srep19124

Zhao, Y., Hoeltgebaum, L.E.B., Kukal, M.S., Zhao, M., 2025. Leading Satellite-Based Evapotranspiration Products Insufficiently Capture Interannual Variability: Evidence From GRACE/FO and In Situ Observations. Geophys. Res. Lett. 52, e2025GL116784. https://doi.org/10.1029/2025GL116784

Zheng, H., Bian, Q., Yin, Y., Ying, H., Yang, Q., Cui, Z., 2018. Closing water productivity gaps to achieve food and water security for a global maize supply. Sci. Rep. 8, 14762. https://doi.org/10.1038/s41598-018-32964-4

Zhong, L., Xu, K., Ma, Y., Huang, Z., Wang, X., Ge, N., 2019. Evapotranspiration estimation using surface energy balance system model: A case study in the Nagqu River Basin. Atmosphere 10, 1–13. https://doi.org/10.3390/atmos10050268

Zhou, Y., Pan, X., Yang, Z., Wang, Z., Guluzade, R., Yuan, J., Liu, S., Ding, X., Ma, W., Yang, Y., 2025. The comparison between single-point method and footprint-integrated validation method of the remote-sensing retrieval of evapotranspiration: a case study at Daman site. Int. J. Remote Sens. 0, 1–21. https://doi.org/10.1080/01431161.2024.2412803

Zhu, W., Tian, S., Wei, J., Jia, S., Song, Z., 2022. Multi-scale evaluation of global evapotranspiration products derived from remote sensing images: Accuracy and uncertainty. J. Hydrol. 611, 127982. https://doi.org/10.1016/j.jhydrol.2022.127982

Zoebl, D., 2006. Is water productivity a useful concept in agricultural water management? Agric. Water Manag. 84, 265–273. https://doi.org/10.1016/j.agwat.2006.03.002

Zolkos, S.G., Goetz, S.J., Dubayah, R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sens. Environ. 128, 289–298. https://doi.org/10.1016/j.rse.2012.10.017

Zwarteveen, M., Smit, H., Guzmán, C.D., Fantini, E., Rap, E., 2019. Questions of Environmental Representation in a Nonmodern World 23, 24.

CODE AND DATA AVAILABILITY

CHAPTER 2

The systematic categorization and analysis of the reviewed articles are available at https://doi.org/10.4121/797dcaff-56e3-45ae-a931-f6f4a3135d26.v2

CHAPTER 3

The reported RMSE data from the reviewed articles that used eddy covariance to validate satellite-based estimates of evapotranspiration are available at https://doi.org/10.4121/e6e1713a-0c2b-4775-a7f4-9e6e0b2cf40f.v2

The reported PBIAS data from the reviewed articles that used eddy covariance to validate satellite-based estimates of evapotranspiration are available at https://doi.org/10.5281/zenodo.17900376

CHAPTER 4

Code and results for evaluation of reanalysis datasets and uncertainty propagation in this chapter are available at https://doi.org/10.5281/zenodo.13970798

CHAPTER 5

Code and results for in-situ validation and triple collocation are available at https://doi.org/10.5281/zenodo.17901733

APPENDICES

Appendix A SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A-1. Methods to measure Evapotranspiration (ET) or Latent Heat Flux (LE). N/A: Not Applicable.

Method	Instruments	Measurement	Intermediate calculations	Method to derive ET_a or LE
Eddy covariance (EC)	Infrared Gas Analyser (IRGA)	Water vapor concentration	Calculation of mixing ratio of water vapor and dry air	Derivation of LE from eddy covariance using Reynolds averaging (Foken et al., 2012; Montgomery, 1948)
	Sonic anemometer	3-direction wind speed	Calculation of covariance of vertical wind velocity and mixing ratio	
Lysimetry	Weighing lysimeter, Drainage lysimeter or Soil water content sensor (e.g. neutron probe)	Soil weight or Soil water content	Calculation of change in soil water storage	Derivation of ET as residual of soil water balance (e.g., Teuling, 2018)
	Rain gauge	Rainfall	Calculation of water inflow	
	Flow meter	Irrigation		
Scintillometry	Large Aperture Scintillometer (LAS) • Transmitter • Detector	Log-variance of the intensity variations of the received light beam signals	Calculation of structure parameters of refractive index, temperature, specific humidity	Derivation of sensible heat (H) and Latent heat flux (LE) based on Monin-Obukhov Similarity Theory (Beyrich et al., 2021; Frehlich, 1992; Hill, 1997)
	Temperature probe	Air temperature at different heights	Determination of sensible heat (H) direction based on temperature gradient	

Bowen ratio energy balance (BREB)	Temperature probe, Relative humidity probe, Barometer	Gradient of atmospheric temperature, air moisture content (actual water vapor pressure)	Calculation of Bowen ratio	Derivation of H and LE from Bowen ratio and surface energy balance
	Net radiometer, Wind speed and direction sensor	Net radiation	Calculation of sum of H and LE from surface energy balance equation	
	Soil heat flow plates, Soil moisture probe	Soil heat flux		
Surface renewal (SR)	Fine-wire thermocouple	High frequency temperature fluctuation	Calculation of sensible heat (H) based on a solution of the scalar conservation	Derivation of LE as residual of surface energy balance (Paw U et al., 1995)
	Net radiometer, Wind speed and direction sensor	Net radiation	Calculation of sum of H and LE from surface energy balance equation	
	Soil heat flow plates, Soil moisture probe	Soil heat flux		
Advection-aridity (AA)	Anemometer Pressure sensor Thermometer Hygrometer	Windspeed Air pressure Air temperature Relative humidity	Calculation of wet surface ET (P-T), potential ET (Penman), and relative evaporation	Derivation of ET based on complementary relationship (Brutsaert and Stricker, 1979)
Combinatory method (CM)	Anemometer Pressure sensor Thermometer Hygrometer	Windspeed Air pressure Air temperature Relative humidity	Calculation of LE and H without stability correction	Derivation of LE using stability correction (Thom et al., 1975; Zhong et al., 2019)
	Temperature probe, Net radiometer	Net radiation Soil heat flux	Calculation of stability function	

FAO-56 Crop coefficient	Net radiometer Anemometer Pressure sensor Thermometer Hygrometer	Net radiation Windspeed Air pressure Air temperature Relative humidity	Calculation of FAO-56 potential ET for a reference crop	Derivation of ET using crop coefficients (*need crop information) (Allen et al., 1998)
Evaporation pan	Class A pan	Depth of water evaporated	Calculation of evaporation of open water	N/A
Sap flow	Thermal dissipation probe	Movement of xylem sap	Calculation of plant transpiration (Lu et al., 2004)	N/A

Appendix B SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Table B-1. Metrics used for uncertainty assessment in the reviewed articles in the systematic quantitative literature review.

Name	Formula	Unit	Best value
Standard deviation (σ_y)	$\sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}$ y_i: estimate from remotely sensed data x_i: estimate of in-situ data $\bar{y}$: average of y_i	Unit of x	0
Variance (σ_y^2)	$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$	Squared unit of x	0
Mean error (ME), Mean bias error (MBE), Bias	$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N} y_i - x_i$ x_i: estimate of in-situ data	Unit of x	0
Relative error (RE) or Mean error percentage (MEP) or Percent Bias (PBIAS)	$\mathrm{PBIAS} = \mathrm{MEP} = \frac{\mathrm{ME}}{\bar{x}} \times 100$	%	0
Normalized mean bias (NMB)	$\mathrm{NMB} = \frac{\mathrm{ME}}{\sigma_x}$	-	0
Mean absolute error (MAE) or mean absolute difference (MAD)	$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i - y_i\rvert$	Unit of x	0
Mean absolute percentage error (MAPE) or mean absolute percentage difference (MAPD)	$\mathrm{MAPE} = \frac{\mathrm{MAE}}{\bar{x}} \times 100$ $\mathrm{MAPE} = \frac{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i - y_i\rvert}{\bar{x}} \times 100$	%	0

Pearson's coefficient of correlation (ρ)	$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$	-	1
Regression coefficient: slope (a)	$a = \frac{N\sum_{i=1}^{N}(x_i y_i)\sum_{i=1}^{N}(x_i)\sum_{i=1}^{N}(y_i)}{N\sum_{i=1}^{N}(x_i)^2 (\sum_{i=1}^{N}(x_i))^2}$	-	1
Regression coefficient: intercept b	$b = \frac{\sum_{i=1}^{N}(y_i)\sum_{i=1}^{N}(x_i)^2 - \sum_{i=1}^{N}(x_i)\sum_{i=1}^{N}(x_i y_i)}{N\sum_{i=1}^{N}(x_i)^2 (\sum_{i=1}^{N}(x_i))^2}$	-	0
Coefficient of determination (R^2)	$R_1^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$	-	1
	$R_2^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$	-	1
	$R_6^2 = \rho^2 = \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})\left(y_i - \bar{y}\right)\right]^2}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)\left(\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right)}$	-	1
	$R_7^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(x_i)^2}$	-	1
Mean squared error (MSE)	$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2$	Unit of x	0
Systematic MSE	$\mathrm{MSE}_s = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - x_i)^2$ $\hat{y}_i$: estimate of y_i based on the ordinary least-squares regression $\hat{y}_i = a + bx_i$	Squared unit of x	0
Unsystematic MSE	$\mathrm{MSE}_u = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$ $\hat{y}_i$: estimate of y_i based on the ordinary least-squares regression $\hat{y}_i = a + bx_i$	Squared unit of x	0
Root mean squared error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}$	Unit of x	0
Systematic RMSE	$\mathrm{sRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - x_i)^2}$ $\hat{y}_i$: estimate of y_i based on the ordinary least-squares regression $\hat{y}_i = a + bx_i$	Unit of x	0

Unsystematic RMSE	$\mathrm{usRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$ $\hat{y}_i$: estimate of y_i based on the ordinary least-squares regression $\hat{y}_i = a + bx_i$	Unit of x	0
Normalized RMSE or fractional RMSE	$\mathrm{nRMSE} = \frac{\mathrm{RMSE}}{\sigma_x}$	-	0
Relative RMSE	$\mathrm{rRMSE} = \frac{\mathrm{RMSE}}{\bar{x}} \times 100$	%	0
Centred or Unbiased RMSE	$\mathrm{uRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(x_i - \bar{x}) - (y_i - \bar{y})]^2}$	Unit of x	0
Coefficient of variation (CV)	$CV_y = \frac{\sigma_y}{\bar{y}}$	-	
Nash-Sutcliffe Efficiency (NSE)	$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$	-	1
Index of agreement (d_2)	$d_2 = 1 - \frac{\sum_{i=1}^{N}\omega_i(y_i - x_i)^2}{\sum_{i=1}^{N}\omega_i(\lvert y_i - \bar{y}\rvert + \lvert x_i - \bar{x}\rvert)^2}$ ω_i: irregularly weight that represent relative size of the i[th] interval or cell size.	-	1
Modified index of agreement (d_1)	$d_1 = 1 - \frac{\sum_{i=1}^{N}\omega_i\lvert y_i - x_i\rvert}{\sum_{i=1}^{N}\omega_i(\lvert y_i - \bar{y}\rvert + \lvert x_i - \bar{x}\rvert)}$ ω_i: irregularly weight that represent relative size of the i[th] interval or cell size.	-	1
Taylor skill score (TSC)	$\mathrm{TSC} = \frac{4(1+\rho)}{\left(\hat{\sigma}_f + \frac{1}{\hat{\sigma}_f}\right)^2 (1+\rho_0)}$ $\hat{\sigma}_f = \frac{\sigma_y}{\sigma_x}$ ρ_0: Maximum attainable correlation ($\hat{\sigma}_f \to 1,\ \rho \to \rho_0$)	-	4
Kling-Gupta Efficiency (KGE)	$\mathrm{KGE} = 1 - \sqrt{(\rho-1)^2 + \left(\frac{\sigma_y}{\sigma_x} - 1\right)^2 + \left(\frac{\bar{y}}{\bar{x}} - 1\right)^2}$	-	1
Standard error (SE)	$\mathrm{SE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N-1}}$	Unit of x and y	0

Similarity index (Ω)	$\Omega = \frac{m\sigma_b^2 - \sigma^2}{(m-1)\sigma^2}$ m: number of ensemble members σ^2: total variance of all members concatenated σ_b^2: variance of the time series that results from calculating the ensemble mean of each time step	-	1
Spatial Efficiency (SPAEF)	$SPAEF = 1 - \sqrt{(\rho-1)^2 + \left(\frac{CV_y}{CV_x} - 1\right)^2 + (\gamma-1)^2}$ $\gamma = \frac{\sum_{i=1}^{N} \min(K_i, L_i)}{\sum_{i=1}^{N} K_i}$ γ: histogram matching term K: histogram of reference map L: histogram of simulated map	-	1
Median symmetric accuracy (MSA)	$\text{MSA} = 100\left(e^{M(\lvert(Q)\rvert)} - 1\right)$ M: median of the data $Q = \frac{y}{x}$	%	0
Degree correlation (r_l)	$r_l = \frac{1}{\eta}\sum_{m=0}^{l}\left(C_{A_{lm}}C_{B_{lm}} + S_{A_{lm}}S_{B_{lm}}\right)$ $\eta = \sqrt{\sum_{m=0}^{l}\left(C_{A_{lm}}^2 + S_{A_{lm}}^2\right)}\sqrt{\sum_{m=0}^{l}\left(C_{B_{lm}}^2 + S_{B_{lm}}^2\right)}$ $C_{A_{lm}}$ and $S_{A_{lm}}$ are spherical harmonic coefficients of degree l and order m of dataset A $C_{B_{lm}}$ and $S_{B_{lm}}$ are spherical harmonic coefficients of degree l and order m of dataset B	-	1

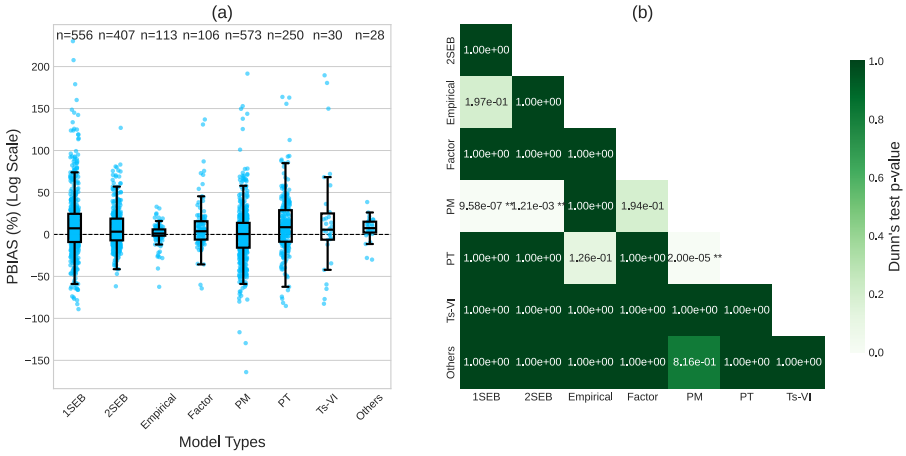


Figure B-1. PBIAS (%) in satellite-based ET_a estimates categorized by model type. (a) The box-and-whisker plot represents the 5th, 25th, 50th (median), 75th, and 95th percentiles of the distribution. (b) Dunn's post-hoc test results at a significant level of $\alpha = 0.05$ (significant *p*-values marked with **).

Table B-2. Comparison of PBIAS (%) in satellite-based ET estimates resulting from different treatments of energy balance closure of EC flux measurement.

Reference	Energy balance closure treatment	PBIAS (%)				
		N	**mean**	**std**	**min**	**max**
(Liaqat and Choi, 2015)	Unclosed	10	91.5	27.4	50.7	142.4
	Energy residual	10	-29.8	17.7	-58.1	-8.0
	Bowen ratio	10	7.1	15.4	-11.5	33.2
	Linear regression	10	17.7	10.2	6.3	36.8
(Sur et al., 2015)	Unclosed	2	-4.5	2.2	-6.1	-3.0
	Bowen ratio	2	-0.2	0.6	-0.7	0.2
(Byun et al., 2014)	Unclosed	12	106.4	97.5	2.4	278.0
	Energy residual	12	-33.5	14.0	-56.8	-8.7
	Bowen ratio	12	-9.4	22.9	-46.3	22.3
(Hwang and Choi, 2013)	Unclosed	4	-88.5	46.0	-129.5	-26.5
	Bowen ratio	4	-24.2	30.2	-51.1	16.5
(Knipper et al., 2020)	Unclosed	8	45.7	21.2	15.1	71.0
	Bowen ratio	8	24.4	10.3	15.8	40.8
(Guzinski et al., 2020)	Energy residual	4	-3.5	18.0	-23.1	12.2
	Bowen ratio	4	37.8	27.3	7.6	61.2
(Knipper et al., 2019)	Unclosed	3	26.7	6.4	20.4	33.2
	Energy residual	3	-1.5	3.2	-4.6	1.8
(Liaqat and Choi, 2017)	Unclosed	4	58.0	50.3	12.5	125.6
	Energy residual	4	-29.4	9.7	-39.7	-17.9
	Bowen ratio	4	-9.4	19.6	-26.8	18.7
(Michel et al., 2016)	Unclosed	16	46.1	57.8	-17.9	160.2
	Energy residual	16	21.4	53.8	-34.2	148.6
(Feng and Wang, 2013)	Unclosed	4	40.5	31.5	6.7	82.4
	Energy residual	4	-13.1	15.3	-31.1	3.1

Appendix C ETLOOK INPUT DATA DESCRIPTION

Table C-1. The required variables in the WaPOR-ETLook model and the corresponding input data that are pre-processed to obtain these variables. Source: The WaPOR methodology report (FAO, 2020) and WaPOR data manual (FRAME, 2020). A spatio-temporal variable varies in both space and time. A static variable varies spatially but not temporally. *The required variables are derived from reference values of each land cover type in look-up tables, which are roughly estimated in the literature.

Input data		ETLook required variables		
Type	Datasets	Variables	Spatio-temporal	Static
Satellite sensors	Surface reflectance	Normalised Difference Vegetation Index	●	
		Surface albedo	●	
		Latitude, longitude		●
	Land surface temperature	Daily land surface temperature	●	
Static input data	Digital Elevation Model	Elevation		●
		Slope		●
		Slope aspect		●
	Land Cover*	Minimum soil resistance		●
		Minimum stomatal resistance		●
		Observation height		●
		Maximum light use efficiency		●
Climate reanalysis	Air pressure at sea level	Daily air pressure	●	
		Hourly air pressure	●	
	Temperature at 2m	Daily air temperature	●	
		Hourly air temperature	●	
		Yearly air temperature amplitude		●
		Yearly air temperature optimum		●
	Windspeed	Daily windspeed	●	
		Hourly windspeed	●	
	Relative humidity	Daily specific humidity	●	
		Hourly specific humidity	●	
	Precipitation	Daily precipitation	●	
	Shortwave radiation	Daily solar radiation	●	
		Daily atmospheric transmissivity	●	
Globally calibrated	FAO-56	Longwave radiation slope		●
	FAO-56	Longwave radiation offset		●

There are two types of variables in WaPOR-ETLook: static variables and spatio-temporal variables. Static variables refer to quantities or characteristics that are considered constant at all times and varied in space (e.g. elevation), and spatio-temporal variables are those that varied in space and time. The spatio-temporal variables derived from satellite sensors are land surface temperature and surface reflectance, which are used to calculate NDVI and surface albedo. Like most satellite-based ET models, ETLook requires meteorological data (e.g., wind speed, air temperature, air pressure, relative humidity) to constrain the total available energy ($R_n - G$). The operational WaPOR-ETLook model obtains spatio-temporal meteorological variables from global climate reanalysis (e.g., the Goddard Earth Observing System – GEOS5) or global data products (e.g., the Climate Hazards Group InfraRed Precipitation with Stations – CHIRPS). The input data can be obtained from different sources, depending on the desired spatial resolution of output data product. WaPOR-ET data products have 3 level of spatial resolution: Level 1—250m, Level 2—100m, and Level 3—30m. For each level, different sources of input datasets are used to achieve the desired spatial resolution (**Table C-2**)

Table C-2. Changes in input data from version 2 to version 3 of WaPOR ET data production.

Level	Input variables	WaPOR v2	WaPOR v3	Justification of changes
All levels	Air temperature, Windspeed[1], Vapour pressure[2], Air pressure	MERRA2[3] (~56km, hourly); GEOS-5[4] (~28km, 6-hour)	ERA5[5] (~30km, hourly); AgERA5[6] (~11km, daily), GEOS-5[7] (~28km, hourly)	Improved spatial and temporal resolution, temporal consistency in historical data
	Land cover	Copernicus (100m)	WorldCover land cover (10m)	Improved spatial resolution
	DEM	SRTM (90m)	Copernicus GLO-90 (90m)	SRTM DEM is only available between -60° and 60° latitudes
Level 1 v2: 250m, v3: 300m	Albedo, NDVI/fAPAR	MODIS VNIR (250m, daily)	VIIRS VNIR (375m, daily)	MODIS is past the end of its design life (2020)
	Land surface temperature (LST)	MODIS TIR (1000m, daily)	VIIRS TIR[8] (375m, daily)	MODIS is past the end of its design life (2020). Improved spatial and temporal resolution
	Precipitation[9]	CHIRPS (5km, daily)	GEOS-5 (~28km, hourly)	CHIRPS is only available between -50° and 50° latitudes
	Solar radiation	MSG[10] (3km, 15-minute)	AgERA5[6] (~11km, daily); GEOS-5[7] (~28km, 3-hour)	MSG is only available for Africa/Europe and Indian Ocean
Level 2 100m	Albedo NDVI/fAPAR	MODIS VNIR[11] (250m, daily), PROBA-V[12] (100m, 5-day), Sentinel-2[13] (10-20m, 5-day)	Sentinel-2 (10-20m, 5-day)	PROBA-V terminated operational service in 2020. Improved temporal consistency.

	Land surface temperature (LST)	MODIS TIR (1000m, daily) resampled to 100m using bilinear interpolation	VIIRS TIR[9] (375m, daily) downscaled to 100m using Sentinel-2 (20m, 5-day)	MODIS is past the end of its design life (2020). Improved spatial and temporal resolution
Level 3 v2: 30m | v3: 20m	Albedo NDVI/fAPAR	Landsat VNIR (30m, 16-day)	Sentinel-2 (20m, 5-day); Landsat VNIR (30m, 16-day)	Improved temporal resolution and temporal coverage
	Land surface temperature (LST)	Landsat TIR (100m-120m, 16-day)	VIIRS TIR[9] (375m, daily) sharpened to 20m; Landsat TIR (100m, 16-day) for selected areas	Improved spatial and temporal resolution (for selected areas – Libya and Yemen – the historical archive of Landsat was prevalent)

[1]Windspeed is calculated as the mean square root of the sum of the eastward (u) and northward (v) components squared
[2]Vapour pressure is derived from GEOS-5 specific humidity or ERA5/AgERA5 dewpoint temperature.
[3]MERRA is input for WaPOR until 21-2-2014.
[4]GEOS-5 is input for WaPOR from 21-2-2014 onwards.
[5] ERA5 is used only for historical data production due to a latency of 5 days, and the final release is 2 months after the month in consideration.
[6]AgERA5 is used for historical data production due to a latency of 8 days and the final release is 2 months after the month in consideration.
[7]GEOS-5 is used only for Near Real-Time data production.
[8]VIIRS has not published LST data product at 375m resolution. Instead, the Brightness Temperature data product is used to calculate Land Surface Temperature.
[9]Data source that is used as precipitation in the interception model.
[10]MSG shortwave radiation product is used to derive daily transmissivity, from which solar radiation is calculated.
[11]MODIS VNIR is resampled and used for the period before 2014.
[12]PROBA-V is used for the period 2014-2020.
[13]Sentinel-2 is used for the period after 2020.

Appendix D SUPPLEMENTARY INFORMATION FOR CHAPTER 4

Supplementary figures for this chapter can be accessed online at:

https://doi.org/10.1080/02626667.2025.2600682

Appendix E SUPPLEMENTARY INFORMATION FOR CHAPTER 5

E1. Eddy Covariance flux dataset post-processing

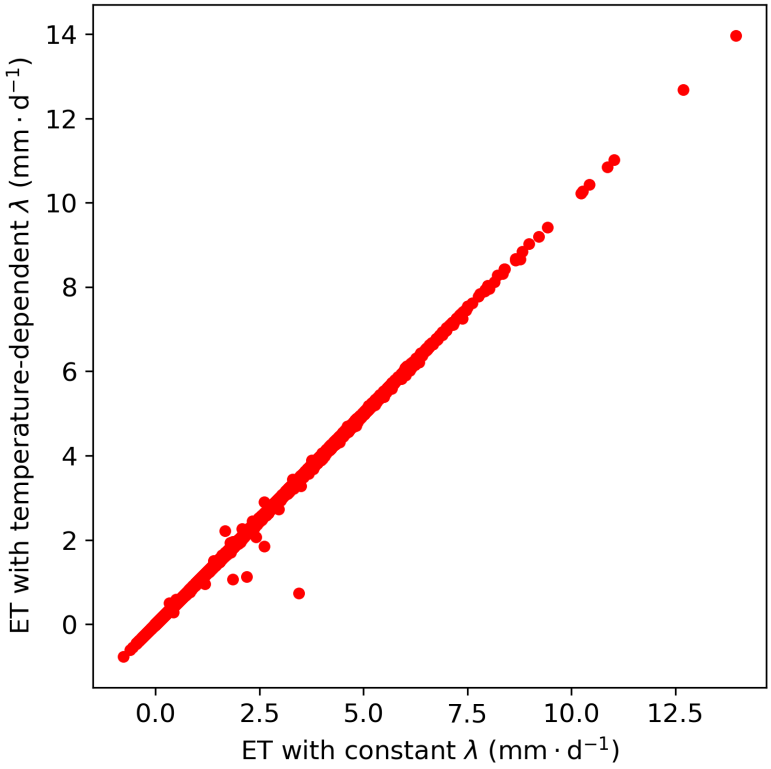


Figure E-1. Comparison of dekadal ET values calculated from latent heat flux by using constant and temperature-dependent latent heat of vaporisation at 75 flux sites.

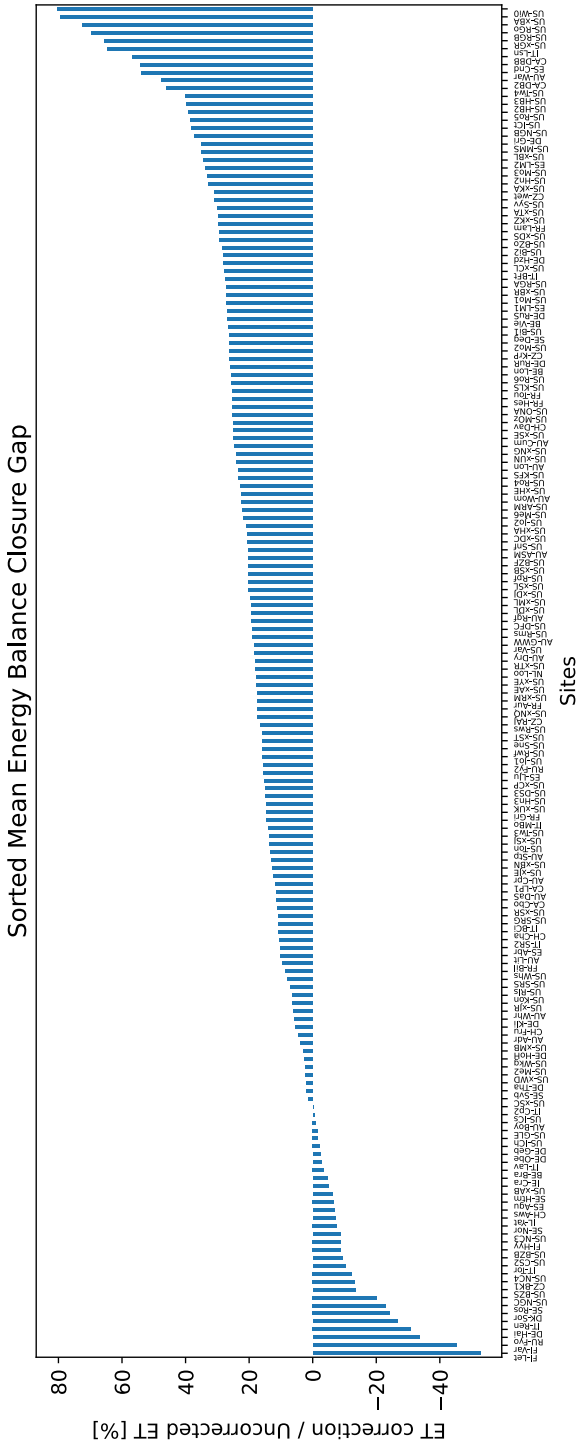


Figure E-2. The average energy balance closure gap relative to the uncorrected ET at 169 flux sites.

E2. OpenET benchmark dataset

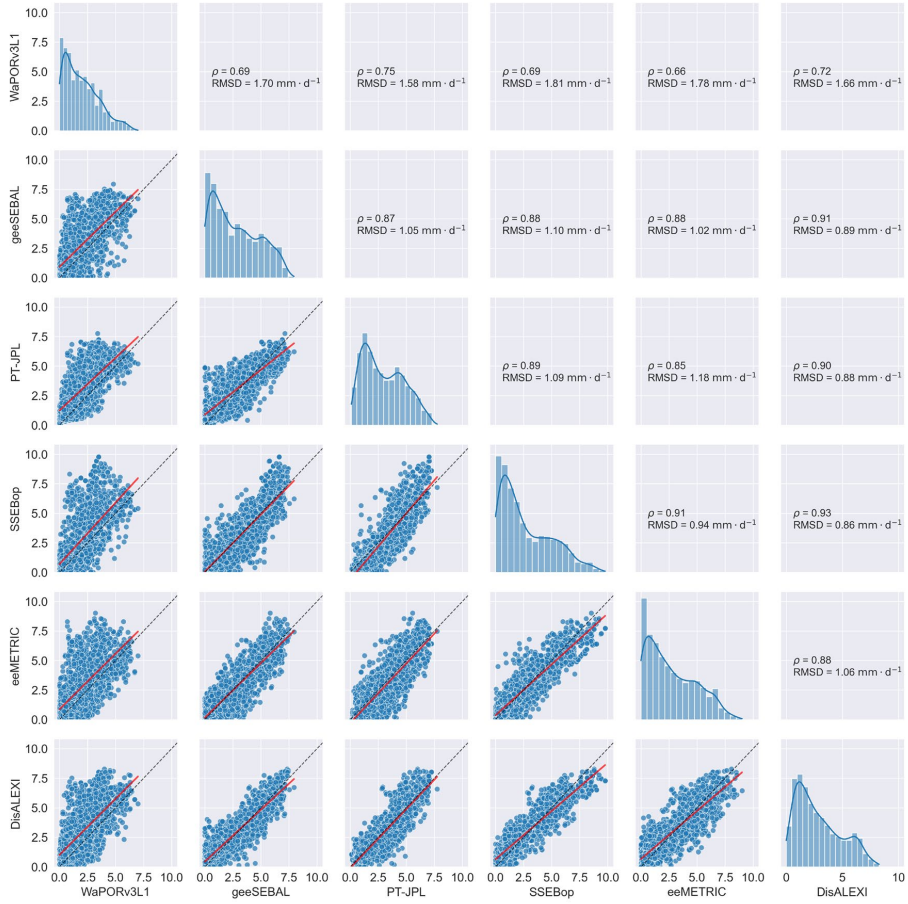


Figure E-3. Cross-correlation of dekadal ET values from the WaPORv3L1 and OpenET ensemble models at the locations of 53 EC flux sites in the contiguous United States.

E.3. In-situ validation of WaPORv3L1 dekadal ET at 169 flux sites with EBC correction of LE

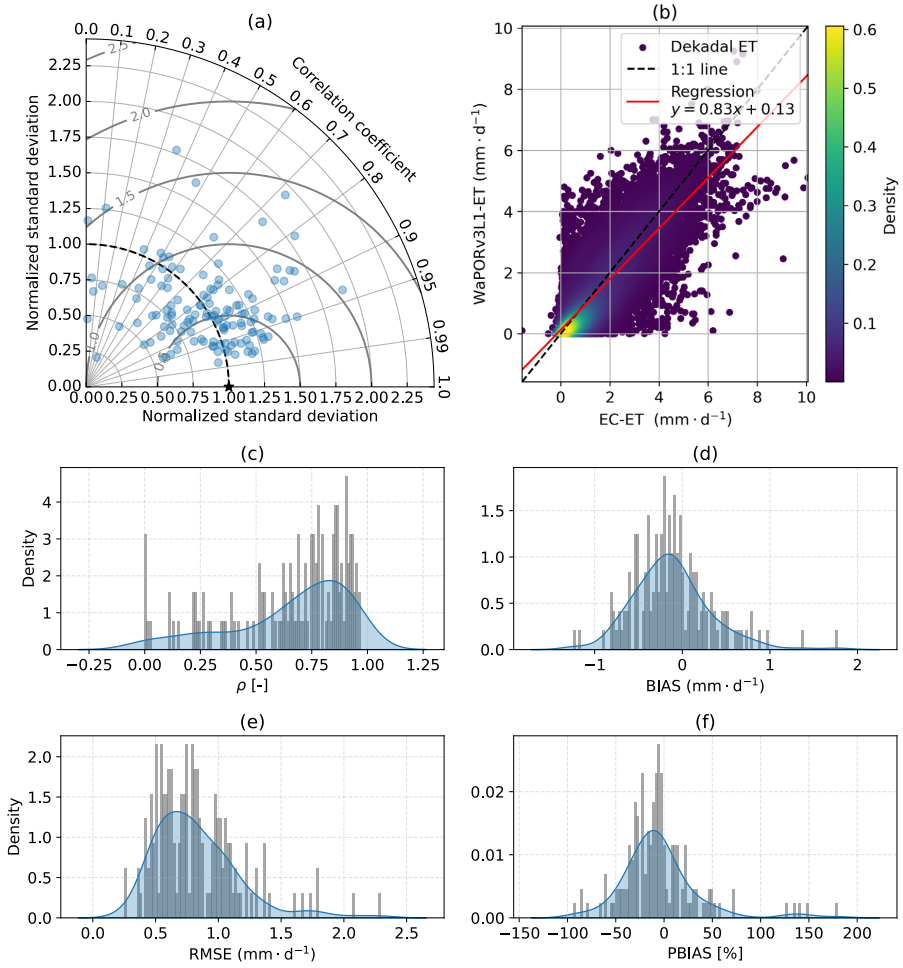


Figure E-4. Performance of WaPOR-ET v3L1 product at 169 EC flux sites with EBC correction between 2018–2022. (a) Taylor diagram, (b) regression analysis of all data points, and kernel density estimation (KDE) of the probability density function for four performance metrics: (c) coefficient of correlation (ρ), (d) bias, (e) RMSE, and (f) PBIAS.

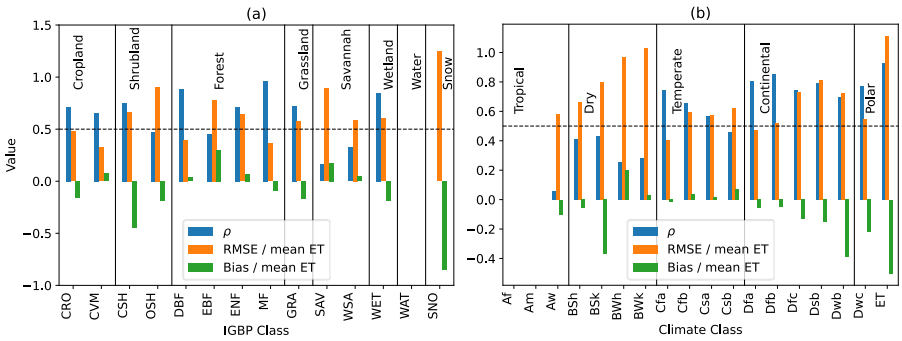


Figure E-5. Average performance metrics of WaPORv3L1 product by (a) International Geosphere-Biosphere Programme (IGBP) land cover classification and (b) Köppen–Geiger climate classification, calculated using EBC-corrected in-situ ET estimates.

E.4. In-situ validation of WaPORv2 and WaPORv3 dekadal ET at flux sites in Africa[41]

Comparison of WaPOR v2L1, v3L1 and v3L2 performance was conducted by in-situ validation against 9 flux tower stations in the South African Environmental Observation Network (SAEON) (Bieri et al., 2022) and African Monsoon Multidisciplinary Analysis - Coupling the Tropical Atmosphere and the Hydrological Cycle (AMMA-CATCH) data portal (AMMA-CATCH, 1990; Galle et al., 2018). The AMMA-CATCH stations locate in Senegal and the SAEON stations locate in South Africa (**Figure E-6**).

[41] This section is partly based on the technical report: IHE Delft, 2024. WaPOR Quality Assessment: Technical report evaluating WaPOR v3. IHE Delft Institute for Water Education, Delft, The Netherlands.

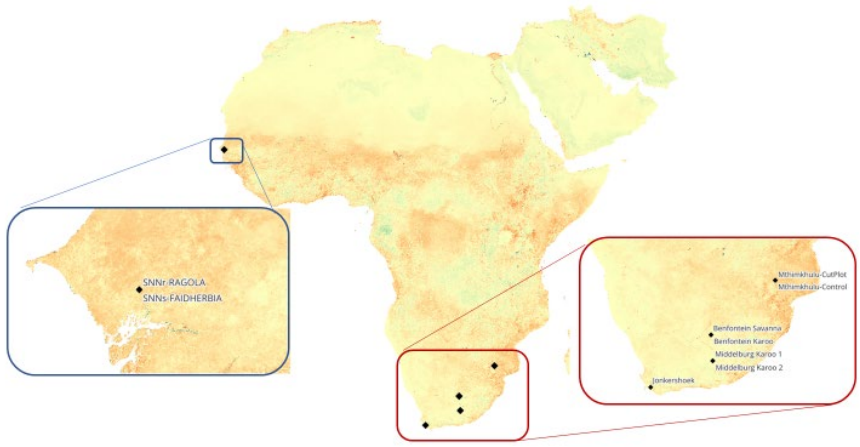


Figure E-6. Locations of the AETI stations in South Africa and Senegal. (Basemap: the difference of annual ET between WaPORv3L1 and WaPORv2L1 for 2022).

For data from Eddy Covariance towers, we filtered low-quality data based on the latent heat flux (LE) quality flag from EddyPro and EasyFlux postprocessing software (quality grade greater than 3), total number of samples per 30 minutes (less than 50 %), and energy balance closure (less than 60 % and greater than 120 %). Then, we calculated ET rate by dividing LE by latent heat of vaporization using Equation (5.7). The 30-min ET was aggregated to daily and then, averaged for every dekad. Since all the sites are homogeneous with footprint sizes range from 100 m to 2 km, we considered the average of WaPOR AETI data within the 100 m distance surrounding the flux towers.

We validated WaPOR v2L1, v3L1, and v3L2 with the dekadal average ET ($mm \cdot d^{-1}$) from flux sites. We do not observe a systematic bias or outperformance of WaPOR v3 over all the stations (**Figure E-7**, **Figure E-8**, and **Figure E-9**). Among the three products, WaPOR v3L1 performed the best over three sites (Middleburg Karoo 1 and 2, Mthimkhulu-Cutplot). WaPOR v3L2 performed the best over four sites (SNNs-FAIDHERBIA, Jonkershoek, Benfontein Karoo, Benfontein Savanna). WaPOR v2L1 performed best over two sites (SNNr-RAGOLA, Mthimkhulu-Control).

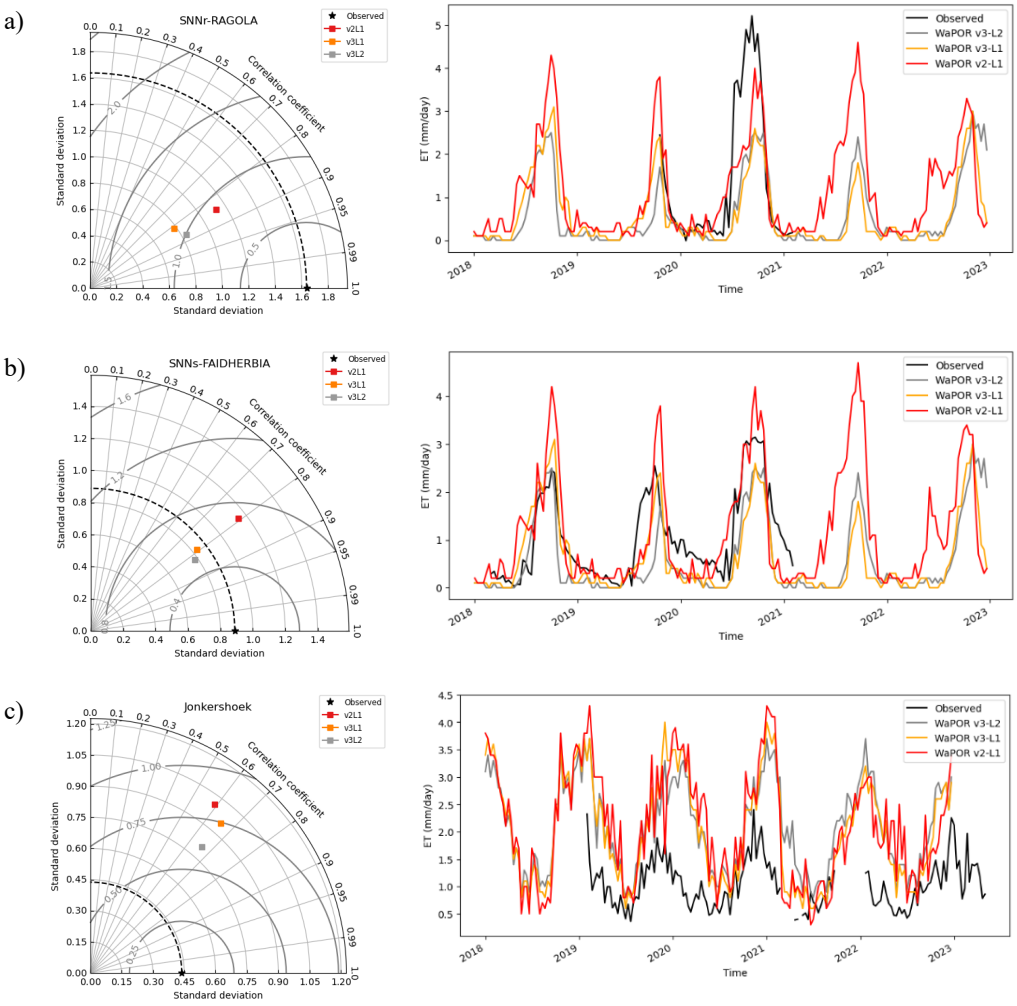


Figure E-7. Taylor diagram (left) and time-series (right) of WaPOR AETI compared to uncorrected observations at SNNr-RAGOLA (a), SNNr-FADHERBIA (b), and Jonkershoek (c). The grey contour lines on Taylor diagram represent centred RMSE.

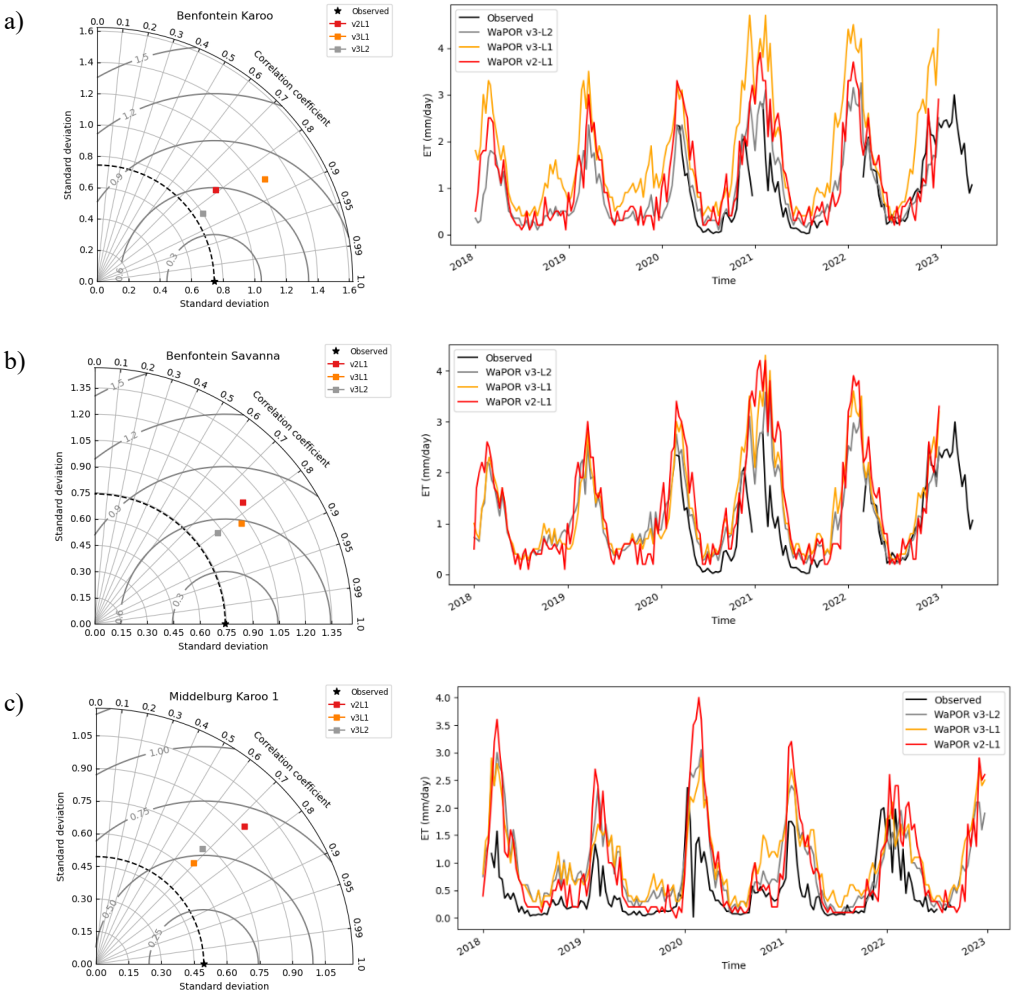


Figure E-8. Taylor diagram (left) and time-series (right) of WaPOR AETI compared to uncorrected observations at Benfontein Karoo (a), Benfontein Savanna (b), and Middelburg Karoo 1 (c). The grey contour lines on Taylor diagram represent centred RMSE.

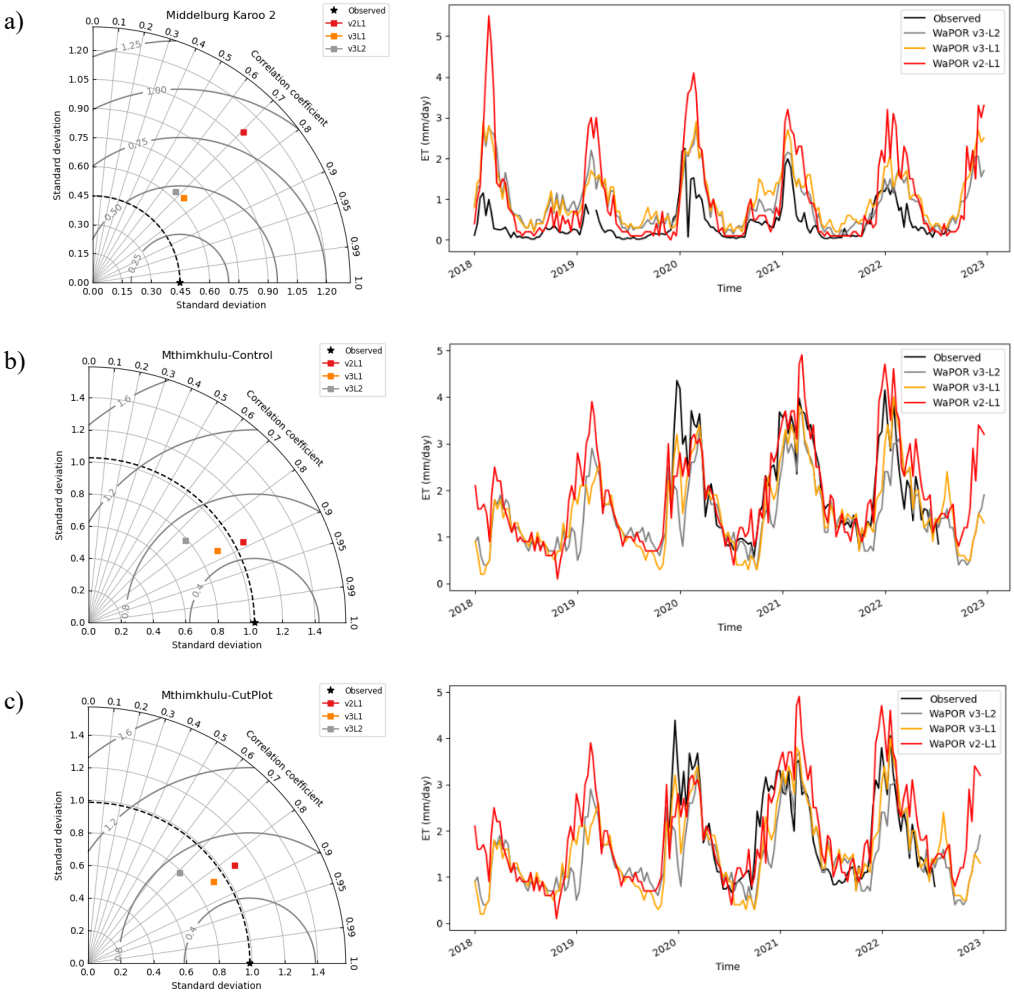


Figure E-9. Taylor diagram (left) and time-series (right) of WaPOR AETI compared to uncorrected observations at Middelburg Karoo 2 (a), Mthimkhulu-Control (b), and Mthimkhulu-Cutplot (c). The grey contour lines on Taylor diagram represent centred RMSE.

E.5. Comparison between WaPORv2 and WaPORv3 monthly and annual ET at basin scale[42]

The ET-WB database contains ET estimates from multi-source datasets (23 precipitation, 7 storage change, 29 runoff datasets) which are derived from satellite products, in-situ measurements, reanalysis, and hydrological simulations for the period 2002-2021 (Xiong et al. 2023). The ET-WB data contains information on 168 river basins globally, of which 24 are in Africa (the head waters of the Tigris and Euphrates basins are not covered by WaPOR v2 and therefore not included). These ET estimates are compared to the WaPOR AETI estimates for the same 24 river basins (**Figure E-10**). The results show that monthly WaPOR data varies less compared to the ET-WB datasets, but in most cases WaPOR v3 data falls within the 25th-75th percentile (8 out of 12 months) whereas WaPOR v2 estimates of AETI generally fall within the higher values with 6 out of 12 months falling in the 75th-95th percentile and one month (September) far exceeding the 95th percentile.

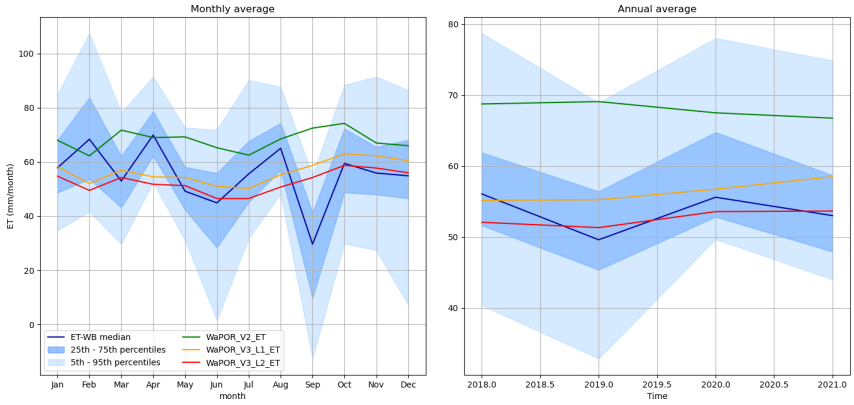


Figure E-10. Monthly average values (mm·month^{-1}) and annual timeseries (average mm·month^{-1}) of the ET-WB and WaPOR-ET over area covered by selected river basins in Africa during the period 2018–2021. The shading shows the spread range among different datasets, with the central solid line indicating the ensemble median value.

[42] This section is partly based on the technical report: IHE Delft, 2024. WaPOR Quality Assessment: Technical report evaluating WaPOR v3. IHE Delft Institute for Water Education, Delft, The Netherlands.

Appendix F SUPPLEMENTARY INFORMATION FOR CHAPTER 6

F.1. Excerpt from blog post "WaPOR Born Out of Necessity!"

"I was visited by a professor on a World Bank mission. He informed me about a training he had given to the Yemeni Ministry of Water on the use of satellite information in agriculture. I linked this to the water supply request and the findings. He wasn't surprised and expected that this wouldn't be the last time I would receive this kind of request. According to him, the shortage of water for domestic use was easy to explain. Satellite information showed that agriculture was the largest user of water. Crops produced in this country consumed three times more water than crops in similar areas and conditions. This information gave me a different perspective on how to approach water scarcity management, agriculture, and domestic water use. I raised the issue of the use of satellite information with the competent authority. They were sceptical about this. It took time, but after several expert meetings, they were convinced that the use of satellite data could have added value. In the end, the ministry even decided to develop an information database system to map water use in agriculture using remote sensing" (Kleijn, 2024).

Appendix G Declaration of use of generative Artificial Intelligence

The author acknowledges the use of generative artificial intelligence (GenAI) to identify errors in programming code and provide suggestions for improving the clarity and academic style of original texts. Any output generated by GenAI was thoroughly reviewed and verified by the author to ensure that the final code and text meet the standards of accuracy, clarity, and relevance. The specific GenAI models used were GPT-4-turbo and GPT-5-mini (developed by OpenAI, Inc.) and Gemini 2.5 Flash and Gemini 2.5 Pro (developed by Google, Inc.). Inputs provided by the author to these GenAI models were withheld from the training and improvement of these models.

LIST OF ABBREVIATIONS

1SEB	One-source Surface Energy Balance
2SEB	Two-source Surface Energy Balance
AETI	Actual evapotranspiration and Interception
ALEXI	Atmosphere-Land Exchange Inverse
BR	Bowen Ratio
BREB	Bowen Ratio Energy Balance
DisALEXI	Disaggregation of the Atmosphere-Land Exchange Inverse
EB	Energy Balance
EBC	Energy Balance Closure
ECMWF	European Centre for Medium-Range Weather Forecasts
eeMETRIC	Google Earth Engine implementation of the METRIC model
ET	Evapotranspiration
ET_a	Actual Evapotranspiration
ET_o	Reference Evapotranspiration
ET_p	Potential Evapotranspiration
ETC	Extended Triple Collocation
ER	Energy Residual
ERA5	The fifth generation ECMWF reanalysis
EC	Eddy Covariance
FAO	Food and Agriculture Organization of the United Nations
GEOS5	The Goddard Earth Observing System version 5
geeSEBAL	Surface Energy Balance Algorithm for Land using Google Earth Engine
GLEAM	Global Land Evaporation Amsterdam Model
GRACE	Gravity Recovery and Climate Experiment
IE	Irrigation Efficiency
IQR	Interquartile Range
JPL	Jet Propulsion Laboratory
LAI	Leaf Area Index

LE	Latent Heat Flux
LST	Land Surface Temperature
MAPE	Mean Absolute Percentage Error
MC	Monte Carlo
ME	Mean Error
METRIC	Mapping Evapotranspiration at high Resolution with Internalized Calibration
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NWP	Numerical Weather Prediction
PBIAS	Percent bias
PET	Potential Evapotranspiration
PM	Penman Monteith
PML	Penman-Monteith-Leuning
PT	Priestley Taylor
PT-JPL	Priestley-Taylor from Jet Propulsion Lab
RET	Reference Evapotranspiration
RMSE	Root Mean Squared Error
RS	Remote Sensing
SEB	Surface Energy Balance
SEBAL	Surface Energy Balance Algorithm for Land
SEBS	Surface Energy Balance System
SIMS	
SSEBop	The operational Simplified Surface Energy Balance
STD	Standard Deviation
TAHMO	Trans-African Hydro-Meteorological Observatory
TC	Triple Collocation
TCA	Triple Collocation Analysis
TCH	Three-Cornered Hat
TIR	Thermal Infrared
TWSA	Total Water Storage Anomaly
VI	Vegetation Index
VNIR	Visible and Near-Infrared

WaPOR	Water Productivity through Open access of Remotely sensed derived data
WB	Water Balance
WP	Water Productivity
WUE	Water Use Efficiency

LIST OF TABLES

LIST OF FIGURES

ABOUT THE AUTHOR

Bich Tran (Trần Ngọc Bích in Vietnamese) was born in August 1994 in Gia Lai, Vietnam. She went to schools in Hanoi and obtained her Bachelor of Science degree in Water - Environment - Oceanography in 2015 at the University of Science and Technology of Hanoi (also known as Vietnam-France University). During her undergraduate studies, a strong interest in hydrology and ecology led her to a research internship at the University of Tours, France, where she focused on the physico-chemical properties of restored vernal pools in the Chinon forest. Following her undergraduate degree, she actively volunteered in several educational and environmental projects in Vietnam. These projects fostered her interest in agriculture and its relationship with the hydrosphere and biosphere. In 2016, she moved to Leuven, Belgium to pursue a Master of Science degree in Water Resources Engineering at Katholieke Universiteit Leuven and Vrije Universiteit Brussel, with the support of the VLIROUS scholarship. She obtained her master's degree in 2018, with a thesis on the application of remote sensing for mapping inundation in the peatlands of Biebrza, Poland.

In 2019, she joined IHE Delft Institute for Water Education in the Netherlands as a project assistant. Through the project "Water Productivity through Open access of Remotely sensed derived data" (WaPOR), she expanded her expertise in water accounting and agricultural water management, alongside developing technical skills in applied remote sensing. From 2021 to 2025, she conducted her PhD research at IHE Delft and TU Delft, which focuses on assessing uncertainties of satellite-based estimation of evapotranspiration. With a passion for higher education, she has mentored several master students, lectured in the courses "Water resources assessment using remote sensing" and "Remote sensing for agricultural water management", and developed several online open courses. As she concludes her PhD research, she starts working as a researcher at Wageningen University and Research.

Photo credit: Regina Hügli

PUBLICATIONS

Journal articles

Tran, B. N., van der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G., and Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. Hydrol. Earth Syst. Sci., 27, 4505–4528, https://doi.org/10.5194/hess-27-4505-2023

Tran, B. N., Dehati, S., Seyoum, S., van der Kwast, J., Jewitt, G., Uijlenhoet, R., and Mul, M., 2025. Evaluating reanalysis datasets as meteorological input for estimating reference evapotranspiration in Africa and Southwest Asia. Hydrological Sciences Journal. https://doi.org/10.1080/02626667.2025.2600682

Yalew, S.G., van der Zaag, P., Tran, B.N., Michailovsky, C.I.B., Salvadore, E., Borgomeo, E., Karimi, P., Pareeth, S., Seyoum, S.D. and Mul, M.L., 2023. Open-access remote sensing data for cooperation in transboundary water management. Water International, pp.1-20. - https://doi.org/10.1080/02508060.2023.2263226

Alba, R., ter Horst, R., Tran, B.N., Klein, A., Unverzagt, K., Godinez-Madrigal, J., Verzijl, A., Rusca, M., Vos, J., Venot, J.P. and Zwarteveen, M., 2025. Situating Hydrological Modeling: A Proposal for Engaging with the Power of Models. Wiley Interdisciplinary Reviews: Water, 12(4), p.e70030. https://doi.org/10.1002/wat2.70030

Dehati, S., Tran, B. N., Karimi, P., & Mul, M. (2025). Comparison and validation of spatial reference evapotranspiration datasets over Africa. Hydrological Sciences Journal. https://doi.org/10.1080/02626667.2025.2600684

Castelli, G., Howard, B., Buytaert, W., Ceperley, N., and The Co-Creating Water Knowledge (CCWK) Working Group., 2025. Co-creating Water Knowledge: From Local Community Engagement to Global Solutions. Hydrological Sciences Journal. https://doi.org/10.1080/02626667.2025.2571065

Articles in preparation

Tran, B. N. and Mul, M.L., *forthcoming*. Blurry eyes from space: A reflexive analysis of uncertainty research on WaPOR's satellite-based model and data. (submitted to Water Alternatives).

Tran, B. N., Seyoum, S., van der Kwast, J., Jewitt, G., Uijlenhoet, R., and Mul, M.L. Assessing uncertainty in WaPOR global evapotranspiration data: insights from in-situ comparison and triple collocation.

Tran, B. N., Hakzi, K., and Mul, M.L. A meta-analysis of relative bias in remote-sensing-based evapotranspiration estimates.

Conference proceedings

Tran, B.N., Van Der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G. and Mul, M., 2022. Uncertainty in Satellite Remote Sensing Derived Evapotranspiration Estimation: Current Status and Assessment Methods. 39th IAHR World Congress from Snow to Sea. https://doi.org/10.3850/IAHR-39WC2521711920221782

Tran, B.N., Mul, M., Seyoum, S., and Wymenga, E., 2022. Monitoring Wetlands Dynamics in the Inner Niger Delta using Open-Access Remotely Sensed Evapotranspiration Data. 39th IAHR World Congress from Snow to Sea. https://doi.org/10.3850/IAHR-39WC2521711920221154

Cover art: Bich N. Tran

Based on the data of Figure 3-3, Figure 5-7, and the PyWaPOR-ETLook model diagram (by Bert Coerver, https://www.fao.org/aquastat/py-wapor/_etlook.html)

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Ngoc Bich Tran

born on 03 August 1994 in Gia Lai, Vietnam

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Delft, 16 April 2026

SENSE coordinator PhD education

Dr Ir Peter Vermeulen

The SENSE Director

Dr Jampel Dell'Angelo

The SENSE Research School declares that **Ngoc Bich Tran** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a workload of 55.5 EC, including the following activities:

SENSE PhD Courses

- A1: Environmental Research in Context (2021)
- A2 Research in context activity: 'Developing open online course on "Concept and Validation of WaPOR"' (2023)

Selection of Other PhD and Advanced MSc courses

- Geostatistics, PE&RC and WIMEK (2021)
- Uncertainty Propagation in Spatial Environmental Modelling, PE&RC and WIMEK (2022)
- Flux Measurement Fundamentals, Karlsruhe Institute of Technology (KIT), Institute for Meteorology and Climate Research - Atmospheric Environmental Research (2022)
- Situated Hydrological Modelling, IRI Thesys (2023)
- Observation Theory: Estimating the Unknown, TU Delft (2021)
- Decolonising Science, IHE Delft (2024)
- Code refinery workshop, TU Delft (2022)
- Managing the Academic Publication Review Process, TU Delft (2022)
- Writing Successful Project Proposals Workshop, IHE Delft (2022)
- Sharing your Research and Work as Simple as a TEDx Talk, TU Delft (2022)
- Cross Cultural Communication Skills in Academia, TU Delft (2022)

Management and Organisational activities

- Member of PhD Association Board of IHE Delft, (2022)
- Chair of PhD Association Board of IHE Delft, (2023)

Teaching and supervision

- Supervising six MSc students (2023-2024).
- Lecturing in MSc course 'Water Resources Assessment using remote sensing data' (2024)
- Lecturing in MSc course 'Remote sensing for agricultural water management' (2025)

Selection of Oral Presentations

- *Towards improving SDG 6.4 Indicators: Assessment of uncertainties in remotely sensed evapotranspiration.* 16th IHE PhD Symposium – Collaborative water resource management for a water secure world, 13/10 -14/10 2022, Delft, Netherlands
- *Rapid Water Accounting Plus Using WaPOR Database*. 1st Water Accounting Plus symposium, 12/7 -13/7 2022, Delft, Netherlands
- *Use of Remote Sensing in estimating ET from agricultural settings.* Symposium: Meeting on the FAO56 method for determining crop reference evapotranspiration and need for revision under arid conditions, 15/6 2023, Delft, Netherlands

IHE DELFT
Institute for Water Education
under the auspices of UNESCO

TU Delft Delft University of Technology

Evapotranspiration (ET) is a major water flux in the terrestrial water balance and a key link between water and surface energy balances. In water sciences and management, the quantification of ET is required but challenging to gauge in situ, leading to the popularity of models based on satellite-derived data. However, uncertainties in satellite-based estimation arise from both methodological and technical factors. This study examines and assesses uncertainties in satellite-based estimation of ET. Part I provides a systematic quantitative literature review, showing the diversity of approaches and constraints arising from the availability and quality of reference data. A meta-analysis of in-situ validations against eddy covariance measurements quantifies the status of uncertainty in terms of reported performance metrics. Part II focuses on the assessment of a satellite-based ET data product for monitoring water productivity from field to global scales. Technical uncertainties through ex-ante and ex-post methods, including error propagation, in-situ validation, and triple collocation analysis are provided. The results highlight spatial variability in uncertainty, limitations of validation data, and challenges in dry and tropical regions providing guidance to users of such products. Finally, this thesis reflects on methodological uncertainties arising from problem framings, model choices, and configurations.

This book is printed on paper from sustainably managed forests and controlled sources

9 789073 445802