Report on Capacity Development Workshop on Python Programming Language

Naivasha, Kenya

Dates: 29th May – 2nd June 2023

Venue: Naivasha, Kenya

1 EXECUTIVE SUMMARY

The Python program training targeting ICPAC staff, was held from 29th May to 2nd June 2023, in Naivasha, Kenya. The main objective of the training is to improve the capacity of ICPAC staff in the use and applications of Python programming for hydrological and climate forecast modelling mainly the Climate into Useful Water And Land Information in Drylands (CUWALID) hydrological model developed under the Down2Earth project. Python is one of the leading programming languages in the world with many climate scientists shifting to Python from NCL and other programming languages. This massive shift is occasioned among others by Python's simplicity compared to other programming languages.

The training was funded by Down2Earth project and facilitated by University of Nairobi-Kenya, Space Science and Geospatial Institute (SSGI)-Ethiopia and ICPAC. The training lays the foundation for use and applications of Python programming for hydrological and climate forecast modelling.

The facilitators with background in data analytics climate modelling and computer science were key resource persons that facilitated the training. Preliminary skill assessment of ICPAC staff on python language was done to guide on training needs. The training covered a wide range of topics, including data cleaning, data wrangling, data visualization, and statistical analysis. Individual and group exercises formed part of the training activities to assess the level of staff mastery of concepts.

The participants appreciated the training in terms of overall theoretical and practical contents and the excellent facilitation and organization. As part of sustainability of knowledge acquired, ICPAC staff regularly meets to assess the progress level and share common challenges arising from python practice. This will include improving and sharpening skills on improving data processing, analyses and visualization techniques.

TABLE OF CONTENTS

2 INTRODUCTION

ICPAC together with a consortium of thirteen other institutions received funding to implement a four-year European Union H2020 research project (Sept 2020-August 2024) known as DOWN2EARTH in the Horn of Africa Drylands (HAD) targeting Kenya, Ethiopia and Somalia. The specific project sites are Isiolo County in Kenya, Meiso in Ethiopia and Odwayne District in Somaliland. The project is envisaged to address the multi-faceted challenges of water scarcity and food insecurity under climate change in HAD, by facilitating community-centered adaptation and resilience to climate change impacts.

ICPAC has direct responsibilities and contribution to all work-packages including e.g., WP3 on 'Improvement of regional seasonal forecasts to support warning systems on food security and water resources', WP4 on 'Regional climate adaptation and resilience through capacity building, informed decisions, and policy', and WP6 on 'Project impact through communication and dissemination'.

Hydrological model outputs and climate data analysis requires tools that are easily accessible, understandable and effective. In this regard, Python is useful because it removes data errors, filling missing values and transforming the data into a format that is easy to work with. It is also useful to perform statistical analyses such as identifying trends, patterns and relationships in the data. Furthermore, it can be used to communicate the findings of the analyses using various visualization options. Python data analyses tools are also extensible, open source and actively developing.

2.1 RATIONALE FOR THE TRAINING

ICPAC with other partners are working in many directions to develop tools and knowledge that will contribute to mitigate the impact of climate hazards (drought/floods) under Down2Earth project.

The D2E project is developing a dryland-specific hydrological forecast model (CUWALID) that support decision making at multiple levels in society. The model currently under development runs on Python programming language. The Python language is a general-purpose and high-level programming language that has been widely used in many applications including data science, data analysis, hydrological and climate forecast models. Nowadays the demand for Python programmers is increasing, there are multiple programming languages that are optimized by companies, but they prefer to use python language.

The ultimate aim is to have modelling infrastructure embedded within ICPAC operational forecasting system. Regional NHMS and other users would finally utilize the information from CUWALID products. ICPAC's knowledge and resident expertise on Python programming language will be an asset in the use and application of the CUWALID model for impact forecasting across key sectors.

An initial engagement in python refresher workshop focusing on improving the skills of key ICPAC's staff from climate, Hydrology and impact sectors was conducted.

2.2 OBJECTIVE OF THE TRAINING

The main objective of the training is to improve the capacity of ICPAC staff in the use and applications of Python programming for hydrological and climate forecast modelling.

2.2.1 Specific objectives:

I. To introduce the basics of Python programming. This included topics such as variables, data types, operators, control flow, and functions

II. To teach the participants how to use Python to manipulate and analyze data. This included topics such as loading, cleaning, visualization, and statistical analysis

III. To demonstrate how Python can be used to develop state of the art visualization and presentation of hydrological and climate forecast model outputs.

3 TRAINING METHOD

In order to acquire knowledge and improve skills of ICPAC's staff in Python programming applications in the field of hydrological and climate forecast, various techniques for training were used to involve participants in learning. The training used various sessions which involved theoretical and practical sessions. The theoretical sessions were led by experienced Python programmers and hydrological and climate modelers. The sessions covered the basics of Python programming, as well as the use of Python for hydrological and climate forecast modelling. The session was interactive, with plenty of opportunities for the participants to ask questions and participate in discussions. Practical session gave participants the opportunity to practice what they had learnt during the theoretical sessions. The participants worked on a variety of exercises and projects, using Python to manipulate and analyze data, develop

visualisations for the outputs from the hydrological and climate forecasts models. The training used a variety of active learning methods to engage the participants and help them learn more effectively. These methods included discussions, question and answer sessions, exercise and practical data analyses. Some of the examples of visualisations and code snippets are shown below.

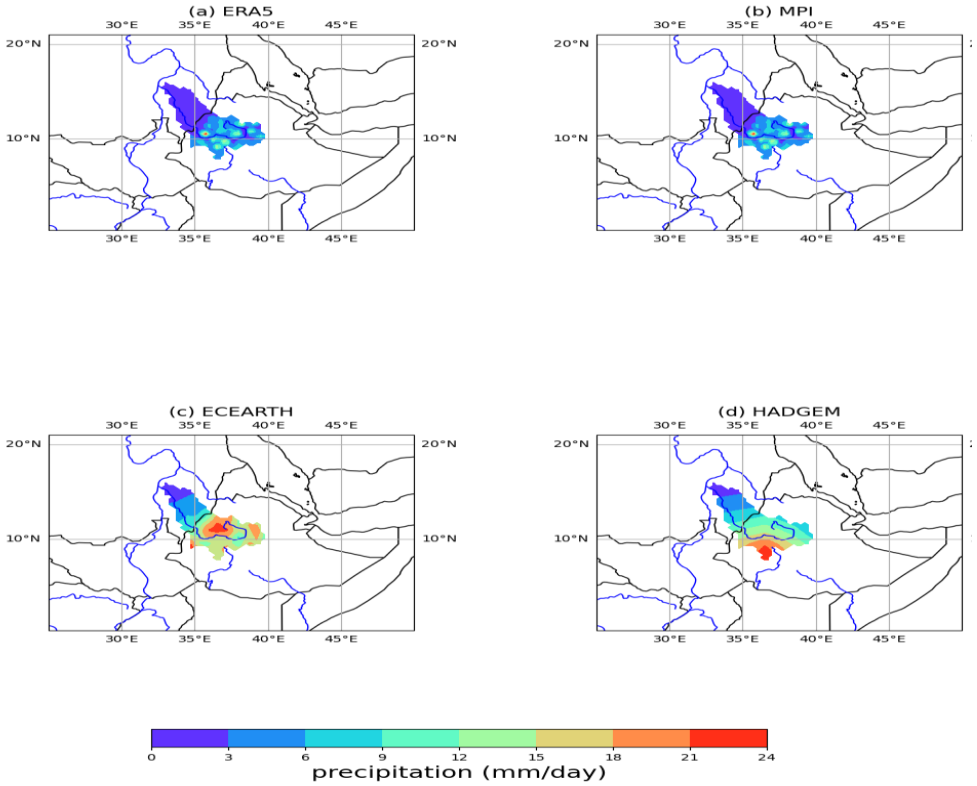

Figure 1. Comparison of CMIP6 models precipitation at the Blue Nile watershed area (the script name from the github repository is cmip6_precip_comparison_watershed.py)

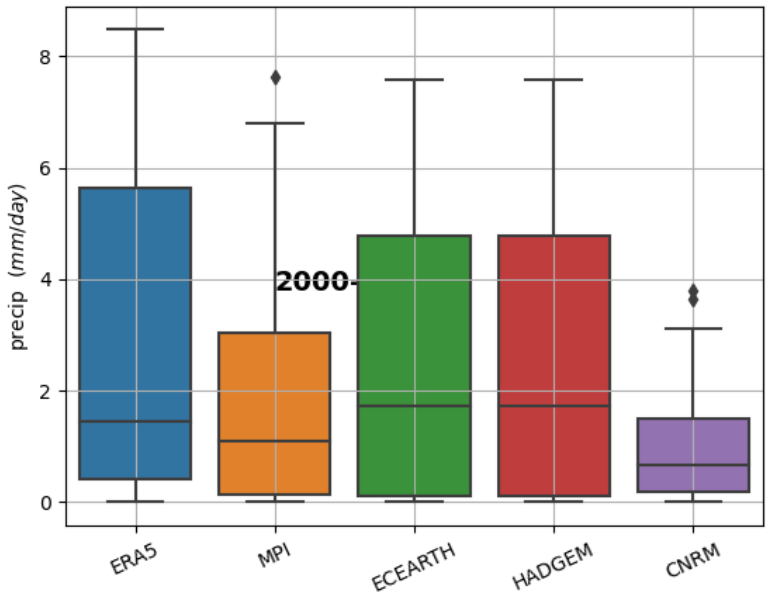

Figure 2. Boxplots of the CMIP6 models at the Blue Nile watershed shown in Figure 1 above (the script name from the github repository is cmip6_precip_comparison_boxplots.py).

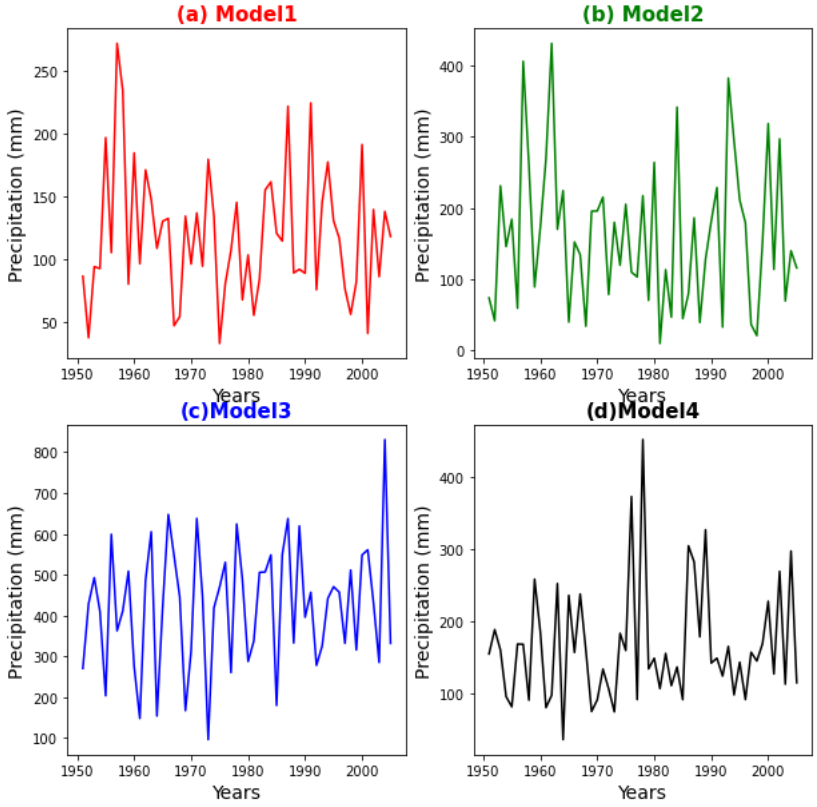

Figure 3: Inter-annual variability of rainfall (An example of Data Analysis based on Python using Pandas)

```
## Our list is of length...

print('The list is of length:', len(all_df))
```

```
The list is of length: 4
```

```
## Loop for analyzing multiple elements in a list

out_dat = [None] * 4                ## Initialize a list of NoneType

for i in [0, 1, 2, 3]:
    ## Convert the Dates column to Pandas Datetime object and make that column an index

    all_df[i]['Dates'] = pd.to_datetime(all_df[i]['Dates']);all_df      ## Convert to a DatetTime object
    df = all_df[i].set_index('Dates')                                   ## Setting the 'Dates' column as an index

    ## Select the months in March to May season using a condition

    df_mam = df.iloc[((df.index.month >= 3) & (df.index.month <= 5))]

    ## Group by year and get the annual MAM rainfall totals. Use mean for temperature

    ann_mam = df_mam.groupby(df_mam.index.year).sum()

    out_dat[i] = ann_mam

## Check the contents of the final output

out_dat
```

Figure 4: Scripting for computation of seasonal means (Pandas) and Multi-panel plots (matplotlib.pyplot), (example of plotting exercise)

4 RESOURCE PERSONS

The team that delivered sessions during the training comprised of 2 International experts and 2 internal experts. A brief description of the trainers along with their sessions conducted by them is provided in Annex3.

5 OUTPUT/OUTCOMES

- Various Python concepts and libraries (netCDF manipulation files, xarray, pandas, rasters, numpy, and shell scripting) have been understood
- Visualization and Data Analysis Plots have been exercised.
- File handling and exception handling techniques implemented.
- Operators, sets, methods, and dictionaries understood in-depth.
- Commands for conditional statements, loops, and functions understood.
- Basic python skills required for the use and applications of CUWALID and in hydrological forecast and climate models obtained.

6 ACCESS TO TRAINING MATERIALS

The workshop materials and notebooks are stored in this github, repsitory https://github.com/icpac-igad/python-workshop. For a hands-on training on git and github, participants are encouraged to clone the repository by fork and update their changes in their respective source version.

7 DISCUSSION AND WAYFORWARD

- The facilitators ensured everyone moved at the same pace. The expansion of the training to other partners is an excellent idea to expand the institution's capacity of staff. This was a well thought out by D2E project
- The python is going to help CONFER project in development of Machine learning systems; its noble idea for D2E to organize such as training, it’s a big boost and transform the programming language at ICPAC.
- Time constraints did not allow us to exhaust everything as planned in the programme.
- The trained participants need to perfect their skills on python so as to become ToT for member states training planned in Feb 2024.
- The key issue is to see how this training helps us perfect data analysis and possible publication.
- More resource people need to be exposed to python programming, improve ICPAC services, and joint publications within institutions.

8 LESSONS LEARNT AND BEST PRACTICES:

The changes in the format of the workshop helped keep it moving smoothly. More interactive sessions and practical exercises were ensured by the trainers, which helped participants underrate the training contents. As the exercises were individual activities, many of the session took more time than allocated. If next time such activity will be planned more time should be allocated to practical sessions.

9 PLAN FOR PYTHON FOLLOW-UP ACTIVITIES

- Bi-weekly meeting on python solving quizzes is seen as recommendation**;** This will boost in transition from other languages that are getting deprecated
- Second phase of the training to consider machine learning amongst others
- Continuous and engagement on python & networking through WhatsApp and MS-teams.

10 ANNEXES

Annex1: List of Training Participants

#	Name	Designation
1	Abebe Tadege	PI-D2E- Climate
2	Khalid Hassaballah	Hydrologist
3	George Otieno	Climate
4	Titike Bahaga	Climate
5	Herbet Misiani	Data (Resource person)
6	Nishadh Kalladath	Data Science specialist (Resource person)
7	Igbal Salah	Groundwater specialist
8	Masilin Gudoshava	Climate
9	Paulino Omay	Climatological specialist
10	Joseph Kiragu	Finance
11	Victoria Minayo	Logistic
12	Elias Maiyo	Driver
13	Philip Omondi	Climate
14	Eunice Koech	Climate
15	Ismael Lutta	Data
16	Erick Okoth	Admin assistant
1	Gemechu Garuma	SSGI (resource person)
18	Elisha Abade	UoN (resource person)

Annex2: Training programme:

DAY 1 –May 29th	
8:00 – 8:30 am	Arrival of participants and registration ***(Victoria Kidiavai)***
8:30 – 9:00 am	Opening session (introducing participants, training objectives and expectations...etc) ***(Mr.Abebe Tadege, Dr.Khalid Hassaballah, Dr.George Otieno)***
9:00 – 10:30 am	• Installation of Linux operation system on Windows ***(All)*** • Installation of Anaconda ***(All)*** • Introduction to Linux operation system ***(Dr.Gemechu Garuma)*** • Installation of Github ***(All)*** • Introduction to version control system- Git, hands on training on git, git add, commit, push, introduction to github, gitlab Exercise 0: Make a repository and push your code to github ***(Dr.Nishadh Kalladath)***
10:30- 11:00 am	**Tea break**
11:00 – 1:00 pm	• Introduction to general programming syntax and Getting started with Python, on Zen of Python, Reading Python codes and understand Python syntax ***(Dr.Elisha Abade, Herbert Misiani)*** • Introduction to Jupyter notebook, docker, Python library package systems, anaconda, pypi. Background of these tools and its necessity, using JupyterHub. Read a Python library or project in github and discussion on Python for self-study, using documentation, use of Visual Studio Code and IDE for reading and understanding the Python codes from basics to moderate complex with more focus on Jupyter notebook ***(Dr.Nishadh Kalladath)***
1:00 – 2:00 pm	**Lunch break**
2:00 – 4:00 pm	• Python Basics: Python data structures; lists, tuples, and dictionaries ***(Dr.Gemechu Garuma)*** • Reading different data types in Python (including text/csv, excel) and writing simple codes to analyze and plot time series data. Indexing in Python. ***(Dr.Gemechu Garuma)*** ***Exercise 1: Read data, do basic arithmetic and plot (line, histograms etc)***
4:00- 4:30 pm	**Tea break (end of the day)**

DAY 2 – May 30th	
8:30 – 8:45 am	Wrap-up of day 1 (interactive)
8:45- 10:30 am	• Reading CSV files, Harness the power of for loops, conditionals and functions ***(Dr.Gemechu Garuma)*** • Reading NetCDF files using xarrays and pamdas ***(Dr.Gemechu Garuma)***
10:30- 11:00 am	**Tea break**
11:00 – 1:00 pm	• Use climatological daily/monthly rainfall data and make box whisker plot for your area of interest (single and multiple files). ***(Dr.Gemechu Garuma)*** • Demonstration on the task with Jupyter notebook, code walk through and elaboration on the methods libraries used in the code ***(Dr.Gemechu Garuma)***
1:00 – 2:00 pm	**Lunch break**
2:00 – 4:00 pm	Participants working on with the task of the day ***(Dr.Nishadh Kalladath)*** Brainstorming points and code snippets related to the task of day: 1. Difference of vector and raster 2. Using vector to subset raster 3. Summarizing the raster into data frame ***(Dr.Elisha Abade***) 4. Climate science data format, CPT, netcdf, zarr ***(Dr.Elisha Abade***) 5. Using the cloud objects from zarr rather than full download CF convention/metadata standards ***Continuing work on the task of the day, code review and conclusion***
4:00- 4:30 pm	**Tea break (end of the day)**

DAY 3 – May 31st	
8:30 – 8:45 am	Wrap-up of day 2 (interactive)
8:45- 10:30 am	Task of the day: Befriend with Data array and Dataframe!, calculate SPI using climate indices library and Impact based forecasting using climada library ***(Dr.Nishadh Kalladath, Herbert Misiani)***

	Demonstration on the task with Jupyter notebook, code walk through and elaboration on the methods libraries used in the code ***(Dr.Nishadh Kalladath, Herbert Misiani)***
10:30- 11:00 am	**Tea break**
11:00 – 1:00 pm	Participants working on with the task of the day Brainstorming points and code snippets related to the task of day: 1. Introduction to pandas dataframe and xarray data array ***(Dr.Elisha Abade***) 2. Dataframe into raster and map plotting 3. Formula computation on dataframe, climada source code read 4. Algorithms apply on dataframe column, the powerplay with apply and lamda methods ***(Dr.Gemechu Garuma)***
1:00 – 2:00 pm	**Lunch break**
2:00 – 4:00 pm	***Continuing work on the task of the day, code review and conclusion (All)***
4:00- 4:30 pm	**Tea break (end of the day)**

DAY 4 – June 1st	
8:30 – 8:45 am	Wrap-up of day 3 (interactive)
8:45- 10:30 am	Task of the day: Hello to CPT from python! Demonstration on the task with Jupyter notebook, code walk through and elaboration on the methods libraries used in the code ***(Herbert Misiani, Dr.Nishadh Kalladath)***
10:30- 11:00 am	**Tea break**
11:00 – 1:00 pm	Participants working on with the task of the day Brainstorming points and code snippets related to the task of day 1. Reading CPT file in python, simple text file read in python 2. Introduction to regex and reading specific headings of CPT

	3. Plotting the CPT, introduction to matplotlib and cartopy 4. Subplot- multi panel mapping/ plotting, animating the maps 5. Mapping template for GHACOF country presentation and way forward to use Python in coproduction workshop ***(Dr.Nishadh Kalladath)***
1:00 – 2:00 pm	**Lunch break**
2:00 – 4:00 pm	***Continuing work on the task of the day, code review and conclusion***
4:00- 4:30 pm	**Tea break (end of the day)**

DAY 5 – June 2nd	
8:30 – 8:45 am	Wrap-up of day 4 (interactive)
8:45- 10:30 am	Task of the day: Exploration into reproducible research and recreating the plots from published studies. Suggested studies 1. Streefkerk, Ileen N., et al. "A coupled agent-based model to analyse human-drought feedback for agropastoralists in dryland regions." *Frontiers in Water* 4 (2023). 2. Nobre, Gabriela Guimarães, et al. "Forecasting, thresholds, and triggers: Towards developing a Forecast-based Financing system for droughts in Mozambique." *Climate Services* 30 (2023): 100344. Setting the stage, procuring the data, code in case of availability, plan the steps, setting the goals for the day ***(All)***
10:30- 11:00 am	**Tea break**
11:00 – 1:00 pm	Participants working on with the task of the day ***(all)***
1:00 – 2:00 pm	**Lunch break**
2:00 – 4:00 pm	(5 minutes presentation from 2 groups) ***(all)*** ***Continuing work on the task of the day, code review and conclusion. Recap of the workshop and wayforward***
4:00- 4:30 pm	**Tea break (end of the day and the training)**

Annex3: Resource Persons' profile

#	Name	Designation	Session delivered
1	Dr. Elisha Abade	Data Sceince and Machine Learning Expert, Department of Computing and Informatics, Faculty of Science and Technology, University of Nairobi, Kenya.	Python data structures (lists, tuples and disctionnaries), pandas, geopandas, cartopy, xarray, numpy, etc.
2	Dr. Gemechu Fanta	Climate Science Expert, Space Science and Geospatial Institute (SSGI), Addis Ababa, Ethiopia	Python data structures (lists, tuples and disctionnaries), pandas, geopandas, cartopy, xarray, numpy, etc
3	Dr. Nishadh Kalladath	Data Sceince and Machine Learning Expert, NORCAP	Verision Control System, Git, Github, how to train your computer for second brain, For loop, SPI calcualtion, Use of IDE for Python coding
4	Mr. Herbert Misiani	Data Management Expert, ICPAC	Data analysis using Pandas