



EGUsphere, referee comment RC1  
<https://doi.org/10.5194/egusphere-2022-427-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on egusphere-2022-427**

Anonymous Referee #1

---

Referee comment on "Using machine learning algorithms to analyze remote sensing and ground-truth Lake Chad's level data" by Kim-Ndor Djimadoumngar, EGU sphere, <https://doi.org/10.5194/egusphere-2022-427-RC1>, 2022

---

In summary this article runs two parallel investigations into the relationship between a selection of atmospheric and land quantities, and two methods of measuring the height of Lake Chad, one form of data is measured in-situ while the other is measured from a remote sensing platform. The collection of data, and the data processing was transparent and thoroughly documented. The investigation then applied a series of out-of-the-box approaches at their default values to regress two lake heights onto the climate variables; a lengthy study of the comparison was made, and some light scientific conclusions from the point of view of relative importance of contributions from different quantities to the lake heights. Some patchy analysis and rough conclusions were drawn to suggest appropriate algorithms for the regression.

I believe the paper was trying to fit within the following scopes of GMD: (i) "new methods for assessment of models, including work on developing new metrics for assessing model performance and novel ways of comparing model results with observational data" and (ii) papers describing new standard experiments for assessing model performance or novel ways of comparing model results with observational data. Unfortunately I do not believe this paper fits within these categories, without major rewriting, and I describe this along with further broad reasons below, along with some suggested changes:

- **Motivation:** The paper is (partially) motivated by saying physical models are not used due to data scarcity, and this is why data driven approaches may be a way forward. And yet in regimes of scarce data, this is precisely where physical models excel, as they can use physics to generalize off-data, while data driven models require far more good quality data. Goals of the investigation was also to "contribute to the general understanding of hydrological processes in the Lake Chad basin", though no scientific conclusions were made in this paper, it was primarily focused on comparing machine learning tools and data exploration.
- **Novelty of methods and assessment:** The assessment came from a series of standard statistical measures such as  $R^2$  or MSE. Likewise the methods all came from the

standard libraries of Sci-kit learn. The methods were taken with their default values and were not tuned to problem performance. The DL method had a more in-depth overview of the construction, but performed very poorly in all categories without explanation, (perhaps from a lack of data or lack of size/layers of the relatively modest size of DL).

- Training: For many of these methods, performance is heavily dependent on tuning parameters. In a case where this parameter space is not explored, it is difficult to know if statements of performance apply to the methods themselves, or the quality of the packages default options. For example the clear overfitting to training data of RT, and likely overfitting of RF could possibly be improved with parameter choices?
- Results: Some results were repeated, e.g. Table 7 summarizes the performance, but Figure 5,6,7 repeat this data with no additional insights gained. Throughout, the fit to training data was given as evidence for performance. In some cases, e.g. Figure 9, this even changes the conclusions - that RF is quoted several times as being considered a better model for the data than LR, despite LR giving consistently lower test errors in both cases. It was not clear that Table 9 is only available in RF; furthermore, other forms of sensitivity analysis or attribution analyses were not carried out for SVR or LR to see if these results were consistent or based on the ML tool chosen.
- Conclusions: I did not understand many of the conclusions. (1) I believe the LR was shown to be more performant than RF on test data, though the authors state RF is a preferred technique (2) I did not understand as to what we should conclude from the use remote sensing and ground-truth data, I feel the authors merely indicated that data is useful for validation, this was consistent through as I could not understand what we were supposed to draw from the parallel investigations, nor what the results helped to explain in this regard (3) I did not see an explanation for Table 9, arguably an important conclusion of why we find different attributions of importance to the different data sources, and whether anything can be learnt from this.

For consideration for publication I would suggest

- Clearer referenced motivations for why it is a good idea to consider data driven approaches even in areas where data quality is poor.
- Making the data exploration more concise, e.g. is showing the calculations of interquartile ranges necessary?
- Critical presentation of results: I do not believe Figure 4(c),4(d) are necessary. Enhancing readability of Table 5 by splitting targets into new table and use of boldface to enhance useful comparisons. I do not believe Figure 5, 6 or 7 are necessary, nor their analysis beyond what is described in table 7. Need for log scales in Figure 8, and the error in  $\$R^2$  values (negative?). Removal of the confusing Figure 9, as in all other plots training and test data are separate, here they are combined.
- New results for robustness: (1) Ensuring that reasonable exploration of Scikit-learn parameter spaces are reported on for each tool were made to ensure robust method performance. (2) Sensitivity analysis for other methods such LR and SVR to compare with Table 9.
- Rewritten clear conclusions evaluating the success of the author's own goals laid out in the introduction: This should include explanation of why these parallel investigations were run, what are the consequences of the results in this respect. Explanation backed up by the robust ML results of which methods are best suited to the data set. Scientific explanation or discussion of why different data lead to different attributions backed up by robust evidence from the multiple methods such as LR, SVR, and RF attribution analysis.
- Discussion: Outlook on the steps and challenges that are required to make predictions

and projections with such models, what scientific steps could be taken on the back of this investigation regarding climate variable attribution, or the use of remote/in-situ data.

- General improvements to the formatting of tables, figures to enhance readability, with more detailed captions. Use of footnotes for URLs. rather than keeping them inline in text