



EGUsphere, referee comment RC2  
<https://doi.org/10.5194/egusphere-2022-535-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on egusphere-2022-535**

Anonymous Referee #2

---

Referee comment on "Potential of natural language processing for metadata extraction from environmental scientific publications" by Guillaume Blanchy et al., EGU Sphere, <https://doi.org/10.5194/egusphere-2022-535-RC2>, 2022

---

### General:

Overall this manuscript fits well with SOIL, and the methodology as well as the results will be of interest to readers. The nature of the study, involving "natural language processing for metadata extraction from environmental {soil} scientific publications" is inherently multidisciplinary, and complex! The necessary methods are well discussed and well referenced, and the appendix of the NLP software will be a big help to researchers in this field. The results relating agricultural practices and soil and site properties are novel and important.

### Specific:

Most SOIL readers are probably substantially unfamiliar with NLP and would benefit from more focused guidance by the authors, which can be accomplished perhaps mostly easily by a trimmed revision. For example the Abstract is overly complex; the Introduction states the objectives of the study on just four lines 96-100, and a trimmed Abstract could focus simply on the achieving of the objectives.

The Material and Methods section is appropriately long, given the emphasis on methods, but could be edited to be more uniformly coherent. Perhaps part of that could be fixed by reformatting the variety of figures, and relegating some of them to just the appendix.

Most of the figures in the Results section are important, but much of the other discussions in Results are really recommendations and can be eliminated or partly moved to Conclusions.

Technical:

I see Reviewer #1 listed some technical issues, most of which I believe can be handled by trimming as suggested.