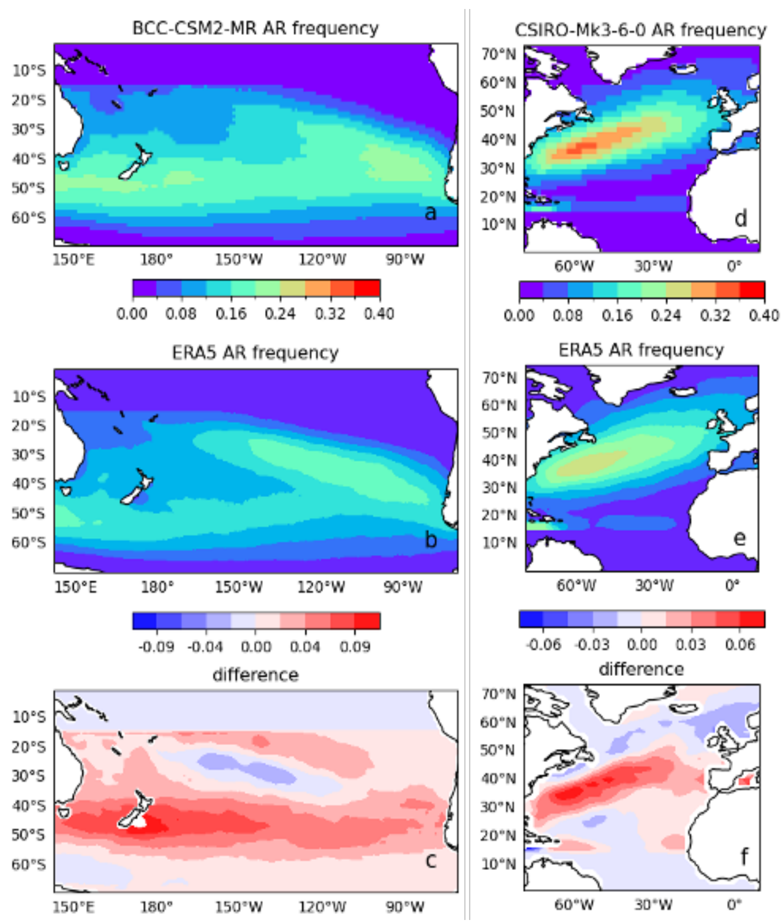
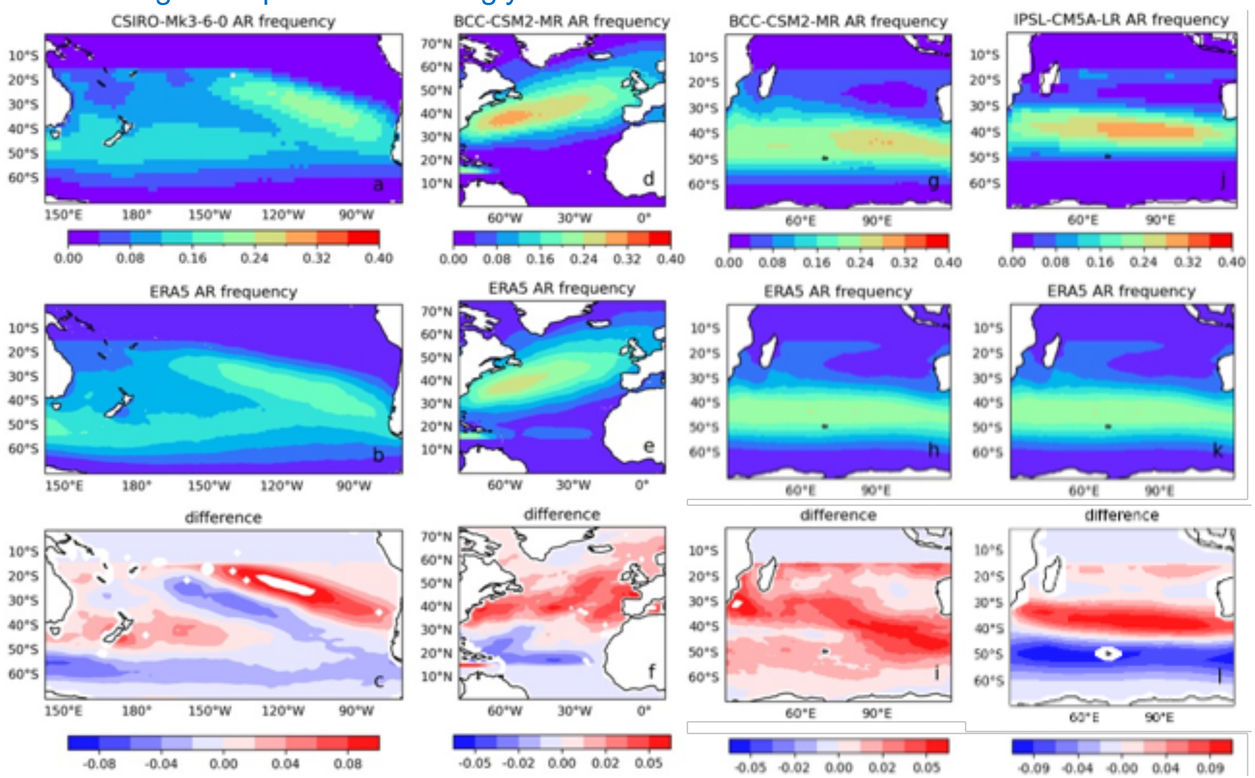
Response to reviewer #2

This paper presents a suite of new atmospheric river (AR) metrics that are designed for quick analysis of AR characteristics and statistics in gridded climate datasets such as model output and reanalysis. The study is very interesting, well organized and written. This work could be published if the following comments are adequately addressed. The author should show more convince evidences for their method robust. For example, for figure 3, what about AR frequency in the South Pacific for BCCCSM2-MR and ERA5, and their differences? What about AR frequency in the North Atlantic for CSIRO-MK3-6-0 and ERA5, and their differences? In addition, what is the advantage of the present method, compared with others'? This should be discussed.

Thank you for your review and comments.
Per your suggestion, we have added a few more figures comparing AR spatial frequency for different models and regions, e.g., AR frequency in the South Pacific for BCC-CSM2-MR and ERA5, and their differences, and AR frequency in the North Atlantic for CSIRO-MK3-6-0 and ERA5, and their differences.



In general, the difference is less prominent than those shown in Fig. 3, e.g., AR frequency in the North Atlantic for BCC-CSM2-MR and in the South Pacific for CSIRO-MK3-6-0 show visible latitudinal shifts compared to ERA5. We've included this figure in the supplementary material. In addition, we made diagnostic plots for the BCC-CMS2-MR model and IPSL-CM5A-LR models over the Indian Ocean, where the pattern correlation shows the largest contrast. Fig. 3 is updated accordingly:



In the discussion we also added: " Another example is the AR frequency distribution over the Indian Ocean for BCC-CSM-MR (Fig. 3g-i) and IPSL-CM5A-LR (Fig. 3j-l) models. Even though, when compared to ERA5, both models show significant spatial correlation in Fig. 2 (r=0.99 and r=0.82 respectively), the spatial bias pattern in IPSL-CM5A-LR exhibits a more apparent latitudinal shift than in BCC-CSM-MR."

Regarding the advantages of our AR metrics package and methods, we are collectively applying a suite of diverse ARDT methods and proposing novel metrics that have not been presented or widely used in previous literature. For example, we use the IoU metric to measure temporal consistency of AR occurrence for different models and ARDTs and use a new method to estimate effective sample size for spatial correlation. It is also the first AR metrics package that incorporates AR peak day calculations, so that users can easily quantify AR landfall seasonality. We have added a paragraph at the end of section 2.2: "One of the advantages of this metrics tool is that it can take data with different resolution, domain (e.g., a list of data files with mixed global and regional spatial extent), and coordinate system (e.g., 180° or 360° longitude coordinates; monotonically decreasing latitude coordinates), significantly saving users' time and efforts preparing the input data files. It is compatible with some non CF-Compliant NetCDF files (e.g., files with latitude coordinate names of "x_coords"). It is also somewhat "intelligent" in handling imperfect data (e.g., data file with corrupted data values, or with incorrect datetime calendar type)."

We have also revised the introduction to emphasize the need for an AR metrics package: "However, as ARDTs are usually designed with targeted research questions in mind, the analysis workflow and codes from one study are not easily adapted to other studies that use different ARDTs. Consequently, studies like intercomparison of ARDTs, or analysis based on an ensemble of ARDTs cannot be efficiently pursued without extensive collaboration or community efforts. Additionally, research of this kind cannot be easily repeated or updated when newer versions of ARDTs have been developed, or newer observational data products become available. As such, a universal analysis framework that is independent of ARDT can help accelerate progress in the AR research community."

And "… routine evaluation of ARs during the model development lifecycle requires a quantitative climate data assessment evaluation workflow that supports comparison between AR characteristics from different ARDTs. We believe progress in improving our understanding of ARs and their impacts could be accelerated with a dedicated tool for calculating AR statistics and evaluation metrics in climate models and gridded data products. Preferably, such an analysis tool is expected able to be comfortably and effortlessly applied to multiple data sources in one job (such as observations, forecasts, reanalysis and across different models), so that users do not have to worry about the inconsistency of format, coordinate system and spatial coverage among different datasets."

We have modified the text throughout the manuscript to emphasize the strength of our AR metrics package.