We thank the reviewer for their helpful comments, our responses are below (in blue)

Although an interesting case study, the study provides limited justification and context, with some broad statements that should be better supported. How does this study inform the sustainable management of groundwater (from the opening line of the abstract)? The description of the study area does not mention groundwater use lower in the catchment, or reference to estimate of sustainable yields on a larger scale. A context of groundwater management issues in the region is not provided.

This was also raised by Reviewer #2. The study took place at catchments instrumented to study the effects of land-use change on groundwater and surface water resources, providing infrastructure and a data record needed for this investigation. The aims of our study were a better understanding of the uncertainties and limitations of commonly-applied recharge methods using these well-instrumented catchments as examples. As we explained, estimating recharge is important to understand the hydrogeology of semi-arid regions and this study has relevance to other regions where these techniques are used. Groundwater is not extensively used in this catchment, and our comments on sustainability were a more general recognition that groundwater can be a vital resource in semi-arid areas. In terms of the specific need to understand recharge in these catchments, defining whether recharge rates changed with land-use changes is important (especially the potential impacts on waterlogging and streamflow caused by changing water table elevations). We will clarify the aims and implications of the revised manuscript.

How do the authors reconcile their view of the importance of recharge estimates with the 'water budget myth'? A related myth that sustainable development of groundwater resources can be defined by groundwater residence times has recently been highlighted by Ferguson et al. 2020, citing classic papers on the water budget myth.

This is beyond what we can easily discuss in this paper. We agree that defining sustainable yields is difficult; however, understanding recharge is a critical part of assessing the impacts of groundwater use (e.g., Gleeson et al., 2012, Nature, 11295). The use of "residence times" also differs between this paper and Ferguson et al. (2020) study. We use it to refer to individual samples rather than the average age of water in the aquifer as a whole (sometimes called the turnover time). We avoid using the term "age" for specific samples as it is not a valid concept for most groundwater (Suckow et al., 2014, Applied Geochemistry, 50, 222–230). Given that the paper does not discuss sustainability in detail, we will reword the introduction and abstract to better reflect this.

The paper is well written and presented, although some additional figures to provide context and explanation would be helpful.

Specific suggestions are provided below.

1) The objectives of the study were to examine uncertainties in varying methods of estimating recharge. However, there is no discussion of how the method comparison is similar or distinct from other recharge studies in semi-arid areas. Have other studies also found the WTF method overestimates recharge for example?

We discussed this at lines 439-441 with reference to previous studies (Cartwright et al., 2007, Journal of Hydrology, 332, 69-92; Crosbie et al., 2010, Hydrology and Earth System Sciences, 14, 2023-2038; Dean et al., 2015, Hydrology and Earth System Sciences, 19, 1107–1123; Perveen, 2016, http://hdl.handle.net/1959.9/560005). We will reinforce this in the discussion and also make this point more generally in the introduction. Several of these studies (including the review of Crosbie et al., 2010, Hydrology and Earth System Sciences, 14, 2023-2038, and Crosbie et al., 2019, Water Resources Research, 55, 7343-7361) show that the WTF method generally overestimates recharge rates (lines 443-444), mainly due to specific yields being overestimated. We will discuss this more fully in the revised manuscript.

2) Comparing methods for recharge rates is interesting, but the authors argue (Line 481) that it is 'fundamentally important to assess the impacts of land clearing'. Why?

Inevitably understanding the impacts of land clearing necessitates using different methods to understand pre- and post-land clearing recharge. Pre-land clearing recharge is commonly estimated using the CMB or longer-lived radioisotopes (e.g., ^{14}C). Modern recharge may be estimated using the WTF method or Tritium. However, this assumes that the rates from those different methods are all broadly correct (or at least comparable), which is probably not the case. This is an important general point that we will emphasise in the introduction and discussion of the revised paper.

3) Section 5.1 on the impacts of reforestation only considers the TRR method, which surprisingly does not find a significant difference in recharge between pasture and forest. Other evidence indicates the forest is using more water, so the study appears to demonstrate the limitations of recharge estimation methods?

This is a good point. The area was partially replanted in the past 20 years. We discussed that the WTF method overestimates recharge rates and CMB method yields long-term recharge rates. ^{3}H activities and the TRR method are applicable to understanding the initial land use changes. We will specify this in the revised paper. There is also a scale issue. Most recharge rates based on groundwater samples from bores average recharge over areas of a few 10's to 100's m^2 (Scanlon et al., 2002, Hydrogeology Journal, 10, 18-39), and commonly the bores in plantation forests are in cleared areas that may be larger than this, possibly underestimating the recharge rates. This was discussed on lines 473-475. The TRR method has a time lag of several years (see Comment 10) that means that it may not yet detect the latest reforestation. We will clarify this in the revision.

4) How do the authors recommend these results inform groundwater modelling? Line 495.

Popular groundwater models, such as MODFLOW, use recharge rates as a boundary condition at the water table. Isotope methods give estimates of recharge rates across large areas over a relatively long period of times dating back hundreds of years. Because the WTF method estimates recharge rates at smaller spatial scales for the years when data are available, it is often considered to be more appropriate to constrain the models. However, the large discrepancies in the values of recharge rates from different techniques and the overestimated recharge rates with the WTF method pose questions on the quantification of boundary conditions for groundwater models. As suggested in our discussion, the use of integrated surface and subsurface hydrologic models might overcome this issue and provide an additional tool for the estimation of recharge rates to support experimental analyses. We will modify this statement in the manuscript to strengthen this part of the conclusions.

5) Both WTF and TRR rely on estimating the effective porosity (or effective specific yield). Mean porosity was previously reported as 0.15 and 0.1 respectively for the pasture and forested catchments but is unclear how this was determined, and how sensitive the WTF and TRR methods are to the range of possible values.

The values of porosity were taken from the previous study in this area (Adelana et al., 2014, Hydrological Processes, 29, 1630-1643). The effective porosity is unlikely to vary significantly and given the nature of the aquifer materials is probably in the range from 0.03 to 0.2 (Morris and Johnson, 1967, U.S. Geological Survey Water-Supply Paper 1839-D, 42p). The uncertainties in the recharge estimates from TRR and WTF (if the specific yield is assumed to be close to the effective porosity) correspond to the uncertainties in the effective porosity. In terms of the TRR method, this is probably less than the uncertainty in h (the height of the upper part of the aquifer in which mixing occurs). In terms of the WTF method, the other issue around specific yield is more important. We can clarify the source of the porosity estimates in Section 3.3 and discuss the uncertainties more fully in Section 4.5.3.

Line 385 states S_y is 'not well known' which is an understatement, as the parameter is highly uncertain. There is also a possibility of semi-confined conditions to develop at very shallow depths and that hydraulic loading could account for part of the water level response to rainfall.

We agree that S_y is highly uncertain, as was discussed in the paper. Given that it is rarely measured and is not an invariant property, it is not surprising that there are errors in the WTF method. This is discussed extensively in the paper (sections 1.1.2 and 5).

If semi-confined conditions developed near the surface, one would expect rapid increases in the levels of WT following large rainfall events. We do not see the fast response to rainfall events, and the changes in WT levels are seasonal. We will make that clear in the revised paper.

6) CMB method is most reliant on the assumption of the long-term rate of Cl delivery and can only be applied in catchments with negligible runoff and sedimentary Cl inputs. How are the results sensitive to 8% runoff measurement from the catchments?

Like most semi-arid catchments, our study sites have minor surface water outflows. Not accounting for the export of Cl in surface water can lead to recharge rates being overestimated using the CMB method (we noted this at lines 357-360). This has a little overall effect on the conclusions as the recharge rates estimated from the CMB method are still lower than from the other two methods, and we can further emphasise that point.

Moreover, the stream discharge has probably increased due to the initial land clearing as is commonly the case throughout southeast Australia (Alison, 1990, Journal of Hydrology, 119, 1-20) and the streamflow prior to land clearing would have most probably been much lower. Additionally, the stream water is saline and most of the water and Cl is probably derived from shallow groundwater discharge - again this is generally the case in these types of streams. Because this is not direct surface runoff, it does not impact the recharge rate estimates. These latter two points are of general importance, and we can discuss them briefly.

7) The limitations of lumped parameter models (LPMs) should be discussed, as the dimensionless ratios assumed to vary over a very wide range (e.g. 0.05 to 1). Are the estimated residence times linearly related to these lumped parameters? Also, can it be clarified why the PEM and DM lumped parameter models were applied and not the exponential-piston flow model?

The lumped parameter models are used for two purposes, to estimate residence times using ^{14}C and to examine mixing using ^{3}H and ^{14}C. In terms of mixing, similar ^{3}H vs. ^{14}C trends are apparent using lumped parameter models and other models that predict the concentrations of the radioisotopes (e.g., the renewal rate calculations; Leduc et al., 2000, Earth and Planetary Science, 330, 355-361; Le Gal La Salle et al., 2001, Journal of Hydrology, 254, 145-156). For the mean residence times, the different lumped parameter models do yield different MRTs. However, this approach is much better than the much-used decay equation that ignores both mixing in the aquifer and variations in the input function. The aims here were to broadly constrain the MRTs of the groundwater (i.e., to demonstrate that much of it is old). Other lumped parameter models (e.g., the EPM or the Gamma model) could have been used but yield MRTs that are similar to the ones that we report (lines 258-260). We will discuss the latter point briefly; however, it is not the main focus of the paper.

8) Clarify Line 295, regarding Cl/Br ratios 'and do not indicate that Cl is predominantly derived from rainfall and concentrated by evapotranspiration'.

Reviewer #2 also commented on this. This sentence was incorrect as written, it should have stated: "The observation that the Cl/Br ratios are significantly lower than those that would result from halite dissolution (10^4 to 10^5: Kloppmann et al., 2001, Geochimica et Cosmochimica Acta, 65, 4087-4101; Cartwright et al., 2004, Applied Geochemistry, 19, 1233-1254; Cartwright et al., 2006, Chemical Geology, 231, 38-56) and do not increase with increasing salinity indicates that Cl is predominantly derived from rainfall". We will correct it in the revised paper.

9) Schematic cross-sections could help explain the relationship between regional vs. riparian groundwater. An additional map that shows the regional catchment context of the catchment divides for groundwater vs. surface water would also be helpful, as the current mapping provides very large scale and small-scale maps.

We will add cross-sections to better show the context and the hydrogeology. The catchments are at the top of a major surface water divide and the groundwater divide will correspond to the surface water divide.

10) Mean residence times, estimated from both ^{3}H and ^{14}C, were ~4K in pasture and ~24K in the forest. Yet the forest was planted the only ~20 years ago, after ~160 years of pasture. The CMB method suggests chloride accumulation over ~10K years of rainfall inputs, to account for relatively high salinities. These differential time scales should be discussed further.

This is an important point, and we will discuss it further. The time taken for the ^{3}H activities to achieve steady-state is ~$1/R_N$ (i.e., the average residence time of the water in the well-mixed zone at the top of the aquifer). The initial land clearing should thus be evident in the TRR estimates, but the later revegetation may not be. Hence, in the cleared catchment, we would expect that WTF and TRR estimates agreed, but in the revegetated catchment, we may still be in the lag period where the ^{3}H activities are showing transient behaviour between different recharge rates. This may also explain the observations in Comment 3. We will discuss this more explicitly.