

## Comment on essd-2020-397

Anonymous Referee #3

---

Referee comment on "ERA5-based database of Atmospheric Rivers over Himalayas" by  
Munir Ahmad Nayak et al., Earth Syst. Sci. Data Discuss.,  
<https://doi.org/10.5194/essd-2020-397-RC4>, 2021

---

The authors created a dataset of atmospheric rivers for the Himalayan region derived from ERA5 data. This dataset could be useful for the community for research into extreme precipitation and flooding. The manuscript is well written and presented.

I think the manuscript could benefit from a bit more explanation on the choice of the AR detection algorithm and the chosen time step. Why did the authors decide to use 6-hourly data despite ERA5 being available on a higher temporal resolution? What made the authors choose this AR identification method over other available methods?

When looking into the dataset I think there could be a bit more additional information on how the data is organised. I am not sure that someone downloading the dataset would be able to understand it in its current form. For example, it took me a while to figure out that a detected AR has a unique id but still has multiple rows as it consists of multiple timesteps. The description in the read me file is very short and could say more about the structure in the .csv files, e.g. that there is a line for every time step in an identified AR. The manuscript and meta data say that the covered period is 1982-2018 while the first detected AR in the files is from January 1979. For one AR timestep the IVT max says one value but when looking into the columns there is a higher IVT value. It seems a bit complicated organised that the longitudes and latitudes corresponding to the AR locations are in different files from the actual IVT values.

Line 246-247: ther is "southward" twice in this sentence, while I think one of them should be "eastward".