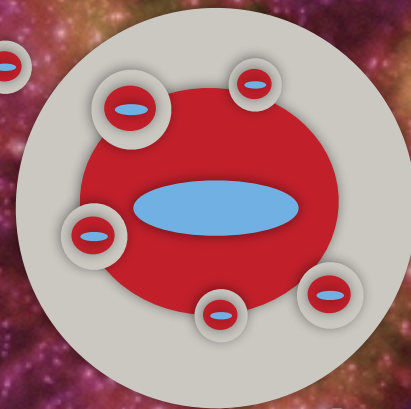


MCMC - first run



MCMC - first run

1. Compiling & Running

In order to perform the MCMC sampling the code should be compiled using `My_Makefile_options_MCMC`. To do so, in Makefile, replace `"My_Makefile_options"` with: `"My_Makefile_options_MCMC"` in lines 23 and 40 of Makefile. Then simply run the executable file generated using an `mcmc input.par` file as a run time argument. MCMC input files have "mcmc" in the name and follow the same naming convention as other input files.

Two input files are currently available:

- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium simulation.
- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_plus_MRll_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium and Millennium II simulations. This allows the model to be constrained over a much broader dynamical range.

There are also bash scripts available to run the code in parallel, in `./AuxCode/Run/`, e.g.: `cosma.bash`. After the burn out phase it is equivalent to have a single MCMC chain with a lot of steps, or a lot of chains with a few steps. Since the galaxy formation model takes a relatively long time to run (from 5 to 10 minutes) it is better to run smaller chains in multiple cores. Typically, one can get convergence after 2 days on 100 cores for 10 parameters.

A fundamental difference between running the code in 'MCMC' or 'normal' mode is the output produced. In 'normal mode', the code runs once in each dark matter sub-volume and one galaxy catalogue is produced per sub-volume. In 'MCMC mode', the code runs in a representative set of dark matter merger trees, for a large number of times, and the output is a single file (per processor), with a list of parameter values and a corresponding likelihood. If the code is not run in parallel, a file is written in the output directory with the comparison between the model and each observational constraint. This can be used for debugging.

MCMC - first run

1. Compiling & Running

In order to perform the MCMC sampling the code should be compiled using `My_Makefile_options_MCMC`. To do so, in Makefile, replace `"My_Makefile_options"` with: `"My_Makefile_options_MCMC"` in lines 23 and 40 of Makefile. Then simply run the executable file generated using an `mcmc input.par` file as a run time argument. MCMC input files have "mcmc" in the name and follow the same naming convention as other input files.

Two input files are currently available:

- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium simulation.
- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_plus_MRll_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium and Millennium II simulations. This allows the model to be constrained over a much broader dynamical range.

There are also bash scripts available to run the code in parallel, in `./AuxCode/Run/`, e.g.: `cosma.bash`. After the burn out phase it is equivalent to have a single MCMC chain with a lot of steps, or a lot of chains with a few steps. Since the galaxy formation model takes a relatively long time to run (from 5 to 10 minutes) it is better to run smaller chains in multiple cores. Typically, one can get convergence after 2 days on 100 cores for 10 parameters.

A fundamental difference between running the code in 'MCMC' or 'normal' mode is the output produced. In 'normal mode', the code runs once in each dark matter sub-volume and one galaxy catalogue is produced per sub-volume. In 'MCMC mode', the code runs in a representative set of dark matter merger trees, for a large number of times, and the output is a single file (per processor), with a list of parameter values and a corresponding likelihood. If the code is not run in parallel, a file is written in the output directory with the comparison between the model and each observational constraint. This can be used for debugging.

MCMC - first run

1. Compiling & Running

In order to perform the MCMC sampling the code should be compiled using `My_Makefile_options_MCMC`. To do so, in Makefile, replace `"My_Makefile_options"` with: `"My_Makefile_options_MCMC"` in lines 23 and 40 of Makefile. Then simply run the executable file generated using an `mcmc input.par` file as a run time argument. MCMC input files have "mcmc" in the name and follow the same naming convention as other input files.

Two input files are currently available:

- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium simulation.
- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_plus_MRll_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium and Millennium II simulations. This allows the model to be constrained over a much broader dynamical range.

There are also bash scripts available to run the code in parallel, in `./AuxCode/Run/`, e.g.: `cosma.bash`. After the burn out phase it is equivalent to have a single MCMC chain with a lot of steps, or a lot of chains with a few steps. Since the galaxy formation model takes a relatively long time to run (from 5 to 10 minutes) it is better to run smaller chains in multiple cores. Typically, one can get convergence after 2 days on 100 cores for 10 parameters.

A fundamental difference between running the code in 'MCMC' or 'normal' mode is the output produced. In 'normal mode', the code runs once in each dark matter sub-volume and one galaxy catalogue is produced per sub-volume. In 'MCMC mode', the code runs in a representative set of dark matter merger trees, for a large number of times, and the output is a single file (per processor), with a list of parameter values and a corresponding likelihood. If the code is not run in parallel, a file is written in the output directory with the comparison between the model and each observational constraint. This can be used for debugging.

MCMC - first run



Max-Planck-Institut für
Astrophysik

1. Compiling & Running

In order to perform the MCMC sampling the code should be compiled using `My_Makefile_options_MCMC`. To do so, in `Makefile`, replace "`My_Makefile_options`" with: "`My_Makefile_options_MCMC`" in lines 23 and 40 of `Makefile`. Then simply run the executable file generated using an `mcmc` input.par file as a run time argument. MCMC input files have "mcmc" in the name and follow the same naming convention as other input files.

Two input files are currently available:

- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium simulation.
- `./input/MCMC_inputs/input_Henriques15_mcmc_MR_plus_MR11_W1_PLANCK.par`: mcmc sampling for the standard Henriques2015 model using a sample of trees from the Millennium and Millennium II simulations. This allows the model to be constrained over a much broader dynamical range.

There are also bash scripts available to run the code in parallel, in `./AuxCode/Run/`, e.g.: `cosma.bash`. After the burn out phase it is equivalent to have a single MCMC chain with a lot of steps, or a lot of chains with a few steps. Since the galaxy formation model takes a relatively long time to run (from 5 to 10 minutes) it is better to run smaller chains in multiple cores. Typically, one can get convergence after 2 days on 100 cores for 10 parameters.

A fundamental difference between running the code in 'MCMC' or 'normal' mode is the output produced. In 'normal mode', the code runs once in each dark matter sub-volume and one galaxy catalogue is produced per sub-volume. In 'MCMC mode', the code runs in a representative set of dark matter merger trees, for a large number of times, and the output is a single file (per processor), with a list of parameter values and a corresponding likelihood. If the code is not run in parallel, a file is written in the output directory with the comparison between the model and each observational constraint. This can be used for debugging.

run the code in MCMC mode:

PARALLEL off, using `input_Henriques15_mcmc_MR_W1_PANCK.par` with `ChainLength =1`

read the comparison with observations, for the selected constraints and redshifts, into the plotting routines - `./output/mcmc_plus_obs*_z*.txt`

MCMC - analysing the output



4. Analysing the Output

For each core, an output file is written into the output directory: `senna_gNN_ii.txt`. Where NN is the last character of the route directory of the model and ii is the core number. In each of these files, one line will be printed per MCMC step containing `weight(normally = 1), -log10(likelihood), log10(P1), log10(P2), . . . , log10(Pn)`.

The output can be analysed using the public COSMO_MC package which contains a number of very useful tests. Alternatively, [mcmc_read_chains.pro](#) has simple idl scripts to produce plots of 1D marginalised regions and of some parameter efficiencies.

`"sort -r -nk2 ./output/senna*"` will sort all the output steps of the MCMC, giving the parameters for the highest likelihood value last.

`"sort -r -nk2 ./output/senna*"` will sort all the output steps of the MCMC, giving the parameters for the highest likelihood value last.