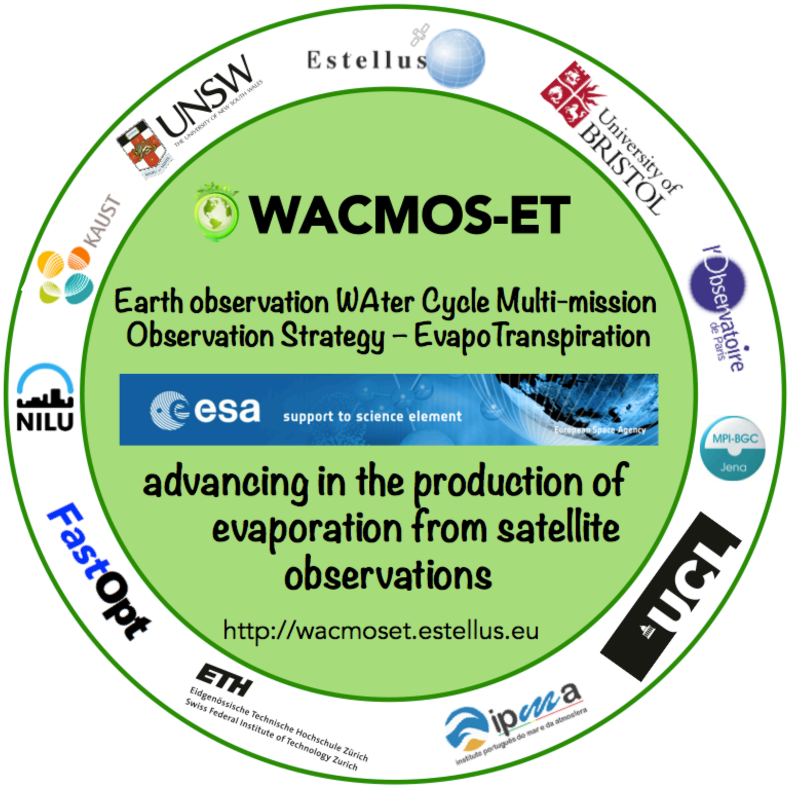

WP3300

Roadmap

Mathew McCabe, Ali Ershadi,

KAUST, Saudi Arabia

Diego Miralles,

Ghent University, Belgium

Carlos Jimenez,

Estellus, France

Martin Jung

MPI-Biogeochemistry, Germany

Dominik Michel

ETHZ, Switzerland

Contents

1 Introduction

WACMOS-ET project is an ESA funded effort to advance the development of evaporation estimates at global and regional scales, having as its main objectives the production of a Reference Input Data Set (RIDS) to derive and validate evaporation estimates and to perform a cross-comparison, error characterization, and validation exercise of a group of selected flux retrieval algorithms driven by the RIDS. The project contributes directly to the GEWEX Global Data and Assessment Panel (GDAP) LandFlux initiative, which is actively engaged in producing a data set of global land surface turbulent fluxes that can be integrated with the existing suite of GEWEX products to allow an observation-based characterization of the water and energy cycles. Here we outline a Roadmap to advance the production and development of evaporation at global and regional scales based on the project findings and past experience from the consortium members in developing such products. The reader is directed to additional supporting documents available form the project website (http://wacmoset.estellus.eu) to review the main project activities related to consolidating the requirements for evaporation estimation, rationale for the evaporation algorithm selection, compiling a Reference Input Data Set to drive the evaporation, design of the project products and the evaluation of the produced flux estimates across different spatial and temporal scales.

2 An Assessment of Global Evaporation Algorithms

Four process-based evaporation models have been evaluated in this project. The models selected for assessment include: the Surface Energy Balance System (SEBS; Su 2002), the Priestley-Taylor Jet Propulsion Laboratory (PT-JPL) model (Fisher et al. 2008), the Penman-Monteith Mu (PM-Mu) model (Mu et al. 2011) and the Global Land surface Evaporation: the Amsterdam Method (GLEAM) (Miralles et al. 2011). We focus here on an overview of their relative performances.

2.1 Overview of model performance and selection

Many of the studied models have been evaluated previously using high-resolution meteorological forcing data and flux tower observations and have proven accurate in simulating evaporative response. However, evaluating these models at the resolutions required of the WACMOS-ET project is fraught with some particular challenges: foremost amongst which is the spatial and temporal scaling issue. Whether it is appropriate to use 3-hourly averaged forcing data to run models and to evaluate simulations, knowing that they will inevitably miss key diurnal drivers of the meteorological variability, is an open question. Likewise, evaluating algorithms against 3-hourly averaged flux retrievals is also a compromise, as the capacity to accurately capture diurnal variability is diminished and instead larger temporal scale influences are being monitored.

Given this, it is perhaps appropriate to assess model performance against a range of space and time scales: including both the instantaneous retrievals as well as temporally averaged responses. To do this, the project has examined tower-scale dynamics, large-scale inter-model agreement and basin-scale water-budgets. One of the significant challenges in these evaluation studies lies in simply comparing the coarse-scale retrievals against tower-based data. While it is the standard approach, it is clearly not ideal given the disparity in footprint and related scale issues, so alternatives for assessing model performance have also been used.

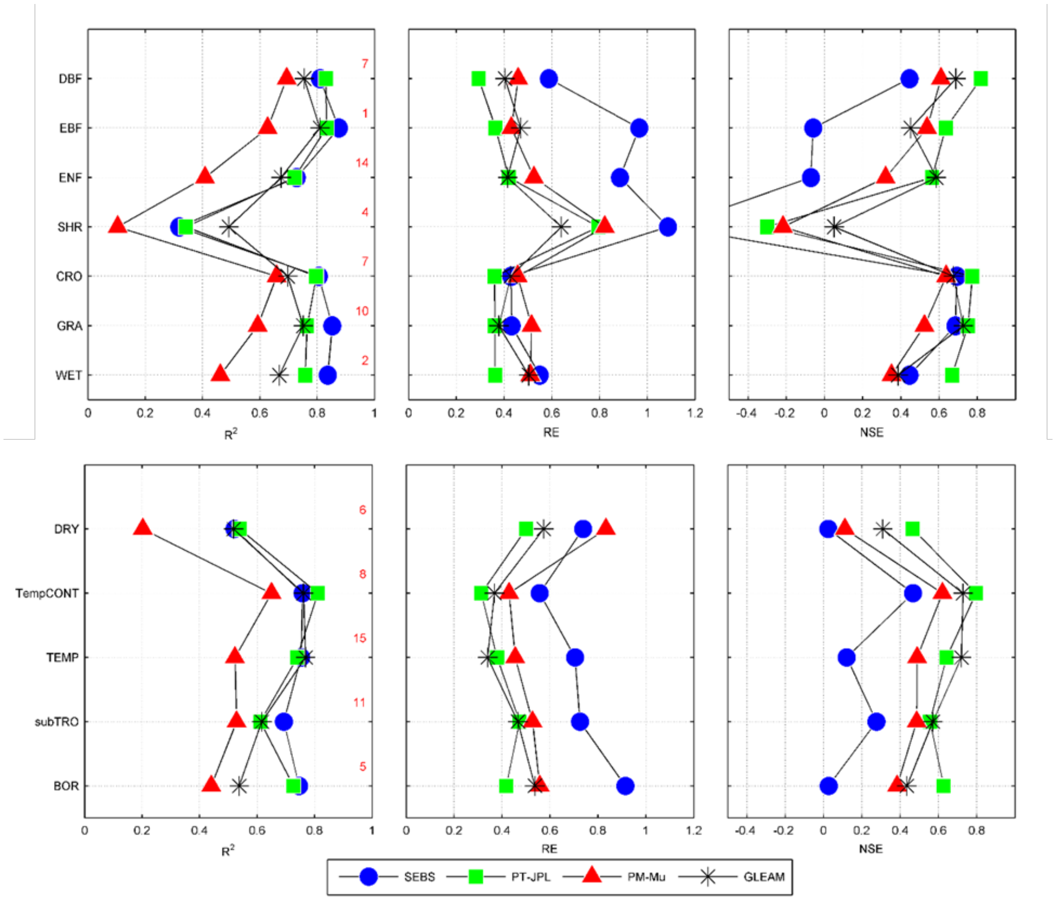

Figure 1: Coefficient of determination (R^2), relative error (RE) and Nash-Sutcliffe Efficiency (NSE) for models across biomes (top) and climate zones (bottom). Each point is for the collection of records of the towers located at the selected biome or climate zone, with the number of towers shown in R^2 plot by red font.

Regardless of these constraints, to summarize: given the limitations of available forcing, sensitivity to data inputs, capacity to perform over a range of land surface and climate types, it seems clear that the Priestley-Taylor (PT) based models (PT-JPL and GLEAM) perform overall better than do the Penman-Monteith (PM-Mu) or surface energy balance approaches (SEBS). However, this is not to say that those PT-type models provide the only way forward for global flux estimates. What has been made clear in this analysis is that different models serve different purposes, and that the most appropriate way forward in developing a robust global product may lie in providing an ensemble of approaches, with model weighting based on validation analyses and uncertainty assessments.

Table 1: Summary of strength and limitation of studied evaporation models

Model	Strength	Limitations
PT-JPL	- Good overall performance in evaporation estimation - No need for soil moisture, wind speed, land surface temperature, and vegetation height	- Estimation of *optimal temperature* requires at least one year of data - Un-validated interception loss estimates
GLEAM	- Good overall performance in evaporation estimation - Consideration of soil moisture and sound calculation of interception - Applicable at large scales with available remote sensing data only	- Requirement of accurate rainfall - Limited functionality for spatial resolutions less than 0.25°, due to reliance on microwave data - Need for continuum gap-free input data
SEBS	- No need for parameterization of resistances - Sound physical representation of heat transfer mechanisms from land to atmosphere, with independent calculation of sensible heat flux	- Overestimation of evaporation - Sensitivity to errors in the vertical temperature gradient - Limited availability of wind speed and canopy height data from remote sensing retrievals - Limited performance over tall canopies due to the presence of the roughness sub-layer - Only for clear-sky conditions
PM-Mu	- No need for wind speed, soil moisture and rainfall data - Sound methodology accounting for the main drivers of evaporation	- Underestimation of evaporation - Requirement of calibration of biome-specific LUTs parameters based on eddy-covariance data

The variability observed in the performance of the models across the tower sites demanded further evaluation across biomes and climate zones to understand if model skill is related to the type of vegetation covering the land surface. As an example, Figure

1 clearly shows the poor performance over shrubland sites, with all models associated with low values of the Nash-Sutcliffe Efficiency (*NSE*). Although PT-JPL and GLEAM performed well over the majority of biomes, the models still exhibit variability in performances across biome types and climate regimes. **The finding that no single model performs well everywhere** highlights the need for considering a biome- or climate-specific composite of flux algorithms for global application.

2.2 Strength and limitations of the candidate models

In evaluating the different WACMOS-ET models against a range of flux tower observations, it was noted that each of the models have their own particular strengths and weaknesses. While some of these are unique to the particular model, others may be common amongst the different approaches. Understanding these positive and negative attributes, particularly in the light of global scale application, is useful in allowing for a reasoned model selection, based on availability of forcing data, robustness of the retrieval approach, appropriateness of the underlying model physics and any other considerations relevant to incorporating these algorithms into an operational global flux context. Table 1 highlights some of the model strengths and weakness that are relevant to this project.

3 On the Development of Evaporation Methodologies

While considerable developments and progress have been made in the application of algorithms for the retrieval of land surface fluxes from space (Kalma et al. 2008; Wang and Dickinson 2012), there are many outstanding challenges and issues that remain to be addressed. Although not a comprehensive accounting of these challenges, the issues can be grouped broadly into two categories: 1) model physics and structural limitations (including access to the data required to force the models); and 2) model assessment and interpretation of simulations.

In the first category, most retrieval techniques used in global flux estimation (and certainly those that performed well within the WACMOS-ET project) are based on the Penman-Monteith or a related form such as Priestley-Taylor: due in large part to the relative simplicity of the techniques and the low forcing data requirements. As will be discussed here, the availability of forcing data at a global scale is a key requisite for broad scale application of any model approach. While other more sophisticated approaches exist for flux estimation, it has not been shown that these are any more accurate or robust in providing global retrievals. Certainly there is a necessary compromise between simplicity and inclusiveness that is required for global application, which means that some processes will either not be described, or be described just poorly. The impacts of such simplifications have been observed throughout this project: particularly in the estimation of fluxes over more complex or heterogeneous environments, or over biomes that represent challenges to even the most sophisticated of modeling approaches (i.e. wetlands and marshes, forests, arid lands and snow covered surfaces). The realization that **no single model construct was capable of consistently outperforming any other** was an important outcome of this work and directs future efforts towards development of ensemble averaging or biome-specific model application.

Apart from improving the model physics and parameterization schemes of retrieval algorithms (or employing approaches different to those examined here: see below), the second category highlights the importance of developing robust and appropriate evaluation metrics for assessing model performance. The question on **how best to**

evaluate model retrievals was a recurring concern throughout the project. While the project performed a comprehensive and multi-scale assessment of flux response, there remain key limitations in the techniques we use to evaluate model responses. Foremost amongst these is the scale mismatch between available tower-based observations and the resolution of satellite-based retrievals, which are generally separated by two or more orders of magnitude in terms of spatial footprint. While eddy-covariance approaches offer the gold-standard for flux validation at the field scale, their use for large scale evaluation remains unclear. Perhaps a new hybrid observation-modeling approach is required, along the lines of the decision tree upscaling approach of Jung et al. (2009) who employ flux tower data in combination with distributed observations to enable large scale flux estimates. Indeed, this approach may not be just a better way to evaluate retrievals, but might also be a better way to produce them – removing the reliance on traditional flux retrieval algorithms completely. Other alternatives to be considered are to map the relationship between drivers and model outputs with connectionist approaches, and using that mapping to investigate the consistency between model outputs and satellite drivers in order to identify potential deficiencies in the modeling approach for specific regions and periods (e.g., Jimenez et al. (2009), Lipton et al. (2014))

These and other issues form the basis for discussions in the following paragraphs. The paragraphs provide a broad scope on possible ways forward to improve our capacity at global scale monitoring.

3.1 Global sensitivity analysis and uncertainty quantification

Further development of the candidate models requires an identification of the most important data and sensitive parameters to focus model modification around these factors. Due to the non-linearity of the models and high level of interactions in model forcing and parameters, application of Global Sensitivity Analysis (GSA) techniques (Sobol 2001; Saltelli et al. 2008) is likely to prove useful across the range of evaporation models. GSA methods aim to quantitatively identify and rank sensitive variable and parameters (i.e. factors) by simultaneously varying them within a defined range. GSA is hence an appropriate technique for models of this study, since all have multiple factors with high-level of interactions.

Another important issue relates to uncertainty quantification (UQ) in data forcing, model parameters, model structure, and response variables of the models. One of a family of methods for resolving uncertainties arising from these components is the Bayesian Total Error Analysis (BATEA) technique (Kavetski et al. 2006), which was recently used in a study for the SEBS model (Ershadi et al. 2013). However, even a simple perturbed forcing experiment would aid in providing improved insight into model behaviour and response. However, few model developers let alone users, seek to undertake such exercises: they are time consuming, technically challenging and are likely to identify limitations and problems within the models being tested – something both the user and model developer are often happy to avoid. In addition, model sensitivity studies, while recognized as being critically important, are often difficult to fund as there is generally little novelty or innovation in applying an existing method to an existing model. Regardless, further development and application of appropriate GSA and UQ methods are critical for enhancing our understanding of global evaporation models. Indeed, it could be contended that no meaningful progress can be made in model development without invoking UQ or model sensitivity studies.

3.2 Ensemble methods for evaporation estimation

A recurring outcome of the analyses performed within the WACMOS-ET project and elsewhere in more recent model intercomparison studies is the observation that no model consistently performs better than any other model. While this may not immediately seem to be a revelation, it is a significant outcome, as it requires a dramatic rethink in the way that global model simulations are currently performed. The motivation behind most evaluation studies is to identify the model that best performs over the widest range of target scenarios. However, this rationale is really only appropriate where models are to be applied at local or other defined locations: certainly not for global scale analysis, where the diversity and variety of land cover types, climate zones and dominant meteorological conditions are beyond the current capacity of any single modeling approach to reproduce. Ultimately, the expectation on any single modeling scheme to provide a robust global estimate is misplaced. One of the key outcomes of this work is the need to embrace an alternative modeling framework: one that is based on considering the strengths (and identifying the weaknesses) of available schemes and exploiting these to enable development of a more accurate global product.

Efforts such as the WACMOS-ET project have provided an ideal test-bed with which to begin exploration of this issue in detail and the findings have been unambiguous. The observed variability of model skill for the studied biomes and climate zones demands that caution is used in applying any single model to any large-scale application in isolation. A possible path forward for the regional to global scale applications of interest here, would be to assemble a mosaicked product based on the predictive skill of the model(s) per biomes and climate zones. In its simplest manifestation, this would entail identification (through traditional evaluation studies) of the best performing scheme, based on a range of performance metrics, for specific biome or land-cover types. The selected model would be optimised for those conditions and then combined with other "best-performing" models for the range of landscapes being considered. This approach has its appeal, as the results from the WACMOS-ET evaluations are already able to guide this model selection to some extent, although further assessment would be recommended to strengthen the selection case. Another solution would be to develop an ensemble product using a suitable multi-model blending technique: for example, a Bayesian Model Averaging technique. Here, competing approaches could be weighted based on their performance against the evaluation set, or even combined as a simple model average, as was done in Ershadi et al (2014). In this and many other cases, application of simple ensemble average has been shown to outperform any single modeling response. However, with the capacity to evaluation models against a range of tower observations, a weighted averaging approach would seem a more robust account of model uncertainties, especially for global application where some models clearly outperform others.

In developing an ensemble technique, there is certainly no limit to the number of modeling systems that could be employed. While the four approaches examined here would provide sufficient depth to allow a robust global product, further addition to this stable of model structures is easily achieved. Indeed, identifying schemes that are based on a range of physical assumptions and methodologies is likely the best way to ensure more accurate spatially distributed reproduction of surface fluxes i.e. averaging four different Penman-Monteith models is probably less useful than averaging three models with different physical basis. Inclusion of other estimation approaches such as i) the MPI tower-based upscaling technique, ii) regionally limited modeling schemes such as ALEXI-disALEXI and those developed from geostationary satellite systems, and iii) even limited area numerical weather prediction output, could all be readily incorporated. Such an ensemble approach (or even the simple tiling scheme mentioned above) is expected to enhance flux estimation in those areas where other modeling systems exist, without affecting those areas only covered by the global simulations developed as part of WACMOS-ET or LandFlux. Indeed, a framework that maximises the inclusion of alternative schemes where available, is likely to outperform one that is based on the limited range of responses available from current global platforms. Further examination and ultimately implementation of such approaches in the context of the evaluation studies undertaken here is warranted.

3.3 The utility of high-resolution flux retrievals

In our efforts to generate and evaluate global evaporation products, there is a significant scale that is often overlooked, but which can have tremendous impact on model estimates of evaporation and surface energy balance: namely, the scale of variability in land use and land cover. Although beyond the scope of the WACMOS-ET project, there is clear interest from both the agricultural and hydrological community on high-resolution retrievals (defined here as sub-kilometer to tens-meter scale). The 0.25 degree retrievals developed here are primarily directed towards the climate modeling and large scale hydrological communities: yet there is a considerable and perhaps more important applied aspect, to delivering high resolution temporal and spatial estimates.

In contrast to the assumed scale invariance of the model schemes employed herein, very few provide a means to explicitly bridge the gap between continental to field scale estimates (Kalma et al. 2008). One such approach is the ALEXI/disALEXI technique of Anderson et al. (2008). The model has been employed successfully within the US and has seen a gradual use in other regions. The capacity to shift between coarse scale (around 3 km) and fine-scale (60 m) in a flux-consistent manner has considerable appeal in meeting the information needs of diverse communities. Developments in the temporal scaling of high-resolution data has also broadened the applications base. In terms of future E-O missions, this and related schemes would certainly benefit from new fine-resolution thermal infrared missions: an area that has a recognized gap in space mission planning.

3.4 Approaches to better partition evaporation

While the project focus was on the determination of evaporation, in many regions the most dominant contribution to this flux is water lost via plant transpiration. In dryland environments, which make up close to 40% of the terrestrial surface, transpiration can account for greater than 80% of the evaporative response. Determining the changes and variability of this term, particularly in the light of increases in anthropogenic CO_2 and biomass changes, has become a topic of considerable recent interest. Currently, there are relatively few approaches for the direct partitioning of evaporation into its soil and plant contributions and certainly no direct techniques at the scale of interest explored here. Indeed, approaches for determining the transpiration flux are almost exclusively undertaken at the leaf to field scale, with specific instruments (often chamber-based) designed for this task. Recent research advocating the potential of stable water isotopes as tracers of plant transpiration have shown some promise, but are yet to realize their potential as operational tools. Likewise, satellite based stable water isotope missions provide global coverage but have yet to provide a mechanism for understanding evaporation partitioning directly.

Techniques to indirectly monitor plant biophysical responses to stress through their spectral response to environmental factors offer some promise, with many techniques developed to relate vegetation indices to stomatal function: a more direct indicator of transpiration (and photosynthesis). Developments from both ground based and remote sensing techniques to monitor chlorophyll fluorescence offer significant promise in this direction, a topic that is discussed in more detail in the following section. However, even with observations of plant stress functions, these need to be formalized and integrated into a modeling system that can relate the observed plant response to subsequent changes in water use. Currently, there are no robust frameworks that readily adapt observations into a coupled modeling framework, so focused effort is required to take advantage of these observational advances.

Discriminating transpiration from evaporation is just one puzzle in the partitioning challenge. One of the key areas of uncertainty in flux estimation is the specification of canopy interception, with the candidate models explored in WACMOS-ET showing considerable variability in the ways in which interception is accounted for, and also in the proportion of total evaporation it represents. While there is a long history of field based collections of interception and empirical approaches to infer this, there is little guidance on how to robustly parameterize this process in large scale modeling exercises. Even at the field scale, representing the process is challenging, with plant physical characteristics, meteorological conditions and precipitation type and rate all impacting its behavior.

In order to determine a more precise understanding of local-to-global patterns and distributions of flux water use, these partitioning challenges will need to be overcome. Since the bio-physical drivers, rates and isotopic composition of transpiration, soil evaporation and interception are different, determining the magnitude of these sources accurately and independently is critical to correctly model the total evaporative flux. Therefore, being able to validate the accuracy of flux modeling approaches in properly representing these separate processes appears crucial to progress towards more accurate evaporation estimates.

3.5 Upscaling tower data

Previous efforts towards evaporation observation-driven products integrated the globally distributed measurements from FLUXNET observations with remote sensing and climate data in a machine learning regression approach (e.g., Jung et al. (2009)). These products were well received by the scientific community, largely because of its independence of prescribed model formulation and observational basis. Nevertheless, any empirical approach is inherently limited by the quantity, quality, and representativeness of observations. Incorporating process-understanding into a hybrid machine-learning-process-based approach has great promise to enhance the accuracy of observation based ET products in conditions that are not well constrained by the measurements. In addition, there are more sources of evaporation observations with complementary information, such as lysimeters, mean annual evaporation derived from observed precipitation and river discharge, and multiple local studies on evaporation partitioning into transpiration, soil evaporation, and interception evaporation. There has been no effort yet to formally integrate all these complementary data streams into evaporation algorithms (e.g. via data assimilation). Such an integration is challenging and requires thorough understanding and quantification of observational uncertainties as well as accounting for the heterogeneous nature of these data streams. Nevertheless, it is possible and provides a promising pathway for evaporation products that are strongly informed by both theory and in-situ ET observations.

3.6 Limitations of the eddy-covariance flux observations

While the flux observations available from distributed global networks such as Fluxnet provide the gold-standard in model evaluation, there are well recognised limitations in the utility of such data for this purpose. These include:

- **Non-closure issues and sensor limitations:** non-closure in energy balance terms is a well-known problem in the application of eddy-covariance tower data, especially at short time scales. The key reason behind the non-closure issue remains unexplained in the literature, but have been attributed to errors in observations, difference in footprint of the sensors, advection, low frequency eddies and stationary secondary circulations (Mauder and Foken 2006; Mahrt 2010). Traditional methods for correcting non-closure include energy residual (Twine et al. 2000) and Bowen ratio (Sumner and Jacobs 2005) techniques, that both have been used in model evaluation studies (Ershadi et al. 2014). However, application of each of the techniques is a compromise to the accuracy of evaluations, as it would result to increased correlations between the modelled and observed evaporation, if the observed available energy is used as input to the models. Another issue is the need for filtering flux observations at rain events, as both the sonic anemometer and gas analyser sensors have limited functionality when relative humidity is high. Such filtering exercise prevents evaluating the quality of rainfall interception estimates.
- **Statistical representativeness of the observed evaporation:** an inevitable limitation that reduced the statistical validity of the evaluations is inconsistency in the period of

tower data, as the length of data records is not uniform across seasons. Moreover, the towers are not uniformly distributed across the studied biomes and climate zones.

- **Footprint of the observed evaporation:** the spatial footprint of the eddy-covariance tower sensors are often much smaller (e.g. hundreds of meters) than the resolution of gridded data from remote sensing and reanalysis platforms. In particular, if coarse resolution evaporation products are of interest, the estimated grid-scale evaporation does not match with local flux observations and would likely result in reduced confidence on skill ranking of the models – even though the model simulations themselves may be perfectly representative of the scale of interest. Obviously the challenge here is evaluating the models at that scale. Scaling techniques are required to relate grid-scale evaporation to the footprint of local observations.

4 On Future Earth Observation Data for Evaporation

The following section seeks to identify a number of target areas that can be focused on to advance the capacity for improving surface flux estimation. The emphasis of this section is on Earth observation (E-O) data that would contribute to this exercise. It should be noted that E-O data does not refer solely to satellite based missions, although some emphasis is placed on these given the funding agency (ESA) and intent of the project.

4.1 The use of numerical weather prediction or reanalysis data

Over the last decade, there have been significant developments in the operationalisation of short term weather prediction and forecasting systems, allowing for near real-time monitoring of coupled land-atmosphere processes. With improvements in the underlying land surface schemes, together with enhanced physical descriptions of atmospheric processes, the large scale simulations derived from these products have been shown to reflect patterns and trends observed in other independent datasets, including those derived from Earth observing platforms (Jimenez et al. 2011). An obvious question is whether dedicated E-O based simulations are required when coupled modeling systems offer commensurate performance. This is especially pertinent when the forcing data required to drive the E-O based modeling systems are derived in part from numerical weather prediction (NWP) schemes or reanalysis data (e.g. ERA)- raising a further issue on the true "independence" of such products. However, it is important to recognise that the question is somewhat ill–posed, as it is not a binary option. The advances made in land-atmosphere modeling have not arisen in isolation of satellite based processes: indeed, satellite data are absolutely critical to the observed improvements, either directly through data-assimilation based integration, or indirectly through offering evaluation metrics of model performance. Likewise, the reliance of satellite based approaches on meteorological data needs to be met from some source, so using the best available observations (in this case, from reanalysis) is appropriate. The question then is whether this invalidates the so-called "independence" of satellite data, given their reliance on modelled data.

One important point in this discussion is that the reanalysis or numerical weather prediction data are not themselves internally consistent. That is, there is no structural requirement that they balance water and energy fluxes. This is important, as one of the recommendations of this work is the development of a consistent forcing dataset to enable a better understanding and characterisation of model structural errors and product accuracy. But it also important in determining whether such data are a viable and reasonable alternative to the types of process-based models being examined here. To answer this requires much more investigation that has been directed to date on the use of NWP and ERA type approaches for flux estimation. Their capacity to do this is based not only on the underlying land surface scheme employed within the coupled modeling system, but also on the extent and nature of the coupling occurring. Furthermore, the assumption that assimilation of linked variables (such as soil moisture) into such modeling frameworks will have a positive effect on surface heat fluxes is not based in any rigorous strong scientific assessment. That it is likely to improve soil moisture is understood, but the flow on effect of this improvement is not well described.

If an over-arching modeling system were developed that could ingest all available observations (ground based and satellite alike), would it perform better than a process-based model focused solely on evaporation? Ultimately, the community is not yet in a position to answer this question. As a general rule, simplicity has been preferred to complexity in Earth system process descriptions, a recognition in part to our limited understanding of complex non-linear systems and their behaviours. While operational weather prediction models provide a potentially valuable source of data for both forcing and estimation of hydometeorological variables, it is premature to declare their primacy over satellite based approaches just yet. At this stage, it is more likely that these two complimentary approaches will converge via the development of improved assimilation schemes. In this regard, developing accurate and independent satellite based estimates of evaporative fluxes is critically important, as it is likely to drive further improvements in the application of coupled modeling systems.

4.2 Developing a consistent forcing dataset

One of the major obstacles towards achieving robust and accurate globally-distributed evaporation estimates is the lack of long-term and high-quality ground-based validation data. While this is a problem that effects a wide range of modeling efforts, the issue is particularly pronounced in the estimation of evaporation, since many key variables such as radiation fluxes and land surface temperature generally do not have a sufficient globally-distributed coverage. The problem is compounded further by the lack of validation sites that adequately reflect a range of climate zones, surface conditions and atmospheric states, all of which severely restrict the capacity to robustly assess candidate models or data products.

Apart from the paucity of distributed data, an equally if not more important issue relates to forcing data sets internal consistency, which is a recurrent problem for land surface products that depend on a large number of datasets. WACMOS-ET achieved a certain degree of consistency in the internally-developed products (e.g., common ancillary data and algorithms to process the radiances from different sensors to produce land surface temperature, or a consistent set of albedo/LAI/FAPAR properly scaled in space to three different resolutions), but much ground for further improvements still exists. For instance, the emissivity used to process the land surface temperature was taken from the Global Infrared Land Surface Emissivity UW-Madison Baseline Fit Emissivity Database developed by Seemann et al. (2008), so there is no connection between the FAPAR used as a proxy for vegetation cover in the SEBS model and the emissivity estimates of the land surface temperature also ingested by SEBS. Perhaps a clearer example is the use of the independent surface radiation product SRB (Stackhouse et al. 2011), instead of an internal product that would have merged existing top-of-the-atmosphere radiances in the relevant spectral bands with the WACMOS-ET albedo, land surface temperature, and atmosphere characterization.

To what degree does this internal lack of consistency in EO products contribute to the errors when estimating a variable such as evaporation is difficult to judge, as in principle there are still not many EO "integrated" data sets that could facilitate such studies. For evaporation estimation, some geostationary platforms may offer an integrated data set over large regions. An example is the suite of products for the Land Surface Analysis Satellite Applications Facility (LSA-SAF, http://landsaf.meteo.pt/), which include downwelling surface radiative fluxes, land surface temperature, surface albedo, fraction of vegetation cover, LAI, and FAPAR products derived from Meteosat Second Generation (MSG) instruments using an internally consistent processing framework. An evaporation product also exists (Ghilain et al. 2011), but it only uses the downwelling radiative fluxes and the albedo from the listed products, so there are still grounds for further developments in terms of a more integrated use of the LSA-SAF for flux estimation (Ghilain et al. 2014).

Regardless, in concert with a thorough assessment and characterization of uncertainty, is the need for the development of an internally consistent forcing data product. At a minimum this would focus on the radiative elements of the models, considering this is a recognized area of uncertainty and sensitivity in all of the candidate models. Even without undertaking an independent sensitivity analysis, having a consistent forcing dataset would provide a mechanism for better understanding the impacts of forcing on model simulations and offer a means to diagnose data induced errors.

4.3 Fluorescence monitoring for informing upon partitioning

The predominant flux in water exchange between the land and the atmosphere is via the process of transpiration. Direct measurements of this variable are challenging in the field, let alone detecting them from space. As such, techniques have been sought to infer transpiration through monitoring the vegetation stress, as stress provides an indirect measure of stomatal response (and hence transpiration). While there are well-developed efforts to determine plant physiological and leaf pigment characteristics such as chlorophyll and carotenoids from multi-spectral approaches, approaches to monitor chlorophyll fluorescence directly are relatively new. With the launch of GOME-2 on board MetOp and OCO-2 has come the capacity to produce spatial maps of fluorescence to enable such process insights. However, there is a current gap in modeling the coupled water-energy-carbon cycles, limiting the uptake of these new datasets for hydrological and related applications. Developing a modeling framework that can integrate plant biophysical response mechanism in a manner that couples with the water and energy cycles represents a key area of needed research that is required to take advantage of these data.

4.4 UAV and Nanosatellites

One of the most exciting domains in the field of remote sensing is the technological advances in nano- and micro-satellites as well as in the research application of unmanned aerial vehicles (UAVs). While operating at a vastly different scale to the focus of the WACMOS-ET project, these systems offer a capacity to meet the demands of a different community that has particular interest in high-resolution retrievals of plant health and water use: the agricultural sector. With developments in both UAVs and sensor technologies has come the capacity to retrieve variables of relevance for agricultural management and production. While the implementation and application of UAVs are beyond the scope of this document to explore, private sector investment in high-resolution (sub-10m) multi-spectral sensing provides a wealth of information potential to both the agricultural sector and the research community in terms of better understanding scaling responses, partitioning and discrimination.

In all of these advances, one key component is lacking: an efficient modeling system that integrates new observations to provide metrics of relevance to the different communities. While technological advances have leapt forward, there has been a lag in the development of science-based applications to take advantage of these new observational platforms. The candidate models here may have some utility in linking with these higher resolution sources, but it is equally likely that a new framework is needed to maximize the information potential of these systems. The development of such a system is a required element in advancing what many consider to be the future of Earth observation.

5 On the Transition from Research to Operations

Few evaporation products exist today that could be considered operational, at least in the sense of delivering flux retrievals over relatively large scales. The only alternative seems to be the use of estimates of surface turbulent fluxes included in the forecast and atmospheric reanalysis from numerical weather prediction centres, which are at relatively coarse spatial resolutions that are dictated by the size of the physical model cells. An example at shorter time and spatial resolutions in Europe is the LSA-SAF ET product that delivers evaporation every 30 minutes and at approximately 5 km resolution over the MSG disk. This product is based on a soil-vegetation-atmosphere-transfer (SVAT) model and uses the downwelling radiative fluxes and albedo from MSG, with the remaining inputs coming from NWP forecasts (Ghilain et al. 2011).

In principle there are no obstacles to implement the evaporation modelling algorithms evaluated during WACMOS-ET in a more operational context, given the needed inputs and with the required data latency. However, given the evaluations carried out during the project one may question the utility of a single-algorithm based operational product and to what purposes it might be used. While it could be argued that for relatively small scale applications connected to agriculture and water management, relative changes in the time evolution of the estimated evaporation can provide useful information e.g., for irrigation purposes (even if the absolute values may be uncertain,), it would be more difficult to justify the relatively large investments required to make one of our evaporation products operational at continental scales for applications where the absolute values have some importance, given the relatively large uncertainty still present in their estimation.

A transition phase is required where further research is directed towards producing more robust evaporation estimates before a more consolidated large-scale operational product could be implemented (e.g., by studying the feasibility of the merging strategies outlined above with an ensemble of single-approach estimates). Such an exercise would also provide an opportunity at integrating other emerging data sets to aid in constraining and discriminating flux components, as well as offering the capacity to explore multi-scale approaches that would meet the requirements of diverse community interests.

6 Concluding Remarks

The issues requiring attention, so that accurate, near real-time retrievals can develop towards operational applications, are certainly not insurmountable. However, it will require a concerted effort involving models and observations representing fluxes at multiple temporal and spatial resolutions to make significant advances in developing reliable global flux products. Although in their infancy and whilst not without problems, the development of these products, especially when driven by strong community need and guided by consistent protocols and assessment strategies, will provide a powerful capacity to expand our knowledge on the Earth and its hydrological regimes into the future.

7 REFERENCES

Anderson, M.C., Norman, J.M., Kustas, W.P., Houborg, R., Starks, P.J., & Agam, N. (2008). A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. *Remote Sensing of Environment, 112*, 4227-4241

Ershadi, A., McCabe, M.F., Evans, J.P., Chaney, N.W., & Wood, E.F. (2014). Multi-site evaluation of terrestrial evaporation models using FLUXNET data. *Agricultural and Forest Meteorology, 187*, 46-61

Ershadi, A., McCabe, M.F., Evans, J.P., Mariethoz, G., & Kavetski, D. (2013). A Bayesian analysis of sensible heat flux estimation: Quantifying uncertainty in meteorological forcing to improve model prediction. *Water Resources Research, 49*, 2343-2358

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 Sensing of Environment, 112*, 901-919

Ghilain, N., Arboleda, A., & Gellens-Meulenberghs, F. (2011). Evapotranspiration modelling at large scale using near-real time MSG SEVIRI derived data. *Hydrology and Earth System Sciences, 15*, 771-786

Ghilain, N., De Roo, F., & Gellens-Meulenberghs, F. (2014). Evapotranspiration monitoring with Meteosat Second Generation satellites: improvement opportunities from moderate spatial resolution satellites for vegetation. *International Journal of Remote Sensing, 35*, 2654-2670

Hoeting, J.A., Madigan, D., Raftery, A.E., & Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science, 14*, 382-401

Jiménez, C., C. Prigent, & F. Aires (2009). Toward an estimation of global land surface heat fluxes from multisatellite observations, *Journal of Geophysical Research: Atmosphere*, 114, D06305, doi:10.1029/ 2008JD011392.

Jung, M., Reichstein, M., & Bondeau, A. (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences, 6*, 2001-2013

Kalma, J., McVicar, T., & McCabe, M. (2008). Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. *Surveys in Geophysics, 29*, 421-469

Kavetski, D., Kuczera, G., & Franks, S.W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resour. Res., 42*, W03407

Lipton, A., P. Liang, C. Jimenez, J. L. Moncet, F. Aires, C. Prigent, R. Lynch, J. G. Galantowicz, R. P. d'Entremont, & G. Uymin (2014). Sources of discrepancies between satellite-derived and land surface model estimates of latent heat fluxes, *Journal of Geophysical Research: Atmospheres,* DOI: 10.1002/2014JD022641

Mahrt, L. (2010). Computing turbulent fluxes near the surface: Needed improvements. *Agricultural and Forest Meteorology, 150*, 501-509

Mauder, M., & Foken, T. (2006). Impact of post-field data processing on eddy covariance flux estimates and energy balance closure. *Meteorologische Zeitschrift, 15*, 597-609

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., & Dolman, A.J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci., 15*, 453-469

Mu, Q., Zhao, M., & Running, S.W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment, 115*, 1781-1800

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd

Seemann, S.W., Borbas, E.E., Knuteson, R.O., Stephenson, G.R., & Huang, H.-L. (2008). Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance measurements. *Journal of Applied Meteorology and Climatology, 47*, 108-123

Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation, 55*, 271-280

Stackhouse, P.W., Gupta, S.K., Cox, S.J., Zhang, T., Mikovitz, J.C., & Hinkelman, L.M. (2011). The NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset. *GEWEX News, 21*, 10-12

Su, Z. (2002). The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci., 6*, 85-100

Sumner, D.M., & Jacobs, J.M. (2005). Utility of Penman–Monteith, Priestley–Taylor, reference evapotranspiration, and pan evaporation methods to estimate pasture evapotranspiration. *Journal of Hydrology, 308*, 81-104

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. *Agricultural and Forest Meteorology, 103*, 279-300

Wang, K., & Dickinson, R.E. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys., 50*, RG2005

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J.B., Zhang, N., Chen, J., Cheng, J., Zhao, S., & Zhang, X. (2014). Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. *Journal of Geophysical Research: Atmospheres, 119*, 4521-4545