Monitoring of evaporation at the global scale using artificial intelligence

Laboratory: LERMA, Observatoire de Paris, 61 avenue de l'Observatoire 75014 Paris
Supervisor: Filipe AIRES, filipe.aires@obspm.fr, 06 20 15 22 98
Requested background: Strong background in applied mathemathics/physics. Good programming skills.

Type of subject:

Theory	~~not at all~~	a little	~~a lot~~
Numerical modeling	not at all	~~a little~~	~~a lot~~
Experimentation	not at all	~~a little~~	~~a lot~~
Data analysis	~~not at all~~	~~a little~~	a lot
Instrumentation	not at all	~~a little~~	~~a lot~~

This internship could be followed by a PhD.

This research subject is truly multi-disciplinary, it is at the intersection of several disciplines: physics, remote sensing, hydrology, climate, applied mathematics, artificial intelligence. Evaporation (*E*) is a major component of the water cycle: It influences the water and energy cycles (Dolman et de Jeu 2010) because of the surface energy flux partitioned into its latent (*E*) and sensible (temperature) components. *E* limits the soil moisture available for vegetation grows so its impact in agriculture and water availability is determinant for society. Unfortunately, *E* is one of the most difficult geophysical variable to measure: Only a few *in situ* measurements are available around the globe, and these point-measurements are hardly representative of their surroundings (Pastorello et al. 2020). Furthermore, no satellite observations can actually measure *E*, so indirect satellite measurements (land surface temperature, wind, soil moisture) are used with an empirical energy-based models (Priestley & Taylor, 1972; Penman, 1948; Monteith, 1965) in order to estimate it (Mu et al., 2011; Zhang Yongqiang et al., 2016; Miralles et al., 2011). However, these satellite datasets of *E* have large uncertainties: In addition to important random errors, they also suffer from large regional biases (Michel et al., 2016). These systematic biases over some environments are a major problem to study the water cycle and it is the objective of this subject to propose a new approach to reduce them.

The need for model-independent *E* estimate in order to benchmark GCM is important. The rationale for this is the maximization of the use of satellite observations while avoiding the use of had hoc and empirical parameters of the current methods. Moving from classical physically-based algorithms towards data-driven methods is foreseen as a solution to improve the accuracy of *E* datasets. Hybrid approaches combining both strategies is also a promising approach.

Our group has developed a data-fusion framework that integrates several satellite datasets monitoring most of the water cycle components (Aires, 2014). This integration uses the Optimal Interpolation (OI) technique (i.e. a Bayesian framework) in order to optimize them and improve their mutual hydrological coherency. Our OI results suggest that *E* has large regional biases, incoherent with the other water cycle components (Pellet et al., 2021a; b). Budget closure approaches such as OI can potentially be used to correct and improve the current *E* products. However, since OI needs *in situ* river discharge estimates it can only be used over monitored basins (rare) and at the basin scale.

We propose to use AI to build an integration Neural Network (NN), trained on OI results, over a large collection of basins around the world (GSIM: Do et al., 2018). Such an approach was initially presented in (Aires 2014) where a NN was proposed for the first time to improve the closure of the water cycle provided by the satellite products. Here, we propose to include as inputs to the NN several environmental variables describing different environments potentially driving the quality of the satellite datasets. We will also investigate how possible it is to introduce physical constraints inside the NN architecture itself. The objective would not be to optimize all the water cycle components, but rather to obtain a state-dependent correction of the *E* datasets. Further work will involve the design of a full hybrid *E*-retrieval approach combining a physical scheme and a machine learning parametrization of it.

In this project, we will first need to build a learning database based on the OI results providing new *E* estimates over a large collection of basins around the world. The obtained OI results need to be analyzed to make sure that the new estimates are reliable enough. Then, new NN architectures will be tested, with different learning schemes and inputs. An important point is to estimate the uncertainties of the NN estimates by following for instance a previously developed method (Aires and Pellet, 2021). A large activity will concern the evaluation and validation of the results using global *in situ* database such as FLUXNET, and other geophysical variables (TS, MODIS IR).

Bibliography:

- Aires (2014). Combining datasets of satellite-retrieved products. Part I: Methodology and water budget closure. J. Hydrometeorol., 15 (4), 1677–1691.
- Aires & Pellet (2021), Estimating retrieval errors from neural network inversion schemes - Application to the retrieval of temperature profiles from IASI, *IEEE Trans. on Geosc. and Rem. Sens.*, 59, 8, pp. 6386-6396.
- Do, Gudmundsson, Leonard, & Westra: The Global streamflow indices & metadata Archive (GSIM), Earth Syst. Sci. Data, 10, 765–785.
- Dolman, de Jeu, (2010) Evaporation in focus. *Nature Geosci* 3, 296.
- Monteith (1965). Evaporation and environment. (Vol. 19).
- Michel, Jiménez, Miralles, Jung, Hirschi, Ershadi, Fernández-Prieto (2016). The WACMOS-ET project - Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. Hydrol. Earth Syst. Sci., 20 (2), 803–822.
- Miralles, De Jeu, Gash, Holmes & Dolman (2011). Magnitude and variability of land evaporation and its components at the global scale. Hydrol. Earth Syst. Sci, 15, 967–981.
- Mu, Zhao, & Running (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ., 115 (8), 1781–1800.
- Pastorello, Trotta, Canfora *et al.* (2020), The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci Data* 7.

- Pellet, Aires, Yamazaki & Papa (2021). Coherent satellite monitoring of the water cycle over the Amazon. Methodology & initial evaluation. *Water Resources Research*, 57.
- Pellet, Aires & Yamazaki (2021). Coherent satellite monitoring of water cycle over the Amazon. Total water storage change & river discharge estimation. *Water Ress. Res.*, 57.
- Penman (1948). Natural evaporation from open water, hare soil and grass. Proc. R. Soc. Lond. A. Math. Phys. Sci., 193 (1032), 120–145.
- Priestley & Taylor (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. Mon. Weather Rev., 100 (2), 81–92.