

From: John Fasullo <fasullos@gmail.com>
Subject: Re: Collaboration on ENSO variance modulation
Date: September 10, 2020 at 9:44:30 AM EDT
To: Ben Goldman <bg502257@live.wpcsd.k12.ny.us>

Great - along with that we'd like to finalize the figures as much as possible to tell the storyline clearly. Have a look at some of the figures in papers published in GRL journals. Note the font size, line thicknesses, and layout. We'd like to adjust our figures to be more like the best ones you can find. It can be a tedious process but one that will pay dividends in the long run. Reviewers also appreciate it as it helps getting a manuscript accepted. Once your code is set up you can have a single routine that sets up your figure attributes that you reference from other routines. I'm not sure how to do this in python, in IDL it is just an independent routine that passes settings. NCL also has a nice gallery of (mostly) publication quality figures: <https://www.ncl.ucar.edu/gallery.shtml> The one change I'd note from some of these is that the title rarely goes on top of the plot in a paper but is usually inlaid within the axes such as here:

On Sep 9, 2020, at 6:17 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Thanks for checking in. Classes start tomorrow.

We really don't need to rush on the paper; as far as the science research class is concerned, the final copy isn't due until a full year from now. (Much of my class hasn't even started their experiments, mostly because all the labs are closed.) I just needed to get the intro drafted because the teacher assigned it to show that our background research is well-developed. Of course if you have a deadline for this paper, I will definitely adhere to it.

I would be happy to work on outlining the paper. For the science research class, we are required to select a few tasks to complete every 2 weeks, and making an outline could certainly be one of them. Other assignments I get include reading journal articles and putting together presentations.

On Sep 9, 2020, at 7:10 PM, John Fasullo <fasullos@gmail.com> wrote:
Just touching base on this work. I imagine your high school classes have begun again? I've been busy with other work but am going to get an outline set and being putting our paper together.

On Aug 31, 2020, at 2:28 PM, Benjamin Oren Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Never mind, found the figure. That is xghg, xaer, etc. I have an updated figure for that data, with letters that I will refer to in the caption. For now, a is ff, b is xghg, c is xaer, d is xbmb, e is xlulc, f is fixedO3. Control is the same for all.

I have the ff-xghg... in this figure, with the results from the 10000 random selections (bootstrap test). The dates are uneven between the 2 columns, which I can fix.

On Aug 31, 2020, at 4:21 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Sorry for the delay. Which figure do you mean?

On Aug 27, 2020, at 4:03 PM, John Fasullo <fasullos@gmail.com> wrote:

This looks good. Could you remind me if the titles on this figure are correct? Is "Greenhouse Gas" the allbutGHG or the FF-allbutGHG?

On Aug 27, 2020, at 12:03 PM, Benjamin Oren Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Here is a copy of the updated introduction. I tried to be more focused on past research directly related to our project, and I described our goals/hypothesis. Thanks so much for your feedback earlier!

On Aug 25, 2020, at 1:06 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Hi Dr. Fasullo,

I made a few more changes to the intro. Wanted to let you know that I am switching to saving it on GitHub because I need to use a few offline apps to compose it (bibdesk, vim, etc), and the only other option was buying an Overleaf plan. Let me know if you want access to view or edit the source, so I can invite you to the repository. Alternatively, I could email the PDF it to you or periodically copy the pdf to the shared dropbox folder.

On a side note, I've decided to learn R. The correlation computations were difficult to set up in Python, and I think I might have better luck with R. I will still use Python to prepare the datasets, but I might translate most of the math, or at least use R in the future. I think learning its stronger statistical api will pay off in

the long run.

Sorry if these changes cause any inconvenience. Please let me know if you need anything.

PS. The reason I am focusing on the writing as opposed to more analysis is that my introduction is due when school starts and I want it to be ready by then. Once school starts, I will probably have a few more written assignments, and then I can continue processing the data.

Sincerely,
Ben

On Aug 19, 2020, at 7:39 PM, John Fasullo <fasullos@gmail.com> wrote:

Read through the Overleaf document - a bit overly focussed on ENSO/paleo which I'm afraid doesn't provide much guidance for us aside from saying that we expect changes might occur. But a nice start.

On Aug 19, 2020, at 10:27 AM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Ok, great. I only have a few paragraphs so far, mostly about previous research. Do you mind if I share it with you on overleaf? If you don't have overleaf, I can send a PDF.

On Aug 19, 2020, at 10:38 AM, John Fasullo <fasullos@gmail.com> wrote:

On the intro, sure I'd love to see it. I don't think you have to list Peter as a coauthor yet but feel free if you want.

On Aug 18, 2020, at 11:10 AM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Ok. By the way, I am beginning to work on the introduction for the paper for a science research class assignment. I can send a draft or the latex source if you would like. I am using the AGU's template. Should I list Dr. Gent as a coauthor?

I believe that I am required by the class to do most of the writing of the paper, if you are ok with that.

On Aug 18, 2020, at 12:59 PM, John Fasullo <fasullos@gmail.com> wrote:

For now let's do it this way and perhaps go into finer detail for anything we go to publication with.

On Aug 18, 2020, at 10:11 AM, Ben Goldman

<bg502257@live.wpcsd.k12.ny.us> wrote:

Ok, great. Just to be sure, it is adequate to take the p-values given by the Pearson's correlation computation and overlay white for $p > .2$, grey for $p > .1$ and color for the correlation coefficient for $p < .1$?

This is a big relief. At first, I thought I would have to calculate the correlation coefficient at each grid-point for every ensemble member.

On Aug 18, 2020, at 12:03 PM, John Fasullo <fasullos@gmail.com> wrote:

That's right. The p-value is suitable but note that your estimation of that likely assumes each data point is independent. Should be OK.

On Aug 18, 2020, at 9:58 AM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Could I just use the p-value from the correlation function and mask where it is < 0.05 , < 0.1 , and < 0.2 ? Or is the significance calculation more complex? I'm not sure what to do with the degrees of freedom. Online, it says that to get confidence intervals, you transform r to z' and then multiply by the standard error. Is that what you are describing? Or I should do a t-test using r , which I think uses the number of degrees of freedom?

On Aug 18, 2020, at 10:21 AM, John Fasullo <fasullos@gmail.com> wrote:

Not a problem. Really all you have to do is compute the # of degrees of freedom and the correlation. Since the # of degrees of freedom is effectively constant for each point in our domain this will equate to just plotting the correlation values above/below a minimum absolute value - it doesn't mean redoing your computations. For the 1920-2080 period we have what should be about 5 degrees of freedom for 30-yr windows - though you can test this by looking at the autocorrelation of the piControl runs. I guess this would also be a decent way to do this - compute correlations across 1000 randomly chosen intervals of the control of similar length and see what the bounds on the distribution of correlation are. If you just want to take a quick look, use levels of 0.4 and 0.6.

On Aug 17, 2020, at 11:43 AM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Hi, Dr. Fasullo. Sorry I don't have the new figures. I am still having trouble calculating the confidence intervals. I tried getting the correlation arrays for entire ensembles, but that takes too long to execute. I have a few questions about confidence intervals. Are the confidence intervals the same over the entire array/ cross section, or is it specific to each grid-point? Also, do you need to assemble a set of separate correlation coefficients to calculate it, or can you derive it from the r or p value?

My apologies for this interruption to our project.
Thank you for all your help!

On Aug 11, 2020, at 2:15 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Ok, great. I will be using the python data file format because it is more efficient.

On Aug 11, 2020, at 2:14 PM, John Fasullo <fasullos@gmail.com> wrote:

I think we can actually just link to the Earth System Grid. No need for submitting our processed data.

On Aug 11, 2020, at 12:12 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Thanks for clarifying, one other question: I saw that the journal requires us to submit all of our data. Does it matter what format I use? I can do text files or Python's binary file type.

On Aug 10, 2020, at 11:45 PM, John Fasullo <fasullos@gmail.com> wrote:

Yup. And significance.

On Aug 10, 2020, at 5:41 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Ok - sounds good. I am pretty sure the Matplotlib program that I use can do that. I will still have to calculate the correlation grid for each member though, right?

On Aug 10, 2020, at 7:39 PM, John Fasullo <fasullos@gmail.com> wrote:

no just screen out (with white or grey) any correlations for which significance is <80% <90%. One way to do this is to set such values to zero or missing values. Even better is to apply a transparency to a figure without doing so (in Illustrator for example) and then overlay the contours that are screened for significance - but that is a bit more advanced. the key point is to emphasize regions that are significant while covering the broader patterns and omitting values that are negligibly small.

On Aug 10, 2020, at 5:02 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Oh, sorry! I misinterpreted what you meant. All of my calculations have been using the ensemble mean so far. So I should calculate the correlations for each

member, and then display the confidence limits for the entire ensemble at each grid point, correct?

On Aug 10, 2020, at 6:54 PM, John Fasullo <fasullos@gmail.com> wrote:

I'm not sure these are right. Recall that I was alluding to 80% and 90% confidence limits. Not 0.8 and 0.9 correlation coefficients. Perhaps read up on these. Do you have a book in statistics? If not, I found an old site here that describes the basics. <https://www.sportsci.org/resource/stats/pvalues.html>

On Aug 7, 2020, at 2:12 PM, Benjamin Oren Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Here are the new figures:

As you can see, the correlation coefficients for all but the aer ensemble are nearly all negative.

I have also included the timeseries for the surface and subsurface layers for each ensemble for each cross-section. Surface is on the left column, subsurface is on the right. Temperature is in blue, Nino 3.4 is in green.

Keep in mind that all timeseries are linearly detrended and smoothed with a 30-year moving average.

I have also included the correlation cross-sections with the colormap you mentioned in your last email. Very little of the data is correlated >90%, so much of the pattern is not displayed.

<Screen Shot 2020-08-07 at 3.11.11 PM.png>

<Screen Shot 2020-08-07 at 3.15.23 PM.png>

<Screen Shot 2020-08-07 at 3.20.16 PM.png>

<Screen Shot 2020-08-07 at 3.24.34 PM.png>

On Aug 5, 2020, at 6:04 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Yeah, I'll fix that asap. I will see if I can make that work. It will probably involve defining a custom colormap. Right now, the figures are somewhat misleading as the red areas are much less significant than the blue regions.

On Aug 5, 2020, at 5:52 PM, John Fasullo <fasullos@gmail.com> wrote:

Just noticed the range of the contours. Should go from -0.9 to 0.9 with blues/reds at the extremes and insignificant levels in white and then grey - perhaps all significance levels < 80% in white, grey for 80-90. Does this make sense?

On Aug 5, 2020, at 11:48 AM, Benjamin Oren Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Here are the updated figures:

I limited the depth at 700 m, and used a finer gradient for the contour. Be aware that the range of the colormap varies between plots. I will fix that in the next "iteration".

On Aug 5, 2020, at 1:22 PM, Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Ok, that clears everything up. Sorry for the confusion!

On Aug 5, 2020, at 1:22 PM, Peter Gent <gent@ucar.edu> wrote:

Ben,

The depth unit in POP is cm, not m. So the centers of the top two grid cells, which are both 10m thick, are at 5m and 15m.

Cheers, Peter.

On Wed, Aug 5, 2020 at 10:40 AM Ben Goldman <bg502257@live.wpcsd.k12.ny.us> wrote:

Hi Dr. Gent!

Thank you so much for helping us! Please let me know if you need any information or data.

I'm pretty sure that the plots are the right way up. I'm unsure how to only plot the upper 700 meters, as the cross-sections begin at 500 meters, and then the next layer of data is at 1500 meters. Maybe I have the units wrong. The depth dimension is in meters, correct?

I should have the timeseries plots soon. Just to be sure, what depths should I be using for the surface and subsurface boxes? Also, aside from the surface and subsurface layers for each cross-section, should I include any other regions?

Thanks for interpreting these results for me. How does stratification regulate upwelling?

> On Aug 5, 2020, at 10:02 AM, John Fasullo <fasullos@gmail.com> wrote:

>

> Peter and Ben - I wanted to introduce you to each other and so I've included you both on this email. Attached is a picture of Peter sporting his stern British demeanor - perhaps Ben you have one on hand (without the sternness is fine)? As we've discussed, Peter is going to lend his oceanographic insights to our project. I have briefed him with a short powerpoint summarizing some of our results.

>

> On the plots you recently sent, we need to plot them out with the vertical axis oriented with the surface on top and I think also want to restrict our attention to the ocean's upper 700 m since that is where the physical mechanisms likely have to be acting. With any correlations that we think are important it is an important next step to take a look at the timeseries to see where the shared variance is coming from (e.g. variability or trends) and be sure that our broader interpretation of mechanisms is correct. Right now, it seems to me that the stratification in the central Pacific ocean is regulating the efficacy of upwelling in driving a surface temperature change (that is for a fixed vertical velocity anomaly, the greater stratification leads to a greater temperature change). But there is much we have left to do to justify / discredit this interpretation.

>

> Let's take a look together at your new plots (showing both TEMPDT and TEMPCEP/EEP/WEP) once you've made the alterations above.

>

> Best to you both,

> John

>

>

> <gent.jpg>

<19.pdf><21.pdf><22.pdf><23.pdf>

<19.pdf><21.pdf><22.pdf><23.pdf><20.pdf><24.pdf><25.pdf><26.pdf>

<paper.pdf>

<PastedGraphic-1.pdf>

<2.pdf><4.pdf>