



# FDRI Accessing gridded time series data with Python and R Notebooks

May 2026

Kit Macleod, Matt Fry, Mike Brown, Matt Dalle Piagge, Clare Jones, Anna Rose Klaptocz, and Faiza Samreen

Executive summary .....	3
<b>1. Background and purpose .....</b>	<b>4</b>
1.1. Floods and Droughts Research Infrastructure (FDRI).....	4
1.2. Rationale for this feedback session .....	4
1.3. Participants .....	4
1.4. Objectives .....	5
1.5. Gridded data, data formats and notebooks.....	5
<b>2. Session outline.....</b>	<b>6</b>
2.1. Activities and their purposes.....	7
<b>3. Results and discussion .....</b>	<b>7</b>
3.1. Summary of the participants and our session.....	7
3.2. Awareness and use of gridded data formats and UK datasets .....	8
3.3. Awareness and use of computer languages to access and explore data.....	9
3.4. Awareness and use of notebook and related technologies.....	9
3.5. Understanding user needs for gridded datasets and notebooks .....	10
<b>4. Next steps.....</b>	<b>11</b>

# Executive summary

As part of our development of the Floods and Droughts Research Infrastructure (FDRI) digital offering, in April 2026, we held an interactive feedback session with key users on gridded time series datasets and notebooks to aid access. For this session, we defined 'key users' as UK hydrological researchers, UK hydrological consultants, and other technical experts in organisations that span UK and devolved governments and their agencies.

We designed an interactive feedback session to meet three objectives:

1. Give participants an understanding of FDRI's proposed offering for notebook-based access to existing gridded datasets (see Section 1.5)
2. Understand user needs for gridded data and notebook-based access
3. Collect feedback on FDRI's proposed approach to notebook-based access

Over 80 people registered for the session and 46 attended on the day, with 36 taking part in the interactive surveys (not answering every question). Our key findings were:

- There was greater awareness and use of NetCDF and text-based formats than Zarr format.
- ERA5 was the most frequently used dataset for gridded data analysis followed by HadUK-Grid.
- The most common pain points with gridded data formats and datasets were related to the size of datasets and storage, inconsistencies in data, and a lack of technical skills.
- Participants showed a strong preference for Python and R. At registration, 25 said they preferred to use Python, 20 said they preferred to use R, and 15 said they used both.
- We asked about their use of notebooks and related technologies, 22 responded, with Jupyter-Python (11) being used sometimes or frequently followed by R Markdown (six). With only six aware of Jupyter-R.
- Google Colab was the most widely used platform for running Jupyter notebooks. Use of other platforms such as JupyterLite and Binder was very limited.
- All eight respondents said they would like notebooks for each dataset. With five wanting NetCDF and Zarr support. Only three wanted more existing gridded datasets in Zarr format.
- We asked what they would like a notebook to contain to demonstrate a gridded dataset, and nearly all agreed with our suggestions to include e.g. some form of an introductory overview at the start, how to open a dataset, how to explore the metadata and how to plot the dataset.
- When we asked what notebook features were important for exploring datasets (24 respondents) 20 said they agreed that they need to run on their local computer, 19 said run in the cloud, and 19 said be easy to use.

These findings will directly inform the next phase of our work on the provision of gridded data and notebook-based access. Priorities include refining notebook design based on user feedback, expanding Zarr support and building a community resource for shared notebooks. Further engagement opportunities will be made available through the FDRI newsletter and website.

# 1. Background and purpose

## 1.1. Floods and Droughts Research Infrastructure (FDRI)

Floods and droughts cause significant economic, social, and environmental impacts and they are predicted to increase in intensity, frequency, and duration in response to changes in climate and human activity. The need for new science to underpin the UK's preparedness and resilience to these extreme events has never been more pressing. The Floods and Droughts Research Infrastructure (FDRI) project is a capital investment by the UK Government enabling essential science and innovation to improve the country's resilience to hydrological extremes. It will enable focused observations of water input, movement, and storage in three river basins, one in England, Wales and Scotland, with mobile instruments available for UK-wide deployment. A significant focus for FDRI will be the delivery of this new observational data, alongside other UK-wide hydrological monitoring data, to users via a new digital infrastructure for hydrological research.

## 1.2. Rationale for this feedback session

To ensure FDRI's digital and data products meet the needs of users, we've run a series of feedback sessions throughout the project's implementation phase. In January 2024, we held two collaborative workshops with 50 individuals on 'Enabling reproducibility in hydrological research'. In May 2025, we held two online interactive workshops, with 15 individuals each, on 'Digital user interfaces and experiences'.

This session builds on these previous engagement activities, with the aim of gathering user feedback on the provision of gridded time series datasets and accompanying data access notebooks.

We will continue to work with users to co-design and iteratively refine FDRI's data and digital products.

## 1.3. Participants

Target participants for this session were FDRI's key digital platform user groups UK hydrological researchers, UK hydrological consultants and other technical experts in organisations that span UK Government and its agencies.

Participants were recruited through a combination of UKCEH social media and mailing lists (including the FDRI newsletter and the British Hydrological Society Circulation). During sign-up,

participants were asked a series of questions, including about their preferences for accessing data using Python, R or other computer languages/tools, to give session organisers a strong idea. Over 80 people registered for the session and 46 attended on the day, with 36 taking part in the live surveys (Appendix 1 contains a list of organisations represented).

## 1.4. Objectives

Our objectives were to:

1. Give participants an understanding of FDRI's proposed offering for notebook-based access to existing gridded datasets
2. Understand user needs for gridded data and notebook-based access
3. Collect feedback on FDRI's proposed approach to notebook-based access

## 1.5. Gridded data, data formats and notebooks

A range of hydrological use cases, across sectors, require finding and accessing increasingly large, gridded time series datasets. One of the most used types are meteorological data which include variables such as precipitation and air temperature.

Traditionally these gridded time series datasets have been provided in NetCDF format (Network Common Data Form). NetCDF is a self-describing binary data format and associated software tools for creating, sharing, and accessing array-oriented scientific data<sup>1</sup>. Active development of the CF Metadata Conventions (Climate and Forecasting) support NetCDF files through defining metadata that provides a definite description of what the data in each variable represents, and the spatial and temporal properties of the data.

Conventionally, these files are structured as a series of two-dimensional arrays for each timestep, often with separate files for a given period of time (e.g. month or year) which means they are optimised for rapid access to all data for that given period of time. However, many hydrological use cases for this data, e.g. extraction of a long time-series for a given location or catchment do not suit this optimisation, as data must be extracted from multiple files.

An example of a gridded meteorological dataset is HadUK-Grid<sup>2</sup>, it is produced annually at 1 km x 1 km grid resolution based on the Ordnance Survey's National Grid. The dataset is provided, as NetCDF files, at several spatial (1, 5, 12, 25, and 60 km) and temporal resolutions (daily, monthly, seasonal, and annual) to support a range of use cases.

---

<sup>1</sup> <https://www.unidata.ucar.edu/software/netcdf>

<sup>2</sup> <https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb/>

Over the past 10 years there has been development and increasing use of a new data format, Zarr<sup>3</sup>, that overcomes some of the limitations of single-file formats like NetCDF. The Zarr data specification sets out storing chunked, compressed, N-dimensional arrays designed for cloud-native, high-performance computing. In this way, Zarr enables these large datasets to be structured to optimise access for specific use cases, and for data to be directly accessible over the web. Our previous work investigated how Zarr-based data can be optimised for a wider range of use cases<sup>4</sup>. Increasingly meteorological and geospatial data are available in Zarr format. The recent rise of Zarr formatted data is due to widespread provision of cloud-based object storage for storing large datasets, and development of key packages for accessing cloud-based data in the Python community (such as Zarr-Python<sup>5</sup> and Xarray<sup>6</sup>). In the R community, there is less support, with the recent Zarr package<sup>7</sup> providing initial Zarr access since the end of 2025.

Computational notebooks let people run software code directly in a web browser, combining documentation, code, and graphics in a clear and easy-to-read format. This has made notebooks popular for tutorials and other presentational purposes, as well as for rapid prototyping. Jupyter notebooks are a popular notebook option, supporting a range of computer languages including Python and R.

A notebook is accessed through a web browser and runs compute either on your local machine (for example JupyterLite) or on a remote server. Many notebook services now exist, including platforms such as Google Colab<sup>8</sup> and Binder<sup>9</sup>. Commercial notebook services like Google's Colab provide free and paid plans, the later including more compute units and faster GPUs. For UK research within the remit of NERC, the NERC JASMIN Notebook Service<sup>10</sup> provides more compute than the free commercial offerings.

## 2. Session outline

We designed the session to be interactive using an online tool (Mentimeter.com) that allowed questions to be shared with participants and the visualisation of their responses (see Appendix 2); we also encouraged the participants to share their questions using the Q&A feature of Microsoft Teams, and to raise hands and share points verbally. The session had three main presenters with

---

<sup>3</sup> <https://zarr.dev/>

<sup>4</sup> [https://github.com/NERC-CEH/object\\_store\\_tutorial](https://github.com/NERC-CEH/object_store_tutorial)

<sup>5</sup> <https://zarr.readthedocs.io/>

<sup>6</sup> <https://xarray.dev/>

<sup>7</sup> <https://r-cf.github.io/zarr/>

<sup>8</sup> [https://developers.google.com/colab?\\_gl](https://developers.google.com/colab?_gl)

<sup>9</sup> <https://mybinder.org/>

<sup>10</sup> <https://help.jasmin.ac.uk/docs/interactive-computing/jasmin-notebooks-service/>

three additional colleagues to ensure variety in our delivery, and to reflect on and discuss participant responses to our questions.

## 2.1. Activities and their purposes

The two-hour session was designed to meet the three objectives, through a series of five activities. These were:

### 1. Introduction to the session

To make sure participants were aware of the session objectives and how it was being run, and opportunities for two-way engagement.

### 2. Overview of FDRI digital work

To bring all participants up to a similar level of understanding of the data and digital work in FDRI.

### 3. Learning about participants and their needs

To learn about the participants in terms of awareness and use of gridded time series data formats and datasets, and the pain points they have. Their awareness and use of computer languages to access and explore data. And their familiarity with notebook related technologies for accessing and exploring data.

### 4. Demonstration

To share our understanding of challenges related to accessing and exploring gridded time series data, and show examples of how innovations and best practices in data formats and notebook-based access are improving data access.

### 5. Learning about participant needs for FDRI gridded data and notebooks

To learn from the participant's use cases what aspects of the gridded data work was of most use to them including spatial and temporal aggregation of data. Explore what they would like from FDRI work related to gridded data and notebook-based access over the next 6-12 months.

## 3. Results and discussion

### 3.1. Summary of the participants and our session

We had 46 participants and 36 took part in our interactive responses. About half were from academia with the rest split equally between industry and public sector. Over 80 people had signed up to our session. Of the 46 participants, 17 stayed for the full two hours. The remaining 29

left at various points: five dropped out between 100 and 120 minutes, five left during the comfort break before the demonstration activity, five left between 30 and 60 minutes, five left between 10 and 30 minutes, and four stayed for less than five minutes.

We were happy with the level of engagement in relation to the interactive questions, with 20-30 participants answering our questions during the first session 'Learning about participants and their needs'. This reduced to about 10 during the second interactive session on 'Learning about participant needs for FDRI gridded data and notebooks'. This reduction and the limited verbal and Teams Q&A responses will help guide the design of future sessions.

Of the six participants that completed our evaluation survey, three were very satisfied with our organisation of the event, two were satisfied and one was neither satisfied nor dissatisfied. Three participants were satisfied that they were able to share their views and contribute to discussions, another one was dissatisfied and another very dissatisfied.

Participant comments were mainly positive, for example "thanks a lot, it was very informative" and "the event was informative and well-structured, particularly in introducing gridded data access through notebooks". A couple commented on the UK focus for example "the project is focused on Great Britain but should include guidance for participants in other regions".

## 3.2. Awareness and use of gridded data formats and UK datasets

We asked about awareness and use of gridded data formats. There was greater awareness and use (participants used sometimes or frequently) of NetCDF (14) and text-based formats (21) with only two having explored Zarr format and a further two who used Zarr data sometimes.

The participants (25 who answered this question) said they used ERA5<sup>11</sup> most frequently, with eight saying they used it sometimes or more frequently, followed by HadUK-Grid<sup>12</sup> (six), and CEH-GEAR<sup>13</sup> and UK MetOffice UKV<sup>14</sup> (both two). With none having explored the CHES-SCAPE<sup>15</sup> gridded dataset.

Examples of gridded datasets they used included ERA5 (five mentions) and CHIRPS<sup>16</sup> (three mentions). ERA5 is the European Centre for Medium-Range Weather Forecasting's ECMWF

---

<sup>11</sup> <https://climate.copernicus.eu/climate-reanalysis>

<sup>12</sup> <https://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb/>

<sup>13</sup> <https://catalogue.ceh.ac.uk/documents/fc9423d6-3d54-467f-bb2b-fc7357a3941f>

<sup>14</sup> <https://registry.opendata.aws/met-office-uk-deterministic/>

<sup>15</sup> <https://catalogue.ceda.ac.uk/uuid/8194b416cbee482b89e0dfbe17c5786c/>

<sup>16</sup> <https://www.chc.ucsb.edu/data/chirps3>

Reanalysis 5<sup>th</sup> Generation. CHIRPS stands for the Climate Hazards Group InfraRed Precipitation with Station data and is a 35+ year quasi-global rainfall dataset.

We explored pain points with gridded data formats and datasets, and the most common themes (based on 19 responses) were related to size of datasets and storage (eight), inconsistencies (six) e.g. in data structures and coordinate systems, lack of technical skills (three) e.g. to handle these large data sizes, and not enough metadata (two).

### 3.3. Awareness and use of computer languages to access and explore data

We learned about their preferences for programming languages to explore gridded data. Participants showed a strong preference for Python and R. At registration, 25 said they preferred to use Python, 20 said they preferred to use R, and 15 said they used both. Out of the 27 responses during the session then Python was used by 18 sometimes or more frequently, R by 13, and Matlab<sup>17</sup> by seven.

In terms of the packages/libraries they used to open and explore gridded datasets then Python packages were most frequently mentioned, with Xarray<sup>18</sup> (six), Pandas<sup>19</sup> (six), and Netcdf4<sup>20</sup> (five) being the most frequently mentioned. In terms of R packages, ncd4<sup>21</sup> was most frequently mentioned (three), followed by terra<sup>22</sup> (two), the latter being a general-purpose spatial data analysis package.

### 3.4. Awareness and use of notebook and related technologies

We asked about their use of notebooks and related technologies, 22 responded, with Jupyter-Python (11) being used sometimes or frequently followed by R Markdown (six). With only six aware of Jupyter-R, and three of them using it sometimes, and only two participants being aware of Quarto including one using it frequently.

Of the 22 respondents who answered the question about Jupyter notebook options, Google Colab<sup>23</sup> was the most widely used platform for running Jupyter notebooks, with 16 aware of it and

---

<sup>17</sup> <https://uk.mathworks.com/products/matlab.html>

<sup>18</sup> <https://docs.xarray.dev/en/stable/>

<sup>19</sup> <https://pandas.pydata.org/>

<sup>20</sup> <https://unidata.github.io/netcdf4-python/>

<sup>21</sup> <https://cran.r-project.org/web/packages/ncdf4/index.html>

<sup>22</sup> <https://cran.r-project.org/web/packages/terra/index.html>

<sup>23</sup> <https://developers.google.com/colab>

of these nine used it sometimes or more frequently. Use of other platforms such as JupyterLite<sup>24</sup> and Binder<sup>25</sup> was very limited, with only two participants having used either sometimes or more frequently.

### 3.5. Understanding user needs for gridded datasets and notebooks

All eight respondents said they would like notebooks for each dataset. With five wanting NetCDF and Zarr support. It was surprising that only three wanted more existing gridded datasets in Zarr format. However, this finding may reflect unfamiliarity with the format rather than a lack of interest. As shown in section 3.2, very few participants currently use Zarr, so it is unsurprising that demand for it was not strong.

We asked what they would like a notebook to contain to demonstrate a gridded dataset, and nearly all agreed with our suggestions to include: intro/purpose/overview text at the start, how to open a dataset, how to explore the metadata, how to plot data (timeseries), and how to plot data (spatial).

We explored the spatial and temporal aggregations of gridded datasets that best met their use cases, with 5 km (seven out of nine respondents), daily (10 out of 11) and monthly (7 out of 11) being most desired of the existing 1 km hourly datasets. When we asked what specific temporal and spatial groupings/aggregations of UK data were required, the limited (five) responses most frequently mentioned temporal aggregations of 15 minute, hourly, and daily, with no common use case in the responses.

When we asked what notebook features were important for exploring datasets (24 respondents) 20 said they agreed that they need to run on their local computer, 19 stated they wanted to run them in the cloud, and 19 said be easy to use. These three features were followed by three further features, with 11 agreeing that persistence of notebooks (storage), version control, and installed packages were important. Though six to eight participants strongly disagreed with the importance of this second set of features. This makes sense when you consider that the group included people with quite different levels of coding experience. For those with less relevant technical expertise, ease of use is what counts — they just want something that works without a complicated setup. FDRI notebooks should cater for this by being straightforward to use out of the box, while still offering more advanced options for users who want them.

In general, the 20 respondents agreed or strongly agreed with our suggestions for what a notebook could contain to demonstrate a gridded dataset: intro/purpose/overview text at the

---

<sup>24</sup> <https://jupyterlite.readthedocs.io/en/stable/>

<sup>25</sup> <https://mybinder.org/>

start, how to open a dataset, how to plot data (timeseries), and how to plot data (spatial). With 19 out of 20 agreeing or strongly agreeing to include: how to explore the metadata. And 16 agreeing or significantly agreeing with the need for code to show how to see a tabular view of the first rows, and only 13 agreed or strongly agreed there was a need for text on the purpose of each cell.

In relation to our question about what additional content they would want in a notebook to demonstrate and explore a gridded dataset. Then nine responded, with six indicating some form of tutorial material including how to see notebook content, how to use them, how to convert data, and how to export the data. Another suggestion was how to load and work with things in chunks. A couple of responses were related to similar datasets from other sources (maybe relates to earlier ERA5 and CHIRPS suggestions).

In our final question about what they would like from this FDRI work, then two of the five respondents said: API access, cloud optimisation, and ability to add/link datasets. It should be noted that this is a small number, and the responses may not be representative of the wider group. That said recurring themes included API access, cloud optimisation and the ability to add or link datasets which were consistent with the responses elsewhere in the session. One respondent said they wanted a start guide to demonstrate how to start using these types of data, and tutorials which would be useful for teaching students.

We asked if respondents would like to contribute notebooks to a FDRI community resource, and four out of the seven respondents said yes in six months or more, and three were unsure if they would.

## 4. Next steps

This report will be widely circulated through FDRI communication channels including the newsletter and shared with the participants. We will gain further feedback through multiple channels including the GitHub repository, updates in the FDRI newsletter, and follow up emails.

The findings from this session will inform the next phase of our work on the provision of gridded data and notebook-based access. To help address the low level of awareness and use of Zarr based versions of datasets we plan to provide additional Zarr based versions of key UK gridded meteorological data to support spatial (UK extent, one time point per chunk) and temporal (one year of data with small spatial tiles) use cases. We currently provide access to a Zarr version of hourly CEH-GEAR and getting started notebooks in Python and R. We will work towards Zarr versions of other UK datasets of interest (e.g. daily 1 km resolution HadUK-Grid rainfall). Once the Zarr datasets are ready and supporting Python and R notebooks have been created, we will advertise their availability through FDRI communication channels. We will consider providing additional tutorial material related to these datasets and data formats.

Our notebook design will incorporate the content elements most valued by participants (which supported our initial suggestions), including dataset overviews, metadata exploration, and timeseries and spatial plotting. We will continue to provide notebooks that can be run simply on Google Colab and other platforms. The findings around spatial and temporal aggregation preferences will inform decisions about which resolutions to prioritise in the initial release of gridded datasets.

A further engagement session is planned for September as part of the wider soft launch of the FDRI digital platform, which will allow participants to review and respond to updated notebook prototypes informed by this session's findings.

## Acknowledgments

The authors gratefully acknowledge the provision of data collected from the Floods and Droughts Research Infrastructure (FDRI) funded by the Natural Environmental Research Council (NERC-UKRI).

## Contact

Kit Macleod (UKCEH), [kitmac@ceh.ac.uk](mailto:kitmac@ceh.ac.uk)

# Appendices

## Appendix 1. Organisations represented in the participants who signed up to our event

National Oceanography Centre, UK

Stockholm University, Sweden

Rolls-Royce SMR

Tongji University, China

Universidad Nacional de Ingeniería, Peru

National Institute of Hydrology- Roorkee, India

Northumbria University, UK

PICON GmbH, Germany

Gebze Technical University, Turkey

Environment Agency, UK

Intertech, Turkey

Leibniz University Hannover, Germany

National University of Singapore, Singapore

University of Bejaia, Algeria

Kwame Nkrumah University of Science and Technology, Ghana

Rolls-Royce SMR

Lancaster University, UK

Jahangirnagar University, Bangladesh

Niras, UK

ADAS, UK

Ulster University, UK

AtkinsRéalis, UK

Anglian Water, UK

Edinburgh Napier University, UK

Northern Ireland Environment Agency, UK

University of Glasgow, UK

Bannau Brycheiniog National Park Authority, UK

UKCEH, UK

Climate Change Committee, UK

REECS, Philippines

Amazi, Burundi

Jacobs, UK  
Natural Resources Wales, UK  
University of Bath, UK  
University of Aberdeen, UK  
HES, UK  
Yorkshire Water, UK  
CEDEX, Spain  
University of Leeds, UK  
University of Granada, Spain  
Ausenco and UFMG, Brazil  
2u1k, Turkey  
Ethiopian Public Health Institute, Ethiopia  
Universitat Rovira i Virgili, Spain  
National Institute of Technology, Delhi, India  
Africa Center of Excellence for Water Management, Ethiopia  
Université Cadi Ayyad Marrakech, Morocco  
Climingo, UK  
Bangladesh Space Research Remote Sensing Organization, Bangladesh  
Dilla University, Ethiopia  
Haramaya University, Ethiopia  
Highland Hydrology, UK  
ISRIC - Soil World Information, Netherlands  
Queen Mary University of London, UK  
Giki, Pakistan  
Formerly at BfG (German Federal Institute of Hydrology), Germany  
NORZEC, Mozambique  
Durham University, UK

## Appendix 2. Presentation including questions

[Access the slide deck here](#)

# Partners

Funder:

- UKRI-NERC

Lead partner:

- UK Centre for Ecology & Hydrology

Other partners:

- British Geological Survey
- Imperial College
- University of Bristol

