

4. SmallData Analysis to Cube Production

SmallDataAna is a python modules to help with the analysis of the smallData. It is also integrated with a sister module (SmallDataAna_psana) that can be used to set up the "userData" in the smallData file.

Starting the analysis

If you are working in your experiments "smallData" directory (/reg/d/psdm/xxx/yyyy/results/smalldata_tools/ with xxx being the hutch and yyy the experiment name), you start the analysis session using `./examples/runSmallDataAna -r <#>`

This will start an ipython session where you can play with the data. An alternative approach is to use jupyter notebooks. Several examples can be found in the `small_data_tools/examples` directory.

If you are not on a LCLS machine and/or have not set up the analysis release, this code should also work with limited functionality (i.e. you CANNOT look at data that is not in your smallData file). This will need to be tested, please contact Silke if you'd like to use this and it did not just work.

Description of the SmallDataAna design

SmallDataAna is initialized with a datasource. Typically this is a "smallData" hdf5 file, but it can also be a REDIS database for feedback during an ongoing run. A dictionary is filled with keys corresponding to all available data fields. The data in the dictionary includes the shape of the data, where the data originates from and if this data is available onDisk or in memory.

xrData

1-d data in the smallData file is loaded into an xArray object upon initialization. Other fields can be added to the xArray object by requesting them, e.g. by calling `ana.getVar(varname)`. This should be invisible to the user, but if you look at the keys of the xrData object, you will see the in-memory list. This object allows to you use all the python tools that come with xArray/pandas.

user added data

The underlying xArray makes adding user data simple. You can either use `ana.addVar(name='newVar', data=[])` or you can use function `xArray assign` on `xrData`. Upon exiting the ipython session/the notebook, the new data is saved as a netcdf file which will be loaded into the xrData object when you start analyzing this dataset again.

Selection/ Event Filters

The SmallDataAna module contains a list of "Selections". Each selection has a name and a set of filter/cuts and create a sub-dataset by rejecting shots that fail one or more of the selection cuts. These selection can be applied to plots of variables, of the scan data or for the definition of a "cube". It is described in more detail in [SmallDataAna - Filters/Event Selection](#).

Binned data

To e.g. plot a scan, you need to bin the data in a given variable. This is what "cube" does. The Cube-object contains information about how the dataset should be binned, in particular which event selection should be used, which variable to bin in and what bins to use. It also contains a list of variables that should be binned. The standard deviation of the binned data is also saved. Using SmallDataAna_psana, you can also add data that is only in the xtc file: this is described in [4. Binned Data Production](#).

"Picked" data (e.g. for Ptychography)

`smalldata_tools/examples/PickedEventsCube.py` is code where for each point of the scan, a single image is picked using a user defined variable. This could be the strongest shot or any quality that is in the smallData file or can be calculated based on variables in the smallData files.

- [SmallDataAna - Getting Started](#)
- [SmallDataAna - Filters/Event Selection](#)
- [SmallDataAna - Binned Data / Cube Production](#)
- [SmallDataAna - Example Data & Scripts](#)
- [SmallDataAna_psana - makePedestal](#)