USING PYTHON® LANGUAGE FOR THE VALIDATION OF THE CCI SOIL MOISTURE PRODUCTS VIA SM2RAIN

Luca Ciabatta[1], Christian Massari[1], Luca Brocca[1], Christoph Reimer[2], Sebastian Hann[2], Christoph Paulik[2], Wouter Dorigo[2], Wolfgang Wagner[2]

[1] Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy

[2] Vienna University of Technology, Vienna, Austria

ABSTRACT

Remote sensing techniques provide a new way to obtain hydrological variables (i.e. rainfall and soil moisture), mainly in poorly instrumented areas that are fundamental for natural hazard assessment and mitigation. The even increasing availability of satellite derived products characterized by high temporal and spatial coverage requires the development of techniques and instruments for big data volume managing. Moreover, the use of open source systems is highly encouraged in order to increase their use by the scientific community. In this study, the application of the SM2RAIN algorithm to the CCI soil moisture product is proposed as case study. A number of Python® classes and methods have been developed for this purpose, with the aim of creating an open-source web validation tool for SM dataset, within the Earth Observation Data Centre for Water Resources Monitoring (EODC).

INTRODUCTION

Accurate estimates of rainfall and soil moisture are of paramount importance for geo-hydrological (floods and landslides) risk assessment. These variables are usually obtained through a monitoring network or by running land surface models. The first method is impacted by spatial representativeness issues (Kidd and Levizzani, 2011) and requires a big effort in the set up and maintenance phases. The second method instead suffers from some issues like the spatial resolution and seasonal dependent performance (Ebert et al., 2007, Dee et al., 2011).

To overcome these issues, satellite-derived precipitation and soil moisture products may be a valuable alternative. The retrieval of rainfall from satellite is obtained through the inversion of the atmospheric signals reflected or radiated by atmospheric hydrometeors (Kucera et al., 2013). However, if microwave sensors do not pass when it rains, these algorithms are unable to capture the rainfall events thus providing a significant bias in the estimates. Indeed, these methods provide underestimation of the frequency of light rainfall (Huffman et al., 2000, Tapiador et al., 2012) and the overestimation of the frequency of heaviest precipitations which determine a large bias especially for near-real-time products (that do not use gauge observations for bias adjustment).

The even increasing availability of satellite estimates, their increasing spatial and temporal resolution and the development of long-term products, however, involve the use of big data volume analysis and managing platforms and software. As way of example, the European Space Agency Climate Change Initiative (ESA-CCI) developed and provided three different satellite soil moisture (SM) products, by merging several satellite estimates (Liu et al., 2012). These three datasets, namely, "Active", "Passive" and "Combined" (obtained by combining the SM information retrieved from active, passive and active plus passive sensors), provide daily SM estimates on a global scale, with 0.25° of spatial resolution from 1978 to 2014. The big data volume poses a serious problem and can limit the use of such products for climatological and natural hazards related studies. A way to overcome this issue could be the use of cloud computing environments that allow to manage and analyse huge volumes of data. With this aim, the Earth Observation Data Centre for Water Resources Monitoring (EODC) was recently established in Vienna. The EODC provides a powerful open-source cloud computing environment for Earth Observation (EO) data analysis and is now used for Sentinel 1 data pre-processing and provision. On this basis, the main purpose of this study is to present a new set of open-source tools for analyse long-term (>30 years) SM datasets within the EODC platform. However, a long term validation of the products is challenging since ground soil moisture observations are scarce and, except some cases, lack of long term recording.

To overcome this issue, Brocca et al. (2013, 2014) have developed a method – SM2RAIN – that allows to estimating rainfall using only satellite SM observations. The method has shown to be particularly suitable for estimating accumulated rainfall amount. Indeed, at each satellite overpasses SM2RAIN records the SM value and relates it to the amount of water fallen into the soil via the inversion of the soil water balance equation. After been developed and applied worldwide (Brocca et al., 2013, 2014) the method has been successfully tested in flood prediction applications (Massari et al., 2014 and Ciabatta et al., 2016) and for rainfall correction (Ciabatta et al., 2015). These studies have shown that the accuracy of the retrieved rainfall is strictly dependent on the quality of the soil moisture dataset used as input into SM2RAIN. Given that, since different long-term rainfall products are available worldwide, the retrieved rainfall obtained through soil moisture via SM2RAIN offers a unique opportunity to evaluate the quality of the soil moisture observations. In this respect, SM2RAIN is a valuable tool for testing the accuracy of the ESA-CCI SM products, due to the well-known relationship between rainfall and soil moisture conditions. The main idea is that the perfect SM product can record all the variation in SM condition due to rainfall. By inverting the relationship, one can obtain a rainfall estimate and then assess the quality of the SM product by comparing the estimated rainfall with a benchmark. Indeed, by providing SM for three main different products, SM2RAIN has the chance to evaluate their relative performance in an alternative way. In addition, this is a chance for testing the capability of SM2RAIN to producing rainfall by using a continuous, homogenous, long-term SM time series.

DATA AND METHODS

In this study, SM2RAIN is applied to the ESA-CCI SM datasets. These satellite products are characterized by daily temporal resolution and 0.25° of spatial resolution. The Passive and Combined datasets span from 1st November 1978 to 31st December 2014, while the Active dataset from 5th August 1991 to 31st December 2014. The SM2RAIN algorithm is based on the inversion of the soil water balance equation and uses three parameters which are calibrated by using rainfall from the Global Precipitation Climatology Centre (GPCC) full data daily product at 1° of resolution (Schamm et al., 2015) during the period 2008-2010. SM2RAIN has been calibrated on a pixel-by-pixel basis by selecting the closest land GPCC pixel selected with the Nearest Neighbour Algorithm. The algorithm parameters have been estimated by minimizing the Root Mean Square Error (RMSE) between the estimated and the observed rainfall for 5 days of accumulated rainfall. For a detailed description of the SM2RAIN algorithm the reader is referred to Brocca et al. (2013, 2014).

The performances are assessed in terms of correlation coefficient (R) and RMSE, for five days of accumulated rainfall during the calibration period. The estimated rainfall datasets are then assessed by considering their native resolution (0.25°) and by regridding them at 1° of resolution, in order to perform the evaluation on the same grid of the considered benchmark. The regridding procedure has been carried out by averaging the 25 closest pixels to each 1° grid centroid. Python® language has been used to create classes and methods for the geographical, statistical and cal/val analysis steps and has been chosen for the availability of packages for geographical analysis and big data handling and for the strong support provided by the vast Python scientific community. The developed routines take advantages of the analysis libraries developed by the TUWIEN Remote Sensing Research Group, like the Python Toolbox for the Evaluation of Soil Moisture Observations (pytesmo, https://pypi.python.org/pypi/pytesmo, Paulik et al., 2014). All the developed routines will be implemented into the EODC platform in order to create a valuable satellite SM data validation tool. Figure 1 draws the analysis framework highlighting the used Python® packages.

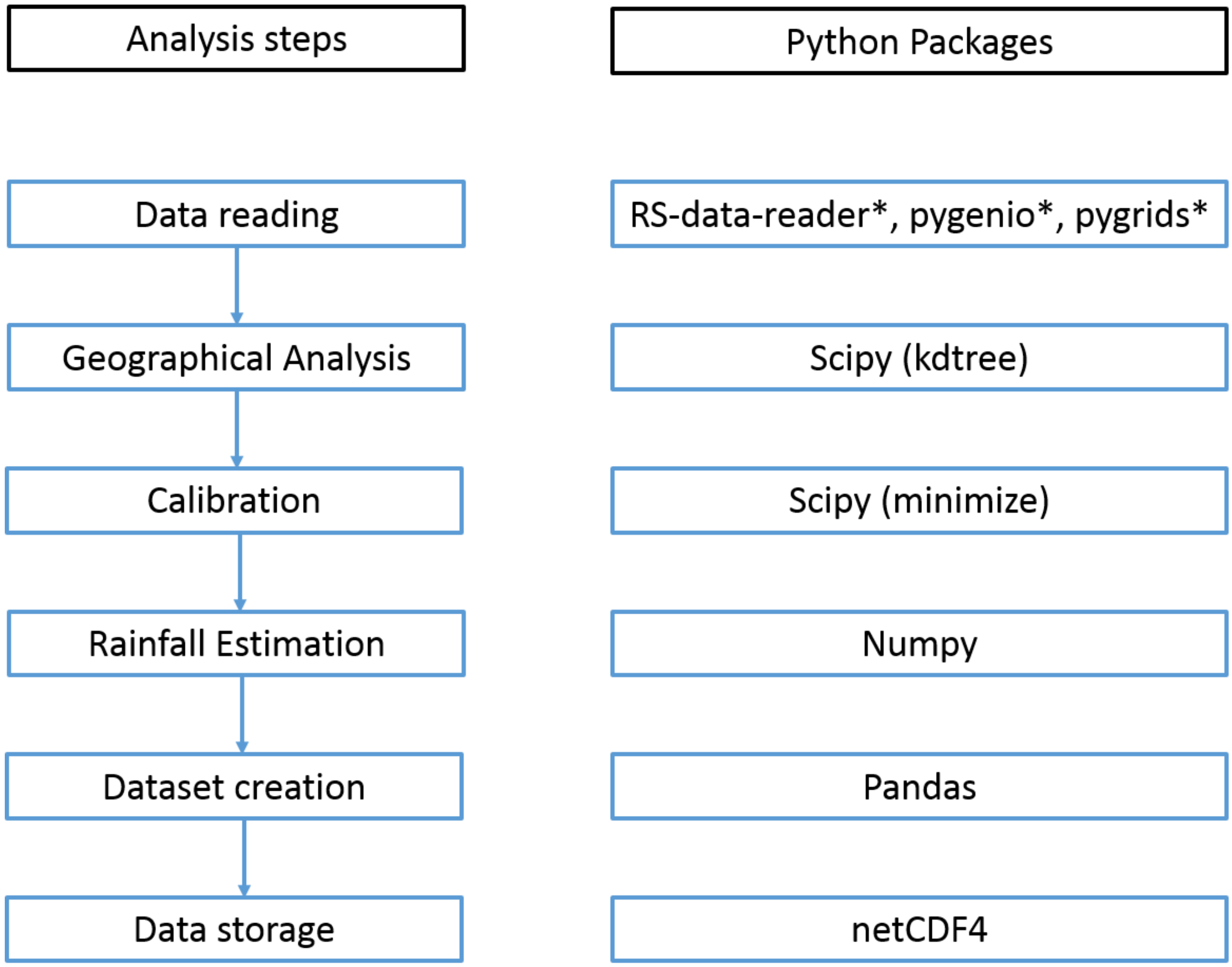


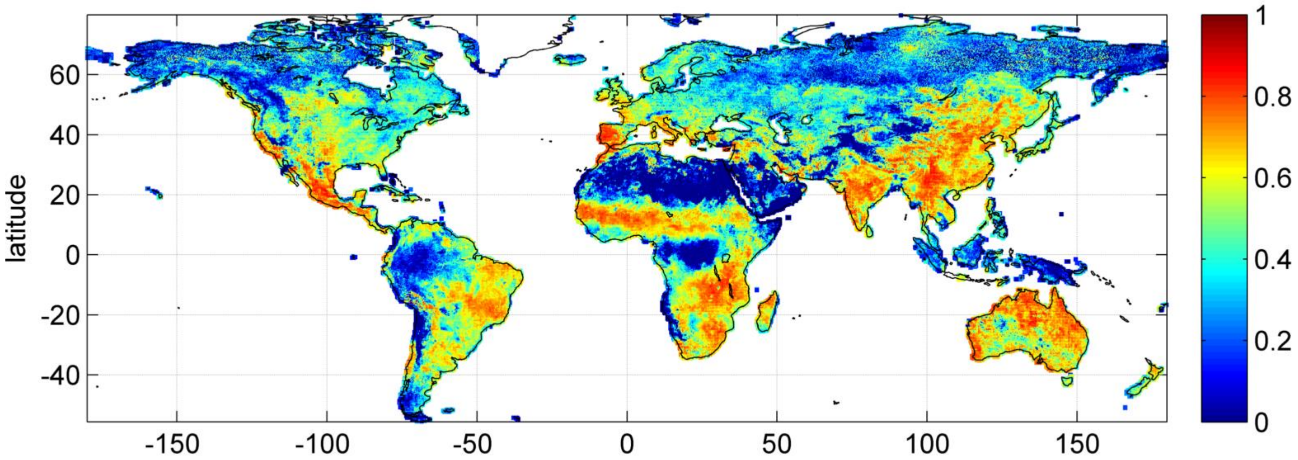
Figure 1 – Analysis framework and Python packages used for each steps. The * indicates Python packages developed by the TUWIEN Remote Sensing Research Group.

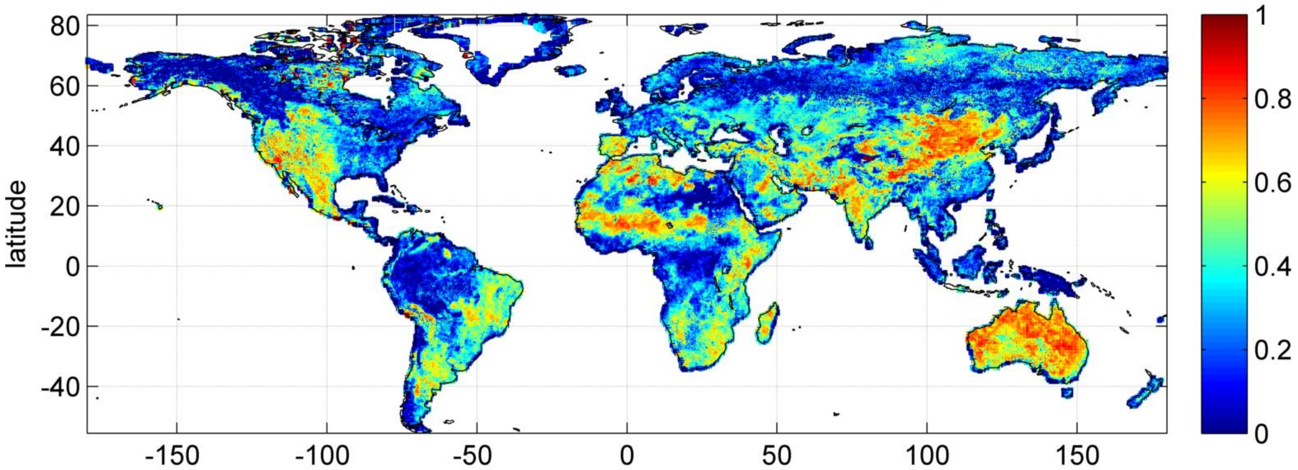
RESULTS AND CONCLULSIONS

The SM-derived rainfall datasets are in good agreement with the observed benchmark. As it can be seen in Figure 2, the performances are depending on the SM input data quality, i.e. the correlation is lower over deserts, over vegetated areas, over mountainous regions and over areas characterized by frozen soil. The rainfall obtained from the passive dataset shows the lowest correlation, due to the quality of the input data, while the "active" and the "combined" rainfall show similar patterns and R median values. The passive dataset seems to perform better over the desert areas and over Australia than the other two rainfall estimates. The combined product shows the highest R values, taking advantages of the two different parent datasets. In terms of RMSE the three rainfall datasets provide a different scenario, with the "active" rainfall showing the best score and the "combined" showing the highest RMSE error. The analysis of the 1° datasets (not shown for the sake of brevity) provides the same trend, with better performance values, probably due to an averaging effect during the regridding procedure. Table 1 summarizes the results obtained for the considered datasets in terms of median R and RMSE by considering the two different analysis grids. It is worth to underline that the median performances do not reach extraordinary values due to the presence of areas where the satellite retrieval is highly impacted, as discussed above. In this framework, the obtained results can be considered very satisfactory.

Product	R		RMSE	
	0.25°	1°	0.25°	1°
Active dataset	0.42	0.56	9.48	8.77
Passive dataset	0.32	0.44	9.72	9.24
Combined dataset	0.46	0.57	12.02	10.74

Table 1 – Median correlation coefficient (R) and Root Mean Square Error obtained for the Active, Passive and Combined ECV-SM derived rainfall during the calibration period, at 0.25° and 1° of spatial resolution and for 5 days of accumulated rainfall.





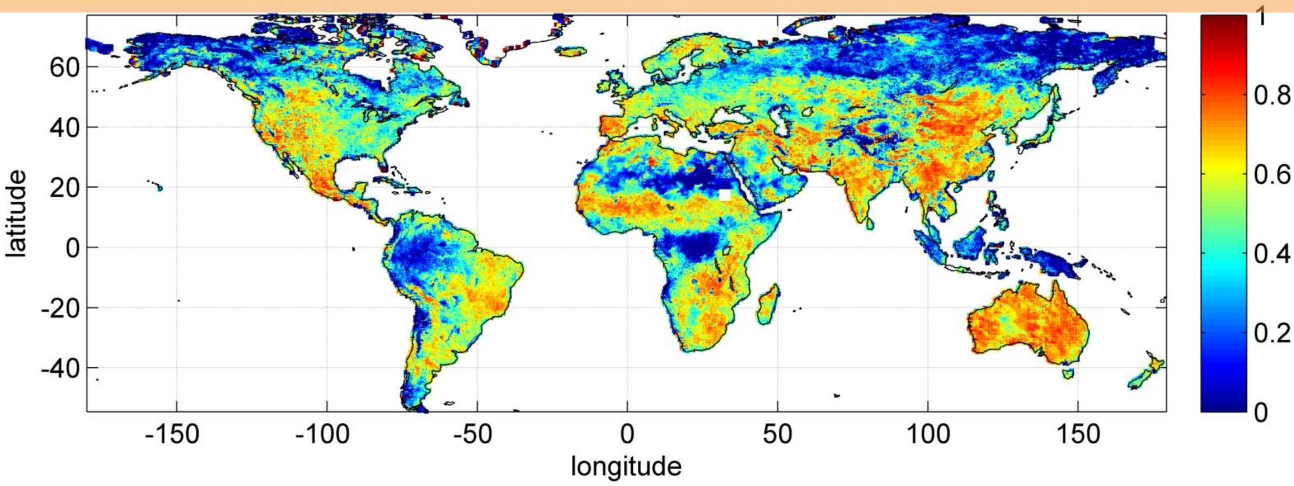


Figure 2 – Correlation maps for 5 days of accumulated rainfall obtained for the active (up), passive (middle) and combined (down) ECV-SM dataset.

Base on the results obtained in this analysis the following conclusion can be drawn:

- Cloud computing facilities can be very beneficial for analyzing huge amount of data and they are becoming a fundamental environment for these kind of analysis, due to the increasing volume of EO data;
- A Python® validation and big data analysis tool is presented. The validation tool will be exported in other open source languages in order to test their capabilities and to find out the best software structure;
- SM2RAIN algorithm can be used for estimating rainfall and for assess the quality of SM dataset, due to the relationship between rainfall and soil wetness conditions. An assessment carried on via SM2RAIN does not need long observed SM records, which can be hardly obtained on a global scale for more than 30 years;
- During the analysis period, the “combined” rainfall outperforms the “active” and the “passive” estimates;

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the ESA Climate Change Initiative for supporting this study. EODC Team is gratefully acknowledged for the technical support.

REFERENCES

Brocca, L., Melone, F., Moramarco, T., Wagner, W. (2013). A new method for rainfall estimation through soil moisture observations. Geophysical Research Letters, 40, doi:10.1002/grl.50173.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hann, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., Levizzani, V. (2014). Soil as a natural rain gauge: estimating global rainfall from satellite soil moisture data. J Geophys Res, 119(9), 5128-5141.

Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Puca, S., Rinollo, A., Gabellani, S., Wagner, W. (2015). Integration of satellite soil moisture and rainfall observations over the Italian territory. J Hydrometeorol., 16 (3), 1341-1355.

Ciabatta, L., Brocca, L., Massari, C., Moramarco, T., Gabellani, S., Puca, S., Wagner, W. (2016). Rainfall-runoff modelling by using SM2RAIN-derived and state-of-the-art satellite rainfall products over Italy. International Journal of Applied Earth Observation and Geoinformation, 48, 163-173.

Dee, D.P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quart. J. Roy. Meteor. Soc., 137, 553-597.

Ebert, E. E., J. E. Janowiak, and C. Kidd. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. Bull. Amer. Meteor. Soc., 88, 47– 64, doi:10.1175/BAMS-88-1-47.

Huffman, G.J, R.F. Adler, M.M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2000: Global precipitation at one-degree daily resolution from multisatellite observations. J. Hydrometeorology, 2, 36-50.

Kidd C., Levizzani V. (2011). Status of satellite precipitation retrievals. Hydrol Earth Syst Sci, 15, 1109-1116.

Kucera, P. A., E. E. Ebert, F. J. Turk, V. Levizzani, D. Kirschbaum, F. J. Tapiador, A. Loew, and M. Borsche (2013). Precipitation from space: Advancing Earth system science. Bull. Amer. Meteor. Soc., 94, 365–375.

Massari, C., L. Brocca, T. Moramarco, Y. Tramblay, and J.-F. Didon Lescot. (2014). Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall. Adv. Water Resour., 74, 44–53.

Paulik, C., Steiner, C., Hann, S., Melzer, T., Gruber, A., Wagner, W. (2014). Open source toolbox and web application for soil moisture validation. Proceedings of the IGARSS 2014, (2014), ISBN: 978-1-4799-5774-3; Paper ID 3331, 3 pages.

Schamm, K., Ziese, M., Raykova, K., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U. (2015). GPCC Full Data Daily Version 1.0 at 1.0°: Daily Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. DOI: 10.5676/DWD_GPCC/FD_D_V1_100

Tapiador, F. J. et al. (2012). Global precipitation measurement: Methods, datasets and applications. Atmos. Res., 104-105, 70-97.