Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC1
https://doi.org/10.5194/hess-2021-126-AC1, 2021

Reply on RC1

Daeha Kim et al.

Author comment on "A comparative evaluation of the calibration-free complementary relationship with physical, machine-learning, and land-surface models" by Daeha Kim et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-126-AC1, 2021

We greatly appreciate the efforts of the referee for reviewing our study. The constructive comments will help us to improve the manuscript. To consider the referee's comments, we will update our CR ET calculations with locally preferable forcing (the SILO data). We found that the ERA-interim forcing inputs are not very different from the SILO data. It is also found that the SILO Precipitation (P) and the GPCC P data are not considerably different at the 0.5° grid scale (see Figure S2). Hence, we believe that the evaluation of the CR method would not considerably change even though the datasets are replaced with the locally preferable ones.

However, we will update our calculation with the locally preferable ones. It will improve confidence of our analysis at least somehow. Since the referee has a concern about ETwb, we will use the SILO P data for estimating ETwb. And, in revision, we will add a comparison between the ensemble of the modeled runoff (Q) product and the Q observations given by the referee. Since the GRUN and the LORA datasets were already validated by global Q observations, we do not expect that the ensemble of the two would not substantially biased from the observations. However, due to the scale-mismatch between the modeled Q and the observations, only sufficiently large catchments could be considered in the comparison. This revision will improve the confidence of the comparative evaluation.

It should be noted that the accuracy of ETwb is mostly determined by the quality of P data, not by Q. Hence, even though some biases are found with the modeled Q, that would exert minor influences on the accuracy of ETwb. In Australia, 90% of P evaporates (Glenn et al., 2011). Our responses to specific comments are following. For instance, the LORA and the VIC simulated mean annual Q of the Murray-Darling river basin as much as 15 mm/a and 42 mm/a, respectively (Hobeichi et al., 2019), and the difference between the two seems large. Nevertheless, ETwb (P – Q) estimates corresponding to the mean annual P (470 mm/a) of the basin are 455 mm/a and 428 mm/a, respectively. The relative difference between the ETwb estimates is only 6%. Indeed, we did use Q products from multiple models for the evaluation, so that potential biases in the modeled flows might be reduced.

Our responses to specific comments are following.

Comment 1: The most questionable issue in this MS is that the authors still use ERA-

Interim forcing to drive the CR in the MS submitted in 2021 rather than 2011, the latter is the year when the widely-known ERA-Interim paper was published in Quarterly Journal of the Royal Meteorological Society. In 2021, this is an obviously out-of-date meteorological forcing which should not be used (and also ERA-Interim's ET output). Its successor, ERA5, with a spatial resolution of 0.25 degree, should be considered as a significant improvement, evidenced by a few recent papers when both ERA-Interim and ERA5 were used to drive the ET model (Martens et al., 2020) and hydrological model (Tarek et al., 2020). In general, ERA-Interim has larger errors than ERA5, which would propagate into the CR-model simulated ET values. It is very strange that the authors did not use locally-preferable meteorological forcing when they focus only on Australia. In Australia, the SILO forcing (https://www.longpaddock.qld.gov.au/silo/), produced by interpolation of ground observations from the Australian Bureau of Meteorology and also other sources, should be much better than those forcing developed for a global coverage. While SILO has no net radiation nor wind speed, its air temperature, vapor pressure, air pressure, and solar radiation should be much more reliable than those from ERA-Interim/ERA5 for Australia. For this reason, the authors should try to drive the models using these.

Response: We will update the ERA-interim temperature forcing data with the ones provided by the Bureau of Meteorology (BoM) of Australia, since it would not take long for us to update the calculations. However, we do not agree that the ERA-Interim inputs are unreliable forcing to the CR method. Even though they are updated to the ERA-5, the ERA-Interim reanalysis system performed even better than the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA) in reproducing observed variations of precipitation (Acharya et al., 2019). Ostensibly, one can assume that the local data sources are more reliable than global products; however, global datasets, too, could be reliable for a hydrometeorological analysis.

The objective of this study is to evaluate the CR method at a common grid scale at which physical, machine learning, and land surface models are developed. At the grid scale, we believe that the ERA-Interim data have been reliable (e.g., Kim et al., 2021).

In addition, the referee's argument that global data sources would not be as good as local data is a hypothesis that should be tested. Hence, we compared the mean vapor pressure deficit (VPD) of the ERA-interim, which is a major input for the CR method, with those from the locally preferable SILO archive. Figure S1 shows that except some overrated values in the north-western part, the mean VDPs from the ERA-Interim were not considerably deviated from the SILO data (see).

Indeed, since a higher VPD indicates a lower ET in the CR principle, the CR ET estimates will increase at least somehow when the ERA-Interim forcing is replaced with the SILO data, being more deviated from ETwb. Therefore, our conclusion on the calibration-free CR method (i.e., higher than ET_{wb}) would not change.

Nonetheless, to consider the comment, we will update our calculations with the preferable ones (temperatures and vapor pressure datasets from the SILO archive), and this revision will reduce concerns about input-data quality. Since the SILO archive does not provide wind speed and net radiation data, we will use the ERA-5 datasets for the recalculation. This revision will not take long, and we do not expect largely different outcomes.

Comment 2: Net radiation is often regarded as the most sensitive input for most ET models (see e.g., Figure 3 in Fisher et al. 2017). For this reason, I have to remind that net radiation from any atmospheric reanalysis dataset is essentially from the model simulations (for upward short- and long-wave radiation), which may have greater uncertainties than satellite observations. While satellite-based net radiation often has relatively coarse spatial resolution (one degree), the authors used 0.5 degree for their simulations, which is not far from it. Therefore, I wonder if the CR model could be improved with satellite-based net radiation data driving it. I suggest trying to use net radiation from CERES (https://ceres.larc.nasa.gov/) and/or GEWEX SRB (https://www.gewex.org/data-sets-surface-radiation-budget-srb/).

Response: We disagree. To predict ET, the CR method mainly uses the response of VPD to soil moisture deficiency rather than depending largely on radiation data. Any methods that assume the proportionality of ET to net radiation (e.g., the PT-JPL and GLEAM) could produce ET products being sensitive to changes in net radiation. However, the predictor of the CR method is the ratio between E_p and E_w, which is a normalized variable. Hence, while Ma and Szilagyi (2019) found outstanding performance of the CR method with a reanalysis radiation dataset, its similar performance was found even with net radiation estimated by the simple standard method (Kim et al., 2019).

Even though the ERA-interim radiations are modeled values, they could become forcing data to the CR method and can lead to outstanding performance (e.g., Kim et al., 2021).

On the contrary, the satellite radiation data would lead to scale-mismatch with the other ET models and the ERA forcing inputs, while precision of the remote sensing observations are not always guaranteed. The satellite radiation is unlikely good in our assessment.

Instead, we will re-calculate CR ET with the ERA-5 radiation datasets to benefit from the updated ERA reanalysis system.

Comment 3: As the water-balance-based ET is key for assessing the ET models, the weakest point in ETwb of this MS is the grid-based runoff data (GRUN and LORA) which has much larger uncertainties than the station-measured ones at the outlet of a basin. In Australia, the most popular runoff data is “Zhang et al., 2013. Collation of Australian modeller's streamflow dataset for 780 unregulated Australian catchments. CSIRO Land and Water, 115 pp.” (available at: https://publications.csiro.au/rpr/pub?pid=csiro:EP113194). Note also that it is more appropriate to involve only the unregulated basins with minimum human activities for validation purpose. While the authors did compare their continent-averaged ETwb with previous similar studies in Line 230-239, this does not necessarily mean that ETwb is accurate at the grid or basin scales.

Response: This is an arguable comment. The Q data for 780 unregulated catchments only can evaluate ET in the gauged catchments. They are too coarse to evaluate the modeled ET for the entire continent. Importantly, the catchment Q observations will lead to scale-mismatches with the modeled products, because many of the 780 catchments are smaller than the 0.5° grid resolution. The Q observations are accurate, but cannot lead to general conclusions on performance of the CR method for the arid continent.

In this case, we believe that an acceptable evaluation reference with a larger spatial coverage could be a better option. The grid ETwb could be acceptable when the P and the synthesized Q data are of good precision. Even though the ETwb is not true values, it could become a reference for cross-evaluation of the modeled ET, possibly providing practical information for ungauged basins. For example, Pan et al. (2020) assessed numerous ET models against an ensemble of modeled ETs.

Approximately, in Australia, 90% of precipitation evaporates, and thus reliability of ETwb depends mostly on the quality of P data. Figure S2 is comparison between the locally preferable SILO P and the GPCC P data. Only 2.1% difference was found between the

SILO and the GPCC precipitation for 2002-2012. We do not agree that this slight difference makes the GPCC data and the ETwb unreliable. The objective of this study is a comparative evaluation, not an absolute evaluation.

To improve confidence of our evaluation, we will compare the modeled Q data with the Q observations of catchments larger than the grid resolution.

Comment 4: Another key deficiency is the precipitation data for calculating ETwb. For the same reason explained above, the authors should use precipitation data with a regional/continental focus rather than the one developed for a global coverage. The BILO precipitation data is often regarded as the most reliable one for Australia (https://data.gov.au/data/dataset/67749ef0-7223-437a-851a-573edde09567), which should be used to replace GPCC for a more accurate ETwb. I do understand that the authors want to use grid-based ETwb data for evaluations, but the authors should also test its reliability with basin-scale ETwb. The latter could be derived using measured runoff data from the above-mentioned 780 basins.

Response: We believe that the SILO data based on local observation would be good. Beesley et al. (2009) validated the daily SILO precipitation using the leave-one-out-cross-validation, and showed acceptable performance. As replied, we will add the comparison between the catchment Q observations and the ensemble of the modeled Q.

Comment 5: I would further argue that the authors' ETwb data are at least 10% smaller than the real values most likely are due to the suboptimal choice of the precipitation as well as the gridded runoff data. The reason I am saying this is that FLUXCOM and CR ET yields pretty similar values seen in Fig. 4. It should be noted that FLUXCOM must provide (and it does) one of the most accurate ET data available today since it is based on actual ET measurements by eddy-covariance, even though its inter-annual variance is somewhat subdued as seen in Fig. 3 in comparison with the CR ET values, but this temporal smoothing feature of FLUXCOM is well known from previous studies.

Response: We disagree. Since it could be regionally biased and inaccurate, the FLUXCOM might not be accurate in Australia, where flux observations for training are insufficient. Please see the number of the FLUXNET2015 stations in Australia here (https://fluxnet.org/sites/site-summary/). Even though the FLUXCOM is based on eddy covariance flux data, the towers are usually installed in accessible locations only. And, large part of Australia is almost inhabitable for humans. Operating a flux tower in such a location is very difficult. It is very likely that the training data were insufficient in the arid continent. Hence, the quality of the FLUXCOM is questionable in Australia, and the referee's argument is hypothetical.

Comment 6: So all in all, it is argued here that most of the difference between CR ET and ETwb is most likely due to unsatisfactory choices in model and water-balance forcing rather than to the need of spatially changing the alpha value of the CR model. (This is not to say that the CR would not overestimate ET rates near the sea since there the air moisture is significantly decoupled form the underlying land surface). As another choice, the authors could also apply several different sources for the forcing in the CR model as well as in the water-balance and see how they affect the outcomes (this would also serve as a sensitivity analysis). Chances are they do in a significant way.

So before one just replaces a unique calibration-free model with one needing calibration, one must make sure that the original model was evaluated correctly and exhaustively. I do not feel at all this is the case in this study.

Response: This comment is based on a hypothesis that the ERA-Interim forcing for the CR method and the P data for ETwb are unreliable. But, here we showed their acceptability by the direct comparisons with the locally preferable datasets. The referee's argument could be an overstatement.

However, as replied, we will update our calculations with locally preferable ones to prevent readers from any prejudice. New calculations would not take long, and evaluations on the CR method is unlikely to change much.

Comment 7: Note also that there is a significant difference between Brutsaert's alpha and the alpha value of the CR employed in this study. The Priestley-Taylor (PT) equation is evaluated at the measured air temperature in the former case, while in the latter at the required (but mostly unknown) wet-environment air temperature (estimated via Tws). Without the latter, the PT equation naturally overestimates the wet-environment ET rates (and thus the actual ET rates as well) the more significantly, the drier and hotter the environment has become, therefore a correction (typically based on some measure of aridity) in the alpha value is necessary in the Brutsaert model, but not in the CR model employed in this study. So in this study the alpha value is meant to be the best available estimate of the real PT alpha value and not some weak analog of it, as in the Brutsaert model (i.e., Brutsaert et al., 2020), the latter taking up values much below the physically still interpretable value of one.

Response: We understand the temperature correction could change the magnitude of Ew, and will add this point in the discussion section. However, the essential difference between Brutsaert (2015) and Szilaygi et al. (2017) is the rescaling variable. Thus, the equation for alpha developed in Brutsaert et al. (2020) is unlikely to work for the calibration-free formulation of Szilagyi et al. (2017).

We agree that the alpha in the CR framework is an analog of the PT coefficient. However, the alpha still has the same physical meaning to that of the PT coefficient. It quantifies the proportion of the aerodynamic component of the Penman equation when the surface is with ample water. A potential reason for the alpha deviating from 1.26 would be the chosen equation for E_p. The traditional Penman equation could overrate E_p as shown in Yang et al. (2019). We will add this point in the discussion section.

Comment 8: Another model application issue is that in the CR model of this study Tws is estimated only one way, while the original authors of this CR model also described another method for the Tws estimation, yielding somewhat smaller Tws values (as mentioned in this MS, and therefore potentially resulting in a higher alpha-value estimate, most probably bringing the derived alpha value into the often quoted, typically observed 1.1 – 1.32 interval). In fact, most of this CR model's applications use the latter, so it would be worth to check how it affects model outcomes and the constant alpha value estimation.

Response: We disagree. The choice for Tws would exert very minor influences on the CR ET. As shown in Table 1 in Szilagyi et al. (2017), the choice for Tws led to small differences in the alpha within the order of 0.01 or 0.02.

Comment 9: Line 238: I do not think GPCC is a reanalysis precipitation data.

Response: It is a gridded gauge-analysis product. We will revise it.

Comment 10: Please check the text for the numerous typographical errors. For example: correctly ‘Priestley’.

Response: We will check all the typographical errors in revision.

References

Acharya, S. C., Nathan, R., Wang, Q. J., Su, C.-H., and Eizenberg, N.: An evaluation of daily precipitation from a regional atmospheric reanalysis over Australia, Hydrol. Earth Syst. Sci., 23, 3387–3403, https://doi.org/10.5194/hess-23-3387-2019, 2019.

Beesley, C. A., Frost, A. J., and Zajaczkowski, J.: A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets, 18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009.

Brutsaert, W.: A generalized complementary principle with physical constraints for land☐surface evaporation, Water Resour. Res., 51, 8087–8093, https://doi.org/10.1002/2015WR017720, 2015.

Glenn, E.P., Doody, T.M., Guerschman, J.P., Huete, A.R., King, E.A., McVicar, T.R., Van Dijk, A.I.J.M., Van Niel, T.G., Yebra, M. and Zhang, Y.: Actual evapotranspiration estimation by ground and remote sensing methods: the Australian experience, Hydrol. Process., 25, 4103–4116, https://doi.org/10.1002/hyp.8391, 2011.

Hobeichi, S., Abramowitz, G., Evans, J., and Beck, H. E.: Linear Optimal Runoff Aggregate (LORA): a global gridded synthesis runoff product, Hydrol. Earth Syst. Sci., 23, 851–870, https://doi.org/10.5194/hess-23-851-2019, 2019.

Kim, D., Ha, K.-J., and Yeo, J.-H.: New drought projections over East Asia using evapotranspiration deficits from the CMIP6 warming scenarios, Earth's Future, 9, e2020EF001697, https://doi.org/10.1029/2020EF001697, 2021.

Kim, D., Lee, W.☐S., Kim, S. T., and Chun, J. A.: Historical drought assessment over the contiguous United States using the generalized complementary principle of evapotranspiration, Water Resour. Res., 55, 6244–6267. https://doi.org/10.1029/2019WR024991, 2019b.

Ma, N., and Szilagyi, J.: The CR of evaporation: a calibration-free diagnostic and benchmarking tool for large-scale terrestrial evapotranspiration modeling, Water Resour. Res., 55, 7246–7274. https://doi.org/10.1029/2019WR024867, 2019.

Martens, B., Schumacher, D. L., Wouters, H., Muñoz-Sabater, J., Verhoest, N. E. C., and Miralles, D. G.: Evaluating the land-surface energy partitioning in ERA5, Geosci. Model Dev., 13, 4159–4181, https://doi.org/10.5194/gmd-13-4159-2020, 2020.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling, Hydrol. Earth Syst. Sci., 24, 1485–1509, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Szilagyi, J., Crago, R., and Qualls, R: A calibration☐free formulation of the complementary relationship of evaporation for continental☐scale hydrology. Journal of Geophysical Research: Atmospheres, 122, 264–278. https://doi.org/10.1002/2016JD025611, 2017.

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., Donohue, R. J.: Hydrologic implications of vegetation response to elevated CO2 in climate projections. Nature Clim. Change 9, 44–48, https://doi.org/10.1038/s41558-018-0361-0, 2019.

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2021-126/hess-2021-126-AC1-supplement.zip