Models for the clustering of far-infrared and submillimetre selected galaxies

Jo Short* and Peter Coles
School of Physics & Astronomy, Cardiff University, 5 The Parade, Cardiff CF24 3AA

Accepted 2011 January 7. Received 2010 December 10; in original form 2010 November 1

ABSTRACT
We discuss and compare two alternative models for the two-point angular correlation function of galaxies detected through the submillimetre emission using the Herschel Space Observatory. The first, now-standard Halo Model, which represents the angular correlations as arising from one-halo and two-halo contributions, is flexible but complex and rather unwieldy. The second model is based on a much simpler approach: we incorporate a fitting function method to estimate the matter correlation function with approximate model of the bias inferred from the estimated redshift distribution to find the galaxy angular correlation function. We find that both models give a good account of the shape of the correlation functions obtained from published preliminary studies of the Herschel Multi-tiered Extragalactic Survey (HerMES) and the Herschel-Astrophysical Terahertz Large Area Survey (H-ATLAS) performed using Herschel, and yield consistent estimates of the minimum halo mass within which the submillimetre galaxies must reside. We note also that both models predict an inflection in the correlation function at intermediate angular scales, so the presence of the feature in the measured correlation function does not unambiguously indicate the presence of intra-halo correlations. The primary barrier to more detailed interpretation of these clustering measurements lies in the substantial uncertainty surrounding the redshift distribution of the sources.

Key words: cosmology: large-scale structure of the Universe – submillimetre: galaxies.

1 INTRODUCTION

Current theoretical models predict that galaxies form and evolve in cold dark matter (CDM) haloes. Galaxies consequently tend to trace the distribution of mass, although the manner in which they do this may be biased (Kaiser 1984; Bardeen et al. 1986; Coles 1993; Mo & White 1996). In principle, therefore, once the bias is allowed for, it is possible to use measurements of the clustering of galaxies to determine the clustering properties of the dark matter, especially if measurements can be made as a function of redshift. Because clustering evolution is sensitive to the parameters underlying the background cosmological model, such observations can provide another (independent) test of the concordance cosmological model; see e.g. Coles (2005). In addition, clustering observations can be used to constrain properties of the galaxies themselves, such as the minimum halo mass within which they reside, which may yield clues about the processes of galaxy formation and evolution.

A steadily increasing number of surveys of large-scale galaxy clustering are now becoming available. In the optical wavebands, there are projects such as the UKIDSS Ultra Deep Survey (Hartley et al. 2010) and the Sloan Digital Sky Survey (SDSS) Redshift Survey (Connolly et al. 2010; Ross & Brunner 2009) which are being used to extract information on clustering as a function of redshift. The Herschel Space Observatory was launched in 2009 and is the only space observatory to cover a spectral range from the far-infrared to submillimetre (submm), and therefore provides a new and unique window through which to study high-redshift galaxy clustering. Two surveys of particular interest to this article, Herschel Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2010) and H-ATLAS (Eales et al. 2010), have already released angular correlation results (Cooray et al. 2010; Maddox et al. 2010) from their science demonstration phase (SDP) which we will discuss further later, so it is timely to raise the issue of modelling submm galaxy clustering.

The so-called Halo Model has been used extensively over the past few years in modelling galaxy clustering in a variety of contexts. The Halo Model combines approximations of the dark matter profile within individual haloes, the mass function and bias models to estimate the correlation function for given cosmological parameters and characteristic halo masses. It has been shown to provide accurate and reliable predictions of clustering measurements, at the price of a certain degree of complexity and modelling freedom. The main focus of this Letter is a comparison of the Halo Model and a fitting function method which was initially introduced by Hamilton et al. (1991), and subsequently developed by Peacock &
Dodds (1994), Jain, Mo & White (1995) and especially Peacock & Dodds (1996), to estimate the matter correlation function. Work by Matarrese et al. (1997) followed by Moscardini et al. (1998) and Coles et al. (1998) showed how to incorporate this idea into a technique for providing detailed predictions of angular correlation in high-redshift galaxy surveys. In this Letter we compare the predictions of this much simpler approach with results from the Halo Model.

The Letter is organized as follows. In the next section, we outline the methodology for the fitting function method. Following that in Section 3, the models are compared to recent Herschel data and the best-fitting values of the free parameters are extracted and compared. In Section 4, we conclude with a summary.

2 MODELS OF GALAXY CLUSTERING

A model for the angular correlation function of galaxies basically involves a model for the three-dimensional clustering of galaxies as a function of epoch (i.e. redshift \( z \)) combined with a machinery (derived from a model of the background cosmology) for projecting this information down the observer’s light cone.

2.1 Angular correlation function

The observed angular correlation function, \( \omega_{ab}(\theta) \), is calculated from the spatial power spectrum, \( \Delta^2(k, z) \), using

\[
\omega_{ab}(\theta) = \int \frac{d^2 \xi(r, \theta, z)}{d u d \theta},
\]

where \( N(z) \) is the normalized distribution of galaxies over redshift, \( G(z) = (dz/dx)^{-1} \) where \( x(z) \) is the comoving distance to redshift \( z \), the separation \( r(u, \theta, z) = \sqrt{u^2 + x^2(z) \theta^2} \) Mpc h\(^{-1} \) and \( \xi(r, z) \) is the two-point (galaxy–galaxy) correlation function, where

\[
\xi(r, z) = \int_0^\infty \Delta^2(k, z) \frac{\sin(kr)}{kr} \frac{dk}{k}.
\]

The galaxy power spectrum can be approximated using different methods such as those described below.

2.2 Power spectrum: Halo Model

The Halo Model estimate for the galaxy power spectrum represents it as arising from two distinct components:

\[
\Delta^2(k, z) = \Delta^2_{\text{lin}}(k, z) + \Delta^2_{nl}(k, z).
\]

The first term represents contributions from galaxies residing within the same dark matter halo, while the second is generated by contributions from galaxies in separate haloes. The one-halo term models small-scale highly non-linear clustering using a recipe that involves a number of different components, including the dark matter profile of the haloes, the mass function of haloes and the halo occupation distribution (HOD); the two-halo term depends on the underlying matter power-spectrum as well as the bias of galaxy clustering relative to mass clustering. In this Letter, the dark matter halo density profile is approximated using the NFW profile suggested by Navarro, Frenk & White (1997). The mass function is effectively the number density of haloes at a given mass and for this we use the approximation found by Sheth & Tormen (1999). The HOD is a simple way of relating the distribution of galaxies from the dark matter distribution, by estimating how many galaxies are expected within a dark matter halo of given mass. This is split into two estimates – the number of central galaxies (which would either be 1 or 0) and the number of satellite galaxies; a summary of the procedure can be found in Ross & Brunner (2009). There are extensions to the model (e.g. Gil-Marín, Jimenez & Verde 2010) but we just consider the basic version here. Cooray & Sheth (2002) provide a detailed review of the Halo Model, so we will not go into any further details.

2.3 Power spectrum: fitting function

The linear matter power spectrum can be calculated as

\[
\Delta^2_{\text{lin}}(k, z) = \frac{P_{knl}(k, z)}{2\pi^2},
\]

where the transfer function, \( T(k, z) \), and the growing mode of linear perturbations, \( D_z \), can be approximated as described by Eisenstein & Hu (1999).

The non-linear power spectrum (\( \Delta^2_{nl} \)) is known for extreme values of the wavenumber \( k \) (i.e. at small \( k \), \( \Delta^2_{nl} \approx \Delta^2_{\text{lin}} \), and at large \( k \), \( \Delta^2_{nl} \approx \Delta^2_{\text{org}} \)), the latter derived using the so-called stable clustering ansatz.

In the light of these two asymptotic regimes, a fitting function was proposed to extrapolate \( \Delta^2_{nl} \) at all other \( k \) directly from the linear counterpart:

\[
\Delta^2_{nl}(k, z) = \mathcal{F} \left[ \Delta^2_{\text{lin}}(k, z) \right],
\]

where \( k_{nl} = k_{\text{lin}} \left( 1 + \Delta^2_{nl}(k_{\text{lin}}, z) \right)^{1/3} \). Peacock & Dodds (1996) found a fit for this for low-density universes with CDM and a cosmological constant:

\[
\mathcal{F}(x) = x \left[ 1 + B \beta x + \frac{Ax^3}{1 + (Ax)^{3\beta}} \right]^{1/\beta},
\]

where

\[
A = 0.482(1 + n_{\text{eff}}/3)^{-0.947},
\]

\[
B = 0.226(1 + n_{\text{eff}}/3)^{-1.778},
\]

\[
\alpha = 3.101(1 + n_{\text{eff}}/3)^{-0.244},
\]

\[
\beta = 0.862(1 + n_{\text{eff}}/3)^{-0.287},
\]

\[
V = 11.55(1 + n_{\text{eff}}/3)^{-0.423},
\]

and

\[
n_{\text{eff}} = \frac{\ln \left( \frac{2\pi^2}{k_{\text{lin}}^2} \cdot \Delta^2_{\text{lin}}(k_{\text{lin}}) \right)}{\ln k_{nl}}.
\]

This method can be used to approximate the non-linear dark matter power spectrum (\( \Delta^2_{\text{org}} \)). Under very general conditions, proved by Coles (1993), this can be used to approximate the non-linear galaxy power spectrum, \( \Delta^2_{g} \), via a simple linear relationship of the form:

\[
\Delta^2_g(k, z) = h^2_{\text{eff}} \Delta^2_{\text{org}}(k, z).
\]

We use an approximation for the bias proposed by Moscardini et al. (1998), which is based on work by Mo & White (1996) and suggested the following relation between the bias and mass density fluctuations:

\[
b(z, m) = 1 + \frac{1}{\delta_c} \left[ \frac{\delta^2(z, m)}{\sigma^2_{\text{lin}}(z, m)} - 1 \right],
\]

where \( \delta_c \) is the critical linear density required for collapse and \( \sigma^2_{\text{lin}} \) is the linear variance in the smoothed density field. The effective bias can then be derived as

\[
b_{\text{eff}}(z) = \int_{\ln(M_{\text{min}})}^\infty n(z, m) \, dm,
\]

where \( n \) is the normalized mass function, given by Sheth & Tormen (1999). Note for simplicity we assume here that \( M_{\text{min}} \) is constant over redshift; given the larger redshift ranges a minimum mass varying with redshift may provide an improved fit but this approximation is sufficient for the comparison of models in this work.
Apart from the cosmological parameters, which we constrain to match other observations in this analysis, this fitting function method has just one free parameter (the minimum halo mass \(M_{\text{min}}\)) whereas the Halo Model has three: the minimum halo mass, the average mass of a halo with one satellite galaxy (\(M_{\text{sat}}\)) and the slope of the first moment of the satellite galaxy HOD (\(\alpha_s\)).

### 3 FITS TO THE DATA

In this section, the models described above are compared with the observed angular correlation from recent Herschel surveys and are used to put estimates on the minimum halo mass associated with these sources. The models we choose use current cosmological parameters from the Wilkinson Microwave Anisotropy Probe (Bennett et al. 2003; Komatsu et al. 2009; \(\Omega_m = 0.274\), \(\Omega_\Lambda = 0.726\), \(h = 0.705\), \(n = 0.96\), \(\sigma_8 = 0.812\)).

The redshift distribution of the sources plays a vital role in modelling the observed angular correlations. Fig. 1 shows the normalized redshift distributions, \(N(z)\), which were used in these calculations for both the HerMES and H-ATLAS data sets. The estimates for the redshift distribution for the HerMES data sets are consistent with those used by Cooray et al. (2010) in their analysis. Several estimates of the redshift distribution of objects in the H-ATLAS survey have been made. First, Amblard et al. (2010) used colour–colour diagrams to estimate the redshift distribution; however this method includes only a subset of the sources used in the angular correlation analysis and is known to be slightly biased towards higher redshift objects. Secondly, Dye et al. (2010) used optical counterparts of submm sources to estimate the redshift distribution; but again this method only uses a subset of all the objects and is biased towards lower redshift objects. So far, the best estimates of the redshift distribution, in that they use all the available sources, are those by Eales et al. (in preparation) who use spectral energy distribution fitting to find a best-fitting redshift distribution. We use the estimates by Eales et al. (in preparation) in this analysis, but the values for \(N(z)\) used here are for illustrative purposes only and should not be regarded as definitive given the preliminary state of the data. We have tried a variety of alternative models and find that, for reasonable choices, the results for \(\omega_{\text{obs}}\) are not especially sensitive to a particular mean redshift in \(N(z)\); this is probably because of the relatively slow evolution of the power spectrum at low redshift in the concordance cosmology. On the other hand, the results are sensitive to the width of the distribution in \(z\); the wider the distribution over \(z\) the lower the amplitude of the angular correlation function. This is consistent with what you would expect if you consider that the clustering signal is more concentrated in narrower redshift bands.

The HerMES (Oliver et al. 2010) will cover approximately 70 deg\(^2\) of the sky in three different wavelength bands. The Lockman–SWIRE field, which is one of the shallower fields, was covered by the initial science run and results of the clustering have recently been published in Cooray et al. (2010) along with a Halo Model analysis. Here we show just three of the available results, in Fig. 2, to compare with the theoretical models we have discussed.

![Figure 2](https://academic.oup.com/mnrasl/article-abstract/412/1/L93/1008631/1008631)

**Figure 2.** These plots show the angular correlation function \(\omega_{\text{obs}}(\theta)\) measured in the HerMES survey in the Lockman–SWIRE field. Sources were divided into three sets by flux density: \(S_{350} > 35\) mJy (top); \(S_{350}/S_{250} > 0.85\) (middle) and \(S_{350}/S_{250} < 0.85\) (bottom). Data are compared to four different theoretical models: the non-linear approximation using the fitting function (solid line), a power law (short dashed line), the Halo Model (dotted line) and the linear approximation (long dashed line). The Halo and fitting function methods fit the data well, as does the linear approximation. The power law overestimates at large scales and underestimates at mid-angular scales.

**Figure 1.** The normalized distribution of galaxies as a function of redshift for data sets from HerMES and H-ATLAS. HerMES estimates are from Cooray et al. (2010) and are shown in thin lines: \(S_{250} > 35\) mJy is a solid line, \(S_{350}/S_{250} > 0.85\) is the dotted line and \(S_{350}/S_{250} < 0.85\) is the dashed line. H-ATLAS estimates are from Eales et al. (in preparation) and are shown in thick lines: \(S_{250} > 33\) mJy is the dotted line and \(S_{350} > 35\) mJy is the solid line.

© 2011 The Authors, MNRAS 412, L93–L97

Monthly Notices of the Royal Astronomical Society © 2011 RAS
above: $S_{250} > 35$ mJy, $S_{350}/S_{250} > 0.85$ and $S_{350}/S_{250} < 0.85$. From the observations, we see that the angular correlation has similar amplitudes in each of the three data sets; the middle panel has the highest amplitude and this most likely because the redshift distribution of the component sources is the most narrow of the three data sets making the clustering signal more concentrated and appear stronger.

In Fig. 2, four models are compared to the observations: the short dashed line shows a simple power-law fit, the dotted line shows the fit from the Halo Model, the solid line shows the non-linear approximation using the fitting function approach, and the long dashed line shows just the linear spectrum. In the top two plots, the power-law fits the data well at low angular scales, but underestimates at mid-scales and overestimates at higher scales. This is simply the case for the last plot, but it also overestimates clustering on small scales. The Halo Model provides a better fit; the low and intermediate angular scales both fit well, although in the top two data sets the model is slightly overestimating at large angular scales. The fitting function approximation also fits the data well. The fit at intermediate scales is perhaps not quite as good as the Halo Model, but then this approach has fewer free parameters; the differences in the behaviour at large angular scales is a consequence of slightly different best-fitting values for the bias parameter. However, an important point to note here is that the linear angular correlation function also provides a pretty good fit to each of the data sets. By comparing the linear and non-linear models, we see that the transition between linear and non-linear regimes does not become evident until quite small scales (~1 arcmin). The limited resolution of the Herschel telescope makes it difficult to probe the clustering regime on scales much smaller than this.

In these models, the amplitude of large-scale clustering depends on the bias and that, in turns, boils down to the minimum halo mass through equation (9). We can therefore use the measured clustering amplitude to get a rough estimate of the mass of haloes hosting these galaxies. The values of the minimum halo mass corresponding to the theoretical models in Fig. 2 are shown in Table 1. We see that for each of the data sets, the minimum halo masses found from both the fitting function and Halo Models are in agreement (within the errors) and suggest a value of $M_{\text{min}} \sim 10^{12.2\pm0.4} M_\odot h^{-1}$. The results using our Halo Model are consistent with those found in the Halo Model analysis by Cooray et al. (2010).

The other results we have analysed are from the Herschel-ATLAS survey (H-ATLAS; Eales et al. 2010) which will cover 550 deg$^2$ of the sky, in five different wavebands covering the far-infrared to submm. The SDP of the survey has just been completed (which covers roughly 1/30 of the final H-ATLAS data set) and Maddox et al. (2010) have released measurements of the angular correlation function of the galaxies observed. Fig. 3 shows different flux cuts for $S_{250} > 33$ mJy (top) and $S_{350} > 36$ mJy (bottom). The results are quite noisy so none of the models fit particularly well – a detailed analysis will have to wait until the completion of the survey. This noise is thought to be due to galactic cirrus in the H-ATLAS data as the number of detections is very similar to the HerMES data sets and both surveys are similarly confusion limited. In this case, therefore, we have just plotted the fitting function results against the data. It was used to find approximate values of the corresponding minimum halo masses, which are $10^{12.5}$ and $10^{13.2} M_\odot h^{-1}$ for the examples plotted in Fig. 3, top to bottom, respectively. The scale of the $\omega_{\text{obs}}$ results for the $S_{250} > 33$ mJy flux cut is significantly smaller that those in the other examples. This is predominately due to the wide bimodal redshift distribution (van Kampen et al., in preparation; Smith et al. 2010), although it does still suggest a slightly lower minimum mass $\sim 10^{12.5} M_\odot h^{-1}$. This minimum mass is consistent with that found in the cross-correlation analysis by Guo et al. (2010).

### Table 1. Best-fitting parameters for the halo and fitting function models for the HerMES results. $M_{\text{min}}$ is the minimum halo mass for which galaxies can form, $M_{\text{halo}}$ is the average mass of a halo with one satellite galaxy and $\alpha_s$ is the slope of the first moment of the satellite galaxy HOD. All masses are in units $M_\odot h^{-1}$. We see that the $M_{\text{min}}$ found for each data set are consistent between models, given an accuracy of $10^{13.0} M_\odot h^{-1}$ on the masses.

| Data set      | $M_{\text{min}}^{DF}$ | $M_{\text{halo}}$ | $M_{\text{min}}$ | $\alpha_s$ |
|---------------|------------------------|-------------------|-----------------|------------|
| $S_{250} > 35$ mJy | $10^{13.4}$              | $10^{13.0}$     | $10^{13.0}$     | <1.2       |
| $S_{350}/S_{250} > 0.85$ | $10^{13.2}$              | $10^{13.2}$     | $10^{13.4}$     | <1.2       |
| $S_{350}/S_{250} < 0.85$ | $10^{13.4}$              | $10^{13.0}$     | $10^{13.2}$     | >1.4       |

### 4 DISCUSSION AND CONCLUSIONS

In this Letter, we have compared models of the angular correlation function to data from the SDP of Herschel. We highlight a fitting function method which provides an improved fit to the data than a power law, and similar to that of the Halo Model. It has just one free parameter, the minimum halo mass, compared to the two and three for a simple power law and the Halo Model, respectively. The halo mass is more meaningful physically as a parameter than those involved in the power-law fit, so the fitting function is a much better method than the power law for a quick-and-simple...
Clustering of FIR/sub-mm selected galaxies

analysis. Although neither as sophisticated nor as flexible as the Halo Model, it remains a useful tool that is perfectly adequate for modelling currently available data. For example, the minimum halo mass found using our fitting function model is consistent with that found using the Halo Model.

Another point of interest is that, in fact, the linear angular correlation function also provides a reasonable fit to the data. Owing to the limited resolution of the Herschel telescope, it is difficult to identify pairs of galaxies sufficiently close together on the sky to probe anything but the mildly non-linear regime. The currently available data provide some evidence of a transition between the linear and non-linear regimes, but they provide no unambiguous detection of the change of shape in $\omega_{\text{obs}}$ that the Halo Model predicts. This does not mean the Halo Model is incorrect, of course, but what it does mean is that, at least for the time being, the paraphernalia involved in modelling intra-halo correlations is rather superfluous for these objects; simpler models can yield perfectly adequate results.

Finally, we stress that the data sets to which we have applied these models are preliminary. The biggest stumbling block to a more complete analysis relates to the considerable uncertainties in the redshift distribution $N(z)$ of the sources involved. The choices we adopted for this analysis are for illustration only, so the results should not be regarded as definitive. Further data, especially ancillary data providing spectroscopic redshifts, will be needed before the precise nature of these galaxies can be determined.

ACKNOWLEDGMENTS

JS receives funding from an STFC studentship. For the purposes of this work, PC is supported by STFC Rolling Grant ST/H001530/1. We thank Steve Eales, Loretta Dunne and Steve Maddox for helpful discussions, and are grateful to them for allowing us to use data from Eales et al. (in preparation) in advance of publication.

REFERENCES

Amblard A. et al., 2010, A&A, 518, L9
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bennett C. L. et al., 2003, ApJS, 148, 1
Coles P., 1993, MNRAS, 262, 1065
Coles P., 2005, Nat, 433, 248
Coles P., Lucchin F., Matarrese S., Moscardini L., 1998, MNRAS, 300, 183
Connolly A. et al., 2010, ApJ, 579, 42
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Cooray A. et al., 2010, A&A, 518, L22
Dye S. et al., 2010, A&A, 518, L10
Eales S. et al., 2010, PASP, 122, 499
Eisenstein D. J., Hu W., 1999, ApJ, 511, 5
Gil-Marín H., Jimenez R., Verde L., 2010, MNRAS, submitted (arXiv:1008.4583)
Guo Q. et al., 2010, MNRAS, submitted (arXiv:1011.3048)
Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, ApJ, 374, L1
Hartley W. et al., 2010, MNRAS, 407, 1212
Jain B., Mo H. J., White S. D. M., 1995, MNRAS, 276, L25
Kaiser N., 1984, ApJ, 284, L9
Komatsu E. et al., 2009, ApJS, 180, 330
Maddox S. et al., 2010, A&A, 518, L11
Matarrese S., Coles P., Lucchin F., Moscardini L., 1997, MNRAS, 286, 115
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Moscardini L., Coles P., Lucchin F., Matarrese S., 1998, MNRAS, 299, 95
Navarro J. F., Frenk C. F., White S. D. M., 1997, ApJ, 490, 493
Oliver S. J. et al., 2010, A&A, 518, L21
Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020
Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
Ross A. J., Brunner R. J., 2009, MNRAS, 399, 878
Sheth R. K., Tormen G., 1996, MNRAS, 308, 119
Smith D. J. B. et al., 2010, MNRAS, submitted (arXiv:1007.5260)

This paper has been typeset from a TeX/LaTeX file prepared by the author.