HOD.x: User's Manual

1. Basic Usage

To execute HOD.x, simply type `HOD.x [hod.bat]` at the command prompt. The code will then execute all the tasks laid out for it in the `hod.bat` file. The code will not run without the bat file.

2. Input Variables

The code will accept a large number of input parameters, some of which are required and some that are optional (or dependent on other parameters). Basically, this breaks down into the cosmology you have chosen, the HOD parameterization you have chosen and the values for the HOD parameters, and what you want the code to do.

The parameters of the `hod.bat` file can be in any order. Comments can be placed in the file after a '%' character. My thanks to Volker Springel, since I nicked this code from GADGET.

2.1. Cosmology and Transfer Function Options

• OMEGA_M	This is the matter density in units of the critical density.
• SIGMA_8	This is the amplitude of the matter power spectrum.
• OMEGA_B	This is the baryon density Ω_b (for ITRANS=5).
• HUBBLE	Hubble parameter—"little" $h \equiv H_0/100\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ (For ITRANS=5).
• GAMMA	growth index of structure: $d\ln D/d\ln a = \Omega_m(z)^\gamma$
• SPECTRAL_INDX	Spectral index n_s of the primordial power spectrum.
• ITRANS	Variable for selecting the type of transfer function you want.
	TRANSFER FUNCTION OPTIONS
• ITRANS=0	No transfer function. Scale-invariant power spectrum.
• ITRANS=4	Efstathiou et al. (1992) $T(k)$
• ITRANS=5	Eisenstein & Hu (1999) $T(k)$ (**Best option**)
• ITRANS=11	Takes a tabulated file for $T(k)$.
	Must be in CAMB format.

2.2. Halo Occupation Parameterizations

In the HOD, the PDF $P(N|M)$ is broken in to a mean occupation function, $\langle N\rangle_M$, and a dispersion about that mean $P(N|\langle N\rangle_M)$. Since we're generally only concerned with two-point

statistics, the second moment of the PDF is all that's needed. The mean occupation function is further broken into central and satellite galaxies. This is key: because central galaxies have a preferred location, they have different clustering properties than satellites (and dark matter). They also have a different dispersion about the mean, since there can only be one of them per halo. For satellites, the code assumes Poisson dispersion. You can look in Zheng et al. (2005) for discussion of this topic.

Here are the options for the `hod.bat` file.

- pdfc — $\langle N_{\rm cen}\rangle_M$ for central galaxies (integer value). Best option = 2.
- pdfs — $\langle N_{\rm sat}\rangle_M$ for satellite galaxies (integer value). Best option = 11.
- CVIR_FAC — ratio between concentration of satellite galaxies and the dark matter halo (float).
- EXCLUSION — How the small-scale halo pairs are treated. More below. Best option = 4.
- M_min — Minimum mass scale for central galaxies.
- M1 — Satellite mass scale.
- M_cut — Cutoff scale for satellite galaxies.
- alpha — Power-law for satellite occupation function.
- sigma_logM — $\sigma_{\log M}$: "Softness" of central occupation cutoff.
- GALAXY_DENSITY — Space density of the galaxy sample.

For `GALAXY_DENSTY`, the code can compute this number from a full HOD since

$$\bar{n}_g = \int \frac{dn}{dM} \langle N\rangle_M, \tag{1}$$

where dn/dM is the halo mass function. If `M_min` is specified, the code will compute `GALAXY_DENSITY` for you (although you still need to put something in the `hod.bat` file). If `M_min=0`, the code will compute `M_min` given the value of `GALAXY_DENSITY`.

Here are some options for central occupation functions. There are plenty more in the file `hod_function.c`, but these will do for most all purposes. Interested users can look in the C code and eve implement their own occupation functions. Most of these assume that you're fitting a luminosity threshold sample, but we'll tale about a bin presently.

$$\langle N_{\rm cen}\rangle_M = 1 \ \ (M > M_{\rm min}) \tag{2}$$

pdfc=1. This is a simple step function where $\langle N_{\rm cen}\rangle_M = 0$ below $M_{\rm min}$, which is defined as the minimum halo mass that can house a galaxy in the given sample. This assumes no scatter between mass and luminosity.

$$\langle N_{\rm cen}\rangle_M = \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{3}$$

pdfc=2. This is a smooth transition between 1 and 0 galaxies. By construction, $M_{\rm min}$ is now the mass at which half the halos have a central galaxy in the sample. The parameter $\sigma_{\log M}$ physically relates to the scatter of galaxy luminosity at fixed mass (and assumes that the distribution is a lognormal; see the discussion in Zheng et al. 2007). For bright samples ($L > L_\star$), this scatter is important for fitting the amplitude of the $\xi(r)$ in the two-halo regime. Scatter brings down the amplitude by including lower-mass halos, providing a better fit to observations (see, eg, Zehavi et al. 2005). This is your best option for fitting data. If you have ratty data and don't want to add an extra free parameter, just set $\sigma_{\log M}$ to 0.2.

$$\langle N_{\rm cen}\rangle_M = 1 \;\; (M_{\rm min} < M < M_{\rm max\_cen}) \tag{4}$$

pdfc=7. This is for a bin in magnitude rather than a threshold sample. In a bin sample, there is a maximum halo mass at which central galaxies can exist; at higher masses, central galaxies will be too bright to be in the sample. In practice, it's hard to constrain $M_{\rm max\_cen}$ from one sample; it's easier to use $M_{\rm min}$ from the next brighter sample.

For satellite galaxies, there are also many options that can be seen in `hod_functions.c`. We list some of the more pertinent examples here, and also note that users can implement their own in the code.

$$\langle N_{\rm sat}\rangle_M = \left(\frac{M}{M_1}\right)^\alpha \tag{5}$$

pdfs=1. Simple power law, where M_1 is the mass at which halos have, on average, one satellite. The parameter α is the power-law index. In the code, $\langle N_{\rm sat}\rangle_M$ is not allowed to be larger than $\langle N_{\rm cen}\rangle_M$. Thus, if `pdfs=1` is used in conjunction with `pdfc=1`, halos below $M_{\rm min}$ are not allowed to have satellites.

$$\langle N_{\rm sat}\rangle_M = \left(\frac{M - M_{\rm cut}}{M_1}\right)^\alpha. \tag{6}$$

pdfs=2. This adds the additional freedom of a cutoff in the satellite occupation function at a mass scale $M_{\rm cut}$. At $M < M_{\rm cut}$, $\langle N_{\rm sat}\rangle_M = 0$. Strictly speaking, M_1 no longer is the mass where $\langle N_{\rm sat}\rangle_M = 1$ unless $M_{\rm cut} \ll M_1$, but since the parameter `M1` in the `hod.bat` file is what's used for this variable in the code, we'll stick to that notation.

$$\langle N_{\rm sat}\rangle_M = \left(\frac{M}{M_1}\right)^\alpha \exp\left(\frac{-M}{M_{\rm cut}}\right) \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right] \tag{7}$$

pdfs=11. The gigantic mess makes sense when you parse it up. One of the problems with `pdfs=2` is that the shape of the cutoff is quite abrupt; this can cause $\langle N_{\rm sat}\rangle_M$ functions that rapidly rise from

0 to > 1 in a very short mass range, giving rise to occupation functions that look unphysical. The exponential cutoff is much softer and compares well with results from semi-analytic modeling. The far-left term in this function is the same as `pdfc=2`. Essentially this means that we're multiplying the satellite occupation function by the central occupation function, guaranteeing that the satellite occupation cuts off at a mass higher than the central occupation cutoff (see the discussion in Wake et al. 2008).

$$\langle N_{\rm sat} \rangle_M = \left(\frac{M}{M_1} \right)^{\alpha} \frac{1}{2} \left[1 + {\rm erf} \left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}} \right) \right] \tag{8}$$

pdfs=10. This is the same as pdfs=11, but without the exponential cutoff. The only cutoff enforced is from the central occupation function.

2.3. Tasks the Code will Perform

If the code is run with no tasks set, it will compute things like the galaxy density, bias, satellite fraction, etc, for the input HOD parameters. It will output these to the screen and then terminate. To get the code to do more, set one of the following keywords to 1 in the `hod.bat` file. Tasks it will perform are:

- clustering — Outputs the correlation function (both $\xi(r)$ and $w_p(r_p)$).
- HOD — Outputs the occupation function.
- wp_minimization — Fits input data to determine an HOD.
- populate_simulation — Take a halo catalog and populate with galaxies.
- MCMC — Run a Monte Carlo Markov Chain on input data.

More on these below, but the code will also output many different aspects of the dark matter distribution as well.

- matter_xi — Outputs $\xi(r)$ (linear and non-linear, using Smith et al. 2003)
- matter_pk — Outputs $P(k)$ (linear and non-linear, using Smith et al. 2003)
- sigma_r — Outputs matter variance $\sigma(r)$ and $\sigma(M)$.
- cvir — Outputs halo concentrations as a function of halo mass.
- halostats — Outputs the differential mass function, dn/dM, and the bias function $b(M)$.

For all the above tasks, the format of the output file is listed in the C file `tasks.c`, which is fairly well documented. For each task, an output file will be created. Each file will have a common `root_filename`, to be set in `hod.bat`. ie, `HOD 1` with `root_fileame test1` creates a file called `test1.HOD`. See more documentation in `tasks.c`.

3. Fitting Data

For either `wp_minimization` or `MCMC`, several additional options and parameters exist. **If you are fitting data, the code expects you to supply a value for `GALAXY_DENSITY`. If you set `M_min` in the `hod.bat` file, it will set it to zero in expectation of re-setting it from the data.**

- fname_wp — Filename for $w_p(r_p)$ data (or $\xi(r)$).
- fname_covar — Filename for covariance matrix (even if not using it).
- COVAR — Flag for using the full covariance matrix.
- DEPROJECTED — Flag for fitting $\xi(r)$ rather than $w_p(r_p)$.
- pi_max — Maximum π for integrating $\xi(r)$ to get $w_p(r_p)$.
- POWELL — Flag for using Powell's method for χ^2 minimization.

Notes: Observationally, $w_p(r_p)$ is obtained by integrating the two-dimensional redshift-space correlation function $\xi(r_p, \pi)$ along the line-of-sight direction π. Because $\xi(r_p, \pi)$ can only be measured robustly output to a value $\pi_{\rm max}$, the HOD code must have this same value. For SDSS data, $\pi_{\rm max} = 40\ h^{-1}\,$Mpc (which is also the default value in the code). For 2dFGRS data, this value is usually 70 $h^{-1}\,$Mpc. See, eg, Norberg et al. 2002; Zehavi et al. 2005.

File Formats: The format for `fname_wp` is three columns:

1. radius [comoving $h^{-1}\,$Mpc]
2. $w_p(r_p)$ or $\xi(r)$
3. diagonal error bar

If `COVAR==0`, the error bar in the `fname_wp` will be used to calculate χ^2. The format of the `fname_covar` file is:

```
for i=1,ndata
    for j=1,ndata
        read C(i,j)
    endfor.
enfor
```

Free Parameters: Users may wish to keep some HOD parameters constant and others leave free. When minimizing, the code will only vary parameters with `free[n]==1`, set in the `hod.bat` file:

Minimization method: There are two options for χ^2 minimization: Powell's method (`POWELL==1`) and Amoeba (`POWELL==0`). Both are taken from Numerical Recipes. Powell's method usually converges more slowly but is more robust. However, sometimes Powell can get caught in a local minimum while Amoeba is more likely to find its way out to the true minimum.

free[1]	M_min
free[2]	M1
free[3]	alpha
free[4]	M_cut
free[5]	sigma_logM
free[6]	CVIR_FAC

MCMC: The Monte Carlo Markov Chain method (`MCMC==1`) is not a fast method for finding the χ^2 minimization, but it allows one to estimate the errors on the parameters given the input data. If this option is chosen, a file called `[root_filename]_[iseed].MCMC` is created which contains the following:

col 1:	Element in the chain (acceptance).
col 2:	Trial number.
col 3-3+n:	The n free parameters.
col 4+n:	The χ^2 for that model.

Note that masses (ie, $M1$ and $M_{\rm cut}$) as well as $\sigma_{\log M}$ are stepped in $\log_{10}$. Normally in MCMC, when a trial is rejected the previous position is simply added again to the chain. The code doesn't do this to reduce the size of the output file, but users can use the trial number in column 2 to determine how to weight each element.

The code has a default `iseed=555`, but can accept a different iseed at the command prompt: `HOD.x hod.bat iseed`. This way, you can start multiple chains with the exact same hod.bat file and not have to change the root_filename within it.

4. Defining a Halo

DELTA_HALO: All halo occupation calculations require models for the halo mass function, the halo bias function, scale-dependent bias, and a model for the matter correlation function. Halos are defined at a given overdensity Δ, ie,

$$M_\Delta = \frac{4}{3}\pi R_\Delta \rho_{\rm crit} \Omega_m \Delta, \tag{9}$$

where I have defined Δ to be with respect to the **mean matter density**. To make the halos a constant Δ with respect to the critical density, simply divide $\Delta_{\rm crit}$ by Ω_m. The code requires that you specify Δ with the parameter `DELTA_HALO`. For the halo mass function, the code uses the results of Tinker et al. (2008). The code uses the bias function calibrated from the same simulations from Tinker et al. (2010). For any value of $180 < \Delta < 1600$, the code will calculate the appropriate mass function for the desired halo definition. The will also calculate the halo bias function for any

value of Δ as well, based on the results of Tinker et. al. (in preparation). Halo bias is non-linear as scales inside of a few $h^{-1}\,\mathrm{Mpc}$. The code incorporates scale dependent bias similar to the fitting formula of Tinker et al. (2005), but adjusted slightly to better match new simulations (see Tinker et al. 2012).

If you look in the files `halo_mass_function.c` and `halo_bias.c`, you will see other options for the mass function and bias relation; eg, Jenkins et al. (2001), Warren et al. (2006), Sheth & Tormen (1999). If you prefer to use those (**only applicable for $\Delta \sim 200$**), the code can be easily adjusted to use one of these options.

Halo Exclusion: The parameter `EXCLUSION` defines how close halo pairs are treated. It is not physically reasonable to have two $10^{15}\ h^{-1}\,\mathrm{M}_\odot$ halos at a separation of a few hundred $h^{-1}\,\mathrm{kpc}$. `EXCLUSION==1` is the most common type of halo exclusion, used in Zheng (2004) and Zehavi et al. (2005), integrates over all halos with $R_\Delta \leq r/2$, where r is the pair separation for which $\xi(r)$ is being calculated. This underestimates the number of close halo pairs. A better option (`EXCLUSION==2`) is to include all pairs for which the *sum* of their radii is less than the separation, $R_{\Delta,1} + R_{\Delta,2} \leq r$. What I feel is the best option is to allow halos to overlap but without allowing the center of one halo to be within the radius of another halo. This option `EXCLUSION==4` is similar to option 1, but now the integral is over all halo pairs with $R_\Delta \leq r$. This definition corresponds to how halos are defined in the spherical overdensity approach used in Tinker et al. (2008).

5. Using the code

Like all things, `HOD.x` is a garbage in/garbage out program. If you provide an unphysical occupation function to the code, some of the integrals will not converge and the code will crash. I have attempted to provide some error checking for this, but it's not possible to cover all possibilities. During χ^2 minimization, the code will some times hit on one of these unphysical HODs. When this happens, the code will slow to a crawl and output like `''Too many steps in qromo.''` will appear to stderr. If this occurs, try restarting the minimization from a different starting location[1] One can use published HODs like those found in Zehavi et al. (2005), Tinker et al. (2005), and Zheng et al. (2007) as starting models, matching up the number density of the sample you are modeling with the SDSS or DEEP2 sample with the closest number density.

The code provided is based mostly the models presented in Zheng (2004), Tinker et al. (2005), and Tinker et al. (2012). If you use this code for part of a published project, please cite these papers.

[1] I admit this sounds a bit like "have you tried rebooting your computer?" But it works a large fraction of the time.

REFERENCES

Efstathiou, G., Bond, J. R., & White, S. D. M. 1992, MNRAS, 258, 1P

Eisenstein, D. J. & Hu, W. 1999, ApJ, 511, 5

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372

Norberg, P., Baugh, C. M., Hawkins, E., Maddox, S., Madgwick, D., Lahav, O., Cole, S., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Lewis, I., Lumsden, S., Peacock, J. A., Peterson, B. A., Sutherland, W., & Taylor, K. 2002, MNRAS, 332, 827

Sheth, R. K. & Tormen, G. 1999, MNRAS, 308, 119

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., & Couchman, H. M. P. 2003, MNRAS, 341, 1311

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. 2008, ApJ, 688, 709

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., Klypin, A., Warren, M. S., Yepes, G., & Gottlöber, S. 2010, ApJ, 724, 878

Tinker, J. L., Sheldon, E. S., Wechsler, R. H., Becker, M. R., Rozo, E., Zu, Y., Weinberg, D. H., Zehavi, I., Blanton, M. R., Busha, M. T., & Koester, B. P. 2012, ApJ, 745, 16

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41

Wake, D. A., Sheth, R. K., Nichol, R. C., Baugh, C. M., Bland-Hawthorn, J., Colless, M., Couch, W. J., Croom, S. M., de Propris, R., Drinkwater, M. J., Edge, A. C., Loveday, J., Lam, T. Y., Pimbblet, K. A., Roseboom, I. G., Ross, N. P., Schneider, D. P., Shanks, T., & Sharp, R. G. 2008, MNRAS, 387, 1045

Warren, M. S., Abazajian, K., Holz, D. E., & Teodoro, L. 2006, ApJ, 646, 881

Zehavi, I., Zheng, Z., Weinberg, D. H., Frieman, J. A., Berlind, A. A., Blanton, M. R., Scoccimarro, R., Sheth, R. K., Strauss, M. A., Kayo, I., Suto, Y., Fukugita, M., Nakamura, O., Bahcall, N. A., Brinkmann, J., Gunn, J. E., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Loveday, J., Meiksin, A., Schlegel, D. J., Schneider, D. P., Szapudi, I., Tegmark, M., Vogeley, M. S., & York, D. G. 2005, ApJ, 630, 1

Zheng, Z. 2004, ApJ, 610, 61

Zheng, Z., Berlind, A. A., Weinberg, D. H., Benson, A. J., Baugh, C. M., Cole, S., Davé, R., Frenk, C. S., Katz, N., & Lacey, C. G. 2005, ApJ, 633, 791

Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760