

User guide for WEAVEify (v1)

WEAVE-QSO Collaboration



Contents

1	Overview of WEAVEIFY	1
1.1	Purpose of the code	1
1.2	Installation	1
1.3	Introduction	2
2	Quick Start	3
3	Modifying WEAVEIFY parameters	6
3.1	Continuum placement parameters	6
3.2	Instrumental broadening parameters	6
3.3	Wavelength sampling parameters	9
3.4	Telescope throughput parameters	9
3.5	Noise variation parameter	11
4	Automating WEAVEIFY for large number of spectra	12
4.1	Single core WEAVEIFY automation	14
4.2	Multi core WEAVEIFY automation using MPI	14
5	Reading WEAVEIFY output	15
5.1	Output file names	15
5.2	Data format	17
5.3	Header format	18
5.4	Script to read WEAVEIFY output	18

6	Detail description of free parameters in WEAVEIFY	20
6.1	QSO parameters	20
6.2	Wavelength sampling parameters	20
6.3	Continuum placement parameters	20
6.4	Noise parameters	21
6.5	Cosmological parameters	21
6.6	MPI job submission parameters	22
6.7	Input / Output file saving parameters	22
6.8	Miscellaneous parameters	22
7	Acknowledgement and Citations	24
8	Frequently Asked Questions (FAQs)	24

1 Overview of WEAVEIFY

1.1 Purpose of the code

WEAVEIFY (together with ASTRO-MOCKS) is developed to forward-model the QSO absorption spectra in simulations that look similar (in a statistical way) to those that will be observed by the WEAVE instrument. WEAVEIFY can efficiently generate large number of ($\sim 10^4 - 10^6$) mock QSO absorption spectra that accounts for realistic noise, continuum placement and instrumental resolution effects of WEAVE instrument. Even though WEAVEIFY is written to model spectra for WEAVE survey, the code is modular and flexible enough to mock the spectra for variety of instruments by changing parameters. We refer the reader to the paper describing WEAVEIFY for details. The code is publicly available at <https://ingbitbucket.ing.iac.es/projects/WVQSO/repos/weaveify>

1.2 Installation

- User can download the code using following command. The size of the zip file is ~ 50 Mb.

```
git clone ssh://git@ingbitbucket.ing.iac.es:7999/wvqso/weaveify.git
```

- Extract the zip file at your location of choice. The typical file structure would look like shown in Fig. 1

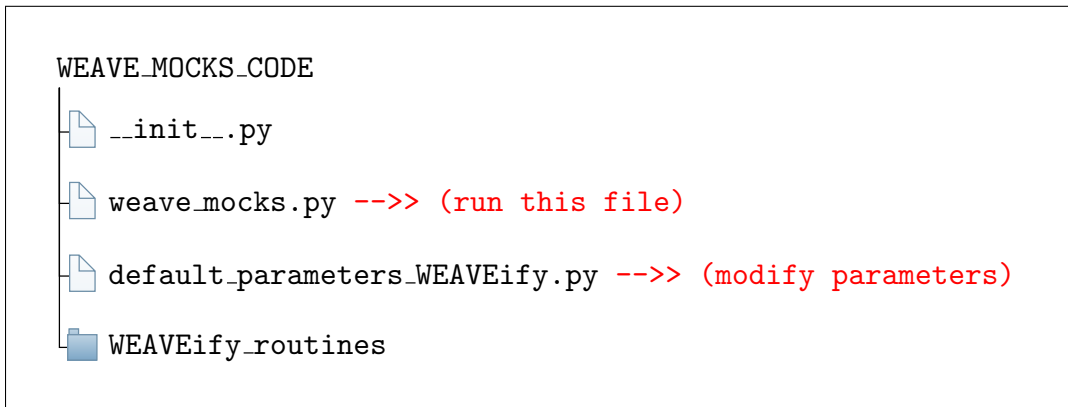


Figure 1: Basic structure of the code directory. Users need to run the file `weave_mocks.py`. All the parameters of the code are specified in the file `default_parameters.WEAVEify.py`

- WEAVEIFY is written in python and is compatible with both *python2* and *python3*
- Following packages are required
 - *numpy*
 - *scipy*
 - *matplotlib*
 - *astropy* (Optional) package is needed if the desired output format is **fits**.
 - *h5py* (Optional) package is required if the desired output format is **hdf5**.
 - Either *astropy* or *h5py* should be installed.
 - *mpi4py* package is required if spectra needs to be generated on multi-core machines in parallel. *mpi4py* is not needed if users want to run the code for few spectra on single core.

- Users can refer to §8 if they have any difficulty in installing these python packages.
- All the parameters needed to mock the spectra are given in the file `default_parameters_WEAVEify.py`. User may modify this parameter file later for their custom needs (see §4 for details). We recommend users to make a copy of this file before modifying it.

1.3 Introduction

- To execute the code, go to directory `WEAVE MOCKS CODE` and run the python script `weave_mocks.py` (see Fig. 1)
- WEAVEIFY can be run either with command like interface or by modifying the parameter file `weave_mocks.py`.

```
python weave_mocks.py <argument-list> # Command like Interface
```

- We recommend to use the command like interface to test the code behaviour for small number of mocks (~ 100 spectra).
- While for generating large mock spectra, we recommend to modify `weave_mocks.py` file.

```
python weave_mocks.py # Script like Interface for Automation
```

- We provide few examples of `weave_mocks.py` file to illustrate how automation can be achieved with script approach.
- We refer users to §4 on how to use Message Passing Interface (MPI) with script.

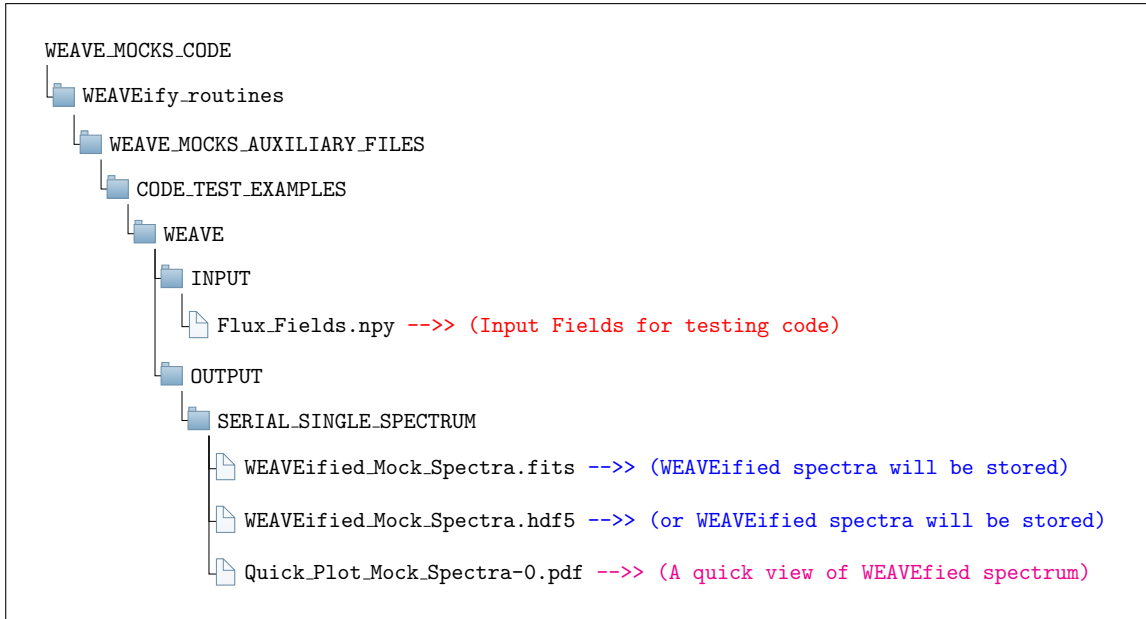


Figure 2: Location of input and output directory for test examples of the code.

2 Quick Start

- To begin, we provide few test cosmological mock flux fields with the code. The tests examples are kept in directory `WEAVE` directory (see Fig. 2)
- In this directory there are 2 subdirectories `INPUT` and `OUTPUT`. `INPUT` directory contains spectra that are generated from cosmological Sherwood simulation suite.
- For the following tutorial, the code output, e.g., mock spectrum and plots, would be stored in `OUTPUT/SERIAL_SINGLE_SPECTRUM/` directory.
- After each successful execution of code, user can view mock spectrum from `Quick_Plot_Mock_Spectra-0.pdf`
- For quick start, we can run the code using following command,

```
python weave_mocks.py flag_make_quick_plot=y
```

- Once the code is executed successfully, we can go to directory `OUTPUT/SERIAL_SINGLE_SPECTRUM`

The code has generated several files with fits extension and a pdf file showing the mock spectrum. The main output mock spectrum is in file `WEAVEfied_Mock_Spectra.fits`. We can view part of the mock spectrum by opening file `Quick_Plot_Mock_Spectra-0.pdf`

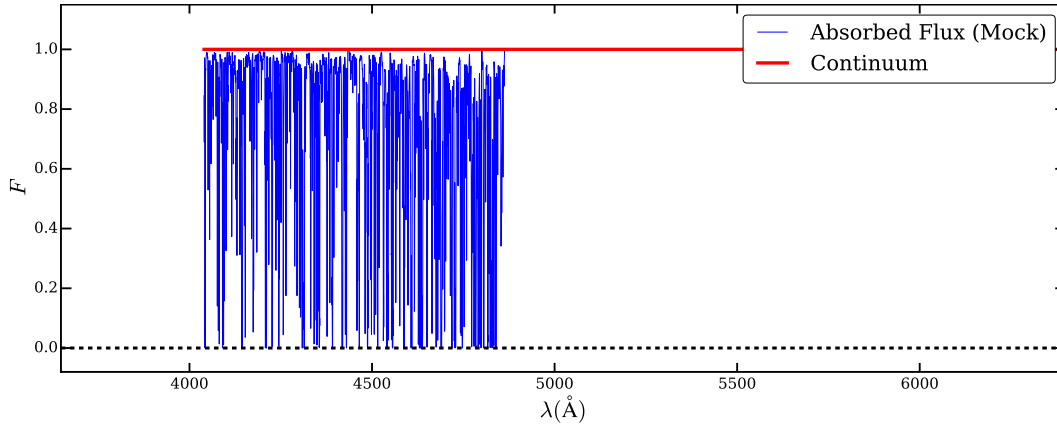


Figure 3: Example of mock spectrum without any WEAVEification

- We can change the resolution of the spectrum to LR mode of WEAVE as (default is HR),

```
python weave_mocks.py flag_make_quick_plot=y resolution_mode=LR
```

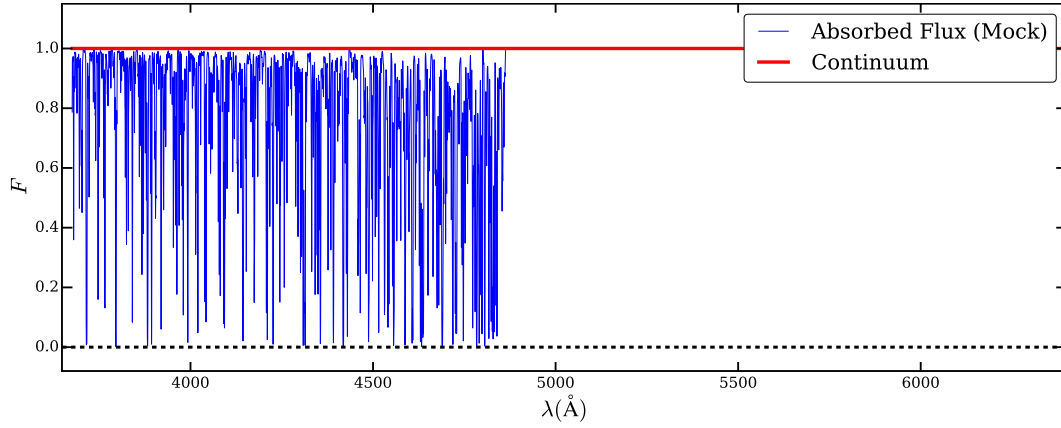


Figure 4: Example of mock spectrum for WEAVE LR mode

- We can place continuum using `flag_cont_method` variable (default is PCA),

```
python weave_mocks.py flag_make_quick_plot=y resolution_mode=LR
flag_cont_method=PCA
```

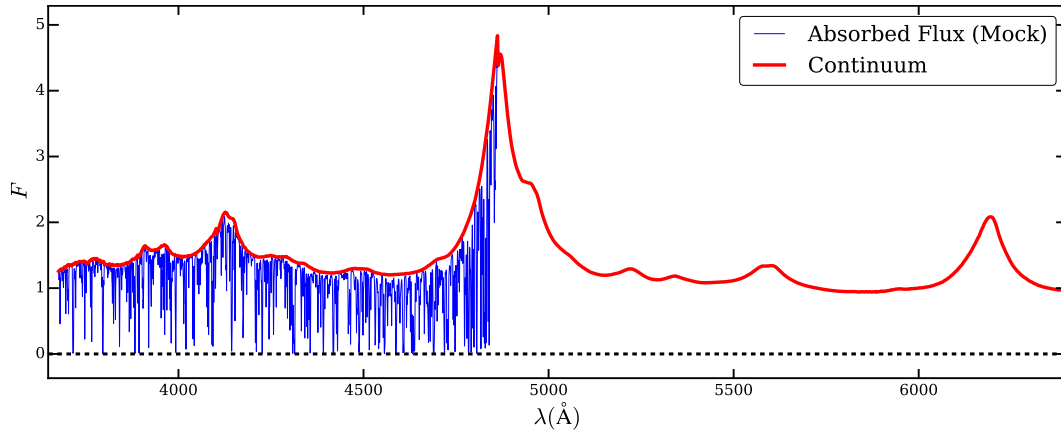


Figure 5: Example of mock spectrum with continuum placement

- To add realistic noise to the spectra, we can use `flag_noise` variable to OpR3 as

```
python weave_mock.py flag_make_quick_plot=y resolution_mode=LR
flag_cont_method=PCA flag_noise=OpR3
```

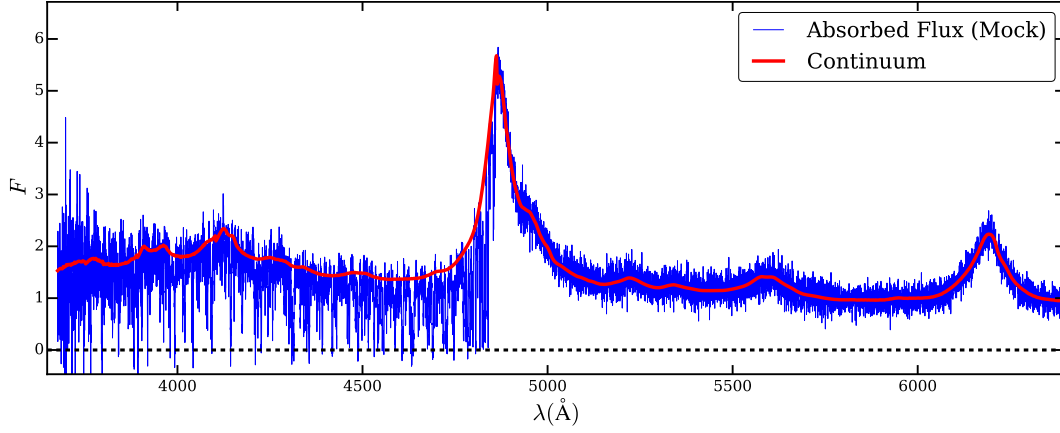


Figure 6: Example of mock spectrum with realistic noise

- To change magnitude of the source, we can execute following command

```
python weave_mock.py flag_make_quick_plot=y resolution_mode=LR
flag_cont_method=PCA flag_noise=OpR3 m_source=18
```

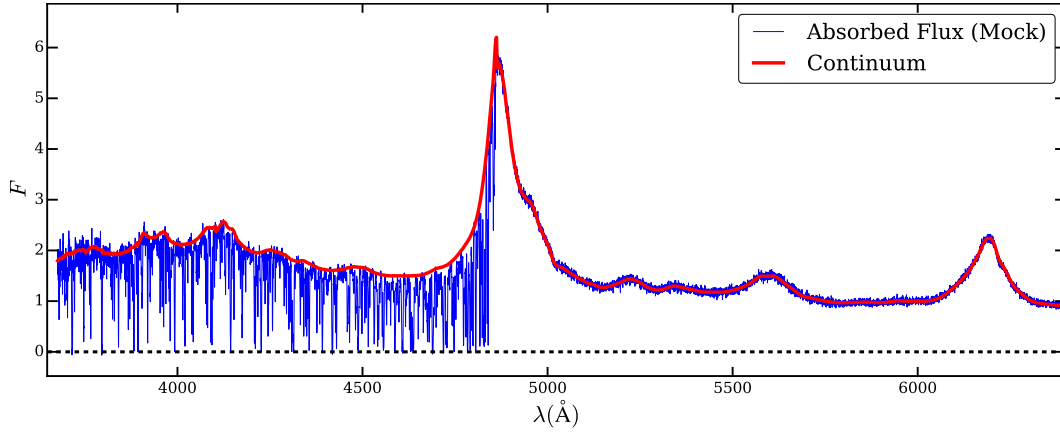


Figure 7: Example of mock spectrum with background QSO of 18 magnitude

- To change sky magnitude, we can execute following command

```
python weave_mocks.py flag_make_quick_plot=y resolution_mode=LR
flag_cont_method=PCA flag_noise=0pR3 m_source=20 m_sky=18
```

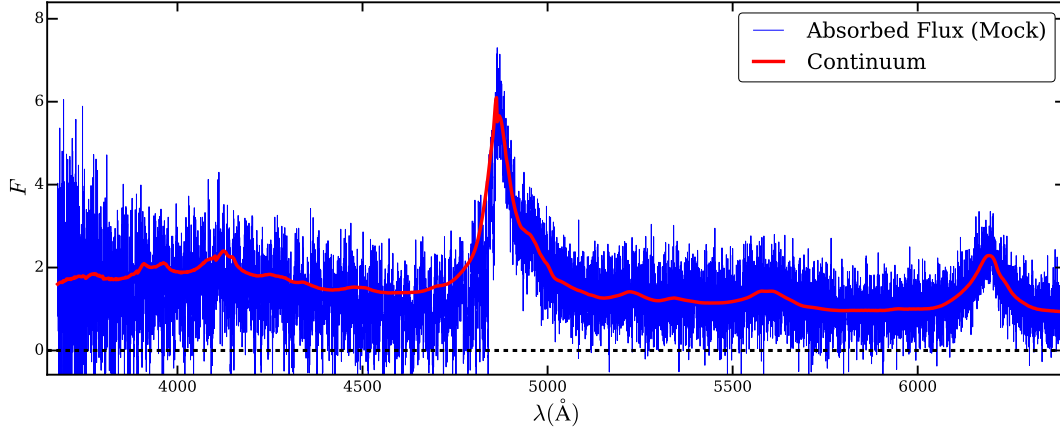


Figure 8: Example of mock spectrum with sky brightness of 18 magnitude and source magnitude of 20

3 Modifying WEAVEIFY parameters

3.1 Continuum placement parameters

In WEAVEIFY, the continua are generated randomly using Principle Component Analysis (PCA) method. In this method, a random realization of continuum is generated from a composite continuum in red side (with respect to Ly α emission) of the QSO spectrum. Using the random continuum realization, composite spectrum and PCA components, the continuum is generated at all other wavelengths. There are 3 main variables in WEAVEIFY that controls the continuum placement which are described below.

- **flag_cont_method** : This variable tells WEAVEIFY a method to generate continuum. Currently only ‘PCA’ method of continuum generation is implemented in WEAVEIFY.
- **random_seed_cont** : As mentioned earlier, a realization of continuum is generated from composite spectrum. However, sometimes users may want to test the code using same realization of continuum. If **random_seed_cont** is set to some integer value then it would generate a pseudo-random numbers such that if same value of **random_seed_cont** is used again the continuum will be same. By default **random_seed_cont=None** indicating that the continua are generated using true random numbers.
- **alpha_cont** : It is possible that users would like to test the extreme cases of variation in continua in Ly α forest. The power-law index (α) of QSO spectra can be used to generate a large variation continuum $F_{\text{cont}} \propto \lambda^\alpha$. Setting the value of variable **alpha_cont** provides a convenient way to generate large variation in continua. The typical values of α varies from -0.5 to -3.0 but users can chose more extreme range like 0.0 to -5.0

3.2 Instrumental broadening parameters

- Many observations of quasar absorption spectra are taken from different instruments and/or telescopes. Often the instrumental broadening (also known as line spread function) of the

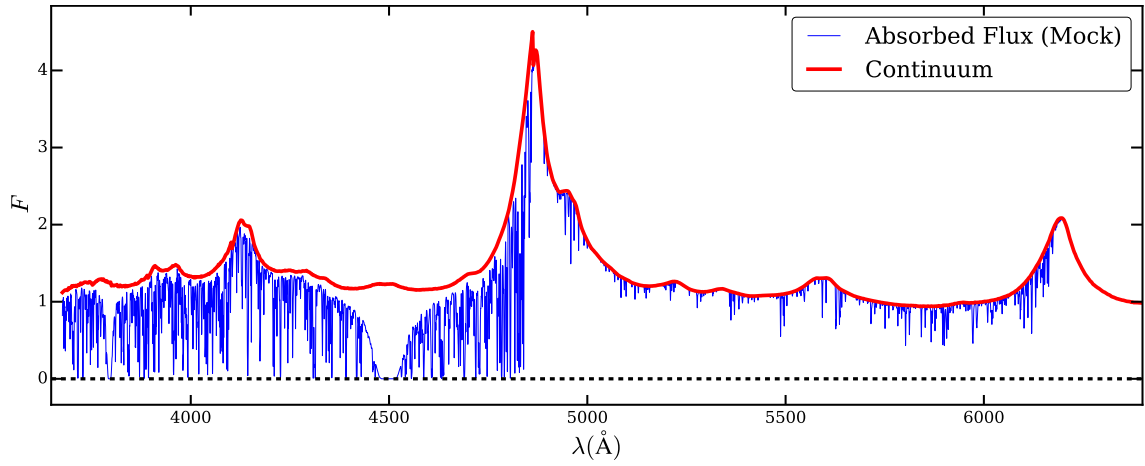


Figure 9: Mock spectrum with another continuum (Continuum templates are randomly generated from PCA). QSO continuum power-law index used in above example is $\alpha = -0.5$

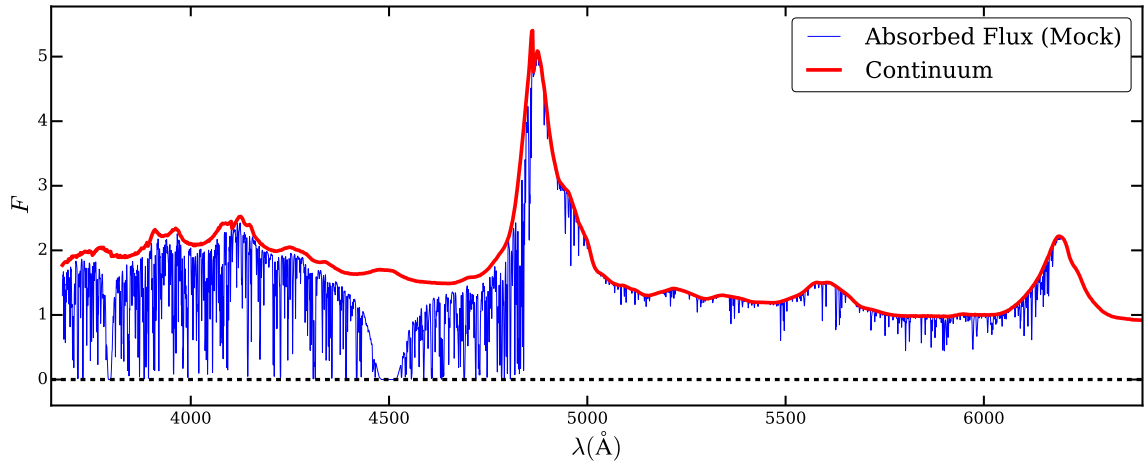


Figure 10: Mock spectrum with another continuum (Continuum templates are randomly generated from PCA). QSO continuum power-law index used in above example is $\alpha = -3.0$

absorption line is different for each instrument.

- For example, in WEAVE-LR sample the quasar absorption spectra would be obtained with $R = \lambda/\Delta\lambda \sim 5000$ while WEAVE-HR would have $R \sim 20000$.
- For standard WEAVE-LR and WEAVE-HR setting, we provide line spread function in files `instrumental_broadening_parameters_WEAVE_LR.txt` and `instrumental_broadening_parameters_WEAVE_HR.txt`
- The location of these files is shown in Fig. 11. Users can modify the files as per their needs.
- First two columns in these files are (lower and upper) wavelength ranges. The third column is FWHM of Gaussian LSF in velocity units i.e., km s^{-1} . We assume WEAVE LSF is Gaussian in shape (see Listing 1) for details.

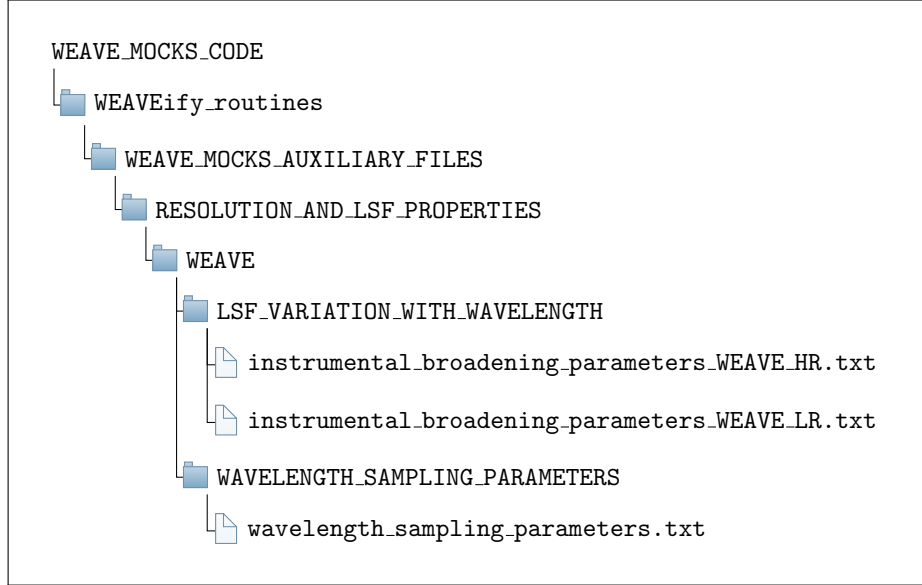


Figure 11: Location of instrumental broadening and wavelength sampling parameter files

Listing 1: Snippet of instrumental broadening parameter LR file

```

# Column-00 -->> Lower Wavelength (Angstroms) of wavelength range
# Column-01 -->> Upper Wavelength (Angstroms) of wavelength range
# Column-03 -->> FWHM of Gaussian

3000.00      3500.00      60.0
3500.00      4000.00      60.0
4000.00      4500.00      60.0
:            :            :
:            :            :
8500.00      9000.00      60.0
9000.00      9500.00      60.0
9500.00      10000.00     60.0

```

3.3 Wavelength sampling parameters

- Similar to instrumental broadening parameters, we provide wavelength sampling parameters in file `wavelength_sampling_parameters.txt` (see Fig. 11). The snippet of this file is shown in Listing 2
- Since wavelength coverage of WEAVE instrument is different in two resolution modes, we have tabulated the properties of each resolution mode and wavelength coverage separately in each row.
- In particular, users need to specify starting wavelength (CRVAL1), wavelength separation (CDEL1), Number of pixels, whether the wavelength are equally spaced in logarithmic or linear scale and central wavelength of the wavelength range.
- The default numbers specified in wavelength sampling parameters file are for WEAVE low resolution (LR) and high resolution (HR) mode. Users can modify these numbers as per their preferences.

Listing 2: Snippet of `wavelength_sampling_parameters.txt` file

```
# This file contains wavelength sampling parameters for WEAVE
# These parameters are taken from OpR3 Mocks
#
# Column-00 --> Resolution Mode (HR or LR)
# Column-01 --> Arm of the spectrograph Mode
# Column-02 --> CRVAL1 (Starting wavelength in angstroms)
# Column-03 --> CDEL1 (Wavelength separation in angstroms)
# Column-04 --> N_pixel_obs (Number of Pixel in the spectrum)
# Column-05 --> lambda_scale (LIN or LOG)
# Column-06 --> Central Wavelength of arm (Used to calculate
#               grating dispersion in noise addition)
#
# The wavelength array is obtained as
#
# if lambda_scale is LIN
# lambda_arr_obs = CRVAL1 + CDEL1 * np.arange(N_pixel_obs)
#
# if lambda_scale is LOG
# lambda_arr_obs = 10** (CRVAL1 + CDEL1 * np.arange(N_pixel_obs))
#
# -----
LR      Blue      3676.0      0.25      9649      LIN      4900.0
LR      Red       5772.0      0.25     15289      LIN      7950.0
# -----
HR      Blue      4040.0      0.05     12201      LIN      4250.0
HR      Green     4730.0      0.05     14401      LIN      5130.0
HR      Red       5948.0      0.05     18041      LIN      6450.0
# -----
```

3.4 Telescope throughput parameters

- Telescope throughputs specifies the efficiency of the telescope to capture the number of photons from source. Telescope throughput parameters are essential for realistic addition of the noise to the spectra.

- Since WEAVE instrument is still in installation phase, we have estimates of telescope throughputs. Thus telescope throughputs are subjected to change in future.
- The telescope throughputs are specified in file `telescope_throughput_V01.txt`. The telescope throughputs are taken from official WEAVE WEAVE-SYS-007 document. The location and sample content of this file is shown in Fig. 12 and Listing 3 respectively.

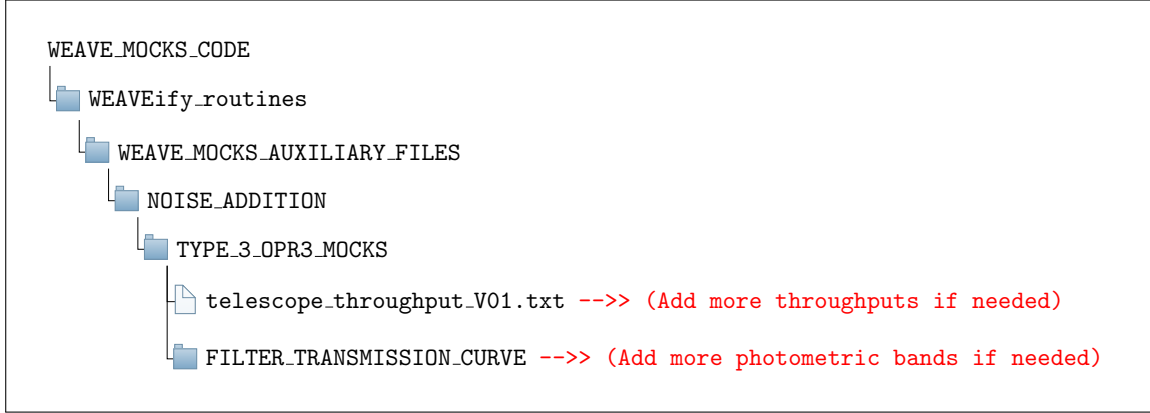


Figure 12: Location of telescope throughput file.

- There are 6 types of throughput to be specified in `telescope_throughput_V01.txt`, (i) Mirror reflectivity throughput, (ii) Central obscuration throughput, (iii) Prime focus corrector throughput, (iv) Fibre throughput, (v) Spectrograph throughput and (vi) CCD detector throughput
- It is important to note that wavelengths are specified in nano-meters while throughputs are always specified in percentages in this file. WEAVEIFY internally converts these quantities in to their standard units. The throughputs of various sources are multiplied to get effective throughput of the telescope.
- The ‘Essential’ type throughput refers to the minimum throughput that each system is required to meet. The ‘Optimal’ type throughput refers to the goal for each system to achieve this throughput. The noise routines of WEAVEIFY code currently use ‘Optimal’ throughput values.
- If users need to add more throughput they can append new rows to the table. In such case, one need to give an unique name to throughput type (Column-02). Once the new values are added in this file, users need to modify `throughput_type` variable to new unique name in `default_parameters_WEAVEify.py` file (see Fig. 1). This variable is currently set to ‘Optimal’.

Listing 3: Snippet of `telescope_throughput_V01.txt` file

```

# Column-00 -->> Resolution Mode (HR or LR)
# Column-01 -->> Arm of the spectrograph Mode
# Column-02 -->> Throughput type (Essential or Optimal or Observed)
# Column-03 -->> Wavelength in nano-meters
# Column-04 -->> M1 Reflectivity throughput (percent)
# Column-05 -->> Obscuration throughput (percent)
# Column-06 -->> Prime Focus Corrector throughput (percent)

```

```

# Column-07 -->> Fibres throughput (percent)
# Column-08 -->> Spectrograph throughput (percent)
# Column-09 -->> Detector throughput (percent)
#
# -----
LR  Blue  Essential  370.0  85.5  78.7  71.3  21.7  21.0  66.3
LR  Blue  Essential  400.0  85.8  78.7  79.1  53.7  46.3  92.1
:   :      :      :      :      :      :      :      :
:   :      :      :      :      :      :      :      :
HR  Red   Optimal   680.0  89.7  78.7  80.8  83.3  26.1  92.1
HR  Red   Optimal   685.0  89.6  78.7  80.8  83.4  21.0  91.4
# -----

```

3.5 Noise variation parameter

We estimate the noise for a given observation by first counting the total number of source photons received by the telescope as given below,

$$N_{\text{source}} = 10^{-m/2.5} \times N_{\text{photons, } m=0} \times f_{\text{atm}}(\text{airmass}) \times t_{\text{exp}} \times \eta_{\text{telescope}} \times A_{\text{unobst}} \quad (1)$$

where m is magnitude of source in given wave band, $N_{\text{photons, } m=0}$ is number of photons/sec/Å/m² received from source with $m = 0$ at the top of atmosphere, f_{atm} is transmission of atmosphere at a given airmass, t_{exp} is exposure time in seconds, $\eta_{\text{throughput}}$ is total throughput of telescope and A_{unobst} is unobstructed area of main mirror in m².

We count the total number of sky photons (N_{sky}) similar to N_{source} by replacing the source brightness magnitude (m_{source}) by sky brightness magnitude m_{sky} . Since the phenomenon of counting the photons on CCD is governed Poisson statistics, we calculate the noise contribution of source (or sky) as $\sigma = \sqrt{N}$. Assuming the noises are uncorrelated, the total noise contribution and signal-to-noise ratio (SNR) is given by,

$$\begin{aligned} \sigma_{\text{tot}} &= \sqrt{\sigma_{\text{source}}^2 + \sigma_{\text{sky}}^2 + \sigma_{\text{dark}}^2 + \sigma_{\text{read-out}}^2} \\ \text{SNR} &= N_{\text{source}} / \sigma_{\text{tot}} \end{aligned} \quad (2)$$

where σ_{dark} , $\sigma_{\text{read-out}}$ are noise due to dark current and CCD read-out respectively.

Some of the noise (miscounting of photons) can also come from thermal radiation of the detector. If $n_{\text{e,dark}}$ is number of ‘dark’ electrons detected per unit time and N_{pix} be the number of pixels used to observe an object then the Dark shot noise (σ_{dark}) is given by,

$$\sigma_{\text{dark}} = \sqrt{N_{\text{pix}} n_{\text{e,dark}} t_{\text{exp}}} \quad (3)$$

Read out noise is slightly different than the Shot noise. While reading the dectector, the amplifiers contribute some additional noise to the signal. This is a characteristic of the chip and of the read out mode used. Let $n_{\text{read-out}}$ be read-out-noise per read-out and N_{pix} be the number of pixels read in single exposure time. If there are N_{epoch} exposures then the total read-out noise ($\sigma_{\text{read-out}}$) is given by,

$$\sigma_{\text{read-out}} = \sqrt{N_{\text{pix}} N_{\text{epoch}}} \times n_{\text{read-out}} \quad (4)$$

Note that $n_{\text{read-out}}$ is not in the square root.

Following are the main variables in WEAVEIFY that can be changed to vary SNR of the mock spectra.

- `flag_noise` : This variable defines the method of adding noise to the spectra. The default value of `flag_noise= 'OpR3'` . If `flag_noise` is set to None, then noise will not be added to the spectra.
- `m_source` and `wave_band_source` : The source magnitude and the wavelength filter is defined by the variables `m_source` and `wave_band_source` respectively. WEAVEIFY assumes that the source magnitudes are specified in UBVRI photometric systems.
- `m_sky` and `wave_band_sky` : Similar to source, the sky magnitude and the wavelength filter is defined by the variables `m_sky` and `wave_band_sky` respectively. WEAVEIFY assumes that the sky magnitudes are specified in UBVRI photometric systems.
- `t_exp` : Total exposure time in seconds
- `airmass` : Airmass of the atmosphere. The default value is 1.107.
- `seeing_profile` : Atmospheric seeing effect is controlled by three parameters. `seeing_profile` defines the functional form of seeing. The options for `seeing_profile` is either 'Gaussian' or 'Moffat'. The default seeing profile is Moffat.
- `beta_moffat` : The shape of the Moffat profile depends on the β parameter. The default value is 2.5. This variable is only used if seeing profile is set to Moffat.
- `fwhm_seeing` : The variable `fwhm_seeing` defines the full width at half maximum of seeing profile expressed in arcsec. The default value is 0.75 arcsec.
- `read_noise` : Read-out noise per electron. The default value is 3.3.
- `dark_current` : The noise due to the dark current. The default values is 0.0.
- `throughput_type` : WEAVE scientific instrument team has provided the variation of telescope throughput with wavelength for LR and HR resolution mode. Currently, there are two options to chose for `throughput_type` . The 'Essential' type throughput refers to the minimum throughput that each system is required to meet. The 'Optimal' type throughput refers to the goal for each system to achieve this throughput. The noise routines currently use 'Optimal' as default throughput values (see §3.4 for details).
- `dia_telescope` : Diameter of the telescope in meters. The default value is 4.2m for WHT.
- We show an example of changing source magnitude on noise properties of the WEAVEIFY mock spectrum. `Output:Quick_Plot_Mock_Spectra-0.pdf`.

4 Automating WEAVEIFY for large number of spectra

The examples provided in §2 are suitable to produce few 10's of mock spectra. Often users would like to use WEAVEIFY to generate large number of ($\sim 10^4 - 10^6$) mock spectra. In this section we describe how WEAVEIFY can be used in script mode to generate spectra for WEAVE like survey. We provide two sample scripts to demonstrates how WEAVEIFY can easily be used for generating large number of mock spectra. The file `weave_mocks_serial_multiple_spectra.py` shows an example of automating the WEAVEIFY on single core while file `weave_mocks_parallel_mpi_multiple_spectra.py` shows an example of using WEAVEIFY on multi-core machines.

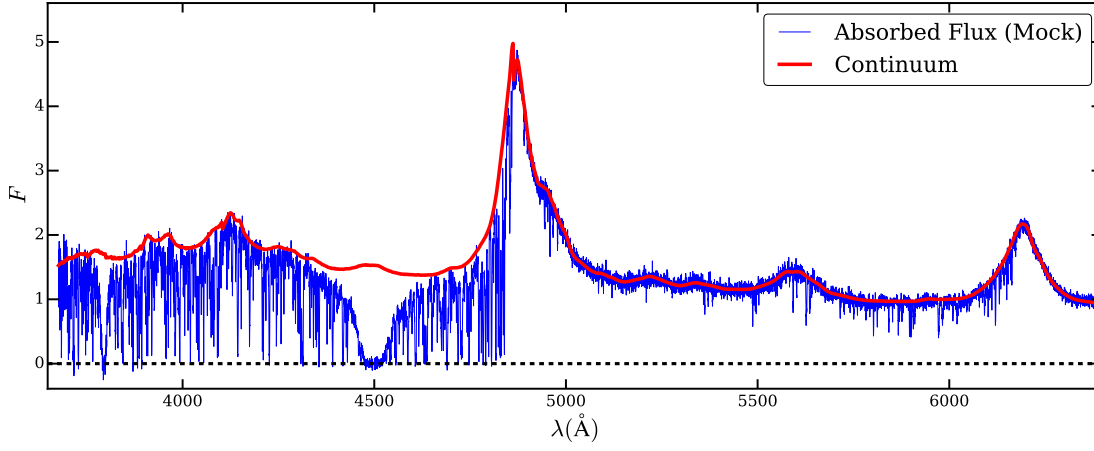


Figure 13: Example of mock spectrum with H I forest + DLA for a source magnitude of 18.5.

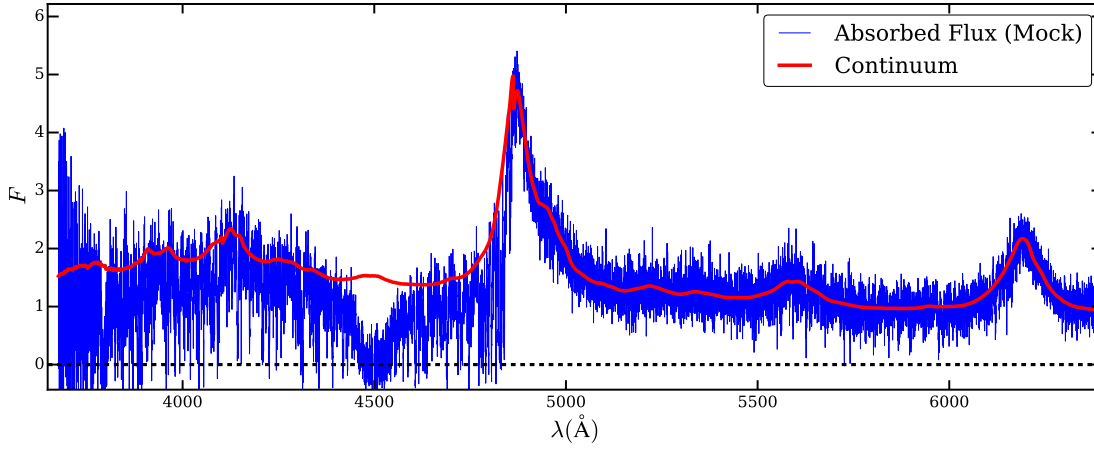


Figure 14: Example of mock spectrum with H I forest + DLA for a source magnitude of 20.5.

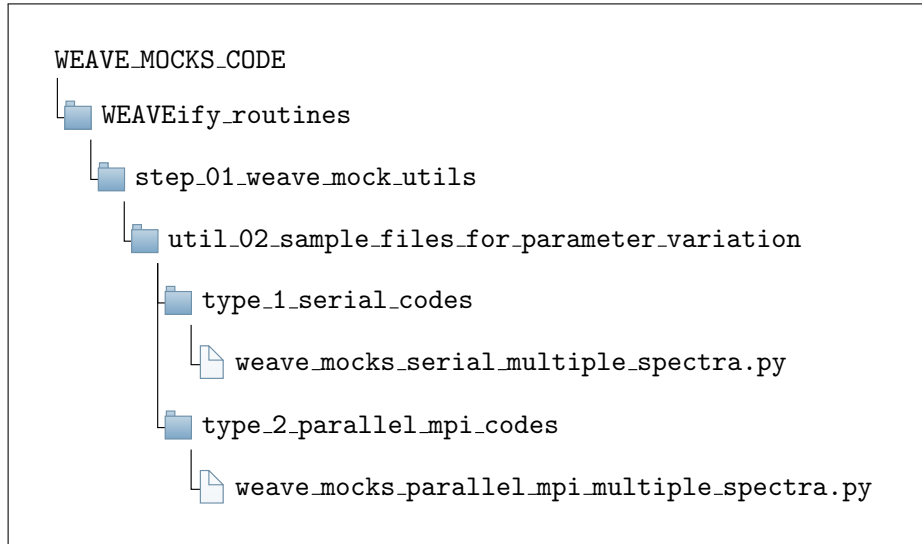


Figure 15: Location of sample codes to automate WEAVEIFY for large number of spectra. The file `weave_mocks_serial_multiple_spectra.py` shows an example of using single core machine while file `weave_mocks_parallel_mpi_multiple_spectra.py` illustrates use of WEAVEIFY on multi-core machine.

4.1 Single core WEAVEIFY automation

In this section, we describe the steps to automate WEAVEIFY for multiple spectra.

1. First we identify the parameter that needs to be varied. Identify the variable name corresponding to this parameter. Refer §6 or file `default_parameters_WEAVEify.py` to see list of all variable names. For example, if we want to vary magnitude of sources, then the corresponding variable name is `m_source`.
2. WEAVEIFY code uses one single main function `WEAVEify_function_calls_main`. This function has been imported in all `weave_mock.py` files. We need to supply the variable name users want to change to the function `WEAVEify_function_calls_main` as a keyword argument (`**kwargs`). *In this step we need to set the variable `flag_mpi='n'` to tell the code that it is a single core script.*
3. All other parameters would take default value as set in file `default_parameters_WEAVEify.py`.
4. In the final step, we need to call this function multiple times in `for` loop in which the parameter is changed. In this way WEAVEIFY can generate multiple number of spectra.
5. We provide a more detailed example of automating WEAVEIFY on single core machine in file `weave_mock_serial_multiple_spectra.py` (see Fig. 15).
6. Users can copy the file `weave_mock_serial_multiple_spectra.py` to `WEAVE MOCKS_CODE` directory and run following command to test how WEAVEIFY works

```
python weave_mock_serial_multiple_spectra.py
```

Note that the `spec_id` is one of the most important variable that should be varied within for loop. This variable is used in WEAVEIFY to create unique hdf5 group or fits extension name in the final output files.

4.2 Multi core WEAVEIFY automation using MPI

Now we use the same example given above, and show how to automate WEAVEIFY on multi-core machines using Message Passing Interface (MPI). The basic idea behind automation is same as that of single-core machine except that the `for` loop is executed simultaneously on different cores. This essentially reduces the number of iterations that the `for` loop performs. A prerequisite to use multi-core functionality is that the `mpi4py` is installed. The steps to automate WEAVEIFY using parallel MPI libraries is as follows

1. *First we modify the `weave_mock.py` file for single core machine as explained in previous section. We recommend to test if the code is working on single core. The code that runs on single-core machine can simply be run on multi-core machine. For this we just need to change variable `flag_mpi='y'`*
2. We provide a more detailed example of automating WEAVEIFY on multi core machine in file `weave_mock_parallel_mpi_multiple_spectra.py` (see Fig. 15).
3. We recommend reader to compare the content of file `weave_mock_parallel_mpi_multiple_spectra.py` with `weave_mock_serial_multiple_spectra.py`. All the variables are identical except `flag_mpi` and `out_path`. We changed `out_path` to save the output at different location. However, users do not need to change this.

4. Users can copy the file `weave_mock_parallel_mpi_multiple_spectra.py` to `WEAVE MOCKS_CODE` directory and run the program. The syntax for running MPI codes is as follows.

```
mpirun -np <number-of-processors> python <file-name.py>
```

Following command would work with the examples provided.

```
mpirun -np 5 python weave_mock_parallel_mpi_multiple_spectra.py
```

Note that syntax to submit the MPI job may differ on different machines especially if the HPC uses Slurm resource management systems. HPC system admin team usually provides instruction on how to submit the slurm scripts. Users do not need to worry about distributing the equal number of jobs on different processors. Even if the number of spectra to mock are not divisible by number of processors, WEAVEIFY code would try to distribute them as evenly as possible.

5 Reading WEAVEIFY output

In this section we describe the basic format of WEAVEIFY output files. For fits format, the data is arranged in extension name and BinaryHDU tables. The data of each mock spectrum is stored under different extension name. All data arrays are stored with single precision accuracy to optimize the file sizes. A header is attached to each extension separately. For hdf5 format, the data is arranged in groups, dataset and Header blocks. Similar to fits format, the data of each mock spectrum is stored under different hdf5 group. A header block is explicitly added to each group. *The output data structure and naming convention is same for fits and hdf5 file format.*

5.1 Output file names

By default, the WEAVEIFY outputs (i.e., optical depths and fluxes) are saved in file strating with prefix ‘WEAVEified_Mock_Spectra’. The variables `out_file_flux_base` and `out_file_tau_base` controls the prefix of the output files. Both variables currently set to `WEAVEified_Mock_Spectra`. As a result optical depths and fluxes are saved in the same file. If users want to change the prefix of the output files, they need to change these two variables. It is also possible to store the optical depths and fluxes in two different files. In such case, users need to chose different values of the two variables. When WEAVEIFY is run in parallel, the outputs from different processors (ranks) are stored in different files and a suffix ‘Rank_<processor-number>’ is attached to each file. This way of storing the file optimizes the performance of resources while maintaining the output file size. The extension of output files is controlled by variable `flag_output_format`. WEAVEIFY supports two file format fits or hdf5 which requires different python modules to be installed (see 1.2 for details).

```
For flag_mpi = 'n', flag_output_format = 'fits',  
Output File Name: WEAVEified_Mock_Spectra.fits
```

```
For flag_mpi = 'y', rank = 5, flag_output_format = 'hdf5',  
Output File Name: WEAVEified_Mock_Spectra_Rank_5.hdf5
```

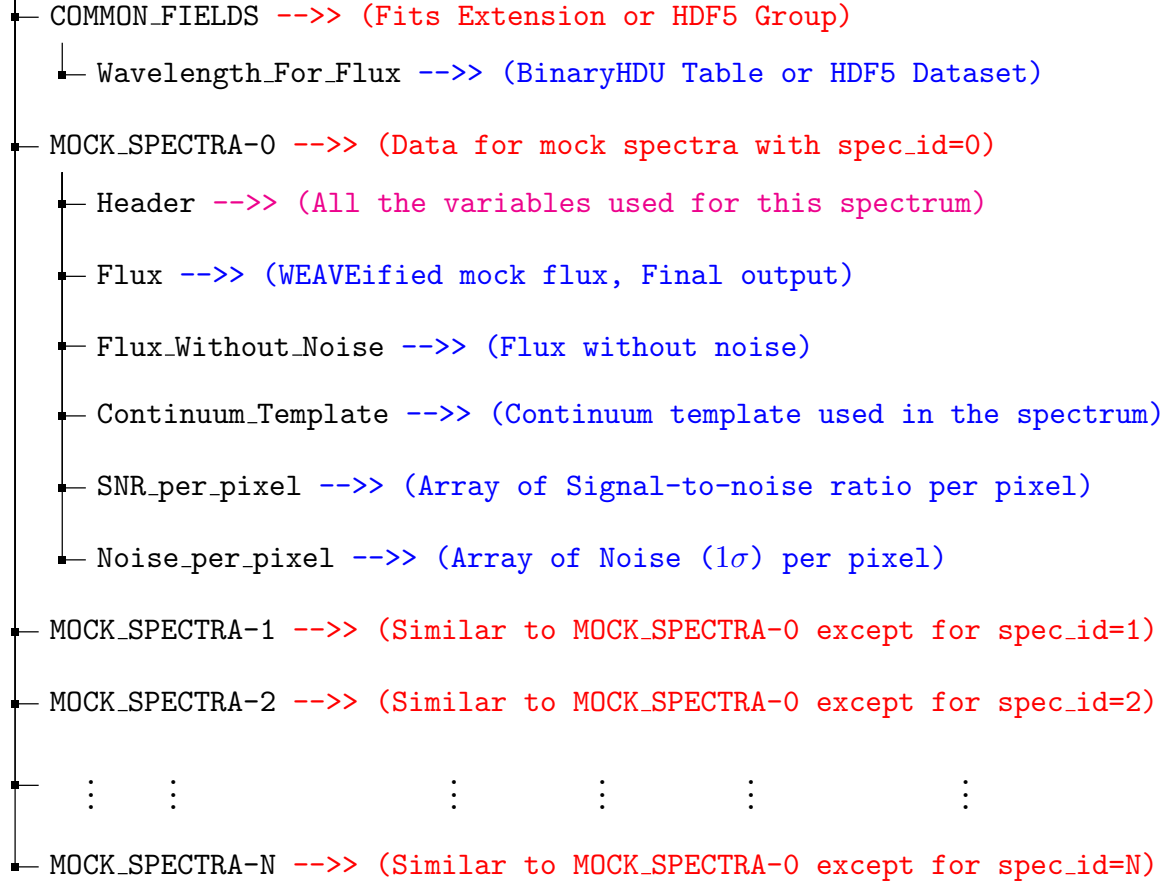


Figure 16: **The figure shows basic structure of WEAVEIFY output files.** Data is arranged in extension name (for fits format) and group name (hdf5 format). Each extension / group name contains its own header. The header stores all the input parameters supplied to WEAVEIFY. Note that header variables would be written for each mock spectrum separately. Data of optical depth for DLA and metals may not be stored if sightline does not contain DLA or metals as specified by users. Note that dimension of optical depth array and flux array need not be same. Hence separate wavelength array is provided for optical depth and flux. The output data structure is same for fits and hdf5 format. **Variable spec_id is used to create an unique extension / group name that distinguishes data of one mock spectrum from another. We provide a script read_weaveify_output_all_format.py to read the WEAVEIFY output files.**

5.2 Data format

Fig. 16 shows the basic data structure of typical WEAVEIFY output file (fits or hdf5). The WEAVEIFY output file contains following fits extension (hdf5 group) names, BinaryHDU tables (hdf5 dataset) and headers.

- **COMMON_FIELDS:** (Fits Extension or HDF5 Group, Block, String)
This extension / group contains the arrays common to all the spectra. For example, for a given resolution mode the wavelength array of a flux would remain same for all the spectra. Hence to optimize the output file size it is important to save the array only once in file.
- **Wavelength_For_Flux:** (BinaryHDU Table or HDF5 Dataset, Array, Float32)
This table/dataset contains wavelength array for the fluxes. The shape of this array could be equal to or smaller than the **Wavelength_For_Optical_Depth** table/dataset.
- **MOCK_SPECTRA-*(spec_id)*:** (Fits Extension or HDF5 Group, Block, String)
This extension / group contains the output flux and optical depth arrays for spectrum with unique identification variable *spec_id*. The mock spectra are distinguished from each other by *spec_id*. Hence it is important to vary *spec_id* for every spectrum.
- **Header:** (FITS or HDF5 Header, Dictionary, Mixed data type)
The header contains all the variables supplied to WEAVEIFY for generating the spectrum. Since the parameters varies from one spectrum to another, Header is specified for every mock spectrum separately to reflect changes. More details about Header format can found in §5.3.
- **Flux:** (BinaryHDU Table or HDF5 Dataset, Array, Float32)
This table/dataset contains final WEAVEIFY flux. The Flux table/dataset includes all the observational effects (such as continuum, noise, resolution mode etc.) that are specified by the user. The shape of this array is same as that of **Wavelength_For_Flux** array.
- **Flux_Without_Noise:** (BinaryHDU Table or HDF5 Dataset, Array, Float32)
This table/dataset contains noiseless flux. The Flux table/dataset includes all the observational effects (except noise) that are specified by the user. Subtraction of **Flux** table/-dataset from **Flux_Without_Noise** gives noise realization. This field would be important for users if they want to test the effect of noise levels on properties of QSO absorption spectra. The shape of this array is same as that of **Wavelength_For_Flux** array.
- **Continuum_Template:** (BinaryHDU Table or HDF5 Dataset, Array, Float32)
This table/dataset contains continuum template used to generate the mock spectrum. The continua are generated randomly for each spectrum, unless random seed is specified. This field would be important for users if they want to test the effect of continuum estimate on properties of QSO absorption spectra. The shape of this array is same as that of **Wavelength_For_Flux** array.
- **SNR_per_pixel:** (BinaryHDU Table or HDF5 Dataset, Array, Float32)
This table/dataset contains Signal-to-Noise ratio used to add the noise the spectra. Since the sensitivity of spectra often changes with wavelength, the SNR per pixel changes with wavelength. The SNR per pixel array is used to generate a random realization of noise that is added to the final spectrum. The shape of this array is same as that of **Wavelength_For_Flux** array.
- **Noise_per_pixel:** (BinaryHDU Table or HDF5 Dataset, Array, Float32)
This table/dataset contains Noise per pixel that is used to add contiminate the spectra.

Since the sensitivity of spectra often changes with wavelength, the noise per pixel changes with wavelength. The noise per pixel array is used to generate a random realization of noise that is added to the final spectrum. The shape of this array is same as that of `Wavelength_For_Flux` array.

5.3 Header format

The header contains the information of all the variables used to generate the mock spectrum. Since the variables change from one spectrum to another, we save the given list of parameters for each spectrum individually. Fits file format does not allow the length of Header keyword (also known as cards) to be more than 8 characters. In Table 1, we provide the list of header keywords and their meaning. Even though hdf5 format allows user to store header keyword to be of any length, we follow the same procedure as fits format. This ensures the consistency of WEAVEIFY output irrespective fits or hdf5 format.

5.4 Script to read WEAVEIFY output

To facilitate users, we provide a python script to read the WEAVEIFY output files. The location of this file is shown in Fig. 17. The file contains a function `read_weaveify_output_fields`. Users can call this function in their python script. The function returns a dictionary that contains header, wavelength arrays and mock spectra arrays. We refer reader to the docstring of the function in file `read_weaveify_output_all_format.py` that explains the argument list in details. This data can be access using following commands,

```
from path.to.location.read_weaveify_output_all_format import \
read_weaveify_output_fields

out_dict = read_weaveify_output_fields(out_path,\
                                       flag_output_format='fits',flag_mpi='n',rank=None,\
                                       spec_id=None,flag_print_output='n')

# Syntax: out_dict[<extension or hdf5 group name>][<field name>]

# For example, to read flux from spec_id=0, we can use
flux_arr = out_dict['MOCK_SPECTRA-0']['Flux']

# To read wavelength from COMMON_FIELDS, we can use
wavelength_arr out_dict['COMMON_FIELDS']['Wavelength_For_Flux']

# To read header dictionary
header_dict = out_dict['MOCK_SPECTRA-0']['Header']
```

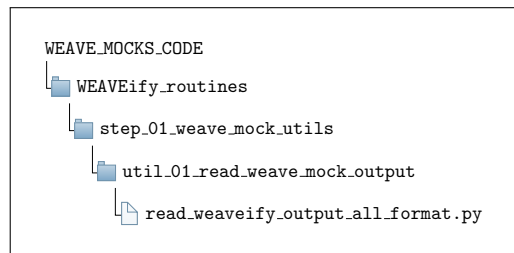


Figure 17: The location of sample python script to read the WEAVEIFY output files.

Table 1: Header keywords / cards in WEAVEIFY output

Header Keyword	WEAVEIFY Variable	Description
OBJMAG	<code>m_source</code>	Source magnitude
OBJWBAND	<code>wave_band_source</code>	Wavelength filter of source magnitude
OBJTYPE	<code>source_type</code>	Type of source (Point or Extended)
INSTR	<code>instrument</code>	Instrument Name
RESMODE	<code>resolution_mode</code>	Resolution mode for WEAVE survey
SPECTYPE	<code>spectra_type</code>	Type of Spectra
DIATEL	<code>dia_telescope</code>	Diameter of telescope in meters
PIXSIZE	<code>ccd_pix_size_micron</code>	Size of CCD pixel in microns
FIBREDIA	<code>dia_fibre_arcsec</code>	Diameter of fibre in arcsec
CRVAL1	<code>CRVAL1</code>	Start of wavelength array in angstroms
CDEL1	<code>CDEL1</code>	Wavelength spacing in angstroms
NPIXEL	<code>N_pixel_obs</code>	Number of pixels in wavelength/flux array
WSCALE	<code>lambda_scale</code>	Wavelength spacing scale (LIN or LOG)
CONTINUM	<code>flag_cont_method</code>	Continuum generation method
LSFTYPE	<code>lsf_type</code>	Functional form of Line Spread Function
NOISTYPE	<code>flag_noise</code>	Type of noise added to spectra
SKYMAG	<code>m_sky</code>	Sky magnitude
SKYWBAND	<code>wave_band_sky</code>	Wavelength Filter of sky magnitude
EXPTIME	<code>t_exp</code>	Exposure time in seconds
SEEING	<code>fwhm_seeing</code>	Atmospheric Seeing
BEAMSIZE	<code>beamsize</code>	Beam size
DARKCURR	<code>dark_current</code>	Amount of Dark current
SEEPRFLE	<code>seeing_profile</code>	Functional form of Seeing Profile
ROFFSET	<code>r_offset</code>	Offset of source from center of fibre in arcsec
READNOIS	<code>read_noise</code>	Amount of read noise
AIRMASS	<code>airmass</code>	Atmospheric airmass
BETAMOFT	<code>beta_moffat</code>	Power law index of MOFFAT Seeing profile
THRTYPE	<code>throughput_type</code>	Type of throughput
FLUXNORM	<code>flux_norm_(erg/s/cm^2/A)</code>	Normalization factor for flux
ZEM	<code>z_em</code>	Emission redshift of QSO
SPECID	<code>spec_id</code>	Spectrum Identification number or String

6 Detail description of free parameters in WEAVEIFY

In this sections, we describe all the parameters and fields that are required by WEAVEIFY to generate mock spectra. We have also provided the default value of the variables.

6.1 QSO parameters

Table 2: List of QSO Parameters

WEAVEIFY Variable	Default Value	Description
<code>z_em</code>	3.0	Emission redshift of QSO
<code>spec_id</code>	None	Unique identification number or string to save data

QSO parameters decides if the emission redshift of the QSO and name of the QSO. List of variables and their description is provided in Table 2. The most important variables is `spec_id`. The variable `spec_id` could be string or number. The data for given spectrum is saved in the output file under `MOCK_SPECTRA-` `spec_id`. Hence user always need to change this variable while making mocks for multiple spectra.

6.2 Wavelength sampling parameters

The variables changing the DLA properties are mentioned in Table 3. A more detail description of these parameters is given in §3.2 and 3.3.

Table 3: Wavelength sampling parameters

WEAVEIFY Variable	Default Value	Description
<code>instrument</code>	WEAVE	Name of the instrument
<code>resolution_mode</code>	LR	Resolution mode HR or LR
<code>CRVAL1</code>	None	Start of wavelength array in angstroms
<code>CDEL1</code>	None	Wavelength spacing in angstroms
<code>N_pixel_obs</code>	None	Number of pixels in wavelength/flux array
<code>lambda_scale</code>	None	Wavelength spacing scale (LIN or LOG)
<code>lsf_type</code>	GAUSSIAN	Functional form of Line Spread Function

6.3 Continuum placement parameters

The variables changing the continuum placements are mentioned in Table 4. A more detail description of these parameters is given in §3.1.

Table 4: Continuum placement parameters

WEAVEIFY Variable	Default Value	Description
<code>flag_cont_method</code>	PCA	Method to generate random continuum
<code>random_seed_cont</code>	None	Seed to generate random continuum
<code>alpha_cont</code>	None	Power-law slope of the continuum $F_{\text{cont}} \propto \lambda^\alpha$

6.4 Noise parameters

The variables changing the noise properties are mentioned in Table 5. A more detail description of these parameters is given in §3.5 and §3.4.

Table 5: Noise parameters

WEAVEIFY Variable	Default Value	Description
<code>flag_noise</code>	OpR3	Type of noise to be added to spectra
<code>source_type</code>	point	Type of source point or extended
<code>dia_telescope</code>	4.2	Diameter of Telescope in meters
<code>dia_fibre_arcsec</code>	1.3045	Diameter of Fibre in arcsec
<code>m_source</code>	20.0	Apparent magnitude of source
<code>wave_band_source</code>	R	Wavelength filter of source magnitude
<code>m_sky</code>	20.92	Apparent magnitude of sky
<code>wave_band_sky</code>	B	Wavelength filter of sky magnitude
<code>t_exp</code>	3060.0	Exposure time in seconds
<code>airmass</code>	1.107	Airmass value for the atmosphere
<code>seeing_profile</code>	Moffat	Functional form of seeing Gaussian or Moffat
<code>beta_moffat</code>	2.5	Moffat profile index
<code>fwhm_seeing</code>	0.75	Full Width at half maximum of seeing in arcsec
<code>read_noise</code>	3.3	Read noise per electron
<code>dark_current</code>	0.0	Noise due to dark current
<code>ccd_pix_size_micron</code>	15.0	CCD pixel size in microns
<code>r_offset</code>	0.1	Offset of fibre from object center in arcsec
<code>throughput_type</code>	optimal	Type of telescope throughput

6.5 Cosmological parameters

Cosmological parameters used for generating mocks are usually fixed for a given simulation box. The cosmological parameters are initialized in function `cosmological_parameters` in file `default_parameters_WEAVEify.py`. *Note: Cosmological parameters can only be changed by modifying the file default_parameters_WEAVEify.py. The cosmological parameters can not be changed by providing as an argument in command line..* A more detail description of cosmological parameters along with their default values is given in Table 6.

Table 6: Cosmological Parameters

WEAVEIFY Variable	Default Value	Description
<code>omega_m</code>	0.308	Matter density parameter Ω_m
<code>omega_l</code>	0.692	Dark Energy density parameter Ω_Λ
<code>omega_r</code>	0.0	Radiation density parameter Ω_γ
<code>omega_k</code>	0.0	Curvature density parameter Ω_k
<code>omega_b</code>	0.0482	Baryon density parameter Ω_b
<code>hubble_param</code>	0.678	Hubble parameter h
<code>sigma_8</code>	0.829	Density fluctuations parameter σ_8
<code>Y</code>	0.24	Primordial Helium abundance by mass Y
<code>rho_c_by_hsq</code>	1.8791e-29	Critical density parameter ρ_c/h^2 , $\rho_c = 3H_0^2/(8\pi G)$
<code>ns</code>	0.961	Index of primordial power spectrum

6.6 MPI job submission parameters

The variables described in Table 7 are used to submit a parallel job using MPI processes. The most important parameter to submit a job in parallel is `flag_mpi`. Users need to set this parameter to 'y' if they want to run job in parallel. The variables `los_indx` and `N_los_total` are used to print the progress of the code. `los_indx` presents the spectrum on which WEAVEIFY is currently working on. While `N_los_total` is total number of spectra for which mock spectra needs to be generated. The details on how to submit a multi-core MPI job is described in §4.2. Note that there are two more variables `N_proc` and `rank` that are internally set when MPI job is submitted. Users do not need to modify these two variables (neither for serial or parallel jobs).

Table 7: MPI job submission Parameters

WEAVEIFY Variable	Default Value	Description
<code>flag_mpi</code>	<code>n</code>	Set this parameter to 'y' if you are using the code in parallel using MPI
<code>N_los_total</code>	<code>1</code>	Total number of mock spectra to be generated. This variable would be internally changed.
<code>los_indx</code>	<code>1</code>	Current mock spectra on which code is working. This variable would be internally changed.

6.7 Input / Output file saving parameters

The input / output directory and file names are set by variables as described in Table 8. Users always need to set the variable `out_path` to the directory where outputs will be stored. The prefix of the output files are set by `out_file_flux_base` and `out_file_tau_base` variables. Users can chose different names of output files. *Users need to make sure that the values of two variables are exactly the same if the optical depths and fluxes need to be saved in same file. WEAVEIFY provides a flexibility of saving the optical depths and fluxes in seperate files. In such case, users need to set the value of `out_file_tau_base` different than the variable `out_file_flux_base`* If the value of variable `flag_make_quick_plot` is set to `y`, the output spectrum is plotted and saved in a file. The variable `out_file_plot_base` is prefix for output plot files. The output of spectrum with `spec_id` would be saved in file `<out_file_plot_base>-<spec_id>.pdf` file.

6.8 Miscellaneous parameters

In addition to abover mentioned parameters / variables, there are additional variables described in file `default_parameters_WEAVEify.py`. These variables do not need to be change. We refer reader to `default_parameters_WEAVEify.py` file for more details.

Table 8: Parameters related to Input / Output

WEAVEIFY Variable	Default Value	Description
<code>out_path</code>	<code>None</code>	Path of directory where output needs to be stored
<code>out_file_flux_base</code>	<code>WEAVEified_Mock_Spectra</code>	Prefix of flux data file
<code>out_file_tau_base</code>	<code>WEAVEified_Mock_Spectra</code>	Prefix of optical depth data file. If OD needs to be saved in separate file, change this variable
<code>flag_output_format</code>	<code>fits</code>	Format of the output file fits or hdf5
<code>flag_make_quick_plot</code>	<code>n</code>	If y makes quick plot
<code>out_file_plot_base</code>	<code>Quick_Plot_Mock_Spectra</code>	Prefix of output plot file

7 Acknowledgement and Citations

8 Frequently Asked Questions (FAQs)

1. *I am running WEAVEIFY on single (or multi) core machine for 2 or more number of spectra. But the output is saved only for one spectrum (the last one)?*

You are not varying the variable spec_id in for loop. This variable tells the code to store the different spectrum in different blocks or HDUBinary Table. Note that spec_id can be a string or integer variable.

2. *How do I install the python packages mentioned in §1.2 ?*

If users have difficulty in installing python packages, they can use following steps to install packages.

- (a) Download and install miniconda using the scripts provided at <https://docs.conda.io/en/latest/miniconda.html>
- (b) In the next step we need to create a conda environment for installing the desired packages.
- (c) We have provided two *yml* files as shown in Fig. 18. Goto the above mentioned directory using cd command.

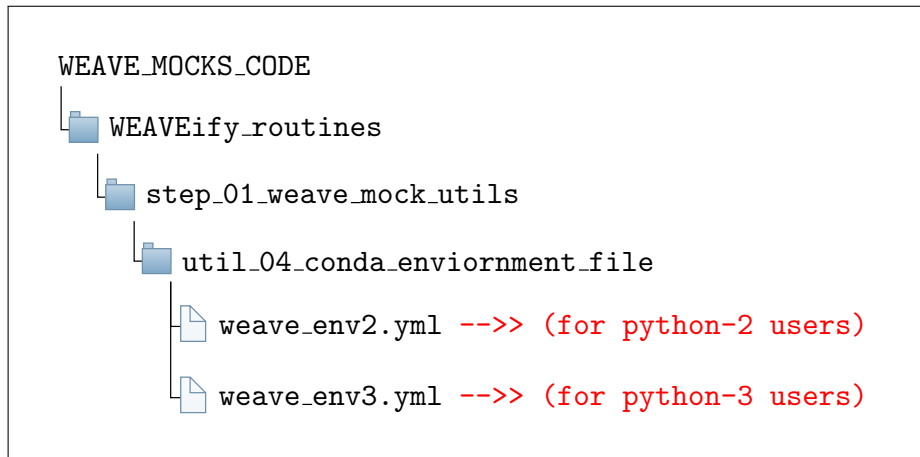


Figure 18: Location of conda enviornment files. The code has been tested for using all the packages mentioned in two yml files.

- (d) To create conda envrionment from the yml files, enter following command in terminal

```
conda env create --file weave_env3.yml -n weave_env3
```

- (e) Python2 users need use *weave_env2.yml* file in the above command to install python2 and associatated packages
- (f) This will install all the packages needed to run WEAVEIFY code.
- (g) To use the recently installed version of python, users need to activate conda enviornment using

```
conda activate weave_env3
```

(h) Similarly users can leave conda environment (after running the code) using

```
conda deactivate
```

(i) The code has been tested on both packages mentioned in two yml files and seems to perform well.