LOFAR International Single Station Metadata Definition

Griffin Foster

January 22, 2018

Abstract

This is a technical description of the metadata format defined for use with LOFAR international single station data products including: total correlation matrices (ACC), beamlet statistics (BST), subband statistics (SST), and single subband correlation matrices (XST). A python module (`issformat`) has been written which implements this definition. Metadata is stored as JSON files. HDF5 is used to provide a metadata wrapper for raw data files.

1 Introduction

The statistics files produced by a LOFAR station contain no observational metadata to describe the array configuration. Only time is encoded in the standard file name. In order to provide the essential metadata we have defined a set of keys which can be written to a JSON file as an auxillary file to the raw data. It is useful to keep a common naming convention for this file by replacing the .dat extension with a .json extension. In addition, an HDF5 format is defined to wrapper the raw data and provide the metadata as attributes. In this document the metadata is defined, along with the JSON format, HDF5 format, and a python module which can be used to interface with these file types.

2 Statistics File Types

Total Correlation (ACC): This is a correlation matrix of all antenna pairs in the configured array, for all subbands. For each integration a single subband is correlated, cycling through all 512 subbands. Each file contains a single integration per subband. The binary file is made up of 32-bit floats (real and imaginary, to make a 64-bit complex number). The data is of the form (512 subbands $\times$ nantpol $\times$ nantpol), where nantpol is 192 (96 antenna elements, 2 polarizations) for an international station. Note, KAIRA has only 48 antenna elements. The standard naming convention is `YYYYMMDD_HHMMSS_acc_512xNANTPOLxNANTPOL.dat`, e.g. `20120611_124534_acc_512x192x192.dat`. See `acc2npy()` for how to read in an ACC file to a numpy array.

Beamlet Statistics (BST): The detected and integrated power from each station beamlet. This is the most complex of the statistics files. Depended in the bitmode (4, 8, or 16) there are (976, 488, or 244) beamlets respectively. Each beamlets is made up of a set of RCUs beamformed in a specific direction for a giver coordinate system for a single subband. Each file contains the total number of beamlets for multiple integrations. The binary file is made up of 32-bit floats. The data is of the form (nintegrations × nbeamlets). The standard naming convention is `YYYYMMDD_HHMMSS_bst_00POL.dat`, e.g. `20170217_111340_bst_00X.dat`. See `bst2npy()` for how to read in an BST file to a numpy array.

Subband Statistics (SST): The subband statistics of a single RCU, i.e. the integrated power spectrum for an antenna element. Multiple integrations from the same RCU can be stored in a single file. The binary file is made up of 32-bit floats. The data is of the form (nintegrations × 512 subbands). The standard naming convention is `YYYYMMDD_HHMMSS_sst_rcuRCUID.dat`, e.g. `20140430_153356_sst_rcu024.dat`. See `sst2npy()` for how to read in an SST file to a numpy array.

Subband Correlation (XST): Similar to an ACC file in content, contains the correlation matrix for a single subband. Multiple integrations of the subband can be included in the file. The binary file is made up of 32-bit floats (real and imaginary, to make a 64-bit complex number). The data is of the form (nintegrations × 1 subband × nantpol × nantpol), where nantpol is 192 (96 antenna elements, 2 polarizations) for an international station. Note, KAIRA has only 48 antenna elements. The standard naming convention is `YYYYMMDD_HHMMSS_sbSBID_xst.dat`, e.g. `20170728_184348_sb180_xst.dat`. See `xst2npy()` for how to read in an XST file to a numpy array.

3 Metadata Key Definitions

A minimum set of keys have been defined that are necessary to describe most (if not all) possible observing modes using a LOFAR station. A common set of keys are used for all statistics data types (Section 3.1). There are also keys specific to file types BST (Section 3.3), SST (Section 3.4), and XST (Section 3.5).

3.1 Common Keys

- **Station:** `string`, station ID, 5 characters, e.g. SE607, UK608, IE613, ...
- **RCUmode:** `int` or list of `int`, mode of each RCU, valid values: 1-7. If only one entry is used then it is assumed all RCUs are the same mode. Otherwise, an entry for each RCU is required.
- **Timestamp:** `string`, date and time of the file.
- **HBAElements:** `string`, when using the HBA in a non-standard mode by disabling elements in the tile, e.g. HBA 'All-sky' mode, then this key is used to store the setup of each tile. A tile state is encoded in a 4-digit hexidecimal string. Each hexidecimal character represents a row of the tile. (optional)

- **Special:** `string`, Extra entry to include comments for the observation. (optional)
- **Rawfile:** `string`, Filename of the raw data file. (optional))
- **Integration:** `int`, integration length in seconds, default: 1.

3.2 Total Correlation (ACC) Keys

3.3 Beamlet Statistics (BST) Keys

- **Bitmode:** `int`, beamlet bit mode, 16, 8, or 4 bit resulting in 244, 488, 976 possible beamlets respectively.
- **Pol:** `string`, polarization of beamlet, X or Y.
- **beamlets:**
  - **ID:** `int`, beamlet ID number
  - **Pointing:** (`float`, `float`, `string`), pointing in given coordinate system (theta, phi, coord). Coordinates are in radians. Valid coordinate systems: J2000, HADEC, AZELGEO, ITRF, B1950, GALACTIC, ECLIPTIC, JUPITER, MARS, MERCURY, MOON, NEPTUNE, 'PLUTO, SATURN, SUN, URANUS, VENUS
  - **Subband:** `int`, subband ID.
  - **RCUs:** list of `int`, RCUs in the beamlet.

3.4 Subband Statistics (SST) Keys

- **RCU:** `int`, RCU ID.

3.5 Subband Correlation (XST) Keys

- **Subband:** `int`, subband ID.

4 JSON Definition

4.1 Examples

5 HDF5 Definition

5.1 Examples

6 `issformat` use

6.1 Install

6.2 Functions

6.3 Examples