

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2021-494-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-494

Anonymous Referee #2

Referee comment on "Reconstructing climate trends adds skills to seasonal reference crop evapotranspiration forecasting" by Qichun Yang et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-494-RC2>, 2021

1. General comments

This paper presents a method to improve monthly seasonal forecasts of potential evapotranspiration using a trend-aware statistical model (BJP-ti). This model builds on previous work by the authors on the BJP model combining a data transform with a multivariate normal distribution.

The topic of trend-aware forecasts is fundamental in a changing climate where the use of long historical time series to calibrate statistical forecast and post-processing models becomes questionable. This paper provides a valuable contribution to the field by showing how an existing statistical model can be extended to include trends with limited additional complexity. The model performance is thoroughly analysed using well established metrics that target a wide range of forecast attributes. Finally, the paper is well written, clear and to the point concise, with figures that provide strong visual evidence to support the authors' analysis.

We do not see any major issues with the paper and recommend it to be published with minor revisions. The two main items that could be improved by the authors aside of the detailed points raised in the following section relate to:

- [cross validation scheme] The authors used a traditional leave-one-out cross validation scheme where a single month is left aside for validation and the model is calibrated against the remaining data points. This an optimistic cross-validation scheme because the validation month is likely to show a similar trend and is not completely independent from the calibration data. A more conservative approach would be to split the data set in two parts, although this would not solve the problem completely. This an important issue but would require complex theoretical developments that are probably beyond the

scope of this paper. However, we recommend a bit of discussion around this point.

- [risk of overfitting when there is no observed trend] The authors demonstrate that the BJP-ti model outperforms BJP and raw forecasts when there is a trend in observed data. However, some of the results shown by the authors suggest that its performance is worse than BJP when the trends are not significant. This result is to be expected because of the higher number of parameters of BJP-ti which may increase the risk of overfitting and counter-performance over validation data. We recommend highlighting this point in the manuscript to better identify the strengths and weaknesses of BJP-ti.

2. Detailed comments

[Line 26] "Reference crop evapotranspiration (ET_o) measures the evaporative demand of the atmosphere": Please provide additional details regarding the definition of ET_o. We suggest the following: "Reference crop evapotranspiration (ET_o) measures the evaporative demand of the atmosphere for a hypothetical crop of given height, with defined surface resistance factor and albedo. It is generally computed using the Penman-Monteith equation following Allen et al. (1998, see section 2.1), which is known as FAO56. McMahon et al. (2013) provides additional information about the process. "

[Line 29] "In addition, ET_o forecasting also helps constrain the significant uncertainties in streamflow forecasting": please clarify that reference crop evapotranspiration (ET_o) is different from potential evapotranspiration used as input to rainfall-runoff modelling (PET). The sentence highlighted here can be confusing and suggest that both are interchangeable. See McMahon et al. (2013) for further comments on this point.

[Line 94] "we combine the archived re-forecasts and operational forecasts": Please comment briefly on the potential differences in skill between the re-forecast and operational data aside of the number of ensembles generated.

[Line 125] "trends in transformed forecasts and observations are removed to produce detrended data": This is quite an aggressive process because removing trend linearly in transform space, as described in equations 3 and 4, can lead to substantial reduction in untransformed space after a certain time. When trends parameters in BJP-Tri are significant (which seems frequent as suggested by Figure 1), we are a bit concerned that this could lead to forecasts becoming unrealistically large or systematically zero if left unchecked. We suggest commenting briefly on the time needed for the mean unconditional forecast (i.e. considering z_0 only in Equation 5) to depart from the unconditional forecast mean obtained at $t=t_m$ by more than, say, 50% in untransformed space. Perhaps consider showing the distribution of this time across the gridded domain and provide guidance on how frequently BJP-tri should be reviewed to monitor the accuracy.

[Line 132] " t_{δ} is approximately the middle year": does moving t_m has an impact on generated forecasts? I believe not because it is compensated by the value of the mean parameter μ . Please confirm. If this the case, please highlight that the position of t_m is

arbitrary and does not affect the forecasts.

[Line 153] "Equation 8 shows the conditional posterior distribution of parameter δ_{ij} ": We suggest "Equation 8 shows the posterior distribution of parameter δ_{ij} conditional on δ_{ii} ".

[Line 154] "In equation 8, δ_{ii} is the mean and δ_{ii} is the standard deviation for predictors or predictands.": Please move this sentence just after Equation 8. In addition, we suggest the following clarification: " δ_{ii} is the standard deviation for predictors or predictands extracted from the diagonal of covariance matrix S (see equation 5)".

[Line 160] "we adopt a leave-one-year-out cross-validation strategy": for a trend-aware model, this is an optimistic approach to model validation because the model has seen both past and future data during calibration. A more challenging validation would be to split the data in two parts, infer the trend from one part and validate on the other. We understand that this is challenging with a heavily parameterised model such a BJP, consequently it is probably beyond the scope of this paper to solve this question here. However, it is important to flag the potential issue of using traditional leave-out validation for trend analysis.

[Line 166] "The comparison is conducted for months with large areas of statistically significant (at the 95% confidence interval) temporal trends in observed ETo.": this approach is problematic because it does not check the performance of the BJP-ti model when there is no observed trend. BJP-ti is more parameterised than BJP, consequently it is always exposed to the risk of overfitting the data when there is no trend, i.e. when trend parameters cannot be calibrated reliably. Please comment on this point and justify why performance assessment excluded month with no significant observed trend.

[Line 197] " δ_{ij} is raw or calibrated forecasts of ETo (mm month⁻¹)": This is a deterministic metric, so we believe that $x(t)$ is the mean of raw or calibrated forecast. Please clarify.

[Line 221] "Observed ETo shows increasing trends in many parts of Australia in the three selected months": There is a significant body of literature related to trends in evapotranspiration related to climate change (McVicar et al., 2012). Please comment briefly on how this statement relates to current research in the field.

[Figure 1.] We suggest adding the standard deviation of annual ETo in the first column of figure 1 to highlight the significance of trend values. It is important to understand if the observed trends of 6 to 8 mm/decade reported below are large compared to climatological variance.

[line 271] "Slight decreases in r are also found in regions where the observed trends are not statistically significant.": This statement seems to support the comment made against line 166 suggesting that BJP-ti might suffer from over-parameterisation when observed trends are not significant. If confirmed, this is an important limitation of the model that should be highlighted more clearly.

[Figure 2.] We suggest adding in this figure a contour line showing the area where observed trend is not significant. This could help understand better the strength and weaknesses of BJP-ti.

[Line 277-285] Please also report the proportion of the study area where CRPS of BJP-ti is greater than the one of BJP. From Figure 3, it seems that BJP-ti underperforms in large parts of the domain, even if the decrease remains limited.

[Line 290] "with CRPS skill scores lower than -25% in all grid cells": this comparison is informative, but a little bit biased because raw operational forecasts are generally post-processed using techniques such as quantile-quantile mapping. We believe it is useful to show that raw forecasts have serious deficiency to reproduce on-ground observations, but it is also important to highlight that these forecasts would not normally be used for direct estimation of ET_0 . It would be perhaps more interesting to compare the correlation score between raw and BJP-ti forecasts, which discards some the known deficiencies of raw forecasts.

[Line 310] Same comment than for Line 290.

[Line 340] "We recommend that future GCM-based ET_0 forecasting should correct time-dependent errors": this comment should be toned down to include the risk of model overfitting discussed previously in relation to lines 166 and 271.

[Line 361] "Future work for seasonal ET_0 forecasting": We suggest adding the two challenges of model overfitting when there is no observed trend and validation of trend-aware forecast beyond leave-one-out approach.

3. References

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.

McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R., & McVicar, T. R. (2013). Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. *Hydrology and Earth System Sciences*, 17(4), 1331-1363.

McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, A., ... & Dinpashoh, Y. (2012). Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology*, 416, 182-205.