Anonymous Referee #1

We thank the anonymous referee for reviewing our work.

Can the authors explain why the two particular sites are chosen? Are they selected because of their measurement quality? Because they are the only sites available? Because they are at particularly important locations? What is the reason?

We will add the following text in the manuscript in Section 2.1:

> *"The Port Hacking and Maria Island sites were chosen for two reasons. (1) The sites are long term, containing over 70 years of temperature data and enabling long-term trend detection. (2) One site is close to the EAC separation region, while the other is further downstream in the EAC southern extension region. Hence, we can compare trends at locations with contrasting oceanographic conditions. At Port Hacking there are two long term sites (50m and 100m) but only the 100 m site has moored measurements hence we use data from the 100m site."*

I see a discussion about the accuracy of sensors that are on each of the observing platforms. But what about systematic biases?

It is challenging to explore systematic biases between the observing platforms as we require overlap of measurements in time and space. However, there are instances when this is possible.

For example, Satellite surface temperatures agree well with the nearest subsurface mooring measurements (Roughan *et al*., 2022, Malan *et al*., 2021), keeping in mind that there are sometimes differences related to the exact depth of measurements, stratification, and vertical mixing. Further, the satellite surface temperatures agree well with those retrieved by the surface buoys that have been temporarily in place over time at Port Hacking and Maria Island (Figure 1).

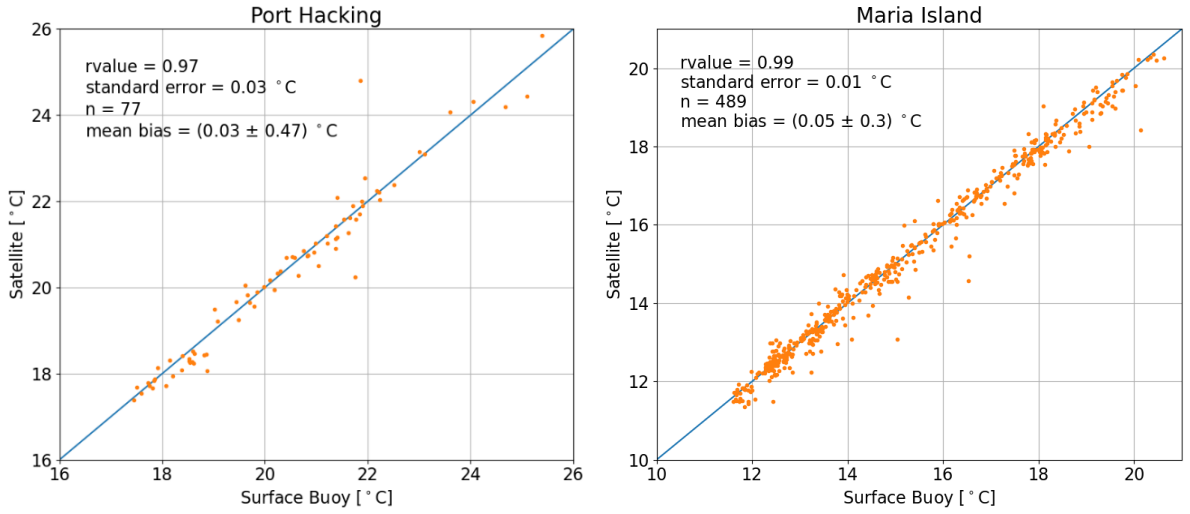

Figure 1: Surface temperature measured at the Port Hacking and Maria Island mooring sites compared with surface satellite temperatures. A 1:1 line is also present coloured blue.

We can also compare CTD profiles (typically collected monthly) with mooring measurements close in time and space at the Port Hacking and Maria Island sites, shown in Figure 2 below.

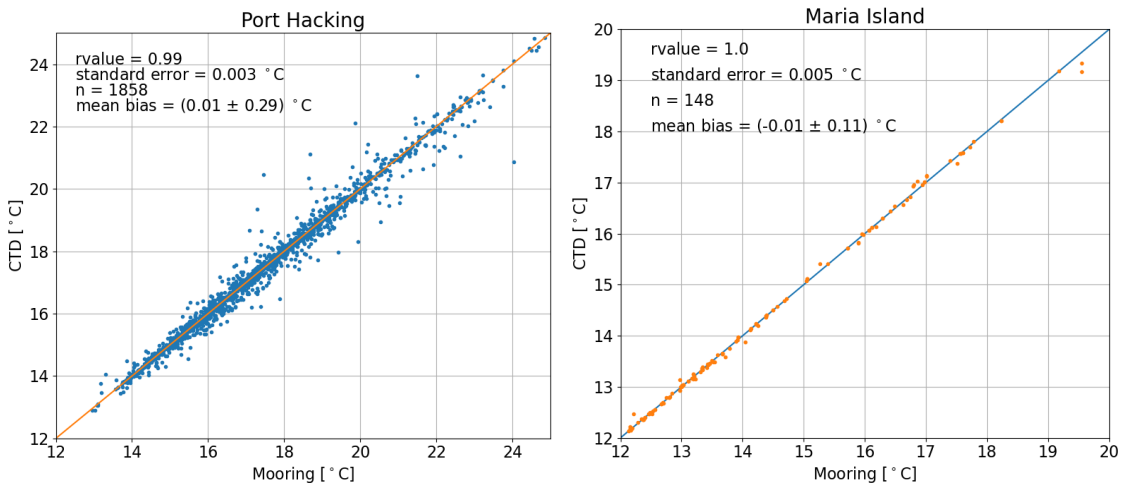

Figure 2: Comparing mooring temperature measurements at Port Hacking and Maria Island within 10 minutes of CTD profiles collected nearby between 2009 and 2022. A 1:1 line is also present.

There is good agreement between mooring measurements and the CTD profiles, with a mean bias of 0.01 (negative at Maria Island).

We have also compared the bottle samples with CTD profiles at the Port Hacking site when there is spatial and temporal overlap (Figure 3).

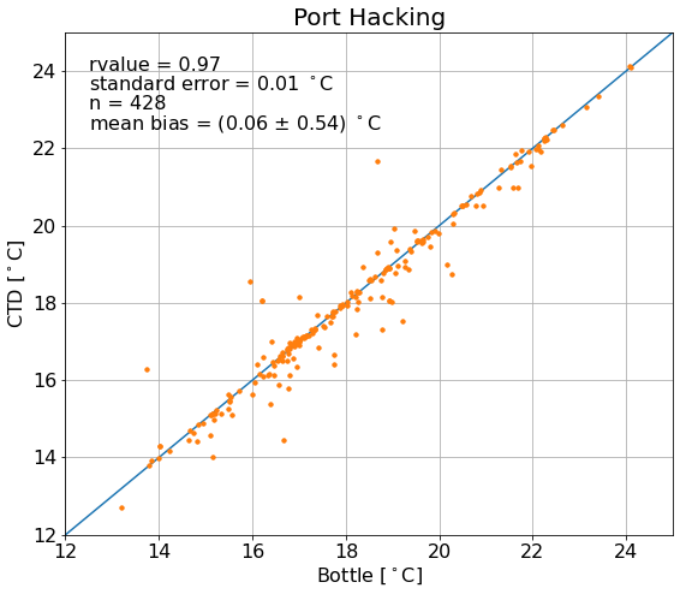

Figure 3: Comparing bottle samples with CTD profiles when available at a similar time (< 12 hours) and location (typically < 5 km) between 1989 and 2004.

These comparisons do not indicate any major biases between the different data sets. We will include the following text in Section 2.1 relating to the referee's comment:

"Where possible, temperatures observed using the various data platforms close in time and space to one another were matched and compared, and our analysis did not indicate any major systematic biases."

I don't see a reference or a discussion of figure 3 prior to the appearance of the figure. Perhaps I missed it?

We reference Figure 3 alongside Figure 4 on L155, Section 3.1. It is here where we also discuss the results shown in Figure 3.

In the second paragraph in Section 3.1, what I think are surprising results are discussed. That the warming rates at deeper layers exceed those at shallower layers. This behaviour is discussed in depth in Section 4.1 with plausible physical mechanisms.

Did you mention any QC that is used on these measurements? If that was done, I missed it.

We did not specifically mention data set QC, however the QC was the same through the water column. Rather, we cited Roughan et al., (2022) several times, and this work includes details on QC. But we will add the following sentences to the manuscript in Section 2.1:

"All temperature data used in this study have been quality controlled. The historical bottle data were initially quality controlled by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) with evolving practice over time, while IMOS CTD and mooring data collected since 2008/2009 were quality controlled using standardised IMOS procedures (Ingleton et al., 2014; Lara-Lopez et al. 2017; AODN, 2023). For satellite data, quality level flags >= 4 were used to select temperature data. Additional QC checks were performed for all data sets, as described by Hemming et al., (2020), and further information on data quality control is described by Roughan et al., (2022)."

(Ingleton *et al*., 2014): https://s3-ap-southeast-2.amazonaws.com/content.aodn.org.au/Documents/IMOS/Facilities/national_mooring/BGCManual_supplementary_documentation/IMOS_NRS_BGC%20Standardised%20Profiling%20CTD%20Data%20Processing%20Procedures_v2.0_Mar2014.pdf

(AODN, 2023): https://github.com/aodn/imos-toolbox/wiki/PPRoutines

(Lara-Lopez *et al*., 2017): https://repository.oceanbestpractices.org/handle/11329/568

I also wonder how valuable it is to compare against quite old studies (Bindoff 1997 for example).

We had included Holbrook & Bindoff (1997) as it is the only inter-decadal temperature trend off southeastern Australia, to the best of our knowledge, that uses a long-term data set (1955 – 1988) from below the surface. We detect a temperature trend between 1953 and 2022 that is very similar to the one described by Holbrook & Bindoff (1997), suggesting that temperature has increased steadily at depth. We therefore believe this reference is valuable to our discussion.

We also think that it is valuable to compare historic trends over a range of time periods to demonstrate the effect of time period length on temperature trends. We discuss this in Section 4.3.

I would think that limiting to studies done in the past decade or so would be wise. Perhaps they are not available?

To our knowledge we have included all recent studies at the time of submission.