Reply to reviewer 1

(our response **in bold font**)

We thank the reviewer for the helpful comments. Actually, for trends computations the effective degrees of freedom of N*=N/T was already used in our matlab routine for the submitted version, however we did not describe it in the text. We edited the text accordingly and now describe in detail how the degrees of freedom are derived. The autocorrelation of each time series is used to determine the statistical independence of measurements in successive years and the half folding width is used to determine the degrees of freedom of the timeseries. The routine used is included at the end of this text.

GENERAL ASSESSMENT

Dr Lothar Stramma and coauthors compiled a mostly public dataset of oxygen, nitrate, silicate and phosphate observations from eight areas of the Pacific Ocean. Using those data, they calculated long-term trends over the entire length of the time series. They also calculated trends over the 1950 to 1976 negative phase of the Pacific Decadal Oscillation (PDO), and after 1976 for the positive phase of the PDO. They showed that in some cases, trend estimates were very different depending on sign of PDO phase.

In addition, they calculated lagged correlation coefficients between oxygen, nutrient and temperature time series from their eight areas with the PDO and with the North Pacific Gyre Oscillation (NPGO) climate indices. They also briefly examined how three other modes of climate variability (ENSO, STC, NPI) may or may not affect the time series in the same eight areas.

This is a worthwhile study. It is important that we better understand "The influence of decadal oscillations on the oxygen and nutrient trends in the Pacific Ocean". But there is a big caveat: the word DECADAL in the paper's title aptly captures the idea that successive years are not statistically independent of each other. For instance, when a given year is above normal for the PDO, the odds of having a below normal value of the PDO the following year is actually lower than 0.5; the dice are rigged. Several consecutive years with above normal values tend to be lumped together, and likewise for below normal values. From the point of view of statistical analysis, this implies the assumption that successive annual values in the oxygen, nutrients and temperature time series are independent of each other is not valid. Consequently, the effective number of degrees of freedom N* will often be much less than the number of years N, so that 95% confidence intervals will be broader than those assumed in the submitted paper, and some trends (Tables 1 and S1) that are considered statistically different from zero at the 0.05 level in the paper probably are not.

Given that trend estimation and determination of their confidence intervals is a key component of this paper, I consider a major revision of the paper will be required to take serial correlation (autocorrelation) of time series into account.

As explained above we had already used the effective degrees of freedom and mention it now more clearly in the text and we changed the serial correlation, as described also below. For the area P e.g. the degrees of freedom for oxygen trends were 13.8 for 23 available years for the period 1954- 1976, 33.9 for 40 years 1977 to 2017 and 50.8 for 63 years 1954-2017. The degrees of freedom depend strongly on the data set and time series analyzed. The method is now described in the text.

MAJOR COMMENTS

The "Data Processing" section of the paper is extremely succinct about the methods used for the computation of trends (p. 9, lines 3-6), the computation of correlation coefficients (p.10, lines 1-3) and their confidence intervals and significance levels. Upon examining the trends and correlation coefficients considered statistically different from zero in the text and Tables, I came to the conclusion that the authors used statistical methods that assume iid (independent identically distributed) variables. This is a major issue, because neglecting autocorrelation (aka serial correlation) in the time series can invalidate the levels of statistical significance and confidence intervals that are found throughout the text and in the Tables of results (von Storch and Zwiers 1999).

The Data processing explanation for the trend was extended with the information on the use of effective number of degrees of freedom for the computation of the confidence intervals, as well as the p-values for the correlations.

The authors did not perform a runs test for randomness to verify the underlying assumption that successive yearly values in their time series are independent of each other. Using the links provided by the authors, I downloaded yearly values (1950- 2017) for the PDO index, and I downloaded monthly values for the NPGO index from which I then computed yearly averaged values from 1950 to 2017. Using a runs test for randomness on the PDO time series, the null hypothesis that the yearly values of the PDO index come in random order is rejected ($p = 2.5e-08$). Likewise, using a runs test for randomness on the NPGO time series, the null hypothesis that the yearly values of the NPGO index come in random order is also rejected ($p = 0.0053$). Given this, it appears that neglecting to take into account autocorrelation in the statistical analyses of trends (for individual time series) and correlation coefficients (for paired time series) is an important flaw in a paper whose intent is to focus on the influence of decadal climate variability on yearly oxygen and nutrient time series. But luckily this flaw can be fixed, as methods that account for autocorrelation by adjusting the number of effective degrees of freedom do exist. For trends calculations, Thomson and Emery (2014) propose dividing the length of the time series N by an integral time scale T in order to obtain an effective number of degrees of freedom $N^* = N/T$. Using their method for estimating an integral time scale by including lags of up to plus or minus 10 years for example, I obtained an integral time scale for the PDO time series of 2.8 years. This implies that on average, we get independent values of the PDO index every 3 years or so.

Autocorrelation of time series is not only a problem for trends estimation. It also affects confidence intervals for correlation coefficients of paired time series. When we have N pairs of (x,y) values, the effective number of

degrees of freedom N* is not equal to N-2 as assumed by the authors, but is generally smaller. A resampling scheme that reduces the impact of serial correlation on inferences made about the correlation coef-
ficient is described by Ebisuzaki (1997), who also provides a review of other techniques that correct for autocorrelation in the estimation of confidence intervals for correlation coefficients.

As mentioned above we made the trend computations using the effective degrees of freedom N*=N/T, however it was not described in the text and now a statement is added in the revised version. Successive years are not assumed to be statistically independent except the autocorrelation function does indicate so. Never the less, with short and non continuous time series, methods and statistical analysis is limited to a degree.

MINOR COMMENTS

1. p.6, line3-4; and 5 CONSECUTIVE months of at least -0.5_C are defined as La Niña Events

'CONSECUTIVE' added

2. p.6, line 16; the discarding OF already sparse data

'OF' added

3. p.8, lines 14-16; the squares and circles are difficult to tell apart from each other, possibly because they are too small. Using circles and triangles (instead of squares) might provide a better visual contrast. This comment about squares and circles also applies to figures 3, 4, 5, 7, S2, S3, S4, S5.

Triangles were also difficult to separate from circles and did not cover the entire crosses, therefore me modified all figures by showing the circles in magenta and the squares in blue.

4. p.9, line 22; Do the spatial scales of 50 m and 100 m apply to both the horizontal and vertical directions? If not, the word "spatial" should be replaced with either vertical or horizontal.

"spatial" was replaced by "vertical"

5. p.15, line 20; replace thEn with thAn

thEn replaced by thAn

6. p.17, lines 25-27; A correlation is significant at r = -0.27, and another is not significant at r = 0.25. But one suspects that the p-value associated with r = -0.27 is barely below 0.05 whereas the p-value associated with r = 0.25 is barely above 0.05. A growing number of scientists insist that giving the actual p-values in the text and in tables is actually preferable to the dichotomy of dividing them in two categories: significant or not significant. See McShane et al. (2019) and other papers in a special issue of The American Statistician entirely dedicated to "A world beyond p < 0.05". I would love to see more p-values throughout the text and tables of this paper.

The p-values were added to table 2 and statistical significance based on p-values were removed in the text. The references to the papers by McShane et al. 2019 and Amrhein et al. 2019 are referenced to explain why no significance statement is given.

7. p.29, line 4; "Trends not within the 95% confidence interval are shown in italics" probably does not convey the intended message. Based on the table, I am guessing the intent was to write "Trends whose 95% confidence interval includes zero are shown in italics", or something like that.

As proposed we modified the text in Table 1 and Table S1 to "Trends whose 95% confidence interval includes zero are shown in italics"

8. p.31; Table 2; Add 2 columns with p-values, one for the PDO-correlation, and the other for the NPGO-correlation. The effective number of degrees of freedom N*-2 (not N-2) should be used to estimate p-values.

p-values are now added in table 2.

9. Figure 6; add the label "Lag (years)" to the the x-axis of the left and right panels.

"Lag (year)"was added to both frames in Figure 6.

REFERENCES

Ebisuzaki, W., 1997. A method to estimate the statistical significance of a correlation when the data are serially correlated. Journal of Climate, 10(9), 2147-2153.

McShane, B.B., Gal, D., Gelman, A., Robert, C. & Tackett, J.L., 2019. Abandon Statistical Significance. The American Statistician, 73:sup1, 235-245.

Von Storch, H. & Zwiers, F. W., 1999. Statistical analysis in climate research. Cambridge University Press, 494p.

Thomson, R. E., & Emery, W. J., 2014. Data analysis methods in physical oceanography. Newnes, 716p. See section 3.15.2 on "Trend estimates and the integral timescale".

The reference McShane et al. 2019 and in addition Amrhein et al. 2019 were included.

Matlab routine used for trend computations:

```
% version for oxygen in micromoles/kg
q=load('wiwi2.dat');
t=q(:,1)'; tm=mean(t); t=t-tm;
y=q(:,2)';
E=[t',ones(size(t))'];
covmat=inv(E'*E);
model=covmat*E'*y';
res=E*model-y';
c_res=xcorr(res,'coeff');
ii=ceil(length(c_res)/2):length(c_res);
tscale=2*max(cumtrapz([0:length(ii)-1],c_res(ii)));
% next line introduced 3 May 2010
tscale = max([tscale,1]);
degfree=(size(E,1)/tscale-size(E,2));
chisqr=sum(res.^2)/degfree;
model_err=sqrt(diag(covmat)*chisqr);
```

```
y_model=t*model(1)+model(2);
y_model_err_95=student(degfree)*sqrt(model_err(2)^2+(t*model_err(1)).^2);
plot(t+tm,y,'r.',t+tm,y_model,'b-',t+tm,y_model+y_model_err_95,'m--',t+tm,y_model-y_model_err_95,'m--')
disp(['The linear trend is ',num2str(model(1),3),' micromoles/kg/yr'])
disp(['The formal 95% confidence intervals for that trend are +/- ',num2str(model_err(1)*student(degfree),3),'
micromoles/kg/yr'])
% next line introduced 30 April 2010
dof=(size(E,1)-size(E,2))/tscale
```