Review by Catherine Prigent:

This paper reports on a program of work, the WAter Cycle Multi-mission Strategy project, launched by ESA, with the GEWEX collaboration. The aim of this program is to develop optimized satellite-derived products to describe the water cycle, taking into account the synergies of the large number of now available Earth observation datasets. The parameters to be examined are the evapotranspiration, the soil moisture, the cloud cover, and the water vapour: for each one, the paper describes the adopted retrieval methodology and briefly present some results.

Global long term datasets of key variables of the global water cycle are still not available with the required accuracy for model evaluation or for the analysis of their inter-annual variability. A large range of satellite observations is now available but their use is not always optimized. The objective of this paper is to propose new multi-satellite datasets of key water cycle parameters, which is a very interesting and challenging subject that fits well with the topics covered by the HESS journal.

The introduction and the background sections insist on the links of the WACMOS program with the other international efforts within GEWEX, on the importance of the development of global products over long time periods, and on the benefit of the synergetic use of multi-satellite observations.

Although the study refers at several occasions to international programs, it seems to ignore the collaborative works that are going on. For instance, the LandFlux activity, launched by GEWEX Radiation Panel is currently evaluating the global turbulent fluxes published by different groups worldwide. The WACMOS products do not seem to be part of this initial comparisons (e.g., Jimenez et al., JGR, 2011; doi:10.1029/2010JD014545, Mueller et al, GRL, 201, doi:10.1029/2010GL046230) and the present study does not refer to it. This study also fails to acknowledge the cloud assessment work performed by Stubenrauch et al. within the GEWEX community. No reference is done either to the GPCP effort, when analysing the precipitation occurrence and intensity.

The program insists on the global nature of the water cycle. However, the capacity of the methodologies is not proved at large scales for all products. For instance, the results for the evapotranspiration should be presented at least at continental scale in order to be convincing. By the same token, evaluation of the cloud products at one station only is clearly not sufficient to validate the method.

This study makes use of a large number of satellite observations, but the way these diverse measurements are exploited is questionable, especially for soil moisture retrieval. Both passive and active satellite observations are used to retrieve the soil moisture. Depending on the environments, the method consists in using one retrieval, or the averaged value of both retrievals. Does this juxtaposition/averaging really benefit from the synergy between the measurements? Both active and passive measurements are sensitive to the soil moisture as well as to other parameters such as vegetation cover, soil texture or roughness. Using the two information jointly could certainly help solve this under-constraint problem, but that would require developing a coupled algorithm that does merge the two pieces of information (Aires et al., JGR, 2006; JGR, 2011).

The very simple solution suggested by this paper has to be thoroughly evaluated to be convincing. As presented in the paper, it is clearly not satisfying (the comparison with in situ measurements in Figure 14 even seems to show that the combination of products is worse than one of the product alone). In conclusion, this study is not up to the expectation it initially raises. It presents some work in progress, based on already published work or on the merging of already published data sets, without any convincing evaluation. It does not provide a thorough analysis of new scientific results. One could support the idea of an overview paper that would report on finalized products described elsewhere, but this is not the case either, as none of these products have been carefully assessed yet. Before publications, more work has to be done, to convince the reader of the added value of these multi-mission products.

We wish to thank Catherine Prigent for her critical and helpful review comments. In the following, the comments are addressed in relevant themes as raised in her comments. The comments are split into the different themes for response.

Reply related evapotranspiration theme

We were aware of the LandFlux activity during the time of writing, but the WACMOS products were not ready to be compared on global scale because the initial phase focused on the refining of algorithms and their validations. We will integrate into future activities.

We have added several relevant publications that became available after the submission of this manuscript. The part related to the evapotranspiration, section 3.1, has been modified to reflect these changes.

"Over the last couple of years several initiatives, like the LandFLUX Initiative (Jiménez et al., 2011; Mueller et al., 2011), have targeted to evaluate and develop large scale evapotranspiration products. In general these products can be divided into four groups: the field measurement upscaling methods, land surface models, empirical methods and energy balance methods. Land surface models, like the European Centre for Medium-range Weather Forecasts reanalysis (ECMWF ERA-interim) (Simmons et al., 2007) and the Global Land Data Assimilation System (GLDAS (Rodell et al., 2004), use field measurements of meteorological parameters to simulate the land surface processes. As these models in general do not use remote sensing data for forcing, they have a coarse resolution (>1.0°). Recently the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite Application Facility on Land Surface Analysis (LandSAF) has tackled this problem by incorporating METEOSAT Second Generation (MSG) data in to the ECMWF TESSEL model (Ghilain et al., 2010). While the resolution has greatly increased the data products are restricted to the MSG disc area. The field upscaling methods use actual measurements from the field to estimate evapotranspiration in initiatives like Fluxnet (Baldocchi et al., 2001). As these measurements in general have a too small footprint for global application upscaling is performed using satellite observations (Jung et al., 2010). The disadvantage of these methods is that the spatial resolution (0.5°) provided is still too coarse for most hydrological applications. As empirical and energy balance models estimate evapotranspiration from remote sensing their resolution (<1km) is much better than the other methods. Empirical methods (Mu et al., 2007; Fisher et al., 2008) rely on the calculation of a potential/reference evapotranspiration and combining this with crop and environmental coefficients. In a recent study (Miralles et al., 2010), these empirical coefficients are replaced through the use of soil-moisture and rainfall observations. Energy balance methods circumvent the use of crop coefficients by characterising the actual processes on the land surface. Several studies (McCabe and Wood, 2006; Jiménez et al., 2011; Mueller et al., 2011; Vinukollu et al., 2011) are performed to investigate the performance of energy balance algorithms, like the Surface Energy Balance System (SEBS) (Su, 2002), the Two Source Energy Balance (TSEB) and ALEXI (Anderson et al., 2007; Kustas and Anderson, 2009). While global evapotranspiration from two source models are not available yet, recently a global evapotranspiration product using SEBS and MODIS data has been made available (Vinukollu et al., 2011)."

Comments:

The program insists on the global nature of the water cycle. However, the capacity of the methodologies is not proved at large scales for all products. For instance, the results for the evapotranspiration should be presented at least at continental scale in order to be convincing.

Reply:
Some completed continental scale validation are added and examples shown in Fig. 4.

Reply related to soil moisture theme

We agree with the reviewer that potentially a superior soil moisture product can be obtained when two or more independent observations that are combined already the retrieval process. However, such an approach premises that observations are concurrently made (given the high temporal dynamics of soil moisture), a condition that mostly is not met, especially not for historic observations with a limited temporal revisit frequency. For this reason we base our product on merging existing soil moisture products. Nevertheless, we included references to the work of Aires et al in Section 3.2, as we think that such an approach is very promising for recent and upcoming missions where active and passive observations are made concurrently (like for SMAP).

Regarding the validation: several major editing were made in the results and validation section. An example showing the improved performance was included in Figure 8. We also added a reference to recent evidence (Jung et al., 2010 in Nature) that the trend map shown in Figure 10 shows realistic values. Also reference made to the work of Liu et al., 2011 which shows the improved performance (especially in terms of temporal coverage) of the merged product.

Reply related to clouds theme

To acknowledge the efforts taken within the GEWEX community, we have added a references to the papers of Stubenrauch et al. (2008) and Stubenrauch et al. (1999):

Stubenrauch, C. J., W. B. Rossow, F. Chéruy, A. Chédin, N. A. Scott, 1999: Clouds as Seen by Satellite Sounders (3I) and Imagers (ISCCP). Part I: Evaluation of Cloud Parameters. J. Climate, 12, 2189–2213.
doi: 10.1175/1520-0442(1999)012<2189:CASBSS>2.0.CO;2

Further, it is indeed an omission to not to mentioning the GPCP effort. The following is added to section 3.3.
"Over the past decades several methods have been developed to detect precipitating clouds and retrieve rain rates. The methods developed for geostationary satellites often use thermal infrared observations, and relate daily minimum cloud top temperatures (Adler and Negri, 1988) or Cold Cloud Durations (CCD) to rain rates (Todd et al., 1995). These methods give fair retrieval accuracies for convective precipitation, but are not for stratiform precipitation. Precipitation retrievals for both stratiform and convective clouds are feasible with the more physically-based microwave radiometer (MWR) based methods (e.g. Wentz and Spencer, 1998). The main drawback of MWR based methods is that they only apply to liquid precipitation and that MWRs are only operated on polar orbiting satellites. Similarly, methods have been developed to derive precipitation from cloud physical properties retrievals of passive imagers (Rosenfeld and Gutman, 1994; Lensky and Rosenfeld, 2006; and

Roebeling and Holleman, 2009). Because several passive imagers are operated onboard geostationary satellites the retrievals of these methods can be made available at high temporal resolution. However, the use of visible and near-infrared radiances limits the application of these methods to daylight periods only. Beside single instrument retrievals, methods have been developed that combine measurement from different sources. The Climate Prediction Centre MORPHing (CMORPH) method provides global precipitation estimates by propagating motion vectors derived from geostationary satellite infrared observations on passive microwave satellite scans (Joyce et al., 2004). While the Global Precipitation Climatology Project (GPCP) merges measurements from three different sources i.e., precipitation estimates from low-orbit satellite microwave data, geosynchronous-orbit satellite infrared data, and surface precipitation gauge observations from the Global Precipitation Climatology Centre (GPCC) (Adler et al., 2003)."

Comments
"By the same token, evaluation of the cloud products at one station only is clearly not sufficient to validate the method."

Reply:

Furhther in section 4.3.2 Cloud products: Results and validation, we have put more emphases to the evaluation of SSI and PRP over various climate regions. The following validation results are added with the ISCCP-Flux dataset.

"The SCIAMACHY Solar Surface Irradiance retrievals were compared against ISCCP-Flux Dataset (FD) Shortwave Downwelling Fluxes (SDF). The difference between monthly mean SCIAMACHY SSI and ISCCP was within -12 W m^{-2}, with a standard deviation of 62 W m^{-2} (Wang et al., 2011). The validity of the SSI products from SCIAMACHY and SEVIRI is determined through a comparison against ground-based measurements of the Baseline Surface Radiation Network (BSRN). The SCIAMACHY product was validated for one year against 19 BSRNs stations, while the SEVIRI product was validated for 5 months against 6 BSRN stations. Figure 14 presents the validation results for three BSRN stations in Europe, which show that instantaneous SSI retrievals from SCIAMACHY and daily averaged SEVIRI retrievals agree well with the respective ground-based irradiance values. The precision (RMSE) is about 75 W m^{-2} for SCIAMACHY and about 25 W m^{-2} for SEVIRI. The accuracy (bias) for SEVIRI and SCIAMACHY is about 3 W m^{-2} and 10 W m^{-2}, respectively."

In addition, we have updated Figure 14 with two more validation sites (Carpentras and Payerne). We have also updated the reference list.

Reply related to water vapour theme
To highlight the added value of the WACMOS water vapour products, the following paragraph was added to the product description in Section 3.4:

"Provided the successful development and validation, the expected improvements of the WACMOS water vapour products compared to existing datasets are the combination of benefits from two different sensors and the availability of the water vapour files plus corresponding error maps. In case of the SEVIRI+MERIS product the spatial resolution is better than the grid size of state-of-the-art regional climate models. Furthermore, the continuation of measurements with similar sensors is very likely."

Some detailed comments:

• Several references are not in the list (e.g. Timmermans et al., 2010; Hollweg, 2005).

Reply: Both references are added.

• P. 11. By saying "most current algorithms" the author seems to neglect another "school" of remote-sensing based evapotranspiration algorithms based on modified Penman-Monteith/Prisley-Taylor approaches. Although it is certainly true that for the moment there are no grounds to establish that one methodology is superior to the other, most of the published global estimates currently come from this alternative approach (e.g., same journal, Miralles et al., 2011, doi:10.5194/hess-15-453-2011).

Reply: This has been modified as given above.

• P. 12. Being evapotranspiration an official MODIS product, would have not been of interest to at least try a comparison of the MODIS fluxes derived for Figure 2 with the MODIS product

Reply: The MODIS evapotranspiration product, MOD16 , has been made available since 11 February 2011, which was before the submission of this manuscript. This product provides maximum spatial and temporal resolutions of respectively 1km and 8 day. A comparison between the WACMOS final product and the MOD16 product will be performed in the future.

• P. 15. Why is the US surface model used as a reference? Why not using the ECMWF model in an ESA project? Any reasons?

Reply: As a reference for the triple collocation and data merging we use GLDAS-NOAH instead of e.g. ERA-40 or ERA-Interim because of its superiority in mimicking soil moisture variations also in very dry areas: the shortcoming of the ECMWF series has been recognized and efforts to improve the land surface model have been undertaken (Balsamo et al., 2009). This explanation has been added to the text.

• P. 16. The explanation of errors in ASCAT in terms of sand dunes is not very convincing: the patterns do not correspond at all with the main sand dunes. . .

Reply: Sand dunes are only part of the causes why scatterometers have difficulties in very dry regions. Generally they suffer from azimuthal effects due to non-randomly distributed objects at the earth surface. In the revised manuscript we also stress this issue in the Section 3.2:

"Nevertheless, the geometric arrangements of objects (vegetation, soil particles, buildings etc.) on the Earth surface have a stronger influence on the backscatter measurements made by active systems than on the emissivity measurements of radiometers."

• P.19 Does the author mean that no "simultaneous" global products of net radiation, ground heat flux or sensible heave flux exist? There are certainly global products of net radiation, such as the NASA/GEWX SRB, or the ISCCP-FD-SRF.

Reply: From our review, it does not appear there are high resolution SRB and ERF products, the ones that are available is of 100 km (SRB) or larger (ISCCP) resolution.

• Figure 2 is of very poor quality (the colour scales have to be changed). The square patterns on A and D have to be explained and the caption needs to be more explicit. The main text mentions fluxes scaled up to daily values, but the fluxes presented at the figure seems to be the instantaneous values at the satellite overpass.

Reply:

Plots in Figure 2 are replaced with newly processed images and added with the following explanation.

"Figure 2. Surface heat fluxes and evapotranspiration in June 2008. The monthly average of daily net radiation (mJ m^{-2} day^{-1}), ratio of soil heat flux to net radiation (-), monthly average of latent heat flux (W m^{-2}) and monthly average of evapotranspiration (mm day^{-1}) is shown in respectively figure A, B, C, D. Note some data are missing near the coasts. The impact of the ECMWF forcing fields can be seen as tiles."

• Table 1. The ET part seems confusing. What do HR, LR stand for? Is it not AATSR, MERIS, MODIS GC coverage and MSG MD? They seem to be mixed up. The 25% reported uncertainty, is it based on preliminary validation efforts or a target precision? ET over oceans does not seem discussed at all in the article, but mentioned in the table. Is it also a WACMOS product?

Reply: Table 1 has been updated. The values for the spatio/temporal resolution and uncertainty listed here are the target requirements. The 25% reported uncertainty is based on literature review, e.g. Kalma et al., 2008. The same SEBS algorithm is applied also to oceans, but so far there is no validation data for the processed years, because SeaFlux data do not cover recent years. The temporal resolution set by the requirements will not be achieved due to limited coverage of the satellite over certain areas and cloud contamination. Instead a L2 product will be created.

• Figure 3. The correlation coefficient on figure 3 should be checked. It seems very high given the plots, especially for LE.

Reply: The results shown in this figure are the same graphs as in the current paper submitted to HESSS (Timmermans et al, 2011) which provides the full details.

Timmermans, J., Van der Tol, C., Verhoef, A., Verhoef, W., Su, Z., Van Helvoirt, M., Wang, L.: Quantifying the Uncertainty in Estimates of Surface Atmosphere Fluxes by evaluation of SEBS and SCOPE models, Hydrol. Earth Syst. Sci. Discuss., 8, 2861-2893, 2011.

• Figure 6. On which time period are these maps calculated? How are explained the various spatial structures on these maps? For instance, the large changes along the Amazon River for both ASCAT and AMSR, or the large gradient in the Arabic Peninsula for ASCAT?

Reply: The triple collocation results in Figure 6 are based on the observation period 2007-2008.We added this clarification to the text. Discussing every single detail in this image would go beyond the scope of this paper. Nevertheless, the main issues are discussed here, i.e. the influence of vegetation and the lower performance of scatterometers in desert areas (which also mainly explain the patterns in the Arabic Peninsula. For a more detailed discussion on the results we refer to Dorigo et al. (2010, HESS).

• Figure 7. Is this map fixed, regardless of the season?

Reply: This is a very viable comment. Yes, the map is fixed, irrespective of the season. Indeed, uncertainties are expected to be variable across the year, obviously depending on the character of the seasonality in the different areas. Unfortunately, in most areas and for most sensor combinations the number of common observations is too small to split them up into different subsets. Nevertheless, the map is not fixed for the whole 30 year period, as the changing availability of sensors leads to different relative errors. This has been added to the text.

• Figure 8. It seems that the combined product is worse than the AMSR one. Is it really the average of the two individual products? Could you provide the rms error? On this example the ASCAT info appears to bring very little information. Why using it then?

Reply: Since the publication in HESS discussion we have improved further the merging strategy (e.g. by using a different correlation threshold). Figure 8 has been updated with a series of new plots that clearly demonstrate the improved performance of the merged product in transition regions between sparsely and densely vegetated areas. Even though the accuracy is not much affected, the number of observation days is clearly improved to almost a daily coverage (after 2007).

• Figure 10. Could you elaborate on the variable that is presented? Why would it be negative?

*Reply: The changes shown are changes in volumetric soil moisture ($m^3 m^{-3}$ * 100%). If on average drying has occurred, the percentages become negative.*