The paper is about climate impacts of hydrogen leakages, a timely and important topic. Even though the authors have good knowledge about previous studies and the introduction is well-written, unfortunately, I cannot recommend publication of this paper in ACP. I do not think this paper substantially contribute with new knowledge in the field and I do not think the results are discussed in an appropriate and balanced way. To me, this paper is a bit misleading, and it looks more like an opinion piece about their metric being much better than GWP100. The figures are a bit confusing, and I would argue, sometimes wrong. Figure 4 and 5 shows ECS (equilibrium, and not transient!) for each year, do the authors assume the climate reach equilibrium in an instant? Also, the paper only interprets already published data and the authors do not state this very clearly. I am also confused about how the authors compare the impacts of hydrogen emissions and CO2 emissions.

1. Not a balanced discussion on GWP100:
The authors state many, many places in the manuscript that using GWP100 is wrong (e.g. L10: “To date, hydrogen’s warming effects have been primarily characterized using the GWP-100 metric—which is misleading for short-lived gases, such as hydrogen, as it obscures impacts on shorter timescales.”), even though this is the standard and official metric by UNFCCC. No metric is perfect, but this is not unique to hydrogen, and I don’t think the authors address this in a well-balanced way, for instance see Boucher and Reddy, 2008 for a description on how to use GWP100 for SLCFs. Since the authors want to bring this discussion to the manuscript, I find it problematic that there is a lack of discussion of other more widely used alternative metrics such as GWP*, GTP or cGTP, the latter two of which are both suggested as suitable metrics for short-lived climate forcers by the IPCC (Forster et. al 2021). GWP100 has an advantage of being comparable to published literature and provides a measure for a time well after the effects of hydrogen have reached steady state. GWP20 will underestimate the long-term effects of CO2 but this is not mentioned at all. Shine et al. (2005) and Allen et al. (2016) also stress that short time horizon lead to overestimates of short lived climate forcers. This is because GWP is an integrated metric in contrast to end-point metrics such as GTP.
Given hydrogen’s known indirect greenhouse gas properties and unknown leak rates, we use a metric for looking at the impacts of energy transitions on net radiative forcing over time called Technology Warming Potential (Alvarez et al., 2012) that considers continuous emissions, providing a more realistic understanding of the climate impacts of fuel switching. I read this as the only possible metric for estimating climate impact of continuous emissions – GWP100 can also be estimated for continuous emissions.

As several previous studies have shown, relying on GWP-100 for understanding the importance of short-lived greenhouse gases relative to carbon dioxide is misleading (Alvarez et al., 2012; Ocko et al., 2017; Ocko and Hamburg, 2019). These ‘several’ papers only refers to papers from the authors themselves and I am not particularly blown away by this list of self-citations on the topic.

Given hydrogen’s short atmospheric lifetime of only a few years, reporting hydrogen’s potency in GWP-100 has limited value. One strategy for indicating the potency of short-lived climate pollutants is to report GWPs for two time horizons – one that conveys near-term impacts (most commonly 20-year time horizon) and one that conveys long-term impacts (100 years) (Ocko et al., 2017). GWPs always have a time horizon where integration stops, GWP20 will not then consider long-term effects of CO2. GWP20 will underestimate the long-term effects of CO2, but the authors do not mention this at all, I find this problematic.

However, assessing the impact of hydrogen through a pulse of emissions is also problematic. This is because continuous emissions are a better representation of actual hydrogen deployment. To better understand the climate effects of hydrogen over all timescales, one would need to consider the radiative effects of continuous emissions over time (Alvarez et al., 2012). But the authors have already converted the radiative effects calculated by Paulot et al 2021 and here they also use continuous emissions.

The benefit of the Technology Warming Potential method is that we can analyse climate impacts over multiple time periods of interest—in the near-, medium-, and long-term—insights that are not available with the use of the GWP-100 metric. This is important when short-lived climate pollutants are emitted as they are often reported and assessed based on the long-term impact of a pulse emission, which overlooks their true impacts during the time they are active in the atmosphere.

However, even the standard GWP-100 approach undervalues the cumulative radiative forcing over a 100-year time period given its reliance on pulse, instead of continuous, emissions (Fig. 3).

2. Figures are misleading
Figure 2: do the authors suggest that GWP_0 can be used? How did you estimate the numbers in Fig. 2? Do you assume the same lifetime for the indirect effects of hydrogen?
here? If that is the case, this figure is not correct.

I think Fig. 3 is very misleading as it shows a timeline with years after technology switch with cumulative radiative forcing and does not take lifetime into account if I understand this correctly.

Figure 4 is even more misleading as it shows a -timeline- of ECS, and not transient. You then assume that the climate has reached equilibrium in an instant? This comment also goes to Fig 5. I do not agree that these figures can be published.

3. Data
The authors are not very clear in their abstract that all the data they interpret is one data set from one model - and that data set is already published. It is totally fine to do that – but it must be stated clearly that this is an interpretation of already published data.

L347: “In the absence of models capable of interactively simulating the chemistry, radiation, and temperature responses in the full atmosphere to hydrogen emissions, we apply the simple approach used by Paulot et al. (2021) to approximate temperature responses to the three hydrogen demand scenarios discussed in Sect. 4.1.1.” Paulot et al 2021 did use a model capable of simulating chemistry and radiation response to hydrogen emissions, that is what their paper is about - the way this is written it seems like you are calculating these effects yourself? Again, this is repeated in L471: “To our knowledge, no model is currently capable of interactively simulating the chemistry, radiative forcings, and temperature impacts from hydrogen emissions into the full atmosphere.” Then I am left to wonder – why do you rely so heavily on the GFDL model from Paulot et al. 2021 in your analysis, if you don’t believe in the results?

Other comments sorted by line number:
L70: “.. approach as there are currently no formal models we are aware of that can simulate the full climate responses to hydrogen emissions”. What would it take to meet the criteria of this sentence and make one apply something else than this simple methodology?

L98: “..but the majority presenting results in terms of GWP-100”. Many of these also state radiative forcing from different sources in addition to GWP100.

L101: “.. the combination of GWP-100 downplaying hydrogen’s true potency and the recent insights into the full atmospheric”. I am not sure that downplaying is clear in this case, the long-term effect of excess carbon in the climate system should not be downplayed either? Also, which recent insights are we talking about here?
Here you should add that the lifetime of methane is much longer than hydrogen, and this will have an effect when integrating their effects.

“However, even a 20-year time horizon is long for a gas that only lasts a few years in the atmosphere”. Hydrogen has no direct climate effects; hence its own lifetime is not the only time value important to this type of consideration. One primary effect of hydrogen is the lengthening of the lifetime of methane. Methane’s lifetime is much longer than that of hydrogen, hence this sentence is somewhat ambiguous.

“When continuous emissions are considered as opposed to just one pulse at time = 0, the potency of hydrogen relative to carbon dioxide is on average double that of the pulse approach (Fig. 2); this is true for long-term effects as well” Are you comparing the appropriate things to each other here?

Table 1: I assume there is a typo here, and the carbon dioxide avoided from H2 consumed should be 11 as in the other columns.

“To determine emissions of methane when considering blue hydrogen production, we assume 3 times the mass of hydrogen is needed in the form of methane for using methane as a feedstock for hydrogen production (Budsberg et al., 2015).” One study - is there any uncertainties here?

“To estimate how much carbon dioxide emissions are avoided from deployment of one unit of hydrogen (which will ultimately depend on the specific technology), we use estimates from the Hydrogen Council (2017) that quantify avoided carbon dioxide emissions from a scenario of replacing 18% of final fossil fuel-derived energy demand in 2050 with hydrogen applications.” Hydrogen has a lot of indirect effects, as the authors also state, but so has CO2. How is that reflected in the numbers for emitted CO2 (11 kg CO2 avoided per 1 kg H2 consumed)? I think the decay functions need to be explained better. Also, how did the authors include the decay of methane into their GWP20 for hydrogen? The Methods needs to be better explained.

“.. based on their decay functions and radiative efficiencies”. Are these decay functions generally well-known for a gas such as H2 that has various indirect effects on climate?

“Methane and carbon dioxide radiative properties and atmospheric lifetimes are taken from Forster et al. (2021), but we do not include climate-carbon feedbacks associated with methane to be consistent with what is included with hydrogen.” In principle one could argue that this too should be compared to the results in the Paulot model to retain consistency.
L351: “The CMIP6 models suggest a best estimate of 3.78 ± 1.08 °C for the ECS and a 3.93 W m-2 effective radiative forcing for a doubling of CO2 (Forster et al., 2021). This suggests a climate efficacy of 0.96 °C (W m2 ) -1.” Would it be more concise to compare to the ECS of the GFDL model, and not to the CMIP6 ensemble at large?

L376: “The benefit of the Technology Warming Potential method is that we can analyse climate impacts over multiple time periods of interest—in the near-, medium-, and long-term—insights that are not available with the use of the GWP-100 metric. This is important when short-lived climate pollutants are emitted as they are often reported and assessed based on the long-term impact of a pulse emission, which overlooks their true impacts during the time they are active in the atmosphere.” As stated above; I am missing a thoroughly discussion of GWP and other possible metrics such as GWP* or cGTP which are more widely discussed in the literature. There is hardly any discussion about the uncertainties using their own metric, especially about how they compare the short lifetime of hydrogen and the long lifetime of CO2, which is a problem about this manuscript.