

Review of paper “Operational Earth system hydrology at ECMWF” by Mazetti et al.

This is an important paper as it describes the world leading operational Earth system model. The hydrological data it is producing will be used in many applications. Thus a good documentation of the assumptions made in the design and the quality of the variables outputted is essential but should also be measured to the highest standards. I would recommend publication of this paper but after major revisions. As detailed below, the paper does not currently provide a scientifically rigorous documentation of where the system stands within the state of the art, how reliable the hypothesis it is built upon are and how that impacts the outputs.

Major Comments

- With the sentence starting in line 46 (“The comprehensive and Physically consistent ...”) two promises are made which are not realised in the paper. To demonstrate that the system is physically consistent the various assumptions made have to be spelled out so that the reader can determine which relations can be expected between the various variables made available. Implicitly throughout this paper the assumption is made that the continental water cycle is pristine, i.e. without human intervention. If that assumption would be spelled out, the authors would not claim in the second part of that sentence that a “realistic representation of freshwater fluxes from river” is achieved. Many studies have shown that the major driver for river discharge changes over the last century in Europe has been human intervention (Vicente-Serrano et al. 2019 or Collignan et al. 2025). The same can be assumed for other mid-latitude continents.
- Placing the ECMWF operational system within the state of the art achieved by the community is simply a matter of scientific honesty. In the references provided at the end of the paper, less than 30% are not from the co-authors or their institutions ! Other Earth system models which strive to provide a “physically consistent treatment of the water cycle” exist but are not referenced : JULES, SURFEX, CLM, WRF-HMS, ORCHIDEE, ... They do not all aim at the medium range weather forecast but that does not make them less ambitious in achieving physical consistency and realism. To maintain our intellectual rigour we have to keep in mind that we are standing on the shoulders of giants !
- Because of the position of ECMWF within the community, it is unavoidable that the data produced by this system will be evaluated in different studies and used to train machine learning models. Who will remember that in this system the assumption of pristine continental surfaces has been made if it is not spelled out here ? What will ML models do if they ingest data from this system and actual observations of the hydrological system ? The user community has to be made aware of the risk of incompatibilities.
- In the description of ecLand (section 3.1) and CaMa-Flood (3.2) nowhere are the chosen conservation laws defined and which terms achieve the coupling. Not in Bousetta et al. 2021 either. ecLand output is known to neglects the water convergence in the lake component (river and groundwater inflows minus water outflow) and thus breaking the catchment level water conservation (Polcher et al. 2026). The energy convergence through lateral water advection is also neglected (Can be significant for some lakes !) At least this assumption is consistent between ecLand and CaMa-Flood ! It is essential to write these assumptions down so that the reader can evaluate which conservation principles can be assumed when using the data. I know all these principles get broken during the data assimilation phase (section 4.6.2) but respecting them in a free running configuration of the system is part of what I would call a “physically consistent treatment of the water cycle”.
- Why would a “A mass-conservative re-gridding” between ecLand and CaMa-Flood be important if the water mass conservation equations are different or inconsistent in both models ? Just a few cases to illustrate the criticality of this point :

- ecLand has a lake water continuity equation without the horizontal divergence term while CaMa-Flood has that term but not the vertical fluxes, how is this made physically consistent ? The interpolation chosen does not change anything to the problem as long as the correct fluxes and state variables are not exchanged.
- ecLand and CaMa-Flood run on different grids and thus do not have the same land/sea mask. How are catchments defined (The basic hydrological closed system on which water conservation can be assume.) if some points of CaMa-Flood are ocean for IFS ? This issue also needs to be addressed in section 4.6.4.
- CaMaFlood computes “water level and inundation extent” how does that inform the continental evaporation in ecLand or ensures that the soil moisture below is filled to saturation first ? Some models have implemented solutions to this (Guimberteau et al. 2012, Decharme et al. 2019, Schrapffer et al. 2023, ...)

If such points are not clearly defined it is not obvious that the system represents a physically consistent water cycle.

- Figure 3 is difficult to follow without further information. I guess the arrows between ecLand and the atmosphere is some sort of implementation of the PILPS interface. What do the arrows from the atmosphere to the rivers represent ? Is this rainfall over lakes and floodplains ? If so, where can this water evaporate and how is consistency ensured ? Some groups have properly describe these coupling issues for their models : Shrestha et al. 2014 for COSMO-CLM-ParFlow, Wagner et al. 2016 for WRF-HMS, Hallouin et al. 2022 for JULES, ...Your graphics are nicer than those in the cited papers, but you do not name the fluxes or state variables which are exchanged. That information is not in the legend of the figure or the manuscript.
- The global evaluation of river discharge is sketchy. Figure 6 shows the flow direction of the major rivers as imposed on CaMa-Flood by the MERIT Digital Elevation Model. The insert is much more interesting. If I read correctly, 90% of the GRDC stations have a KGE below 0.6. A KGE of 0.41 means that the model predictions are not better than climatology and thus that systematic biases exit. This is the case for more than 70% of all stations ! Why does the model perform so poorly ? This should be discussed and linked to the assumptions made or propagation of errors within the system. An essential information for all users of the output from this system !
- The 3 regional applications shown (Australia, Zambia and Pakistan) are perfect for an ECMWF brochure but do they reach the level of a scientific publication ? The base line climatology of the model on which these extreme events are predicted is not evaluated. The quality of the reproduced hydrological events is not quantified (is the extent and timing of expansion and retreat of the flooded are of Pakistan correct ?). There is no reference either to a study which analyses this. A trained eye notes that in the Australian and Zambian cases, the simulated discharge peaks (or water levels) are one to two days earlier than the observations. Is this an error in forecasted precipitation, is this linked to the spin-up of the river model, is the infiltration/runoff partition in ecLand correct, are there parameters of the aquifer to be adjusted ? Some working hypothesis and improvement path should be presented to the future user of this operation forecasting system.
- The CH₄ fluxes displayed for May in figure 10 are quite different. Which one is closer to truth ? How correct are the floodplain extents predicted by the model and does this give arguments to assume that the output of the coupled system is more realistic ? It could also be that the inter-annual variability afforded by the coupling is so large that even a systematic error is acceptable because of the gain brought by the weather/climate feedbacks on the fluxes is valuable ? This should be documented.
- Section 4.6.3 : The runoff generation at different spatial resolutions is only part of the issue. At higher resolution the recharge of soil moisture from the shallow aquifers is critical for the spatial distribution of surface fluxes, as hypothesised by Fan et al. 2019. It has now been demonstrated that a km-scale resolution this contribution of hill-slope process to evaporation

is a missing process in most land surface models (Polcher et al. 2026). This goes back to the above issue of how to define the coupling between the vertical energy, water and carbon fluxes treated in ecLand with the lateral water fluxes predicted by CaMa-Flood.

Minor issues

- Line 39 : Such developments are also Supported by GEWEX/WCRP (see Stephens et al 2023).
- Line 164 : Section 2.3 could move to an annexe. It is expected from any professional Earth system model development team to have quality control procedures in place for the code and the model output.
- Line 224 : How can the 1D simulations be used to evaluate the atmosphere-to-land-to-river ? These cannot be point scale simulations but have to be at the catchment level. Can you please clarify.
- Figure 1 and explanatory text : It is not explained in the paper how IVER and QUAVAR have evolved with the addition of the lateral water transports in the ESM. Surface hydrology is arguably one of the best observed branch of the water cycle and thus offers excellent validation opportunities as outlined later in the paper.
- Line 323 : CaMa-Flood includes a reservoir operation scheme. Does ecLand also include a water usage model (irrigation, domestic, industrial) to estimate the water consumed (i.e. evaporated) ? Is there or is there not an assumption of pristine continents in the system ? Please clarify if this option is activated or not.
- Line 334 : What are “river initial conditions” how can they be separated from groundwater initial conditions in particular for near surface aquifers. How long do these slower reservoirs take to spin-up from a cold or warm start depending on the size of the catchment and its hydrogeology ? This spin-up temporal scale is essential to determine the inertia of the various water reservoir and thus the importance of the initial conditions for this forecasting system.
- Line 350 : Does “coupling land to river” make sense when rivers are an integral part of the land ? It has been shown that redistribution of water by groundwater and rivers shape the landscape and drives evapotranspiration fluxes (Fan et al. 2019, Polcher et al. 2026). So why limit the coupling to rivers ?
- Line 367 : Why is only the flooded fraction passed to ecLand and not the lake fraction ? It is a minimum to ensure consistent assumptions on the water mass conservation. Later in this paragraph the authors understand the limit of their inconsistent assumptions (as “infiltration of the flooded water” is needed) between both models. Posing these assumption up-front would make the sentence starting in line 370 clearer.

I am sorry for being difficult but I think that one of the best Earth system modelling and prediction centres should also be held to the highest scientific standards. It is for the editor to decide if I went too far.

Jan Polcher