

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/hess-2021-383-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-383

Anonymous Referee #1

Referee comment on "Building a methodological framework and toolkit for news media dataset tracking of conflict and cooperation dynamics on transboundary rivers" by Liying Guo et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-383-RC1>, 2021

This paper develops a toolkit to track media data on conflict and cooperation events in transboundary river basins. It is a praising approach to provide systematic collection and mining of mass qualitative data in transboundary river basin management. The paper provided detailed procedures on how such toolkit should be implemented, and further illustrated its applicability with various case rivers. It was presented in a well-structured manner, however, there are some concerns that needed to be addressed.

- clarification of the contribution of this paper is needed. In my opinion, this paper provided a good technical toolkit for retrieval, processing and analysing of keywords related to transboundary water conflict and cooperation, which is a highly specific target. A "methodological framework" (between line 50-60) claimed by the authors, requires inclusions of multiple principles and theoretical components that serve a broader goal. That being said, I think it's also more suitable to move the implication of the toolkit (line 60-75) in the introduction section to section 4.
- there are numerous data sources that collect news media data. What media sources are covered (print and web news? Other social media platforms such as Twitter?) Also, can the authors provide more details on how the LexiNexis data source is input in the toolkit and possibly other data sources? How does this toolkit perform in integrating data collected from multiple data sources?
- In Table 1 the authors summarised the special treatments about basin names. Can you provide more explanations on how to determine the frequency setting when treating river basins with the same names but located in different continents? Also regarding Block 5, are there any principles to determine what key words should be excluded?
- Section 3.1.3 and Figure 5: this section is not clear to me. What is the data used to generate Figure 5, all 286 rivers? 60 Key rivers? Or just Lake Chad as it appears to have the greatest frequency? And more explanation is needed on why presenting the word clouds in both titles and text body.
- Line 315-335: there are too many useless explanations of the figures such as "vertical axis represents...horizontal axis represents..." These are already clearly illustrated on the figures and does not need to be stated in words again.
- The whole paper needs to be grammatically checked again.

