

Review: Ten simple rules for working with high resolution remote sensing data

This paper offers some useful tips such as a new graduate student entering a field that uses remote sensing data might find helpful. These are offered in the form of 10 'rules' under the assumption that using high resolution remote sensing data requires researchers in Earth and Environmental science to be equipped with skills that are seldom taught.

The following critiques are offered with agreement that there would be value in this paper being published, however, it is in need of significant revision.

1. While the reviewer agrees there would be value in the paper concept (Ten simple rules for working with high resolution remote sensing data), the need for such has not been given here. It is simply stated that analysing particularly high resolution remote sensing data is now common yet requires specialised skill and researchers are not taught these skills. In my experience the later is not true. Many free and quality resources exist covering a wide range of the specialised skills needed, from the many software and data carpentry courses through the equally prolific range of short courses most Environmental and Geographical Science Departments and Computing centers run focused on remote sensing data processing, to formal graduate level courses that I know exist in many related departments. Finally NASA, NOAA, the ESA, QGIS, and FOSS4G among others all offer training resources on their data/tools/formats, and Esri offers paid commercial training courses.
2. The paper would be better positioned as a science blog aimed at new graduate students than a peer reviewed journal. The authors do not state the target audience for this paper, however, the content would be of some value to a new graduate student but they are not indicated as a target and publishing under peer review would suggest this is not the intended primary audience. The majority of experienced researchers would not find any significant value in reading this paper.
Additionally, the tone of the paper is more in line with what might be expected in a science blog than a peer reviewed paper. Examples of this are throughout the text but to point to two:
 - *"Show your work and create open workflows to ensure that your effort is also accessible to the community. Weigh the pros and cons of innovation for your particular project. Don't reinvent the wheel"*
 - *"But, consider finishing your plate (answering your science question) before eating dessert."*
3. Multiple of the "rules" presented and elaborations given are guidance that should be covered in any introduction to research course and are not specific to remote sensing data. Examples of such include the recommendations to:
 - Carry out a thorough literature review and develop a robust research plan early on.
 - Clearly define a research question
 - Keep in mind your research goal so as to not be diverted by interesting aspects emerging from the data during analysis, and to rather save such for future work.
 - Begin with a small exploration
 - Allow for/anticipate the unexpected

4. For a formal paper, there is a lack term definition. This is very important generally in all academic literature, but particularly when discussing open science, data, software and reproducibility currently, as these topics are still becoming normalised in the global scientific community and therefore very open to misinterpretation. For instance the terms “open” and “reproducible” and “high resolution” are casually used but not defined.
5. Rule 3 is “3. Don’t use high resolution data”. However, this is not justified beyond a statement that doing so can be costly. It is debatable whether or not this is actually advisable given the ready access researchers in the developed world at least have to extensive computing resources. While there are some conditions under which this would be reasonable these are not given.

Further, it might lead new researchers to use outdated lower resolution data when higher resolution but/and more up to date data is available, this in turn is likely to lead to incorrect conclusions.
6. While rule 6 is “6. Survey the computing landscape”. The discussion of this point is lacking. The need for a thorough literature review (which should include such) is already made earlier in rule 4 (and 2 slightly). However, it would be useful to point a reader to the many research compute infrastructure facilities and their training materials available globally that might not appear in a domain specific focused literature review.

Furthermore, while python, Xarray and Dask are briefly mentioned in the context of explaining data size challenges, is software tooling for remote sensing data analysis is to be discussed there are many other tools that are far more relevant to a wide audience such as QGIS, the plethora of formal published and coded processing procedures and software in many disciplines, Numpy, and Esri’s tools.
7. The citations of Wyngaard et al 2019 and Wilkinson et al 2016 are familiar to the reviewer and both have been poorly used.
 - a. Wyngaard et al 2019 is used as part of the problem statement. Specifically: *“The resulting bespoke approaches that applied researchers develop can be inconsistent, inefficient, and challenging to implement, reproduce, or extend (Wyngaard et al. 2019)”*

Wyngaard et al 2019 is only discussing these challenges as relates to data captured with UAVs as these problems are most prevalent with such data due to the immaturity of the tooling supporting UAV data management. These challenges do not exist to the same extent with for instance Satellite or manned aircraft data which are both very mature fields.
 - b. Wilkinson et al 2016 is used briefly in the discussion of Rule 10 (10. Show your work) to justify some form (this is not discussed) of publishing the method used. However, Wilkinson et al 2016 is the primary source on the FAIR principles which concern the management and publication of data not methods or software.
8. Finally, as indicated the reviewer agrees there would be value in a significantly refactored version of this paper. Recommendations for such would include:
 - Point to and summarise existing specific remote sensing related data ethics policies under the discussion of doing no harm.
 - As suggested in various above items: point to and discuss existing known best practices regarding remote sensing tooling, infrastructure, methods, and publication of

method practices.

- Discuss the complexities of publishing and managing FAIR remote sensing data specifically given that applying FAIR to remote sensing data as geospatial data is notably more complexity than other data. For instance, multiple endorsed recommendations on some of these complexities exist from the Research Data Alliance's (RDA) work.

- The same RDA body has also published best practices on software publication which would improve the value of the point on "showing your work". Discussing these and the many other articles and tools available for making software dependent research reproducible would also be a uniquely valuable contribution.