

---

# **SKA SDP Workflow pre-processing**

**SKA SDP DP HIPPO team**

**May 23, 2023**



## Contents

<b>1</b>	<b>Installation guide</b>	<b>1</b>
<b>2</b>	<b>Workflow steps</b>	<b>3</b>
<b>3</b>	<b>Indices and tables</b>	<b>5</b>



## Installation guide

External libraries required are:

```
DP3
Casacore
```

Please note that currently all the packages required for the DP3 (and the DP3 itself) should be built from source against the same version of casacore. More instructions can be found in: <https://confluence.skatelescope.org/display/SE/Cookbook+for+building+astronomical+packages+from+source>

Please also note that for the DP3 to be importable in python, the following variable should be set:

```
export PYTHONPATH="/home/your_home/where_you_installed_dp3_and_others/lib/
→python3.8/site-packages:$PYTHONPATH"
```

### 1.1 Uninstalling

It can be simply deleted.



## Workflow steps

Preprocessing contains three main stages: preflagger, flagger, and averaging. *Preflagger* is the stage where the RFI masks (e.g. channels known to be prone to RFI because of a known radio source) are applied. RFI *flagging* can be done by one of the flaggers existing in DP3. In the preprocessing pipeline, we have deployed AOFlagger as the more commonly used flagger. However, it can be replaced with the other flagger existing in the preprocessing pipeline (MAD) or possibly a customised flagger. *Averaging* is performed to reduce the size of data without losing the useful information and reducing the unhelpful fluctuations in the data. Averaging can be done in time and/or frequency directions.

More information about the steps of the workflow can be found in: <https://confluence.skatelescope.org/pages/viewpage.action?pageId=205798710>

### 2.1 Execution

To run the pipeline from the src directory, run the following command:

```
python3 pipeline.py --msloc your_MeasurementSet --maskloc your_RFI_mask_pickle_file.pickle --paramsloc your_parameters.json --outloc your_output.ms
```

where `--msloc` is the location of `your_MeasurementSet`, `--maskloc` is the location of `your_RFI_mask_pickle_file`, and `--paramsloc` is the location of `your_parameters` to identify which DP3 parameters you would like to use for each step. The parameters should be provided through a JSON file. Here is an example of the contents in such a file.

```
{
  "preflagger": [
    {
    }
  ],
  "aoflagger": [
    {
      "autocorr" : "True"
    }
  ],
  "averaging": [
    {
      "freqstep": 8,
      "timestep": 2
    }
  ]
}
```

(continues on next page)

(continued from previous page)

```
}  
]
```

```
}
```

The parameters' names are the same as in DP3. For example, `freqstep` is the number of channels to be averaged together. The `--outloc` is the location of the output measurementset, where obviously many of the columns have different dimensions from the original input measurementset because of averaging.



## Indices and tables

- [genindex](#)
- [search](#)