

TIPS ON PEER & EXPERT REVIEWING FOR GEOL 2950

B. FOX-KEMPER

Brown University Department of Earth, Environmental, and Planetary Sciences,
324 Brook St., Providence, RI 02912, USA
baylor@brown.edu

(Updated 7 September 2018)

In addition to writing the proposal and summaries, you will each be performing reviews of each others and the IPCC's work. Unlike most journals, these reviews will not be anonymous. You will be listed by name as an expert reviewer in the IPCC SROCC appendices. This review process will give you an opportunity to read closely about topics other than the one you chose for your proposal and to learn about science writing.

Key Words: Ocean Circulation, Climate, Research Papers.

1. Getting Help!

I am usually available by email. You can make an appointment other times. Just check my calendar at <http://fox-kemper.com/contact> and suggest a time that works for you.

If you are having trouble with the writing, there are lots of places to find help! You can make an appointment at the writing center (http://www.brown.edu/Student_Services/Writing_Center/) You also might ask other students for tips. I can show you past student reviews from related courses if you'd like. If you'd like to examine the expert and government review comments from the fifth assessment report, they are here: <http://www.climatechange2013.org/report/review-comments-disclaimer>. I've put my favorite writing style guides in the bibliography (Turabian 2013; Strunk et al. 2005; Montgomery 2003; Gopen and Swan 1990).

2. Basics

There are a lot of quandaries that arise in peer-reviewing (e.g., one reviewer loves it and one hates it, or a reviewer makes incorrect statements), so you'll get some experience with those issues by practice in a friendly environment during the proposal review process, both as a reviewer and a reviewee. Even though your reviews will be signed and friendly, however, don't let your classmates or the IPCC off too easy! They need your critical input to learn and improve their projects.

2.1. Rubric

For the proposal reviewing, we will be using a rubric based on the AGU guidelines for review. They are a useful guideline to go by, and when you do reviews of your fellow students, I'll expect to get a A1 or B2 or B1 score, etc. The AGU instructions to reviewers follow.

Even if you comment in the margins everywhere on the proposals, you should always

write a short summary blurb of the paper, giving the grade result of the rubric, and restate your primary objections with the paper. For the IPCC expert reviews, you should do the same thing, except you will just turn those in to me rather than uploading them to the IPCC. The IPCC will accept only excel files with comments referenced by line number.

2.2. GENERALLY, DON'T PUT YOUR NAME INSIDE OF A REVIEW

Unless you want the author to know who you are. When reviewing for journals, this can be a problem! In this class, since we are using canvas peer reviewing (which is not anonymous) and the IPCC expert reviewing (also not anonymous), you will know who your reviewers are and they will know when you review them. This changes the tone of the reviews a bit, which we might discuss a bit in class.

2.3. What is Peer Reviewing Good for?

This class and IPCC exercise mimic the process of peer reviewing in a scientific journal, which is one of the key touchpoints of the whole scientific enterprise. Here is an elegant and effective piece written in 2011 for the Guardian by Ben Goldacre that's one of the clearest descriptions of the process (in an editorial about unproven claims made by an Oxford professor informally).

A scientist with enduring concerns about a serious widespread risk would normally set out their concerns clearly, to other scientists, in a scientific paper, and for one simple reason. Science has authority, not because of white coats, or titles, but because of precision and transparency: you explain your theory, set out your evidence, and reference the studies that support your case. Other scientists can then read it, see if you've fairly represented the evidence, and decide whether the methods of the papers you've cited really do produce results that meaningfully support your hypothesis.

Perhaps there are gaps in our knowledge? Great. The phrase "more research is needed" has famously been banned by the British Medical Journal, because it's uninformative: a scientific paper is the place to clearly describe the gaps in our knowledge, and specify new experiments that might resolve these uncertainties.

But the value of a scientific publication goes beyond this simple benefit, of all relevant information appearing, unambiguously, in one place. It's also a way to communicate your ideas to your scientific peers, and invite them to express an informed view.

In this regard, I don't mean peer review, the "least-worst" system settled on for deciding whether a paper is worth publishing, where other academics decide if it's accurate, novel and so on. This is often represented as some kind of policing system for truth, but in reality, some dreadful nonsense gets published, and mercifully so: shaky material of some small value can be published into the buyer-beware professional literature of academic science; then the academic readers of this literature, who are trained to critically appraise a scientific case, can make their own judgment.

And it is this second stage of review by your peers – after publication – that is so important in science. If there are flaws in your case, responses can be written, as letters, or even whole new papers. If there is merit in your work, then new ideas and research will be triggered. That is the real process of science.

If a scientist sidesteps their scientific peers, and chooses to take an apparently changeable, frightening and technical scientific case directly to the public, then that is a deliberate decision, and one that can't realistically go unnoticed. The lay public might find your case superficially appealing, but they may not be fully able to judge the merits of all your technical evidence.

Relatedly, there are well-known psychological biases in assessment of one's own skill level (Kruger and Dunning 1999; Dunning et al. 2003). These Ig Nobel Prize† winning

† In an ironic twist, I found an online site claiming Dunning and Kruger had won the 2000

studies describing the Dunning-Kruger effect. These studies are amusing to read (they feature stories about incompetent bank robbers and humorless students), but they offer important lessons for why peer reviewing is a powerful tool in gaining expertise. People with the lowest levels of competence tend to overrate their own abilities (the bottom quartile rate themselves as being in the third quartile generally). People with the highest skill level tend to underestimate their abilities versus their cohort (or more precisely, they mostly overestimate others). This is demonstrated as being due to a lack of *metacognition* in people with the least experience, i.e., too little experience to be able to judge their own abilities, and an excess of modesty and lack of exposure to others' skill in the most skilled quartile. Beware, there can be emotional trials as you gain skills in metacognition and self-awareness—you might feel smarter or dumber—but correctly calibrating your assessment of your own skill and that of others is critical to being able to improve your skill and recognize whom to study to improve.

So, as scientists, we all need to become good peer reviewers *and* critical readers and thinkers.

3. How do the Grades Work?

An email from a past student:

I have a quick question about the peer evaluations and I thought I should ask before our next evaluations are due. I am a little confused about how the alphanumeric system works with the respective score (and grade ultimately) we are giving the writer. I feel like in reality, no paper from this class should really be worthy of publishing (right?). Especially these first drafts. Therefore, I don't think I will ever give someone a 1, since that is the criteria.

I guess my question is: do the scores out of 100 given to the writer directly correlate with some grade? I felt bad after seeing how things were scored, because I gave both the people I reviewed A3s. However, I felt the score was very dependent on the nature of the assignment, but that both people did exactly what the assignment asked for. I think it would just be helpful when I review (and receive reviews) to understand how this alphanumeric rating is affecting the grade.

My reply (edited for this context a bit), which basically suggests you consider context whenever you review:

You are right, they probably won't reach the standard for publishing, at least not often! But, I think the better way to interpret this is whether they are publishable within the class setting—that is, would it fit in well in a proceedings volume/website of course work. This proceedings would be public, and peer-reviewed (by you), but understandably less complete than you'd see in GRL (well, at least less so than some of the articles in GRL).

But, don't *assume* that these papers aren't publishable, a good idea will go a long way, even without much time to defend it, and many of them may be publishable with only a little more work beyond what is done for class. So read each one with a fresh eye, and realize that you might just see a 2 paper: publishable with revisions. One problem that students have is that they read 'classic' papers or 'important new' papers in classes and journal club. Many, many papers are neither classic nor important, but they are still published! (I'm not saying you should write crummy papers, just don't prejudge).

I do pass on you guys' scores to the final grade, (each reader gets 10 pts to assign, an A1 would be a 10 and a C4 would be a 5). My grades are similar, except I give

Nobel prize in psychology for this work on overconfidence, when in fact they won the 2000 Ig Nobel prize. The Ig Nobel prize is not to be confused with the Nobel prize, as it is a semi-satirical award for "achievements that first make people LAUGH then make them THINK."

from 40 to 80 pts. Thus, if you turn in the assignment you get a 50, and you can work your way up to a 100, with 10 of those gained points from your readers and 40 points from me.

So, you do have an impact, and if you read the scores as being 90+ for an A (although I may curve a tiny bit at the end of semester), then you can't get an A without some support from the readers. However, the readers can't prevent you from failing (60) without me.

Overall, rather than being concerned with how you do the grading, I'd prefer if you focused on your comments. If you gave every single review an A3, but gave good, useful, positive criticism on your reviews it would be immensely helpful. Overall, grades in graduate school are not so critical (even fellowship applications typically don't ask for GPA, nor is a given GPA required to pass prelims or get a PhD), but learning how to write papers, be a good peer reviewer, and think scientifically are critical to your future careers.

4. You will submit your review in plain text if possible.

You will submit your review in plain text. Compose it in a notepad or just in a regular email, then cut and paste it into the box to the right of the paper in canvas/crocodoc. If you need to use pdf for equations or whatever, that's OK to add in addition to the text comments.. .tex .doc and .docx formats are NOT OK FOR REVIEWS. You should also upload a duplicate of your txt or pdf to the Peer Review assignments on canvas, so that I can comment back to you and grade them.

5. The AGU Guidelines

The following pages are taken from the instructions to reviewers for GRL. We will use them as a rubric for our class, with the following adaptations:

- Interpret descriptions of **scientific merit** in the classroom context: e.g., 'important new science at the forefront of an AGU discipline' corresponds to 'important new science at the forefront of what we are learning in class' or 'would convey or exemplify an important topic to a student in this or a similar class'
- Interpret descriptions of **publication quality** in the proceedings volume context: e.g., 'if a submitted manuscript meets GRL standards' equates to 'if a submitted manuscript meets the standards of our proceedings'.
- Presentation categories can be used directly

6. AGU Peer Review Instructions

Editorial Instructions to Reviewers (A reminder of current GRL Policy)

Geophysical Research Letters aims to provide rapid publication of forefront research that has an immediate impact on the science community. The journal features articles from a broad range of geophysical disciplines. We ask your help as a reviewer in evaluating both scientific content (Categories 1-4) and presentation quality (A-C) to determine if a submitted manuscript meets GRL standards.

We welcome and will pass on to authors any specific suggestions which would materially improve the quality of a manuscript. Your comments will be conveyed anonymously unless you specifically note in your review that you wish your identity to be passed on to the authors.

Scientific Quality: Scientific quality of the manuscript is fundamental to publication, and the following Categories 1-4 are meant to aid the reviewer and Editor.

Science Category 1: The manuscript meets one or more of the following criteria:

- Important new science at the forefront of an AGU discipline
- Innovative research with interdisciplinary/broad geophysical application
- Instrument or methods manuscript that introduces new techniques with important geophysical applications

If the manuscript falls in Category 1, please give sufficient detail as to which of these criteria and why.

Science Category 2: The manuscript is potentially Category 1 but significant clarification/revision is needed. For example, the manuscript presents:

- Some unclear or incomplete scientific reasoning
- Inadequate presentation of data
- An instrument/method where the geophysical application is not obvious

If possible, please specify the revisions that might allow this manuscript to meet Category 1 criteria.

Science Category 3: The paper is publishable in the refereed literature but is unlikely to become a Category 1 paper. For example:

- It is a scientifically correct paper but not obviously a significant advance in a geophysical field
- A solid paper with little immediate impact on the research of others (e.g., a routine application of a standard research technique, or a new measurement/laboratory method with limited geophysical application)
- A good but basically incremental improvement to existing data sets, models, or instruments

Science Category 4: This paper is basically unpublishable in an AGU journal:

- There are major scientific errors in the manuscript
- Essentially the same material has been published or is being considered for publication elsewhere
- The technique is not useful
- The research area is not representative of an AGU discipline

Presentation Categories: These categories measure the maturity of the submitted manuscript in terms of language, communication, and GRL criteria.

Presentation A : Manuscripts should meet ALL of the following:

- Abstract is succinct (< 150 words), accurate, and comprehensible to a non-specialist
- Manuscript is generally well-written, logically organized, and adequately illustrated
- Figures and tables are understandable and readable (when sized for GRL)
- English usage and grammar is adequate, with few spelling/typographical errors (please specify any minor fixes)
- Manuscript appears to fit GRL's 4-page limit

Presentation B: Manuscripts are potentially "A" manuscripts with suitable revision. Please give explicit direction as to which sections/features need revision, extension or reduction. For example:

- Abstract needs to be rewritten/shortened
- Manuscript is not well written, is not logically organized, or is inadequately illustrated
- Manuscript needs to be (and can be) shortened
- English usage, grammar, or spelling errors detract from the paper

Presentation C: Manuscripts cannot readily be revised by the authors into Presentation "A" without a major re-write. For example:

- Specific ideas cannot be adequately presented within the 4-page GRL limit
- Organization and illustration of the manuscript make it too difficult to review fairly
- English usage, grammar, and/or spelling errors are endemic and require substantial copy-editing before this manuscript can be reviewed adequately

Highlights: GRL is now highlighting several manuscripts in each issue. Science Category 1

manuscripts are potentially highlightable. If you feel that the manuscript you are reviewing is particularly exciting and deserves to be highlighted, please include in your review a short note as to its importance.

The Editors of GRL rely on the counsel and recommendations of reviewers in order to maintain the quality of the journal and meet GRL's specific criteria. Please recognize that the final decision on a manuscript is made solely by the Editor using the expertise of several reviewers and based on the above GRL criteria.

Michael J. Prather
 Editor-in-Chief
 15 March 1999

7. Example Reviews

Here is a good example of three reviews for a paper from a class of mine like this one. Note their length, level of detail, breakdown of scientific vs. presentation issues, and specific suggestions for how to fix things. Also, notice that they are not free from typos, spelling errors, etc.

A good review does not contain "Conan the Grammarian" suggestions. For many of you, this part will be very unfamiliar. When reviewing a scientific paper, it is not your job to correct grammar, punctuation, spelling, etc. So, it is unlike reading a term paper for a friend! There is a technical editor at any competent journal who will do that for the author after you are done. You should only mention these issues if there are so many as to limit the presentation score, or ones that are just hidden typos that you mention just to be nice.

Your job is to think about the science: Do the numbers seem right? Is the method sensible? Do the conclusions follow from the experiment? Are the hypotheses actually tested? Does the abstract reflect the most important results without wandering off into issues beyond the scope of the paper?

=====

My Review

Summary: You compare the SST found from ARGO and Reid-Mantyla datasets in the Nino3.4 region and find that the mean state differs more than an official ENSO worth.

Strengths:

--Nice choice of problem, it satisfies the pedantic assignment with an interesting application
 --Wow, there's a lot of ARGO data. The histograms are totally different!

Major Problems:

--I think you've got too much cold data, and I expect it is because 20m is too deep to count as SST. Do you have enough Reid-Mantyla data to go shallower?
 --In your discussion of normal distribution, it is apparent from your histograms that they are not normally distributed. Also, from your longitude vs. temp plots, it's clear that you've got some weird cold stuff thrown in with the clustered surface values.
 --I think you need to do some more digging in the data. It is apparent that you don't just have ssts here. You should do some investigating about the depth of the cold tongue vs. season and then make a choice of your upper bin.

Minor Problems (THESE MIGHT BE INCLUDED AS MARGINALIA COMMENTS USING CROCODOC)

--Probably NOAA's ENSO 3.4 index not product.

--Bibliography formatting could be nicer. Use a reference name in the place you want to reference, and then a table of complete references or urls at the end.

--You are giving the st dev of the datapoints, not the st dev of the mean. If each obs was uncorrelated with the others, the latter should be \sqrt{N} smaller.

Overall, this is a nice beginning and a great topic. You need to do some more digging to understand what the data is telling you. It would be much nicer if there was a clearer description of where and why the temperatures are different between the two datasets.

B2

=====

Reviewer 1: B3

Report of Referee on Defining the Mean Temperature State of the Eastern Tropical Pacific: Neutral ENSO by ****A STUDENT****

In this paper, the author calculated the mean sea surface temperature (SST) in ENSO 3.4 region from two different data sets: Reid-Mantyla and Argo. Interestingly, the calculated mean SSTs differ by 0.59 C, which is larger than ENSO threshold of 0.5 C. In my opinion, however, the author does not provide sufficient discussions on his results. For improvement of the paper, the following issues should be addressed in the revised paper;

1. Which of the data sets represents better the mean SST in ENSO 3.4 region? If any, what could be the sources of error?
2. Figure 2 shows the anomaly in Argo data. The anomalous temperatures of around 5 C are too low. This may indicate that the anomaly is due to the instrumental or calculational error.
3. What is the meaning of bias in the last sentence of the results section?
4. The time periods of study for the each dataset should be stated in the abstract.
5. The standard mean SST that NOAA uses to determine ENSO index should be mentioned in the text and be compared with the authors results.
6. The spatial map of ENSO 3.4 area may be informative to readers.
7. The longitude values in Figures 1 and 2 are not consistent. The author needs to choose one of them (-170 or 190). It may be better to use 170W to 120W as x-axis.
8. For convenience in the comparison, the author should use the same temperature range for Figures 1 and 2 and for Figures 3 and 4.

=====

Reviewer 2:

****A STUDENT**** - Defining the Mean Temperature State of the Eastern Tropical Pacific:
Neutral ENSO
Grade: A3

Abstract ----

It would be nice to have SST and ENSO3.4 spelled out. I don't know what the

'3.4' on the end of ENSO; hopefully I'll find out in the paper. With respect to the final sentence, which difference is greater? The ARGO or the R-M data, something else?

Introduction ----

At the end of the first paragraph, how do the greater western Pacific temperatures reflect the Pacific Warm Pool and Western Boundary Current? I may be misinterpreting your intended use of 'reflecting'.

In paragraph 3, I would change "In the ENSO3.4 region, its data spans roughly 40 years and all seasons, although not every season every year (see Figure1.)", to read "The ENSO3.4 data spans roughly 40 years through all seasons; however, there are some gaps in the dataset (see Figure1)." After this you should have the actual Figure1 for easy reference.

Would it be helpful to give a maximum depth of the ARGO floats? I think specifying that they were first deployed in "the year 2000" would help.

I think the last paragraph of the introduction would fit well into the abstract.

Methods ----

It's not exactly clear to me how each dataset was sampled. They seem pretty similar, but the description is not terribly explicit to distinguish between the two. You have a nice justification/explanation of how the depth vs. pressure selection is validated.

I don't think the PDF curves on Figures 3 & 4 really help, just distraction from the histogram.

Theory ----

No need to emphasize the "both" so much. Your reasoning seems reasonable to me.

Results ----

I think you might be a little overkill on the significant figures.

Visual measurement is a bit iffy. Is there some statistic you could use (skewness for example)?

Conclusions ----

The second sentence is a run on. Break it up into distinct bits of information.

Wouldn't the range be a better measure of temperature variation in the tropics instead of standard deviation?

General ----

The citation format should be more succinct in the body of the paper and then in

detail in the reference section.

Make sure not to leave bits and pieces of sections hanging out alone on separate pages when possible.

You should have each figure immediately after you reference it and not before.

All in all a well layed out paper. I think most of the deficiencies are easily revised.

8. Canvas Tips

We will be using Canvas to aid in the peer reviewing process. After assignments are turned in, you will be able to access those to which you are assigned to review. You can click to see the uploaded pdf. You can add text, upload files, etc., on the right hand side of the window. PLEASE DO NOT ATTACH .doc or .pdf FILES FILLED WITH PLAIN TEXT. *Cut and paste plain text into the comment boxes instead.* You can also add comments to markup particular locations in the text you are reviewing. With a little experimentation, you will see how this works. When you are done with your review, you can submit it. Once you submit, I don't think you can go back and see it again. See the example figure below for the layout of some typical reviews. See how the plain text is displayed on the right and the markups are displayed within the document.

REFERENCES

- Dunning, D., K. Johnson, J. Ehrlinger, and J. Kruger: 2003, Why people fail to recognize their own incompetence. *Current directions in psychological science*, **12**, 83–87.
- Gopen, G. and J. Swan: 1990, The science of scientific writing: If the reader is to grasp what the writer means, the writer must understand what the reader needs. *American Scientist*, **78**, 550–558.
URL <http://www.americanscientist.org/issues/pub/the-science-of-scientific-writing/1>
- Kruger, J. and D. Dunning: 1999, Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, **77**, 1121.
- Montgomery, S. L.: 2003, *The Chicago guide to communicating science*. University of Chicago Press, Chicago.
URL <http://www.loc.gov/catdir/description/uchi052/2002072627.html>
- Strunk, W., E. B. White, and M. Kalman: 2005, *The elements of style*. Penguin Press, New York.
URL <http://www.loc.gov/catdir/enhancements/fy0721/2005048913-b.html>
- Turabian, K. L.: 2013, *A manual for writers of research papers, theses, and dissertations: Chicago style for students and researchers*. University of Chicago Press, Chicago, 8 edition.
URL <https://josiah.brown.edu/record=b7338626-S7>

Submission Details

Paper: [redacted] submitted Apr 12 at 3:24pm (late)

1.81 MB View Feedback

Preview of [redacted] Paper3-1.pdf

2 of 21

Comment Draw Highlight Text Strikeout

of the error of Sverdrup transport estimates of WBC transport at 0° erect latitudes and their components (given) actually do not indicate a greater amount of error when applied across Madagascar. Both errors are still substantial at 10 to 40 Sv, depending on the season. Still, reduction of the island circulation might improve Sverdrup estimates at that latitude. One piece of evidence supporting this claim is the reliability of thermal wind in predicting vertical shear in meridional velocity, which is also based on the geostrophic approximation. The correlation of vertical shear and thermal wind is used for tracks drawn across Madagascar as seen in figure 4. Godfrey (1989) [1] introduced simple approximations consistent with the Bjornell model which will be used to estimate island circulation in this work, hopefully improving estimates of WBC transport.

II. THEORY AND METHODS

For oceanic flows away from the equator, Sverdrup transport, the combination of geostrophic and Ekman flows, is often a good approximation of basin circulation away from the western boundary currents (WBC). The meridional Sverdrup transport is given by:

$$V_s = \frac{g}{\rho_0 \beta} \frac{d\bar{\sigma}_\theta}{dy} \quad (1)$$

The thermal wind in the meridional direction can be written:

$$\frac{d\bar{v}}{dy} = -\frac{g}{f \rho_0} \frac{d\bar{\sigma}_\theta}{dy} \quad (2)$$

where $\bar{\sigma}_\theta$ is the barotropic geopotential, g is the acceleration of gravity at Earth's surface, ρ_0 is density averaged over the basin, f is the Coriolis parameter, and \bar{v} is the meridional velocity. The density was computed from salinity and temperature data using the CSF2 ocean toolbox [2]. pressure variations were assumed to be hydrostatic.

An approximate measure of the circulation around islands was introduced by Godfrey (1989)

Grade: [] [] []

out of 10

Title: Sverdrup Transport in the Southern Indian Ocean

Author: [redacted]

Review: B2

Summary: Using data from the ECCOv4 simulation of MITgcm, the author estimated Sverdrup transport in the Southern Indian ocean, specifically the circulation around Madagascar. The monthly Sverdrup transport calculated by integration across a zonal transect was compared with the "Island Rule" calculation established in preexisting literature. The author found that the model calculation deviated significantly from the Island Rule prediction. It is unclear what caused this strong disagreement, but the author suggests it may result from model resolution or incorrect integration bounds.

Strengths:

- Nice plots
- Good description of the Island Rule equation. Also, a cool idea to compare model data to the Island Rule calculation.

Major Problems:

- I found it very difficult to follow some parts of the paper simply from having no diagram to reference for where currents are and how they relate to each other.
- The correlation coefficient seems out of place in the paper. You talk about the results of the analysis briefly in the introduction, before the equation has been defined, and then it isn't mentioned again. I think the paper would be stronger either without the correlation coefficient, or with a more detailed description/analysis of the findings (placed after the theory section).

Minor Problems:

- Some typos, including an incorrect equation

Overall: You asked some interesting questions related to your findings, but

FIGURE 1. Example Peer Reviews in Canvas