Avcorr - instruction

Contents

Description

Parallelized and multi-functional code for computing area- or space- averaged correlation functions $\bar{\xi}_J$ of orders 2nd-9th, along with cumulants s_J, as functions of scale(s).

It allows for the calculation in four different modes (VERSIONs):

- **angular**: angular counts in $[RA, DEC]$ catalog
- **BOX**: count in spheres within cubic $[X, Y, Z]$ simulation box
- **BOX_ellipses**: counts in ellipsoids within cubic $[X, Y, Z]$ simulation box
- **LC_ellipses**: counts in ellipsoids within $3D[X, Y, Z]$ lightcone

for more details check section about code versions.

1 Requirements, compilation and running

Requirements

- c++17 or newer
- mpi.h library for c++
- libhdf5-cpp

To compile the code, type: **make**
To run, type (X stands for the number of desired parallel jobs):

Running

```
mpiexec -np X ./Avcorr.exe
```

Or use bash script ./*run* by typing: **./run X**
firstly ensuring if it has permissions (chmod +700 ./*run*)

- **Computations can be restarted in any moment if user decided to stop it. The restarted code will just ignore the existing counts.**
- **Input data files need to have the same extension and be located in the same directory as specified in parameter file.**
- **Output scales are always spaced logarithmically**

2 Important files and directories

- Data/ - default location of data for which code calculates the statistics (see description of Datafiles entry in parameter file)
- Results/ - stores results after the run
- Randoms/ - contains optional random catalogs
  (valid if Random_provided=1, see parameter file section)

3 Different code versions, inputs and outputs

angular

Standard counts-in-cells(CiC), counts within circles located on $[RA, DEC]$ sphere.
Input file format: RA, DEC in degrees (datafiles either ASCII or HDF5, see parameter file).

Output files:
Results/[fname]_merrout.txt (contains $\bar{\xi}_J$ results):
θ, $\bar{\xi}_2$, $u_{\bar{\xi}_2}$, $\bar{\xi}_3$, $u_{\bar{\xi}_3}$, $\bar{\xi}_4$, $u_{\bar{\xi}_4}$, $\bar{\xi}_5$, $u_{\bar{\xi}_5}$, $\bar{\xi}_6$, $u_{\bar{\xi}_6}$, $\bar{\xi}_7$, $u_{\bar{\xi}_7}$, $\bar{\xi}_8$, $u_{\bar{\xi}_8}$, $\bar{\xi}_9$, $u_{\bar{\xi}_9}$,

Results/[fname]_merrout_Sn.txt (contains s_J results):
θ, 1, 0, s_3, u_{s_3}, s_4, u_{s_4}, s_5, u_{s_5}, s_6, u_{s_6}, s_7, u_{s_7}, s_8, u_{s_8}, s_9, u_{s_9},

where:
fname - corresponding input datafile without extension (or set of input files, check *Combine_reals* and *Real_template* parameters)
θ - angular scale in degrees, defined by parameters $Rmin, Rmax, nR$ (check parameters section)
$\bar{\xi}_J/s_J$ - J-th order averaged correlated function/hierarchical amplitude
$u_{\bar{\xi}_J}/u_{s_J}$ - error of $\bar{\xi}_J/s_J$, check error calculation section. The 1,0 values at 2nd and 3rd column of _merrout_Sn files stand for consistency of the format since $s_2 = 1$ by definition.
Output files will have additional columns if *ErrPoisson* parameter is set to 1 - check *ErrPoisson* description.

BOX

Counts within spheres in BOX (each axis in $[0, boxsize]$ range)
Input file format: X, Y, Z, units the same as the units of sphere sizes (datafiles either ASCII or HDF5, see parameter file).

Output files:
Results/[fname]_merrout.txt (contains $\bar{\xi}_J$ results):
R, $\bar{\xi}_2$, $u_{\bar{\xi}_2}$, $\bar{\xi}_3$, $u_{\bar{\xi}_3}$, $\bar{\xi}_4$, $u_{\bar{\xi}_4}$, $\bar{\xi}_5$, $u_{\bar{\xi}_5}$, $\bar{\xi}_6$, $u_{\bar{\xi}_6}$, $\bar{\xi}_7$, $u_{\bar{\xi}_7}$, $\bar{\xi}_8$, $u_{\bar{\xi}_8}$, $\bar{\xi}_9$, $u_{\bar{\xi}_9}$,

Results/[fname]_merrout_Sn.txt (contains s_J results):
R, 1, 0, s_3, u_{s_3}, s_4, u_{s_4}, s_5, u_{s_5}, s_6, u_{s_6}, s_7, u_{s_7}, s_8, u_{s_8}, s_9, u_{s_9},

where:
fname - corresponding input datafile without extension (or set of input files, check *Combine_reals* and *Real_template* parameters)
R - scale in the same units as in catalog, defined by parameters $Rmin, Rmax, nR$ (check parameters section)
$\bar{\xi}_J/s_J$ - J-th order averaged correlated function/hierarchical amplitude
$u_{\bar{\xi}_J}/u_{s_J}$ - error of $\bar{\xi}_J/s_J$, check error calculation section. The 1,0 values at 2nd and 3rd column of _merrout_Sn files stand for consistency of the format since $s_2 = 1$ by definition.
Output files will have additional columns if *ErrPoisson* parameter is set to 1 - check *ErrPoisson* description.

BOX_ellipses

Counts within ellipses in BOX (each axis in [0,boxsize] range). Ellipses have axes: axa along X axis, axb along both Y and Z axes. The code creates a grid of axa and axb values with ranges defined in parameter file and for each set of [axa,axb] computes the statistics. Option useful for investigating catalogs with RSD along X axis.
Input file format: X, Y, Z, units the same as the units of sphere sizes (datafiles either ASCII or HDF5, see parameter file).

Output files:
Results/[fname]_merrout.txt (contains $\bar{\xi}_J$ results):
axa, axb, $\bar{\xi}_2$, $u_{\bar{\xi}_2}$, $\bar{\xi}_3$, $u_{\bar{\xi}_3}$, $\bar{\xi}_4$, $u_{\bar{\xi}_4}$, $\bar{\xi}_5$, $u_{\bar{\xi}_5}$, $\bar{\xi}_6$, $u_{\bar{\xi}_6}$, $\bar{\xi}_7$, $u_{\bar{\xi}_7}$, $\bar{\xi}_8$, $u_{\bar{\xi}_8}$, $\bar{\xi}_9$, $u_{\bar{\xi}_9}$,

Results/[fname]_merrout_Sn.txt (contains s_J results):
axa, axb, 1, 0, s_3, u_{s_3}, s_4, u_{s_4}, s_5, u_{s_5}, s_6, u_{s_6}, s_7, u_{s_7}, s_8, u_{s_8}, s_9, u_{s_9},

where:
fname - corresponding input datafile without extension (or set of input files, check *Combine_reals* and *Real_template* parameters)
axa, axb - semi axes of ellipsoid, parallel and perpendicular to line-of-sight (LOS)
$\bar{\xi}_J/s_J$ - J-th order averaged correlated function/hierarchical amplitude
$u_{\bar{\xi}_J}/u_{s_J}$ - error of $\bar{\xi}_J/s_J$, check error calculation section. The 1,0 values at 3rd and 4th column of _merrout_Sn files stand for consistency of the format since $s_2 = 1$ by definition.
Output files will have additional columns if *ErrPoisson* parameter is set to 1 - check *ErrPoisson* description.

LC_ellipses

Counts within line-of-sight-elongated ellipses in 3D catalog made from lightcone. Ellipses have axes: axa along line-of-sight(LOS), axb along two transverse directions. The code creates a grid of axa and axb values with ranges defined in parameter file and for each set of [axa,axb] computes the statistics. Option useful for investigating higher-order statistics in 3D lightcones including RSD effects.
Input file format: X, Y, Z, units the same as the units of sphere sizes (datafiles either ASCII or HDF5, see parameter file).

Output files:
Results/[fname]_merrout.txt (contains $\bar{\xi}_J$ results):
axa, axb, $\bar{\xi}_2$, $u_{\bar{\xi}_2}$, $\bar{\xi}_3$, $u_{\bar{\xi}_3}$, $\bar{\xi}_4$, $u_{\bar{\xi}_4}$, $\bar{\xi}_5$, $u_{\bar{\xi}_5}$, $\bar{\xi}_6$, $u_{\bar{\xi}_6}$, $\bar{\xi}_7$, $u_{\bar{\xi}_7}$, $\bar{\xi}_8$, $u_{\bar{\xi}_8}$, $\bar{\xi}_9$, $u_{\bar{\xi}_9}$,

Results/[fname]_merrout_Sn.txt (contains s_J results):
axa, axb, 1, 0, s_3, u_{s_3}, s_4, u_{s_4}, s_5, u_{s_5}, s_6, u_{s_6}, s_7, u_{s_7}, s_8, u_{s_8}, s_9, u_{s_9},

where:
fname - corresponding input datafile without extension (or set of input files, check *Combine_reals* and *Real_template* parameters)
axa, axb - semi axes of ellipsoid, parallel and perpendicular to line-of-sight (LOS)
$\bar{\xi}_J/s_J$ - J-th order averaged correlated function/hierarchical amplitude
$u_{\bar{\xi}_J}/u_{s_J}$ - error of $\bar{\xi}_J/s_J$, check error calculation section. The 1,0 values at 3rd and 4th column of _merrout_Sn files stand for consistency of the format since $s_2 = 1$ by definition.
Output files will have additional columns if *ErrPoisson* parameter is set to 1 - check *ErrPoisson* description.

Common output files (for every code mode (VERSION))

Output files with final results are described in text above, separately for each code version.

Results/[fname]_moments_nnR_X.fndhst
(contains histogram of counts from X-th scale):
N, n_c,
where N - number of objects within circle, n_c - number of circles with N objects.
Results/log.txt
(contains log of the run)

Results/*_moments_*.txt (temporary files with yet-unmerged results, will be removed at the end if $Clean = 1$ parameter is set)

4 Randoms

Random files (must be located in Randoms/ directory) used in the code are necessary only if *Random_provided* = 1 in parameter file. The format is exactly the same as data files for corresponding VERSION (version is set in parameter file). Randoms define positions of circles/spheres/ellipses for making counts. Therefore it is crucial not to use randoms with positions exceeding the catalog ranges. Check description of Random_provided parameter in parameter file.

5 Parameter file

The param.txt file contains the code settings. For clarity it is divided into parts containing parameters working only for specific code VERSION (angular/BOX/ BOX_ellipses/LC_ellipses). The parameters are:

- **VERSION** - code mode: angular/BOX/ BOX_ellipses/LC_ellipses. Switches between the modes, based on this different data formats and outputs will be used.
- **USE_HDF5** - input data format: 0 - ASCII, 1 - HDF5
- **POS_DSET** - dataset with coordinates (applies only if only if USE_HDF5=1)
- **cols_pos** - columns with coordinates (applies only if only if USE_HDF5=0)

Common parameters (used for **every** VERSION):

- **nreals** - number of sub-probes for errors calculation. Counts are splitted into [nreals] parts, then for each sub-probe the code calculates the statistics. Output result is mean from sub-probes +/- one standard deviation. Check the section about error calculations.
- **kappa** - fraction of statistically independent information contained within one sub-probe, i.e. for $VERSION = BOX$ number of spheres considered for each sub-probe will be: $kappa \times boxsize^3/(4/3\pi R^3)$, where R is the sphere size. Use $kappa << 1$ to avoid error underestimation (otherwise every sub-probe will contain approximately the same information) and $kappa \times nreals > 1$ to extract full information. Check the section about error calculations.
- **Cmin** - minimum number of circles/spheres/ellipses for CiC
- **Cmax** - maximum number of circles/spheres/ellipses for CiC
- **Datafiles** - datafile names from separated by space/tab (with extension). Examples:
  - If Datafiles=*, the code selects all files in Data/ directory
  - If Datafiles=path/*, the code selects all files from path/ directory
  - If Datafiles=file1.h5 file2.h5, the code selects file1.h5 and file2.h5 from Data/ directory, analogically for specified path/
  - [Attention]: all files needs to have the same extension and be located in the same directory
  - [Attention]: if * was used firstly, only files corresponding with that entry will be considered, e.g. for Datafiles=path/* file2.dat, code will read only files in path/

- **Random_provided** - random file(s) provided? 0 - drawing random, 1 - reading random file (s) [check Random_file description].
  Drawing randoms:
  - for $VERSION = angular$: random points within [RA,DEC] rectangle, not closer than current circle size to catalog borders
  - for $VERSION = BOX$: random points in $[R, boxsize - R]^3$ cube, where R - current sphere size
  - for $VERSION = BOX\_ellipses$: random points in $[ax, boxsize - ax]$ ranges, where ax=axa for axis X and ax=axb for axes Y and Z, check parameters for BOX_ellipses and LC_ellipses
  - for $VERSION = LC\_ellipses$: random points between spheres of radii DCMIN+axa and DCMAX-axa. [Attention:] for LC_ellipses random drawer ignores any angular cuts, it just draws them for all directions. If your catalog has angular cuts, use own randoms.

- **Random_file** - random file name (if specific file name imposed here, every datafile will be assigned to the same random file); if $Random\_file = *$, code reads multiple randoms: for each datafile, random filename is assumed to be: Randoms/Randoms_[datafile] will be assigned to each [datafile]. For example, if $Random\_file = *$ and Data/ contains catalogs A.txt and B.txt, then Randoms/ directory has to contain Randoms_A.txt and Randoms_B.txt files. Random file(s) format for ASCII datafiles ($USE\_HDF5 = 0$) is assumed to have only columns with coordinates, while for HDF5 datafiles ($USE\_HDF5 = 0$) the positions are assumed to be within POS_DSET dataset.

- **Recalc** - [0/1] if code is restarted and counts already exist, recalculating the moments [1], or not [0] to save the time

- **Clean** - [0/1] cleaning unnecessary files at the end

- **ErrPoisson** - [0/1] switch whether we want Poisson errors calculation or not. For $ErrPoisson = 0$, there are no columns with Poisson errors in output file. If $ErrPoisson = 1$, the errors for orders 2-9 are added to the file, so the format is like:
  R, $\bar{\xi}_2$, $u_{\bar{\xi}_2}$, $\bar{\xi}_3$, $u_{\bar{\xi}_3}$, $\bar{\xi}_4$, $u_{\bar{\xi}_4}$, $\bar{\xi}_5$, $u_{\bar{\xi}_5}$, $\bar{\xi}_6$, $u_{\bar{\xi}_6}$, $\bar{\xi}_7$, $u_{\bar{\xi}_7}$, $\bar{\xi}_8$, $u_{\bar{\xi}_8}$, $\bar{\xi}_9$, $u_{\bar{\xi}_9}$,
  P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9,
  where P_J is Poisson error for order J.

- **Combine_reals** - [0/1] switch whether our files are different realisations of one model (check Real_template parameter)

- **Real_template** - template telling how are realisations marked if $Combine\_reals = 1$, e.g. if $Real\_template =$ "$box * \_$", and the data files are AAbox1_data, AAbox2_data, AAbox3_data, the code will combine them: the result will be AAboxN_data_merrout.txt: averaged over realisations and errors from standard deviation. One can have any number of files with more than one model, for example: Files: AAbox1_data, AAbox2_data, AAbox3_data, BBbox1_data, BBbox2_data, BBbox3_data, BBbox4_data, will be combined into AAboxN_data_merrout.txt (3 realisations included) and BBboxN_data_merrout.txt (4 realisations included), just Real_template must be the same.
  If $Random\_file = *$, the Randoms/ directory must still contain randoms for each file in Data/ separately, even if these are the same model but different realisations

Parameters for $VERSION$ = **angular** only:

If $Random\_file = *$ and the sky footprint is complicated, right-ascension and declination ranges in param.txt file have to exceed extreme values from catalogs - for pixelization purposes

- **ramin** - catalog right-ascension lower range in degrees
- **ramax** - catalog right-ascension upper range in degrees
- **decmin** - catalog declination lower range in degrees
- **decmax** - catalog declination upper range in degrees
- **Areaf** - catalog sky area in square degrees (for optimization purposes)

Parameters both for $VERSION =$ **angular** and **BOX**:

- **Rmin** - smallest scale considered
- **Rmax** - biggest scale considered
- **nR** - number of scales considered

Parameters both for $VERSION =$ **BOX** and **BOX_ellipses**:

- **Boxsize** - box size in the same units as random spheres/ellipses

Parameters both for $VERSION =$ **BOX_ellipses** and **LC_ellipses**:

- **axamin** - lower range of ellipses axes along X axis (VERSION=BOX_ellipses) / LOS (VERSION=LC_ellipses)
- **axamax** - upper range of ellipses axes along X axis (VERSION=BOX_ellipses) / LOS (VERSION=LC_ellipses)
- **axbmin** - lower range of ellipses axes perpendicular to X axis (VERSION=BOX_ellipses) / LOS (VERSION=LC_ellipses)
- **axbmax** - upper range of ellipses axes perpendicular to X axis (VERSION=BOX_ellipses) / LOS (VERSION=LC_ellipses)
- **naxa** - number of different axes along X/LOS considered - grid size
- **naxb** - number of different axes perpendicular to X/LOS considered - grid size

Parameters for $VERSION =$ **LC_ellipses** only:

- **DCMIN** - minimum comoving distance (from observer) considered, the same units as for ellipse axes
- **DCMAX** - minimum comoving distance (from observer) considered, the same units as for ellipse axes

6 Errors calculation

Combine_reals set to 0

Errors are being calculated by splitting full counts into $[nreals]$ sub-probes (without returning) and computing standard deviation on the statistics calculated for each sub-probe. Additionally, the code provides independently Poisson errors assuming that $error = \sqrt{counts}$ and calculating the errors of $\bar{\xi}_J/s_J$ using error propagation.

Combine_reals set to 1

Combining results from files according to Real_template (check Parameter file section), computing average and standard deviation. The code finds independent models and number of realizations automatically.

7 Example runs and plots

The directory Examples/ contains example results for different configurations along with data and parameter file(s) necessary for recreating them. Each example also contains python plotting scripts for visualization. For more details, check Examples/[example]/ReadMe.txt file. Examples have been computed on data from $COLAVERSE$ simulation prepared by Prof. Wojciech Hellwing and described in Drozda et al. (2025).

8 Troubleshooting and future improvements

The code predicts potential errors in parameter file (mismatching number of columns, nonphysical values i.e. $Rmin < 0$, empty *Data/* directory and many other). If the code breaks or unexpected behavior occurs:

- check *Results/log.txt* file. In case of problems in param.txt it should report errors in lines starting with [Error]
- check if datafiles do not contain NaN, INF or other problems inside the data
- check if proper format is specified - code will break if USE_HDF5 does not match data type
- if ASCII data is used ($USE\_HDF5 = 0$), check if file headers (lines starting with # are only at the beginning of files and are not separated by non - #- starting lines)
- check if data files have the same extension and are located in the same directory

Future improvements to add:

- faster MPI - threads that finish their part need to do remaining jobs assigned to slower ones
- periodicity in $VERSION = angular$
- generalize optimal pixelization into analytical expressions

9 Citation & references

The code uses equations from Gaztanaga (1994) to compute connected and shot-noise corrected moments of counts in cells.
While using this code, please cite Drozda et al. (2025) and refer to:
$https://github.com/Pawel-96/Avcorr$.

References

P. Drozda, W. A. Hellwing, and M. Bilicki. Anisotropic counts-in-cells in redshift space: A new route to cosmological constraints from galaxy surveys, 2025. URL https://arxiv.org/abs/2506.01762.

E. Gaztanaga. High-Order Galaxy Correlation Functions in the APM Galaxy Survey. *Monthly Notices of the Royal Astronomical Society*, 268:913, June 1994. doi: 10.1093/mnras/268.4.913.