Response to Reviewer #3

General Comments：

R1. This is an interesting paper on an important topic. In fact it could be argued the issues discussed in this paper are absolutely fundamental in hydrology. This is particularly true in light of the way “data” is used uncritically in many studies.

Response: Thanks very much for your great efforts to assess our manuscript. We have studied your and reviewers’ comments carefully and will make corrections/revisions as suggested. In the following, we have detailed how these comments (in black) are raised and our responses (in deep sky blue).

R2. The paper is generally well written and structured, although there is a need for a more careful read-through as there is some awkward syntax and grammar (particularly in the Discussion) where the quality of the writing seems to wander a little. My main questions relate to the data set used and the essential paradigm of the PCM. The Redhill site did not have a calibration period. Although the authors suggest that the first period of treatment may be thought of as non-treated in that the trees were very small and not high water using, I feel the implications of this may be important. This then connects to the PCM paradigm; that is, that the length of calibration or the approach developing the calibrations in theory should account for the type of non-stationarity that is discussed (ie. drought). Putting this another way, how can we decide what is non-stationarity and what is variability? This is particularly germane to Australian hydrology where we experience significant variability. I am not suggesting climates are stationary, but disentangling non-stationarity from variability with a relatively short data period (in climate terms) is a question.

Response: Thanks for your constructive comments. We will add more contents about the calibration period, non-stationarity and variability in discussion section in the revised manuscript.

(1) We agree with the question of the calibration period. However, for Red Hill catchment, it is unfortunate that rainfall and runoff data before treatment were not measured. In Zhao et al. (2010), the impacts of the calibration period determined by treatment period and period before abrupt change point of annual runoff on the results at four paired catchments were compared, it was found that runoff changes caused by vegetation change are not sensitive to the different calibration periods. Moreover, considering that the runoff may not change significantly in the first few years after plantation of seedlings of *P. radiata*, we re-estimated the impact of vegetation change on runoff based on a calibration period with the data of the previous three years. The contribution of vegetation change to total runoff changes is 34.2%, and the difference with the result in the manuscript is only 1.4%. Therefore, the calibration period set in this manuscript is reasonable.
(2) In this study, we distinguish climate variability and non-stationarity by considering the influence on the rainfall-runoff relationship. For climate variability, we think that it will not lead to non-stationary changes in rainfall-runoff relationship, that is, rainfall and runoff changes at the same rate. For non-stationarity, it will lead to non-stationary changes in rainfall-runoff relationship, that is, it can be reflected by the significant abrupt change point of double mass curve and the

significant up/down movement of rainfall-runoff linear regression line (Avanzi et al., 2020; Li et al., 2018). For large-scale watersheds, it is really difficult to detect and confirm non-stationarity from variability because of the complexity and regional differences of positive and negative fluctuations or feedback of climate. However, for the two small studied catchments, the impact of climate fluctuations is very intense, and persistent fluctuations below the average are easy to cause non-stationary changes in rainfall-runoff relationship.

R3. The calibration period issue is a vexed one as there is no longer an appetite by funding bodies to set up a paired-catchment experiment and then wait for a lengthy period before anything happens. Bren and Lane (2014, JH 519) explored this issue and proposed a method using daily flows that rather obviously increases the number of data points. Somewhat surprisingly the analysis showed that good calibrations (Nash-Sutcliffe E = 0.8) with 100 days of data, and very little improvement after 3 years. Apologies for the treatise, but I wonder if this approach might be useful in thinking about the PCM. That is, if such an analysis was performed and compared with the other analyses it might be very useful. Data could be randomly pulled out of the Kileys Run data. At the very least it should be discussed.

Response: Thanks for your suggestion. We will add discussions regarding to your concerns in the revised manuscript.

We use the data of seven years before the abrupt change point of annual runoff and three years after the beginning of the data series to explore the change of Nash–Sutcliffe coefficient (N-S) with the increasing length of the observation set. For seven years (1990-1996), the calibration set is from 1990 to 1992 and the verification set is from 1993 to 1996 (Figure R1 (b) and (d)). For three years (1990-1992), the calibration set is from 1990 to 1992 and the verification set is from 1993 to 1996 (Figure R1 (a) and (c)). According to Figure R1, it can be found that the change of N-S is similar to Bren and Lane (2014). It shows that good calibration (N–S > 0.85) is achieved in about 150 days, and then maintains at the high-level value. Similar results are obtained with monthly of flows, calibration (N–S > 0.35) is achieved in about 24 months.

We use the data of the previous three years (1990-1992) as pre-treatment period, and re-estimate the effects of vegetation change on runoff by the PCM, TTM, SBM and new framework. Impacts of vegetation change were 34.2%, 74.2% and 61.0% of total runoff changes by the PCM, TTM and SBM, respectively. The contributions of vegetation change, multiyear drought and climate variability to total runoff changes are 34.2%, 37.4% and 39.0%, respectively, using the new framework. The sum of the three terms is 110.6%, with a difference of 10.6% from 100%. When the pre-treatment period is as pre-change period of runoff (1990-1996) used in the manuscript, impacts of vegetation change were 32.8%, 93.5% and 76.1% of total runoff changes by the PCM, TTM and SBM, respectively. The contributions of vegetation change, multiyear drought and climate variability to total runoff changes are 32.8%, 54.7% and 23.9%, respectively, using the new framework. The sum of the three terms is 111.4%, with a difference of 11.4% from 100% and a difference of 0.8% from 110.6%. It also can be found that the difference in the contribution of vegetation change to total runoff changes estimated by the two different ways to determine the calibration period is only 1.4%. It shows that taking period before the abrupt change point of annual runoff as the calibration period is also reasonable considering a lower mean error.

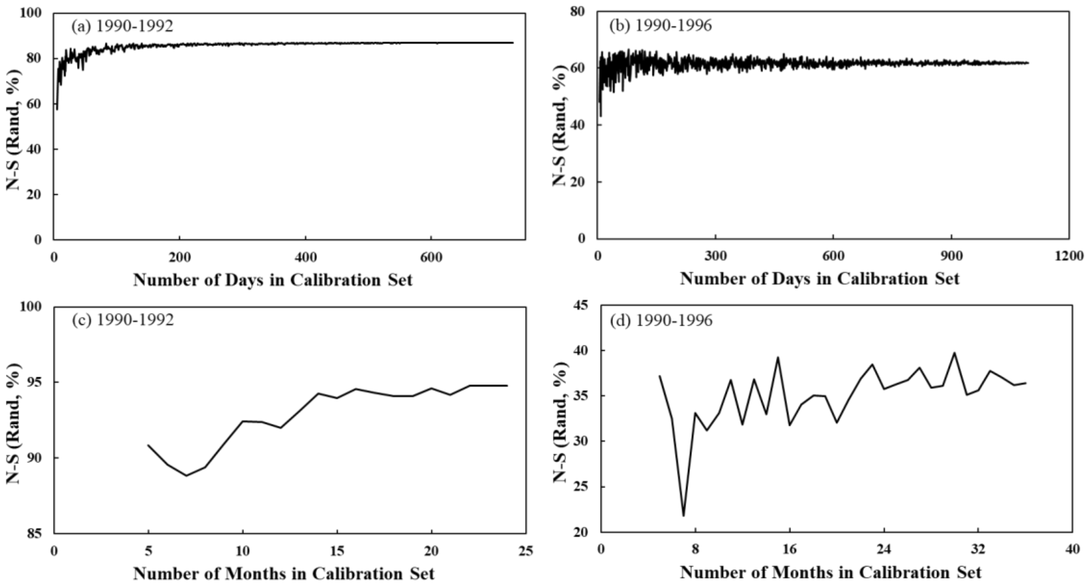

Figure R1. Mean of 10 values of the N–S coefficient of the Coefficient of Determination as a function of the number of days and months in the calibration data set for the Kileys Run and Red Hill data using the 10-fold cross-validation approach to monitor the development of the calibration. ‘‘Rand’’ refers to the data being randomized.

Specific Comments：

R4. Line 40 - there are more updated references for the research in Australian catchment behaviour that are relevant (eg Petersen et al, 2021, Science 372)

Response: Thanks for your advice. Changes will be made as suggested. (see references below).

King, A. D., Pitman, A. J., Henley, B. J., Ukkola, A. M., and Brown, J. R.: The role of climate variability in Australian drought, Nat. Clim. Change, 10, 177-179, https//doi.org/10.1038/s41558-020-0718-z, 2020.

Peterson, T. J., Saft, M., Peel, M. C., and John, A.: Watersheds may not recover from drought, Science, 372, 745-749, https//doi.org/10.1126/science.abd5085, 2021.

R5. L46. I think this statement about PCM and non-stationarity requires more justification. How does it not deal with the issue given that is the paradigm of PCM?

Response: Thanks for your constructive comments. We will add more justification about PCM and non-stationarity in the revised manuscript.

At Red Hill site, the treated catchment suffered from non-stationary changes caused by vegetation change and multiyear drought, and stationary changes caused by climate variability. The control catchment suffered from non-stationary changes only caused by multiyear drought, and stationary changes caused by climate variability. The PCM can offset the effects on both paired catchments induced by multiyear drought and climate variability based on data of the control catchment, and the effects of vegetation change on the treated catchment can then be separated. That is, for the paradigm of PCM, when the treated catchment is affected by N+1 kinds of non-stationary changes and the control catchment is also affected by the same N kinds of non-stationary changes, this method still can separate the effect of the N+1 non-stationary change that does not occur in the control catchment on the treated catchment. The PMC is still the most reliable method compared with other methods.

R6. L 65+ Both TTM and SBM require lengthy records; is this an issue with these analyses at Redhill? There may be an argument to see the drought as a plus in terms of a record with wet/dry periods.

Response: Thanks for your constructive comments. The discussion about this question will be mentioned in the revised manuscript.

(1) The length of data is essentially a problem of reaching an equilibrium state for catchments. Han et al. (2020) provided a global assessment of the steady-state assumption in catchment water balance calculations for 1,057 global unimpaired catchments. Results show that ~70% of the catchments attain steady state within 10 years. The time needed for a catchment to reach steady state shows a close relationship with climatic aridity and vegetation coverage, with arid/semiarid and sparsely vegetated catchments generally having a longer time. For small catchment, it may need a shorter time to reach steady state. It can be seen in Figure 4 (a), the double mass curve becomes a straight line in a very short time. It demonstrates that the length of data used in this study (26 years) is acceptable.
(2) The multiyear drought mentioned in the manuscript is a drought with longer duration and greater intensity, which will cause non-stationary changes in rainfall-runoff relationship of catchments. It is quite different from the wet/dry periods floating near the average line. The effects of multiyear drought on runoff are very significant and cannot be ignored. It is necessary to consider impacts of multiyear drought on runoff in the new framework.

R7. Lane et al. 2005 (JH) used FDCs that included Redhill – might be worth including these results as a comparison from a different method. This paper also has some estimates of time to equilibrium.

Response: Thanks for your advice. Changes will be made as suggested in section 5.2 and 5.4 in the new manuscript.

R8. L 69 – syntax not great "issues about this hypothesis" could be improved

Response: We apologize for these oversights. The lines will be modified to "However, this hypothesis has not been explored and verified, and it is important to examine whether non-stationarity induced by multiyear drought invalidate the PCM and the applicability of the three widely used methods under changing climate with frequent extremes in the future."

R9. L 107 – should be double mass curves, FDCs etc. There are quite a few examples of this, need a careful read.

Response: We apologize for these oversights. We will modify them in the revised manuscript.

R10. L 168- See earlier general discussion. It does trouble me that a site with no calibration is used for this study. In addition, the vegetation effect is dynamic; growing from seedlings to (presumably, given there are no growth data) a closed canopy. I do wonder if this really is the best data set for such a study, or what might be gained from using more data sets.

Response: Thanks for your constructive comments. Except for Red Hill site, there are no paired catchments that have experienced vegetation change and multiyear drought at the same time. The use of data and the calibration period are also further discussed in the responses to comments R2 and R3. It is found that the difference in the contribution of vegetation change to total runoff changes estimated by the two different ways to determine the calibration period (the previous three years after treatment and seven years before the abrupt change point of annual runoff in the manuscript) is only 1.4%. In this study, we not only find the reasons why the results of the three traditional methods are inconsistent, but also think more deeply about the paradigm of PCM. For PCM, when the treated catchment is affected by N+1 kinds of non-stationary changes and the control catchment is affected by the same N kinds of non-stationary changes, this method still can separate the effect of the N+1 non-stationary change that does not occur in the control catchment on the treated catchment. The PMC is still the most reliable method compared with other methods.

R11. Figure 3 is a great figure!

Response: Thanks for your comments.

R12. Table 2 – total flow changes would be useful. They appear later in the text but I think having totals in the table make it easier to evaluate the methods.

Response: Thanks for your advice. The total flow changes have been mentioned in Table 2.

R13. This also brings up another point that I don't think has been discussed properly. Q_{clim} is conceptualised as the climate effect, encompassing wet and dry and mean climate inputs. I am not sure there is adequate discussion of how this does not deal with the climate issue as formulated.

Response: Thanks for your constructive comments. The discussion about this question will be mentioned in the revised manuscript.

Q_{clim} in this study refers to the impact of climate variability (wet and dry spells) on runoff that does not lead to non-stationary changes in rainfall-runoff relationship. It represents the runoff changes caused by climate stability changes and it is estimated by the SBM. The SBM is derived from the Budyko framework. In the formula, the impact of climate change on runoff is estimated by rainfall and potential evapotranspiration changes. In the Budyko curve, it is reflected in moving from one point to another point on the same curve. There is a necessary assumption in the SBM, that is, a transition from one steady state to another with no change in catchment properties (the same curve) (Roderick and Farquhar, 2011; Sun et al., 2014). When the SBM is applied to Red Hill catchment that has experienced multiyear drought, the result is actual the impact of climate variability (without changing the catchment characters/non-stationary changes in rainfall-runoff relationship) on runoff, that is, $Q_{clim.}$ It may ignore the impact of multiyear drought on runoff and indirectly overestimate the impact of vegetation change on runoff. Because catchment properties of Red Hill catchment may have changed due to multiyear drought (Kinal and Stoneman, 2012; Peterson et al., 2021; Saft et al., 2016; van Dijk et al., 2013), which violates the assumptions of the SBM.

R14. 5.2.1 – this paragraph brings up the interesting point (that is the subject of the Saft/Peterson/Fowler etc studies); is it the climate that is non-stationary or is the processes (obviously driven by the climate).

Response: Thanks for your constructive comments. We think that climate change is a combination of non-stationary changes and process. The non-stationary change may be due to the sudden and drastic changes of human activities (explosive increase of industrial activities (Bauska et al., 2015)) and/or the amount of solar energy that gets to earth (Karl and Trenberth, 2003) and other factors. Most of these non-stationary changes occur suddenly and change from one equilibrium state to another through a relatively short period, such as non-stationary change in rainfall-runoff relationship caused by multiyear drought (Fowler et al., 2018; Peterson et al., 2021; Saft et al., 2015). The process is long-term, it may keep the equilibrium state and change continuously, or show a trend change (increase or decrease) over time. For example, the air temperature shows a gradual upward trend in a longer period (years and decades), and in a shorter period (days, months and seasons), the temperature constantly fluctuates around the average temperature, which means that there will still be those days which are cool and those days which are warm (Hoegh-Guldberg et al., 2019).

R15. The Paragraph around Line 375 needs some rewriting, the syntax is jarring. For example “the” control..

Response: We apologize for these oversights. Changes will be made as suggested in the revised manuscript.

R16. L 388 “Because Saft..” this is a poor sentence

Response: We apologize for these oversights. The lines will be modified to “Saft et al. (2016) re-evaluate a large range of factors suggested to be responsible for the additional runoff reductions. Results suggest that the shifts were mostly influenced by catchment characteristics related to pre-drought climate and soil and groundwater storage dynamics, but less affected by the percentage of woody cover.”

R17. L 399 “pines” should not be italicized. P.radiata would be

Response: We apologize for these oversights. The word “*pines*” will be modified to “pines”.

References:

Avanzi, F., Rungee, J., Maurer, T., Bales, R., Ma, Q., Glaser, S., and Conklin, M.: Climate elasticity of evapotranspiration shifts the water balance of Mediterranean climates during multi-year droughts, Hydrol. Earth Syst. Sci., 24, 4317-4337, https//doi.org/10.5194/hess-24-4317-2020, 2020.

Bauska, T. K., Joos, F., Mix, A. C., Roth, R., Ahn, J., and Brook, E. J.: Links between atmospheric carbon dioxide, the land carbon reservoir and climate over the past millennium, Nat. Geosci., 8, 383-387, https//doi.org/10.1038/ngeo2422, 2015.

Bren, L. J., and Lane, P. N. J.: Optimal development of calibration equations for paired catchment projects, J. Hydrol., 519, 720-731, https//doi.org/https://doi.org/10.1016/j.jhydrol.2014.07.059, 2014.

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., and Zhang, L.: Simulating Runoff Under Changing Climatic Conditions: A Framework for Model Improvement, Water Resour. Res., 54, 9812-9832, https//doi.org/10.1029/2018WR023989, 2018.

Han, J., Yang, Y., Roderick, M. L., McVicar, T. R., Yang, D., Zhang, S., and Beck, H. E.: Assessing the Steady-State Assumption in Water Balance Calculation Across Global Catchments, Water Resour. Res., 56, https//doi.org/10.1029/2020WR027392, 2020.

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Guillén, B. T., Bindi, M., Brown, S., Camilloni, I. A., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Hope, C. W., Payne, A. J., Pörtner, H. O., Seneviratne, S. I., Thomas, A., Warren, R., and Zhou, G.: The human imperative of stabilizing global climate change at 1.5°C, Science, 365, eaaw6974, https//doi.org/10.1126/science.aaw6974, 2019.

Karl, T. R., and Trenberth, K. E.: Modern Global Climate Change, Science, 302, 1719-1723, https//doi.org/10.1126/science.1090228, 2003.

Kinal, J., and Stoneman, G. L.: Disconnection of groundwater from surface water causes a fundamental change in hydrology in a forested catchment in south-western Australia, J. Hydrol., 472-473, 14-24, https//doi.org/10.1016/j.jhydrol.2012.09.013, 2012.

Li, Q., Wei, X., Zhang, M., Liu, W., Giles-Hansen, K., and Wang, Y.: The cumulative effects of forest disturbance and climate variability on streamflow components in a large forest-dominated watershed, J. Hydrol., 557, 448-459, https//doi.org/10.1016/j.jhydrol.2017.12.056, 2018.

Peterson, T. J., Saft, M., Peel, M. C., and John, A.: Watersheds may not recover from drought, Science, 372, 745-749, https//doi.org/10.1126/science.abd5085, 2021.

Roderick, M. L., and Farquhar, G. D.: A simple framework for relating variations in runoff to variations in climatic conditions and catchment properties, Water Resour. Res., 47, https//doi.org/10.1029/2010WR009826, 2011.

Saft, M., Peel, M. C., Western, A. W., and Zhang, L.: Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics, Water Resour. Res., 52, 9290-9305, https//doi.org/10.1002/2016WR019525, 2016.

Saft, M., Western, A., Zhang, L., Peel, M., and Potter, N.: The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective, Water Resour. Res., 51, https//doi.org/10.1002/2014WR015348, 2015.

Sun, Y., Tian, F., Yang, L., and Hu, H.: Exploring the spatial variability of contributions from climate variation and change in catchment properties to streamflow decrease in a mesoscale basin by three different methods, J. Hydrol., 508, 170-180, https//doi.org/https://doi.org/10.1016/j.jhydrol.2013.11.004, 2014.

van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, Water Resour. Res., 49, 1040-1057, https//doi.org/10.1002/wrcr.20123, 2013.

Zhao, F., Zhang, L., Xu, Z., and Scott, D. F.: Evaluation of methods for estimating the effects of vegetation change and climate variability on streamflow, Water Resour. Res., 46, https//doi.org/10.1029/2009WR007702, 2010.