This is an interesting paper using a novel approach to quantify global ocean O2 content seasonal to decadal-scale (S2D) variability and trends. The authors use AI and ML in an effort to resolve global and regional ocean S2D O2 variability and trends.The authors combine/aggregate contemporary ship-based Winkler-based O2 data used in GLODAP and sensor-based O2 data from BCG-ARGO.

The authors indicate that the spatial and temporal heterogeneity coverage of the observational O2 data that they chose to use might not be representativeness of S2D variability and trends.They argue that because GOBAI-O2 has no data gaps in time and space (gridded fields), it is more representative of real O2 ocean S2D variability and trends than the observations themselves. The authors also suggest that GOBAI-O2 represents the global ocean O2 mean than other gridding mapping methods (i.e., WOA18-O2).

My concern is that the authors do not quantify (metrics) why GOBAI-O2 is more representative or has greater ability/skill to represent the real ocean O2 mean and S2D variability than the observations they chose to use. What if they had used much additional QC observed O2 profile data coverage from other sources? The paper would benefit from using an objective metric comparison approach. For example, comparing GOBAI-O2's to other mapping methods using the same starting baseline O2 data..

The authors compare GOBAI-O2 to WOA2018-O2 as well as to selected GLODAP sections. I would be surprised to not see differences between these data products. For example, WOA2018-O2 mean climatology is based on a much larger pool of QCed winkler O2 measurements collected over 50 years (1965-2017; about 0.9 million profiles) than the QCed Winkler+BCG-ARGO O2 used by GOBAI-O2 (2004-2021). The authors could also compare GOBAI-O2 to the GLODAPv2 gridded O2 fields. In the end, these comparisons do not resolve GOBAI-O2's ability (metrics) to represent variability and trends better than observations and/or other mapping methods.

Finally, it would be useful if scientists could independently reproduce the GOBAI-O2 results. Are the authors planning on openly sharing the exact data (obs and model) and algorithms used?

Specific line comments and suggestions for consideration

For simplicity, I sometimes use "model" to refer to GOBAI-O2

38. What is the quantifiable metric for indicating that GOBAI-O2 provides a better representation of the real global and/or regional deoxygenation variability and trends than could be estimated from the observations themselves? Please clarify.

63. "have substantially improved the accuracy and reproducibility of optode-based [O2] measurements on Argo floats. In the absence of a reference (i.e., a true known value, a community-adopted certified reference material, or science community consensus reference

data), it is difficult to assess the “accuracy” of O2 field measurements (winkler and sensor based data). Suggestion: “have substantially reduced the uncertainty (or increased the precision?) and reproducibility of optode-based [O2] measurements on Argo floats”

82. GLODAP measurements were largely collected during summer and spaced several years apart. Is the model output biased towards the more abundant ARGO O2 data coverage (Fig 1)?

Combining O2 data measured by Winkler and sensor based is not as straightforward as merging them together. Did the authors conduct preliminary QC checks on the BCG-ARGO O2 for internal data consistency with co-located discrete GLODAP data?

226. What is the uncertainty in deoxygenation content variability as a function of time (assumed constant)?

270. Table 2 has no units. I assume O2 in umol/kg

Fig 2a,b. These figures suggest an envelope of Δ[O2] roughly +- 10-20 umol/kg for relatively higher freq. Is the GOBAI-O2 total uncertainty adequate to resolve decadal-scale deoxygenation trends? In section 3.2.3 Interannual oxygen variability, the authors indicate a relatively small global decadal trend of −1.15 ± 0.26 μmol/kg/decade. Global deoxygenation trends range between 0.6% for models to 2% for observations (Fig 2 in Grégoire et al. 2021; https://doi.org/10.3389/fmars.2021.724913).

Fig 2c, f. Coastal and other oceanic regions have high seasonal to interannual variability. Why are Δ[O2] so small near coasts when compared to the subtropics/tropics?

336. “demonstrates an ability”; ability is a subjective term. Is this ability quantifiable?

337-338: ”This bodes well for the ability of GOBAI-O2, which is trained on actual observational data, to represent decadal scale and seasonal variability in global ocean oxygen in the real world”

What quantifiable metric is being used to indicate that GOBAI-O2 represents the decadal scale and seasonal variability in global ocean oxygen in the real world?

As stated earlier, a large fraction of the ARGO O2 obs were collected in the S. Hemisphere (Fig 2c) and measurements in GLODAP were mostly collected in summer. Global and regional seasonal variability would arguably be difficult to quantify with certainty with a limited observational coverage as used in this case.

I note that in line 345, the authors write ““For example, large Δ[O2] values in the eastern tropical Pacific and Atlantic, coupled with negative correlations in annual mean [O2] and large differences in annual trends and seasonal amplitudes, suggest more observations will be required for GOBAI-O2 to capture variability in that region”

355. I note that in ice-covered regions, there is also little air-sea gas exchange and limited biologically-mediated O2 production adding to undersaturation; particularly in the S. Ocean.

380. “Oxygen concentrations are extremely low in the deep, high-density North Pacific Ocean and North Indian Ocean due to the ages of those water masses' Rather than age specifically, what matters is the net balance of sources and sinks (i.e., air-sea exchange, ventilation/mixing, O2 respiration, redox chemistry).

Fig 5. Is GOBAI-O2 trained using isobars (depth) and isopycnals independently?

Fig 7. Are the model O2 values de-seasoned before depth integration by layers (i.e., subtracting the climatological monthly mean O2 in addition to the long-term mean)? If not, why not?

412: Suggest changing “The spatially weighted rate of deoxygenation in..” to “The spatially weighted decadal rate of deoxygenation in..”

Fig 7d shows that the temperature anomalies below about 500 m are relatively smaller prior to about 2015 than in later years. On the other hand, fig 7e shows relatively high (absolute value) O2 anomalies before and after about 2015 and at all depths and reflected in fig 7a. Is the implication that this is due to mean changes in ventilation to deeper depths?

443. The deoxygenation trends (discussed in 3.2.3 Interannual oxygen variability) seem to be in the 0.5-0.9% range. These trends are in agreement with AR5 model trend estimates (about 0.6%, Bopp et al., 2013). Schmidtko et al. (2017) indicated a global ocean deoxygenation trend of about 2% (See Fig 2 in Grégoire et al. 2021; https://doi.org/10.3389/fmars.2021.724913). Please address this apparent discrepancy.

455. Why do the authors attribute all model (algorithm) variability to natural and/or anthropogenic variability? As shown in Fig 8, model uncertainty is not insignificant..

460. Averaged globally, total uncertainty is 6 umol/kg (line 466). Visual inspection of Fig 8 suggests oceanic regions with total uncertainty values approximately > 10-20 umol/kg.These appear to be due to regional differences in the skill of the algorithm (line 485). Given these regional uncertainties, what would the magnitude of error bars be in Fig 7 for O2 (net anomalies of < 3 umol/kg)?

Fig 8 has no units (umol/kg?). I am surprised to see relatively low uncertainty values along coasts and WBCs where O2 seasonal variability is nominally large and obscures interannual and longer time-scale variability. Why is the algorithm uncertainty largest near the eastern tropical Pacific and Atlantic?

Fig 9. Differences are not unexpected. GOBAI-O2 (2004-2021; Winkler+ARGO O2 sensor) uses a smaller spatial and temporal data coverage than WOA18-O2 (1960-2017; Winkler only). I would argue that an objective comparison would be to compare GOBAI-O2 and other mapping methods including the gridded fields of GLODAP and WOA18-O2.

It is interesting to see that WOA18-O2 minus GOBAI-O2 largest differences seem to follow isopycnals in the N and S. Pacific (F9b) and in the S. Atlantic (F9e). Is this a real feature or an artifact? Comparing GLODAPv2 gridded fields minus GOBAI-O2 would be useful.

510. The authors compare GOBAI-O2 to WOA18-O2; with GOBAI-O2 being about 10 µmol/kg lower than WOA18-O2. GLODAP includes a gridded mean O2 climatology. The authors should also compare GOBAI-O2 to the GLODAP gridded fields. Are the authors indicating that GOBAI-O2 provides a more accurate representation of the global ocean long-term O2 mean than WOA18-O2 and/or other data products? Please elaborate. The GOBAI-O2 global mean total uncertainty as a function of depth is about 4-10 umol/kg (Fig A10). Suggest adding some form of error bars at each depth in Fig A10 (i.e., std, serror, other).

511. WOA18-O2 uses O2 data starting in 1965; not 1955.

513. "... the World Ocean Atlas has been demonstrated to overestimate [O2] in suboxic zones (Bianchi et al., 2012)". Bianchi et al. indicated deviations of about 6 umol/kg in suboxic areas when compared to discrete O2 data profiles in GLODAP (Key et al. 2004). It is not unexpected that a mean O2 climatology spanning 1955-2004 would not exactly represent selected discrete O2 values. Similarly, I would not expect that other mapping techniques such as GLODAP O2 gridded fields exactly match all the discrete O2 data/profiles at any given depth/grid location.The same reasoning applies to GOBAI-O2. For example, Fig 10 shows O2 > 15 umol/kg differences between GOBAI-O2 and O2 values from GLODAP transects in the top 1 km.

GOBAI-O2 uncertainties seem larger than open-ocean O2 observing systems. GOOS Panel-Biogeochemistry-01-EOV-Oxygen Essential Ocean Variables (EOV) version 2.0 (August, 2017) provides uncertainty estimates (ARGO O2: ±2 umol/kg; Bottle Winkler ±0.5 µmol/kg). The figures are improving over time.

https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=17473
https://oceanexpert.org/downloadFile/35904