



Questions & Answers Part 2

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Amita Mehta (amita.v.mehta@nasa.gov) or Sean McCartney (sean.mccartney@nasa.gov).

Binder Hub Contact (for access after September 19):

Please leave your name and email address here:

<https://www.surveymonkey.com/r/HWT7GG7>

Question 1: Why are we using Binder? I understand it is a preinstalled environment, but it doesn't come with additional computing (to my knowledge), so will the binder work on all hardware, and are there any other dependencies we should be aware of before running the installed binder?

Answer 1: All of the computing is running remotely on a cluster on AWS. Access is dependent on your web browser.

Question 2: Why are there only 10-12 models available? What about the rest of the models (total is about 35)?

Answer 2: Those models are available from the intake catalog only. In the future we plan to make all 35 available.

Question 3: How do you calculate for other season definitions like DJF, MAM, JJA, and SON?

Answer 3: There is a parameter (season) you can change in the driver script where you can change the ECCDI. We do not provide an option for customized seasons.

Question 4: How do you use this data for a hydrological model like the Soil and Water Assessment Tool (SWAT)?

Answer 4: SWAT model [weather/climate inputs](#) require temperature, precipitation, humidity, wind speed and surface radiation components. These parameters are available from GDDP-CMIP6. You can spatially and temporarily extract these data for the region of interest and provide as input to SWAT.



**Question 5: Can we access the Jupyter notebooks for examples and practice?
What do we need to do to extend the access to Binder Hub?**

Answer 5: On the homework and exercise page, you can request access after September 19, 2025. The link is in the homework and exercise.

Question 6: If we need to evaluate the model with a local monitoring system (for example, a local model or climate data register), how can we integrate without generating an oversaturation/bootstrapping?

Answer 6: You will have to download the data as well as the script.
The script can be run on your local computer.

Question 7: Why are shapefiles used instead of the more modern single file formats, like GeoJSON or Parquet? Shapefiles are fine, but the file management is more effort, and if it's only used to clip the data from the model the precision of the coordinates can be low.

Answer 7: GDDP data has a resolution of 25 km and it is special for NEX-GDDP-CMIP6 data. We do recognize potential limitations.

Question 8: What is the coordinate system of the "shapefile" and the "model", how are we ensuring that the right area is extracted?

Answer 8: We are plotting and detecting to see if the data is recognized over your application area. WGS84 can work.

Question 9: Is there a way to pull probabilities from these models rather than what I understand to be specific estimates? For example, what is the 1% annual probability extreme event (T or pr) in the future?

Answer 9: There are probabilities that were not calculated in today's examples. You should be able to calculate them in your own time.

Question 10: Are the models bias corrected, if so what is the process for producing bias corrected data?

Answer 10: As discussed in the previous session, they are bias corrected and disaggregated using the daily BCSD statistical downscaling method.

Question 11: Which uncertainty methods are suitable for uncertainty analysis?

Answer 11: This may be dependent on application. If this is in reference to GDDP, you could compare them to the reference dataset as discussed in Part 1.



Question 12: I think there is an error in the "take screenshot" button in the Dashboard. I am not able to take the needed screenshot with different model selection. It is downloading the same screenshot everytime I try to click the button. Can you recheck it once?

Answer 12: It should save it as a .PNG on your computer. We will look into this further.

Question 13: What does the stats table show in the country scale?

Answer 13: It shows the summary statistics of the values plotted on the map.

Question 14: I have used some workflows using GitHub's codespace. Can I download the given workbooks and work locally with Anaconda?

Answer 14: Yes! On the main page of GitHub, you can find all of the libraries to run the script: <https://github.com/aashishpanta0/nasa-arset-training/>

Question 15: On the AWS S3 bucket, I can see the GISS-E2-1-G folder, but not the ECCDI folder. Is there another location where I have to access to download those files?

Answer 15: The AWS bucket only provides the original NEX-GDDP-CMIP6 data. The ECCDI data is in a shared folder on the binder hub.

Question 16: Do the GDDP data we used in this workshop have a resolution of 25 km by 25 km?

Answer 16: Yes, the GDDP data we used in this workshop is 1/4-degree lat/lon, which is approximately 25 km x 25 km.

Question 17: Would it be better to use the Mann–Kendall test and Sen's slope estimator instead of a simple linear regression for analyzing the trend of the ETCCDI indices?

Answer 17: Yes, since there could be several extreme values.

Question 18: Which python version can run the requeriments.txt? There are many conflicts for dependencies; for example, if I use the python version 3.13.1.

Answer 18: The binder was running on 3.10 but it should be able to run on a newer version.



Question 19: Can we run these scripts locally or just using the binder you have provided?

Answer 19: You can run these scripts locally. However, you will need to install all of the packages and source all of the data to be able to do so.

Question 20: I noticed that the analysis focuses mainly on temporal trends of the ETCCDI indices (time series and linear regression). Is there also a plan to analyze the spatial distribution of these trends, for example through grid-based or regional trend maps?

Answer 20: This can be something to look into for a potential publication.

Question 21: Do you have any suggestions for scaling these types of analysis in a python environment. i.e., if trying to run many (100s) of regional analyses? Are there any libraries you would suggest for parallel processing of these types of analysis?

Answer 21: I suggest using XArray and DASK. You will have to have the compute set up for that.

Question 22: Are we able to use this locally without downloading the data? Do we need some kind of user or key to access it without downloading it?

Answer 22: If you are using the catalog, you do not, but if you are using NEX-GDDP-CMIP6 data, you will have to download it. The data will be up permanently, but the Binder will be taken down on September 19.

Question 23: Hi and thanks for good presentations. One question: do you have any tips for people that would like to use an R environment for some of the analysis presented today regarding ETCCDI?

Answer 23: We can support this since we have used this type of analysis using R in the past. Contact us for more information.

Question 24: Do you have a script for ensemble-based analysis?

Answer 24: Not yet due to time and storage constraints. You are able to do so in your own time however. All of the models are downscaled at the same gridpoints.

Question 25: Is the historical NEX-GDDP-CMIP6 downscaled? If yes, is it possible to get the data used for downscaling the historical NEX-GDDP-CMIP6?



Assessing Extreme Weather Statistics using NASA Earth eXchange Global Daily Downscaled Projections (NEX-GDDP-CMIP6)

September 10 & 17, 2025

Answer 25: Yes, the CMIP6 historical experiment was also downscaled for each model. If you refer back to the previous session, you will be able to find the data used for downscaling as well as in GEE.

Question 26: Does the cell 5 of the ETCCDI_pr_indices_driver.ipynb file take a long time to run?

Answer 26: On the server, it will take longer versus running the program locally on your computer due to resource limitations. As the dataset is also very large, it will naturally take longer.